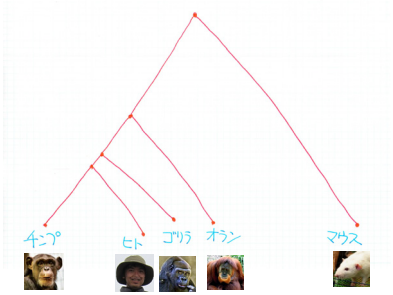
2010/06/30

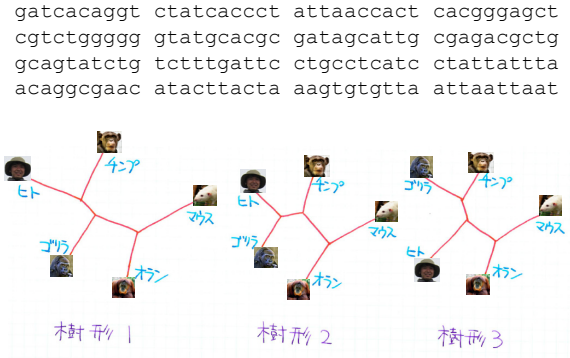
DNA情報からよみとる生物進化とランダムネス

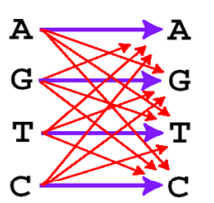
東京工業大学　情報理工学研究科
数理・計算科学専攻
下平英寿

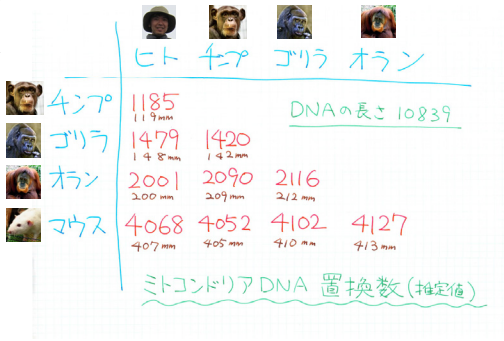
概要

- 系統樹： DNA配列　分子進化　工作



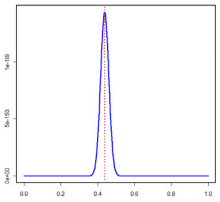


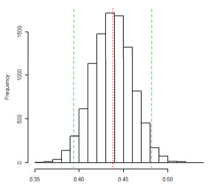


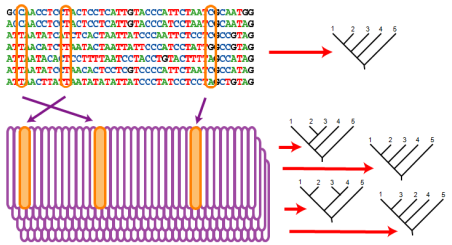


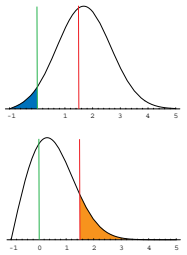
- 統計学： 最尤法　　ブートストラップ法

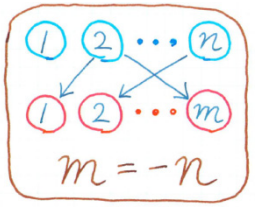
（未知量の推定方法）　　（ランダムネスを調べる方法）









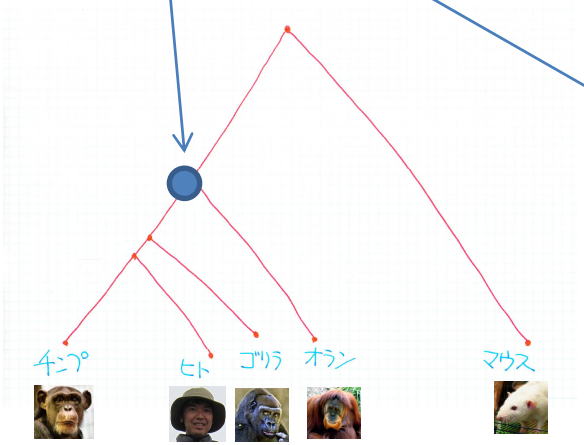


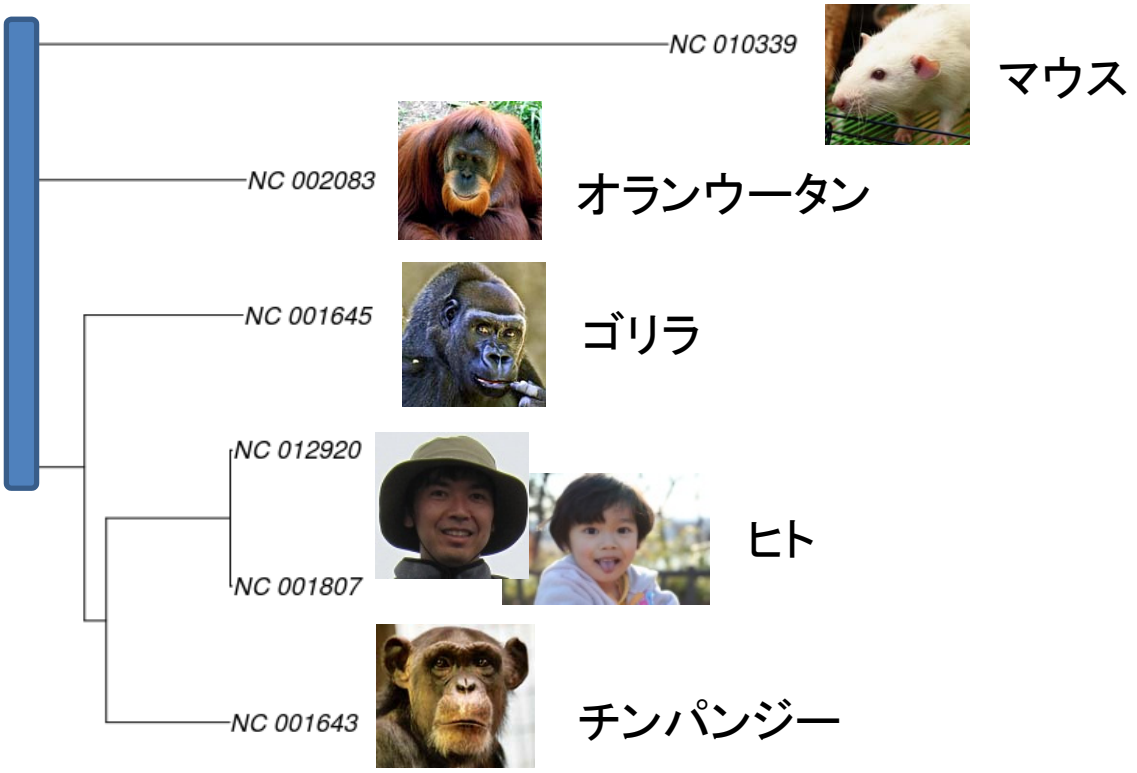
下平研の成果： Shimodaira-Hasegawa test, マルチスケール・ブートストラップ法

系統樹

類人猿の系統樹

類人猿の共通祖先
（1300万年前）

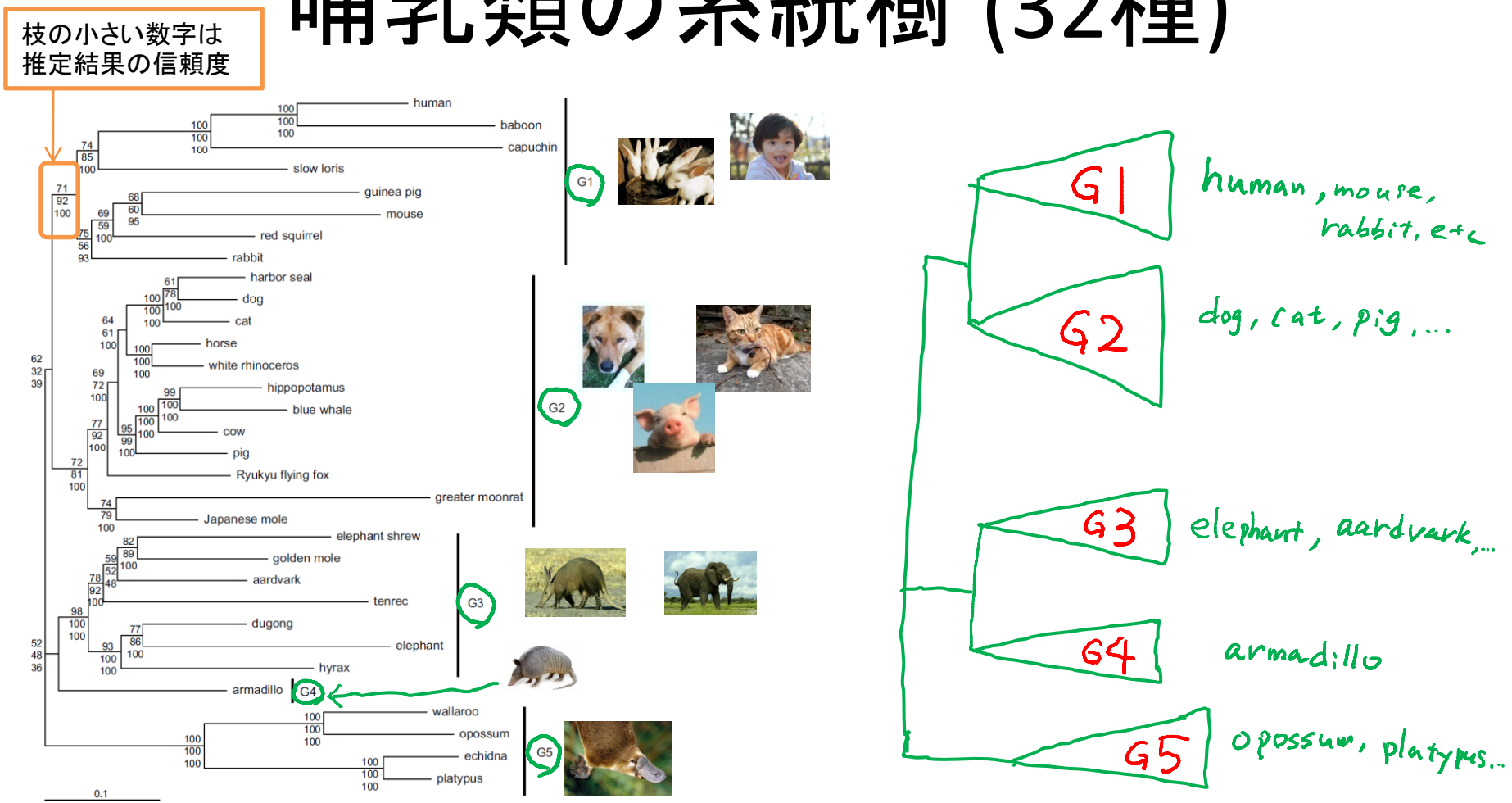




あとで工作します

NC_001807: Homo sapiens: human (old RefSeq)
NC_012920: Homo sapiens: human
NC_001643: Pan troglodytes: chimpanzee
NC_001645: Gorilla gorilla: Western Gorilla
NC_002083: Pongo abelii: Sumatran orangutan
NC_010339: Mus musculus musculus: eastern European house mouse

哺乳類の系統樹 (32種)

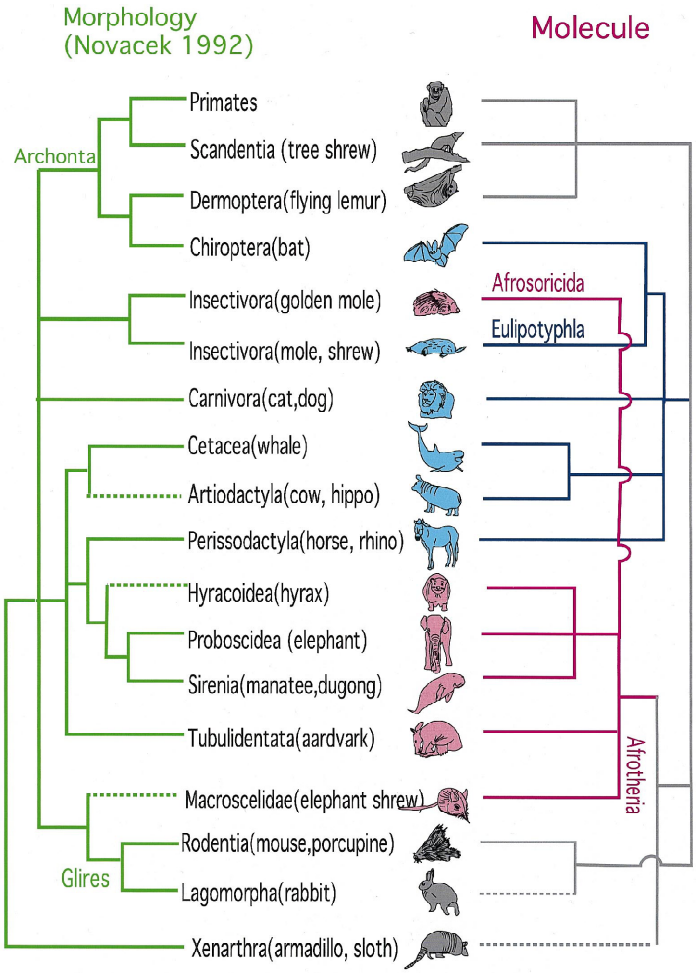


ML tree topology: ((G1,G2), (G3,G4), G5)

Fig 1 of Shimodaira and Hasegawa (2005) from the book (ed. Nielsen)
Data: mt protein sequences of n=3392 amino acids for s=32 species

見た目 vs 分子データ

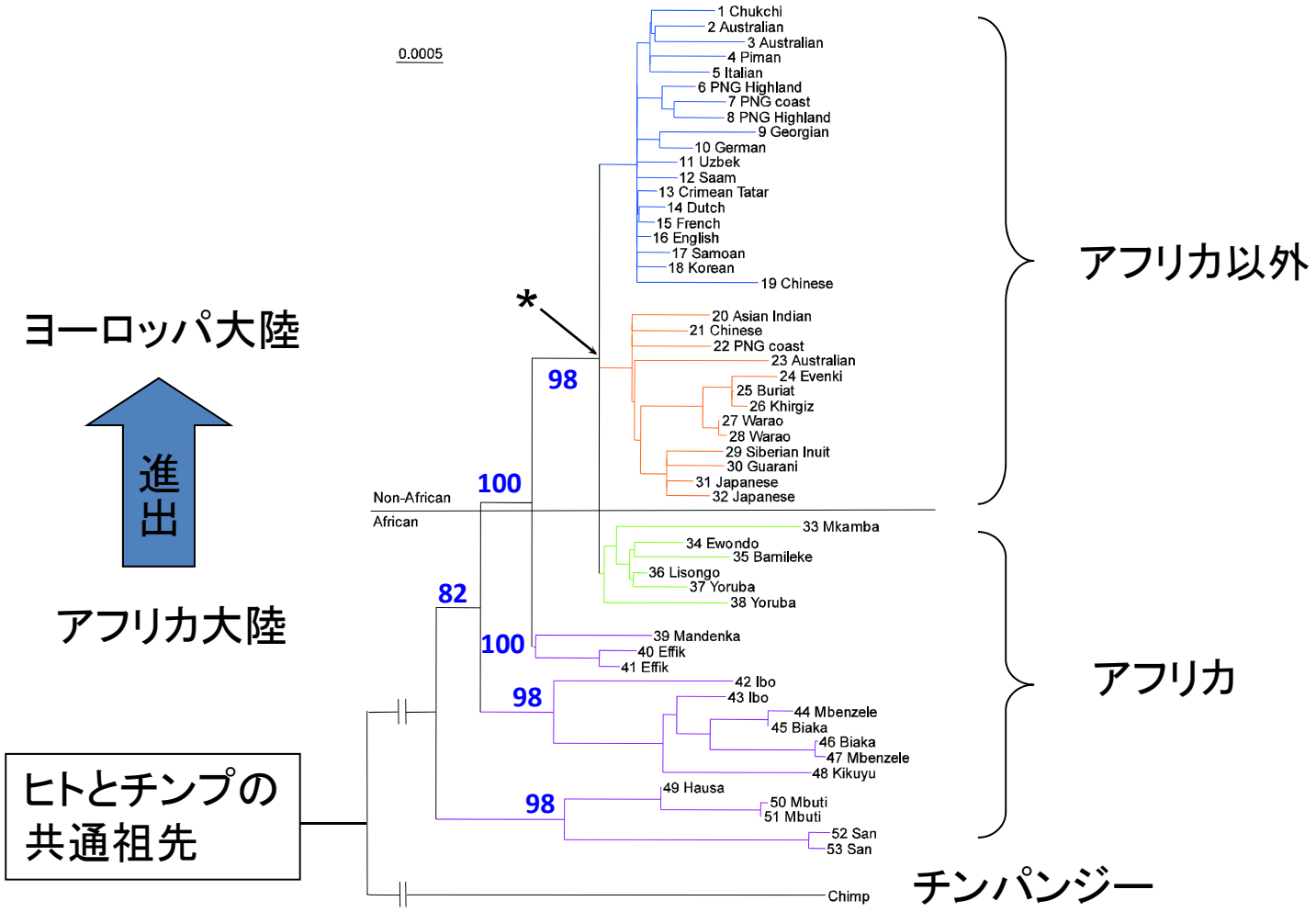
形態学による
古典的な系統樹



ミトコンドリアDNAから
統計手法（最尤法）
によって推定した系統樹

Cao et al. (2000) *Gene*

ヒトの起源はアフリカ



Ingman et al. (2000) *Nature*

ゲノムの情報量

- ある生物種の個体全体を完全な状態に保つために必要な遺伝情報の１セット
- 多数の塩基(A,T,G,C)が並んだDNA分子で遺伝子が記述される
- 1塩基(対)=2ビット，つまり4塩基＝1バイト

- ヒトゲノムは3G塩基 = 0.75Gバイト

(730円)

2Gバイトのメモリに
余裕で入る

(10709円)

2000Gバイトのハードディスクに2500人
のゲノムが圧縮無しでそっくり入る

DNAは小さな細胞核に有ることを考えれば，驚くことではない？

分類	生物種	塩基数	総遺伝子数
哺乳類	Homo sapiens (ヒト)	3Gb	30000
昆虫	Drosophila melanogaster (ショウジョウバエ)	120Mb	13000
植物	Arabidopsis thaliana (シロイヌナズナ)	125Mb	25498
植物	Oryza sativa （イネ）	430Mb	50000?
線虫	Caenorhabditis elegans （線虫）	97Mb	18000
酵母	Saccharomyces cerevisiae （出芽酵母）	12Mb	6286
細菌	Bacillus subtilis （枯草菌）	4.21Mb	4100
細菌	Escherichia coli K12 MG1655（大腸菌）	4.64Mb	4289
細菌	Escherichia coli O157 Sakai（大腸菌 O157）	5.5Mb	5361
細菌	Haemophilus influenzae Rd（インフルエンザ菌）	1.83Mb	1709
細菌	Helicobacter pylori J99（ピロリ菌）	1.64Mb	1491
その他	ミトコンドリア	16.6Kb	13

DNAデータの入手 (NCBI)

http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Tree&id=6072&lvl=3&p=genome&lin=f&keep=1&srchmode=1&unlock

National Center for Biotechnology Information (NCBI)のデータベース

リンクをたどっていく （脊椎動物）

http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Tree&id=7742&lvl=3&p=genome&p=mapview&p=has_linkout&p=blast_url&p=genome_blast&lin=f&keep=1&srchmode=1&unlock

NCBI Taxonomy Browser

Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search for as complete name lock Go Clear

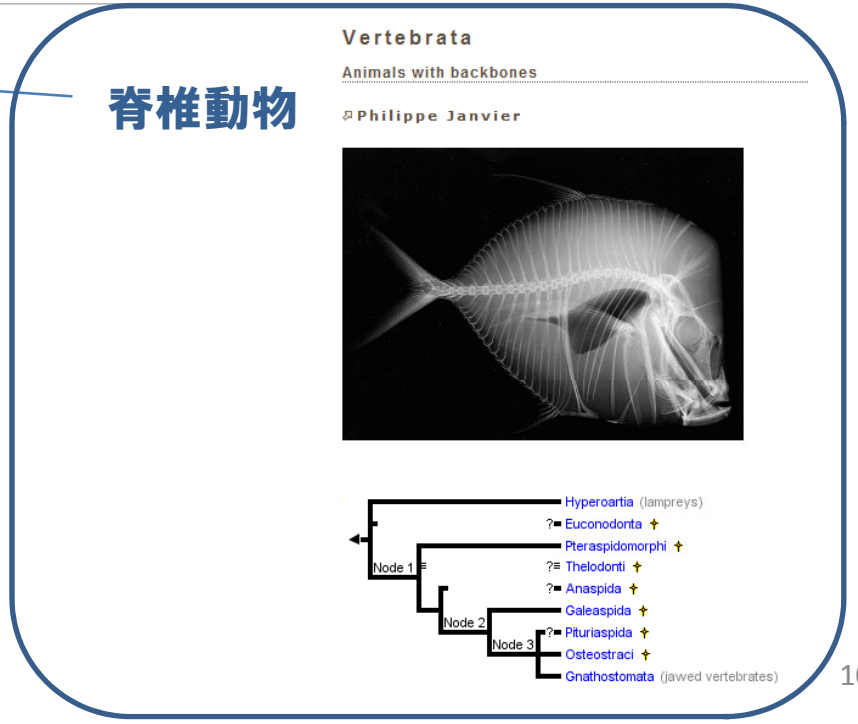
Display 3 levels using filter: none

Nucleotide Nucleotide EST Nucleotide GSS Protein Structure Genome Sequences Genome Projects
Popset SNP 3D Domains Domains GEO Datasets GEO Expressions UniGene
UniSTS PubMed Central Gene HomoloGene OMIA SRA Experiments MapView
LinkOut BLAST TRACE

Lineage (full): root; cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Craniata

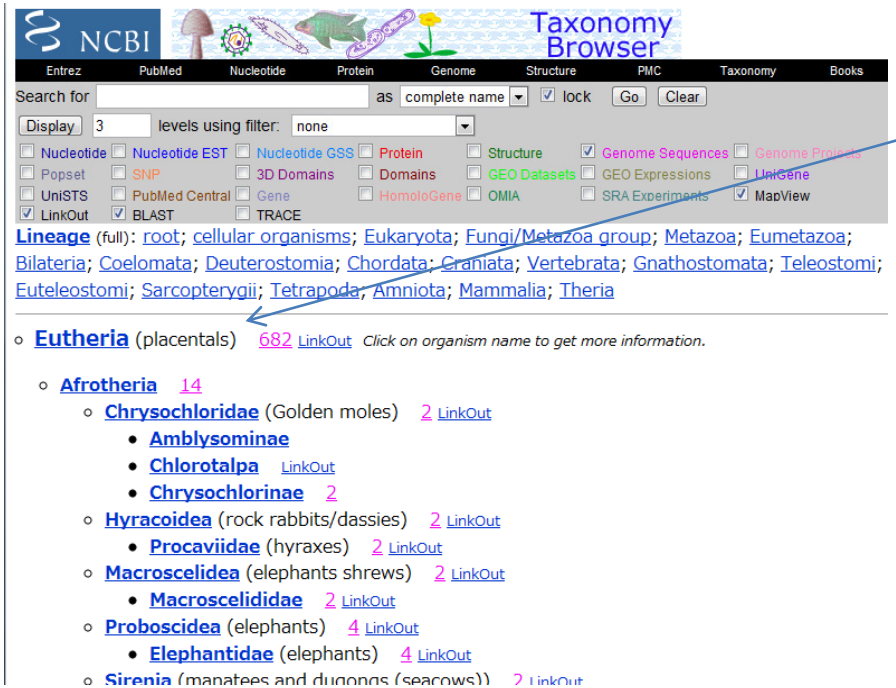
- **Vertebrata** (vertebrates) 1,911 LinkOut *Click on organism name to get more information.*
  - **Gnathostomata** (jawed vertebrates) 1,909 LinkOut
    - **Chondrichthyes** (cartilaginous fishes) 10 LinkOut
      - **Elasmobranchii** (elasmobranchs) 9 LinkOut
      - **Holocephali** 1 LinkOut
    - **Teleostomi** 1,899
      - **Euteleostomi** (bony vertebrates) 1,899
  - **Hyperoartia** (fish) 2 LinkOut
    - **Petromyzontiformes** 2 LinkOut
      - **Petromyzontidae** (lampreys) 2 LinkOut

脊椎動物

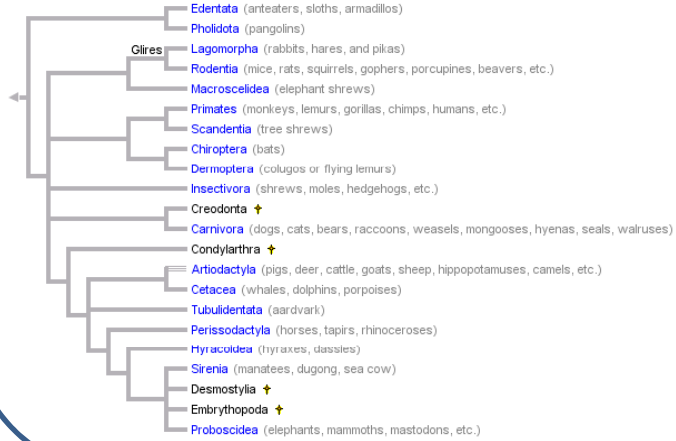


さらにたどって（真獣下綱）

http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Tree&id=9347&lvl=3&p=genome&p=mapview&p=has_linkout&p=blast_url&p=genome_blast&lin=f&keep=1&srchmode=1&unlock



真獣下綱：胎盤を持つ哺乳類



やっとつきました （ヒト上科）

http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Tree&id=9604&lvl=3&p=genome&p=mapview&p=has_linkout&p=blast_url&p=genome_blast&lin=f&keep=1&srchmode=1&unlock

NCBI Taxonomy Browser

Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search for as complete name lock Go Clear

Display 3 levels using filter: none

Nucleotide Nucleotide EST Nucleotide GSS Protein Structure Genome Sequences Genome Projects
Popset SNP 3D Domains Domains GEO Datasets GEO Expressions UniGene
UniSTS PubMed Central Gene HomoloGene OMIA SRA Experiments MapView
LinkOut BLAST TRACE

Lineage (full): root; cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Euarchontoglires; Primates; Haplorrhini; Simiiformes; Catarrhini; Hominoidea

- **Hominidae** (great apes) 132 LinkOut *Click on organism name to get more information.*
  - **Homininae** 106 LinkOut
    - **Gorilla** 2 LinkOut
      - **Gorilla beringei** (Eastern Gorilla) LinkOut
      - **Gorilla gorilla** (Western Gorilla) 2 LinkOut

ゴリラ

    - **Homo** 76 LinkOut
      - **Homo sapiens** (human) 75 MapView LinkOut BLAST page
      - **Homo sp. Altai** (Denisova hominin) 1

ヒト

    - **Pan** (chimpanzees) 28 LinkOut
      - **Pan paniscus** (pygmy chimpanzee) 1 LinkOut
      - **Pan troglodytes** (chimpanzee) 27 MapView LinkOut BLAST page
      - **Pan sp. WO2007026120**

チンパンジー

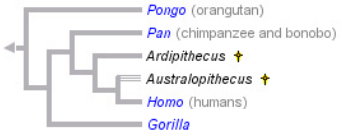
  - **Ponginae** 26 LinkOut
    - **Pongo** 26 LinkOut
      - **Pongo abelii** (Sumatran orangutan) 25 LinkOut BLAST page
      - **Pongo abelii x pygmaeus** LinkOut
      - **Pongo pygmaeus** (Bornean orangutan) 1 LinkOut
      - **Pongo sp.** LinkOut

オランウータン

ヒト上科

Hominidae

Humans, great apes, and their extinct relatives



Homo sapiens (ヒト)

NCBI Taxonomy Browser

Entrez | PubMed | Nucleotide | Protein | Genome | Structure | PMC | Taxonomy | Books

Search for [ ] as complete name ☑ lock Go Clear

Display 3 levels using filter: none

Homo sapiens

Taxonomy ID: 9606
Genbank common name: **human**
Inherited blast name: **primates**
Rank: species
Genetic code: Translation table 1 (Standard)
Mitochondrial genetic code: Translation table 2 (Vertebrate Mitochondrial)
Other names:
common name: **man**
authority: **Homo sapiens Linnaeus, 1758**

Lineage(full)
cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Euarchontoglires; Primates; Haplorrhini; Simiiformes; Catarrhini; Hominoidea; Hominidae; Homininae; Homo

Entrez records

Database name	Subtree links	Direct links
Nucleotide	16,985,984	16,985,959
Nucleotide EST	8,301,515	8,301,515
Nucleotide GSS	1,293,831	1,292,505
Protein	531,420	531,324
Structure	15,730	15,730
Genome Sequences	75	74
Genome Projects	32	32
Popset	18,170	18,170
SNP	31,020,464	31,020,464
3D Domains	61,175	61,175
Domains	8	8
GEO Datasets	9,564	9,564
GEO Expressions	17,689,684	17,689,684
UniGene	123,200	123,200
UniSTS	327,522	327,522
PubMed Central	7,918	7,915
Gene	42,645	42,608
HomoloGene	18,876	18,876
SRA Experiments	8,821	8,821
Taxonomy	2	1

Genome Information

See the NCBI Genome homepage
Go to NCBI genomic BLAST page for Homo sapiens

Genome view: 24 chromosomes

Names 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y

See the Mitochondrion Genome
See the TRACE Assembly

Trace records (raw single-pass reads of DNA sequence)

Sequencing Center Name

Record counts per type

ヒトゲノムの選択ページ

NCBI Genome

All Databases | PubMed | Nucleotide | Protein | Genome | Structure | OMIM | PMC | Journals | Books

Search Genome for txid9606[Organism:exp] Go Clear Save Search

Limits | Preview/Index | History | Clipboard | Details

Display Summary Show 20 Send to

All: 75

Items 1 - 20 of 75 Page 1 of 4 Next

1: NC_012920
Homo sapiens mitochondrion, complete genome
dsDNA; circular; Length: **16,569 nt**
Organelle: **mitochondrion**
Created: **2009/07/08**

ミトコンドリアを選択

2: AC_000156
Homo sapiens chromosome Y, alternate assembly HuRef, whole genome shotgun sequence
DNA; linear; Length: **19,317,006 nt**
Replicon Type: **chromosome**
Replicon Name: **Y**
Created: **2007/09/27**

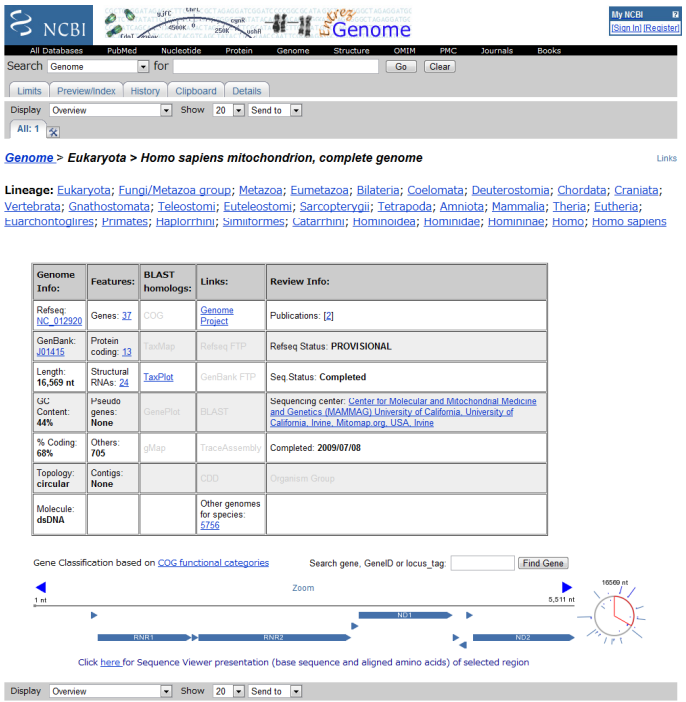
3: AC_000155
Homo sapiens chromosome X, alternate assembly HuRef, whole genome shotgun sequence
DNA; linear; Length: **143,733,266 nt**
Replicon Type: **chromosome**
Replicon Name: **X**
Created: **2007/09/27**

4: AC_000154
Homo sapiens chromosome 22, alternate assembly HuRef, whole genome shotgun sequence
DNA; linear; Length: **34,107,095 nt**
Replicon Type: **chromosome**

Recent activity
Turn Off Clear
IGKV1ORY-1 immunoglobulin kappa variable 1/ORY-1
NC_011137 Homo sapiens neanderthalensis mitochondrion,
» See more...

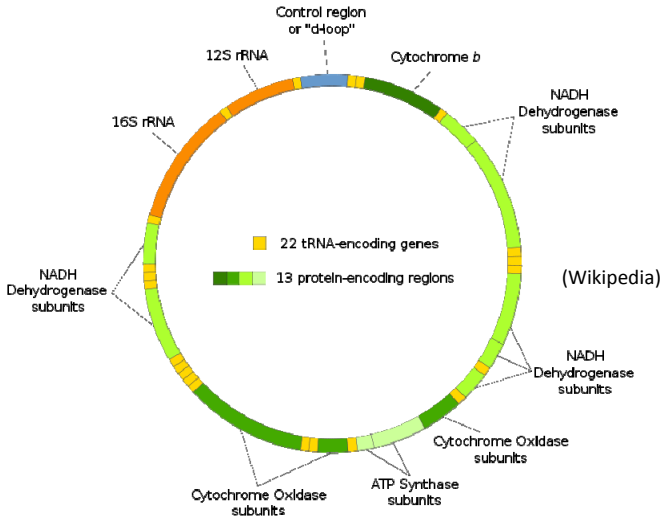
ヒト　ミトンドリア　ゲノム

http://www.ncbi.nlm.nih.gov/sites/entrez?Db=genome&Cmd=ShowDetailView&TermToSearch=24730

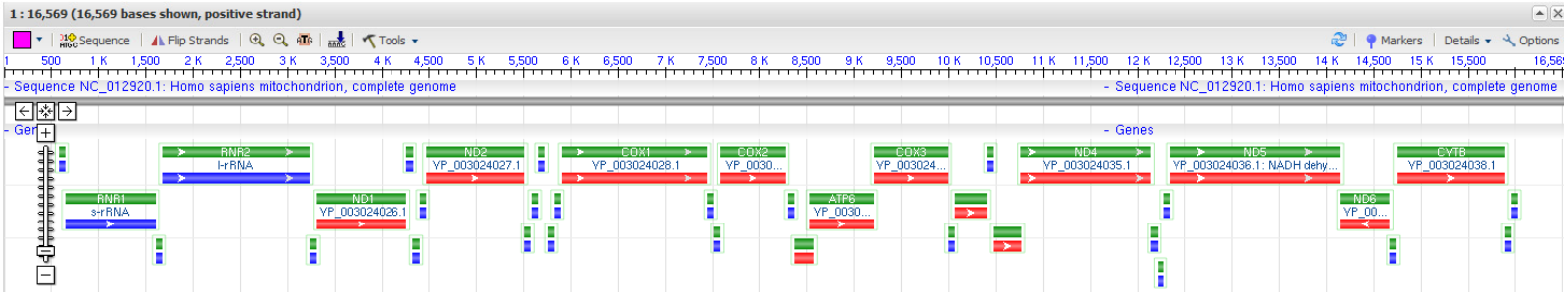


環状ゲノムで長さは16000文字程度
このデータは16569 bp

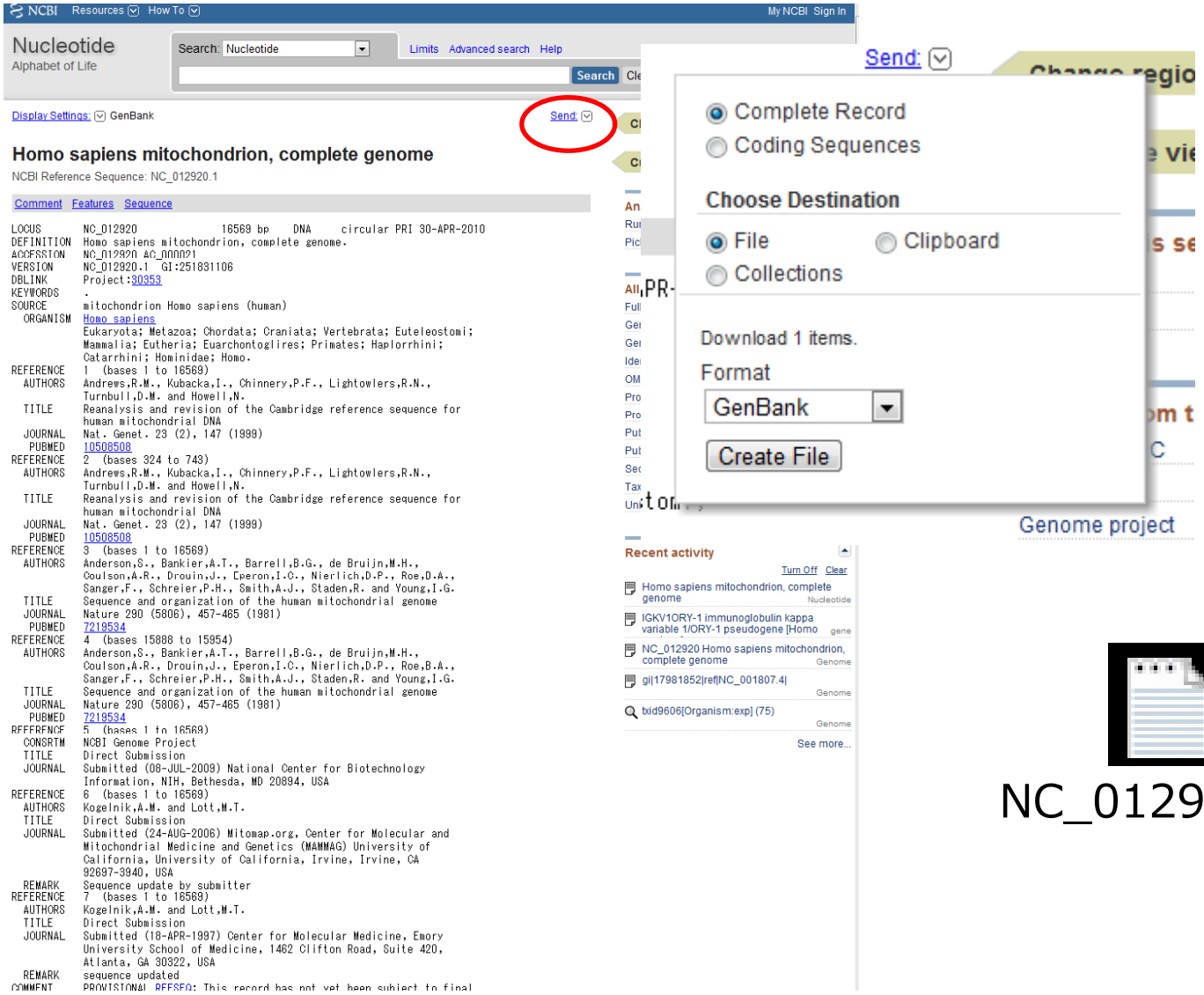
遺伝子の名前：　ND1, ND2, COX1, COX2, …



(Wikipedia)



NC_012920



NC_012920.txt

今回は6本のDNA配列を用意

NC_001807: Homo sapiens: human ヒトA (古いRefSeq)
NC_012920: Homo sapiens: human ヒトB (現在のRefSeq)
NC_001643: Pan troglodytes: chimpanzee チンパンジー
NC_001645: Gorilla gorilla: Western Gorilla ゴリラ
NC_002083: Pongo abelii: Sumatran orangutan オランウータン
NC_010339: Mus musculus musculus: eastern European house mouse マウス

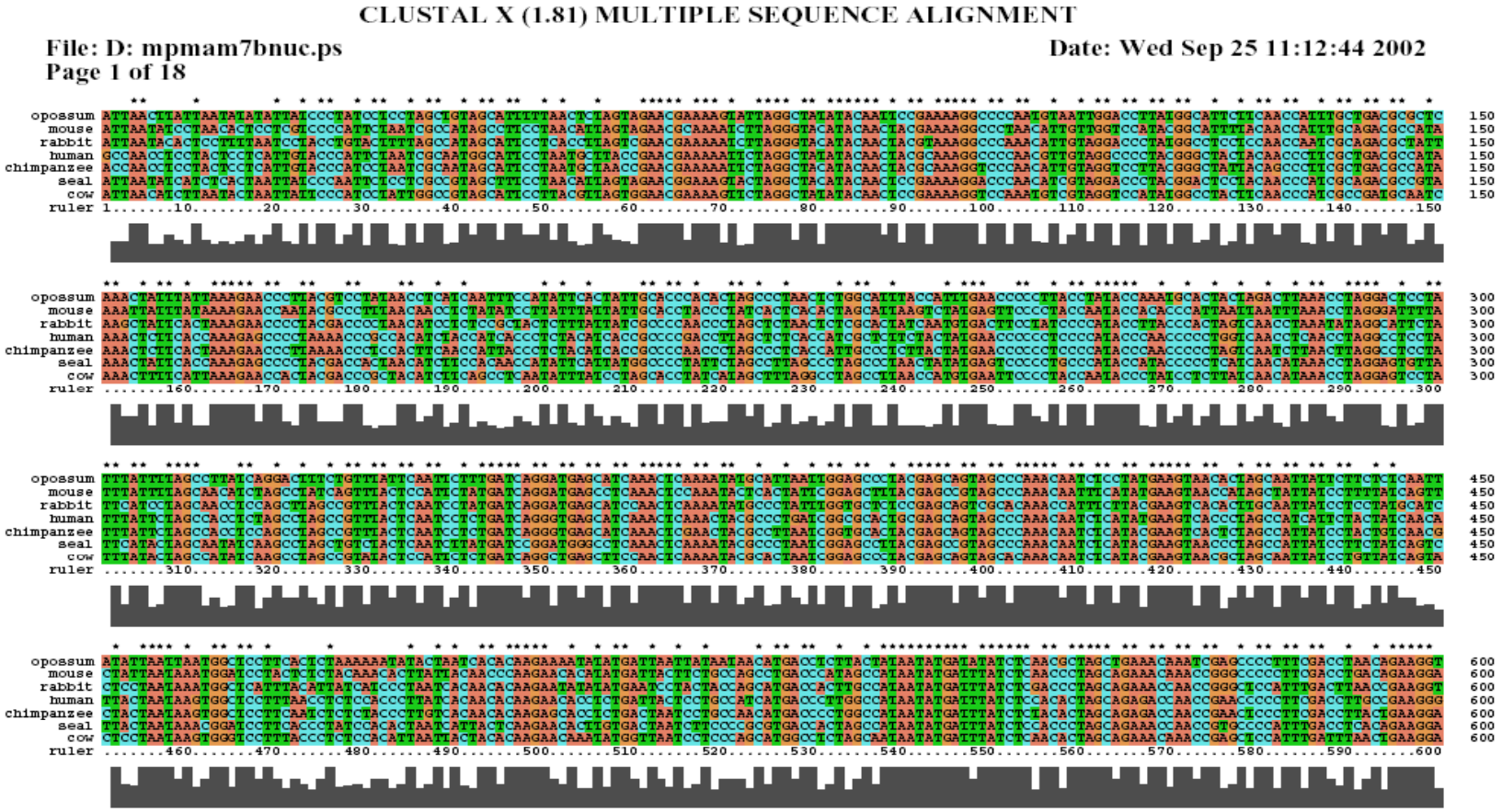
work-conc-nuc.txt

16000 bp 程度あるが，そのうち 10839 bpを利用する.
遺伝子コーディングリージョンを使う.

12個の遺伝子 ND1 ND2 COX1 COX2 ATP8 ATP6 COX3 ND3 ND4L ND4 ND5 CYTB

```
CONSENSUS  ATACCCATG. CCAACCTCCT ACTCCTCATT GTACCCAT.C TAATCGC.AT AGCATTCCTA
NC_001807  .........G .......... .......... ........T. .......A.. G.........
NC_012920  .........G .......... .......... ........T. .......A.. G.........
NC_001643  .C.......A .......... .......... ........C. .......A.. ..........
NC_001645  ...T.....G .T.....T.. .......... .....T..C. .......C.. ..........
NC_002083  ..G.....AA T......... .......... A..T....C. .......C.. ......T...
NC_010339  G.GTT.T.TA TT..TA.... .ACA...C.C ..C..T..T. .......T.. ...C......
                 10          20          30          40          50          60
CONSENSUS  ATGCTAACCG AACGAAAAAT TCTAGGCTAT ATACAACTAC GCAAAGGCCC CAAC.TTGTA
NC_001807  .....T.... .......... .......... .......... .......... ....G.....
NC_012920  .....T.... .......... .......... .......... .......... ....G.....
NC_001643  .......... .......... .........C .......... .......T.. ....A.....
NC_001645  .......... .......... .T........ .......... .T........ ....G.C...
NC_002083  .......... .......... C......C.C .C........ .......G.. ....A....G
NC_010339  .CA...GTA. ....C..... CT....G... .......... .A........ T...A.....
```

DNA配列データ



7種の哺乳類のミトコンドリアDNA配列の一部

塩基配列の解読法

(Wikipedia)

塩基3文字（コドン） → 翻訳 → アミノ酸1個 → アミノ酸を一列につなげたものがタンパク質

情報

物質

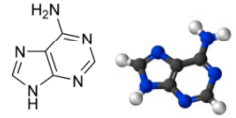
TTT → フェニルアラニン (F または Phe)

GAA → グルタミン酸 （E または Glu)

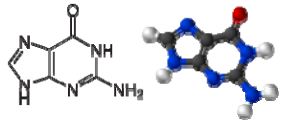
work-conc-nuc.txt　TTTGAACCTGGGGAT… → FEPGD…　work-conc-ptn.txt

塩基は4種類（2グループ）

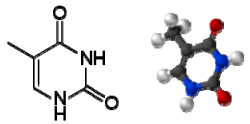
アデニン (A)



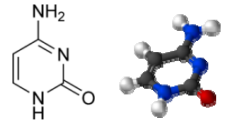
グアシン (G)



チミン (T)



シトシン (C)



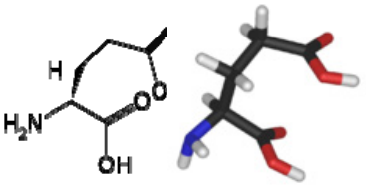
遺伝暗号　変換表2（脊椎動物ミトコンドリア）

TTT F Phe	TCT S Ser	TAT Y Tyr	TGT C Cys
TTC F Phe	TCC S Ser	TAC Y Tyr	TGC C Cys
TTA L Leu	TCA S Ser	TAA * Ter	TGA W Trp
TTG L Leu	TCG S Ser	TAG * Ter	TGG W Trp
CTT L Leu	CCT P Pro	CAT H His	CGT R Arg
CTC L Leu	CCC P Pro	CAC H His	CGC R Arg
CTA L Leu	CCA P Pro	CAA Q Gln	CGA R Arg
CTG L Leu	CCG P Pro	CAG Q Gln	CGG R Arg
ATT I Ile i	ACT T Thr	AAT N Asn	AGT S Ser
ATC I Ile i	ACC T Thr	AAC N Asn	AGC S Ser
ATA M Met i	ACA T Thr	AAA K Lys	AGA * Ter
ATG M Met i	ACG T Thr	AAG K Lys	AGG * Ter
GTT V Val	GCT A Ala	GAT D Asp	GGT G Gly
GTC V Val	GCC A Ala	GAC D Asp	GGC G Gly
GTA V Val	GCA A Ala	GAA E Glu	GGA G Gly
GTG V Val i	GCG A Ala	GAG E Glu	GGG G Gly

アミノ酸20種類

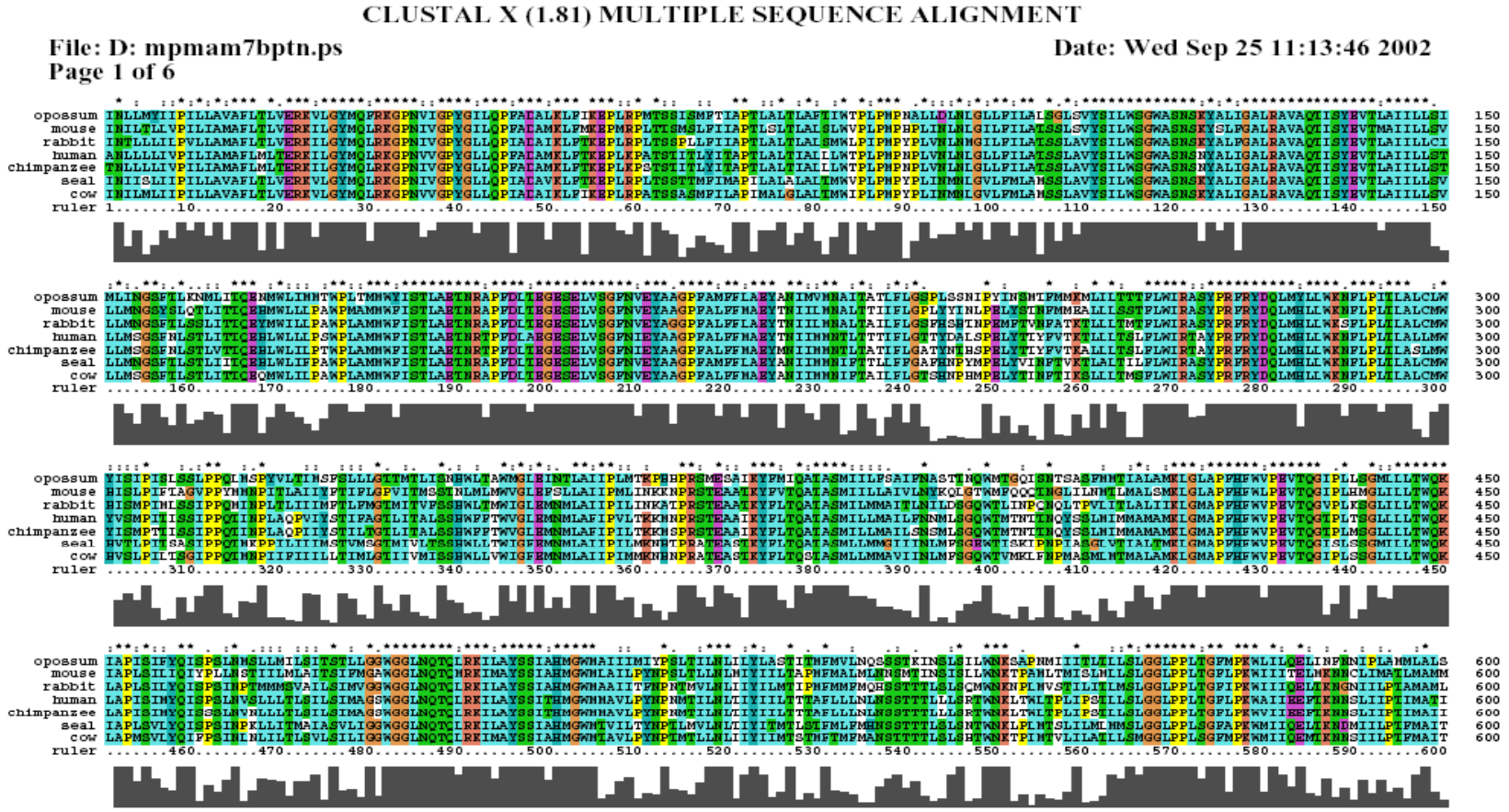
アミノ酸の例

グルタミン酸



(wikipediaより)

アミノ酸配列データ



7種の哺乳類のミトコンドリア・アミノ酸配列の一部

本講演中の哺乳類の系統樹分析は，このアミノ酸配列データを使用した

塩基の変異数
（塩基の違いを数えたもの）

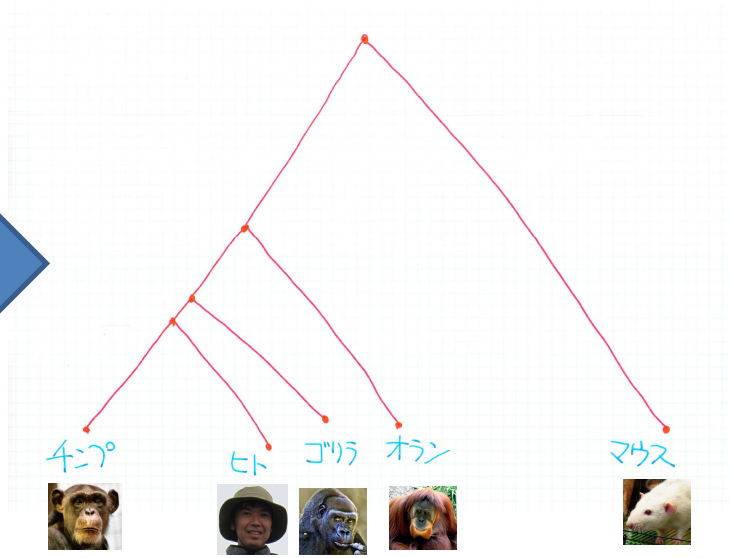
DNA配列の長さ： 10839





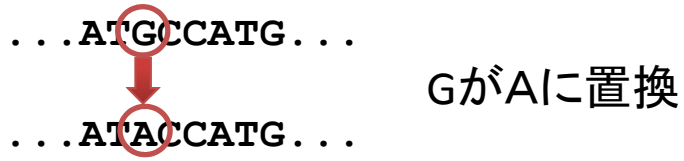
	NC_001807	NC_012920	NC_001643	NC_001645	NC_002083
NC_012920	18				
NC_001643	1063	1061			
NC_001645	1296	1294	1251		
NC_002083	1698	1703	1763	1780	
NC_010339	3148	3147	3134	3158	3180

計算



分子進化の中立説

- 分子進化： DNA配列が変化．文字の置き換え（置換）など．



- DNAレベルでみた進化の大部分は自然選択に「中立的」　(木村資生 1968)

- 分子時計： 分子レベルでみると，進化速度がほぼ一定

個体数Nの集団において，突然変異数はNに比例，各変異の定着率はNに反比例．Ｎ＊（１／Ｎ）＝１であるから進化速度はNに依存しない．

（形態レベルでは成り立たない．例えばシーラカンス．）

分子時計

- 100万年でDNAの1%が変異すると仮定する
- 200万年でDNAの2%が変異
- 1000万年でDNAの9%が変異
- 1億年でDNAの55%が変異

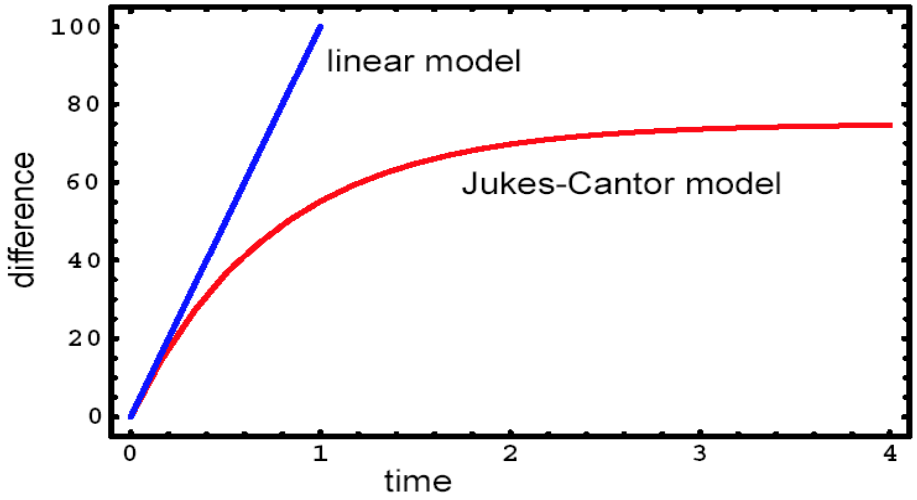
```
time   dif
  0     0% AAAAAAAAAA..........AAAAAAAAAA
0.01    1% AAAATAAAAA..........AAAAAAAAAA
0.1     9% AAAATAAAAA..........AAAAACAAAA
0.2    18% AAGATAAAAA..........AATAACAAAA
0.3    25% AAGAAAAAAA..........AATAACAAAG

1.0    55% AAGTTAACAA..........AATAGCATGC

3.0    74% TACTCGACTT..........CTATAGCAGC
```

分岐時間の推定

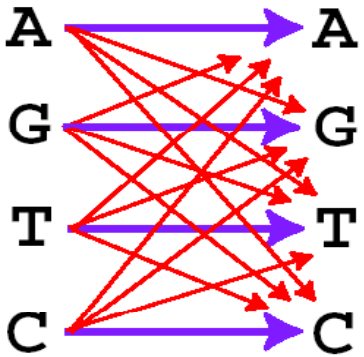
- DNAはランダムに変化
- 確率モデルの一種
- 75%までしか変わらない



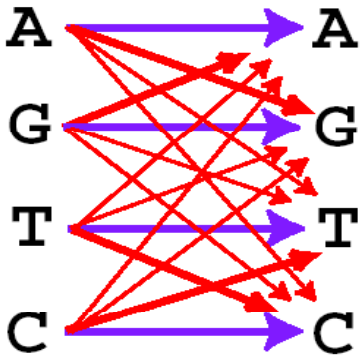
$$P(t) = \frac{3}{4}\left(1 - \exp\left(-\frac{4}{3}\lambda t\right)\right)$$

マルコフ過程

- 塩基置換（A,T,G,Cの４種類）



JCモデル



HKYモデル

（トランジションとトランスバージョン，組成比）

- アミノ酸置換（２０種類）

Dayhoffモデル，Jonesモデルなど

条件付確率

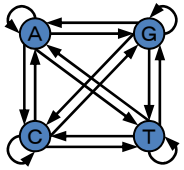
- **Probability of future state-b given current state-a after time-t**

$$P_{ba}(t)$$

note: $a, b \in \{\mathsf{A}, \mathsf{T}, \mathsf{G}, \mathsf{C}\}$ **for nucleotide sequences.**

- **Transition probability matrix**

$$P(t) = \begin{pmatrix} P(A|A) & P(G|A) & P(C|A) & P(T|A) \\ P(A|G) & P(G|G) & P(C|G) & P(T|G) \\ P(A|C) & P(G|C) & P(C|C) & P(T|C) \\ P(A|T) & P(G|T) & P(C|T) & P(T|T) \end{pmatrix}$$

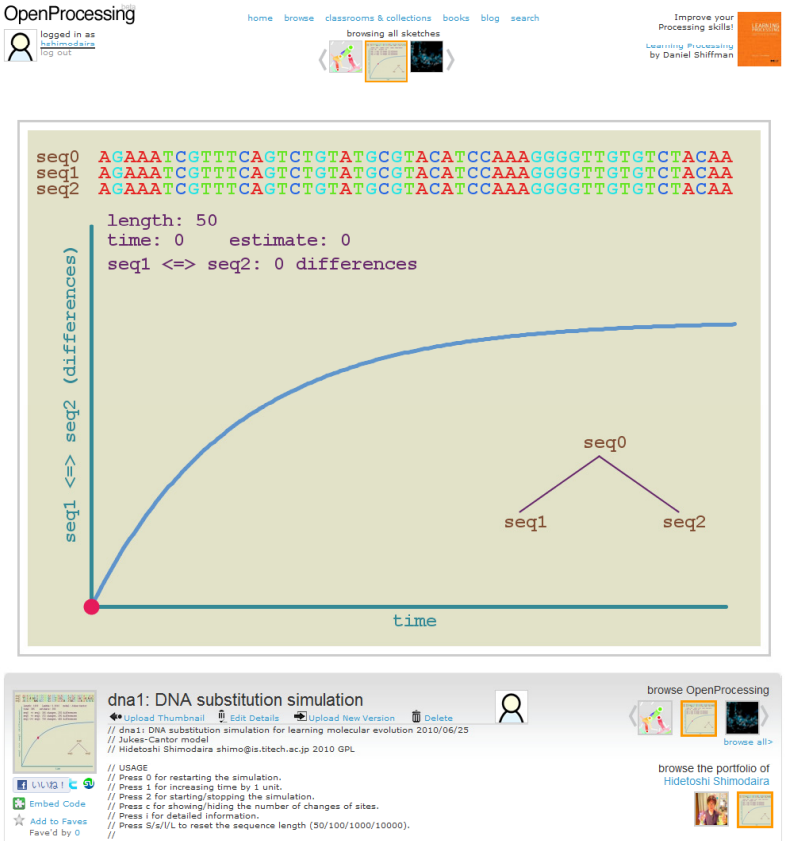


$$P(t) = \exp(tQ) = \sum_{k \geq 0} \frac{t^k}{k!} Q^k$$

note: $Q = 4 \times 4$ **matrix for nucleotide sequences, and** $Q = 20 \times 20$ **matrix for those of amino-acid.**

DNA置換のシミュレーション

http://www.openprocessing.org/visuals/?visualID=10609



キー操作

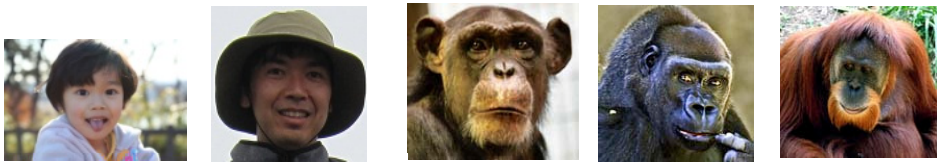
0： 時間を0にする.
1： 時間を1ステップ進める
2： 進行のオン，オフ
c： 置換数のオン，オフ
l (エル): シーケンス長＝1000に変更

javaのインストールが必要になります

変異数から置換数へ

DNA配列の長さ： 10839

変異数
（観測値）



	NC_001807	NC_012920	NC_001643	NC_001645	NC_002083
NC_012920	18				
NC_001643	1063	1061			
NC_001645	1296	1294	1251		
NC_002083	1698	1703	1763	1780	
NC_010339	3148	3147	3134	3158	3180

計算

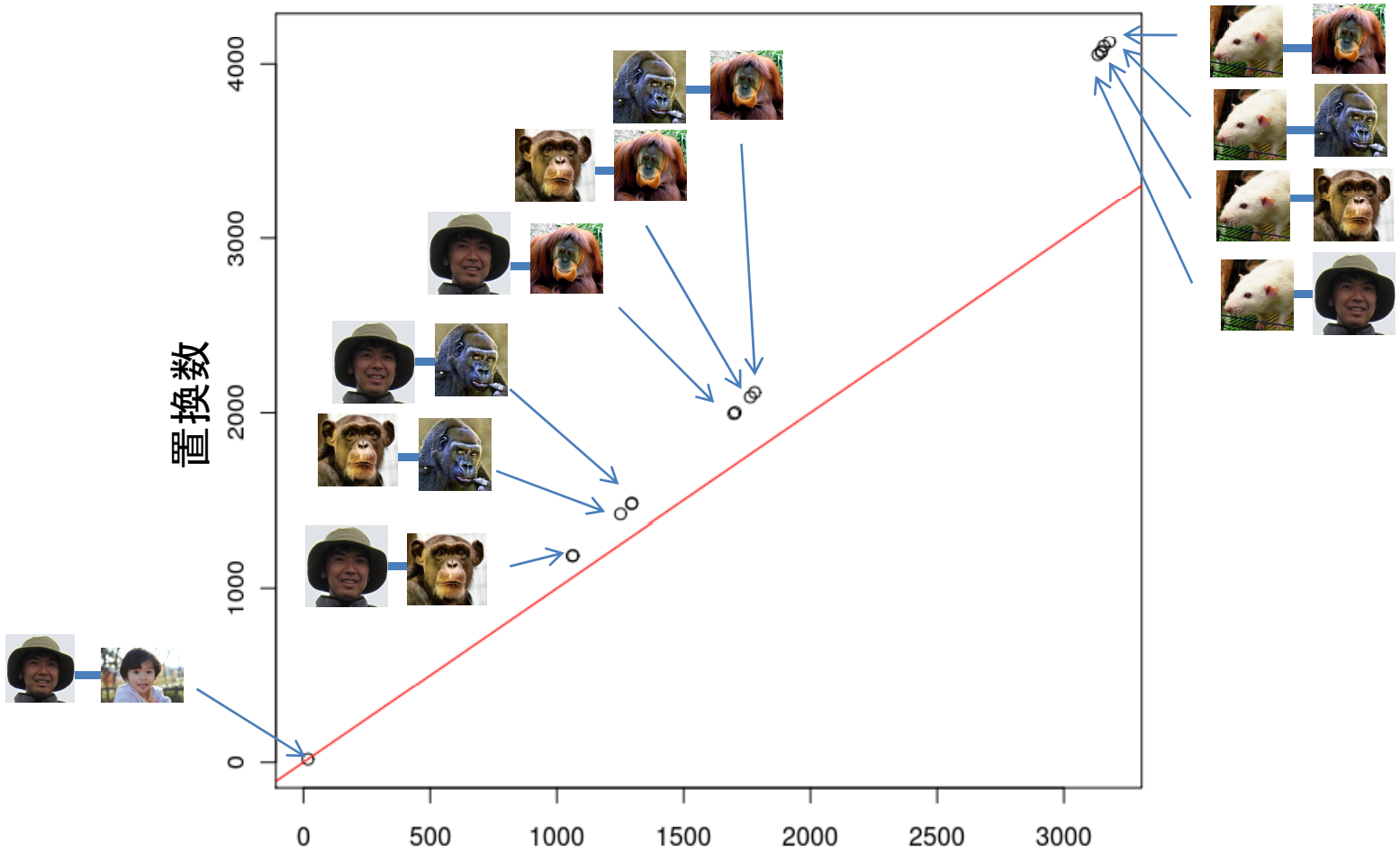
置換数
（推定値）

	NC_001807	NC_012920	NC_001643	NC_001645	NC_002083
NC_012920	18				
NC_001643	1188	1185			
NC_001645	1481	1479	1420		
NC_002083	1994	2001	2090	2116	
NC_010339	4069	4068	4052	4102	4127

注意： 方法の概略を示すために説明を簡略化してある．実際には変異数を変換したのではなく，DNA配列から最尤法で置換数を推定した．R言語のapeパッケージにてdist.dna(dat,“TN93”)を実行．しかし後でJukes-Cantorモデル(“JC”)を試したところ，結果は“TN93”とほとんど同じだった．JCならば変異数を変換して置換数を直接計算できる．

NC_001807: Homo sapiens: human
NC_012920: Homo sapiens: human
NC_001643: Pan troglodytes: chimpanzee
NC_001645: Gorilla gorilla: Western Gorilla
NC_002083: Pongo abelii: Sumatran orangutan
NC_010339: Mus musculus musculus: eastern European house mouse

変異数と置換数の関係

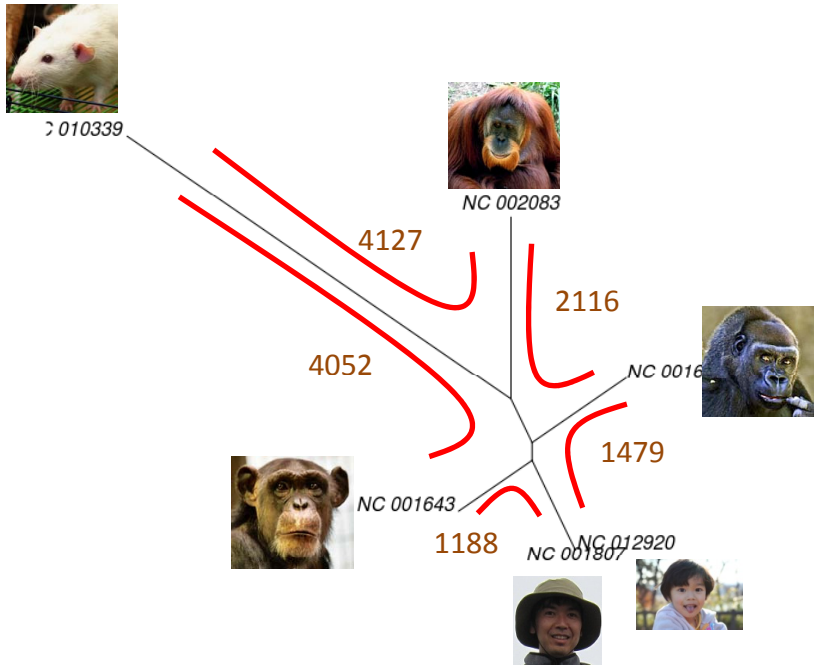


DNA配列の長さ： 10839

置換数から系統樹を推定

置換数
（推定値）

	NC_001807	NC_012920	NC_001643	NC_001645	NC_002083
NC_012920	18				
NC_001643	1188	1185			
NC_001645	1481	1479	1420		
NC_002083	1994	2001	2090	2116	
NC_010339	4069	4068	4052	4102	4127



推定された系統樹 （無根系統樹）

ココに根があると考える（マウス以外）

推定された系統樹 （有根系統樹）

紙テープで系統樹を工作

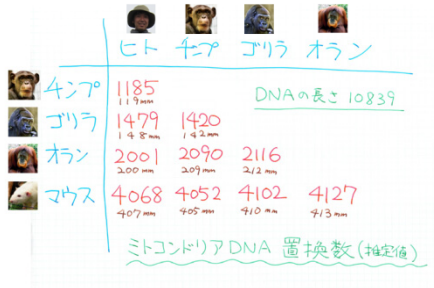
2010/06/27

系統樹の工作

東京工業大学　下平英寿　http://www.is.titech.ac.jp/~shimo/

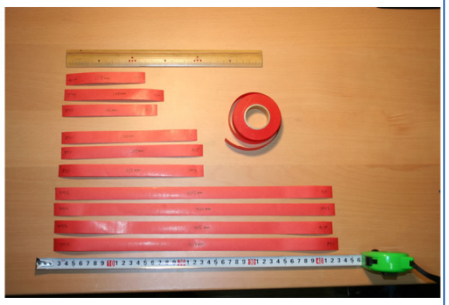
紙テープ（糸でもよい），はさみ，ものさし，のり，ペン，セロテープなど

置換数の長さに紙テープを切る



ここでは置換数10個を1mmにした

1センチがだいたい100万年です



参考：　置換数は最尤法で推定した

- データは類人猿4種（ヒト，チンパンジー，ゴリラ，オランウータン）とマウスのミトコンドリアDNA配列をNCBIのウェブサイトから取得．
- 12個の遺伝子コーディングリージョンをclustalWでアライメントして位置合わせ．長さ10839の配列を得る．
- R言語のapeパッケージにてDNA配列から最尤法で置換数を推定した．dist.dna(dat, "TN93")を実行後，結果を10839倍して整数に丸めた．単純な変異数（塩基が何個異なるか）はdist.dna(dat, "raw")で計算できる．
- 後で確かめたらJukes-Cantor (JC)モデルをつかってもほぼ同じ置換数が得られたので，TN93モデルより解説が容易なJCモデルを工作に使うべきだった．
- 置換数>変異数である．今回のデータでは置換数があまり大きくないので，置換数の代わりに変異数を用いても同じ系統関係が得られた．
- （工作しないで）系統樹まで求めるには，置換数の行列をnj関数に与えれば，近隣結合法の無根系統樹が得られる．root()で根を指定する．

2

YouTubeにも置きました

http://www.youtube.com/watch?v=jqqJODQytzk

統計学:　最尤法

コインが表になる確率 θ の推定

状況設定

コイン投げの実験　　500回くりかえす　　表＝1，裏＝0と書く

実験結果(n=500)　0 0 1 1 1 1 1 1 1 0 0 0 1 1 0 0 1 0 1 1 0 1 1 0 0 0....

i回目のコイン投げをx_iと書くと　$x_1 = 0, x_2 = 0, x_3 = 1, x_4 = 1, x_5 = 1, \ldots, x_{500} = 0$

表になった回数は219回　　$C = x_1 + \cdots + x_n$　　C = 219

推定結果

θ の推定値は表になった頻度　　$\hat{\theta} = \frac{C}{n} = \frac{219}{500} = 0.438$

ほとんど自明に見えるが，どうしてこの推定量でよいのか？

最尤法（コイン）

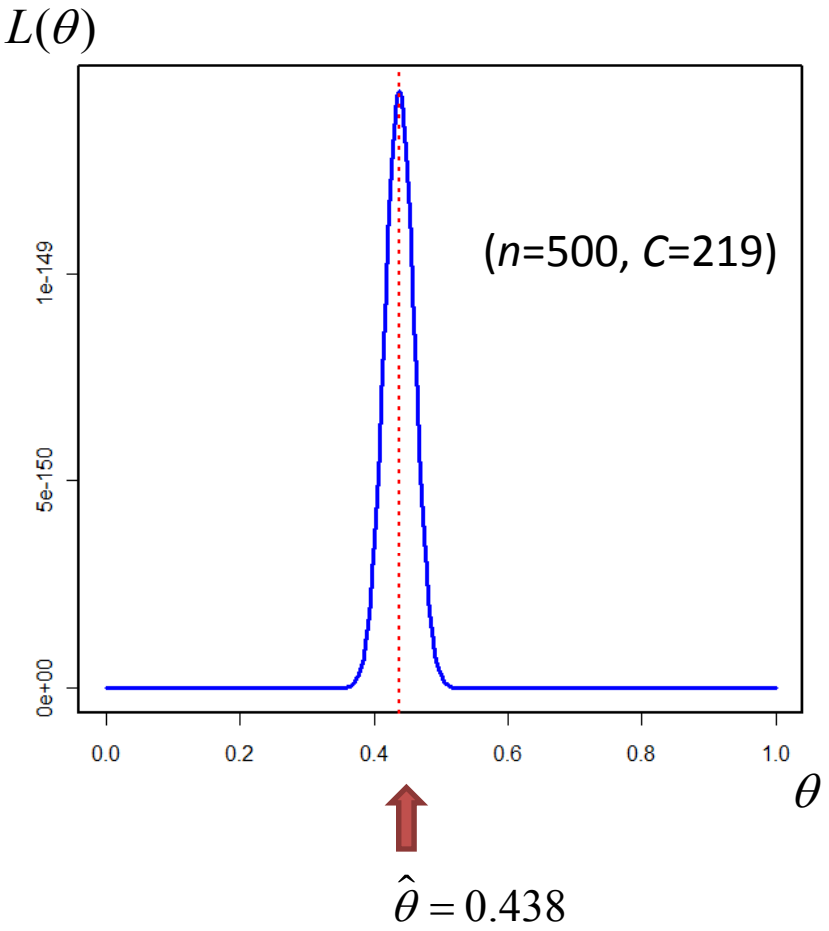
データを観測する確率（尤度）は

$$L(\theta) = \theta^C (1-\theta)^{n-C}$$

確率を最大にするパラメータ値を選ぶ

$$\max_{\theta} L(\theta) \Rightarrow \hat{\theta}$$ 最尤推定量

尤度を最大にするθは $\hat{\theta} = \frac{C}{n}$



最尤法の詳細（コイン）

1回目のコイン投げの結果がx_1である確率は

$$f(x_1;\theta)=\theta^{x_1}(1-\theta)^{1-x_1}=\begin{cases}\theta, & x_1=1\\ 1-\theta, & x_1=0\end{cases}$$

データ($x_1,\ldots,x_n$)を観測する確率（尤度）は

$$L(\theta)=f(x_1;\theta)\cdots f(x_n;\theta)$$

$$=\theta^{x_1}(1-\theta)^{1-x_1}\cdots\theta^{x_n}(1-\theta)^{1-x_n}$$

$$=\theta^{x_1+\cdots+x_n}(1-\theta)^{(1-x_1)+\cdots(1-x_n)}=\theta^{C}(1-\theta)^{n-C}$$

$\log L(\theta)=C\log\theta+(n-C)\log(1-\theta)$　を微分すると

$\frac{d\log L(\theta)}{d\theta}=\frac{C}{\theta}-\frac{n-C}{1-\theta}=\frac{C-n\theta}{\theta(1-\theta)}$ なので $\frac{d\log L(\theta)}{d\theta}=0$ を解くと　$\hat{\theta}=\frac{C}{n}$

最尤法(Jukes-Cantorモデル)

コイン投げと同じところ

塩基が異なる場合=1, 同じ場合=0 と書く

```
...ATGCCATG...
...ATACCCTG...
...00100100...
```

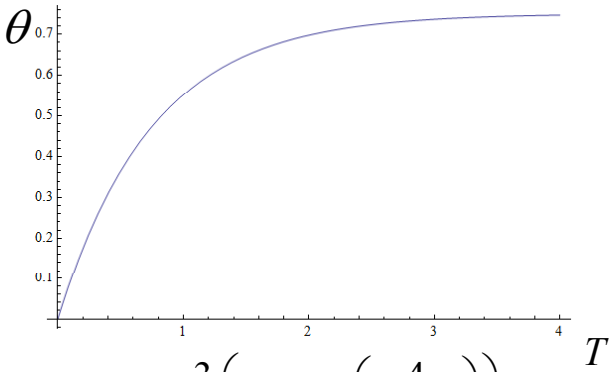
DNA配列の長さ = n

変異数 = C

塩基が異なる確率 = θ

$$L(\theta) = \theta^{C}(1-\theta)^{n-C}$$

塩基が異なる確率と時間の関係



$$\theta = \frac{3}{4}\left(1-\exp\left(-\frac{4}{3}T\right)\right)$$

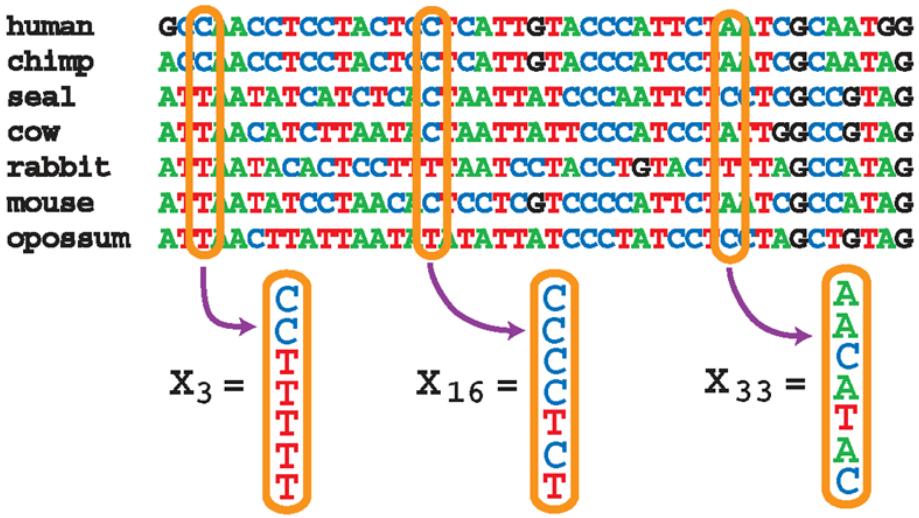
塩基毎の置換数に換算した時間 = T

$$L_{JC}(T) = L\left(\frac{3}{4}\left(1-\exp\left(-\frac{4}{3}T\right)\right)\right)$$ を最大にするT ➡ $$\hat{T} = -\frac{3}{4}\log\left(1-\frac{4}{3}\frac{C}{n}\right)$$

$\left(C \geq 0.75n \text{ なら } \hat{T} = \infty\right)$

DNA配列間の置換数の推定値: $n\hat{T}$

最尤法（系統樹）

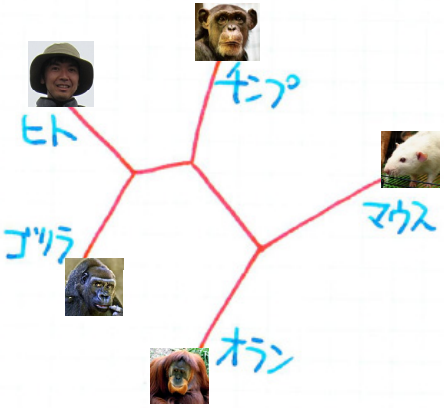


$$L(\text{樹形}, \text{枝長}) = f(x_1; \text{樹形}, \text{枝長}) \cdots f(x_n; \text{樹形}, \text{枝長})$$

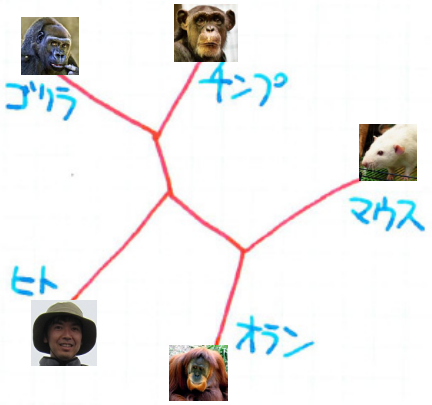
$$\max_{\text{樹形}} \max_{\text{枝長}} L(\text{樹形}, \text{枝長}) \Rightarrow (\widehat{\text{樹形}}, \widehat{\text{枝長}})$$ **最尤推定量**

樹形

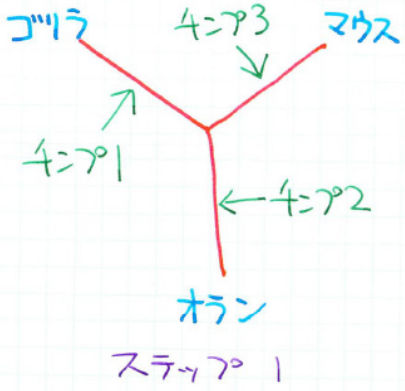
樹形1



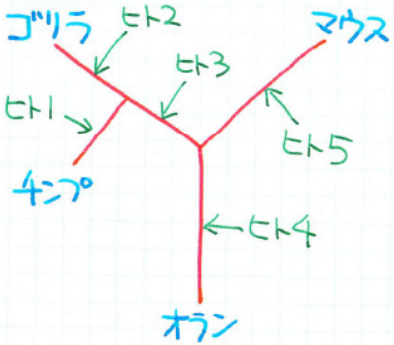
樹形2



樹形3



ステップ1



ステップ2

くみあわせ

3×5

=15とおり

これが樹形の数

系統樹を数える

```
Table[{n, (2 n - 5) ! / ((n - 3) ! 2^(n-3))}, {n, 3, 20}] // TableForm
```

3	1
4	3
5	15
6	105
7	945
8	10 395
9	135 135
10	2 027 025
11	34 459 425
12	654 729 075
13	13 749 310 575
14	316 234 143 225
15	7 905 853 580 625
16	213 458 046 676 875
17	6 190 283 353 629 375
18	191 898 783 962 510 625
19	6 332 659 870 762 850 625
20	221 643 095 476 699 771 875

生物種=*n*のとき，無根系統樹の樹形の数は

$$1\times3\times5\times7\times\cdots\times(2n-5)=\frac{(2n-5)!}{(n-3)!2^{n-3}}$$

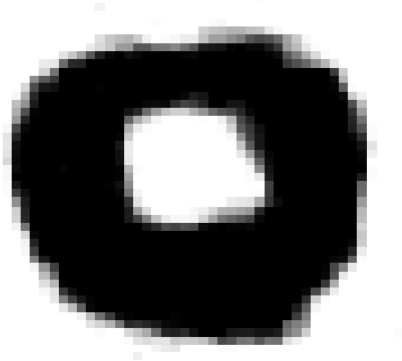
参考： Markov chain Monte Carlo法

情報科学科３年「データ解析」の課題: ランダムネス計算のアルゴリズムをプログラミング

観測画像 Y

画像復元

推定画像 X



ベイズ事後確率を計算 $P(X|Y) = \dfrac{P(Y|X)\,P(X)}{P(Y)}$

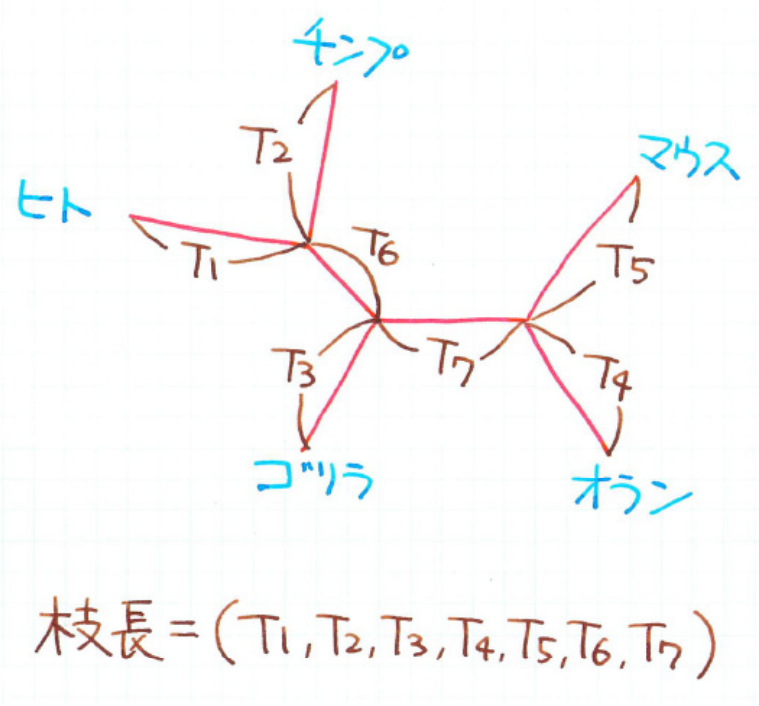
ベイズ
(1702-1761)

50x50＝2500ピクセル

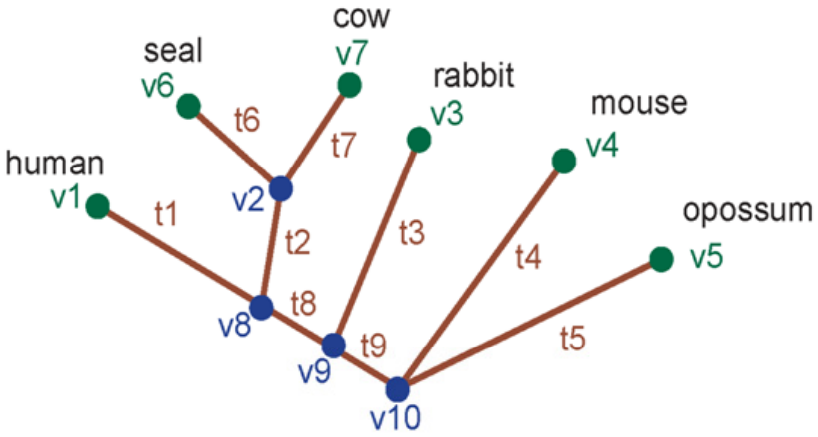
探索する画像の数＝$2^{2500} = 3 \times 10^{752}$

MCMC法（ギブス・サンプラー）

枝長



系統樹の確率モデル



サイト h における配列パタン x_h の確率

$$f(x_h; \text{樹形}, \text{枝長}) =$$

$$\sum_{v2,v8,v9,v10} P(v1|v8;t1)P(v2|v8;t2)P(v3|v9;t3)$$
$$\times P(v4|v10;t4)P(v5|v10;t5)P(v6|v2;t6)$$
$$\times P(v7|v2;t7)P(v8|v9;t8)P(v9|v10;t9)P(v10)$$

各枝でのマルコフ過程(つまり進化)は独立という仮定

- 塩基置換（A,T,G,Cの４種類）

JCモデル　　HKYモデル

(トランジションとトランスバージョン，組成比)

- アミノ酸置換（２０種類）

Dayhoffモデル，Jonesモデルなど

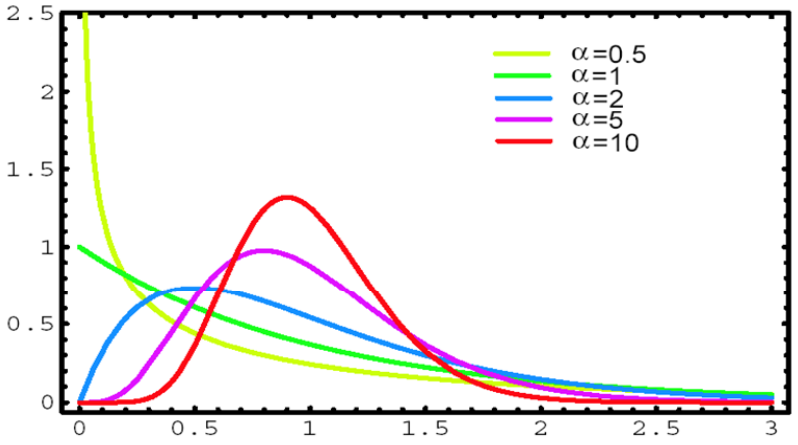
$$P(t) = \exp(tQ) = \sum_{k \geq 0} \frac{t^k}{k!} Q^k$$

進化速度の変動

DNAの座位毎に進化速度が異なる

⇒ ガンマ分布の確率密度関数（平均＝1）でモデル化

$f(\lambda;\alpha)$



λ (進化速度)

$$f(x_h;\text{樹形, 枝長}) = \int_0^\infty f(x_h;\text{樹形, 枝長}\times\lambda)\, f(\lambda;\alpha)\, d\lambda$$

樹形の尤度

$$L(\text{樹形}) = \max_{\text{枝長}} L(\text{樹形}, \text{枝長})$$

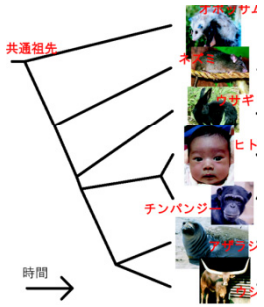
計算には対数尤度を使うことが多い

$$\log L(\text{樹形}) = \max_{\text{枝長}} \log L(\text{樹形}, \text{枝長})$$

$\max_{\text{樹形}} L(\text{樹形})$ ⇒ $\widehat{\text{樹形}}$　最尤推定量

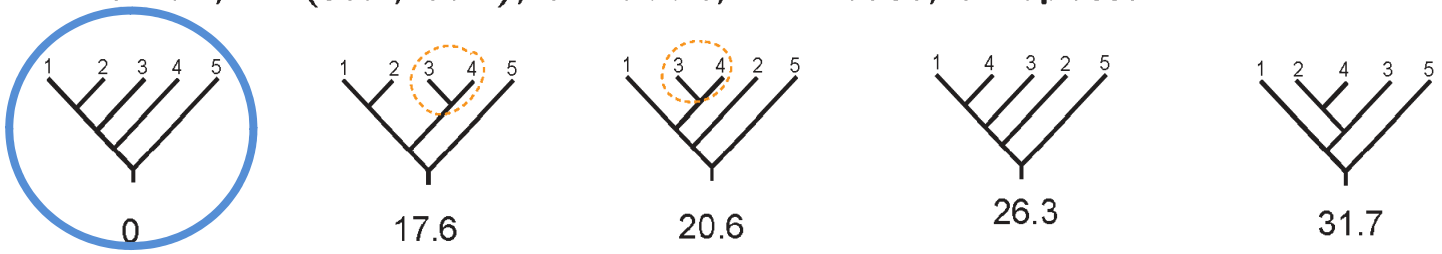
⇕ どちらで考えても同じこと

$\max_{\text{樹形}} \max_{\text{枝長}} L(\text{樹形}, \text{枝長})$ ⇒ $(\widehat{\text{樹形}}, \widehat{\text{枝長}})$

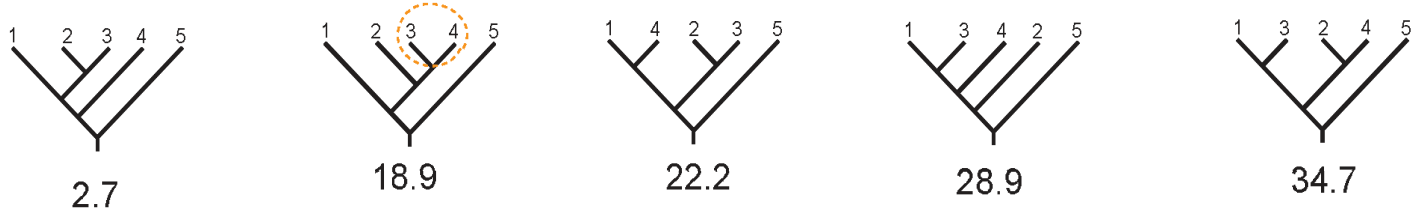


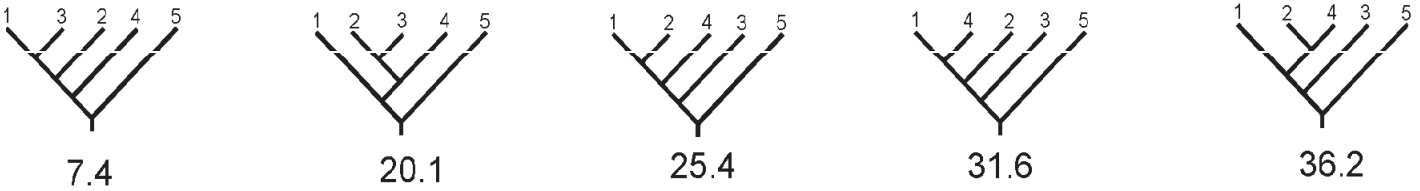
系統樹を尤度の順に並べる

1=human, 2=(seal, cow), 3=rabbit, 4=mouse, 5=opossum



樹形の最尤推定





数値は対数尤度の差　$\log \widehat{L(\text{樹形})} - \log L(\text{樹形})$

最尤法の一般化：情報量規準

Akaike Information Criterion

$$AIC = -2\log L(\theta) + 2\dim\theta$$

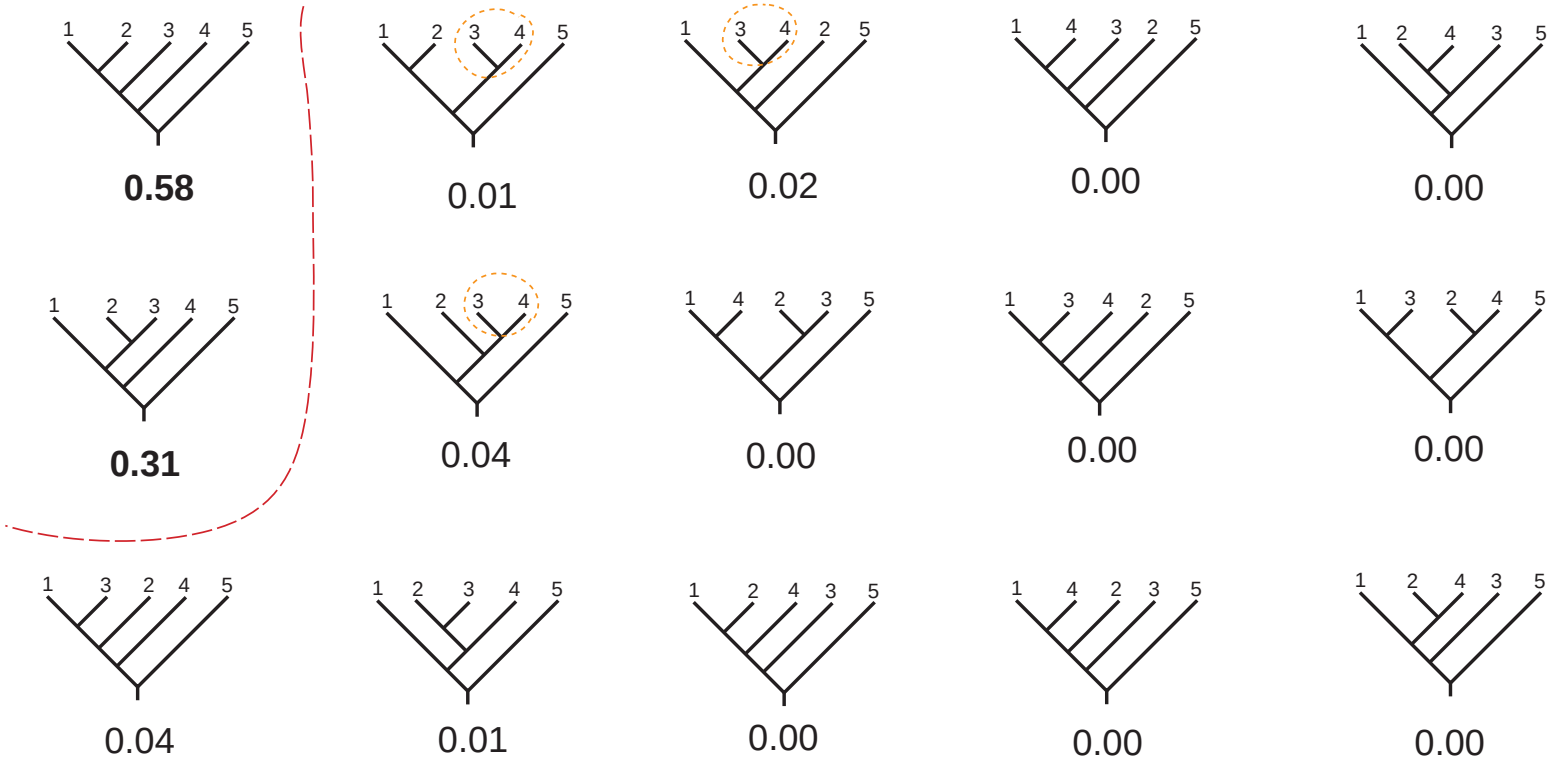
第１項：　対数尤度　＝　確率の対数　＝　エントロピーに相当
第２項：　パラメタ数　＝　確率モデルの複雑さ

2006年　京都賞　基礎科学部門 / 数理科学
赤池 弘次（Hirotugu Akaike）
統計数理研究所　名誉教授

情報数理の基礎概念に基づく、実用性と汎用性の両方を兼ね備えた、統計モデル選択のための規準 Akaike Information Criterion（AIC) の提唱により、データの世界とモデルの世界を結びつける新しいパラダイムを打ち立て、情報・統計科学への多大な貢献をした

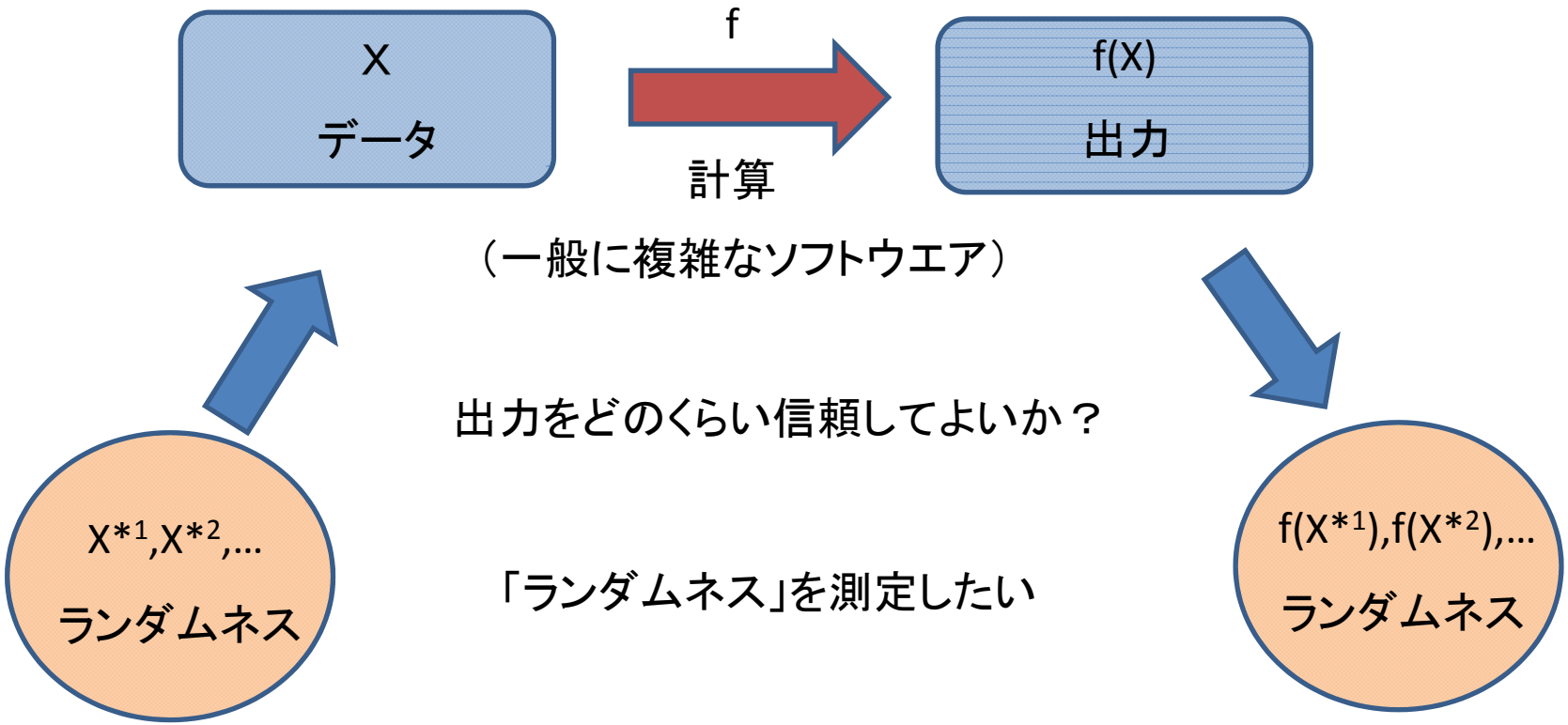
次の話題
系統樹のブートストラップ確率

1=ヒト，２＝（アザラシ，ウシ），３＝ウサギ，４＝マウス，５＝オポッサム



統計学： ブートストラップ法

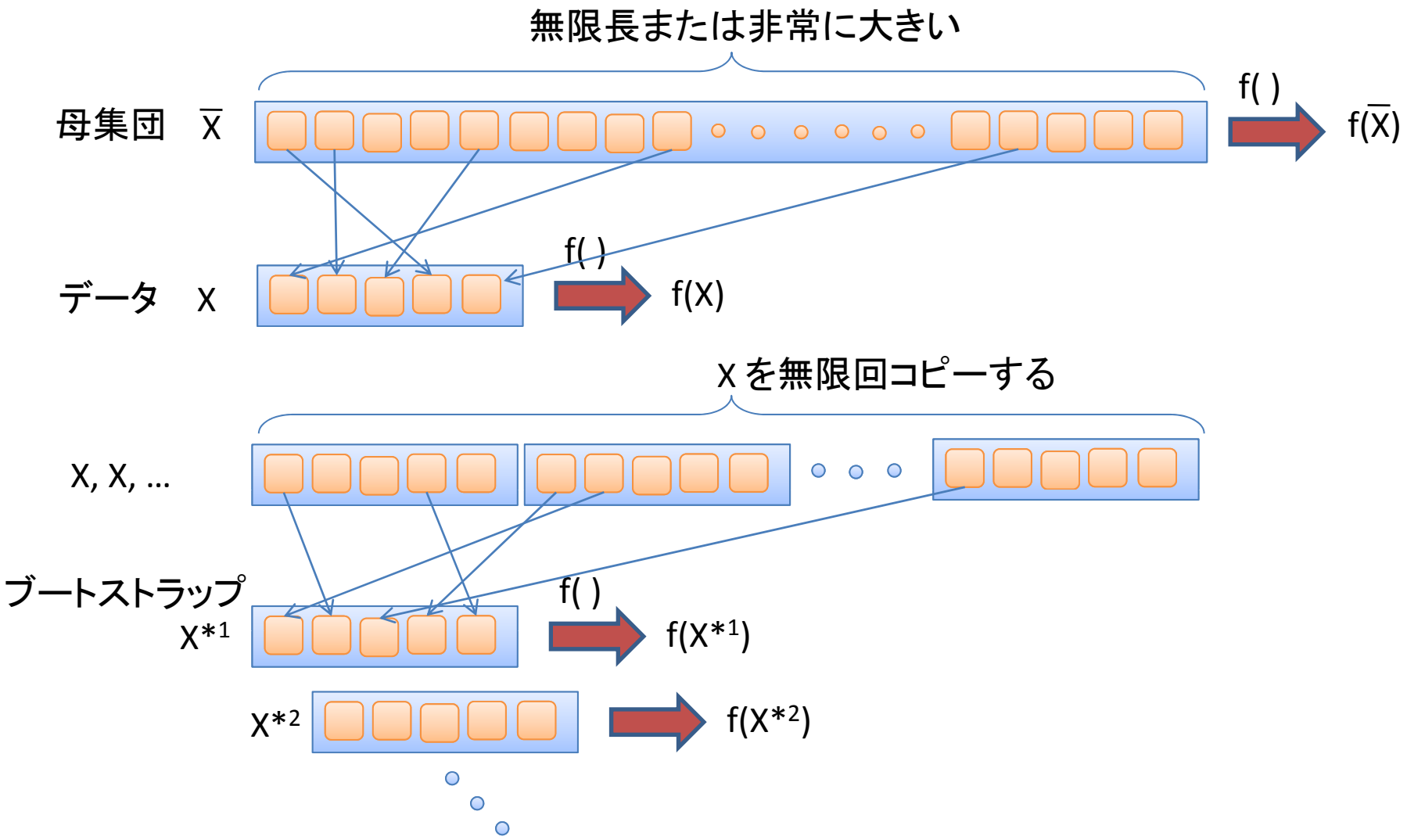
ランダムネスの測定
Measuring the Randomness



ランダムネスのものさし：　統計学の「確率値 (p-値)」または「信頼度」

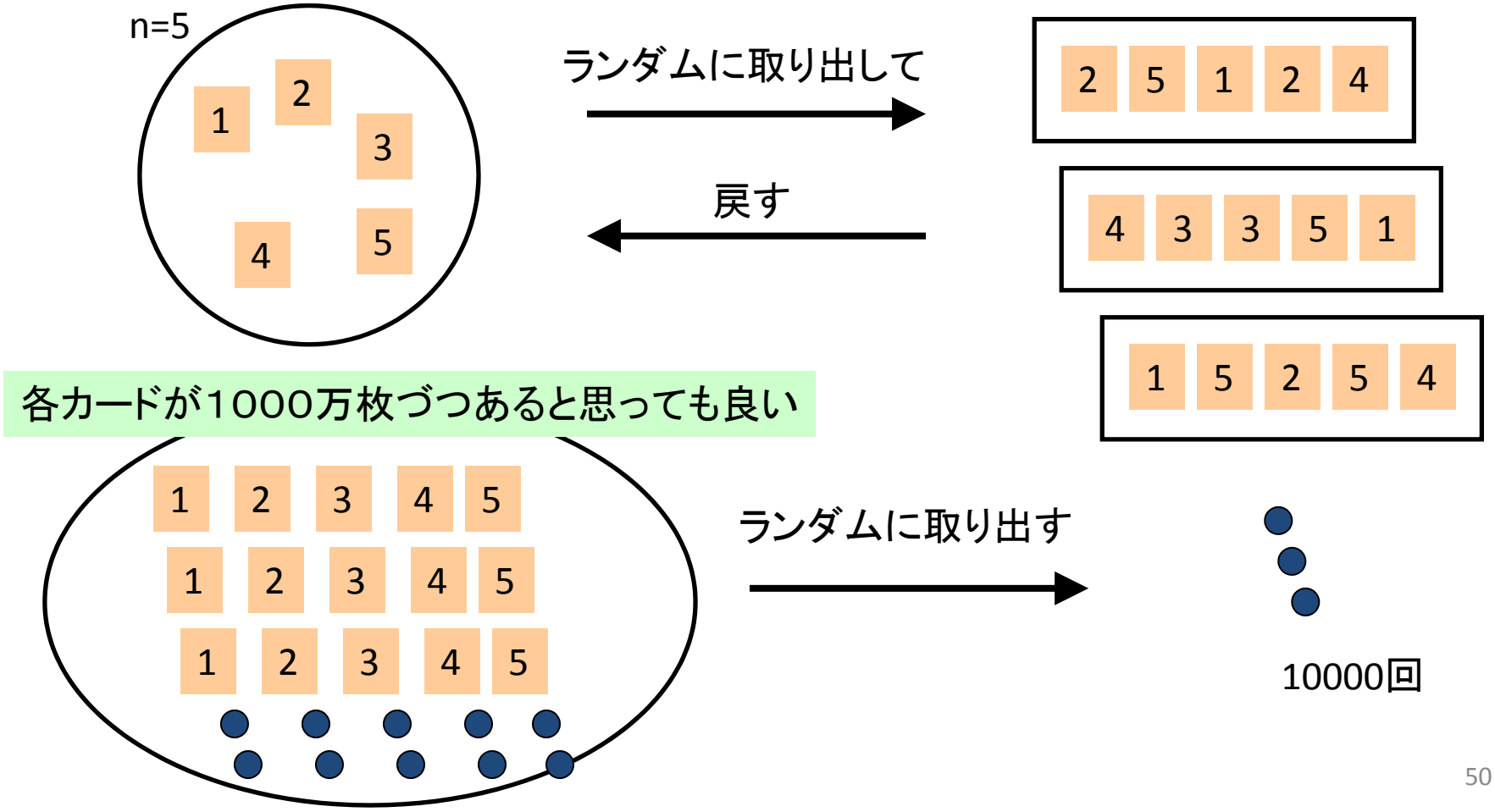
きれいに解が求まらない問題でも使いたい（面倒な式はイヤ）

ブートストラップ法の考え方



復元抽出

ブートストラップ法＝データからの復元抽出（resampling with replacement)



ブートストラップ法の応用場面

- コイン投げ
- 系統樹のブートストラップ確率
- 系統樹の枝に関するブートストラップ確率
- ネットワーク推定における枝のブートストラップ確率（例：遺伝子制御ネットワーク推定）
- 機械学習のバギング（例：手書き文字認識）
- 空間統計（例：酸性雨の分布）

ブートストラップ法（コイン投げ）

シミュレーション

1回目
1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 1 0 0 0 1 0 1 0 0 1 1 1...
$\hat{\theta}^* = 0.418$

2回目
1 0 0 1 0 0 0 1 0 0 0 0 1 0 1 1 1 1 1 0 1 0 1 0 1 0 0 1 0 1...
$\hat{\theta}^* = 0.458$

10000 回目
0 0 1 0 1 1 0 1 1 0 1 1 1 0 0 0 1 1 1 0 1 0 1 1 0 0 0 1 1 1...
$\hat{\theta}^* = 0.420$

信頼区間

$\hat{\theta}^* = 0.418, 0.458, \ldots, 0.420$

を大きくなる順にソート

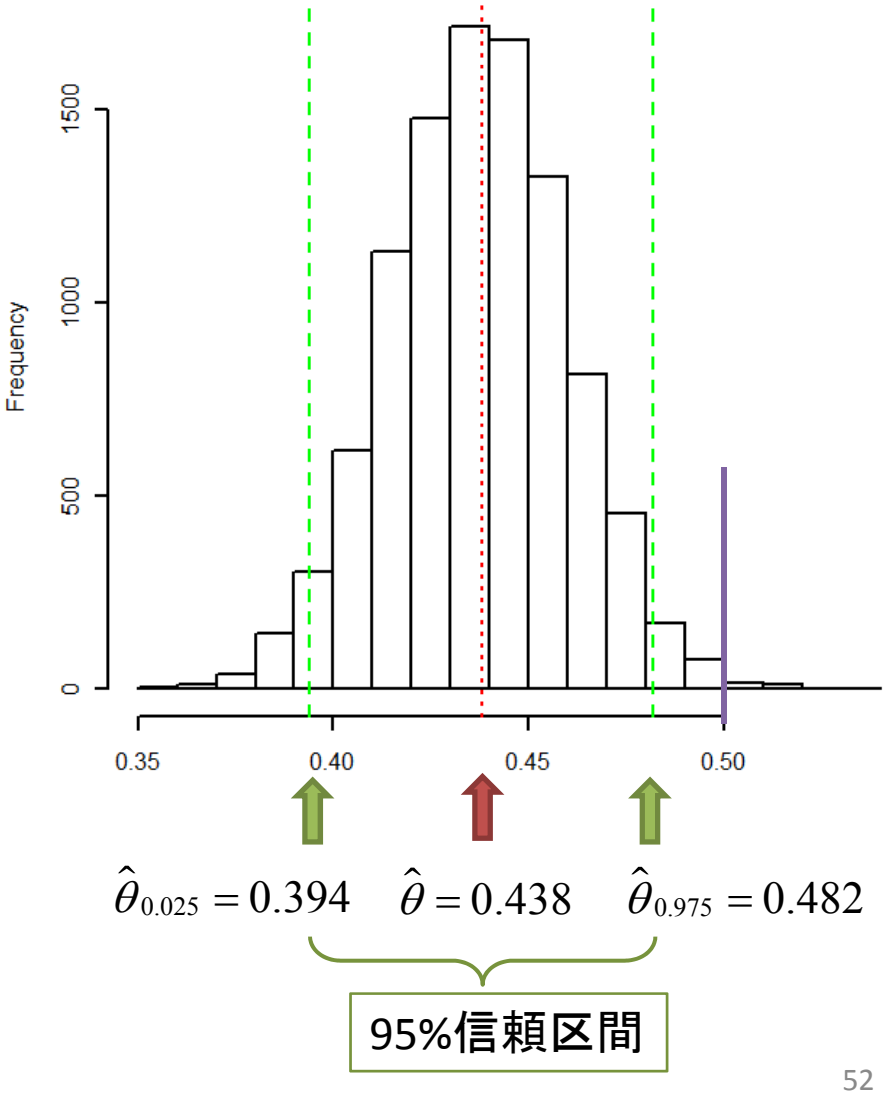
小さい方から数えて2.5%と97.5%の値

$\hat{\theta}_{0.025} = 0.394$　　$\hat{\theta}_{0.975} = 0.482$

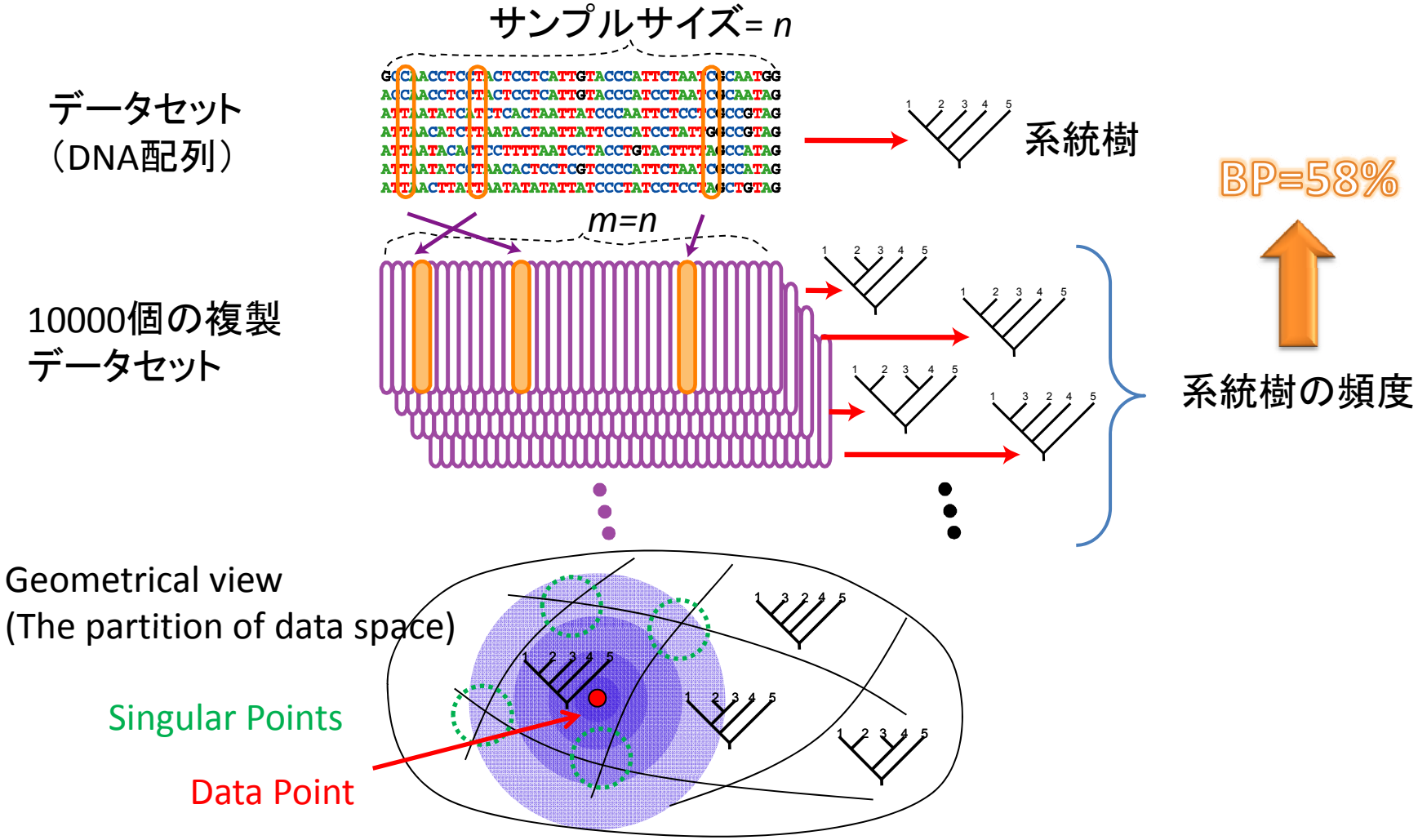
仮説検定

$\#\{\hat{\theta}^* > 0.5\} = 32$

帰無仮説 θ=0.5のp-値は0.0064



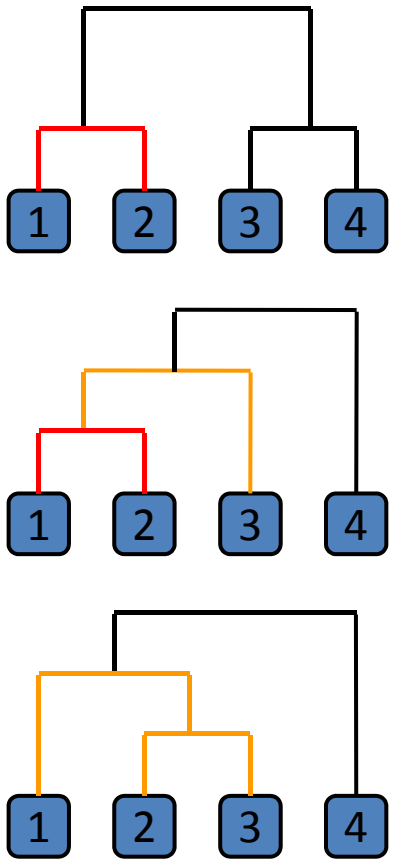
ブートストラップ確率
(Bootstrap Probability を略してBP)



Bootstrap probability = a Bayesian posterior probability (using flat prior dist.)

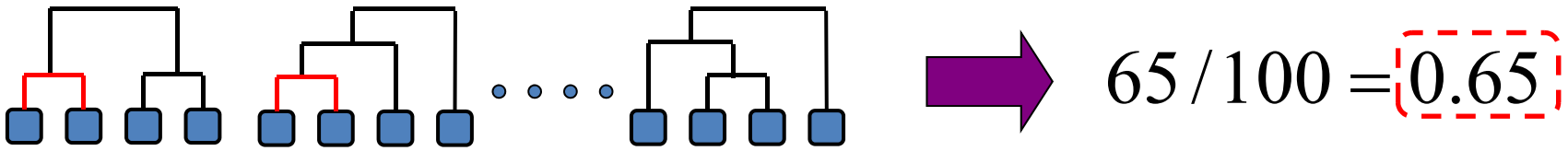
クラスター分析のバラツキ評価

- 樹形図に含まれる群に注目
- 右図の例
  - ２回出現した群
    - {1,2}, {1,2,3}
  - １回出現した群
    - {3,4}, {2,3}
- 群の「出現頻度」を数える



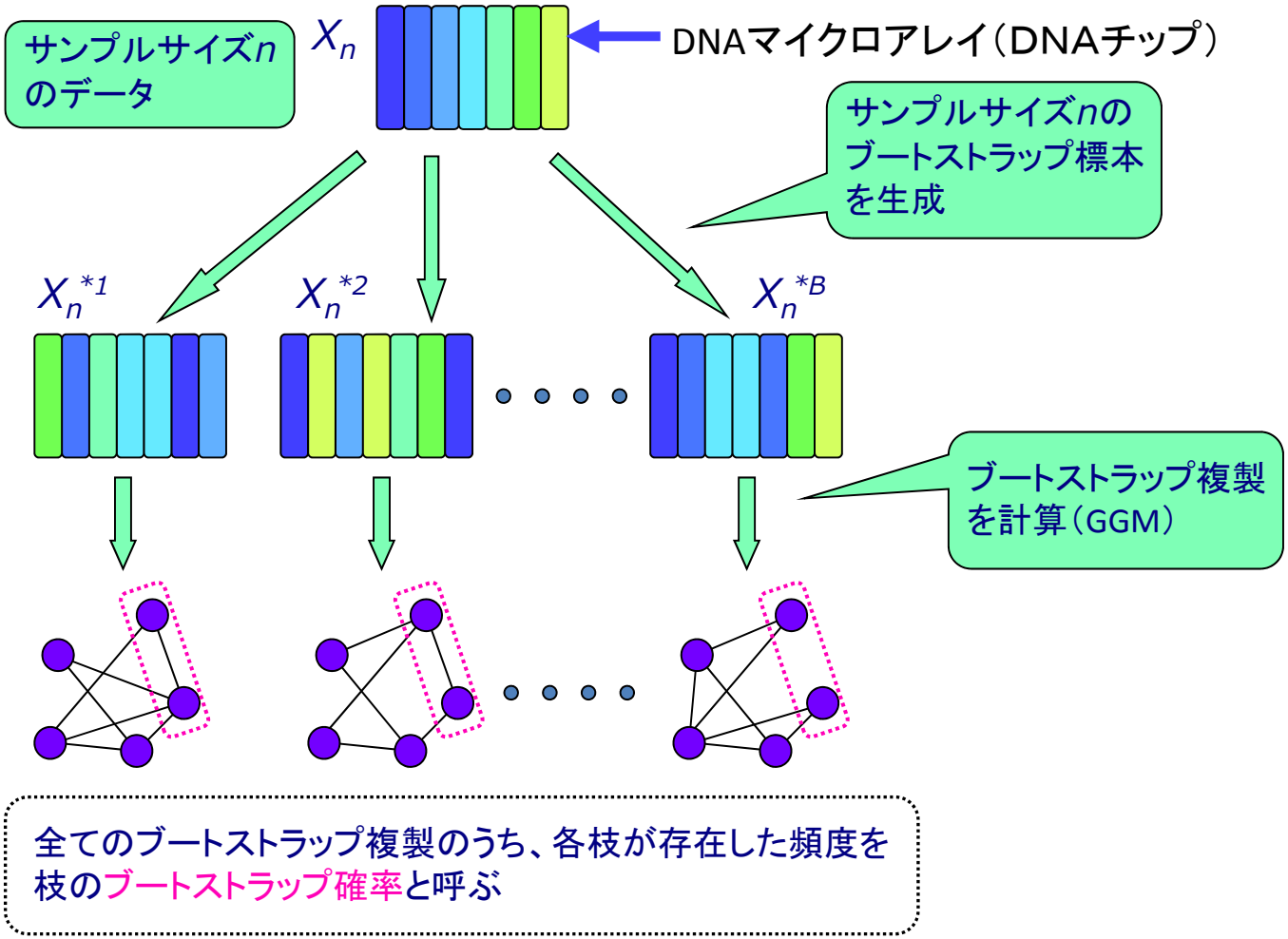
ブートストラップ確率

- ブートストラップ確率
  - モデルが仮説を満たした頻度
  - 例：群が存在した頻度



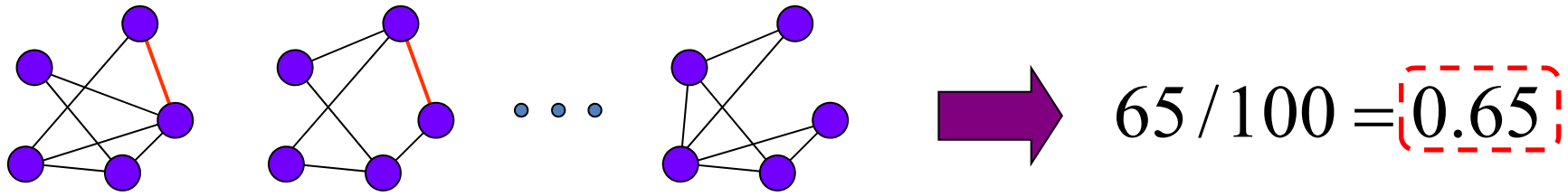
100回のブートストラップ中，ある群が65回存在

ネットワーク推定のランダムネス



ブートストラップ確率

- ブートストラップ確率
  - モデルが仮説を満たした頻度
  - 例：枝が存在した頻度

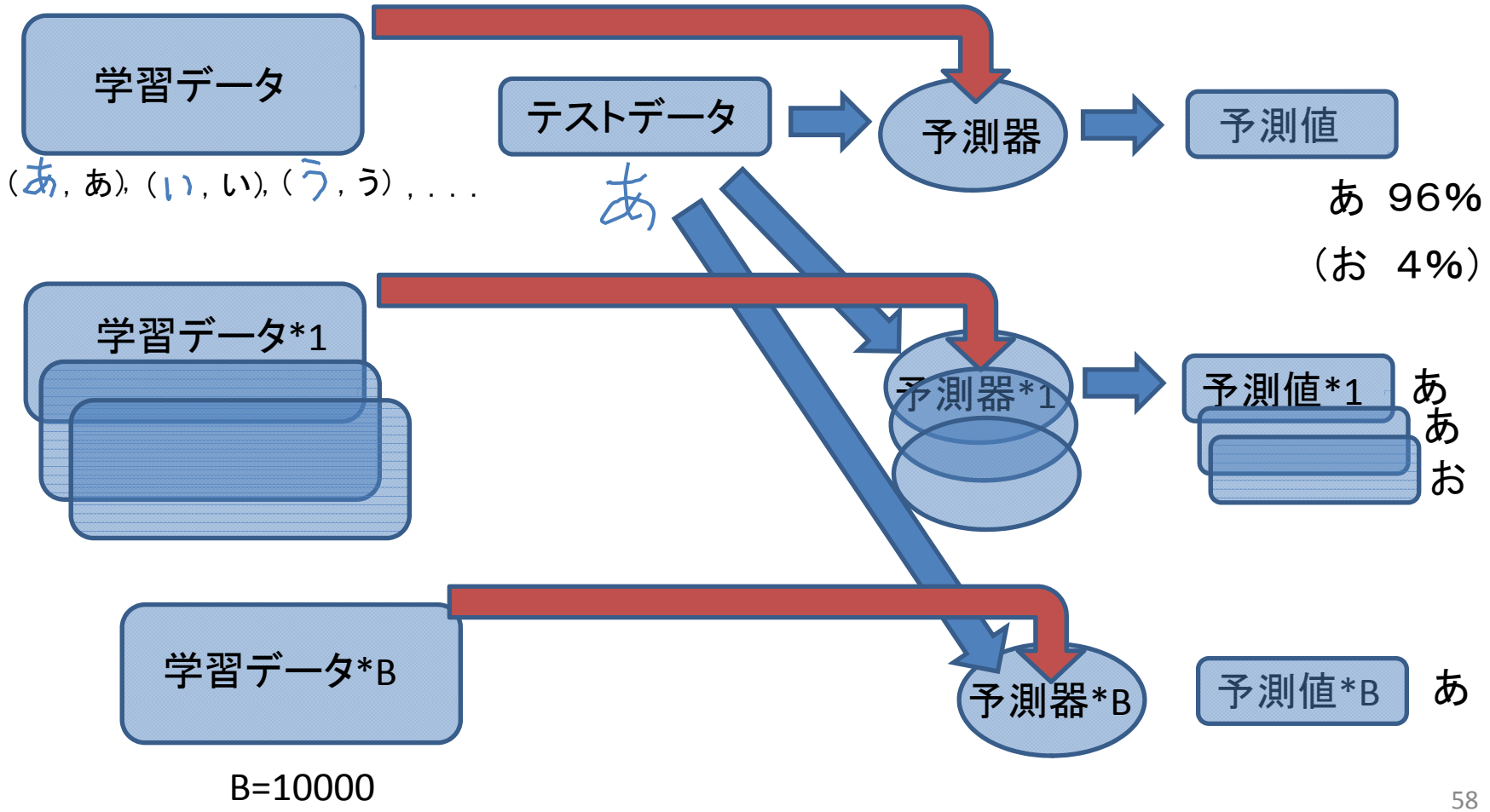


100回のブートストラップ中、ある枝が65回存在

バギング

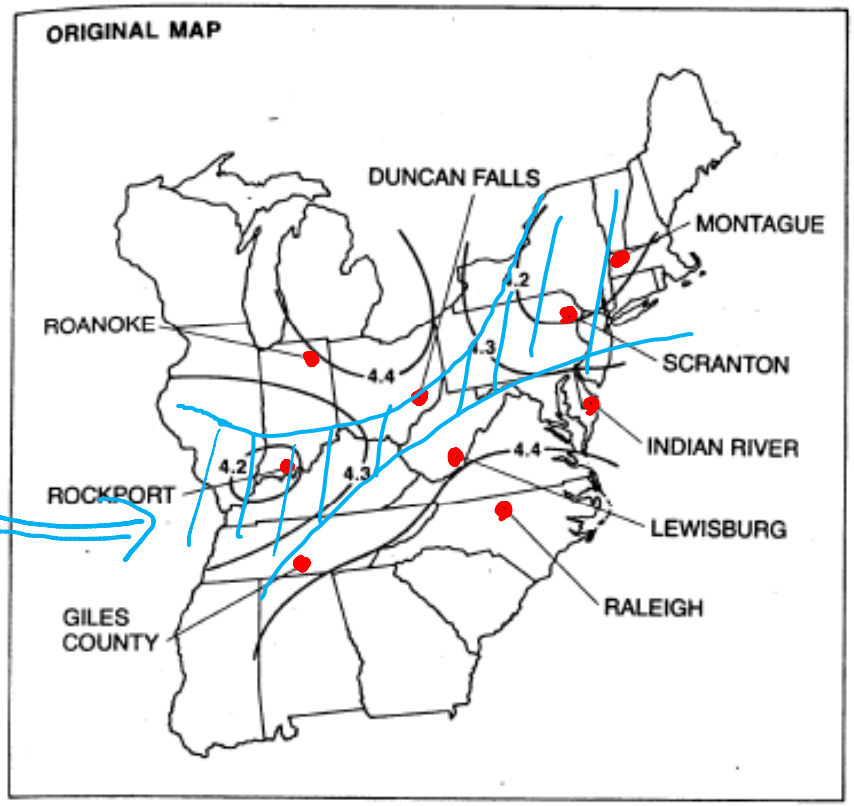
Bootstrap aggregating (bagging)

機械学習の判別問題 （例：手書き文字認識）



酸性雨(pH)の分布

2年間（2000個）の測定値　＠　9カ所の測候所

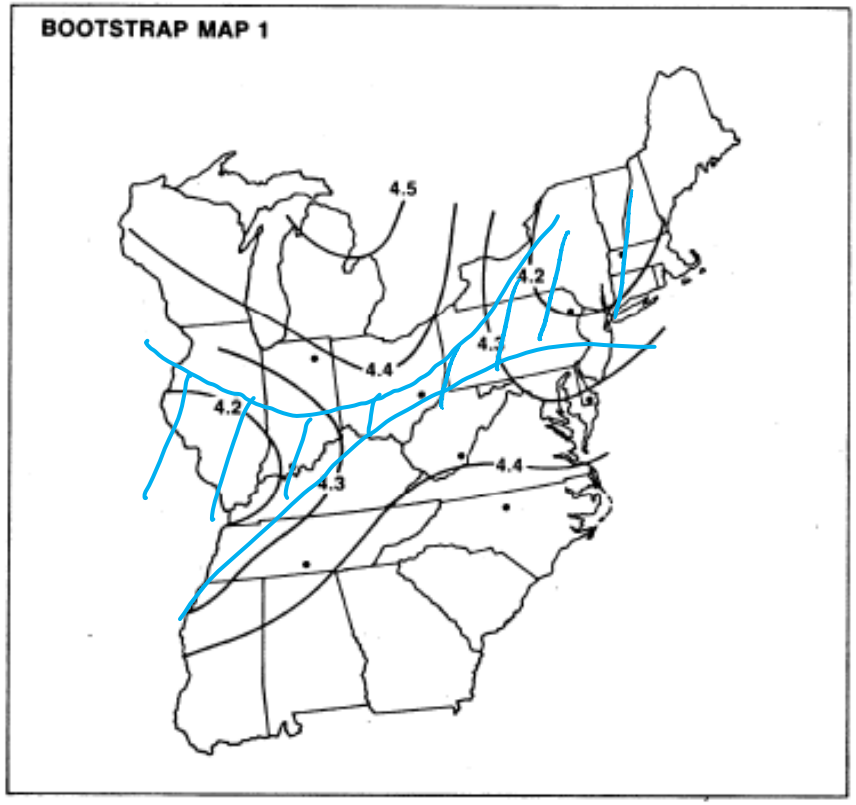


酸性

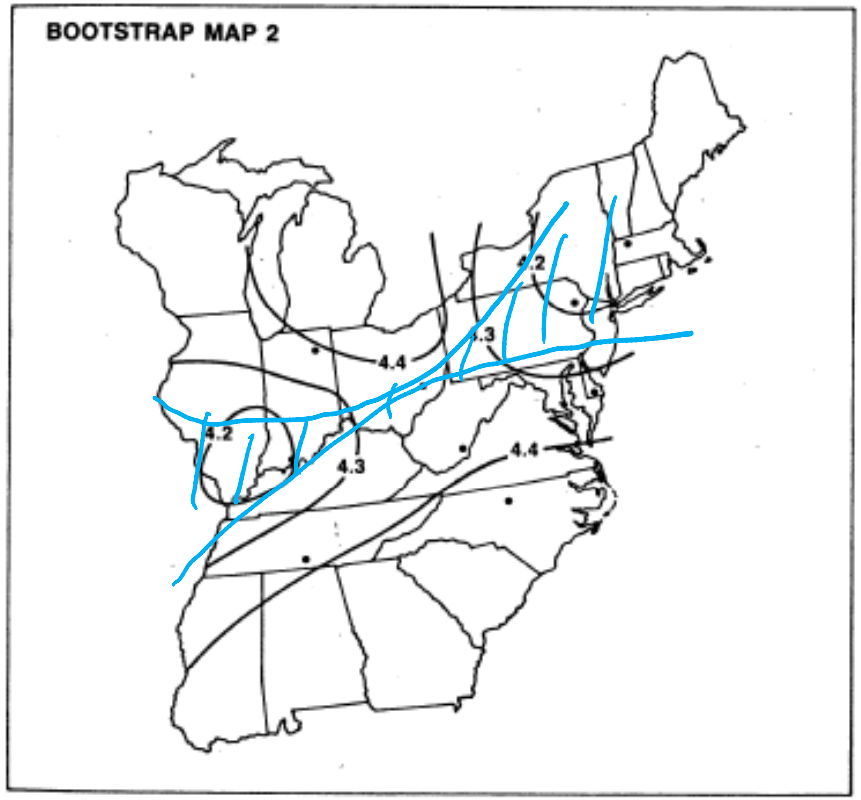
ダイアコニス, エフロン (1983)
「コンピュータがひらく新しい統計学」 サイエンス (松原望 訳)

ブートストラップでデータセットを作って再分析する

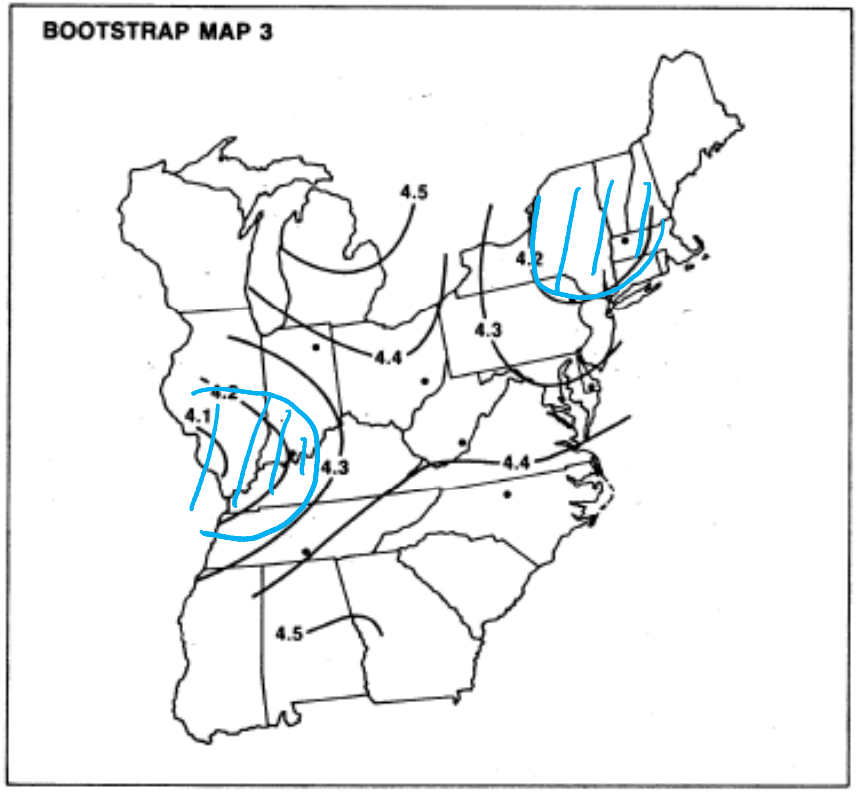
ブートストラップ標本（1）



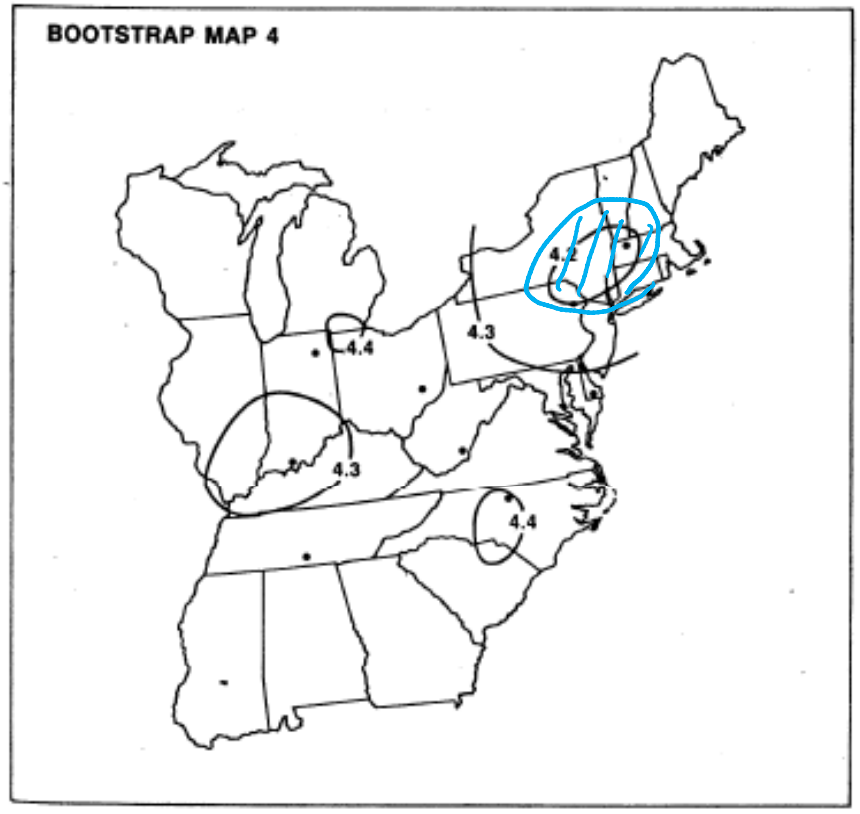
ブートストラップ標本（2）



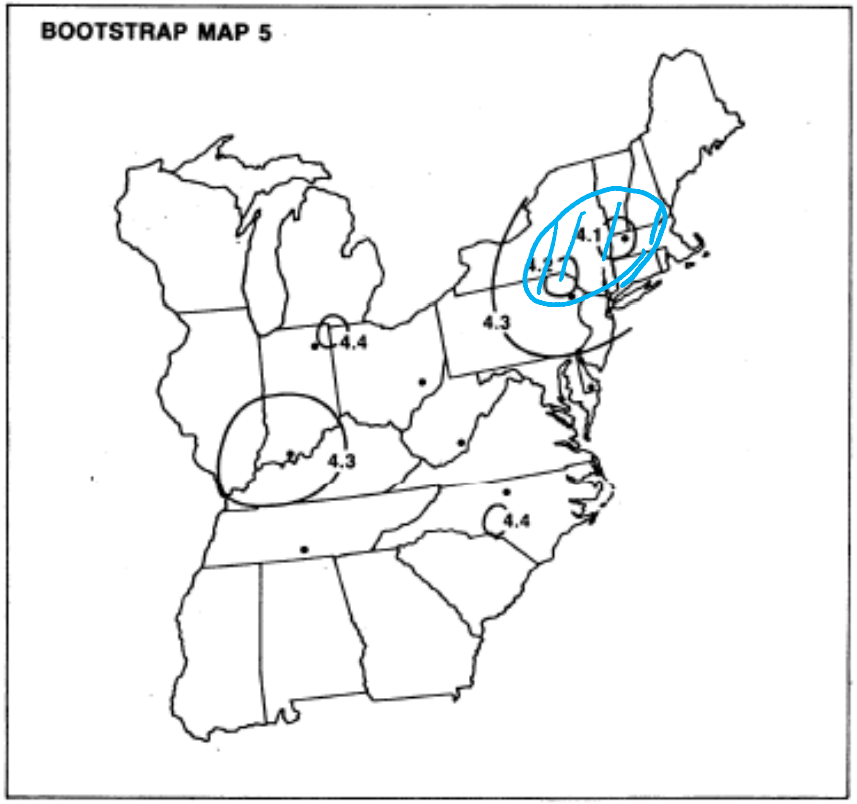
ブートストラップ標本（3）



ブートストラップ標本（4）



ブートストラップ標本（5）

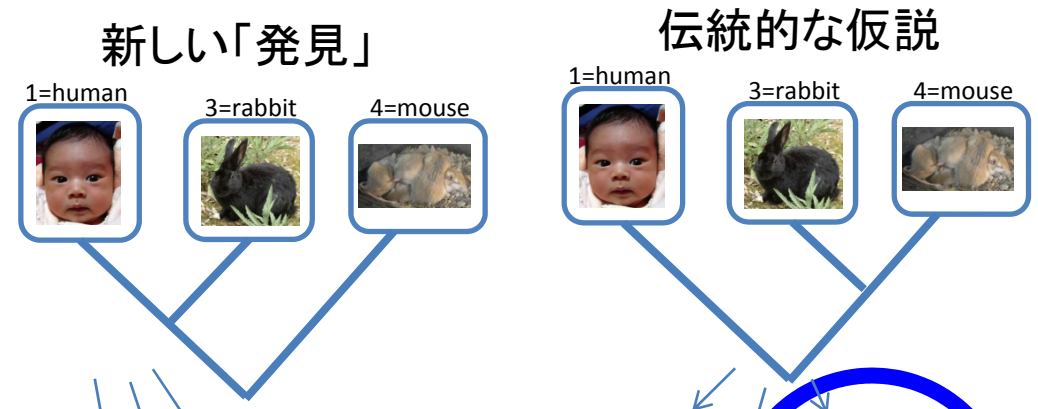


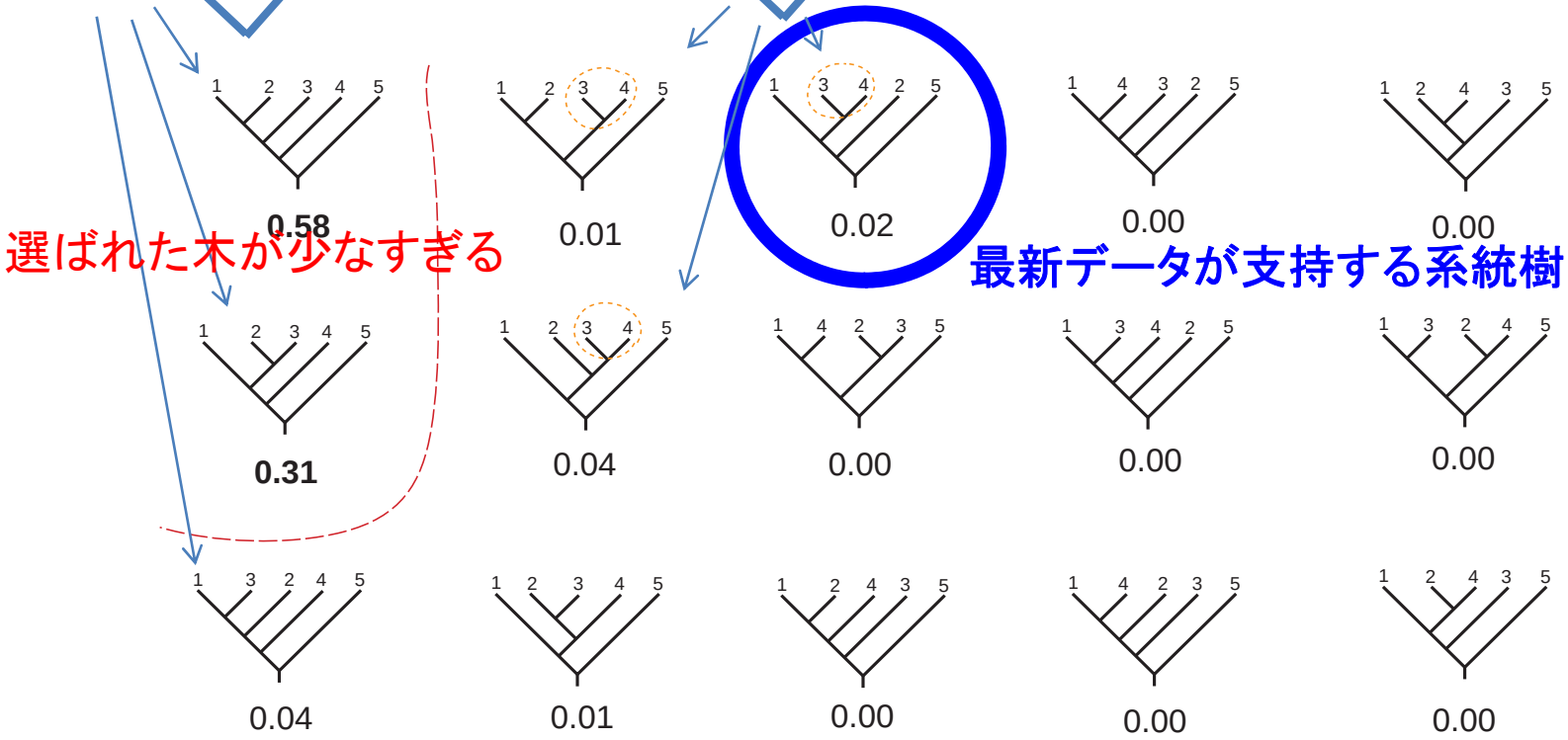
系統樹のランダムネスを測る３つの方法

- ブートストラップ確率
- Shimodaira-Hasegawa 検定
  (多重比較法の一種)
- マルチスケール・ブートストラップ
  （サンプルサイズを*m=-n*にする）

ブートストラップ確率（系統樹）

Efron (1979) *Ann. Stat.*　Felsenstein (1985) *Evolution*





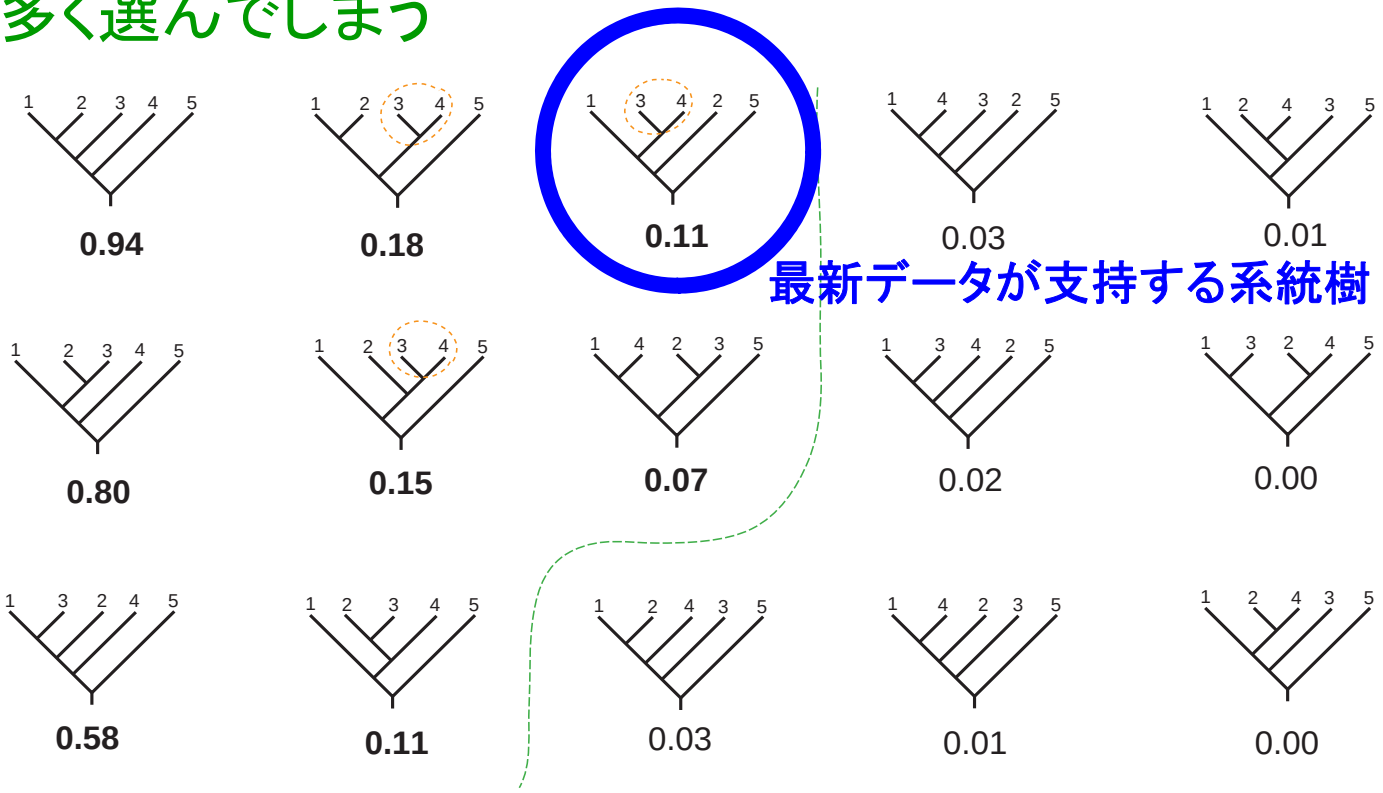
ブートストラップ確率はランダムネスを過小評価 ➡ 偽物の「発見」につながる (false positives)

Mathematical reason: Hypothesis region is convex and the curvature is positive

Shimodaira-Hasegawa 検定 (多重比較法の一種)

木を多く選んでしまう

Shimodaira (1998) *AISM*　　Shimodaira and Hasegawa (1999) *Mol.Biol.Evol.*

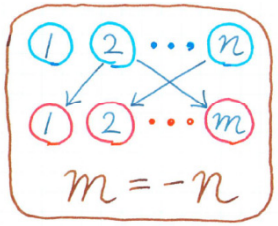


最新データが支持する系統樹

SH検定はランダムネスを過大評価 ➡ 本当の発見を見逃してしまう

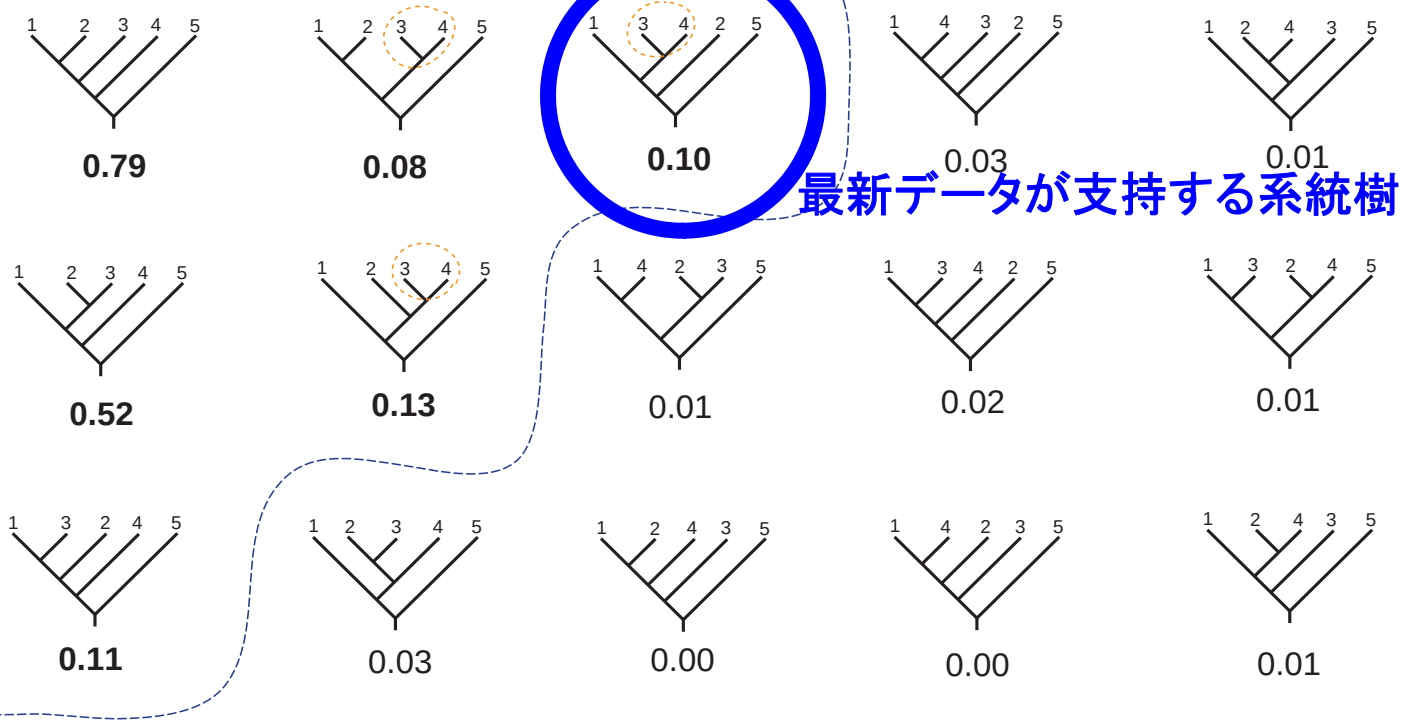
Mathematical reason: Multiple comparisons method evaluates the worst-case scenario and is conservative

マルチスケール・ブートストラップ（サンプルサイズを$m=-n$にする）



Shimodaira and Hasegawa (2001) *Bioinformatics*
Shimodaira (2002) *Syst.Biol.* Shimodaira (2004) Ann. Stat.
Shimodaira (2008) JSPI Shimodaira (2010) AISM

ちょうどいい個数の木が選ばれる



マルチスケール・ブートストラップ法はランダムネスを適切に評価（バイアスゼロ）

➡ ランダムネスは消えないが，無知の程度を正しく知ることができる．

32種の哺乳類の場合

AU：マルチスケール・ブートストラップ法

Table 2. p-values for the fifteen constrained candidate tree topologies.

model	$\Delta\ell$	PP1	PP2	BP	AU	KH	SH	WSH	tree topology
T_1	0.0	**0.28**	**0.61**	**0.23**	**0.69**	**0.55**	**0.97**	**0.95**	((G1,G2),(G3,G4),G5)
T_2	1.5	**0.49**	**0.14**	**0.28**	**0.60**	**0.46**	**0.83**	**0.86**	((G1,(G2,G3)),G4,G5)
T_3	1.7	**0.15**	**0.12**	**0.16**	**0.47**	**0.41**	**0.84**	**0.84**	(((G1,G2),G3),G4,G5)
T_4	1.9	**0.06**	**0.09**	**0.13**	**0.45**	**0.33**	**0.84**	**0.81**	(G1,(G2,(G3,G4)),G5)
T_5	2.6	0.01	0.04	**0.09**	**0.37**	**0.27**	**0.80**	**0.73**	((G1,(G3,G4)),G2,G5)
T_6	6.2	0.00	0.00	0.02	**0.16**	**0.15**	**0.64**	**0.54**	(((G1,G2),G4),G3,G5)
T_7	6.8	0.00	0.00	0.03	**0.25**	**0.28**	**0.58**	**0.61**	((G1,G4),(G2,G3),G5)
T_8	8.3	0.00	0.00	0.01	**0.08**	**0.23**	**0.51**	**0.40**	(G1,((G2,G3),G4),G5)
T_9	8.7	0.00	0.00	0.04	**0.25**	**0.21**	**0.50**	**0.66**	(((G1,G4),G2),G3,G5)
T_{10}	9.9	0.00	0.00	0.02	**0.14**	**0.18**	**0.43**	**0.59**	(((G1,G4),G3),G2,G5)
T_{11}	12.7	0.00	0.00	0.00	0.00	**0.10**	**0.29**	**0.20**	(((G1,G3),G2),G4,G5)
T_{12}	15.9	0.00	0.00	0.00	0.01	**0.05**	**0.17**	**0.27**	((G1,(G2,G4)),G3,G5)
T_{13}	18.6	0.00	0.00	0.00	0.00	0.03	**0.09**	**0.13**	(G1,((G2,G4),G3),G5)
T_{14}	18.8	0.00	0.00	0.00	0.00	0.02	**0.09**	**0.09**	(((G1,G3),G4),G2,G5)
T_{15}	21.5	0.00	0.00	0.00	0.00	0.01	0.04	**0.10**	((G1,G3),(G2,G4),G5)

Note: Only the fifteen candidate tree topologies are considered; the subtree topologies for G1, ..., G5 are specified in Fig. 1. $\Delta\ell$ denotes the log-likelihood difference from the ML topology. The trees are numbered by increasing order of $\Delta\ell$. PP1 denotes the PP calculated by the MCMCMC using MrBayes with clade constraints, and PP2 denotes the PP calculated by the BIC approximation. p-values ≥ 0.05 are in boldface.

Recent Results

③ ← ((G1,G2),(G3,G4),G5)

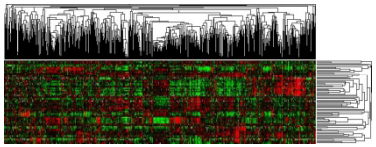
② ← (((G1,G2),G3),G4,G5)

① ← (((G1,G2),G4),G3,G5)

Nishihara et al (2008)

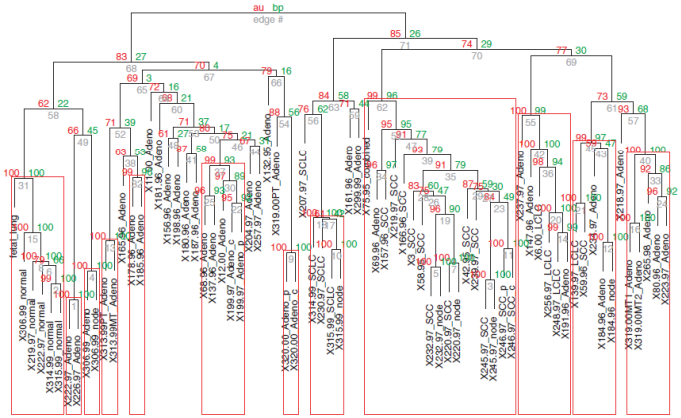
Genome-scale analysis of 1 Mbp

マルチスケール・ブートストラップの応用



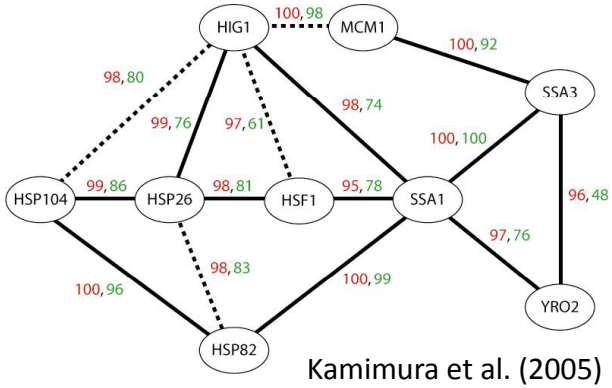
遺伝子発現データ
（DNAチップ）

癌のタイプ分類



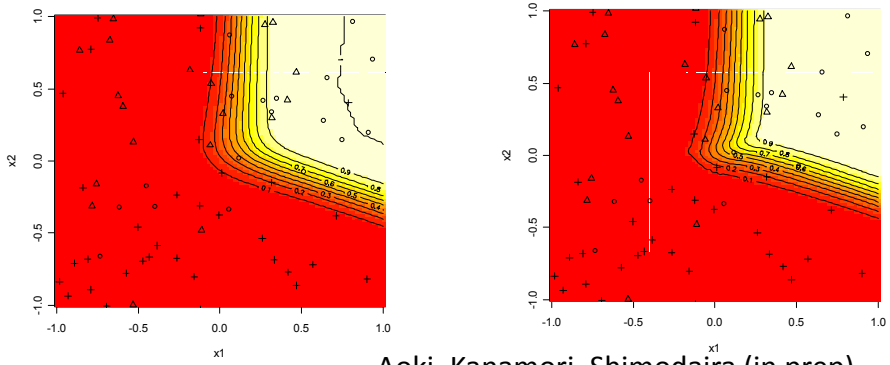
Suzuki and Shimodaira (2006)

遺伝子制御ネットワーク



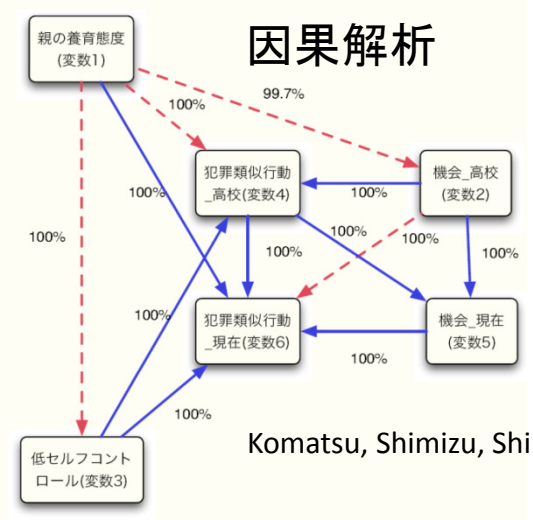
Kamimura et al. (2005)

機械学習



Aoki, Kanamori, Shimodaira (in prep)

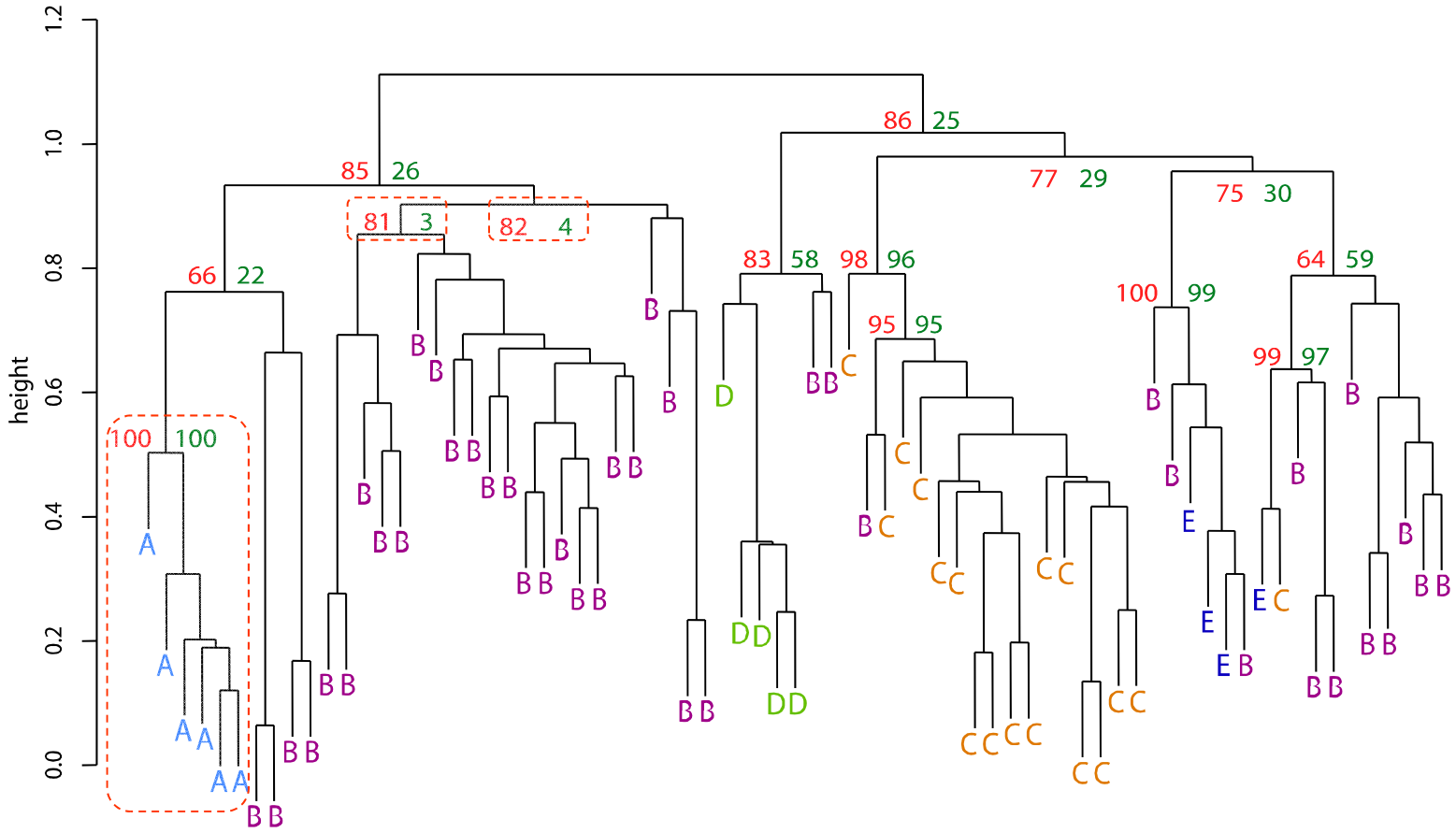
因果解析



Komatsu, Shimizu, Shimodaira (2010)

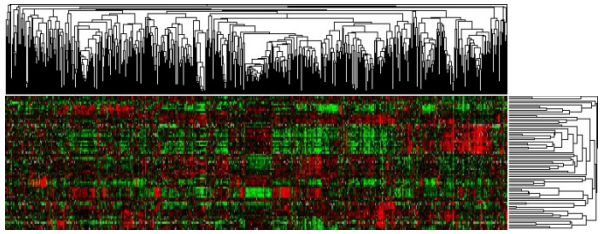
マイクロアレイから癌のタイプ分類

各枝の数値　＝　左：マルチスケール法　右：ブートストラップ確率



並列計算でスピードアップ
(下平研究室　2008/04/10)

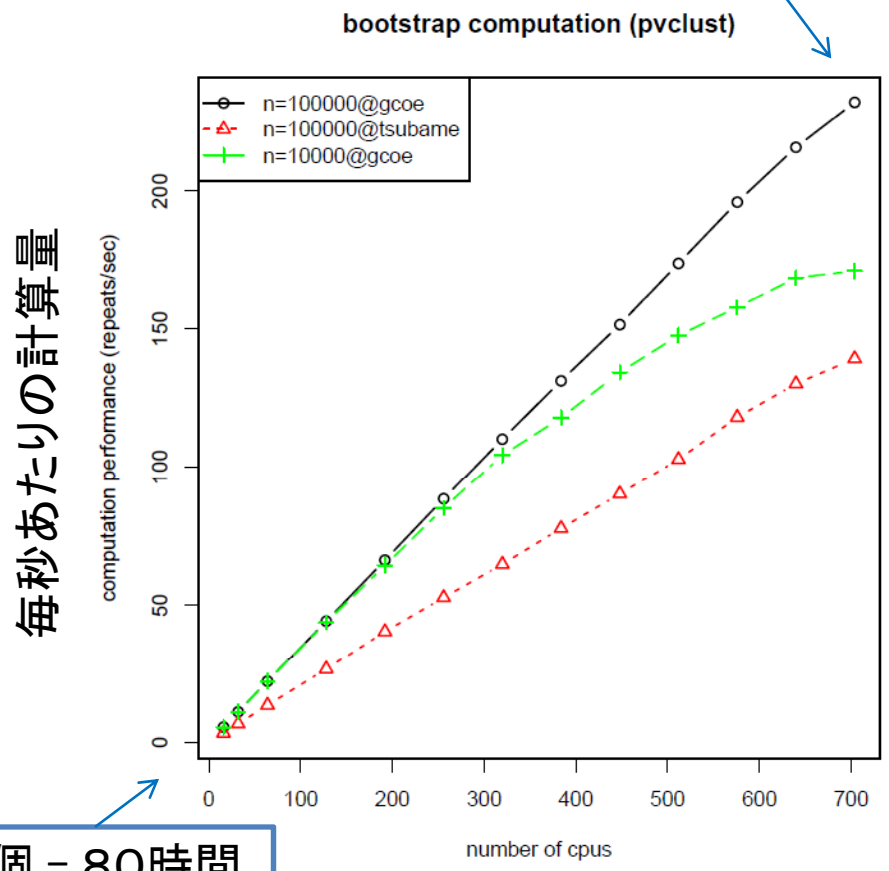
問題：　DNAチップから癌の診断



東工大GSICのスパコン
TSUBAME

10000コア

700個 = 7分



1個 = 80時間

計算プロセッサ（コア）の個数

Bioinformatics April 4, 2006

Data and text mining

Pvclust: an R package for assessing the uncertainty in hierarchical clustering

Ryota Suzuki[1,2,*] and Hidetoshi Shimodaira[1]

[1]Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan and [2]Ef-prime, Inc., 2-17-5 Nihonbashi-Kayabacho, Chuo-ku, Tokyo 103-0025, Japan

Received on August 30, 2005; revised on March 6, 2006; accepted on March 25, 2006
Advance Access publication April 4, 2006
Associate Editor: Satoru Miyano

“pvclust” already in use

Software developed by a master course student at Shimodaira-lab

下平研の学生が作成したソフトウエアが，すぐに他の研究で使われた
ES細胞の研究（話題の iPS細胞に関係する）

PNAS January 24, 2006

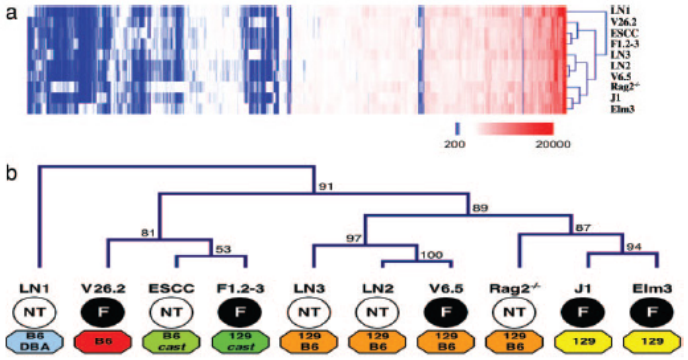
ES cells derived from cloned...

ES cells derived from cloned and fertilized blastocysts are transcriptionally and functionally indistinguishable

Tobias Brambrink*, Konrad Hochedlinger*, George Bell*, and Rudolf Jaenisch*†‡

Brambrink et al.

Gene Expression Profiles (DNA microarray)



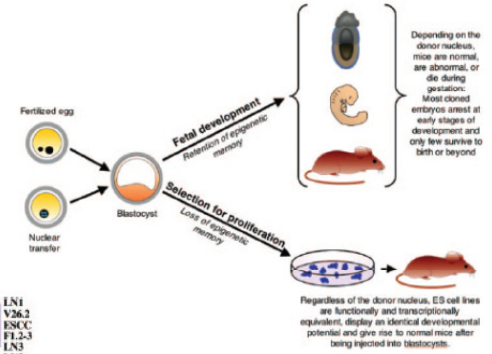
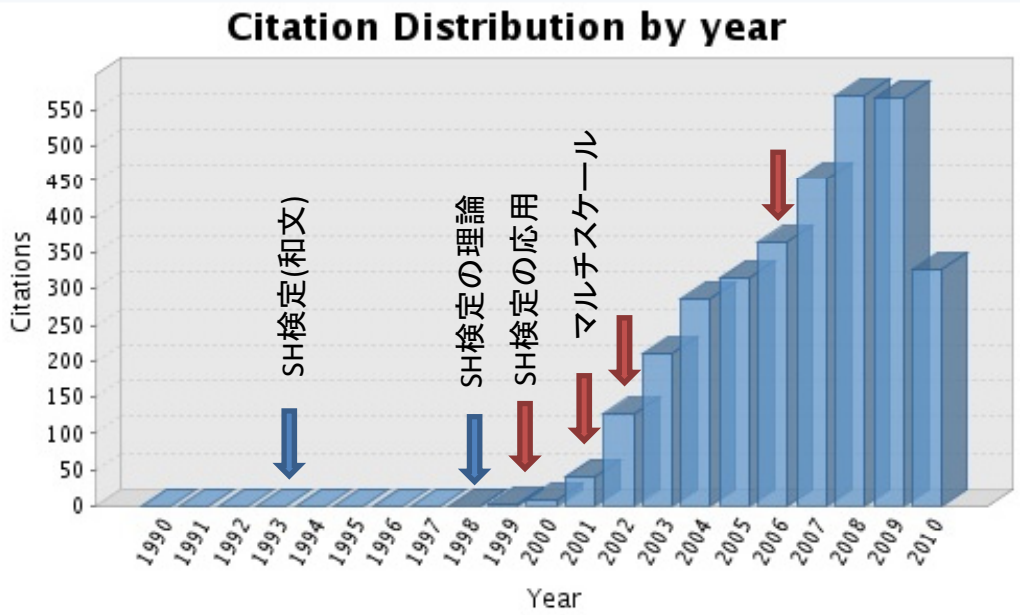


Fig. 2. Hierarchical clustering of individual ESC line expression profiles. (*a*) Heat map of clustering results (blue, no or very low expression; white, low expression; red, high expression). (*b*) Sample tree obtained from hierarchical clustering. ES cell line expression profiles cluster by genetic background (colored octagons) rather than by type of donor blastocyst (NT, cloned; F, fertilized; numbers next to nodes display multiscale bootstrap resampling probability based on 10,000 replications).

論文被引用回数

http://www.researcherid.com/citation/B-9127-2008



1. Title: Multiple comparisons of log-likelihoods with applications to phylogenetic inference
Author(s): SHIMODAIRA, H; HASEGAWA, M
Source: MOLECULAR BIOLOGY AND EVOLUTION Volume: 16 Issue: 8 Pages: 1114-1116 Published: AUG 1999
Times Cited: 1735

2. Title: An approximately unbiased test of phylogenetic tree selection
Author(s): SHIMODAIRA, H
Source: SYSTEMATIC BIOLOGY Volume: 51 Issue: 3 Pages: 492-508 Published: MAY-JUN 2002
Times Cited: 505
DOI: 10.1080/10635150290069913

3. Title: CONSEL: for assessing the confidence of phylogenetic tree selection
Author(s): SHIMODAIRA, H; HASEGAWA, M
Source: BIOINFORMATICS Volume: 17 Issue: 12 Pages: 1246-1247 Published: DEC 2001
Times Cited: 500

4. Title: Mitochondrial genome variation in Eastern Asia and the peopling of Japan
Author(s): TANAKA, M; CABRERA, VM; GONZALEZ, AM; et al.
Source: GENOME RESEARCH Volume: 14 Issue: 10A Pages: 1832-1850 Published: OCT 2004
Times Cited: 156
DOI: 10.1101/gr.2286304

5. Title: Pvclust: an R package for assessing the uncertainty in hierarchical clustering
Author(s): SUZUKI, R; SHIMODAIRA, H
Source: BIOINFORMATICS Volume: 22 Issue: 12 Pages: 1540-1542 Published: JUN 15 2006
Times Cited: 72
DOI: 10.1093/bioinformatics/btl117

Total Articles in Publication List: 19

Articles With Citation Data: 19

Sum of the Times Cited: 3267

Average Citations per Article: 171.95

h-index: 12

Last Updated: 06/21/2010 12:54 GMT

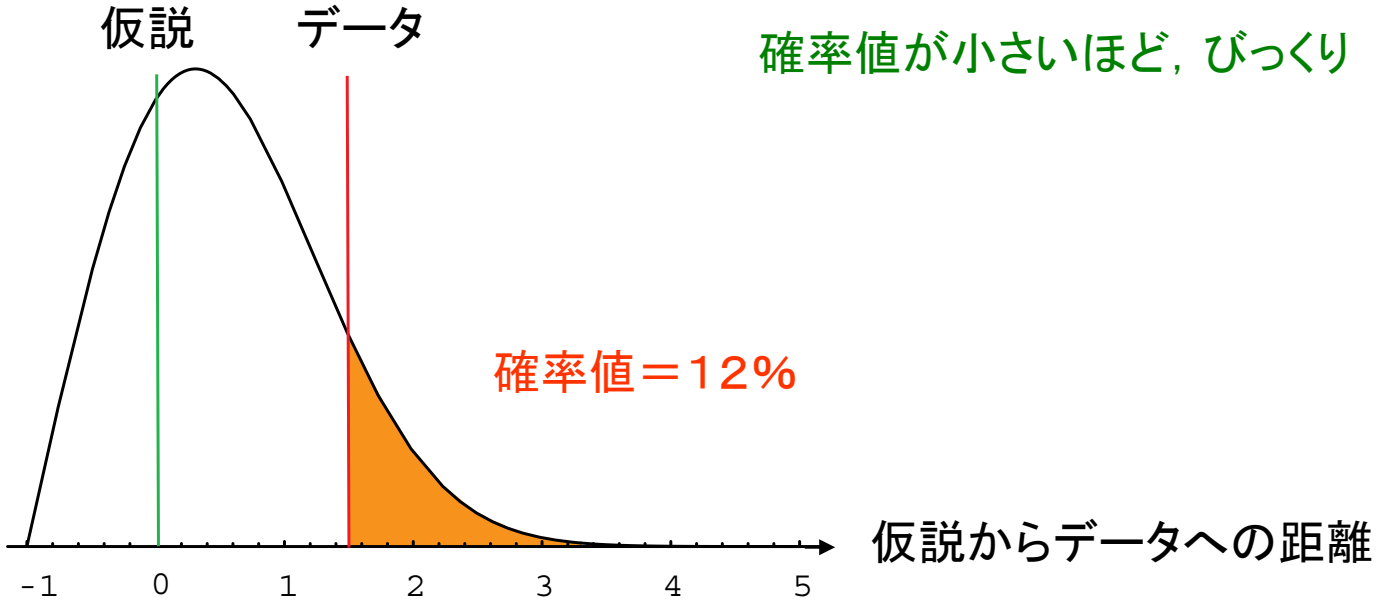
合計回数

マルチスケール・ブートストラップ法の理論

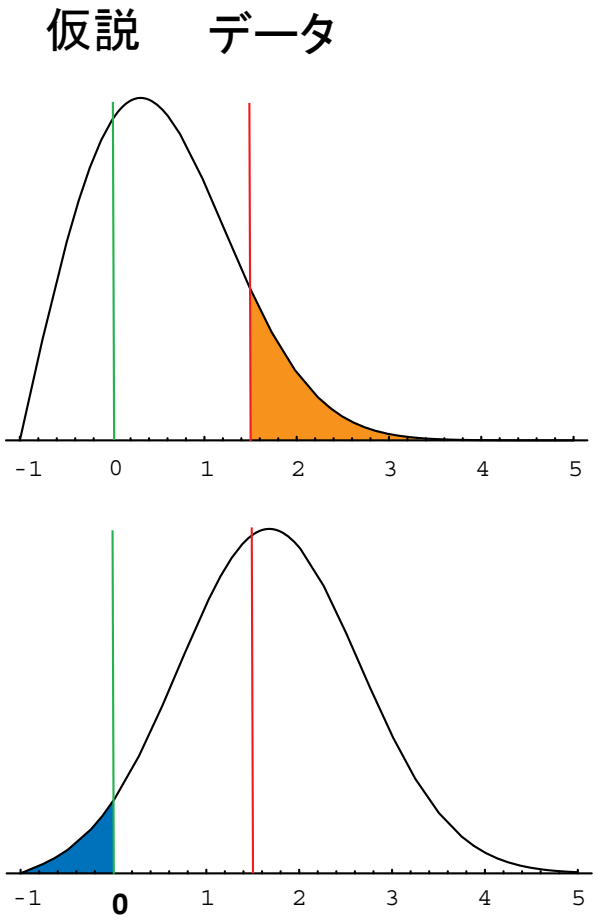
- 仮説検定の確率値
- 確率値とブートストラップ確率の関係
- マルチスケール・ブートストラップ法の原理
- ブートストラップ確率のスケーリング則

仮説検定

- データのバラツキによる見せかけか否かを判定する統計的手続き
- 「驚き」の程度　⇒　確率値 (p-value, p-値)



確率値とブートストラップ確率の関係



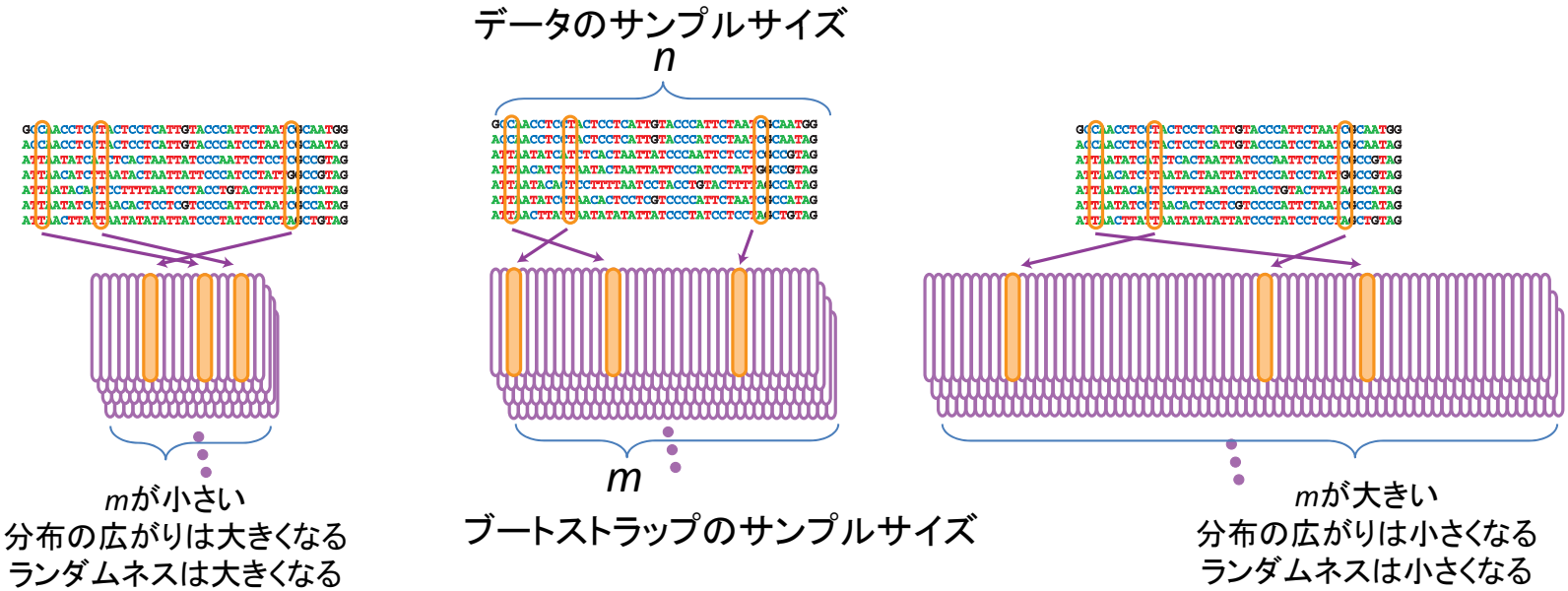
検定の確率値： 12 %

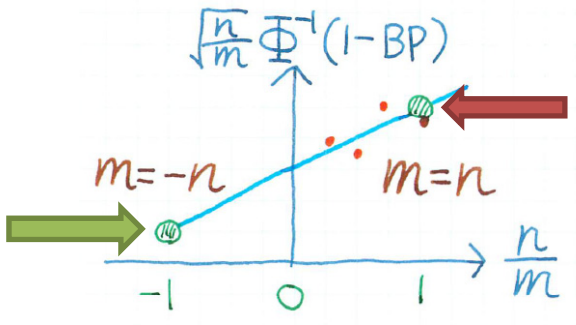
もし分布が左右対称ならば両者は一致する

ブートストラップ確率：3 %

もし下図の分布を左右反転させてからブートストラップ確率を計算すれば上図に近づく

マルチスケール・ブートストラップの原理
$m = -n$ への外挿（“m out of n bootstrap”）



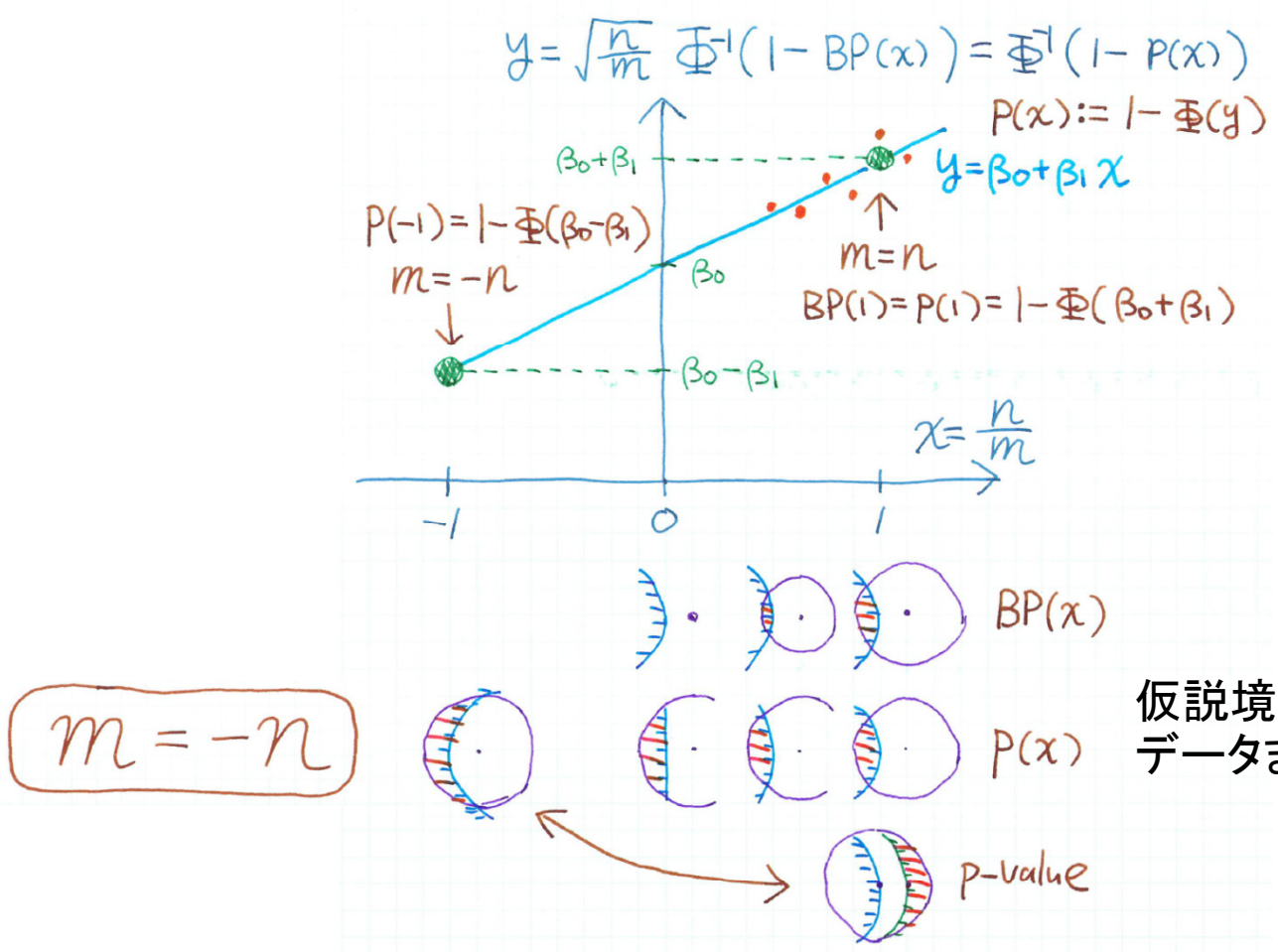


ココが$m=n$の通常の
ブートストラップ

ココが$m=-n$への外挿

分布が左右反転する？
ランダムネスがひっくり返る？

マルチスケール・ブートストラップ法の幾何学

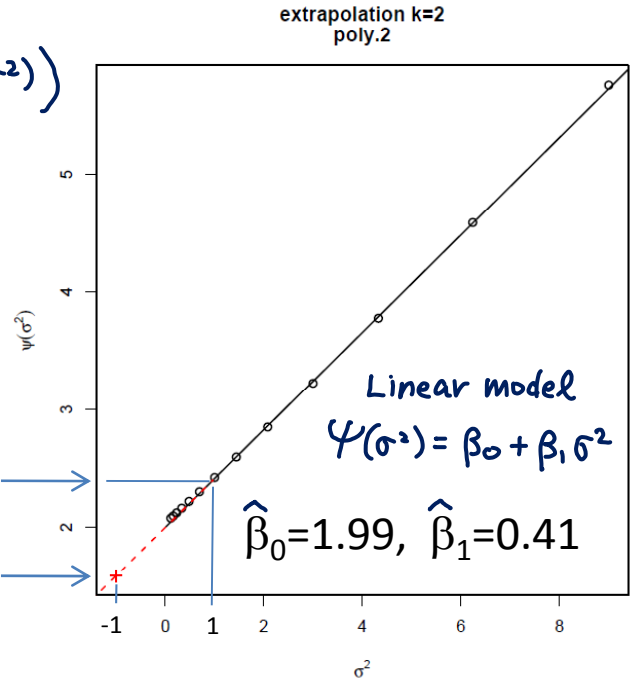


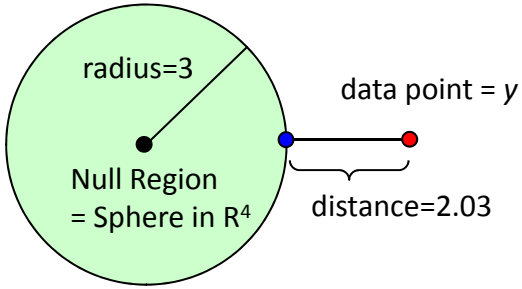
仮説境界を表す曲面の曲率と
データまでの距離が関係している

Example: exact p=0.05

Normalized bootstrap z-value

$$\Psi(\sigma^2) = -\sigma\,\Phi^{-1}(BP(\sigma^2))$$





$\Phi(-\Psi(\sigma^2))$	$\Psi(\sigma^2)$
BP=0.0078 (k=1)	2.40
AU=0.057 (k=2)	1.58

Quadratic model

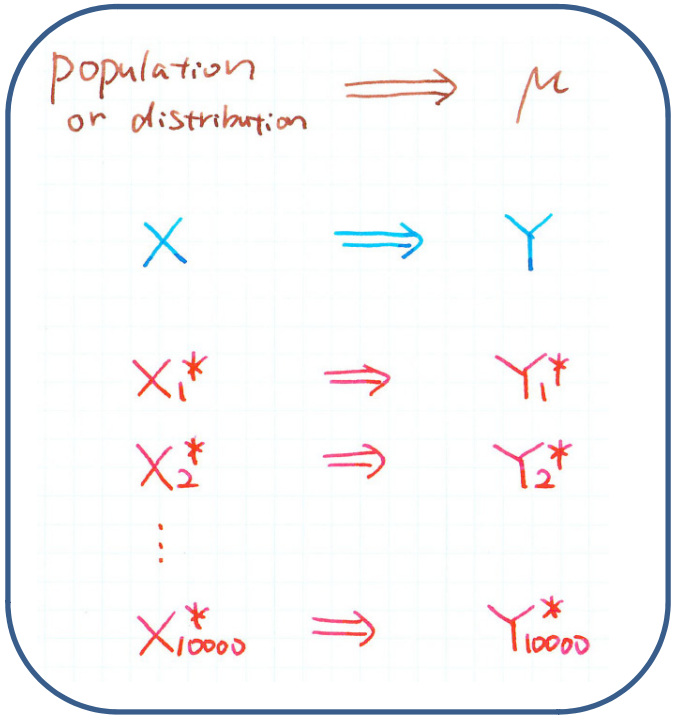
$$\Psi(\sigma^2) = \beta_0 + \beta_1\sigma^2 + \beta_2\sigma^4$$

AU=0.052 (k=3) $\quad \hat{\beta}_0$=2.02, $\hat{\beta}_1$=0.39, $\hat{\beta}_2$=0.024

Geometrical Interpretations of the Coefficients

- β_0 : signed distance
- β_1 : mean curvature
- β_2 : mean 4th derivatives

A Simplified Model



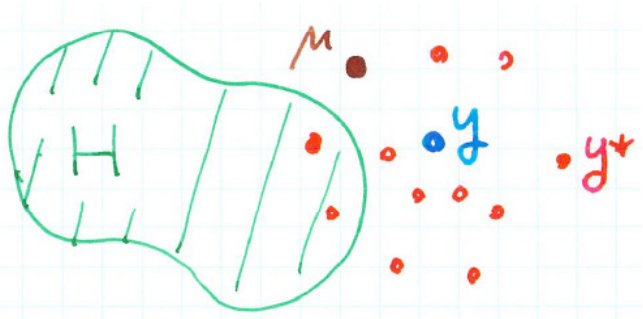
Multivariate Normal Model

$$Y \sim N(\mu, I)$$

$$Y^* \mid y \sim N(y, \sigma^2 I)$$

$$\sigma^2 = \frac{n}{m}$$

$$BP = P_{\sigma^2}(Y^* \in H \mid y)$$



Why m=-n?
(only for illustration purpose)

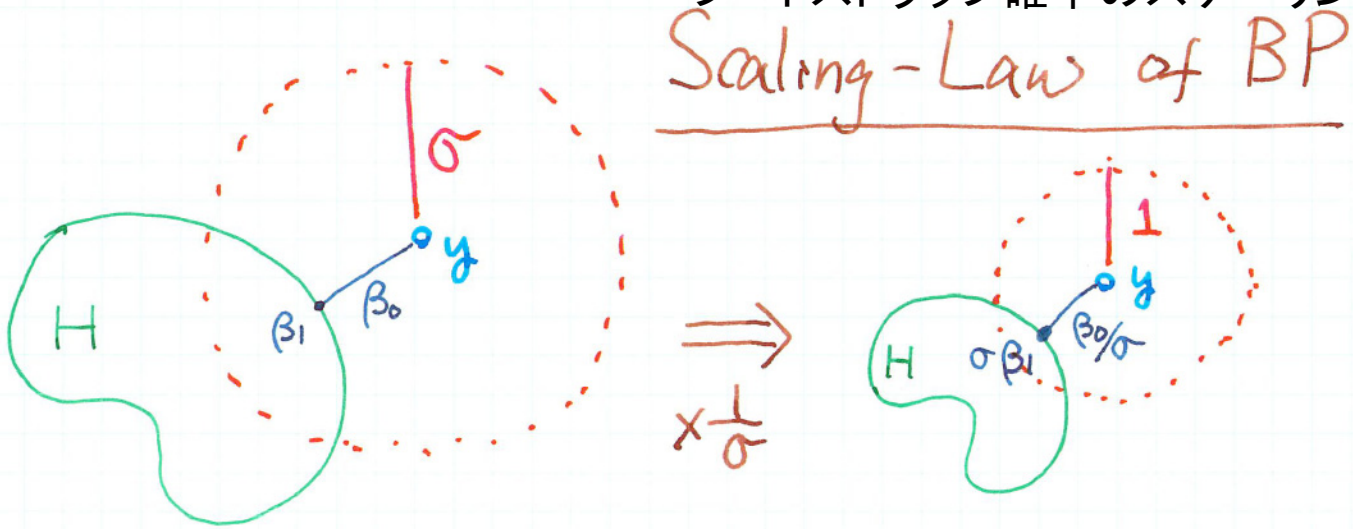
Intuitive Explanation of $m=-n$

$$\left.\begin{array}{l} Y \sim N(\mu, I) \\ Y^* - Y \sim N(0, \sigma^2 I) \end{array}\right\} \Rightarrow \quad Y^* \sim N(\mu, (1+\sigma^2) I)$$

$\therefore \quad Y^* \equiv \mu \quad$ for $\quad \sigma^2 = -1 \quad$ or $\quad m = -n$.

Rescaling the Randomness: bridging Bayesian to Frequentist

ブートストラップ確率のスケーリング則



$$BP = 1 - \Phi[\ \text{distance} + \text{curvature} + \cdots\ ] \quad \text{for } \sigma = 1$$

$$= 1 - \Phi\left[\ \frac{\beta_0}{\sigma} + \sigma\beta_1 + \cdots\ \right] \quad \text{for } {}^{\forall}\sigma > 0$$

$$= 1 - \Phi\left[\ \frac{1}{\sigma}\{\ \beta_0 + \sigma^2\beta_1 + \cdots\ \}\ \right]$$

$$\therefore\ \sigma\Phi^{-1}(1 - BP) = \beta_0 + \sigma^2\beta_1 + \cdots$$

参考文献

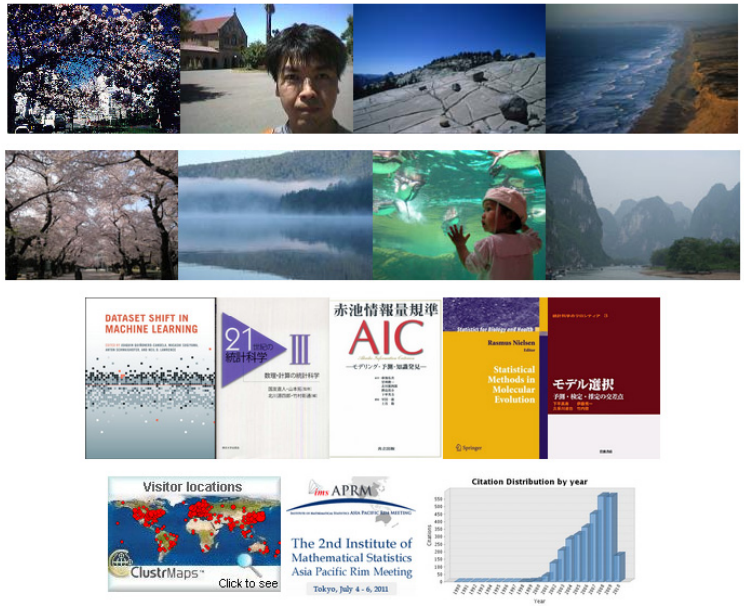
- 分子系統学　長谷川政美・岸野洋久　岩波書店　1996
- そのほか多数あります

amazon.co.jpで「分子系統学」を検索

下平　ホームページ

http://www.is.titech.ac.jp/~shimo/index-j.html

下平英寿　（しもだいらひでとし）



統計科学　&　バイオインフォマティクス

略歴 (Linkedin)，　文献リスト（被引用回数ランキング）

下平英寿　（しもだいらひでとし）

博士（工学), 准教授

〒152-8552　東京都目黒区大岡山 2-12-1-W8-46
東京工業大学　情報理工学研究科　数理・計算科学専攻

shimo (あっとまーく) is.titech.ac.jp
http://www.is.titech.ac.jp/~shimo/