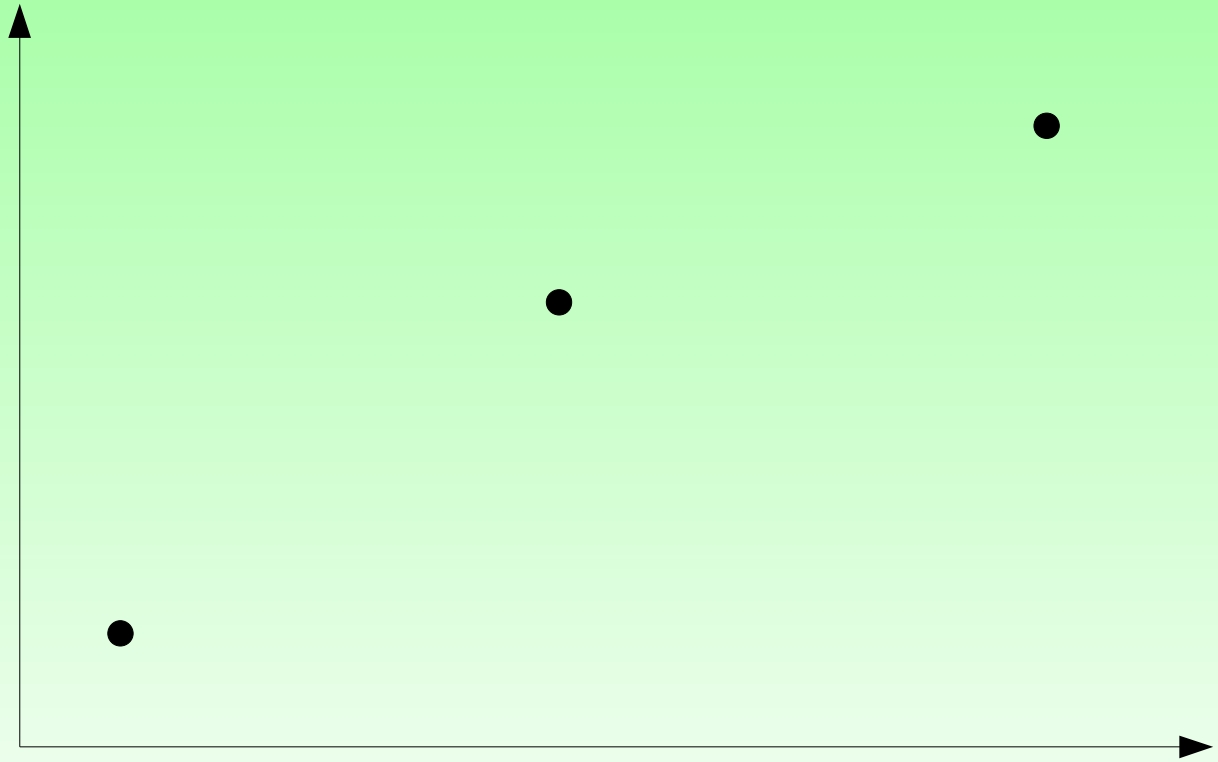
Kakusan3を用いた分子進化モデルの選択

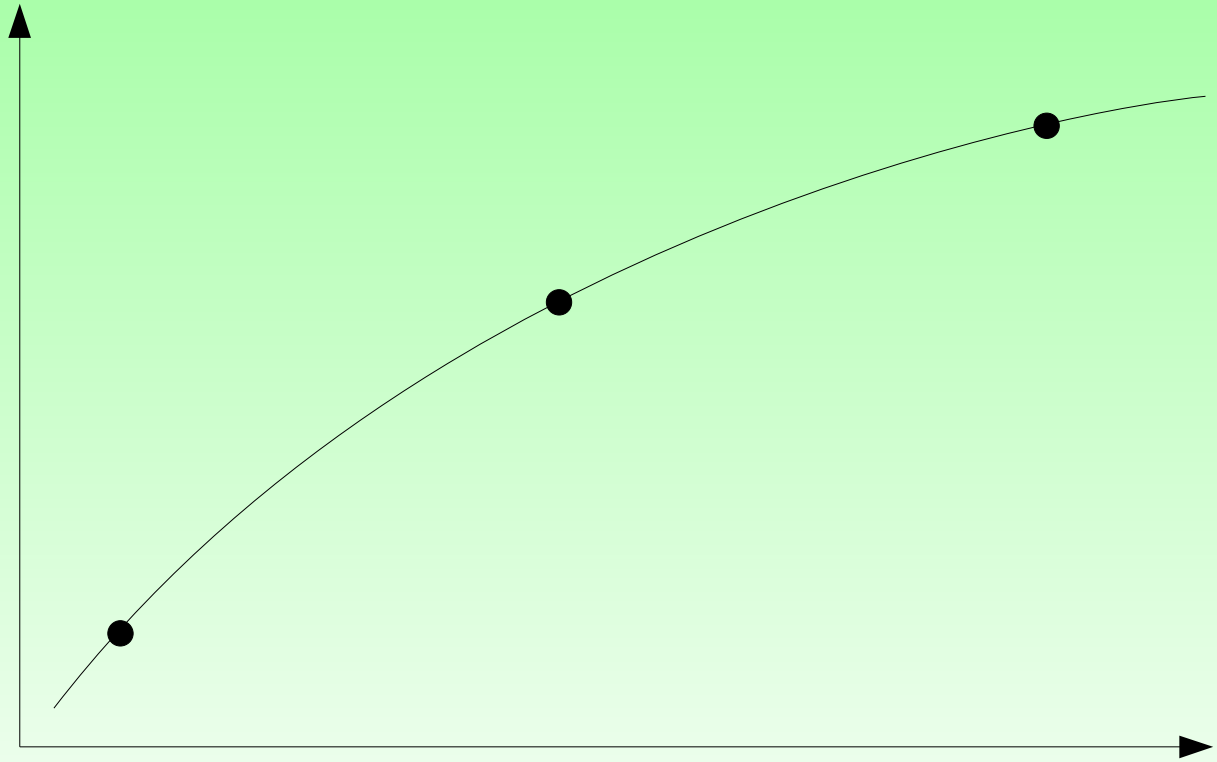
モデル選択って何?

モデル選択って何?

- 3点のデータに対する回帰分析を考える

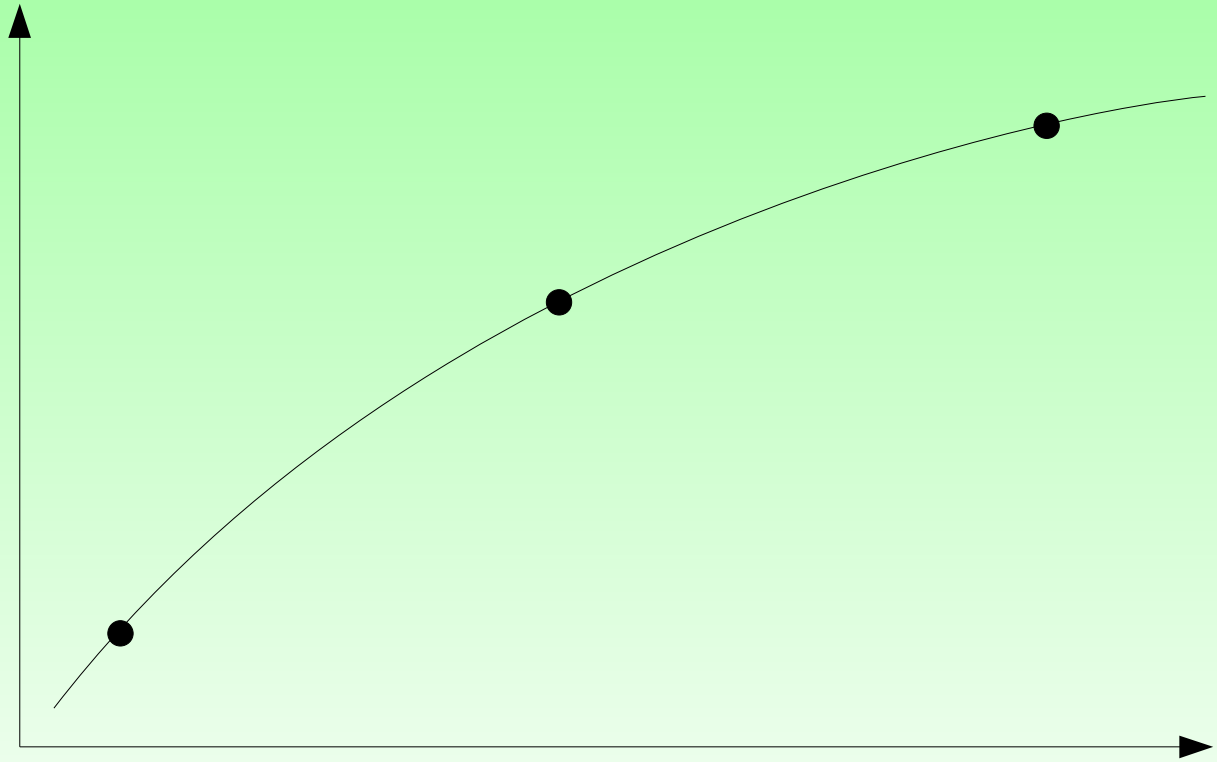


モデル選択って何?



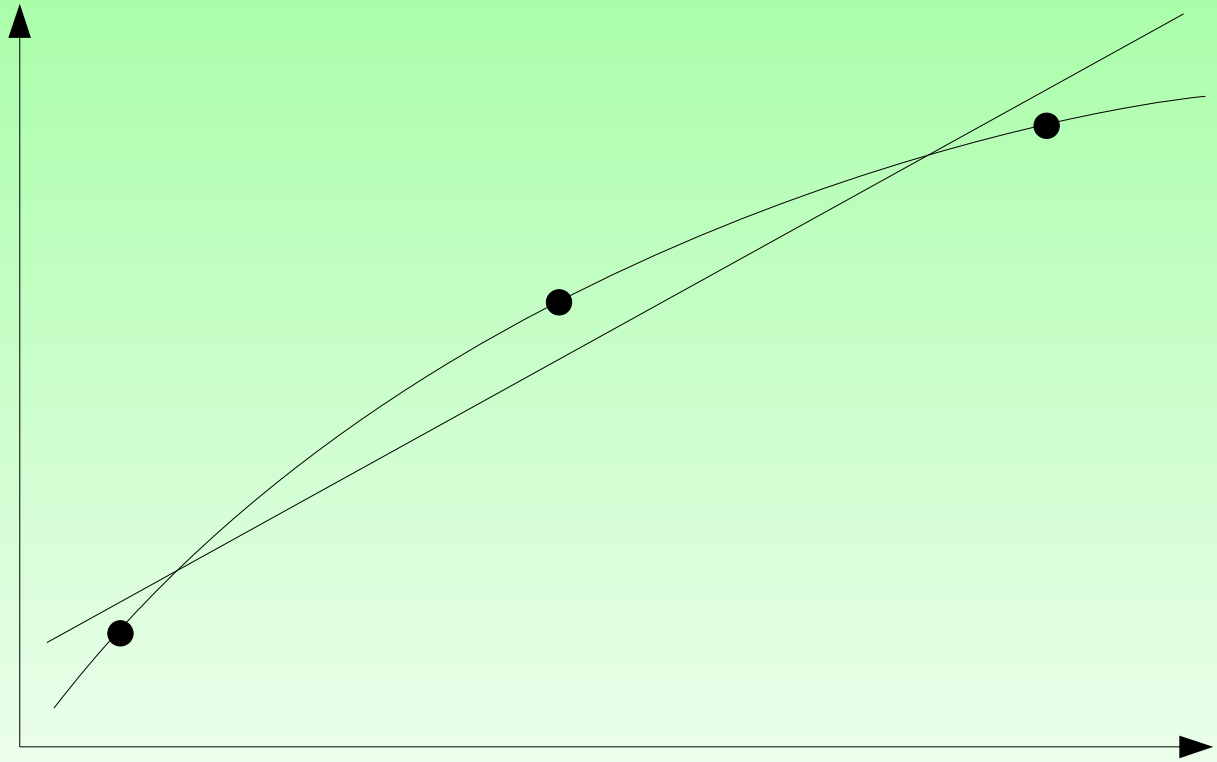
- 3点のデータに対する回帰分析を考える
- 2次曲線を当てはめれば全点を通せる＝尤度最大

モデル選択って何?



- 3点のデータに対する回帰分析を考える
- 2次曲線を当てはめれば全点を通せる＝尤度最大
- ただしデータから推定するパラメータは3つ

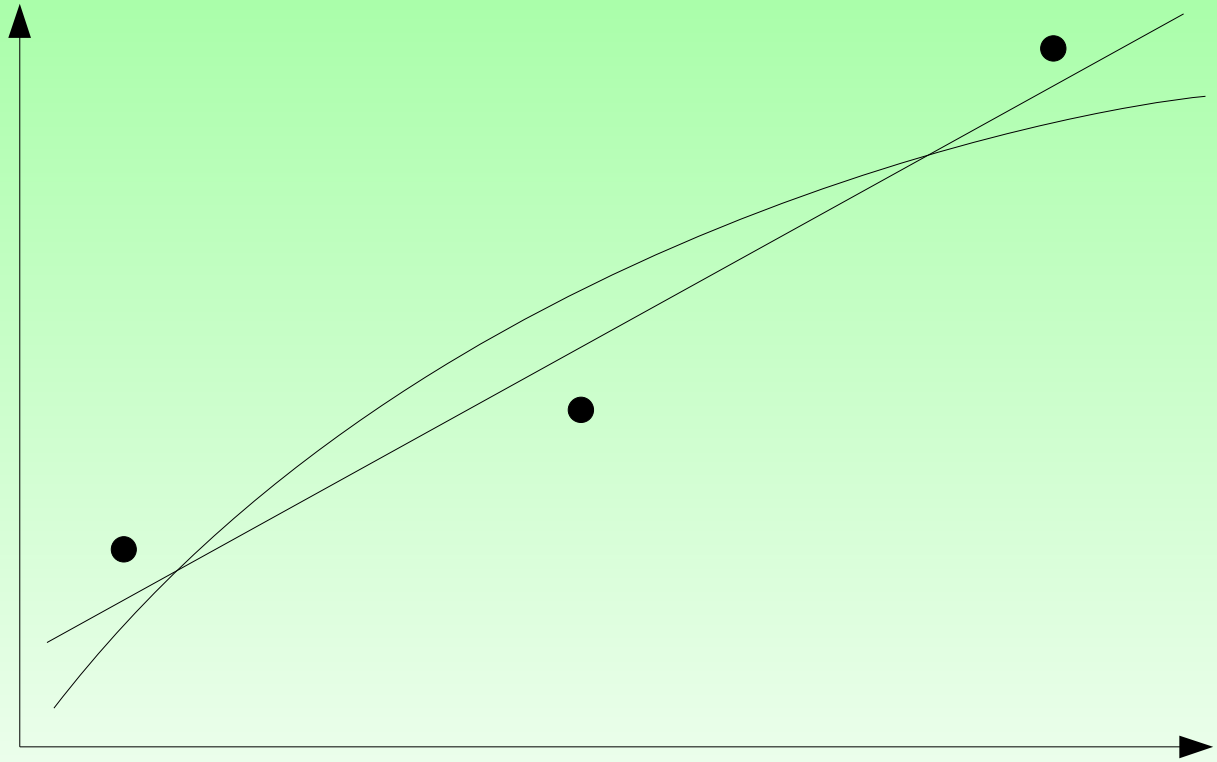
モデル選択って何?



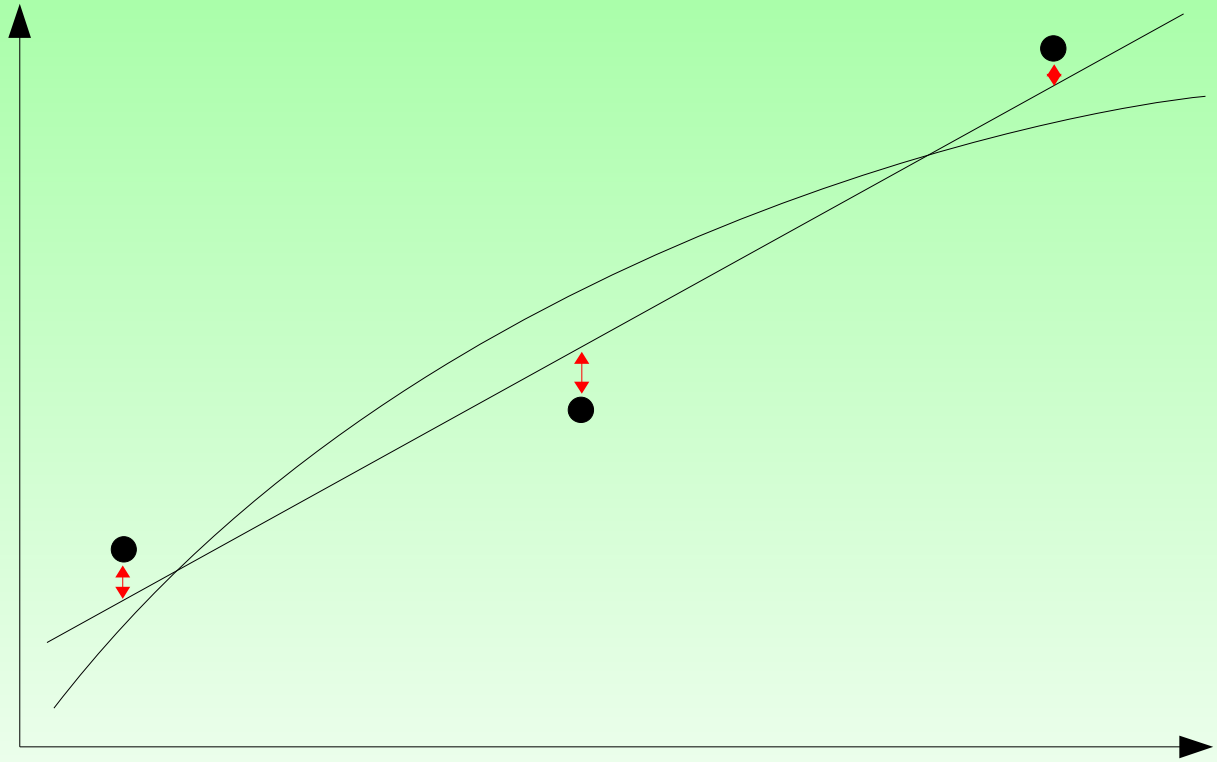
- 3点のデータに対する回帰分析を考える
- 2次曲線を当てはめれば全点を通せる＝尤度最大
- ただしデータから推定するパラメータは3つ
- 1次直線を当てはめれば尤度は低下するがパラメータ数は減少する

モデル選択って何?

- データを取り直した場合を考える

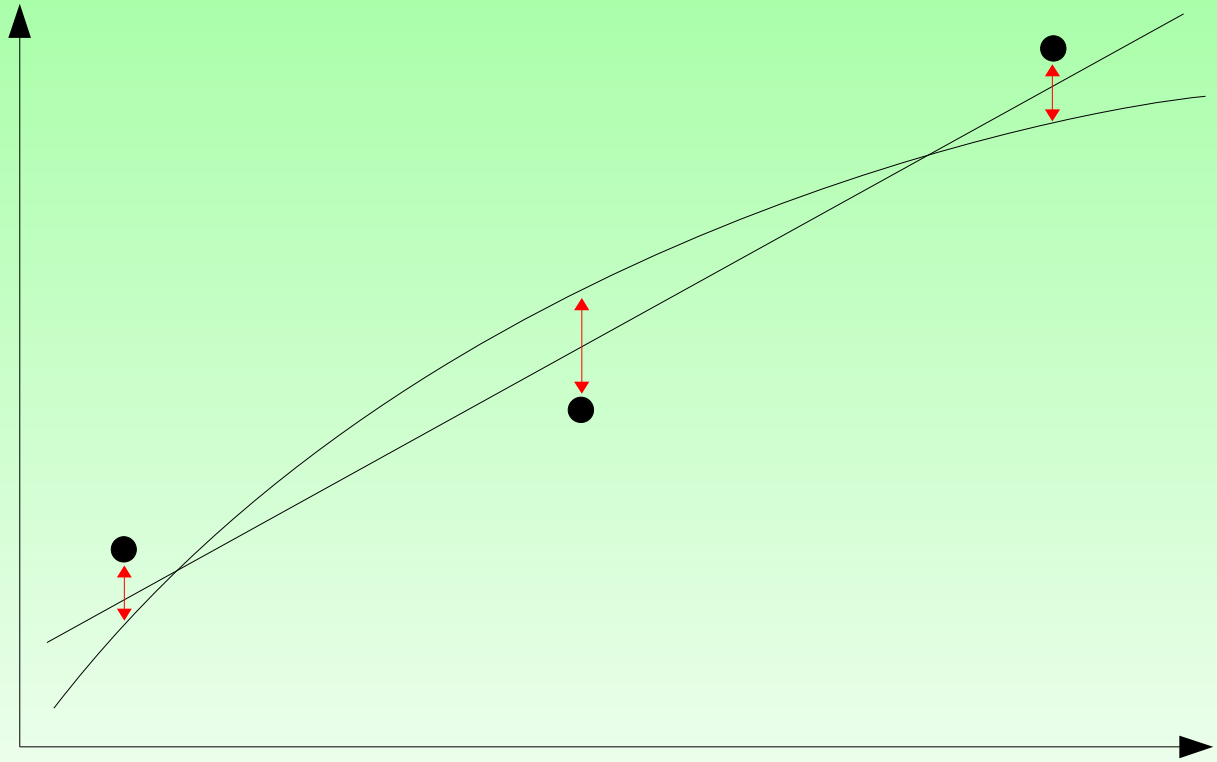


モデル選択って何?



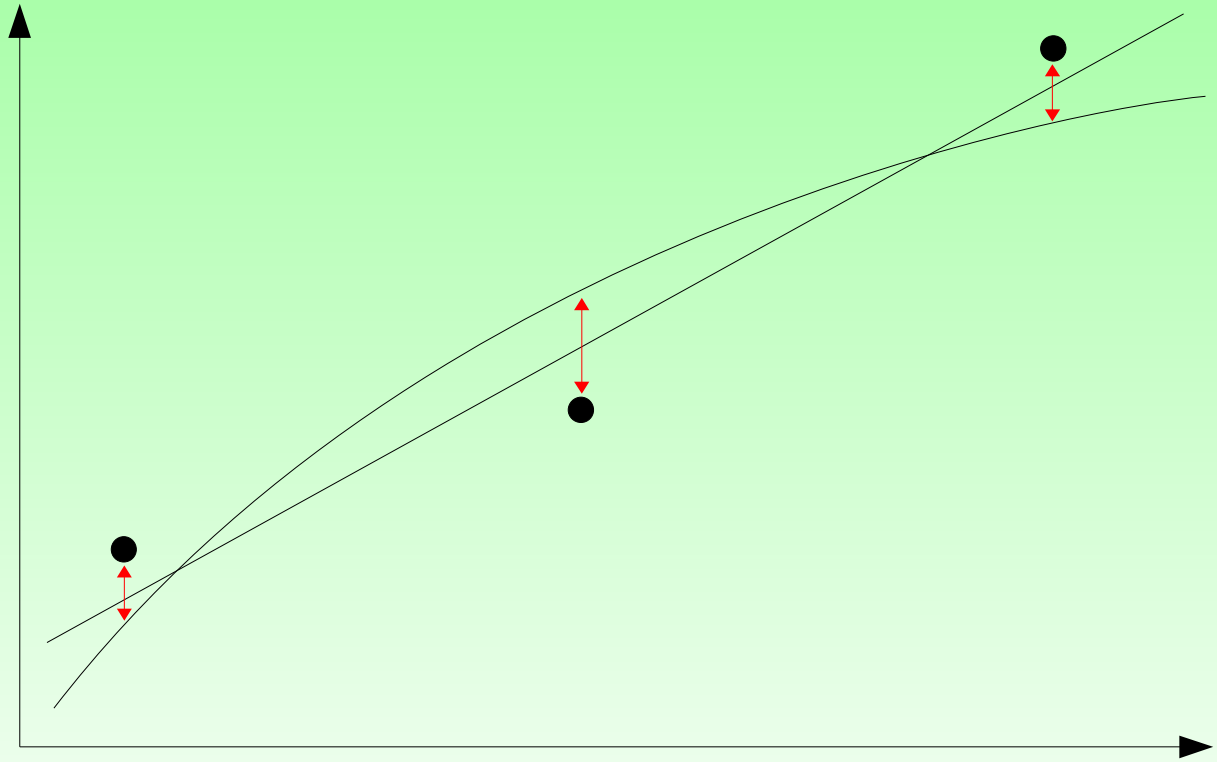
- データを取り直した場合を考える
- 1次直線は当てはまりが大きく変化しない

モデル選択って何?



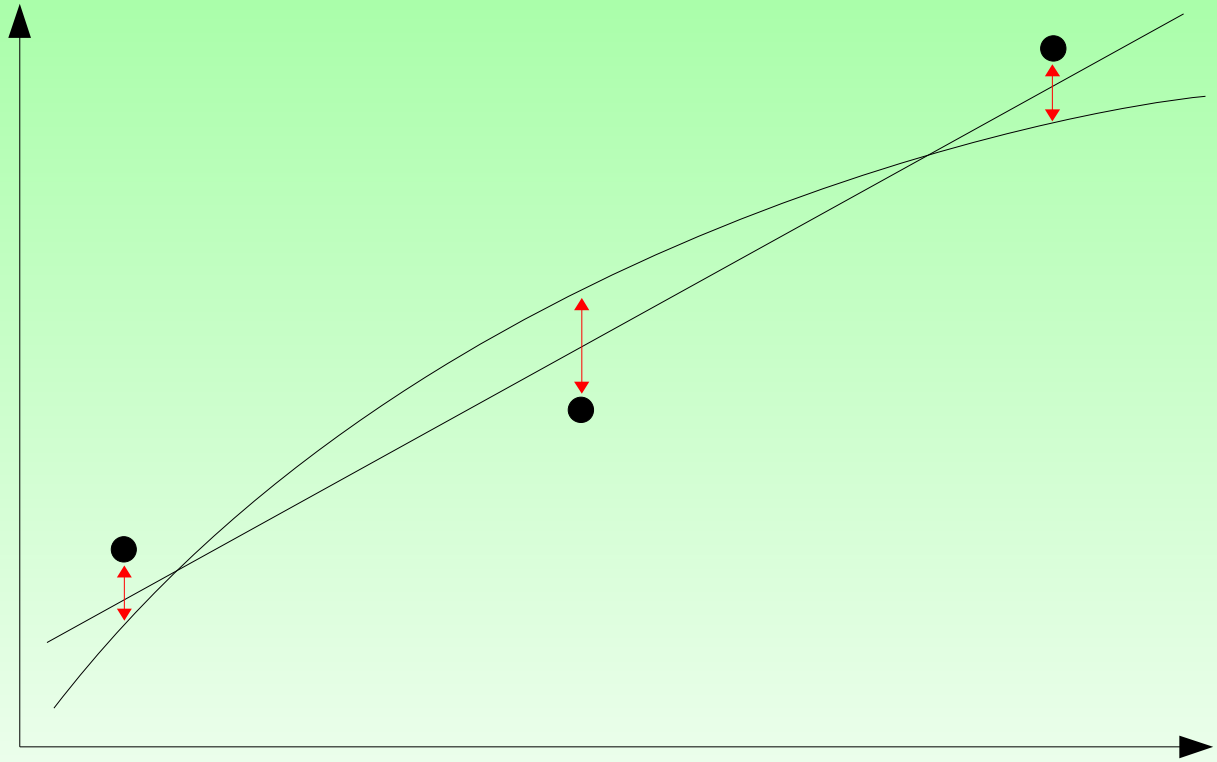
- データを取り直した場合を考える
- 1次直線は当てはまりが大きく変化しない
- 2次曲線の当てはまりは大きく低下する

モデル選択って何?



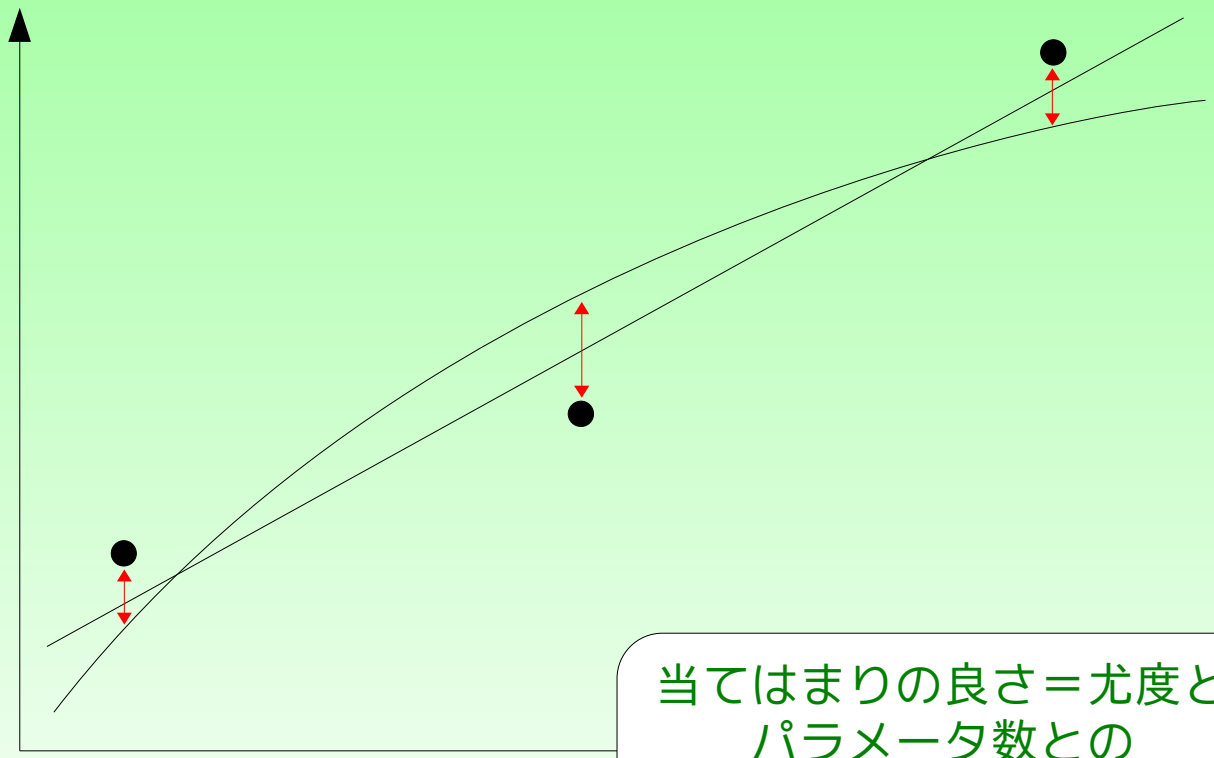
- データを取り直した場合を考える
- 1次直線は当てはまりが大きく変化しない
- 2次曲線の当てはまりは大きく低下する
- 1次直線の方が「母集団のパラメータを推定する能力」が高い

モデル選択って何?



- データを取り直した場合を考える
- 1次直線は当てはまりが大きく変化しない
- 2次曲線の当てはまりは大きく低下する
- 1次直線の方が「母集団のパラメータを推定する能力」が高い
- 2次曲線はデータの「ノイズ」にまでフィットすることで「推定能力」が低下している

モデル選択って何?



当てはまりの良さ＝尤度と
パラメータ数との
バランスを取ることで
推定能力を最大化すべき

- データを取り直した場合を考える
- 1次直線は当てはまりが大きく変化しない
- 2次曲線の当てはまりは大きく低下する
- 1次直線の方が「母集団のパラメータを推定する能力」が高い
- 2次曲線はデータの「ノイズ」にまでフィットすることで「推定能力」が低下している

赤池情報量規準

赤池情報量規準

$$\mathrm{AIC} = -2\ln L + 2K$$

(Lは最大化尤度・Kはパラメータ数)

赤池情報量規準

$$\mathrm{AIC} = -2\ln L + 2K$$

(Lは最大化尤度・Kはパラメータ数)

AIC最小となるモデル(回帰式)が
ベストな＝推定能力が最大化されるモデル

AICcとかBICなどもしばしば使われます

分子系統推定で用いられるデータ

site	12345	...	....M
OTU1	TGTTT	...	TTTTC
OTU2	AGTAC	...	TTTTC
OTU3	AGTAT	...	TTGTC
.	.		.
.	.		.
.	.		.
OTUN	AGTAT	...	ATTTC

分子系統推定で用いられるデータ

site	12345	...	....M
OTU1	TGTTT	...	TTTTC
OTU2	AGTAC	...	TTTTC
OTU3	AGTAT	...	TTGTC
.	.		.
.	.		.
.	.		.
OTUN	AGTAT	...	ATTTC

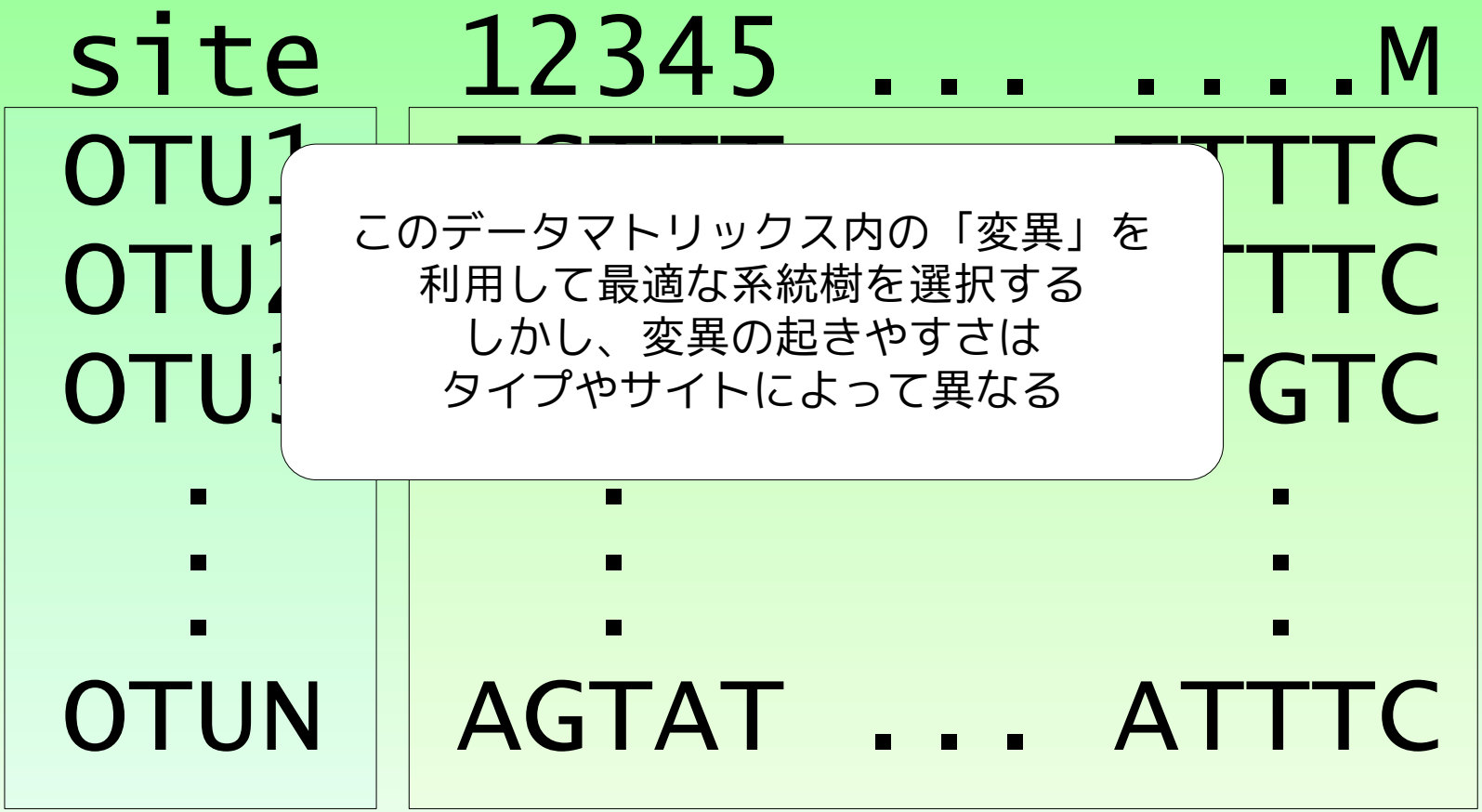
OTU名

分子系統推定で用いられるデータ

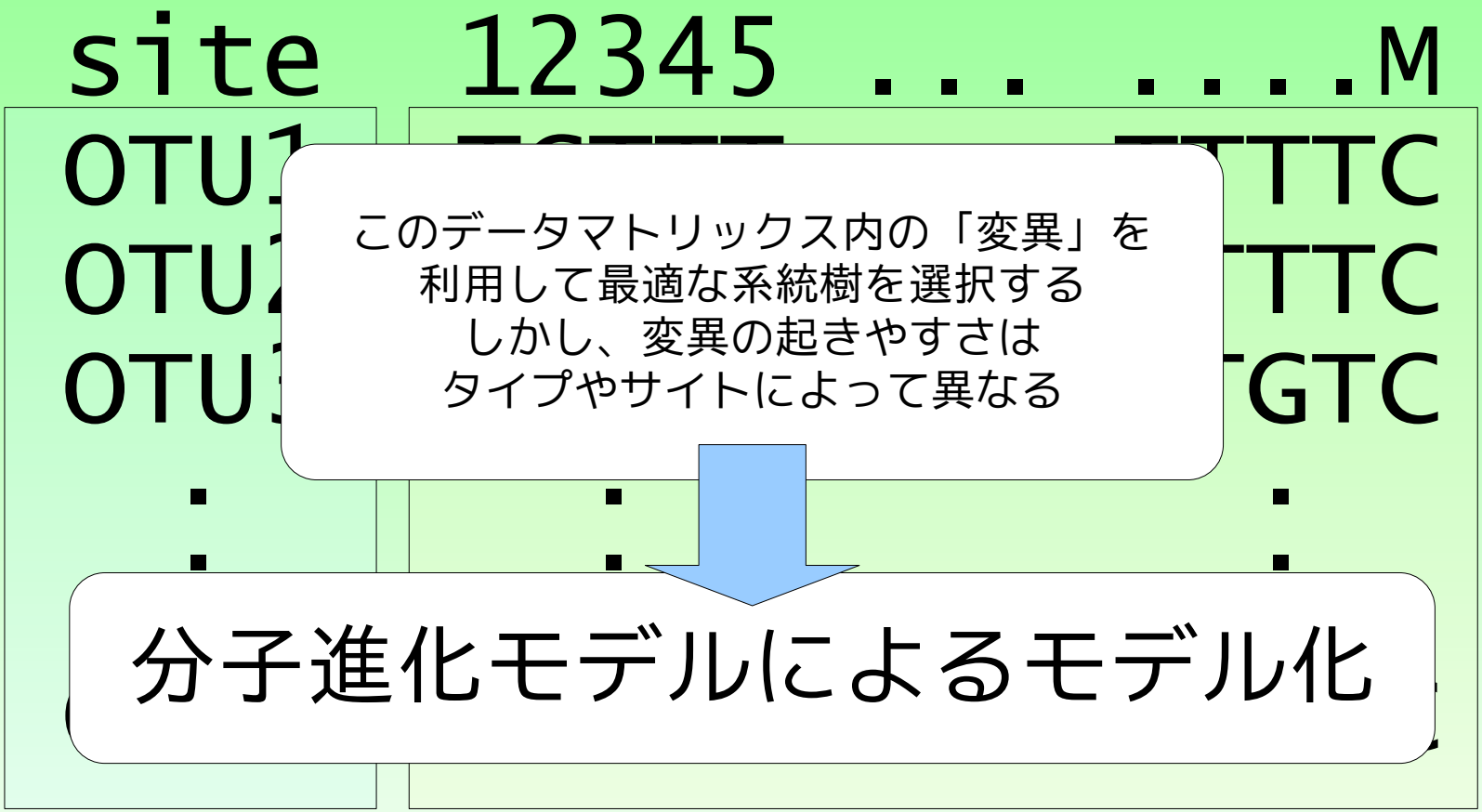
site	12345 ... .M
OTU1	TGTTT ... TTTTC
OTU2	AGTAC ... TTTTC
OTU3	AGTAT ... TTGTC
.	. .
.	. .
.	. .
OTUN	AGTAT ... ATTTC

OTU名　　アライメント済みデータ配列

分子系統推定で用いられるデータ



分子系統推定で用いられるデータ



分子進化モデルがモデル化しているもの

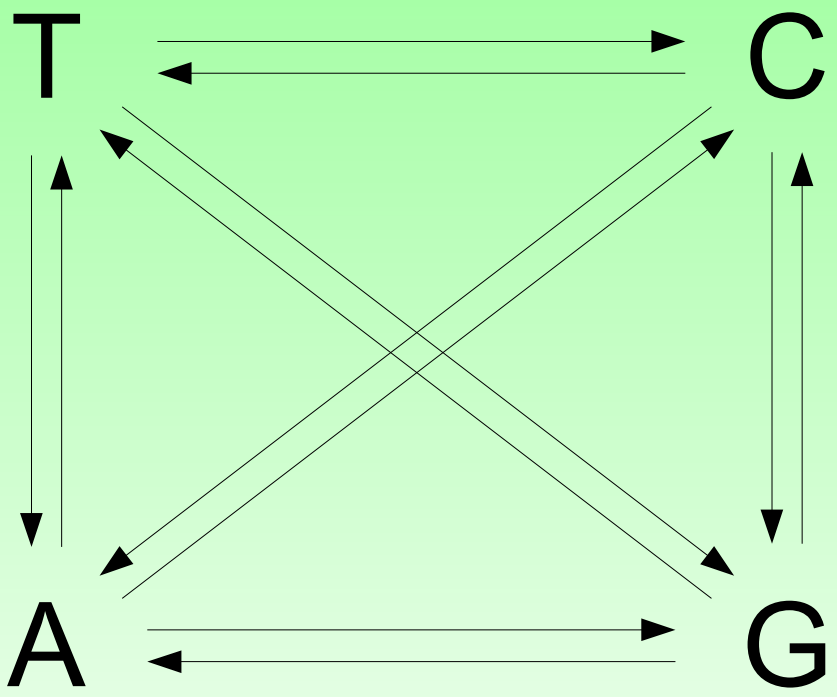
分子進化モデルがモデル化しているもの

- サイト内形質状態間の置換速度不均質性
  - どのタイプの変異が起きやすいかを表す

分子進化モデルがモデル化しているもの

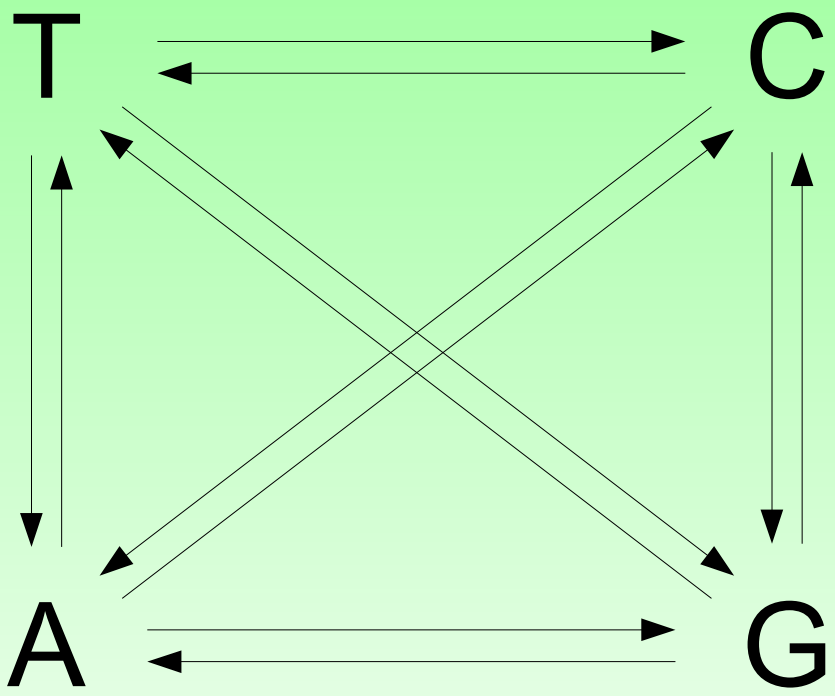
- サイト内形質状態間の置換速度不均質性
  - どのタイプの変異が起きやすいかを表す
- サイト間の置換速度不均質性
  - どこで変異が起きやすいかを表す

DNAの塩基置換パターン

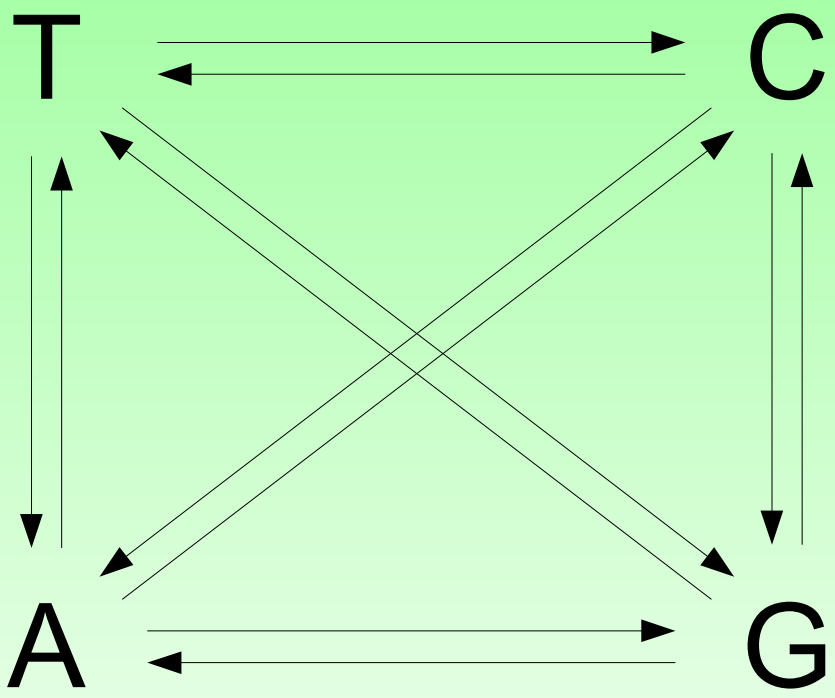


DNAの塩基置換パターン

- 塩基の置換パターンは12通り



DNAの塩基置換パターン



- 塩基の置換パターンは12通り
- それぞれの起きる確率の違い＝形質状態間の推移速度の不均質性をモデル化する

塩基置換確率行列

- r_{XY}は塩基Xから塩基Yへの置換確率

From\To	A	C	G	T
A	-	r_{AC}	r_{AG}	r_{AT}
C	r_{CA}	-	r_{CG}	r_{CT}
G	r_{GA}	r_{GC}	-	r_{GT}
T	r_{TA}	r_{TC}	r_{TG}	-

塩基置換確率行列

From\To	A	C	G	T
A	-	r_{AC}	r_{AG}	r_{AT}
C	r_{CA}	-	r_{CG}	r_{CT}
G	r_{GA}	r_{GC}	-	r_{GT}
T	r_{TA}	r_{TC}	r_{TG}	-

- r_{XY}は塩基Xから塩基Yへの置換確率
- $r_{XY}=r_{YX}$なモデルを時間反転可能という

サイト間の置換速度不均質性

site	12345	...	....M
OTU1	TGTTT	...	TTTTC
OTU2	AGTAC	...	TTTTC
OTU3	AGTAT	...	TTGTC
.	.		.
.	.		.
.	.		.
OTUN	AGTAT	...	ATTTC

サイト間の置換速度不均質性

site	12345	...	....M
OTU1	TGTTT	...	TTTTC
OTU2	AGTAC	...	TTTTC
OTU3	AGTAT	...	TTGTC
.	.		.
.	.		.
.	.		.
OTUN	AGTAT	...	ATTTC

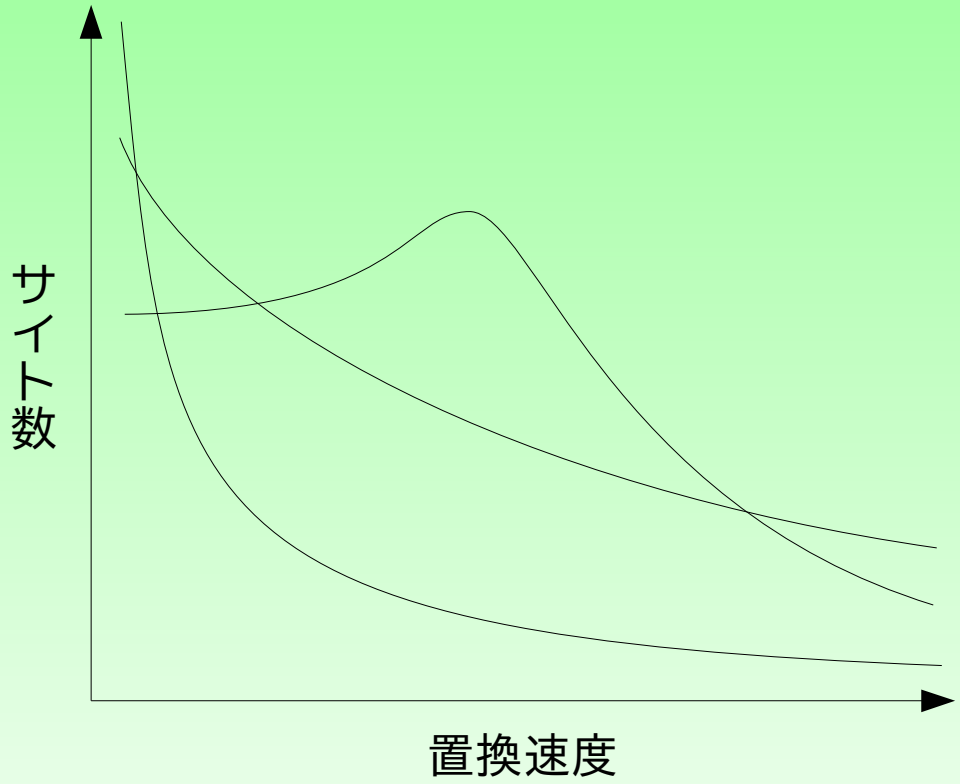
- 変異の多い＝速いサイトと少ない＝遅いサイトがある

サイト間の置換速度不均質性

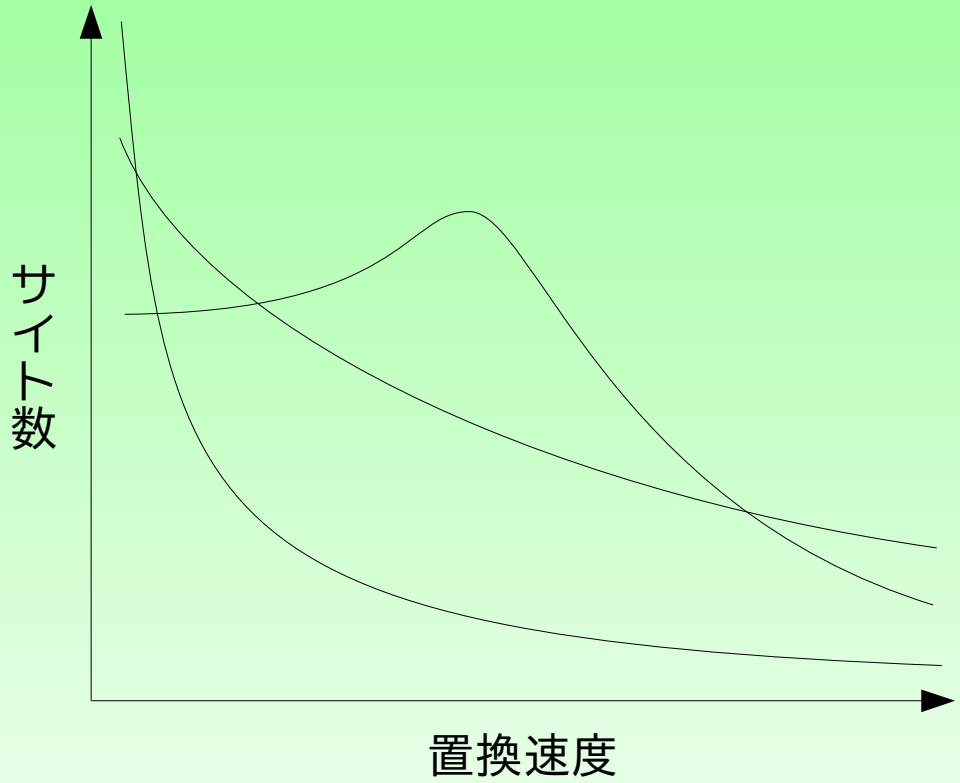
site	12345	...	....M
OTU1	TGTTT	...	TTTTC
OTU2	AGTAC	...	TTTTC
OTU3	AGTAT	...	TTGTC
.	.		.
.	.		.
.	.		.
OTUN	AGTAT	...	ATTTC

- 変異の多い＝速いサイトと少ない＝遅いサイトがある
- サイト間の置換速度の不均質性をモデル化する

サイト間の置換速度不均質性

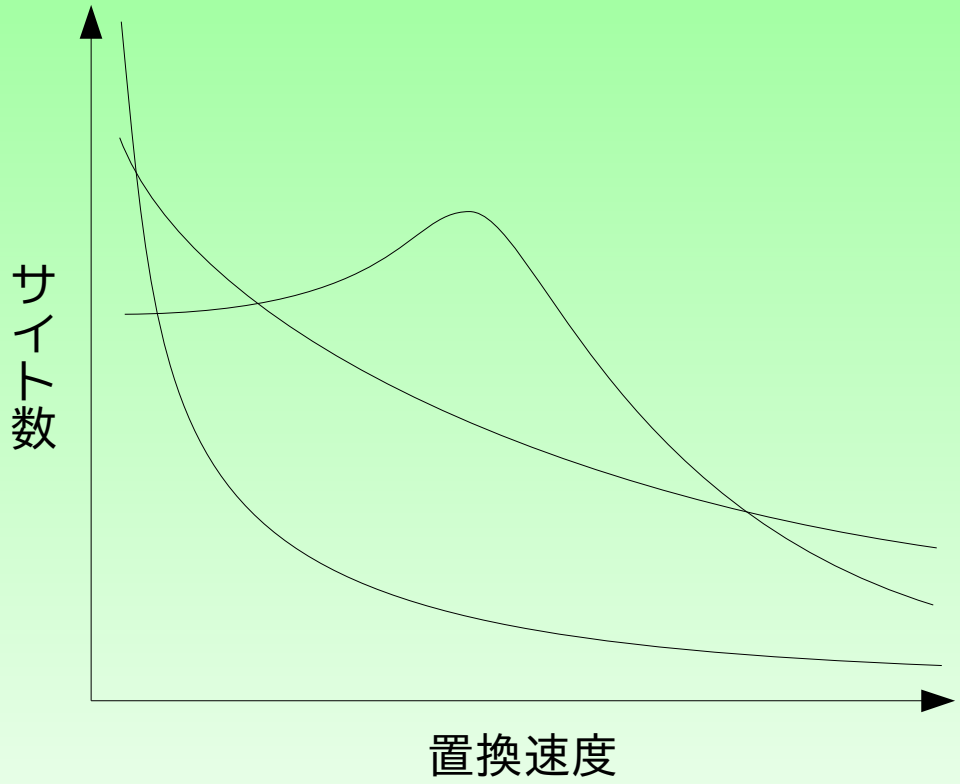


サイト間の置換速度不均質性



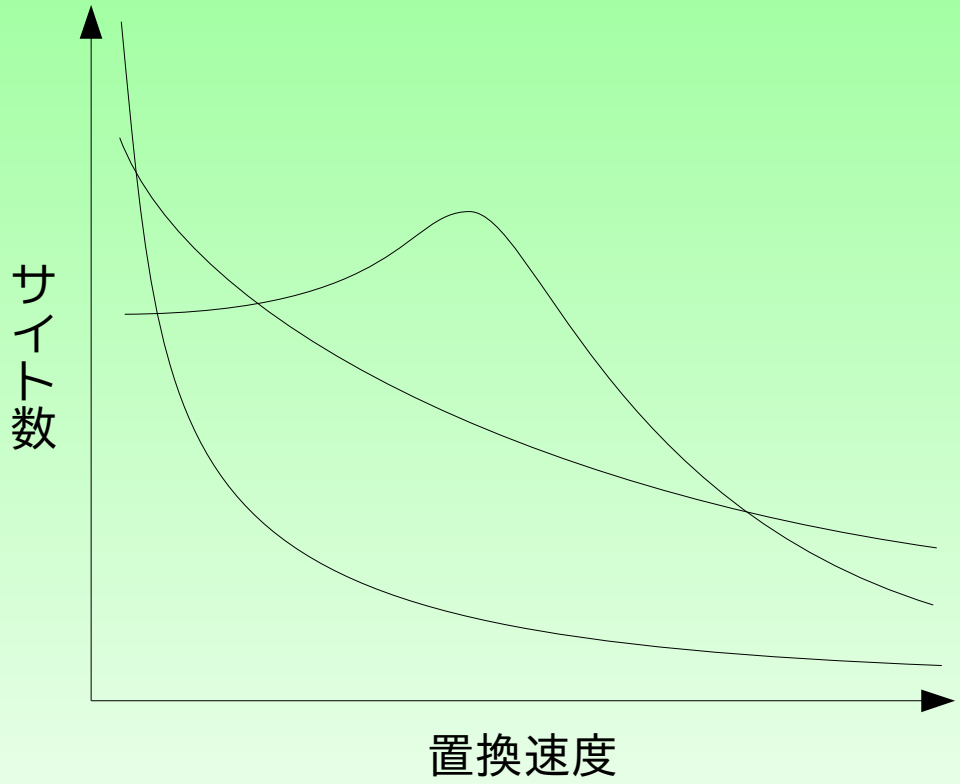
- 置換速度は連続データ
- 分布は非対称

サイト間の置換速度不均質性



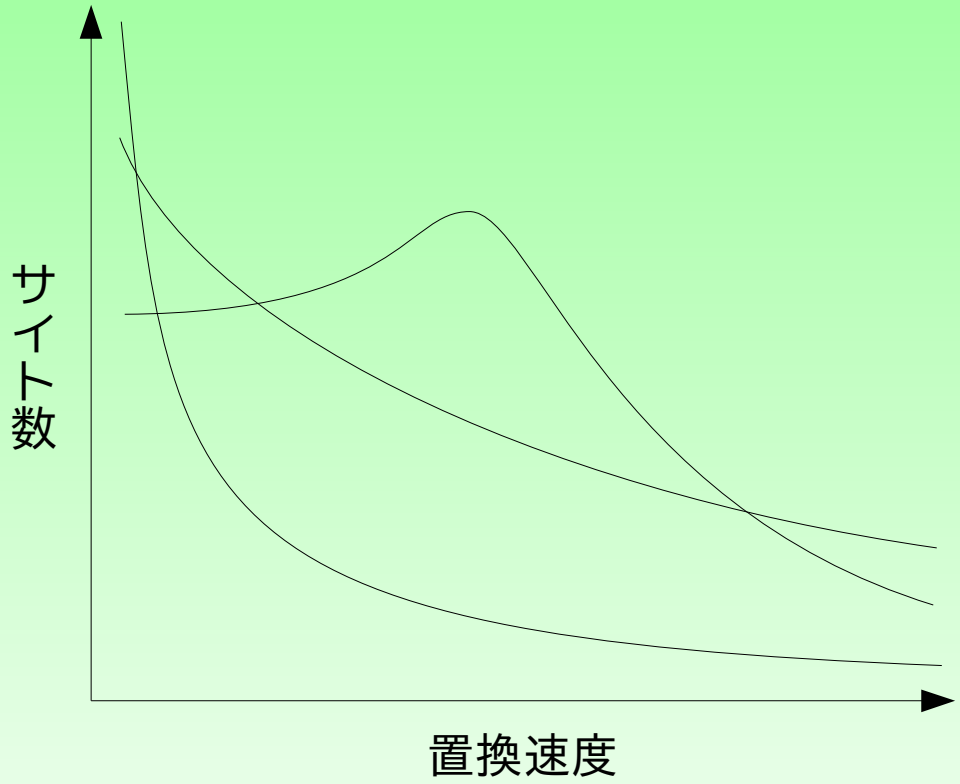
- 置換速度は連続データ
- 分布は非対称
- ガンマ分布を当てはめられる
  - パラメータ数は1

サイト間の置換速度不均質性



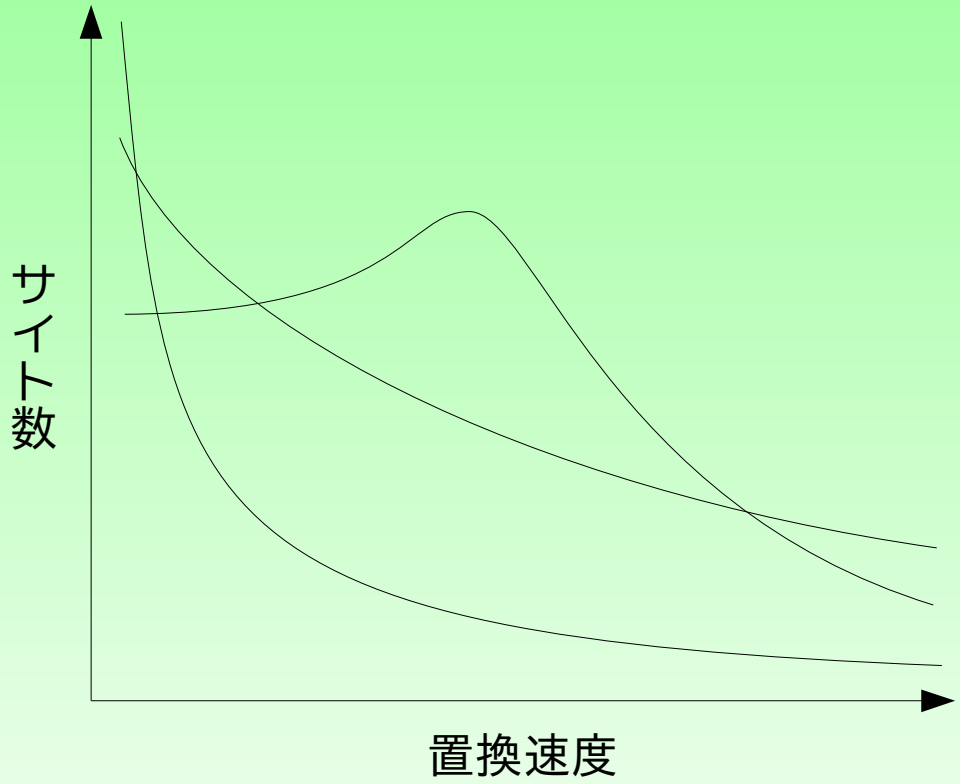
- 置換速度は連続データ
- 分布は非対称
- ガンマ分布を当てはめられる
  - パラメータ数は1
- 変異があるサイトと無いサイトのカテゴリ分けも用いられる
  - パラメータ数は1

サイト間の置換速度不均質性



- 置換速度は連続データ
- 分布は非対称
- ガンマ分布を当てはめられる
  - パラメータ数は1
- 変異があるサイトと無いサイトのカテゴリ分けも用いられる
  - パラメータ数は1
- 変異のあるサイトと無いサイトに分けた上で、変異のあるサイトにガンマ分布を当てはめる
  - パラメータ数は2

サイト間の置換速度不均質性



- 置換速度は連続データ
- 分布は非対称
- ガンマ分布を当てはめられる
  - パラメータ数は1
- 変異があるサイトと無いサイトのカテゴリ分けも用いられる
  - パラメータ数は1
- 変異のあるサイトと無いサイトに分けた上で、変異のあるサイトにガンマ分布を当てはめる
  - パラメータ数は2
- 3つのコドン位置ごとに異なる速度を当てはめる
  - パラメータ数は2

多領域データの場合

- 遺伝子A

```
OTU1  TGTTTTCTTT  ...  TTTTC
OTU2  AGTATTCTTC  ...  TTTTC
OTU3  AGTATTCTTT  ...  TTTTC
OTU4  AATATTTTTT  ...  TTTTC
OTU5  AGTATTTTTT  ...  TTTTC
OTU6  AGTATTCTCT  ...  TTTCC
OTU7  AGTATTCTTT  ...  TTTTC
OTU8  AGTATTTTTT  ...  TTTTC
OTU9  AGTATTCTTT  ...  TTTTC
```

- 遺伝子B

```
OTU1  TCCTCTACTA  ...  AGCTA
OTU2  TCTCCTACTA  ...  AGCTA
OTU3  TCTACTACTA  ...  AGCTA
OTU4  TCTTCCACTA  ...  AGTTA
OTU5  TCTGCCGCTA  ...  AGTTA
OTU6  TCTTTCACTA  ...  AGTCA
OTU7  TCTTTTACTA  ...  AGTCA
OTU8  TTTCCCGCTG  ...  AGCCA
OTU9  ATTTCCACTG  ...  AGCCA
```

多領域データの場合

- 遺伝子A

```
OTU1  TGTTTTCTTT  ...  TTTTC
OTU2  AGTATTCTTC  ...  TTTTC
OTU3  AGTATTCTTT  ...  TTTTC
OTU4  AATATTTTTT  ...  TTTTC
OTU5  AGTATTTTTT  ...  TTTTC
OTU6  AGTATTCTCT  ...  TTTCC
OTU7  AGTATTCTTT  ...  TTTTC
OTU8  AGTATTTTTT  ...  TTTTC
OTU9  AGTATTCTTT  ...  TTTTC
```

- 塩基置換確率行列はGTR
- サイト間の速度不均質性はガンマ分布でモデル化

- 遺伝子B

```
OTU1  TCCTCTACTA  ...  AGCTA
OTU2  TCTCCTACTA  ...  AGCTA
OTU3  TCTACTACTA  ...  AGCTA
OTU4  TCTTCCACTA  ...  AGTTA
OTU5  TCTGCCGCTA  ...  AGTTA
OTU6  TCTTTCACTA  ...  AGTCA
OTU7  TCTTTTACTA  ...  AGTCA
OTU8  TTTCCCGCTG  ...  AGCCA
OTU9  ATTTCCACTG  ...  AGCCA
```

- 塩基置換確率行列はK2P
- サイト間の速度不均質性は可変/不変サイト比でモデル化

多領域データの場合

- 遺伝子A

```
OTU1  TGTTTTCTTT  ...  TTTTC
OTU2  AGTATTCTTC  ...  TTTTC
OTU3  AGTATTCTTT  ...  TTTTC
OTU4  AATATTTTTT  ...  TTTTC
OTU5  AGTATTTTTT  ...  TTTTC
OTU6  AGTATTCTCT  ...  TTTCC
OTU7  AGTATTCTTT  ...  TTTTC
```

- 塩基置換確率行列はGTR
- サイト間の速度不均質性はガンマ分布でモデル化

- 遺伝子B

```
OTU1  TCCTCTACTA  ...  AGCTA
OTU2  TCTCCTACTA  ...  AGCTA
OTU3  TCTACTACTA  ...  AGCTA
OTU4  TCTTCCACTA  ...  AGTTA
OTU5  TCTGCCGCTA  ...  AGTTA
OTU6  TCTTTCACTA  ...  AGTCA
OTU7  TCTTTTACTA  ...  AGTCA
```

- 塩基置換確率行列はK2P
- サイト間の速度不均質性は可変/不変サイト比でモデル化

複数領域のそれぞれに異なる分子進化モデルを当てはめる

多領域データの場合

- 遺伝子A

```
OTU1  TGTTTTCTTT  ...  TTTTC
OTU2  AGTATTCTTC  ...  TTTTC
OTU3  AGTATTCTTT  ...  TTTTC
OTU4  AATATTTTTT  ...  TTTTC
OTU5  AGTATTTTTT  ...  TTTTC
OTU6  AGTATTCTCT  ...  TTTCC
OTU7  AGTATTCTTT  ...  TTTTC
```

- 塩基置換確率行列はGTR
- サイト間の速度不均質性はガンマ分布でモデル化

- 遺伝子B

```
OTU1  TCCTCTACTA  ...  AGCTA
OTU2  TCTCCTACTA  ...  AGCTA
OTU3  TCTACTACTA  ...  AGCTA
OTU4  TCTTCCACTA  ...  AGTTA
OTU5  TCTGCCGCTA  ...  AGTTA
OTU6  TCTTTCACTA  ...  AGTCA
OTU7  TCTTTTACTA  ...  AGTCA
```

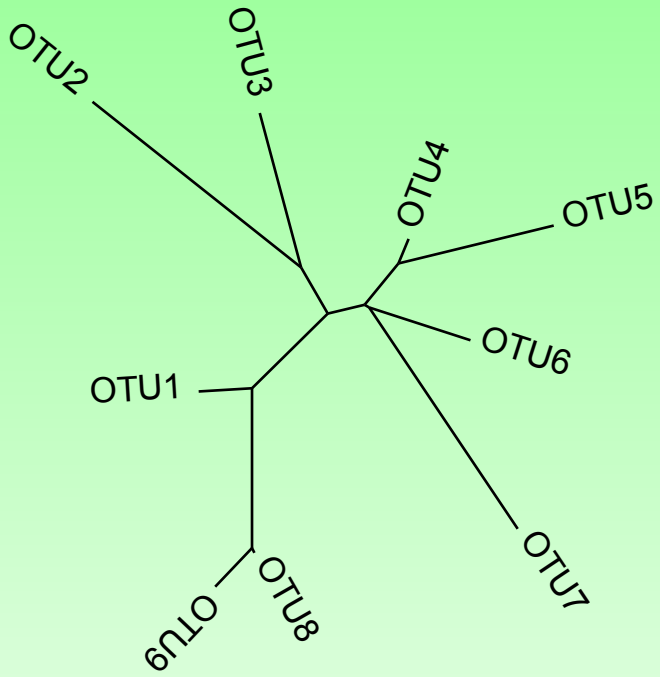
- 塩基置換確率行列はK2P
- サイト間の速度不均質性は可変/不変サイト比でモデル化

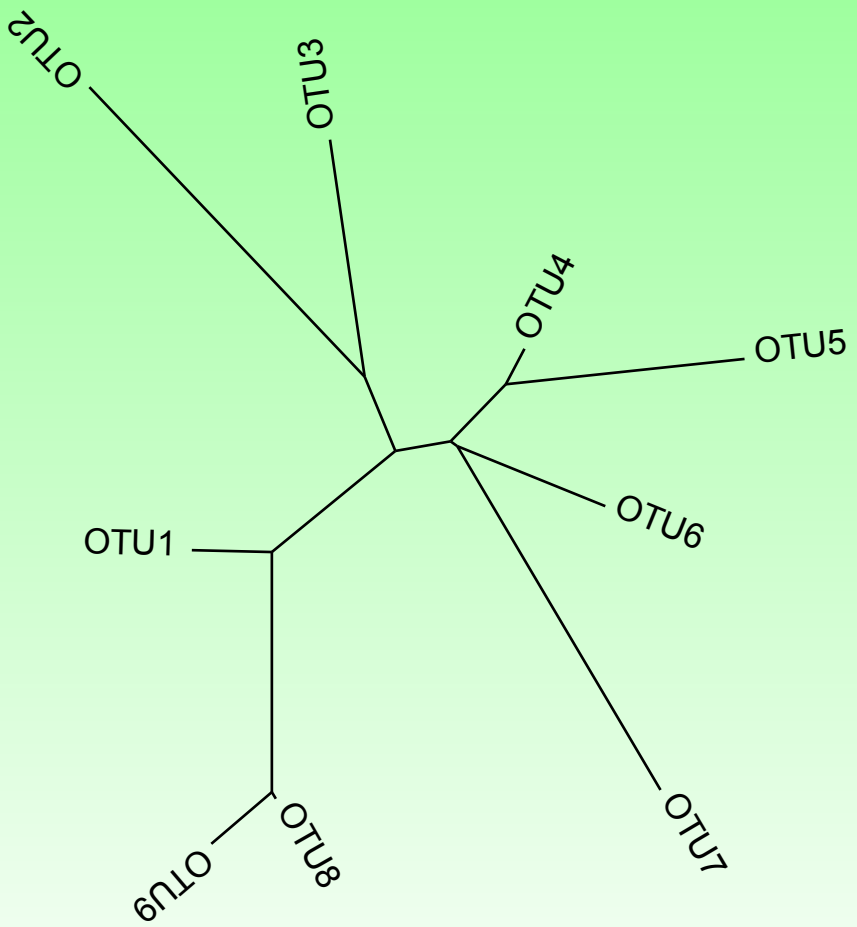
複数領域のそれぞれに異なる分子進化モデルを当てはめる

比例モデルと分離モデルという二通りの当てはめ方がある

比例モデル

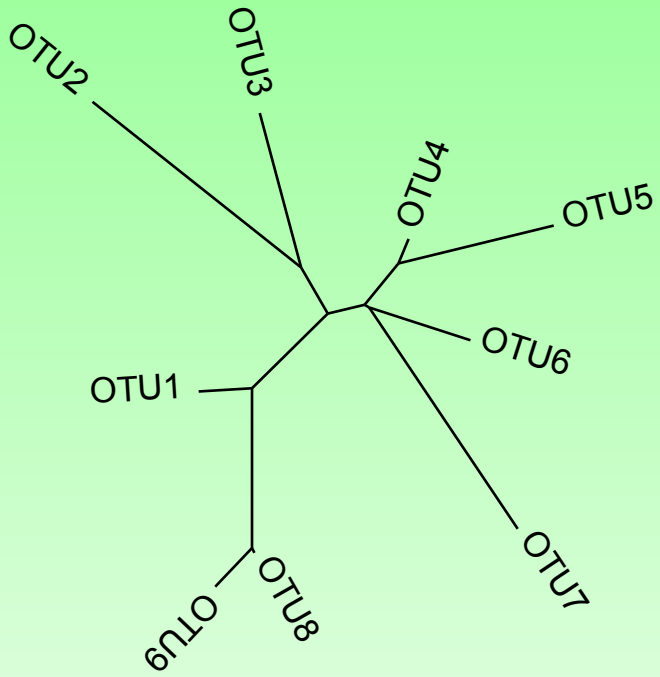
比例モデル

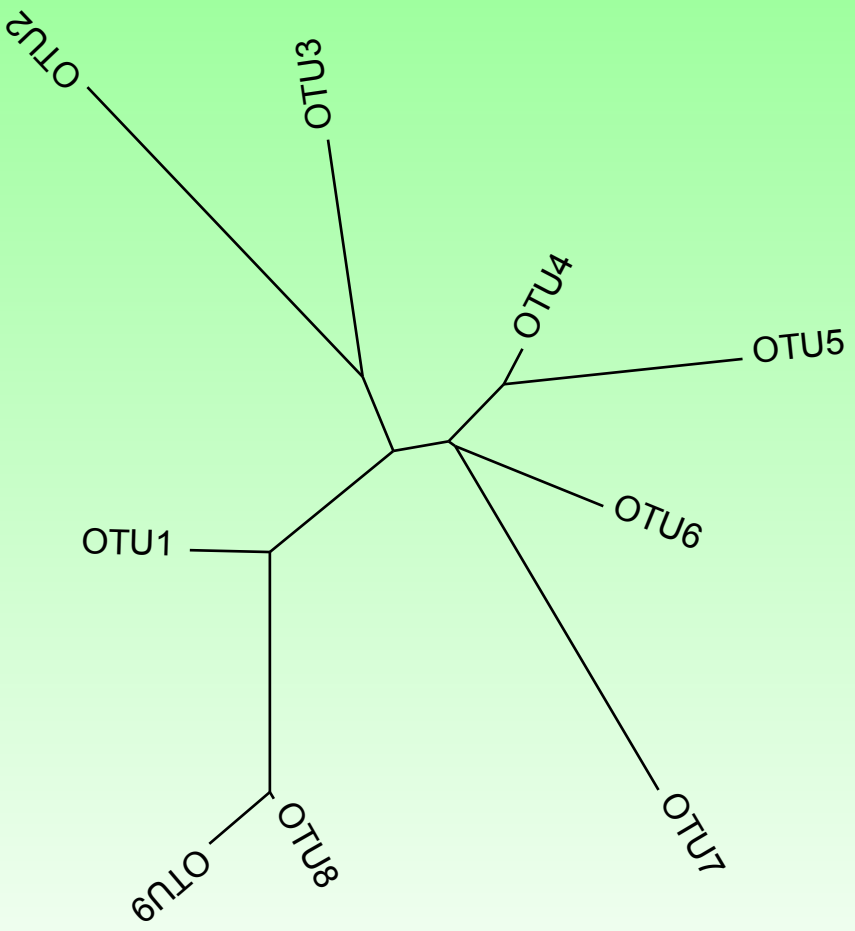




- 比例モデルでは系統樹は相似形

比例モデル

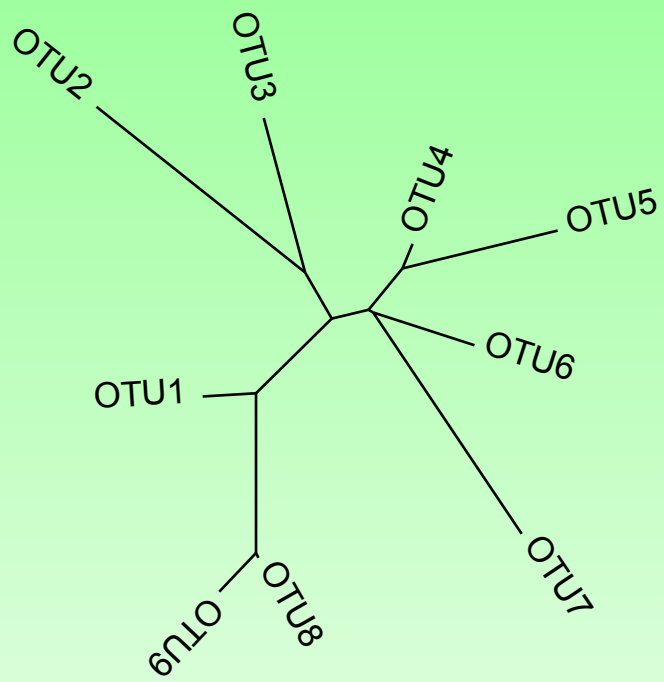


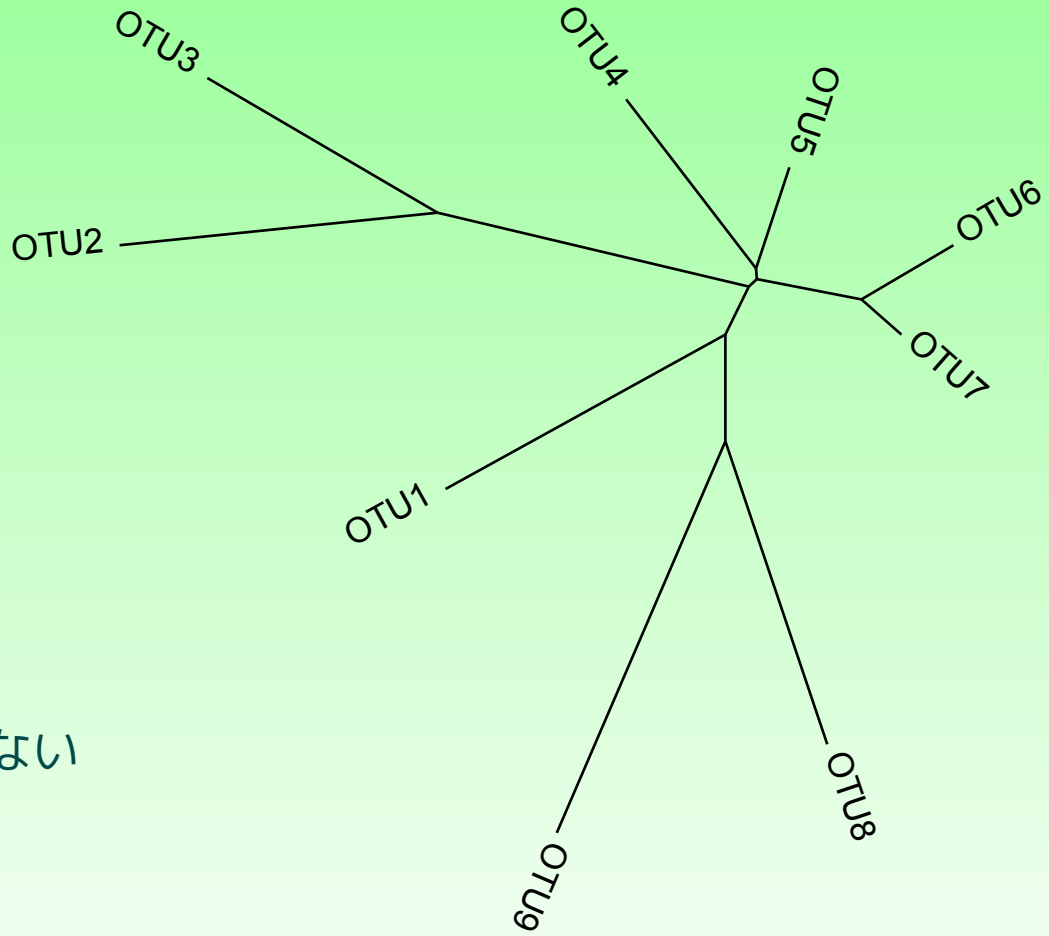


- 比例モデルでは系統樹は相似形
- パラメータ数は
  - 枝長：OTU数×2－3
  - 速度比(枝長比)：領域数－1

分離モデル

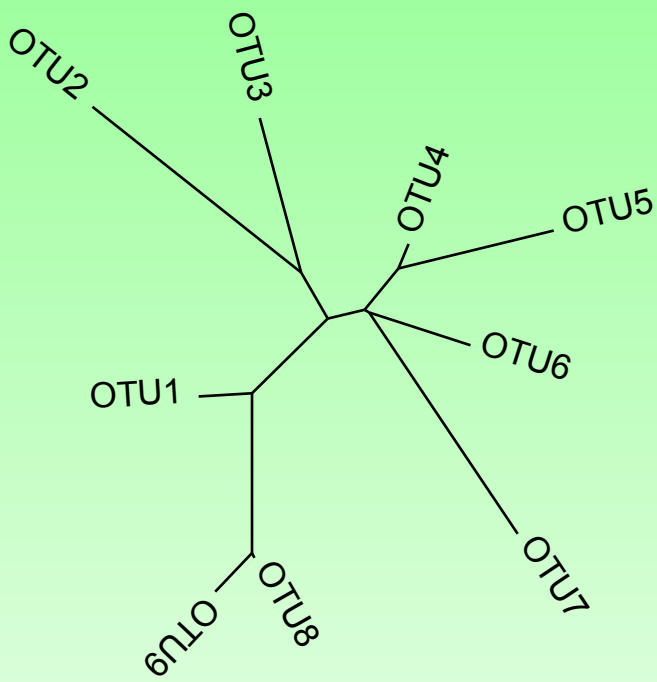
分離モデル

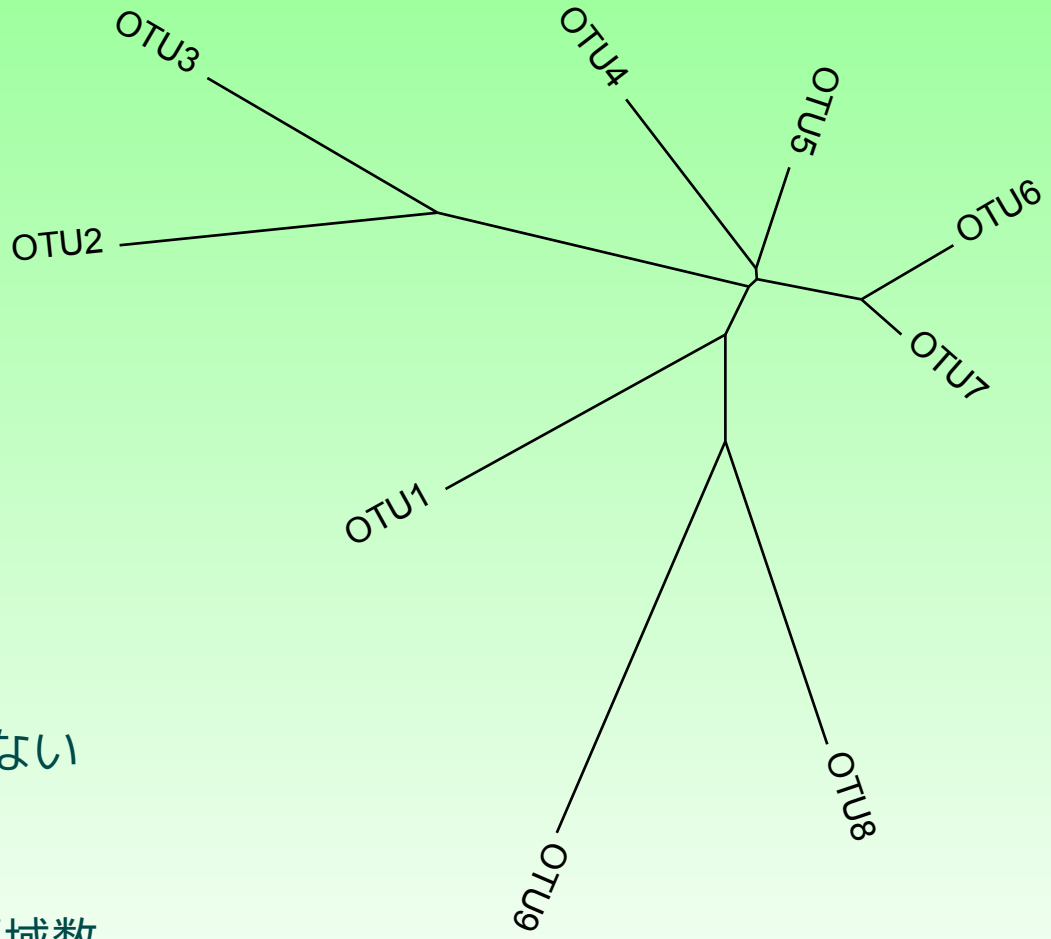




- 分離モデルでは系統樹は相似でない

分離モデル





- 分離モデルでは系統樹は相似でない
- パラメータ数は
  - 枝長：(OTU数×2－3) × 領域数

多領域データにおけるモデルとパラメータ数

多領域データにおけるモデルとパラメータ数

- 分子進化モデル
  - 置換確率行列
    - DNAでは領域ごとに0〜8
  - サイト間置換速度不均質性
    - 領域ごとに0以上

多領域データにおけるモデルとパラメータ数

- 分子進化モデル
  - 置換確率行列
    - DNAでは領域ごとに0〜8
  - サイト間置換速度不均質性
    - 領域ごとに0以上
- 系統樹
  - 単一モデルの場合
    - 枝長：OTU数×2－3
  - 比例モデルの場合
    - 枝長：OTU数×2－3
    - 枝長比：領域数－1
  - 分離モデルの場合
    - 枝長：(OTU数×2－3) × 領域数

多領域データにおけるモデルとパラメータ数

- 分子進化モデル
  - 置換確率行列
    - DNAでは領域ごとに0〜8
  - サイト間置換速度不均質性
    - 領域ごとに0以上

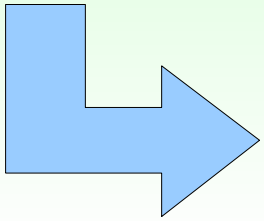
各領域ごとに最適な
分子進化モデルを選択し、
そのモデルを適用して
最適な樹形を選択する

- 系統樹
  - 単一モデルの場合
    - 枝長：OTU数×2－3
  - 比例モデルの場合
    - 枝長：OTU数×2－3
    - 枝長比：領域数－1
  - 分離モデルの場合
    - 枝長：(OTU数×2－3) × 領域数

多領域データにおけるモデルとパラメータ数

- 分子進化モデル
  - 置換確率行列
    - DNAでは領域ごとに0〜8
  - サイト間置換速度不均質性
    - 領域ごとに0以上

各領域ごとに最適な
分子進化モデルを選択し、
そのモデルを適用して
最適な樹形を選択する



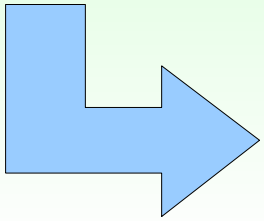
Kakusan3

- 系統樹
  - 単一モデルの場合
    - 枝長：OTU数×2－3
  - 比例モデルの場合
    - 枝長：OTU数×2－3
    - 枝長比：領域数－1
  - 分離モデルの場合
    - 枝長：(OTU数×2－3)
      × 領域数

多領域データにおけるモデルとパラメータ数

- 分子進化モデル
  - 置換確率行列
    - DNAでは領域ごとに0〜8
  - サイト間置換速度不均質性
    - 領域ごとに0以上

各領域ごとに最適な
分子進化モデルを選択し、
そのモデルを適用して
最適な樹形を選択する



Kakusan3

- 系統樹
  - 単一モデルの場合
    - 枝長：OTU数×2－3
  - 比例モデルの場合
    - 枝長：OTU数×2－3
    - 枝長比：領域数－1
  - 分離モデルの場合
    - 枝長：(OTU数×2－3)
      × 領域数

単一・比例・分離モデル間
の比較はしないことに注意

Kakusan3の概要

Kakusan3の概要

- PC・Macからクラスタ・スパコンでも動作可能
  - Windows・MacOS X・その他UNIX対応

Kakusan3の概要

- PC・Macからクラスタ・スパコンでも動作可能
  - Windows・MacOS X・その他UNIX対応
- 多くの形式のデータを読み込み可能
  - FASTA・GenBank・NEXUS・PHYLIP形式に対応

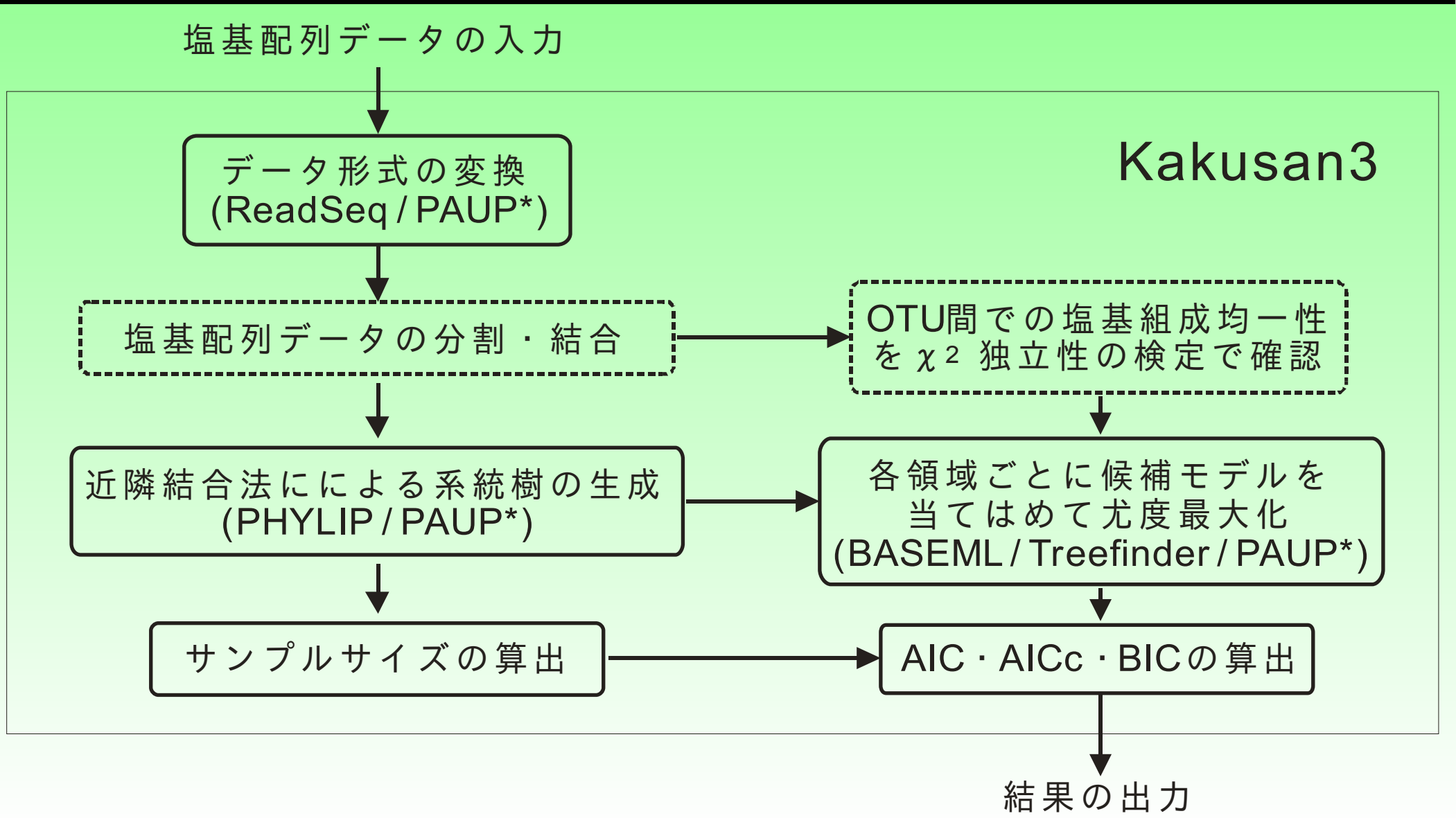
Kakusan3の概要

- PC・Macからクラスタ・スパコンでも動作可能
  - Windows・MacOS X・その他UNIX対応
- 多くの形式のデータを読み込み可能
  - FASTA・GenBank・NEXUS・PHYLIP形式に対応
- 解析ソフト用設定ファイルの出力
  - PAUP*4・MrBayes3・Treefinderに対応

Kakusan3の概要

- PC・Macからクラスタ・スパコンでも動作可能
  - Windows・MacOS X・その他UNIX対応
- 多くの形式のデータを読み込み可能
  - FASTA・GenBank・NEXUS・PHYLIP形式に対応
- 解析ソフト用設定ファイルの出力
  - PAUP*4・MrBayes3・Treefinderに対応
- 遺伝子座・コドン位置ごとにモデル選択
  - 各モデルの尤度最大化を並列処理し高速化

Kakusan3が中でやっていること



入力データファイルの用意

- A.phy

```
 9 15
Outgroup  TGTTTTCTTTTTTTC
Tax1      AGTATTCTTCTTTTC
Tax2      AGTATTCTTTTTTTC
Tax3      AATATTTTTTTTTTC
Tax4      AGTATTTTTTTTTTC
Tax5      AGTATTCTCTTTTCC
Tax6      AGTATTCTTTTTTTC
Tax7      AGTATTTTTTTTTTC
Tax8      AGTATTCTTTTTTTC
```

- B_P.phy

```
 9 15
Outgroup  TCCTCTACTAAGCTA
Tax1      TCTCCTACTAAGCTA
Tax2      TCTACTACTAAGCTA
Tax3      TCTTCCACTAAGTTA
Tax4      TCTGCCGCTAAGTTA
Tax5      TCTTTCACTAAGTCA
Tax6      TCTTTTACTAAGTCA
Tax7      TTTCCCGCTGAGCCA
Tax8      ATTTCCACTGAGCCA
```

入力データファイルの用意

- A.phy

```
 9 15
Outgroup  TGTTTTCTTTTTTTC
Tax1      AGTATTCTTCTTTTC
Tax2      AGTATTCTTTTTTTC
Tax3      AATATTTTTTTTTTC
Tax4      AGTATTTTTTTTTTC
Tax5      AGTATTCTCTTTTCC
Tax6      AGTATTCTTTTTTTC
Tax7      AGTATTTTTTTTTTC
Tax8      AGTATTCTTTTTTTC
```

- B_P.phy

```
 9 15
Outgroup  TCCTCTACTAAGCTA
Tax1      TCTCCTACTAAGCTA
Tax2      TCTACTACTAAGCTA
Tax3      TCTTCCACTAAGTTA
Tax4      TCTGCCGCTAAGTTA
Tax5      TCTTTCACTAAGTCA
Tax6      TCTTTTACTAAGTCA
Tax7      TTTCCCGCTGAGCCA
Tax8      ATTTCCACTGAGCCA
```

- 領域 (遺伝子座) ごとに異なるファイルに分ける

入力データファイルの用意

- A.phy

```
 9 15
Outgroup  TGTTTTCTTTTTTTC
Tax1      AGTATTCTTCTTTTC
Tax2      AGTATTCTTTTTTTC
Tax3      AATATTTTTTTTTTC
Tax4      AGTATTTTTTTTTTC
Tax5      AGTATTCTCTTTTCC
Tax6      AGTATTCTTTTTTTC
Tax7      AGTATTTTTTTTTTC
Tax8      AGTATTCTTTTTTTC
```

- B_P.phy

```
 9 15
Outgroup  TCCTCTACTAAGCTA
Tax1      TCTCCTACTAAGCTA
Tax2      TCTACTACTAAGCTA
Tax3      TCTTCCACTAAGTTA
Tax4      TCTGCCGCTAAGTTA
Tax5      TCTTTCACTAAGTCA
Tax6      TCTTTTACTAAGTCA
Tax7      TTTCCCGCTGAGCCA
Tax8      ATTTCCACTGAGCCA
```

- 領域 (遺伝子座) ごとに異なるファイルに分ける
- OTU名はファイル間で統一する

入力データファイルの用意

- A.phy

```
 9 15
Outgroup  TGTTTTCTTTTTTTC
Tax1      AGTATTCTTCTTTTC
Tax2      AGTATTCTTTTTTTC
Tax3      AATATTTTTTTTTTC
Tax4      AGTATTTTTTTTTTC
Tax5      AGTATTCTCTTTTCC
Tax6      AGTATTCTTTTTTTC
Tax7      AGTATTTTTTTTTTC
Tax8      AGTATTCTTTTTTTC
```

- B_P.phy

```
 9 15
Outgroup  TCCTCTACTAAGCTA
Tax1      TCTCCTACTAAGCTA
Tax2      TCTACTACTAAGCTA
Tax3      TCTTCCACTAAGTTA
Tax4      TCTGCCGCTAAGTTA
Tax5      TCTTTCACTAAGTCA
Tax6      TCTTTTACTAAGTCA
Tax7      TTTCCCGCTGAGCCA
Tax8      ATTTCCACTGAGCCA
```

- 領域 (遺伝子座) ごとに異なるファイルに分ける
- OTU名はファイル間で統一する
- タンパクコード領域のファイルはファイル名の末尾を「_P」にする

入力データファイルの用意

- A.phy

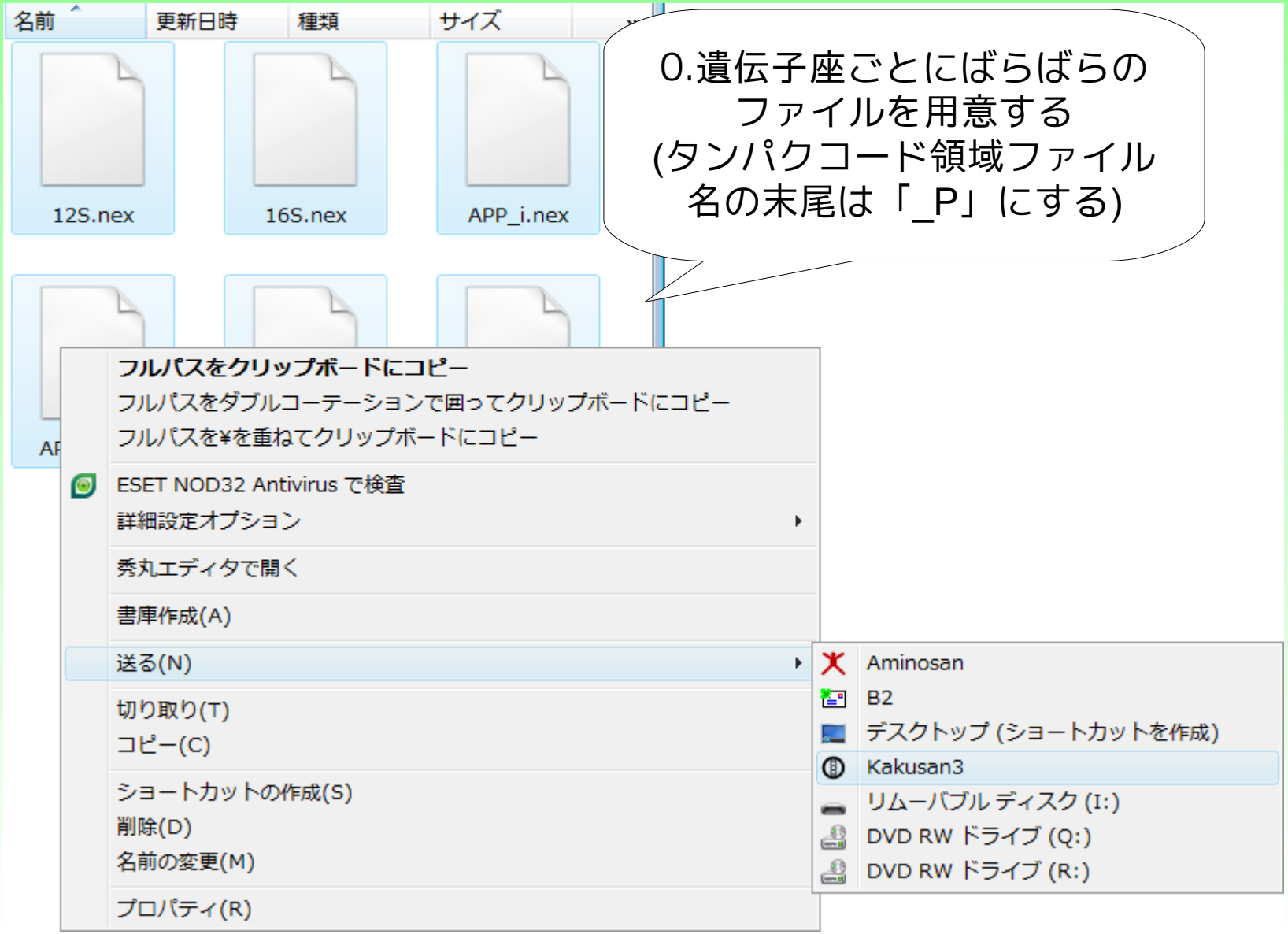
```
 9 15
Outgroup  TGTTTTCTTTTTTTC
Tax1      AGTATTCTTCTTTTC
Tax2      AGTATTCTTTTTTTC
Tax3      AATATTTTTTTTTTC
Tax4      AGTATTTTTTTTTTC
Tax5      AGTATTCTCTTTTCC
Tax6      AGTATTCTTTTTTTC
Tax7      AGTATTTTTTTTTTC
Tax8      AGTATTCTTTTTTTC
```

- B_P.phy

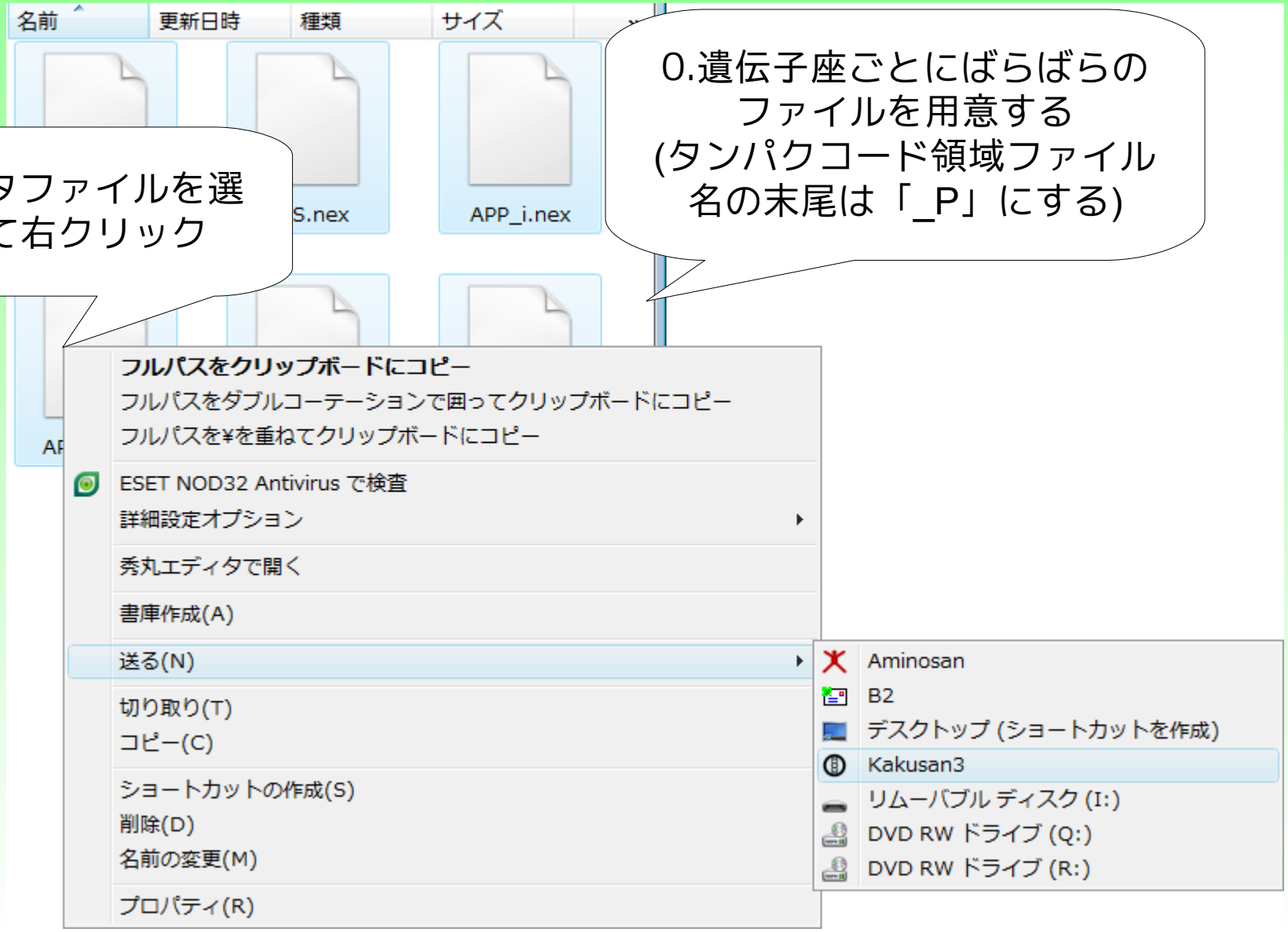
```
 9 15
Outgroup  TCCTCTACTAAGCTA
Tax1      TCTCCTACTAAGCTA
Tax2      TCTACTACTAAGCTA
Tax3      TCTTCCACTAAGTTA
Tax4      TCTGCCGCTAAGTTA
Tax5      TCTTTCACTAAGTCA
Tax6      TCTTTTACTAAGTCA
Tax7      TTTCCCGCTGAGCCA
Tax8      ATTTCCACTGAGCCA
```

- 領域 (遺伝子座) ごとに異なるファイルに分ける
- OTU名はファイル間で統一する
- タンパクコード領域のファイルはファイル名の末尾を「_P」にする
- intronとexonは混ぜるな危険

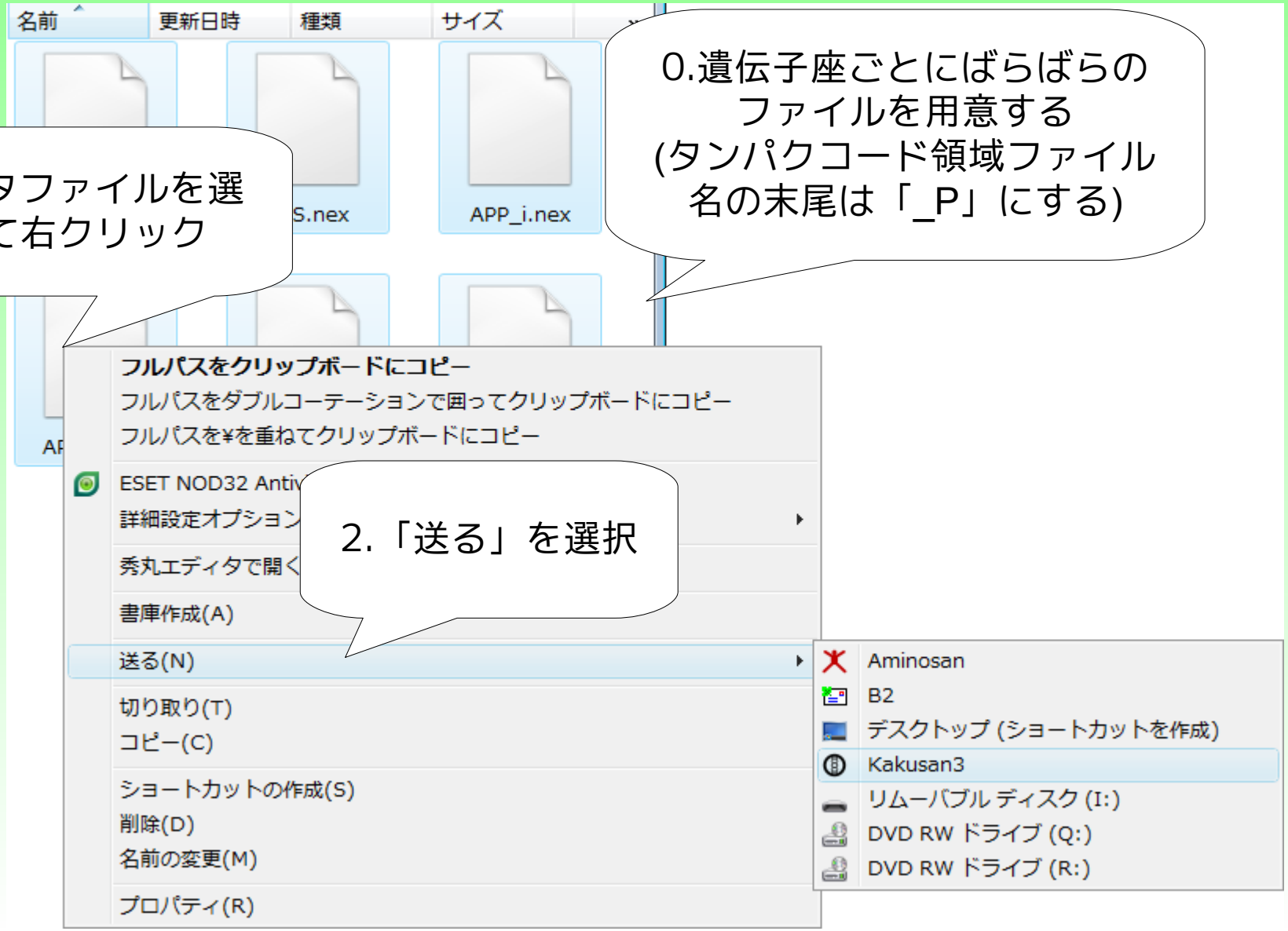
Kakusan3の起動とデータの入力 (Windowsのみ)



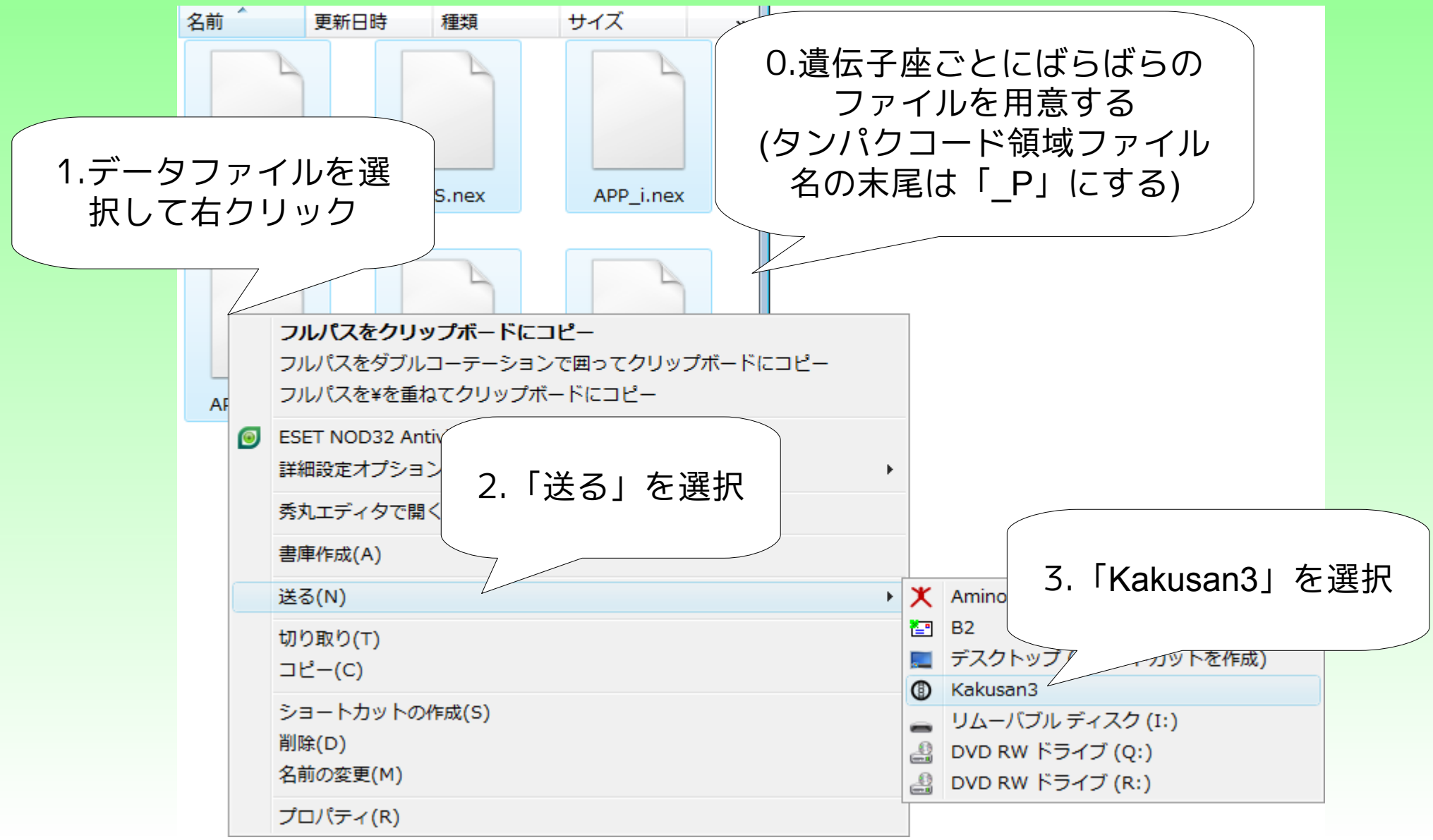
Kakusan3の起動とデータの入力 (Windowsのみ)



Kakusan3の起動とデータの入力 (Windowsのみ)

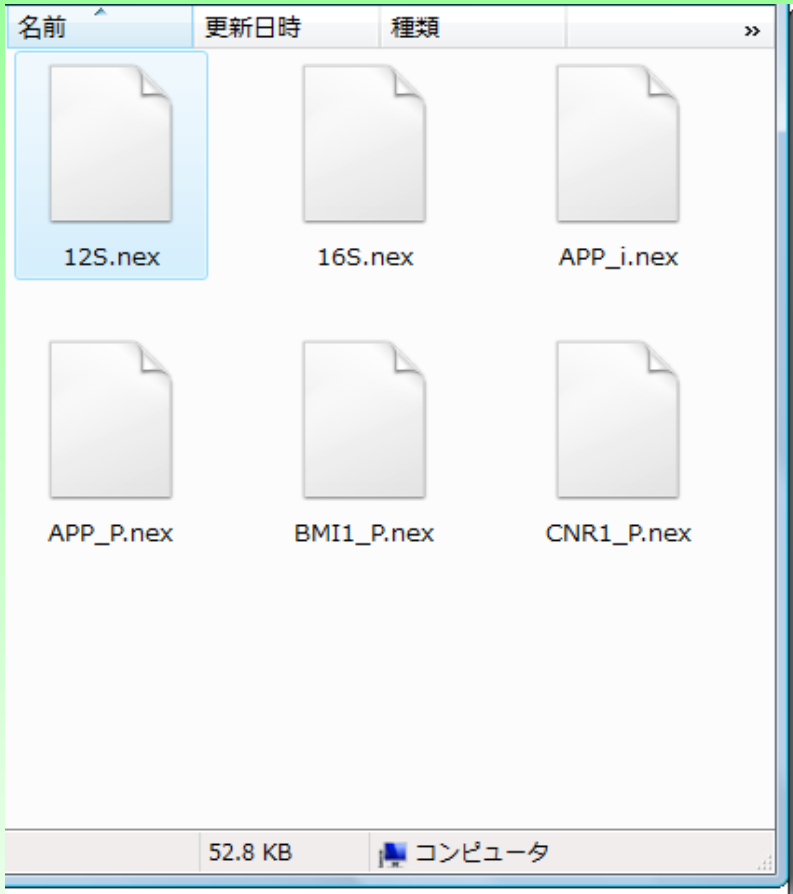


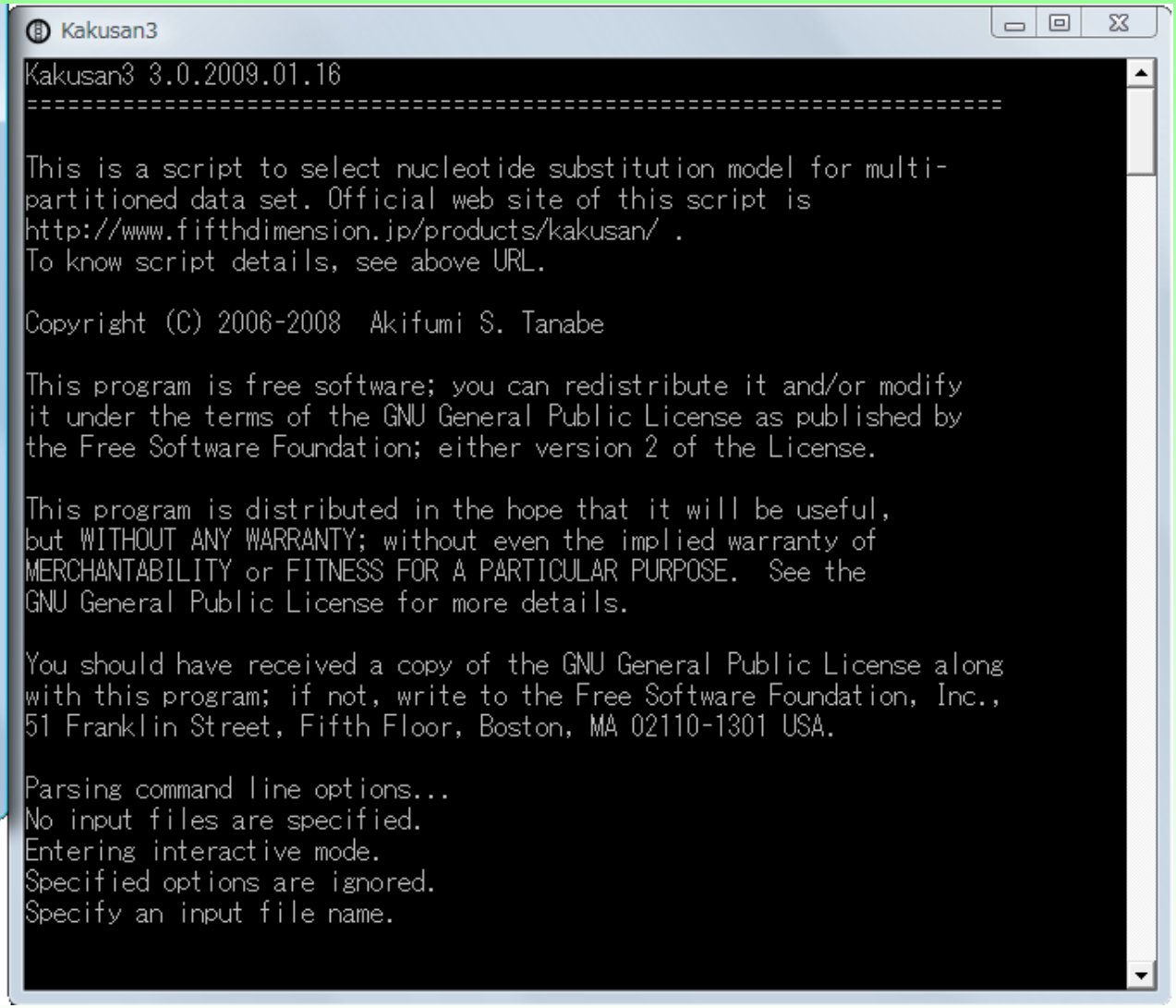
Kakusan3の起動とデータの入力 (Windowsのみ)



Kakusan3の起動とデータの入力 (Windows・MacOS X共通)

(XP以前)

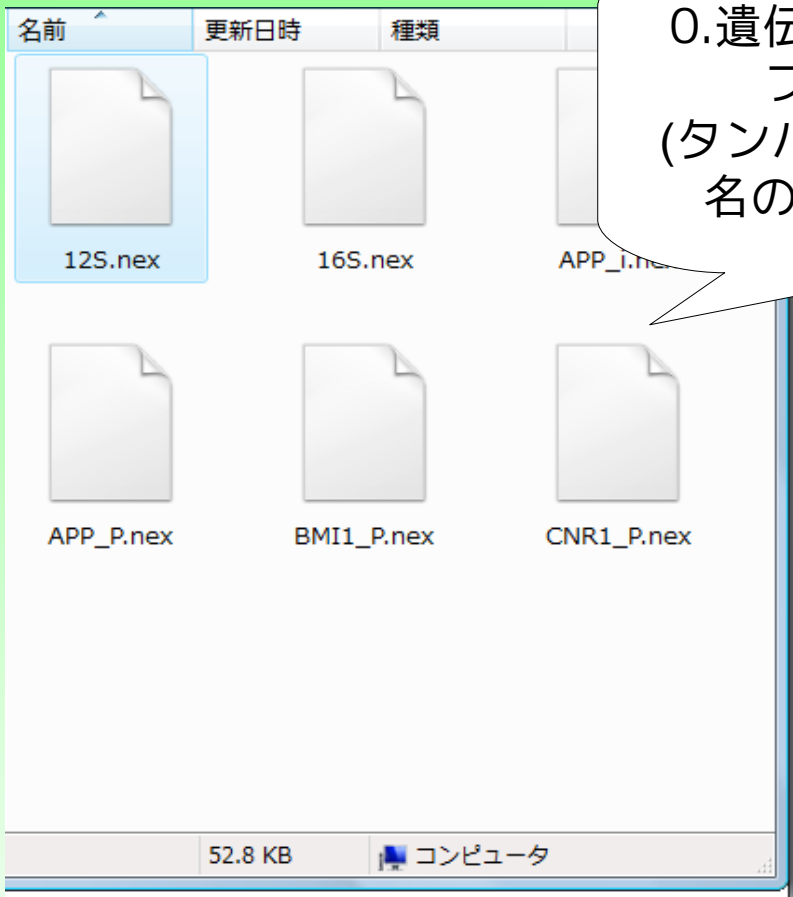


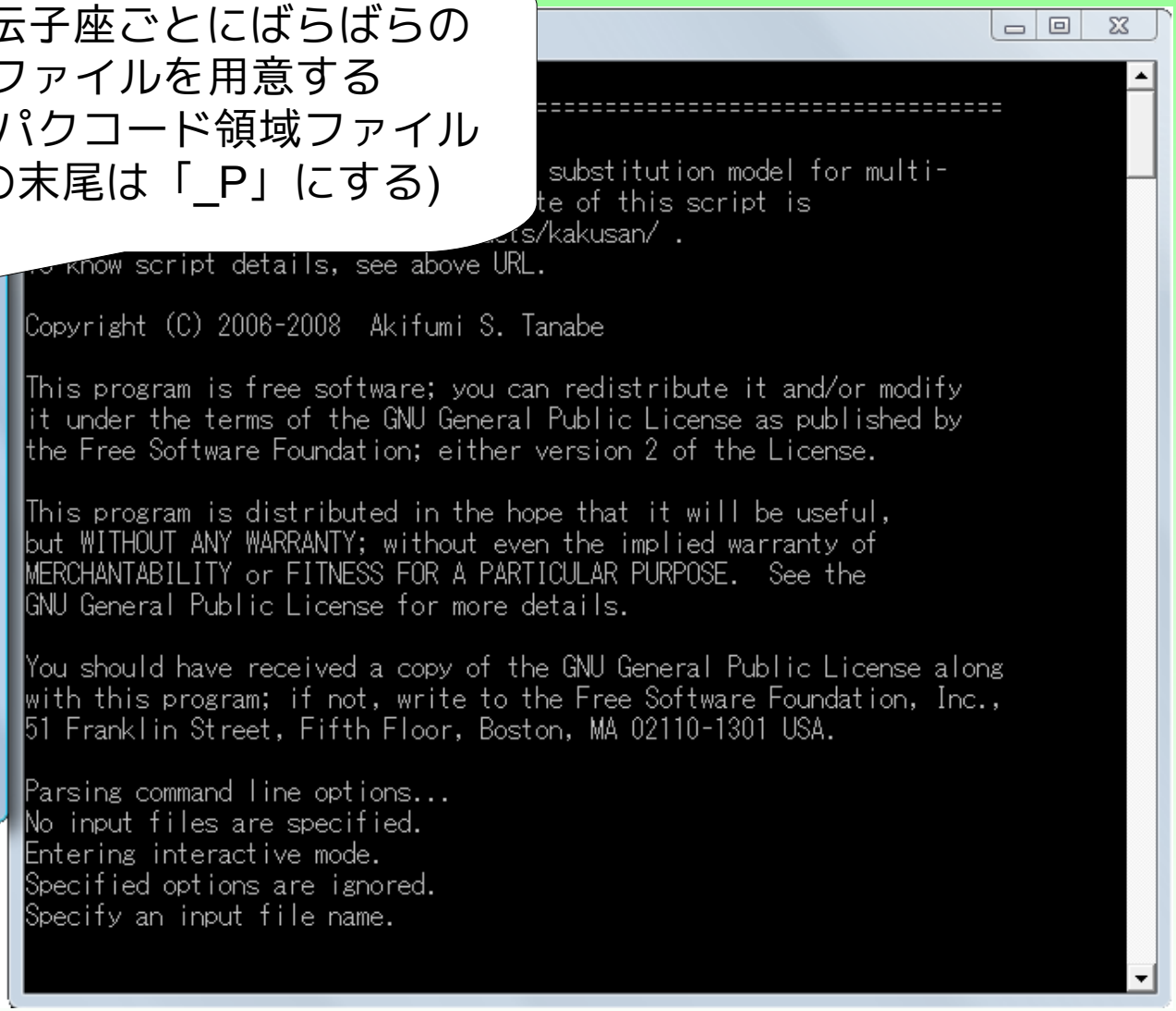


Kakusan3の起動とデータの入力 (Windows・MacOS X共通)

(XP以前)

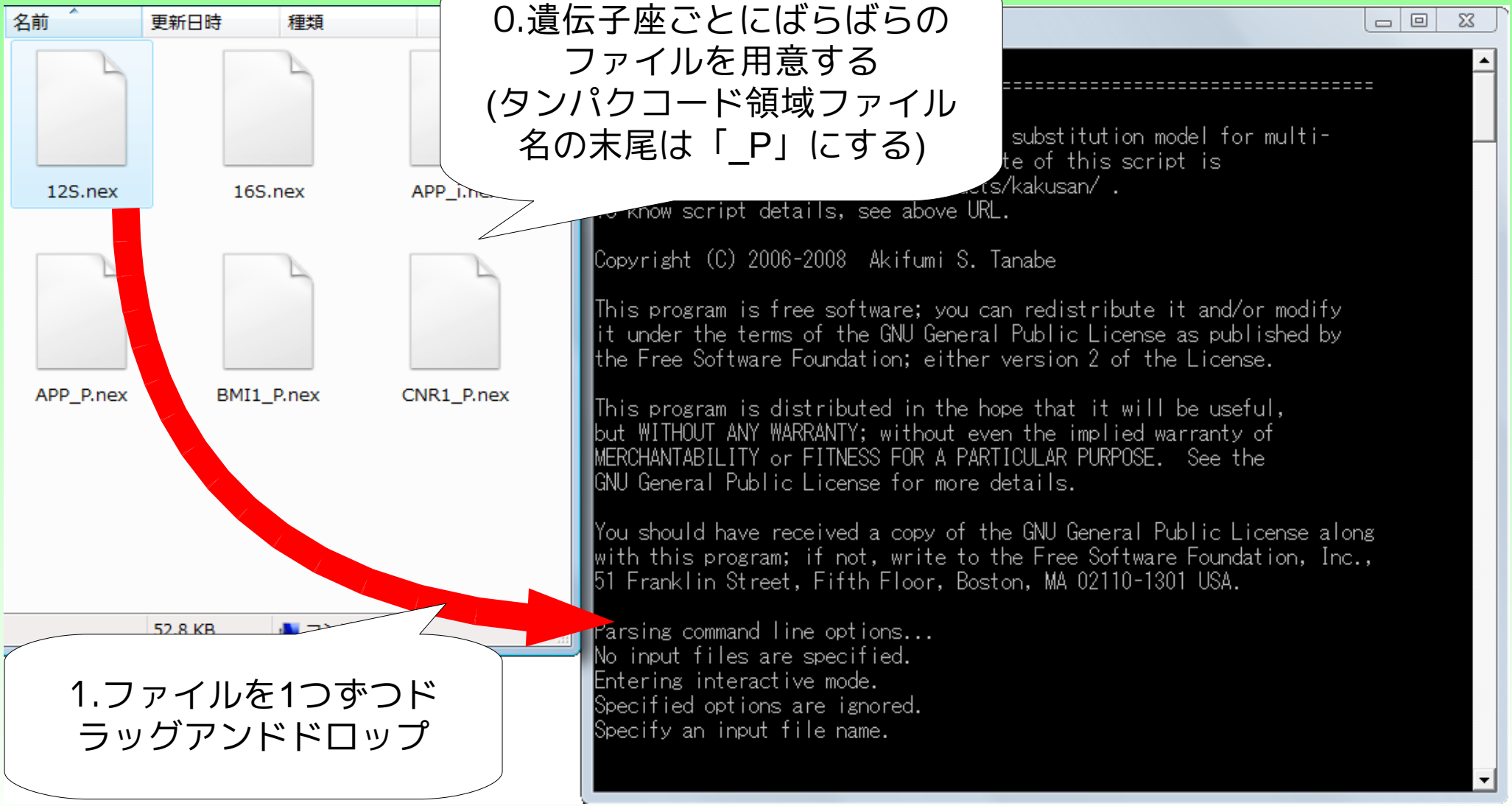
0.遺伝子座ごとにばらばらの
ファイルを用意する
(タンパクコード領域ファイル
名の末尾は「_P」にする)



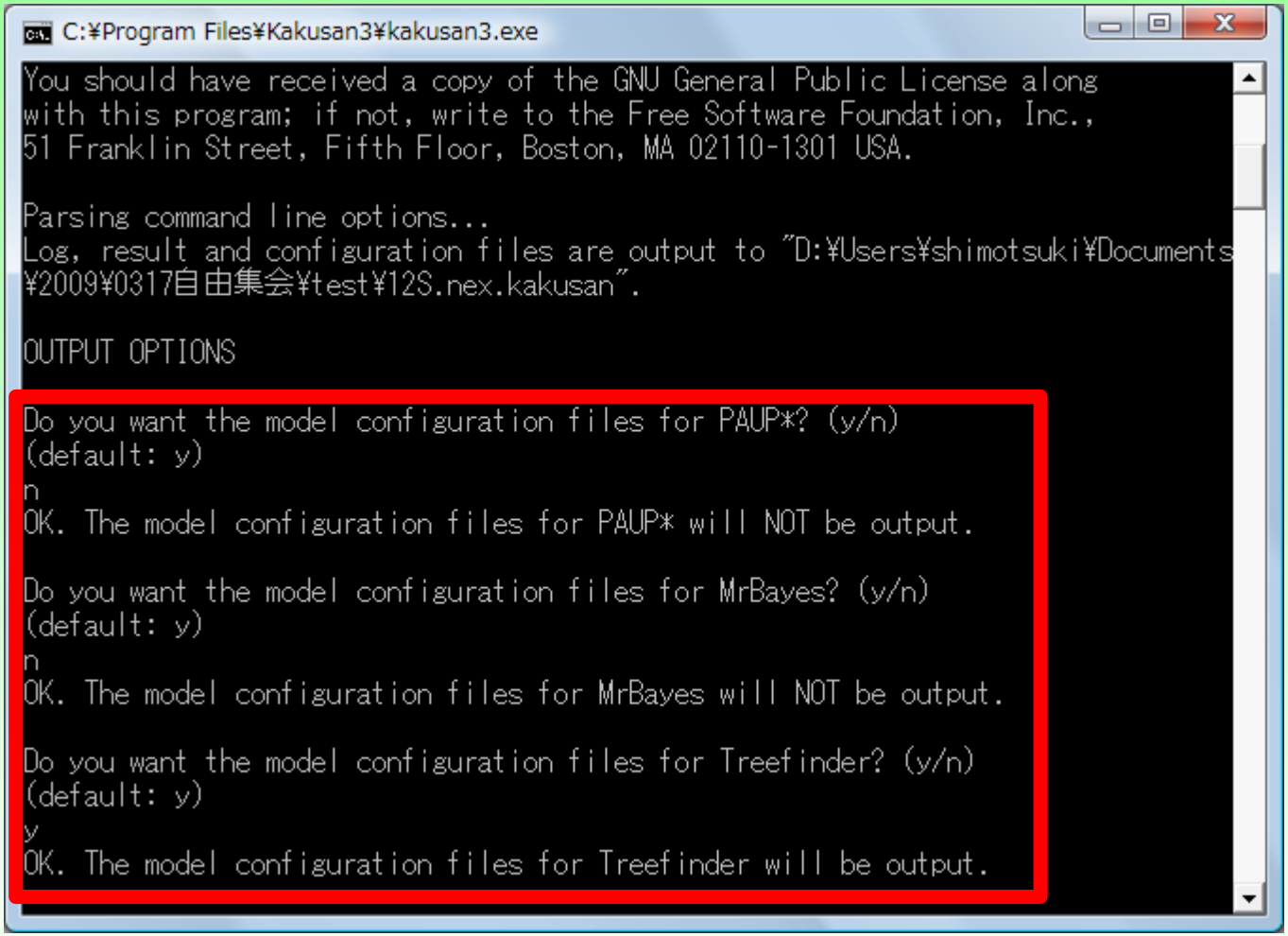


Kakusan3の起動とデータの入力 (Windows・MacOS X共通)

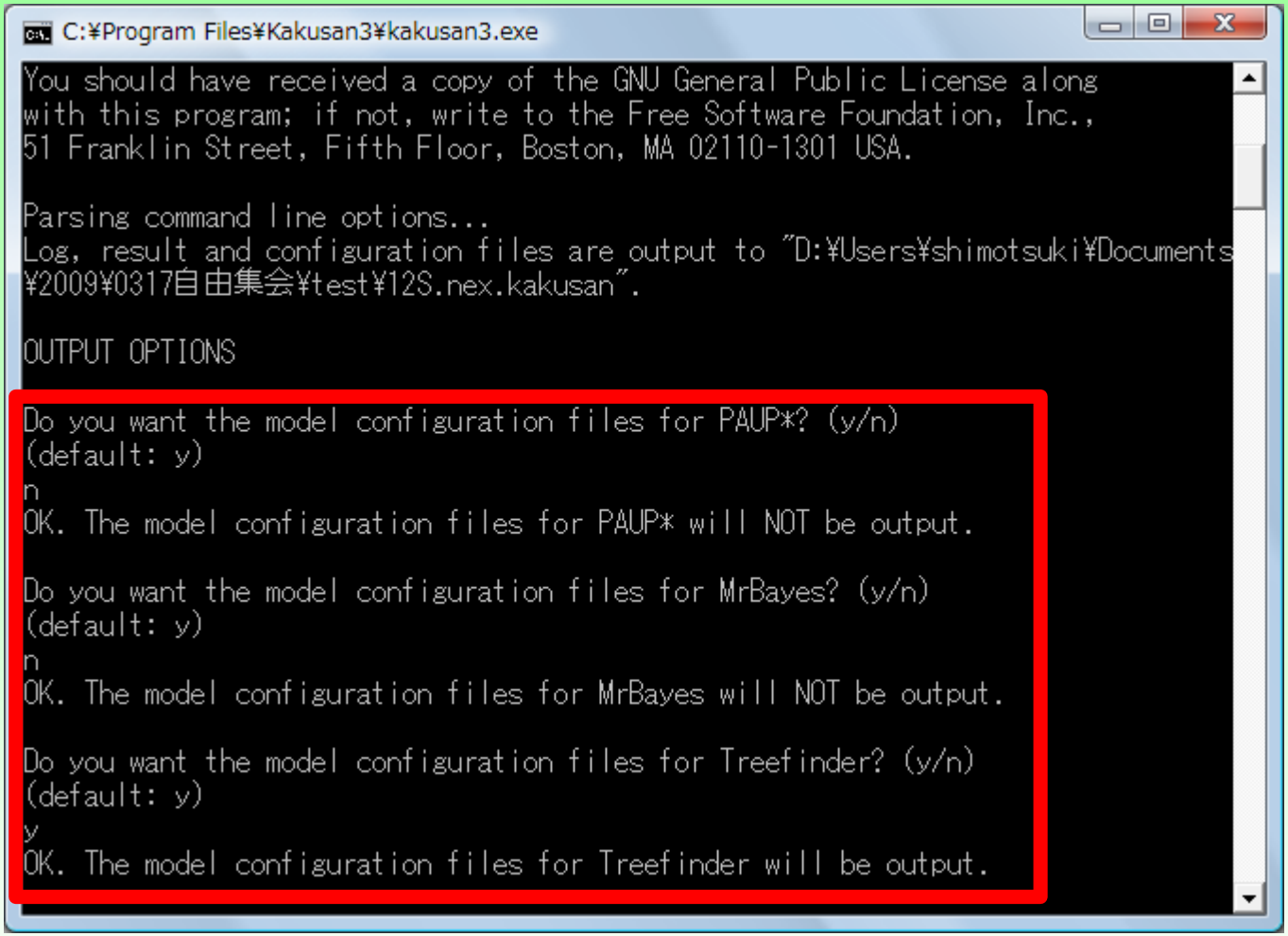
(XP以前)



設定ファイル出力対象の指定

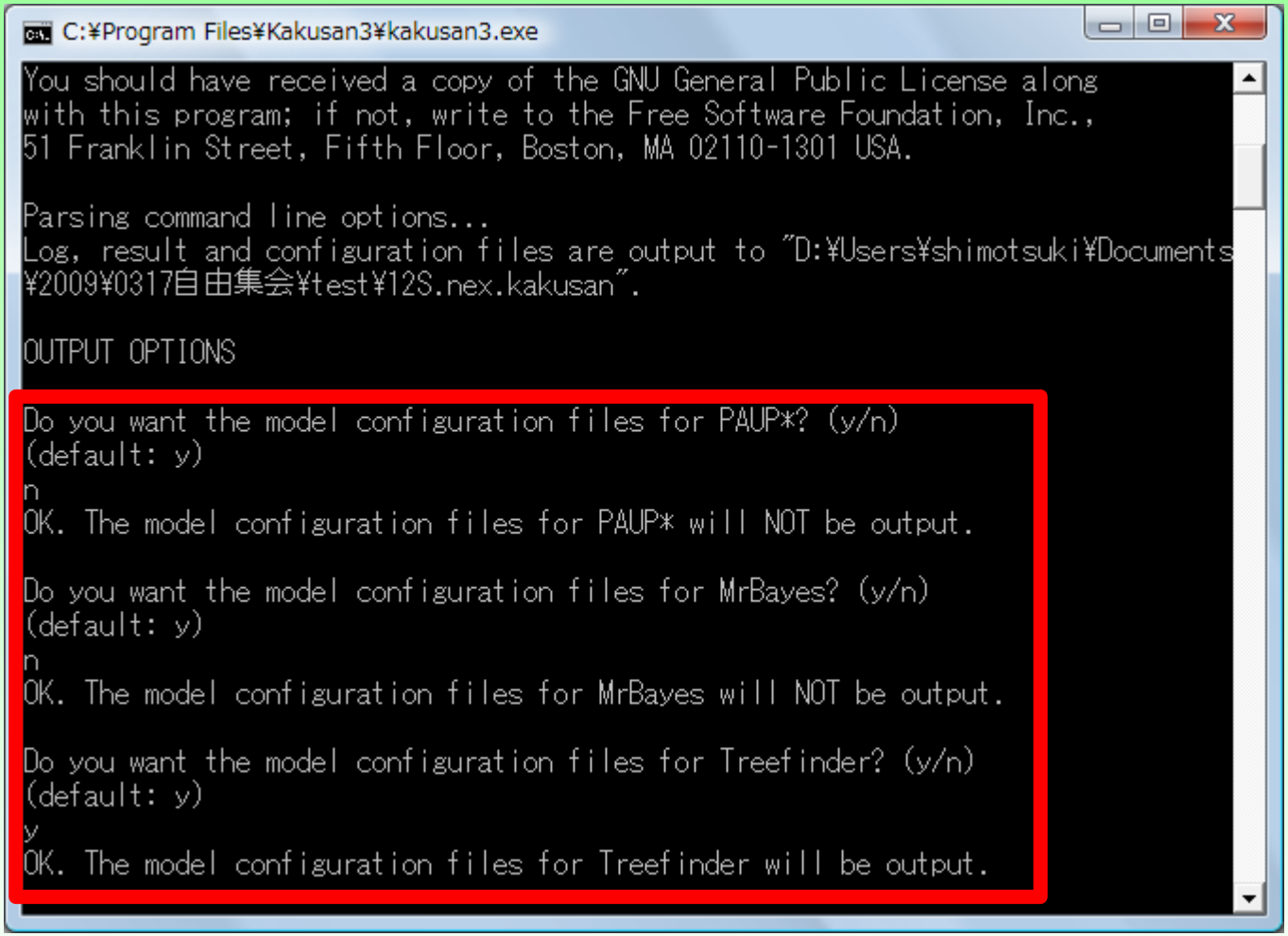


設定ファイル出力対象の指定



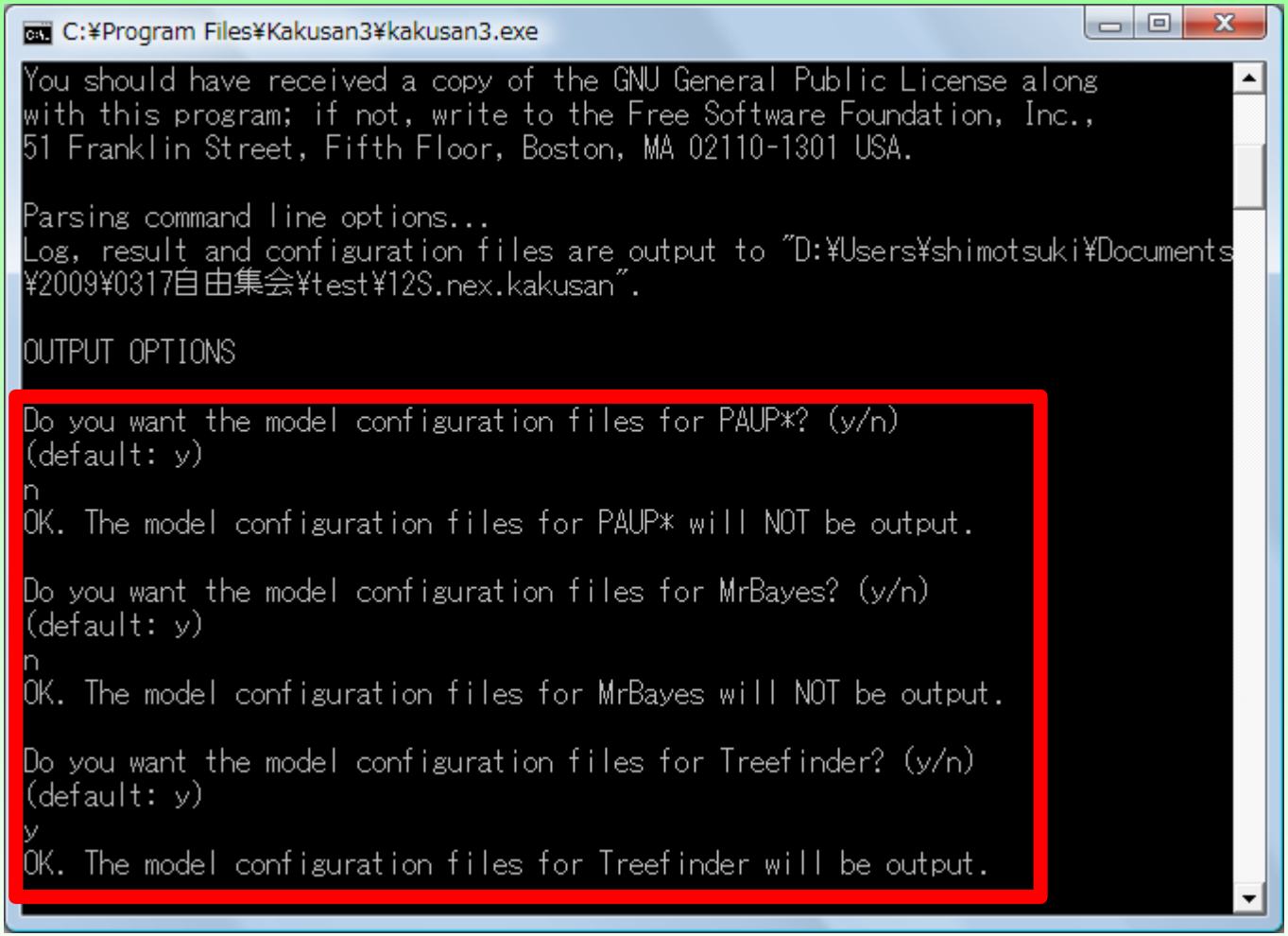
- PAUP*4・MrBayes3・Treefinderのそれぞれに選択されたモデルを当てはめて系統推定を行うための設定ファイルを出力できる

設定ファイル出力対象の指定



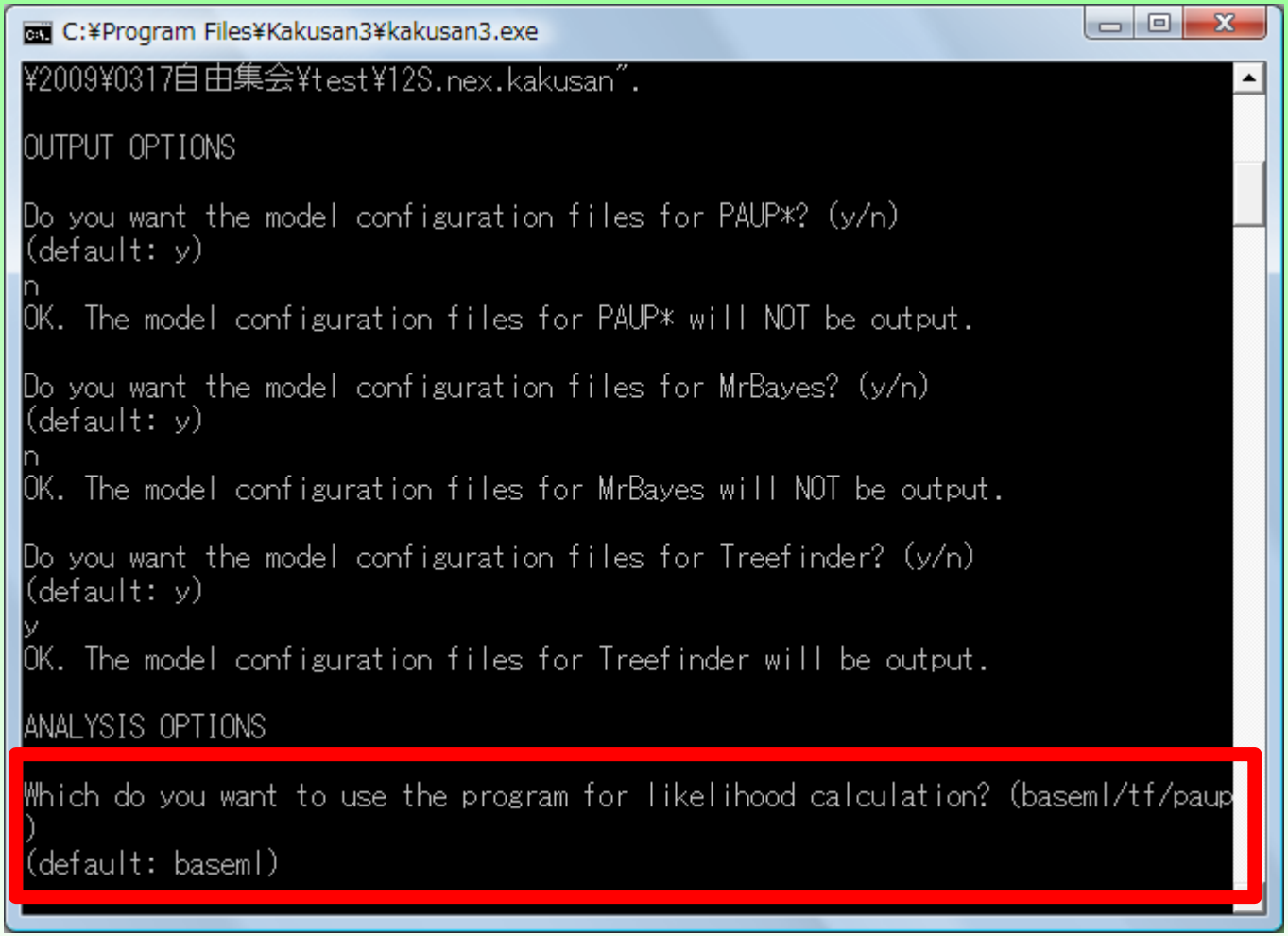
- PAUP*4・MrBayes3・Treefinderのそれぞれに選択されたモデルを当てはめて系統推定を行うための設定ファイルを出力できる
- 対象となるソフトの選択によって後の解析オプションが自動的に設定されます

設定ファイル出力対象の指定

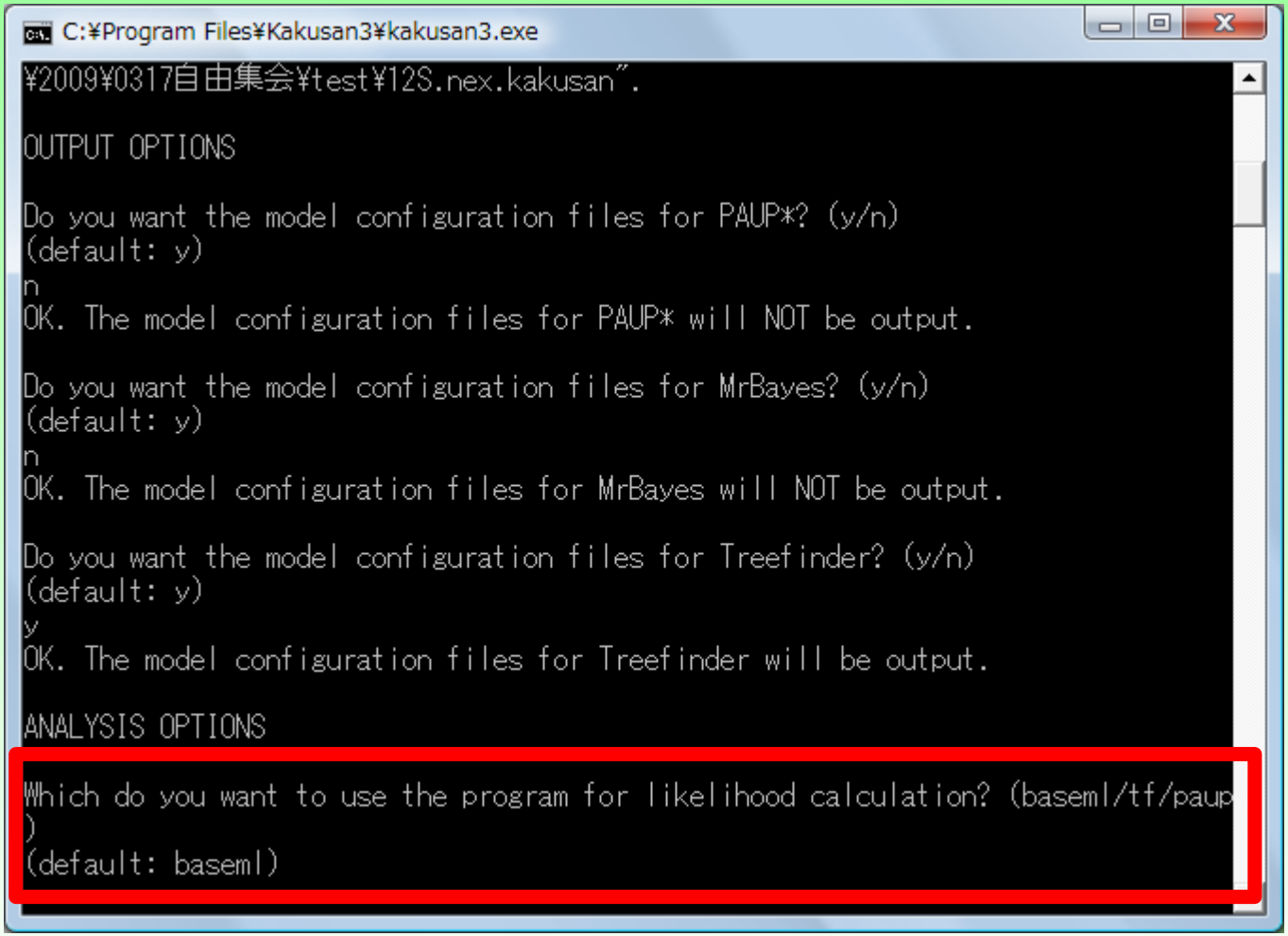


- PAUP*4・MrBayes3・Treefinderのそれぞれに選択されたモデルを当てはめて系統推定を行うための設定ファイルを出力できる
- 対象となるソフトの選択によって後の解析オプションが自動的に設定されます
- そのため、1度の解析ではどれか1つのソフト用出力のみを有効にすることが望ましい

尤度最大化に用いるプログラムの設定

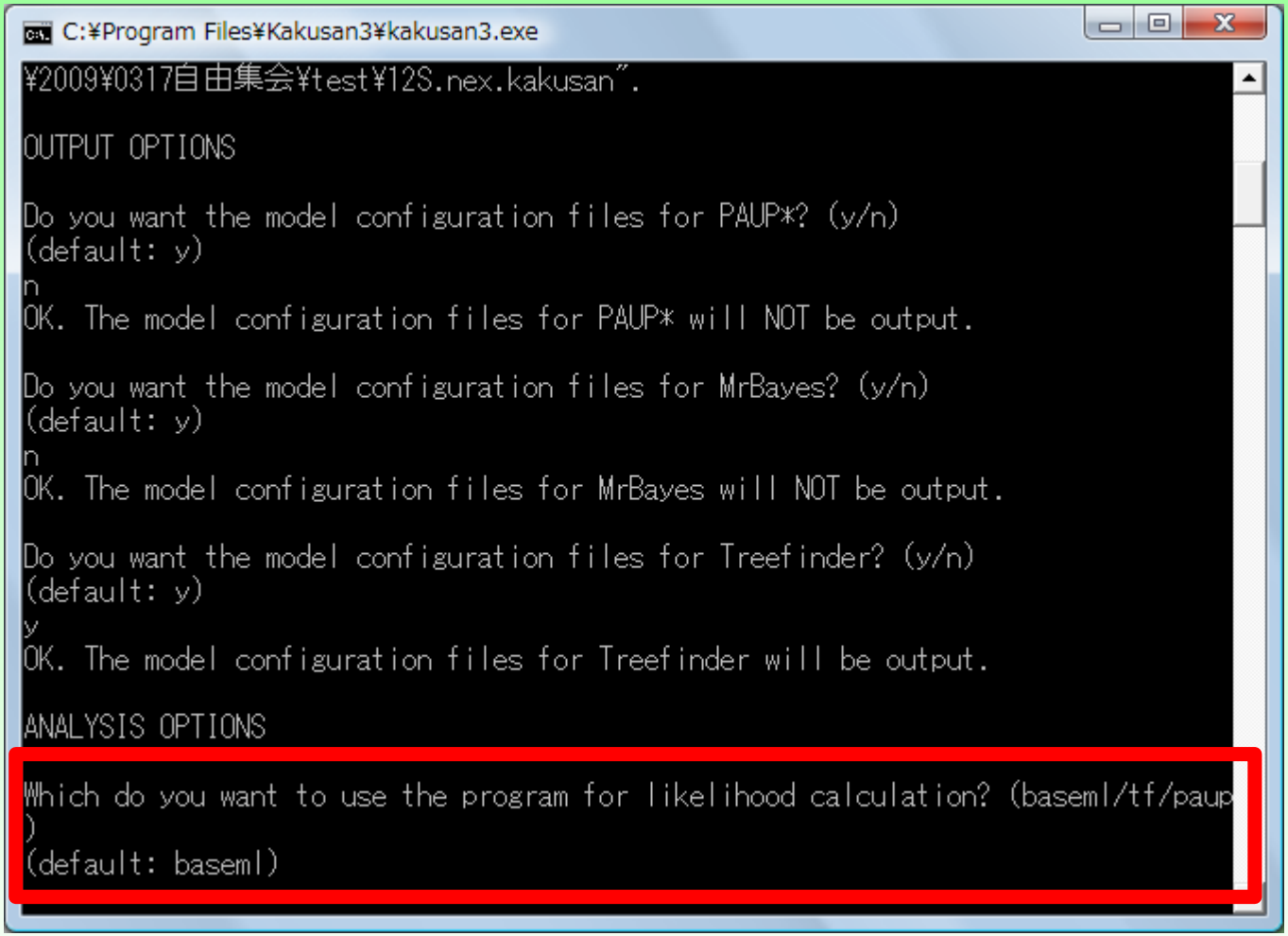


尤度最大化に用いるプログラムの設定



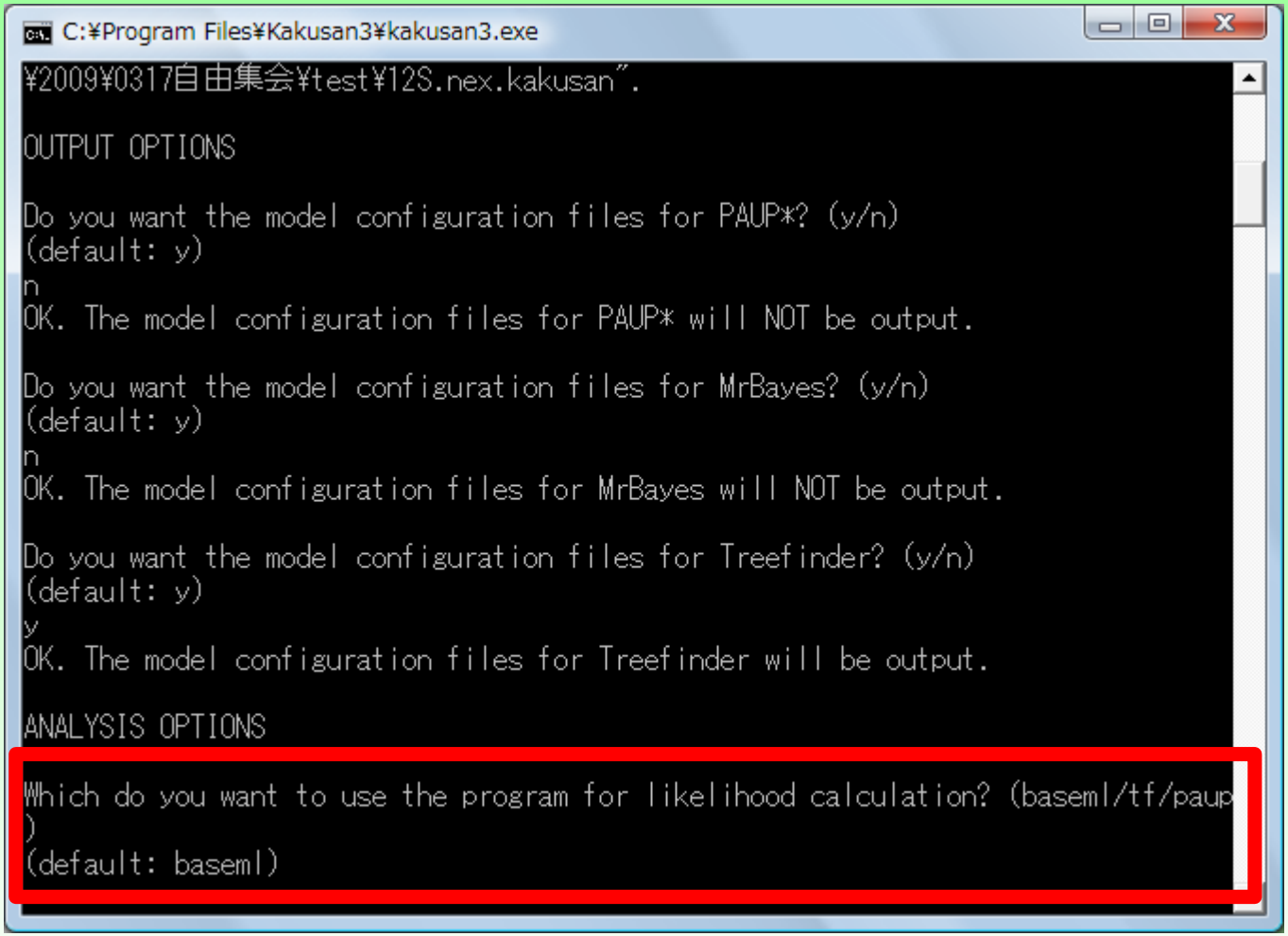
- PAUP*4用設定ファイルを書き出すならpaupを、

尤度最大化に用いるプログラムの設定



- PAUP*4用設定ファイルを書き出すならpaupを、
- MrBayes3用設定ファイルを書き出すならbasemlを、

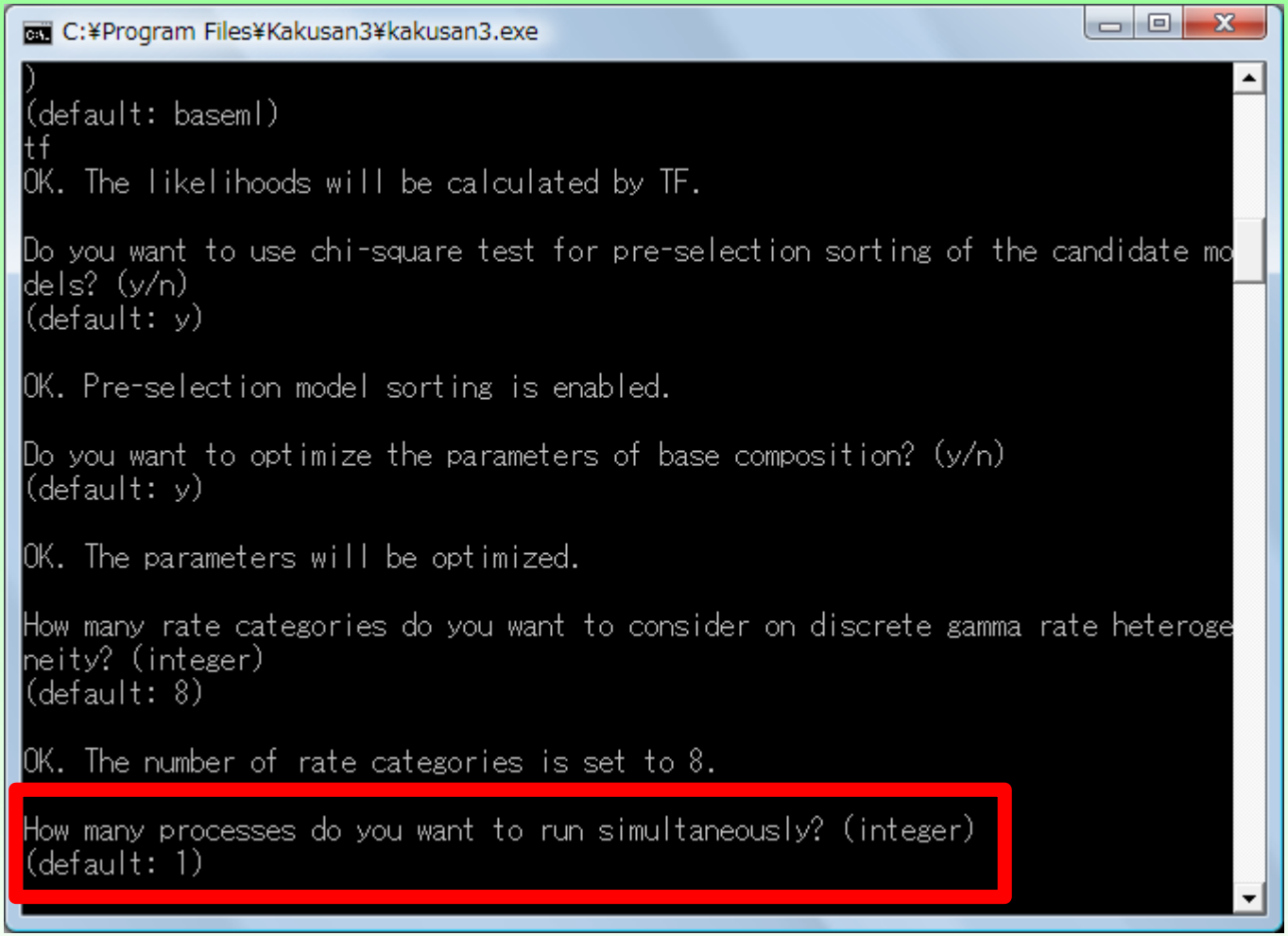
尤度最大化に用いるプログラムの設定



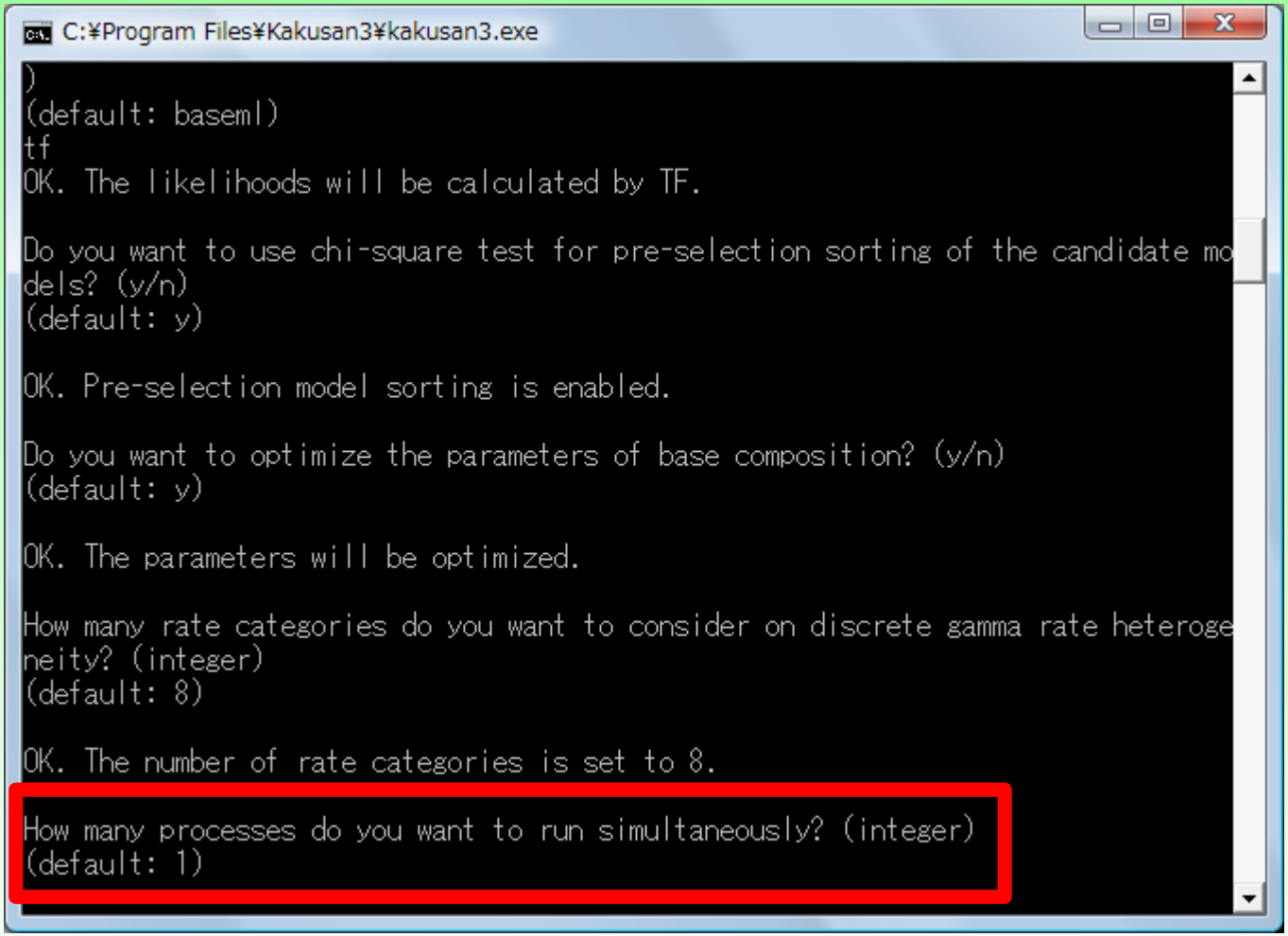
- PAUP*4用設定ファイルを書き出すならpaupを、
- MrBayes3用設定ファイルを書き出すならbasemlを、
- Treefinder用設定ファイルを書き出すならpaupかtfを

  指定することを推奨します

並列処理に関する設定

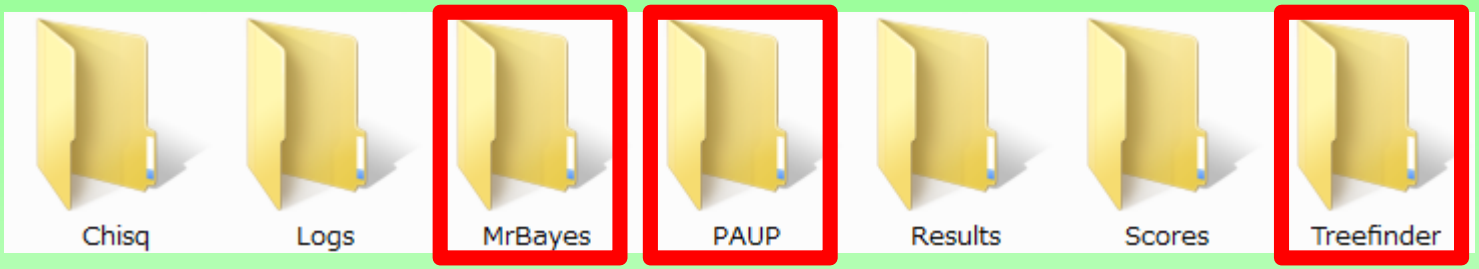

C:¥Program Files¥Kakusan3¥kakusan3.exe
)
(default: baseml)
tf
OK. The likelihoods will be calculated by TF.
Do you want to use chi-square test for pre-selection sorting of the candidate models? (y/n)
(default: y)
OK. Pre-selection model sorting is enabled.
Do you want to optimize the parameters of base composition? (y/n)
(default: y)
OK. The parameters will be optimized.
How many rate categories do you want to consider on discrete gamma rate heterogeneity? (integer)
(default: 8)
OK. The number of rate categories is set to 8.
How many processes do you want to run simultaneously? (integer)
(default: 1)

並列処理に関する設定



- 起動する並列プロセスの最大値を設定
- 搭載しているCPUの数以下にして下さい

出力フォルダの内容



- Chisq - 塩基組成のOTU間均一性の検定結果の出力フォルダ
- Logs - 各種実行ログファイルの保存フォルダ
- MrBayes - MrBayes3用設定ファイル出力フォルダ
- PAUP - PAUP*4用設定ファイル出力フォルダ
- Results - モデル選択結果のテキストファイル出力フォルダ
- Scores - 各モデルの尤度が記されたファイルが出力されるフォルダ
- Treefinder - Treefinder用設定ファイル出力フォルダ

どの設定ファイルを使うべきか - PAUP*4

どの設定ファイルを使うべきか - PAUP*4

- モデル選択規準 (AIC・AICc・BIC) とサンプルサイズ (1〜6)
  - AICc4 (サイト数をサンプルサイズとするAICc)

どの設定ファイルを使うべきか - PAUP*4

- モデル選択規準 (AIC・AICc・BIC) とサンプルサイズ (1〜6)
  - AICc4 (サイト数をサンプルサイズとするAICc)
- 領域の区分 (区分無し・遺伝子座・コドン位置) と枝長の比例・分離
  - そもそもPAUP*4は対応していないので検討不要

どの設定ファイルを使うべきか - MrBayes3

どの設定ファイルを使うべきか - MrBayes3

- モデル選択規準 (AIC・AICc・BIC) とサンプルサイズ (1〜6)
  - BIC4 (サイト数をサンプルサイズとするBIC)

どの設定ファイルを使うべきか - MrBayes3

- モデル選択規準 (AIC・AICc・BIC) とサンプルサイズ (1〜6)
  - BIC4 (サイト数をサンプルサイズとするBIC)
- 領域の区分 (区分無し・遺伝子座・コドン位置) と枝長の比例・分離
  - _proportional_codonshared (コード領域を含む複数領域データの場合)
    - 遺伝子座ごとに異なる塩基置換モデル・枝長は比例・コドン位置区分無し
  - _proportional (コード領域を含まない複数領域データの場合)
    - 遺伝子座ごとに異なる塩基置換モデル・枝長は比例
  - _codonshared (コード領域データの場合)
    - コドン位置区分無し
  - _single (非コード領域データの場合)
    - コドン位置区分無し

どの設定ファイルを使うべきか - Treefinder

どの設定ファイルを使うべきか - Treefinder

- モデル選択規準 (AIC・AICc・BIC) とサンプルサイズ (1〜6)
  - AICc4 (サイト数をサンプルサイズとするAICc)

どの設定ファイルを使うべきか - Treefinder

- モデル選択規準 (AIC・AICc・BIC) とサンプルサイズ (1〜6)
  - AICc4 (サイト数をサンプルサイズとするAICc)
- 領域の区分 (区分無し・遺伝子座・コドン位置) と枝長の比例・分離
  - _proportional_codonproportional (コード領域を含む複数領域データの場合)
    - 遺伝子座・コドン位置ごとに異なる塩基置換モデル・枝長は比例
  - _proportional (コード領域を含まない複数領域データの場合)
    - 遺伝子座ごとに異なる塩基置換モデル・枝長は比例
  - _codonproportional (コード領域データの場合)
    - コドン位置ごとに異なる塩基置換モデル・枝長は比例
  - _single (非コード領域データの場合)
    - コドン位置区分無し

まとめ

まとめ

- より正確な系統推定のためには、モデル選択で選ばれたモデルを当てはめる必要がある

まとめ

- より正確な系統推定のためには、モデル選択で選ばれたモデルを当てはめる必要がある
- 複数領域・コード領域データでは、領域ごと・コドン位置ごとに異なる塩基置換モデルを当てはめることでモデル選択規準が改善することが多い

まとめ

- より正確な系統推定のためには、モデル選択で選ばれたモデルを当てはめる必要がある
- 複数領域・コード領域データでは、領域ごと・コドン位置ごとに異なる塩基置換モデルを当てはめることでモデル選択規準が改善することが多い
- Kakusan3により領域ごとに最適なモデルを選択し、それを当てはめた解析が可能

まとめ

- より正確な系統推定のためには、モデル選択で選ばれたモデルを当てはめる必要がある
- 複数領域・コード領域データでは、領域ごと・コドン位置ごとに異なる塩基置換モデルを当てはめることでモデル選択規準が改善することが多い
- Kakusan3により領域ごとに最適なモデルを選択し、それを当てはめた解析が可能
- Kakusan3は単一・比例・分離モデル間の比較は行わないが、これらを比較したい場合はそれぞれのモデルを当てはめた系統推定結果を比較すればよい