

集団ゲノム学

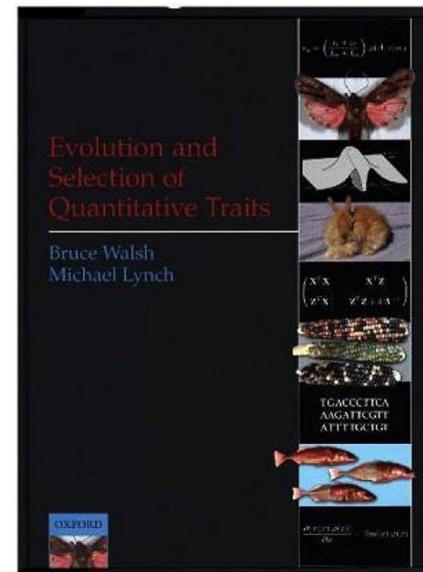
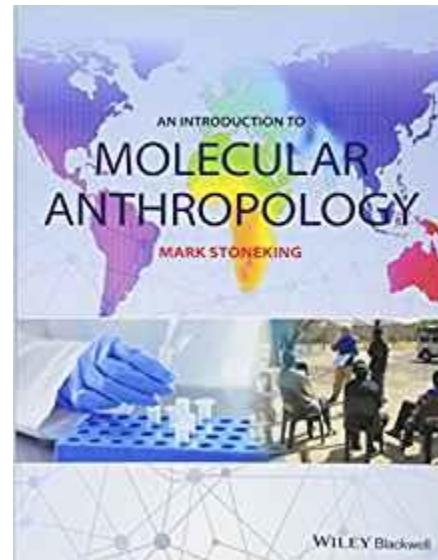
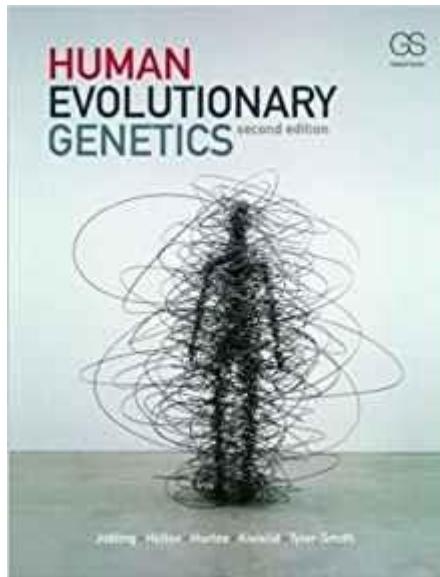
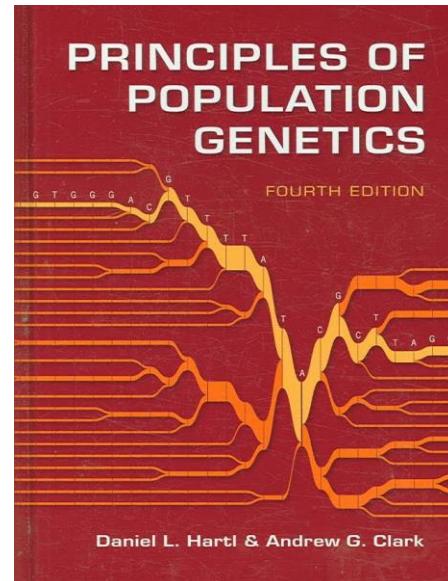
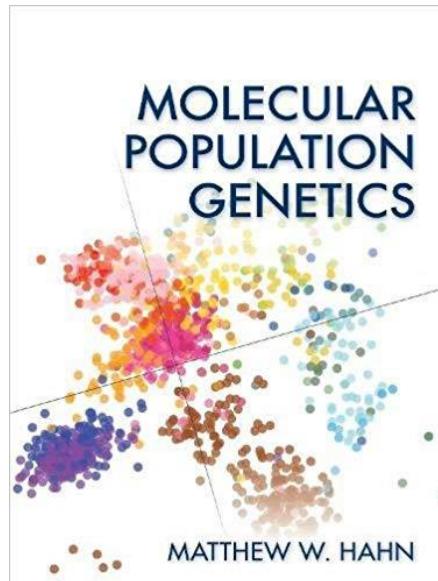
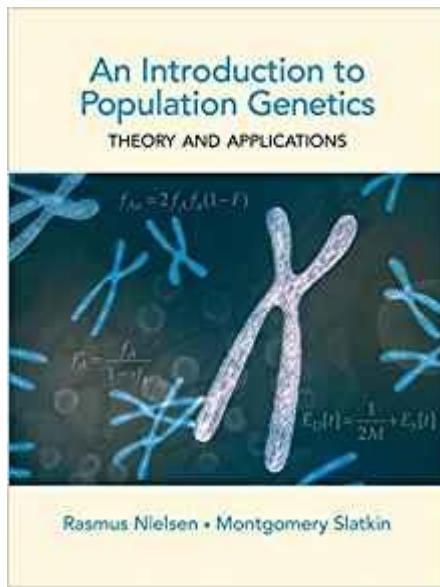
Population Genomics

木村亮介
琉球大学大学院医学研究科
人体解剖学講座

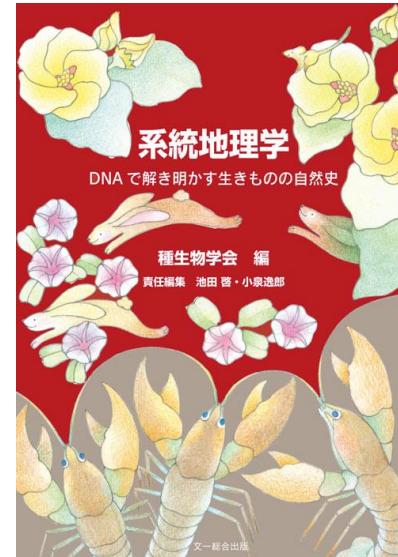
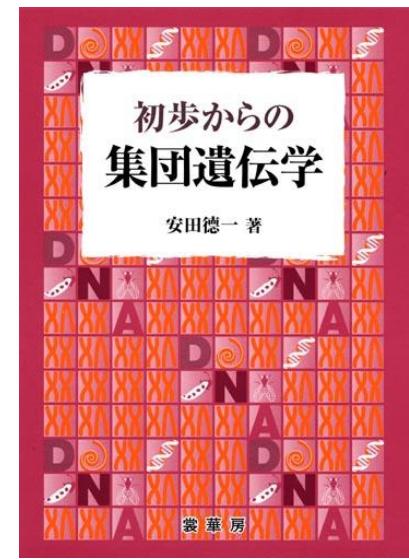
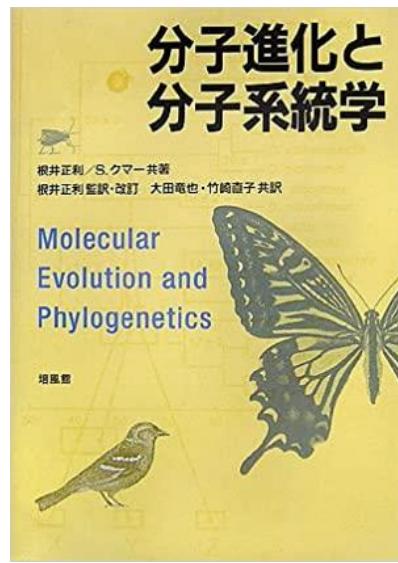
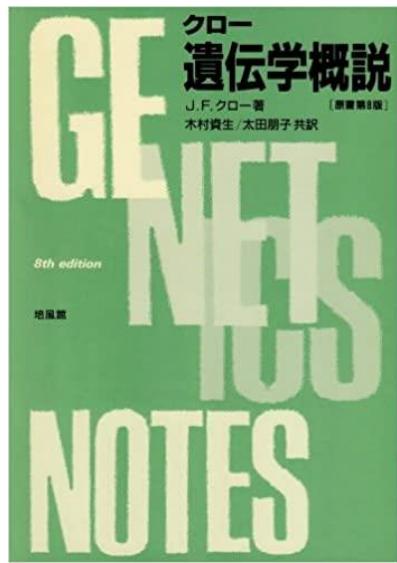
統合生物考古学セミナー
20231209



参考図書（英書）



参考図書（和書）



MENU

1. 集団遺伝学入門
2. ゲノム多様性を表現する
3. 琉球人集団の形成
4. ゲノム多様性と表現型多様性

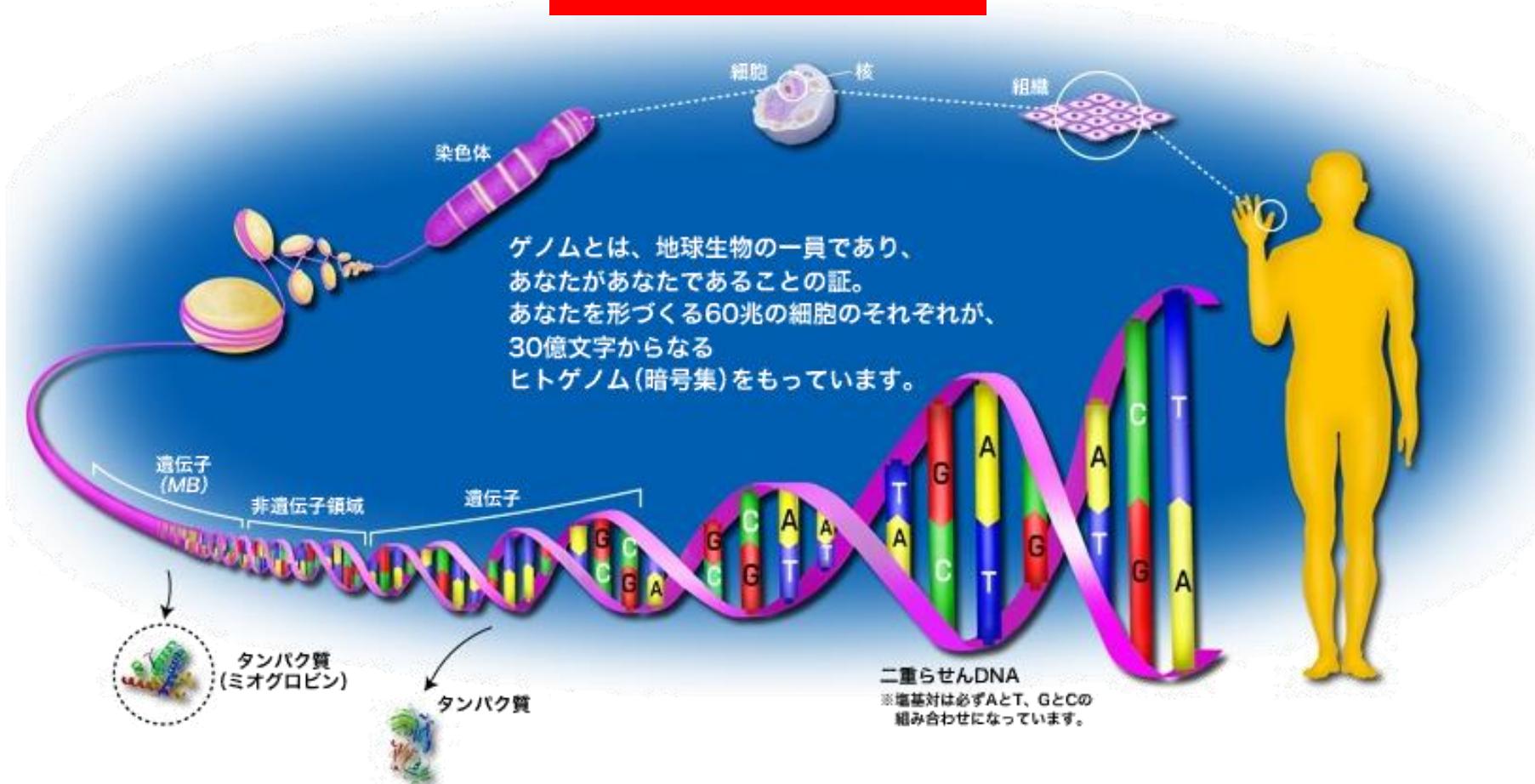
MENU

1. 集団遺伝学入門
2. ゲノム多様性を表現する
3. 琉球人集団の形成
4. ゲノム多様性と表現型多様性

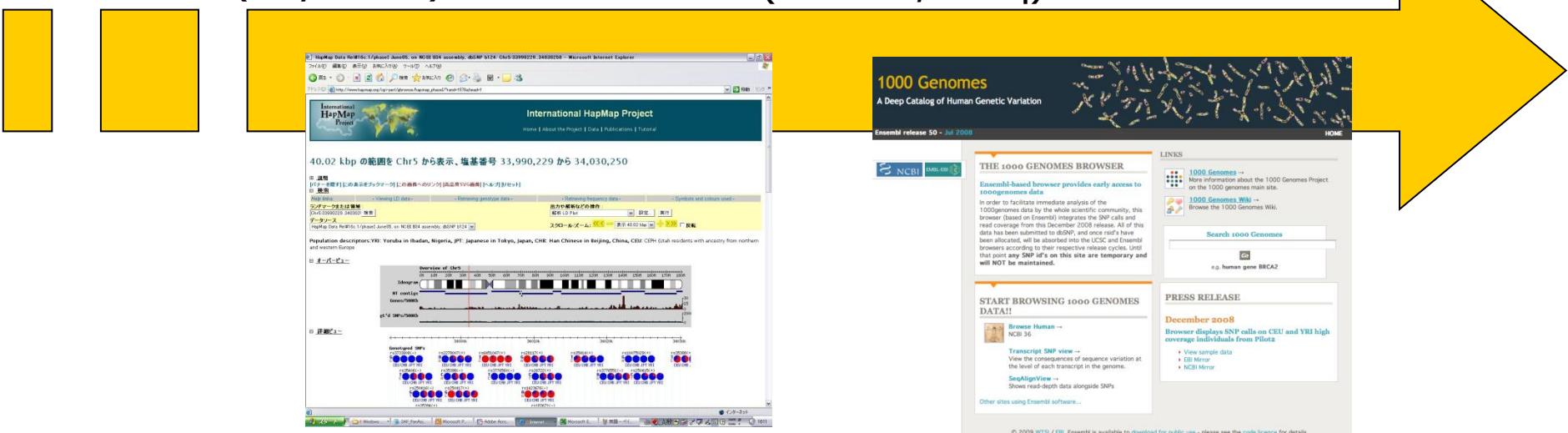
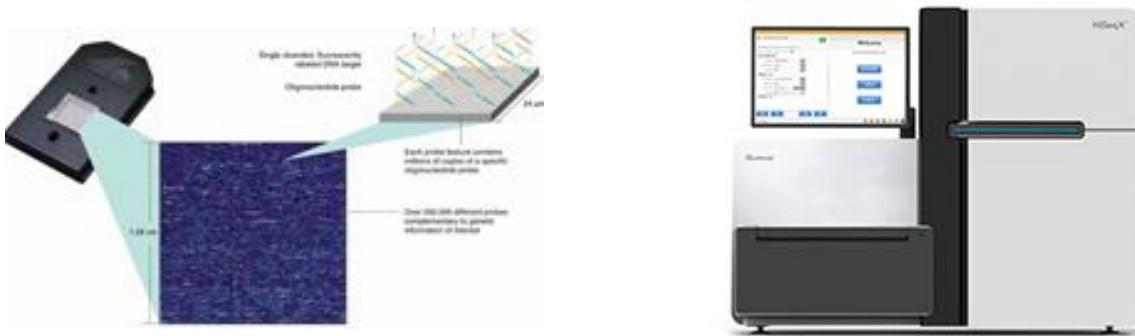
ヒトゲノムプロジェクト

2003年 すべての塩基配列の解読終了

ヒト1人分のゲノム
13年間 約3千億円



DNA解析技術と情報科学の革新



International HapMap Project

1000 Genome Project

ゲノム研究に飛躍的進展をもたらしている

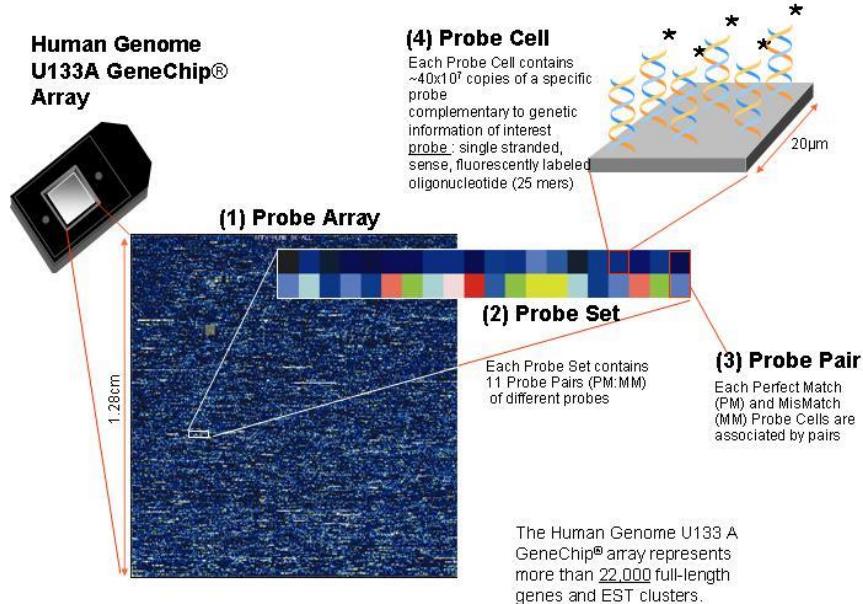
DNAマイクロアレイ (DNAチップ)

遺伝子発現解析：

- RNA発現の網羅的解析

ゲノム多様性解析：

- 研究室レベルでの大規模多型解析 (5万～100万 SNPs)
- ゲノムワイド関連解析：疾患感受性などのヒトの個体差に関する遺伝子多型の探索



マイクロアレイ技術の進歩

・国際HapMapプロジェクト



多型・頻度
連鎖不平衡の状態
ハプロタイプ構造



~2003

2003～
2005

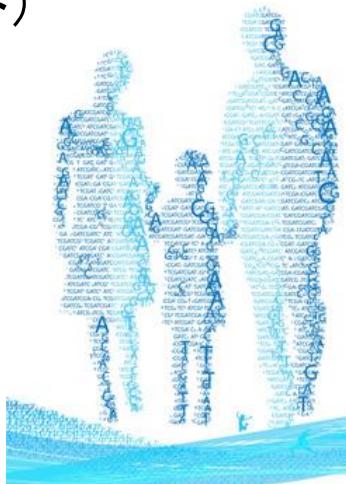
2005

2006

2007

2011

・ヒトゲノム配列解読完了(ヒトゲノムプロジェクト)



・Affymetrix
Mapping
500K Array

・Illumina
Human-
300K

・Affymetrix
SNP Array
6.0
・Illumina
Human-1M

・Illumina
Omni-5

ハイスループット遺伝子型決定のための
技術の発展

DNAシークエンス技術の発展

1975年
サンガー法



1986年
自動DNAシークエンサー



1990年
キャピラリー
DNAシークエンサー



2005年
Roche 454



→ ポリメラーゼ
2012年
PacBio RS



?年
Nanopore
核酸を物理的方法
で検出



2006年
Illumina
Genome Analyzer



2010年
Illumina
Hiseq2000

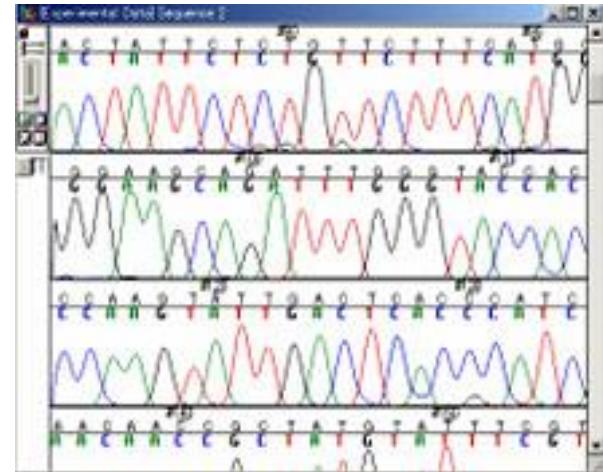
次世代型シーケンサー

キャピラリー型

DNA

PCR or クローニング

サンガー法による末端ラベル

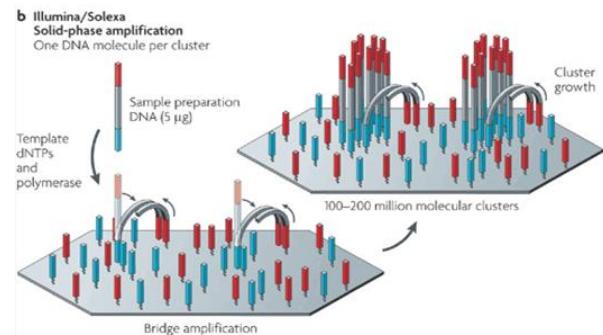


次世代型

DNA

断片化

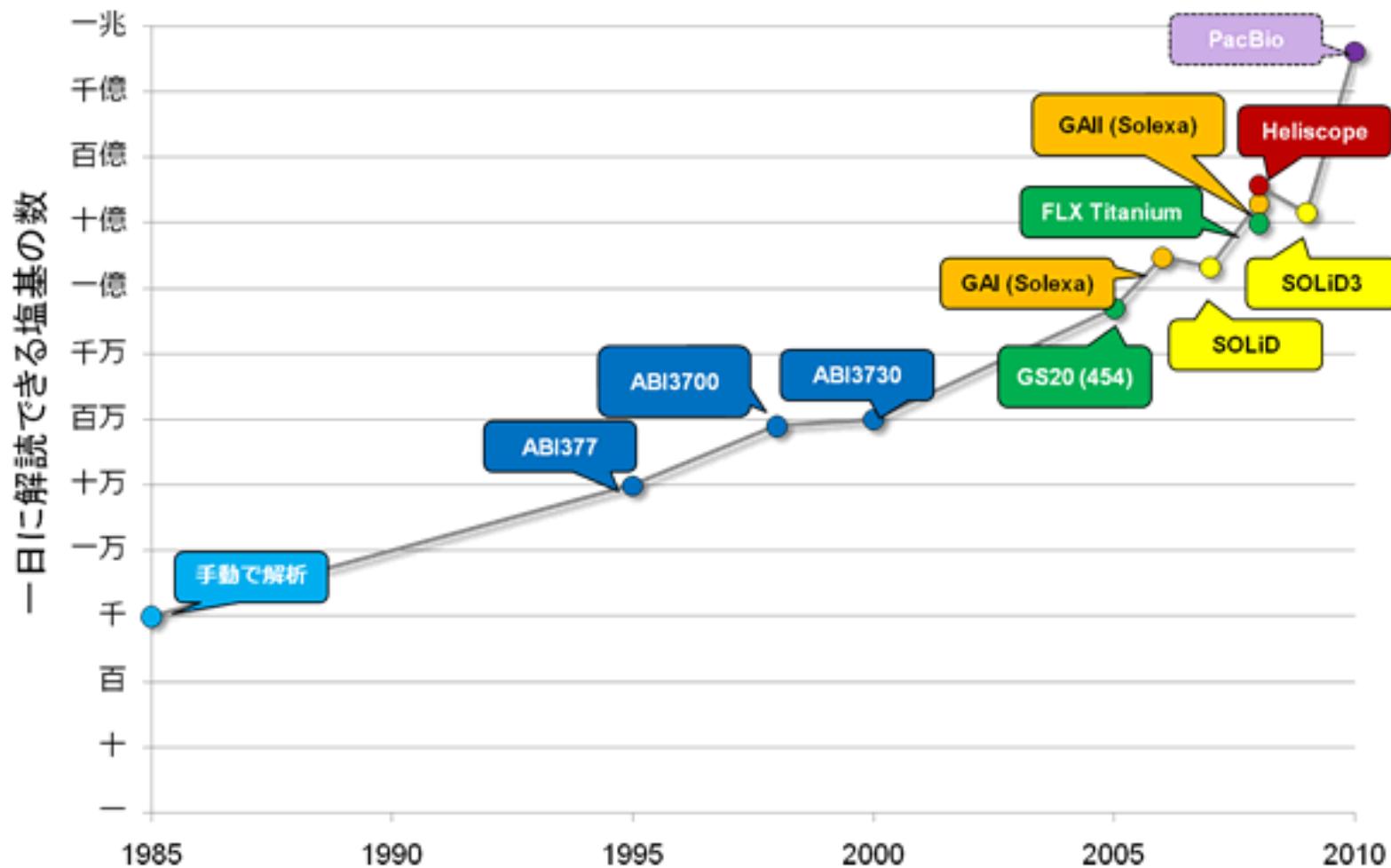
アダプター付加



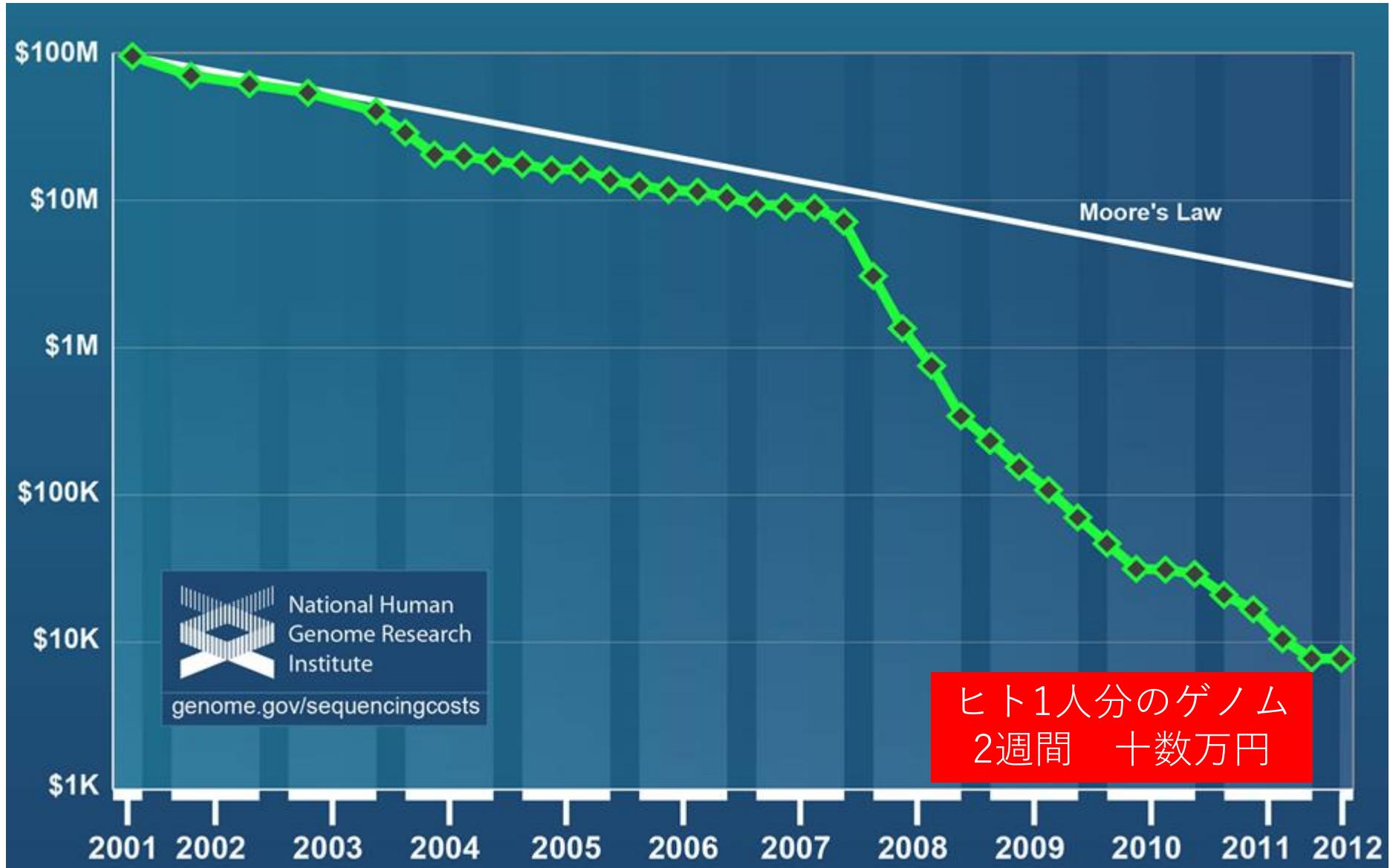
1 塩基伸長反応ごとに蛍光イメージを取り込み蛍光の色から各塩基を決定します。



DNAシークエンサーのスループット



一人分のゲノム解読にかかる費用



ヒト1人分のゲノム
2週間 十数万円



National Human
Genome Research
Institute

genome.gov/sequencingcosts

全ゲノムリシークエンシング -方法-

- 全ゲノムリシークエンシング：
Illumina HiSeqX (マクロジエン・ジャパン)



- 解析パイプライン：GATK Best-Practices



✓ Best-Practices

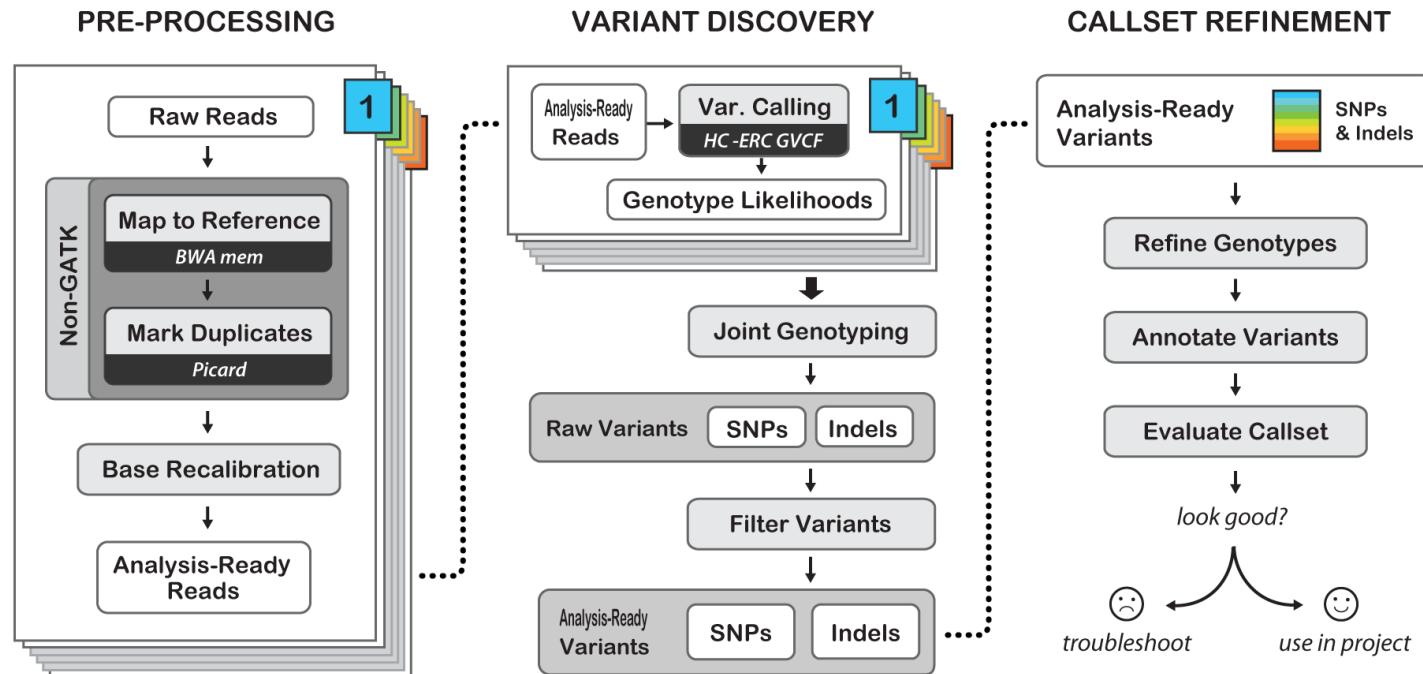
Trimmomatic_0.36

BWA_0.7.15

Samtool_1.5

Picard_2.10.6

Genome Analysis toolkit (GATK) 3.8-0



Best Practices for Germline SNPs and Indels in Whole Genomes and Exomes - June 2016

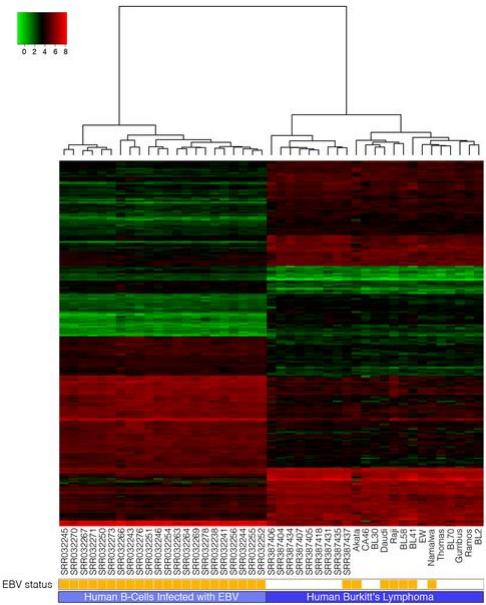
VCFファイル

敢えて例えるなら…

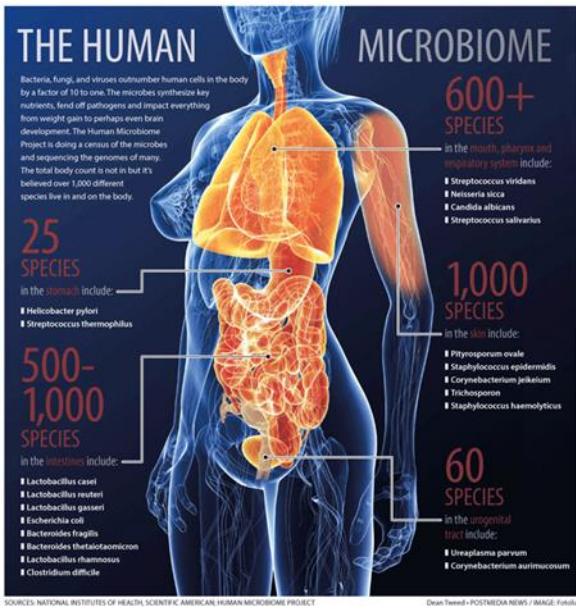
尾行



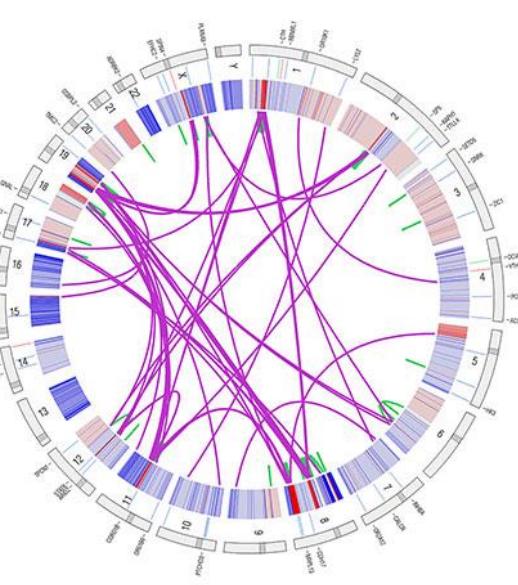
ゲノミクス解析技術の医学研究応用



Transcriptome



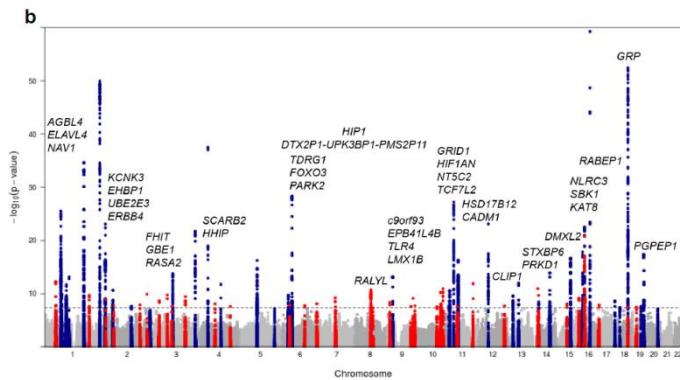
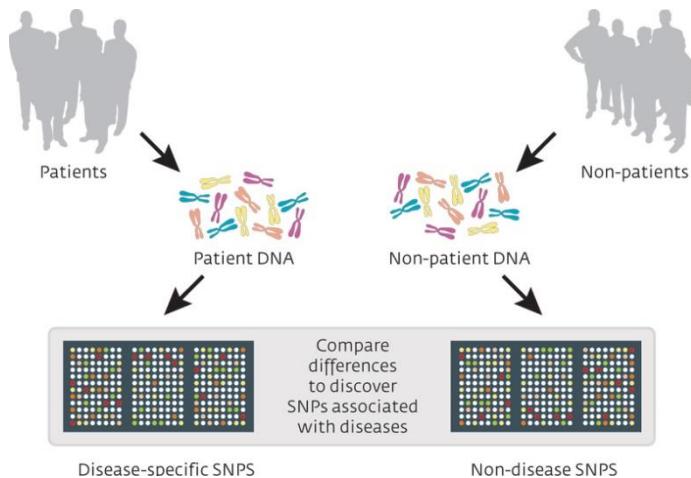
Microbiome



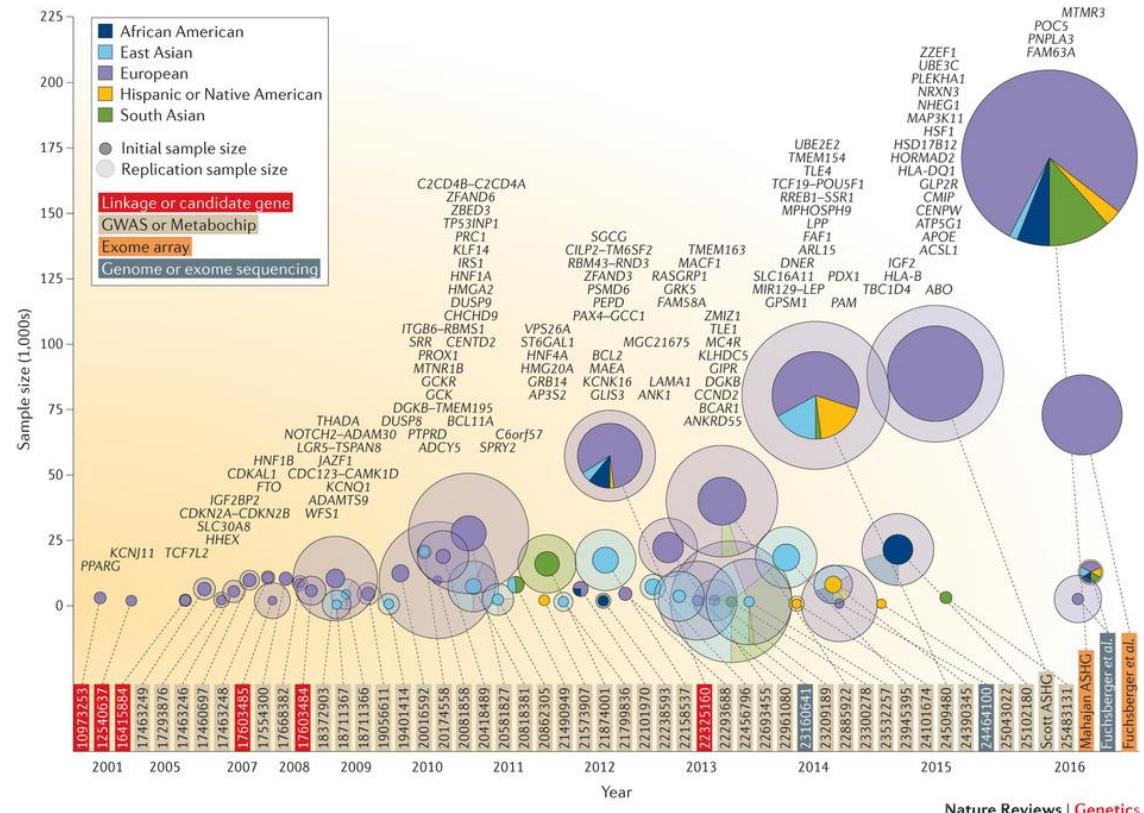
Cancer genomics

etc...

ゲノムワイド関連解析（GWAS）



Locke et al. 2015



Flannick & Florez 2016

ゲノム解析の飛躍的進展とその人類学的応用

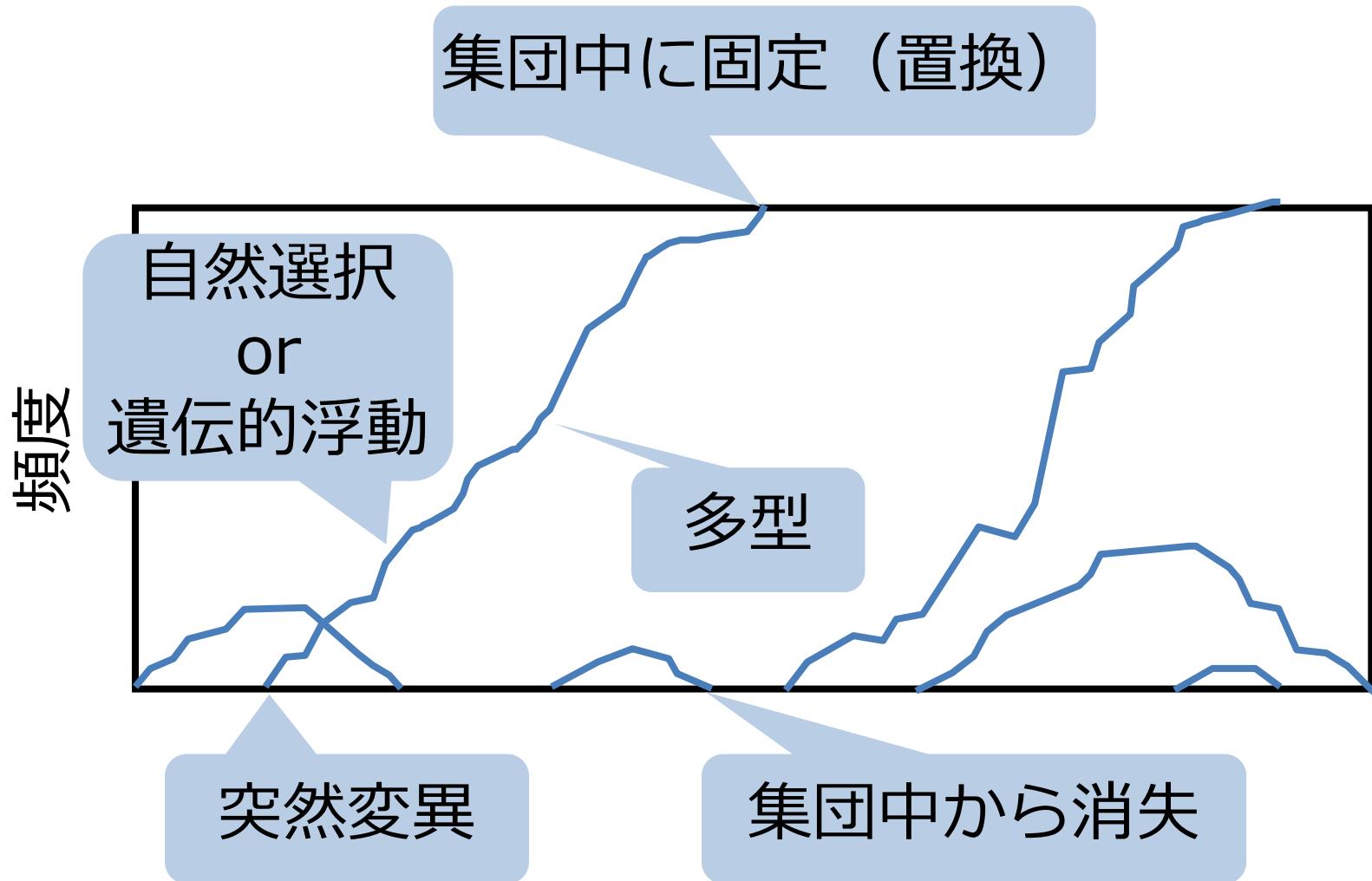
1. 疾患や形質に関連する遺伝子の同定
Genome-wide association study (GWAS)
2. 過去の人口動態や移住 (demography) 、
現在の集団構造 (structure) の高解像解析
3. 自然選択が働いたゲノム領域の探索
4. ネアンデルタール人などの古代ゲノム解析

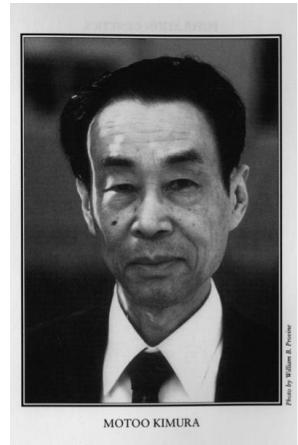
など・・・

研究戦略の大転換！

進化の原理

突然変異 + 集団中への固定
の繰り返し





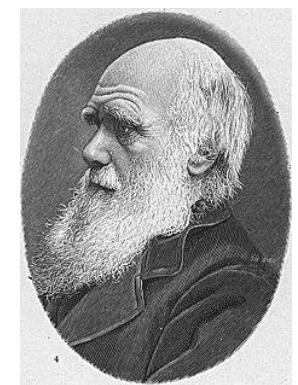
中立進化： 遺伝的浮動による

表現型に影響を及ぼさない突然変異

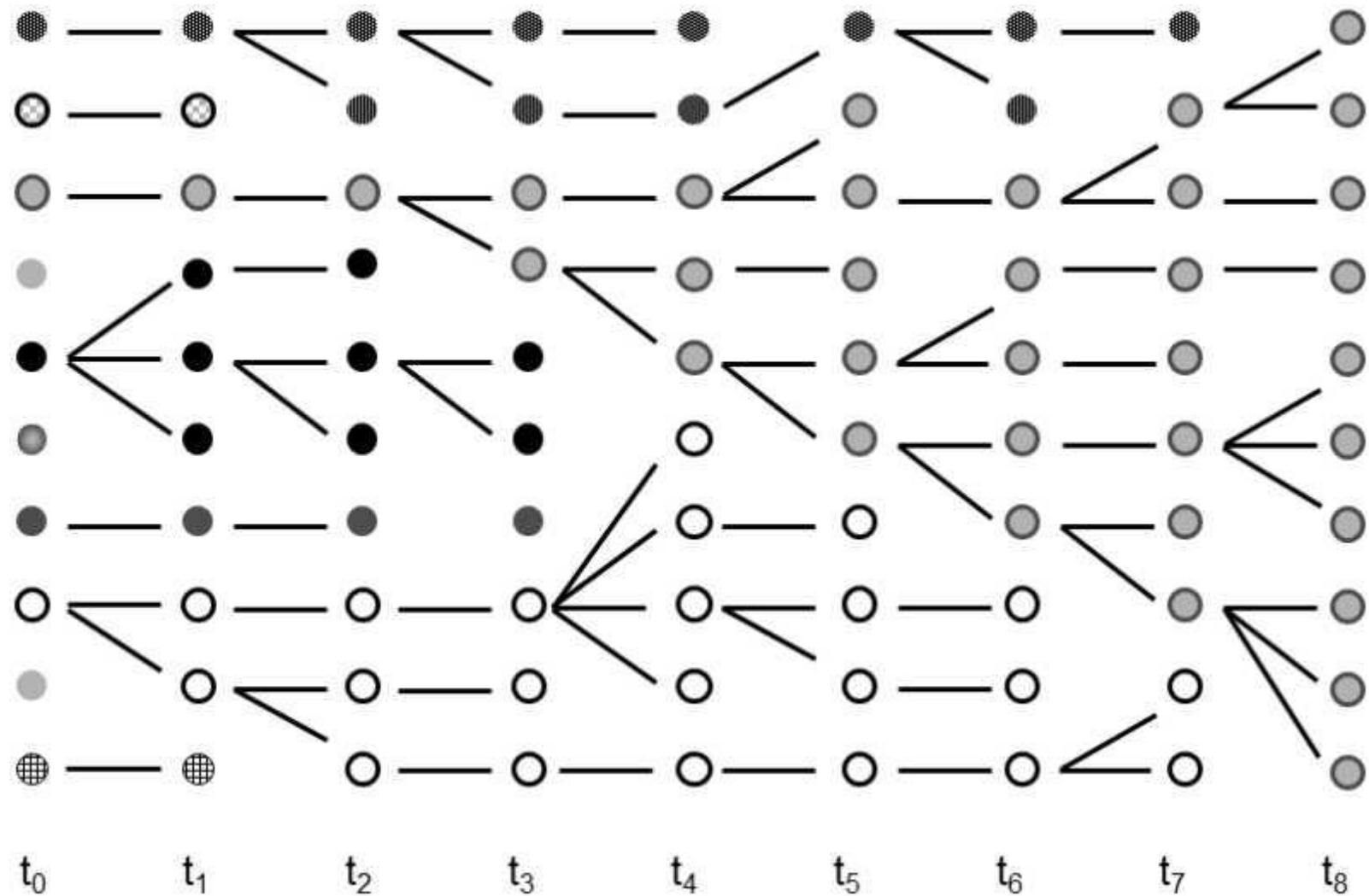
適応度に影響しない表現型を変える突然変異 など

適応進化： 自然（性）選択による

適応度に影響する表現型を変える突然変異

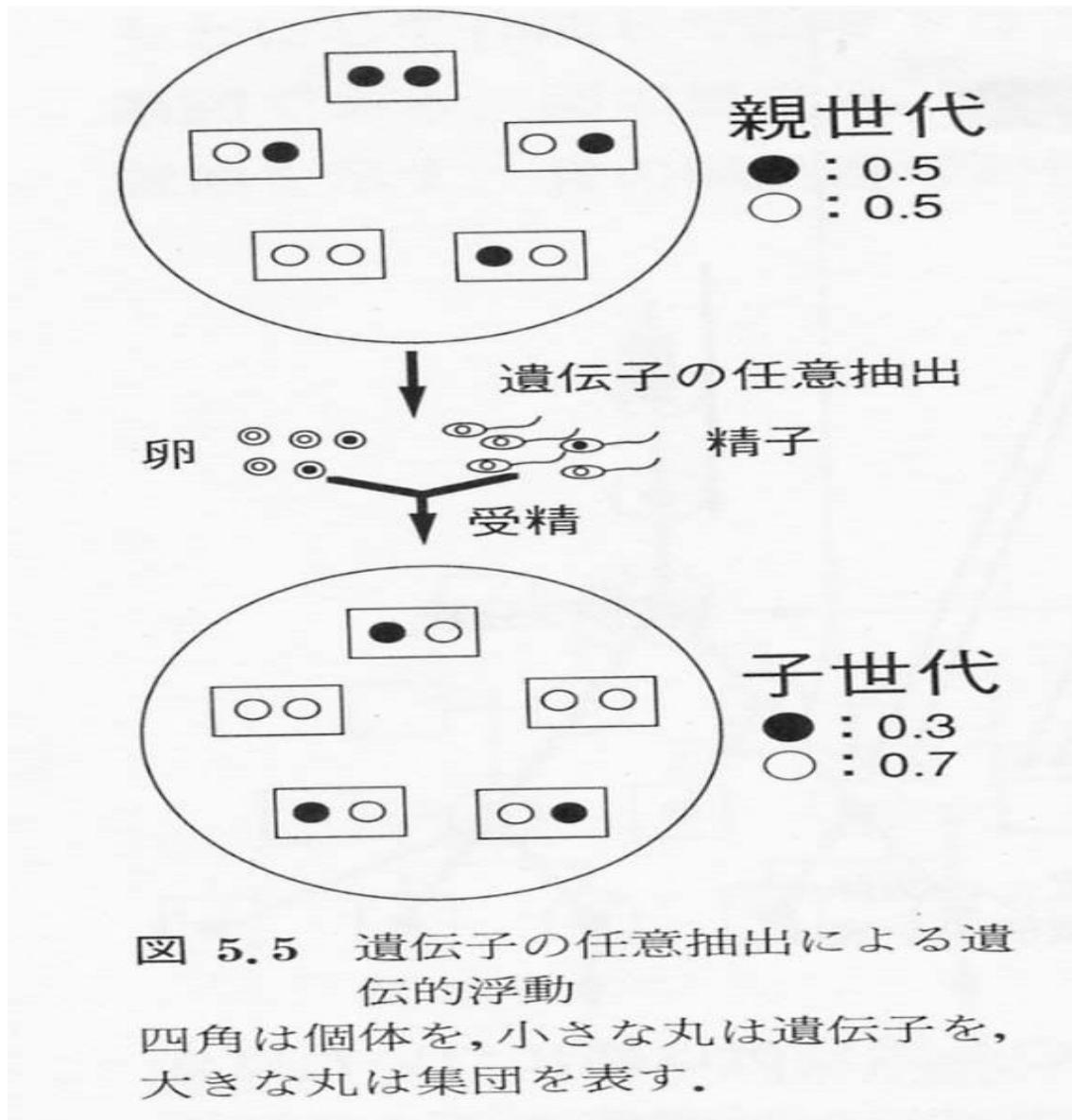


遺伝的浮動 genetic drift



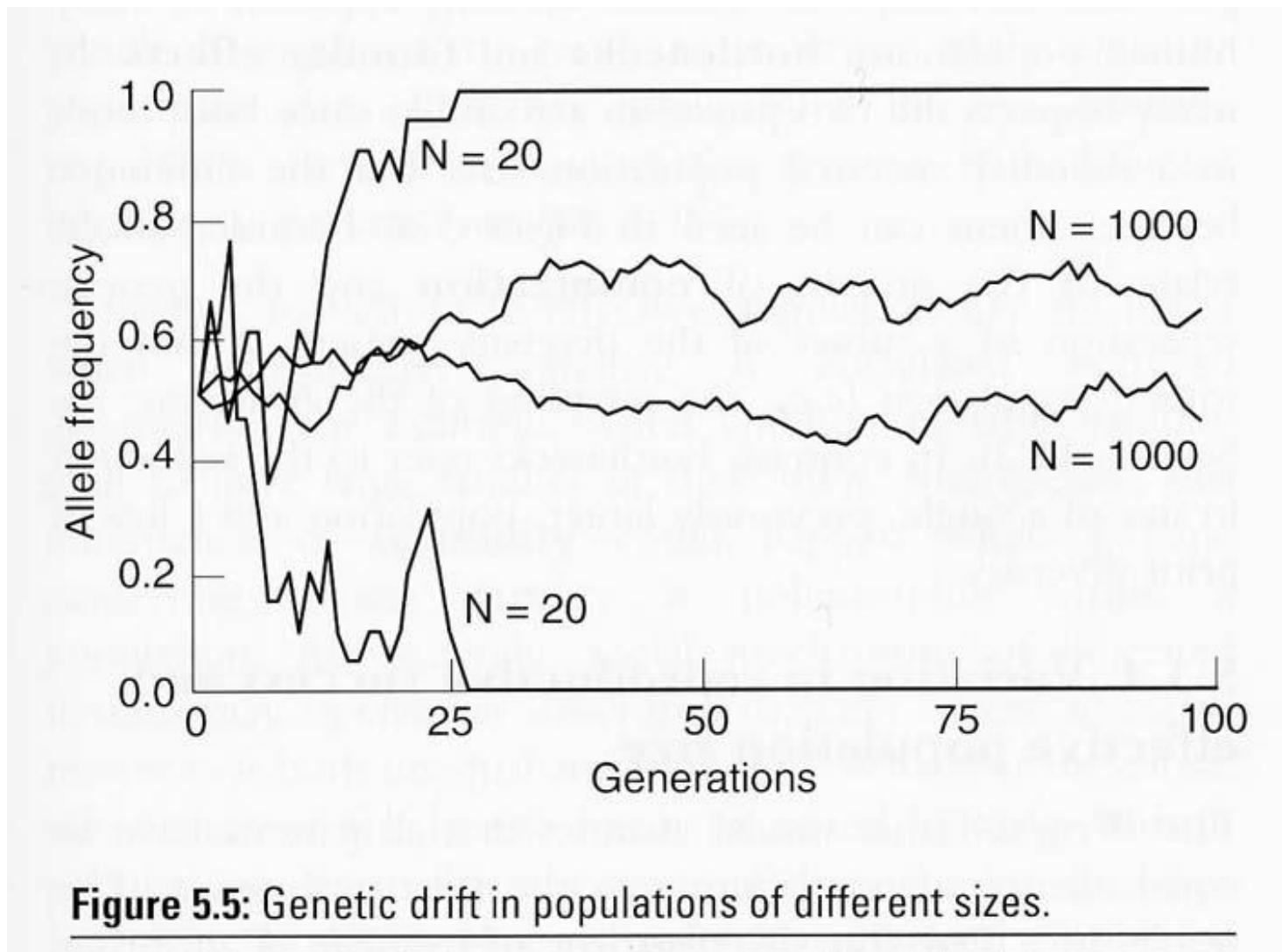
機能的に中立であっても、対立遺伝子（アリル）頻度は確率論的に変動し得る

遺伝的浮動 genetic drift



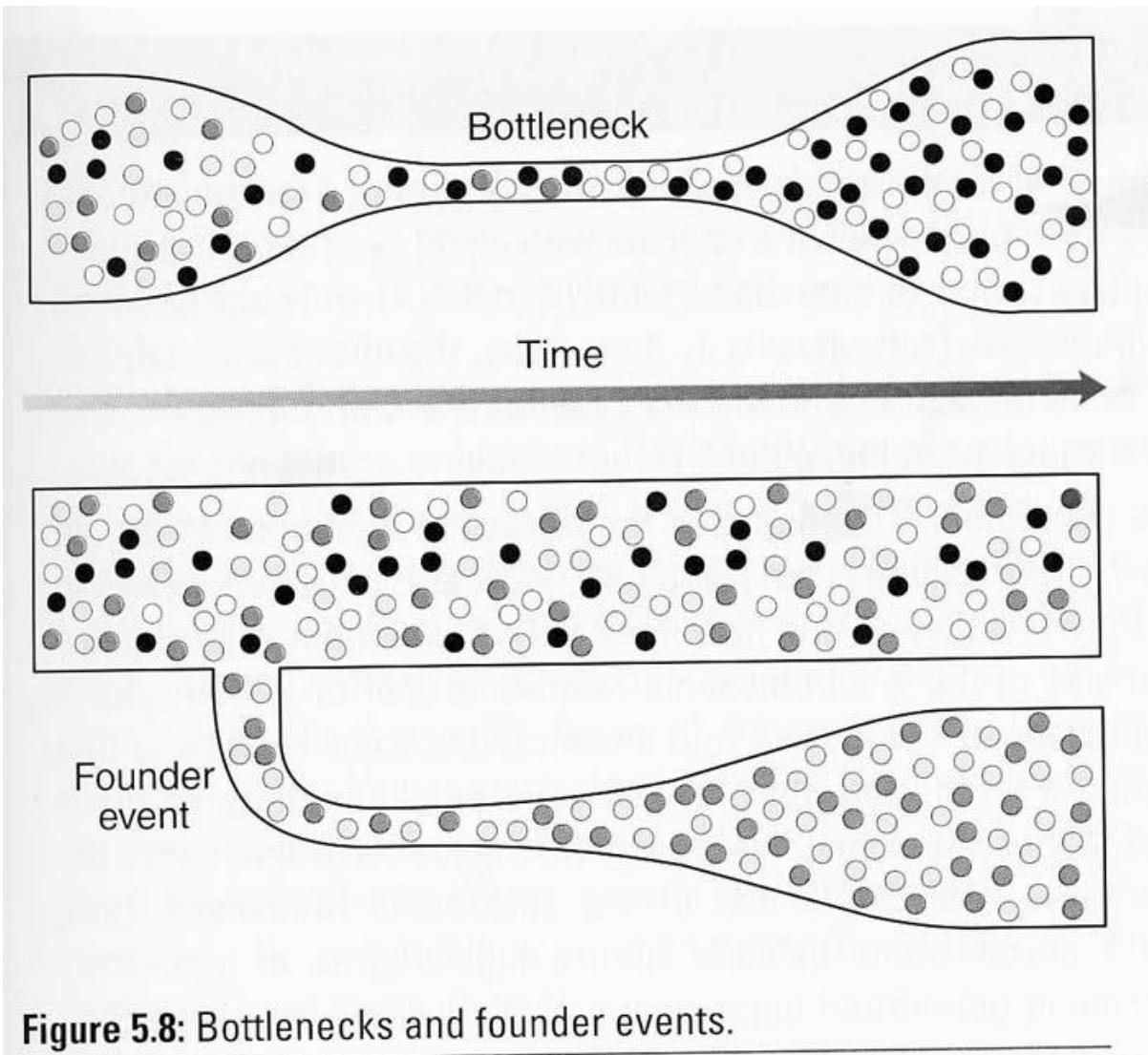
機能的に中立であっても、対立遺伝子（アリル）頻度は確率論的に変動し得る

集団サイズと遺伝的浮動



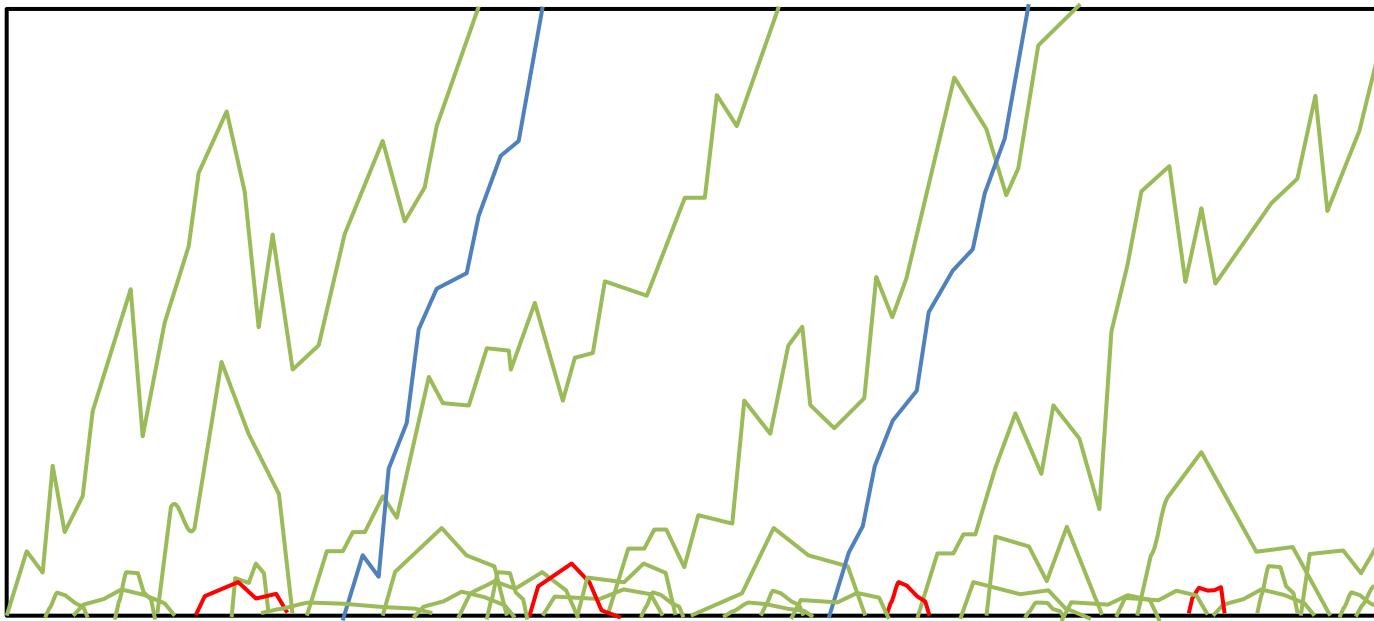
集団サイズが小さいほど、遺伝的浮動の効果が大きい

ビン首効果と始祖効果

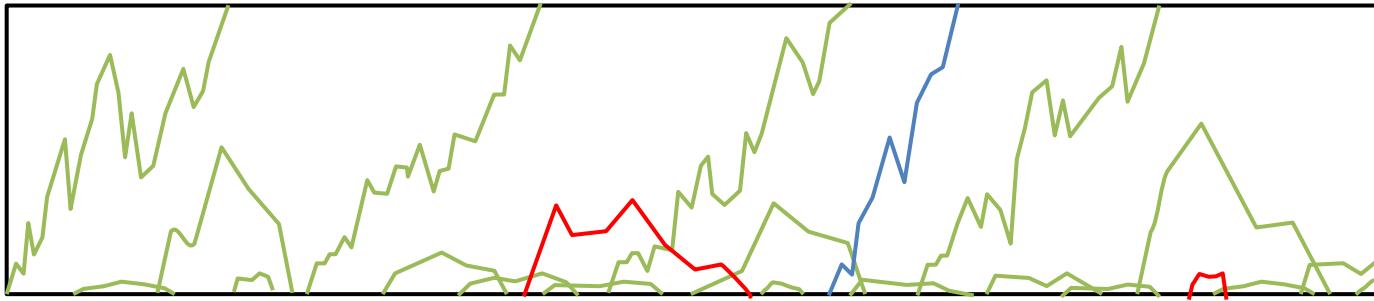


集団サイズおよび選択圧による変異の確率論的振る舞い

集団サイズ：大



集団サイズ：小



突然変異の数：



中立変異：



有利な変異：



弱有害変異：

集団サイズに比例 ($2N\mu$)、ただしそれぞれの初期頻度 ($1/2N$)

集団サイズに寄らず進化速度一定 ($v = \mu$: 固定確率 : $1/2N$)

集団サイズが大きいほど進化速度が速い

集団サイズが小さいと除去されにくい（中立的にふるまう）

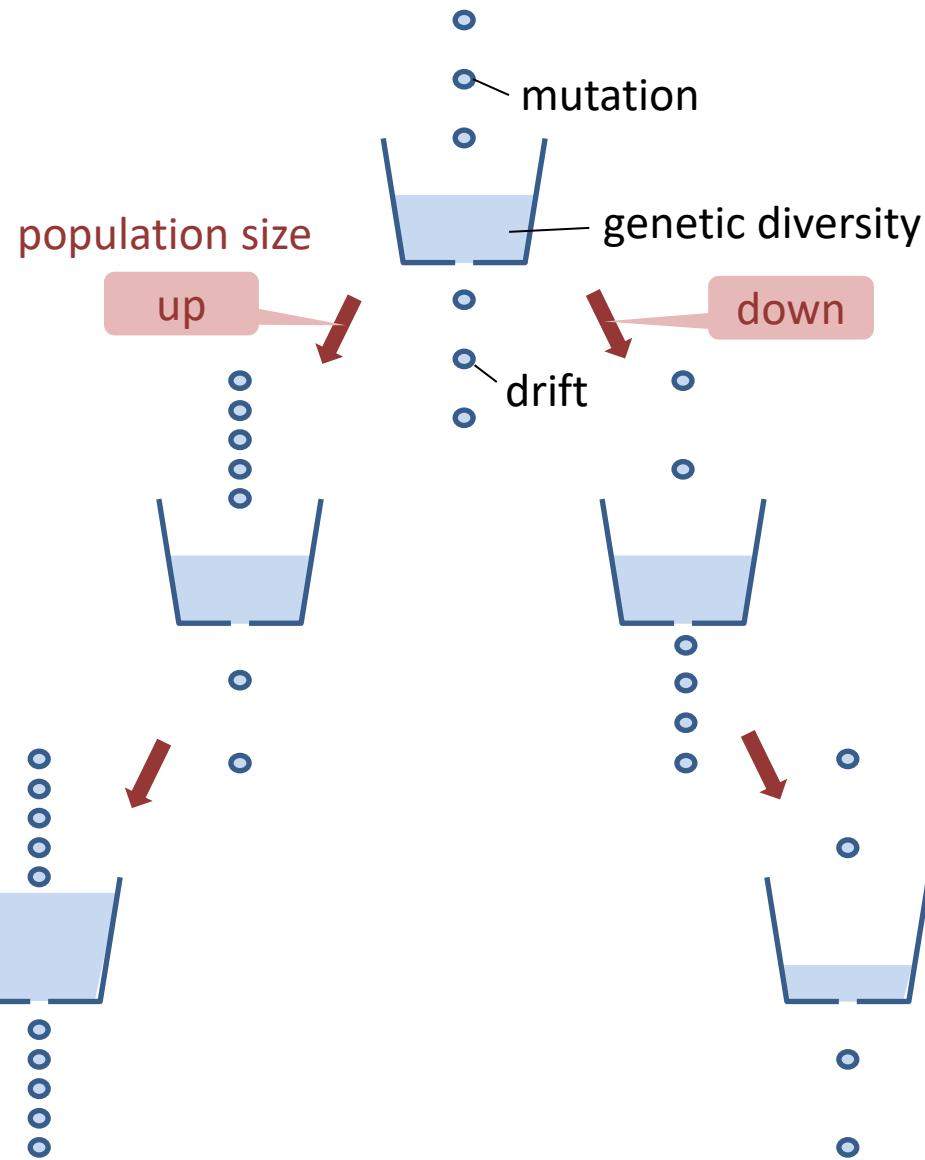
中立のもとでの集団の遺伝的多様性

mutation-drift 平衡

中立のもとでの集団内の遺伝的多様性は起こる突然変異の数と遺伝的浮動で失われる数のバランスで決まる。つまりは、集団サイズに依存する。

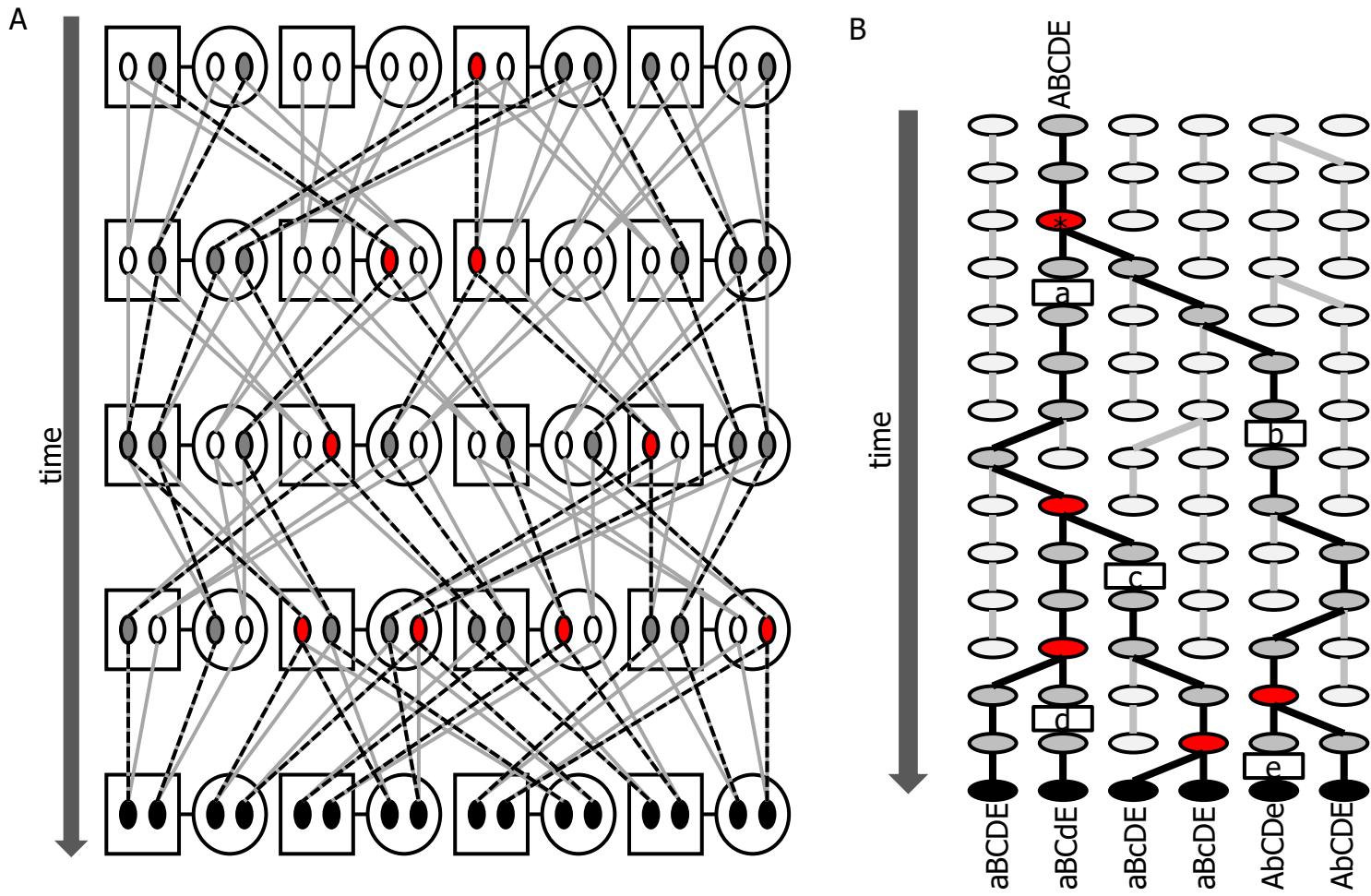
平衡状態のヘテロ接合度：

$$\tilde{H} \approx \frac{4N\mu}{1 + 4N\mu}$$



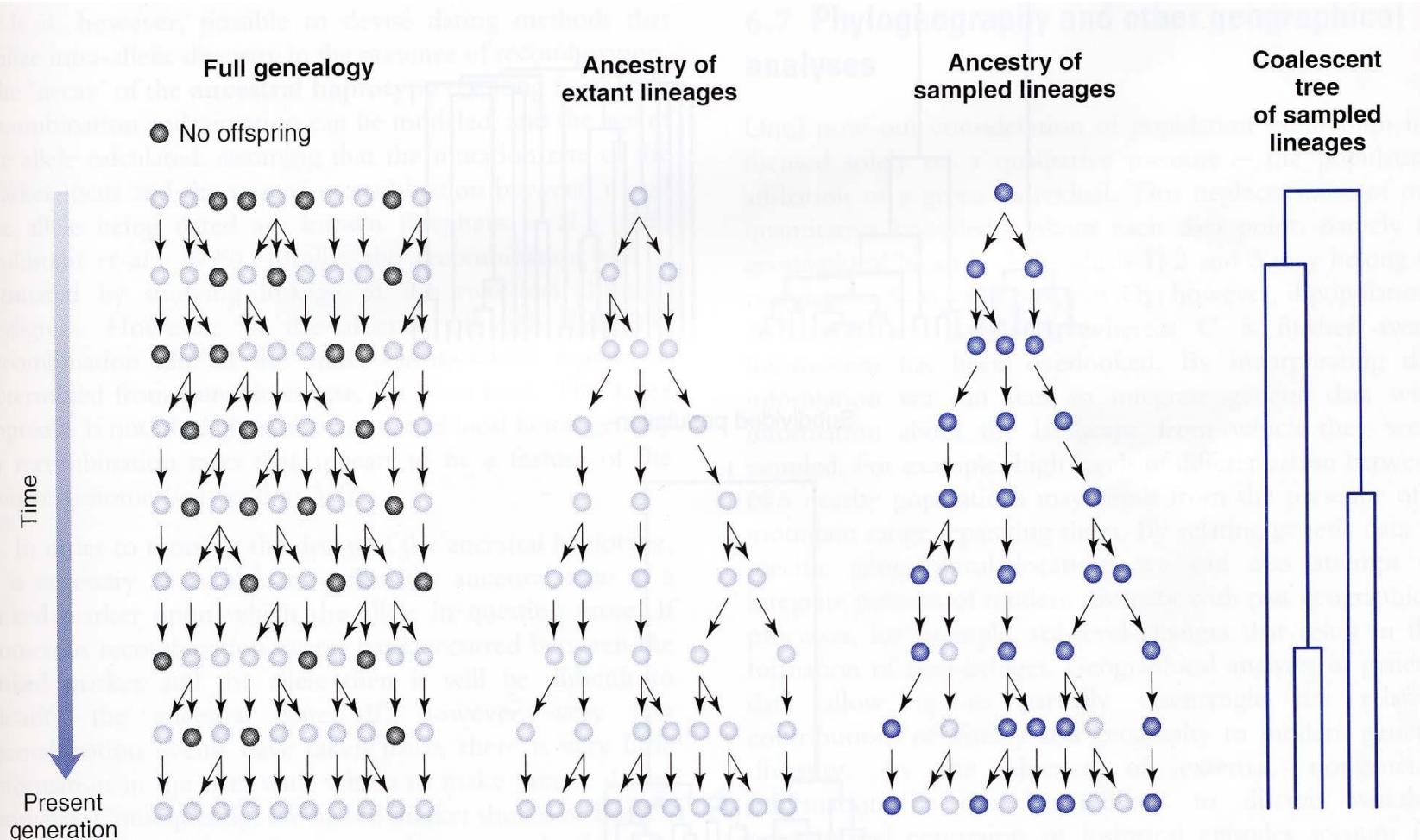
中立のもとでは、集団サイズが大きいほど遺伝的多様性も大きい

集団における遺伝子系図 Gene genealogy



A: 遺伝子の伝達(灰色:取りうる経路、点線:ある遺伝子座における実際の経路)
B: 集団中の全てのアリルは一つの共通祖先に遡ることができる

サンプルの遺伝子系図を観察



(a) Genealogical relations among nine generations of a population with constant size 10.

Considering only those individuals who have contributed to the present generation reveals a most recent common ancestor of all lineages. If only a subset of lineages are sampled (sampled lineages are shaded in dark blue) they share the same common ancestor as that shared by all extant lineages.

continued

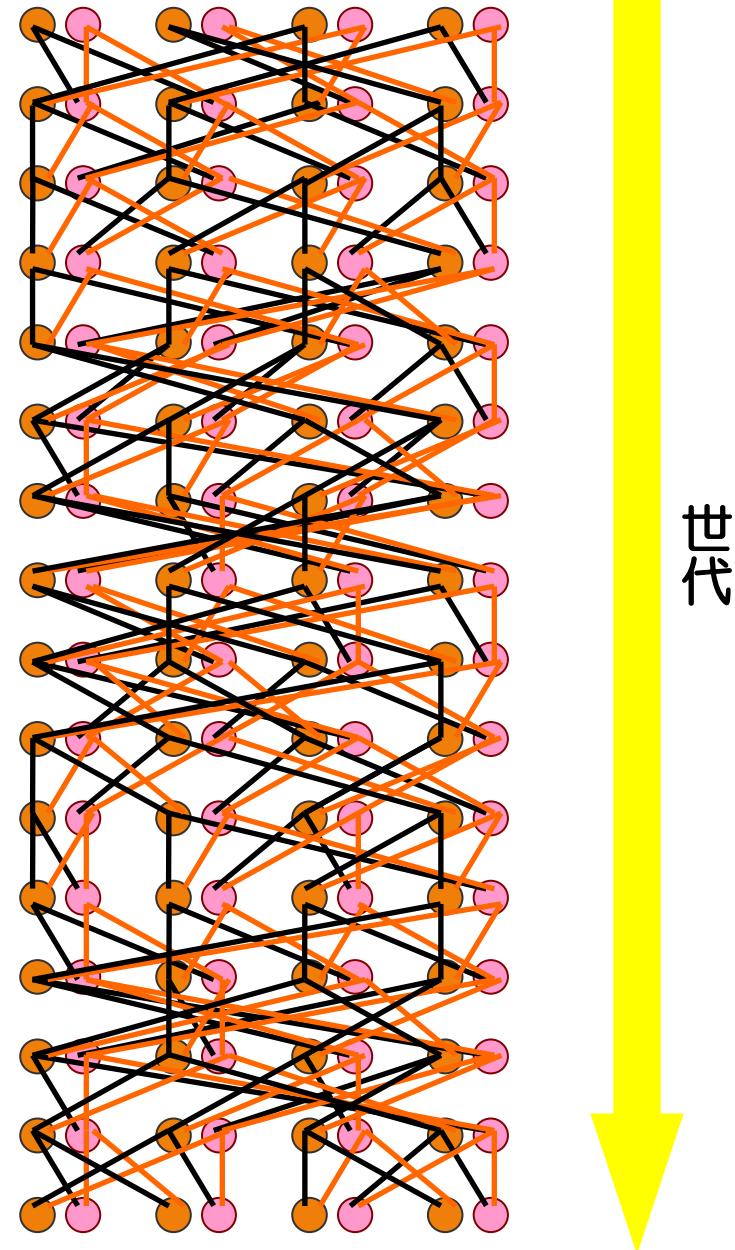
遺伝情報の伝達

各個体の祖先を遡ると。。。

10世代遡ると、重複が無ければ、その世代に祖先は 2^{10} 人。

祖先は複雑に絡み合い、個体間の共通祖先は無数に存在する。

- 男性
- 女性
- ─ 父親→子
- ─ 母親→子



ミトコンドリアDNA (mtDNA) の伝達

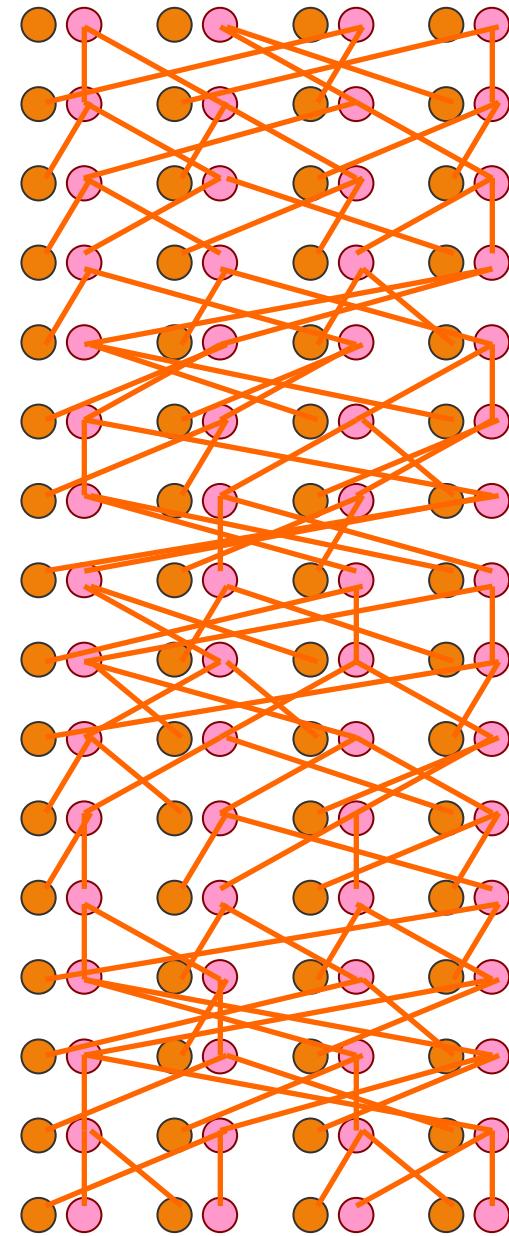
母親→子に伝達

● 男性

● 女性

— 父親→子

— 母親→子



世代

ミトコンドリアDNA (mtDNA) の伝達

母親→子に伝達

現代人の共通祖先★

ミトコンドリアイブであつてイブのミトコンドリアではない！



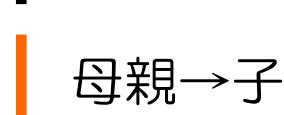
男性



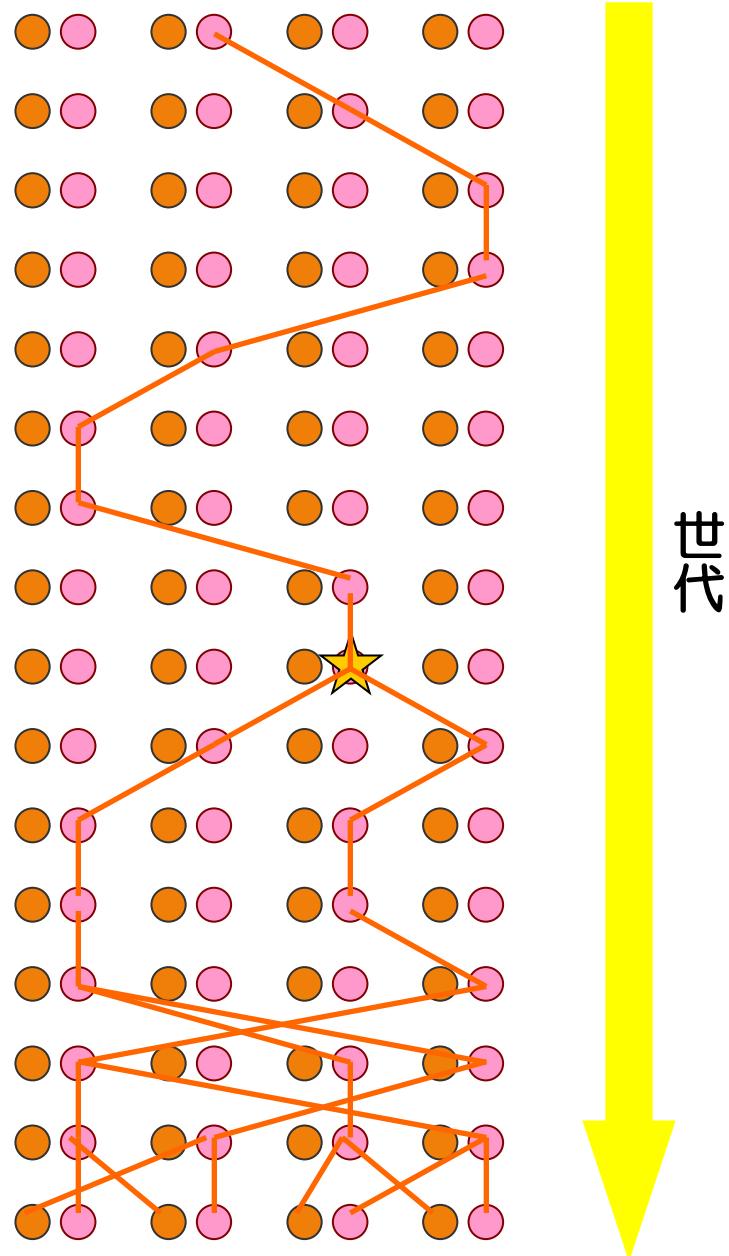
女性



父親→子



母親→子



Y chromosomeの伝達

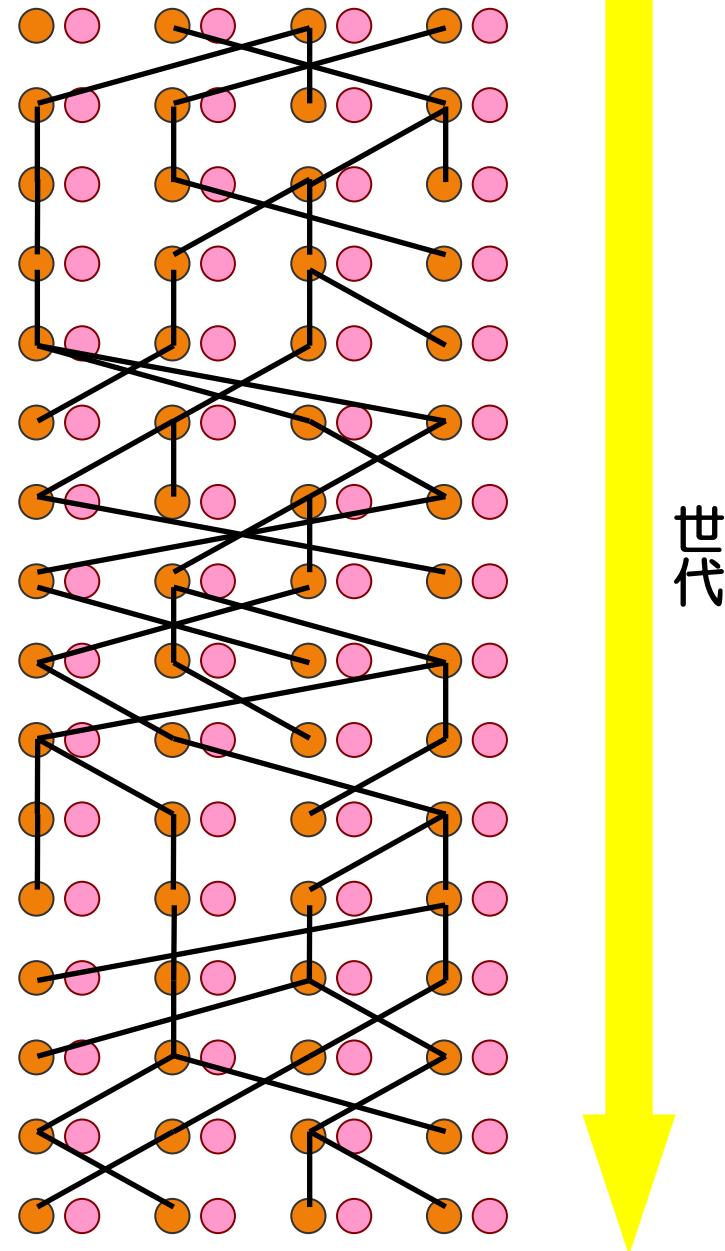
父親→息子に伝達

● 男性

● 女性

— 父親→子

— 母親→子



Y chromosomeの伝達

父親→息子に伝達

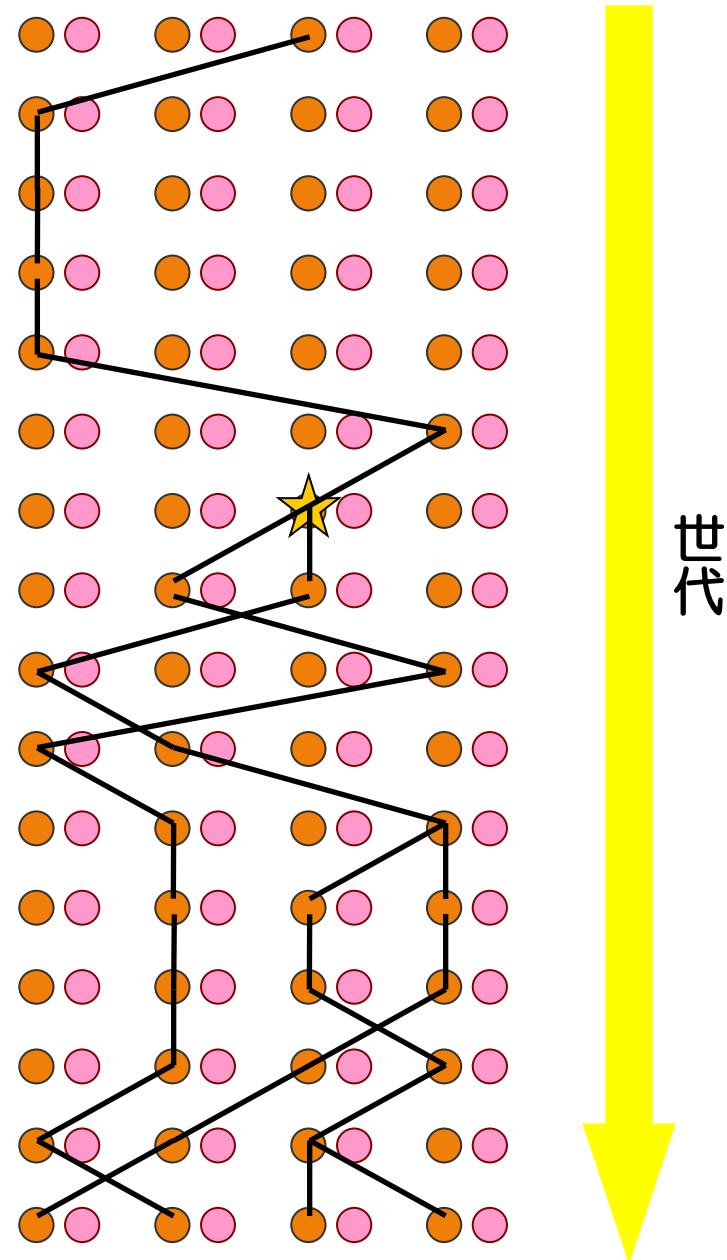
現代人のMRCA ★

● 男性

● 女性

— 父親→子

— 母親→子



常染色体の伝達

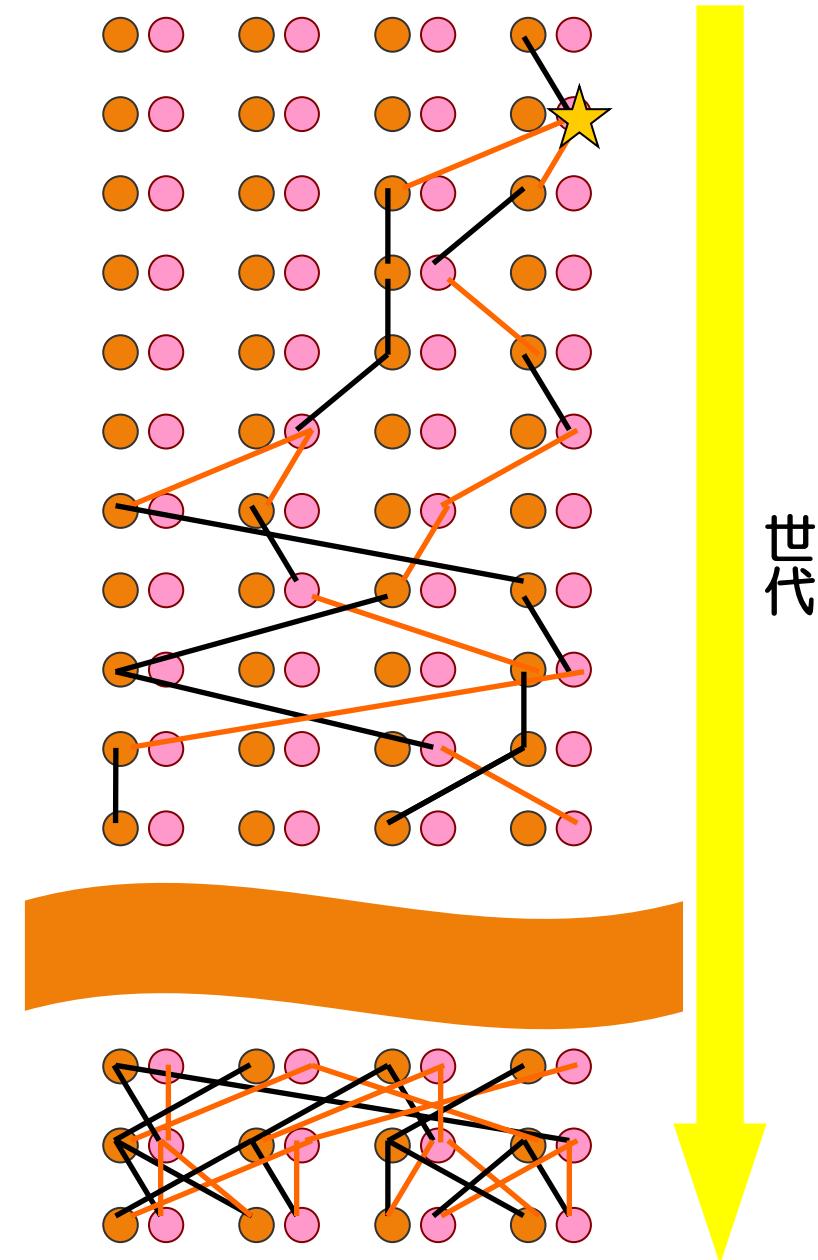
現代人の共通祖先★

たとえ同じ染色体でも、組換えによって領域毎に異なる系図が描け、異なる共通祖先に行き着く



■ 父親→子

■ 母親→子



常染色体の伝達

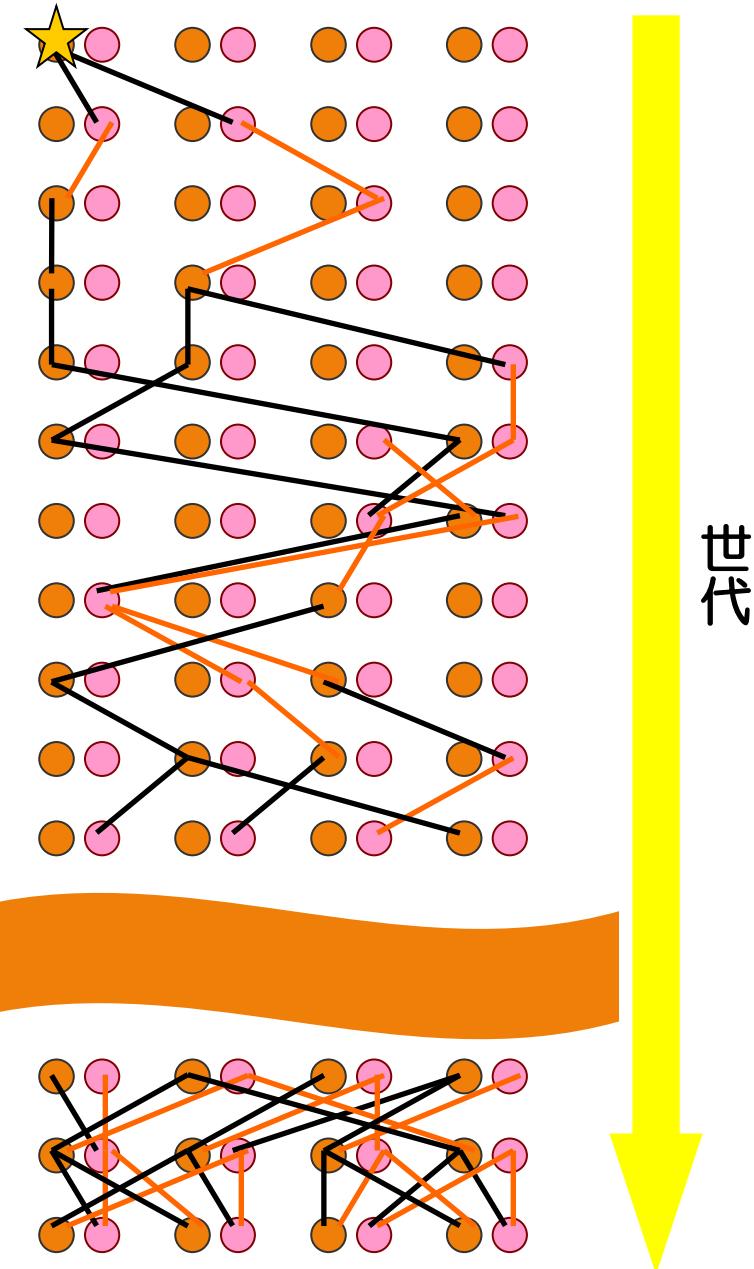
現代人の共通祖先★

たとえ同じ染色体でも、組換えによって領域毎に異なる系図が描け、異なる共通祖先に行き着く

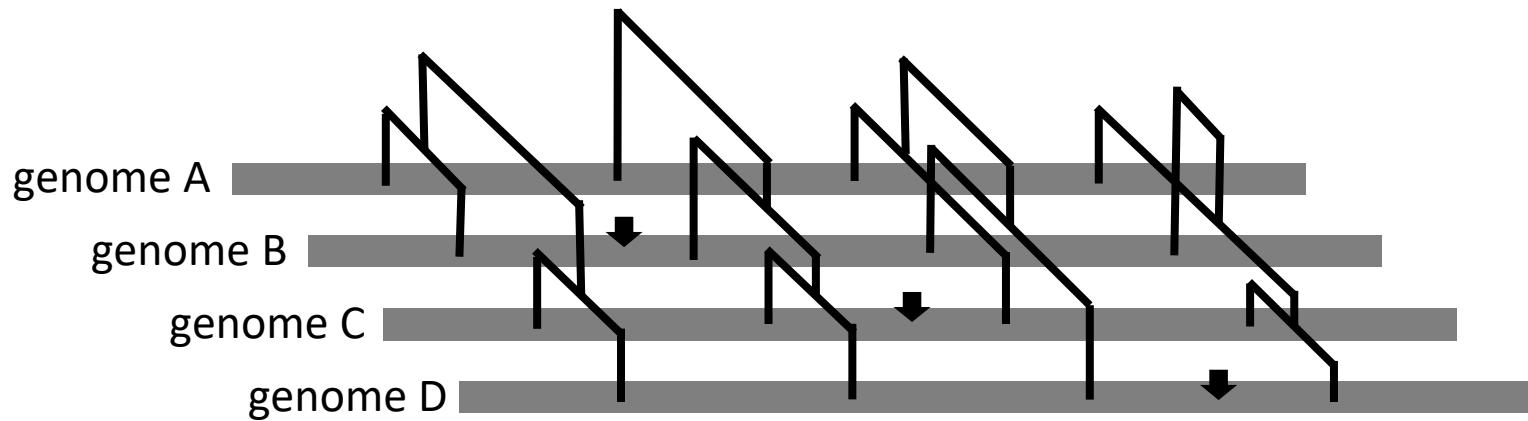


■ 父親→子

■ 母親→子

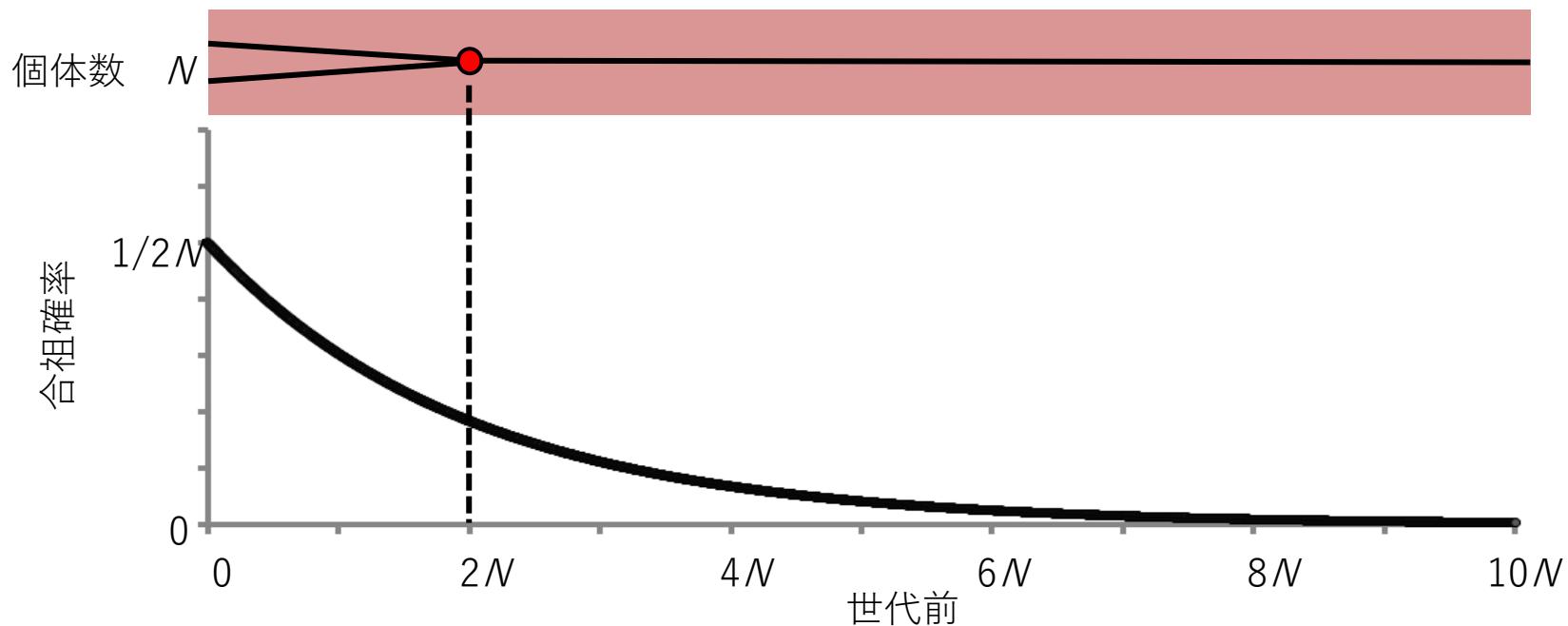


ゲノムと遺伝子系図

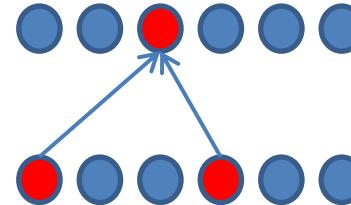


ゲノムは組み換え（矢印）を起こすため、
ゲノム領域ごとに遺伝子系図も変化する。

Coalescence time 合祖時間



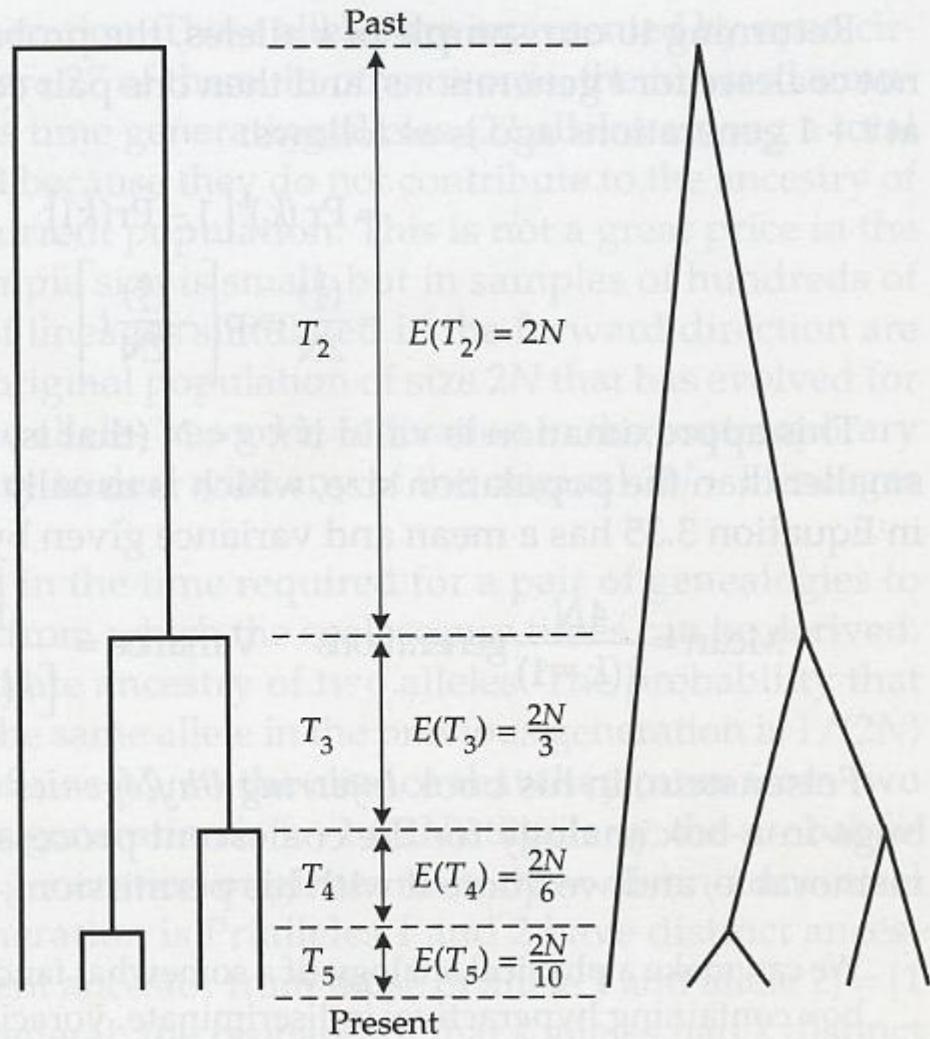
$$f_2(t) = (1/2N)(1 - 1/2N)^{t-1}$$
$$\approx (1/2N)e^{-t/2N}$$



有効集団サイズ一定Nの集団中の合祖時間の平均値
任意の2つの配列: 2N 世代

Coalescence theory

FIGURE 3.15 Two completely equivalent ways of illustrating the coalescences in a gene tree. On the left, the coalescent events are represented as horizontal lines, on the right they are represented as nodes. In any each generation, if there are k alleles present, the expected time back to the next coalescence is given by $4N/[k(1 - k)]$. For example, starting with five alleles, the expected time back to the first coalescence is $4N/[(5)(4)] = 2N/10$. Note that the successive times get longer. When there are only two alleles, the time back to the final coalescence is $2N$ generations.



有効集団サイズ一定 N の集団中の合祖時間の平均値
抽出した n 個の配列: $4N(1-1/n)$ 世代

近交係数と親縁係数

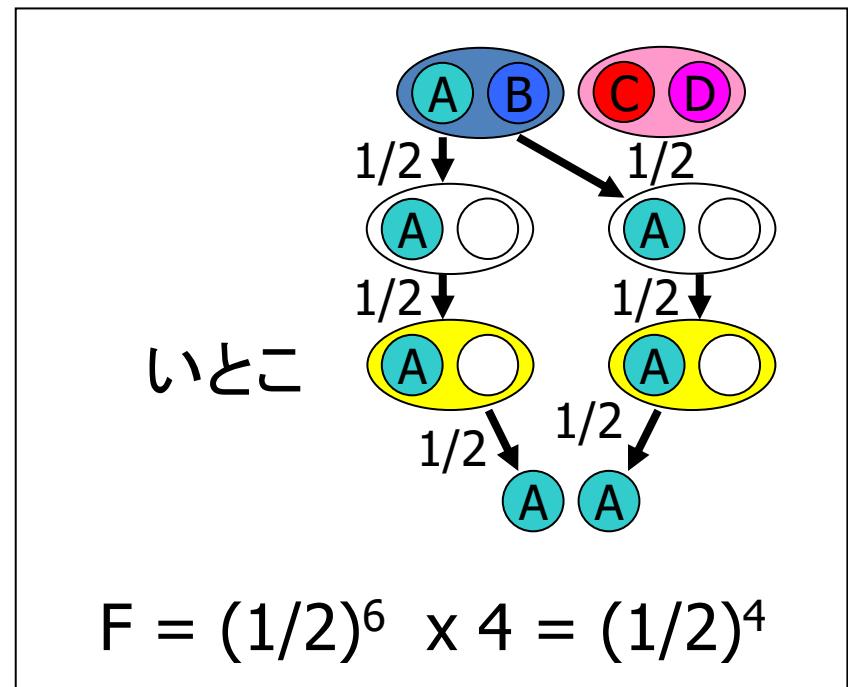
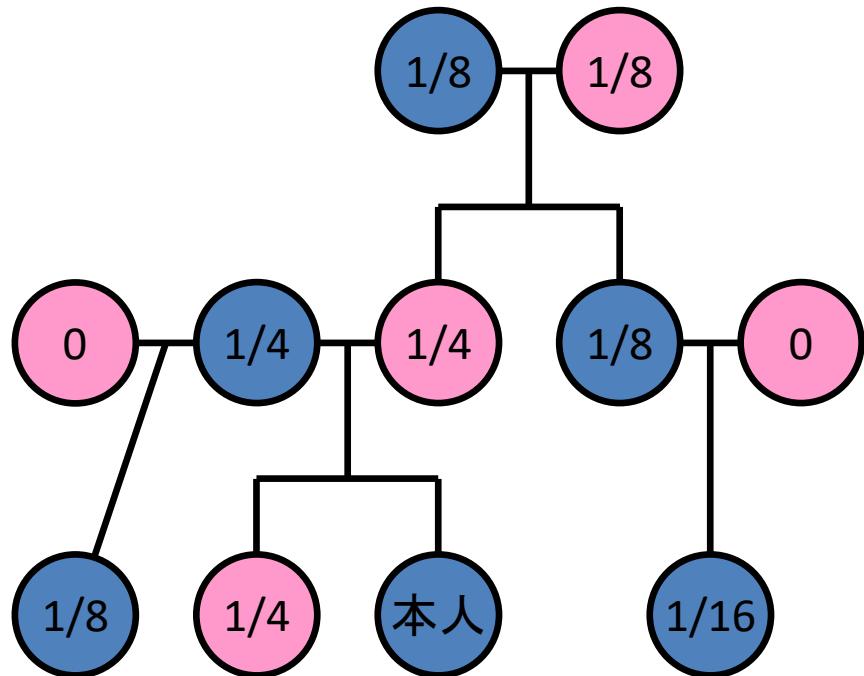
近交係数: inbreeding coefficient

「個体内の相同染色体が同祖的である確率」

親縁係数: coefficient of kinship

「2個体から任意にひとつずつ取り出した相同染色体が同祖的である確率」

2個体の子供の近交係数



いとこの親縁係数
=いとこ婚の子の近交係数

Hardy-Weinberg平衡

任意交配集団では、ふたつのアリルAとaをもつ遺伝子座における遺伝子型AA, Aa, aaの頻度はそれぞれ

$$P_{AA} = p^2$$

$$P_{Aa} = 2pq$$

$$P_{aa} = q^2 \quad \leftarrow aが劣性遺伝疾患アリルであれば疾患の頻度$$

Hardy-Weinberg平衡からのずれは、

- 近親交配
 - 同類交配
 - 分集団化
 - 自然選択
- } 末モ接合を増やす効果

によって生じ得る。

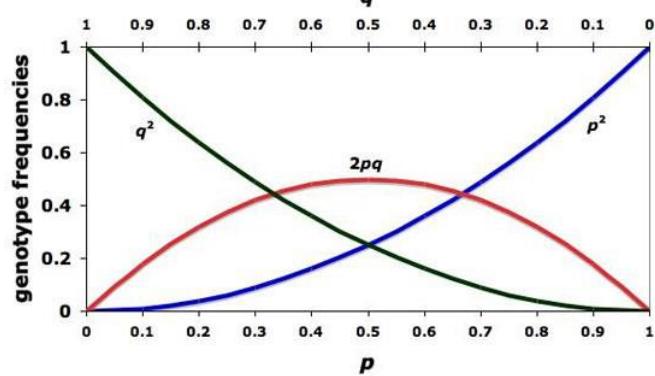
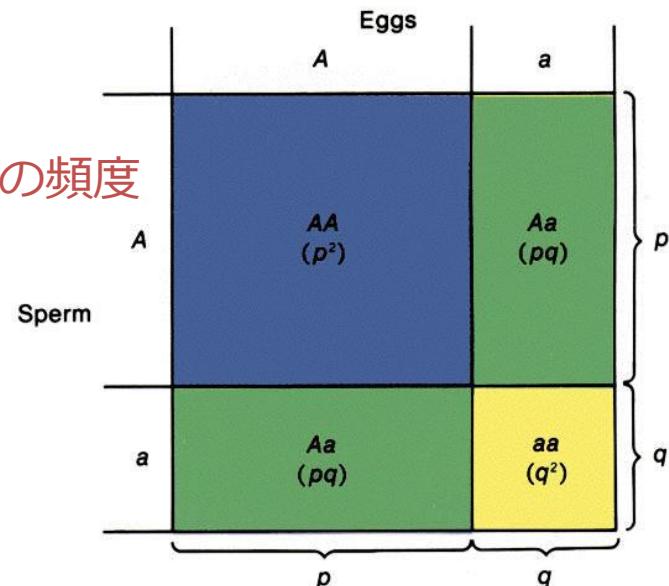
Hardy-Weinberg平衡からのずれ

$\Rightarrow F$: ヘテロ接合が減る効果

$$P_{AA} = (1-F)p^2 + Fp = p^2 + Fpq$$

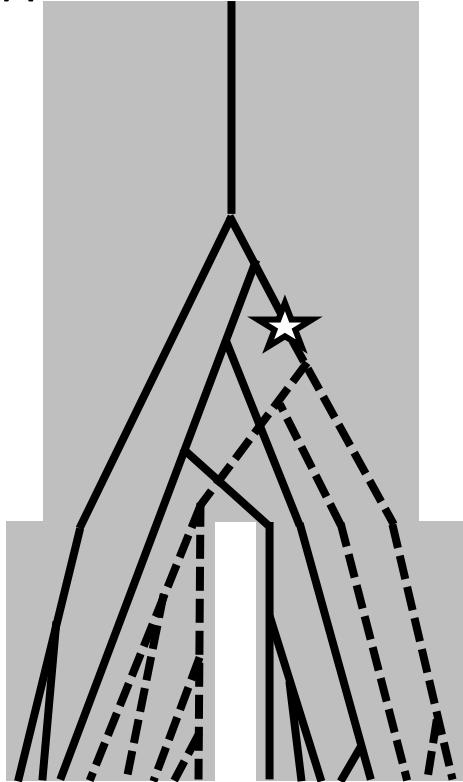
$$P_{Aa} = 2(1-F)pq = 2pq - 2Fpq$$

$$P_{aa} = (1-F)q^2 + Fq = q^2 + Fpq$$

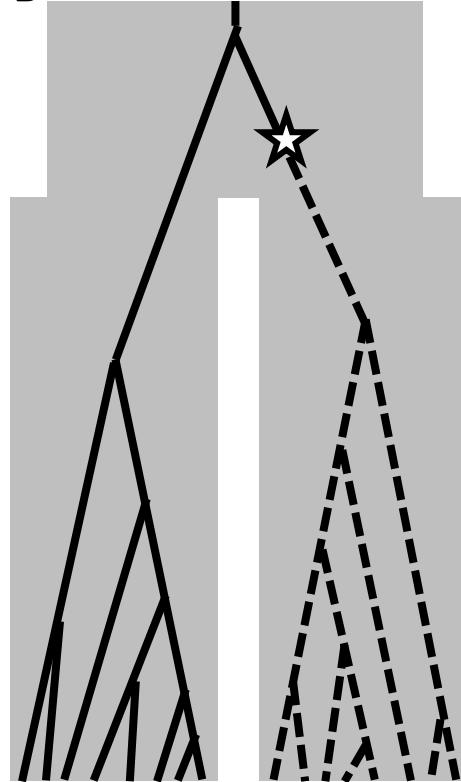


遺伝子の系図と集団の系統関係

A

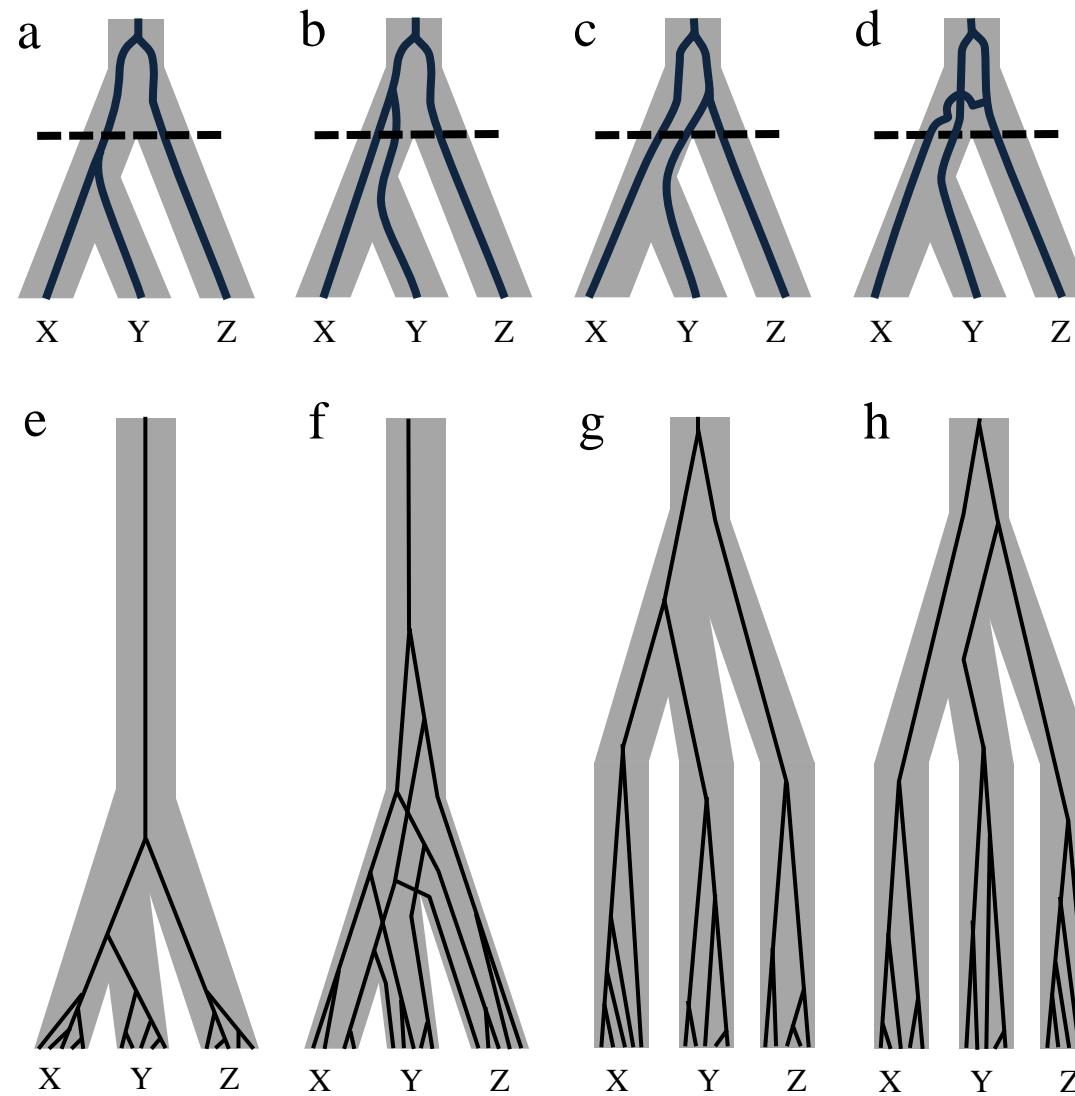


B



- (A) 集団分岐からさほど時間が経っていない場合、集団間で共通の多型をもつ。
- (B) 集団分岐から十分な時間が経てば、それぞれの集団内でのMRCAを集団の分岐後にもち、その結果、集団間で完全に異なるクレードを形成する。

Incomplete lineage sorting 不完全な系統仕分け



遺伝子系図と集団系図の樹形は必ずしも一致しない

ヒト・チンパンジー・ゴリラの系統問題

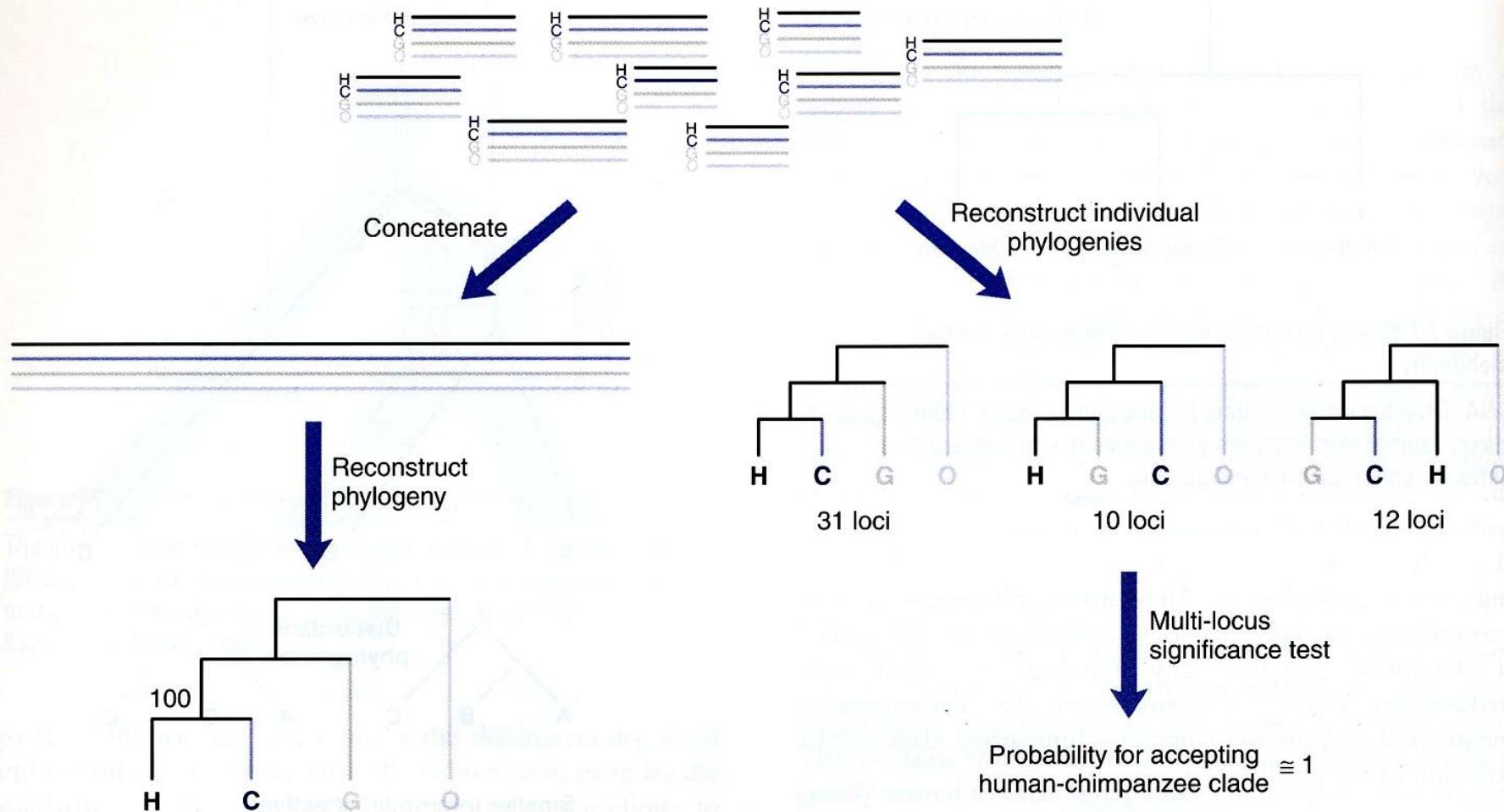
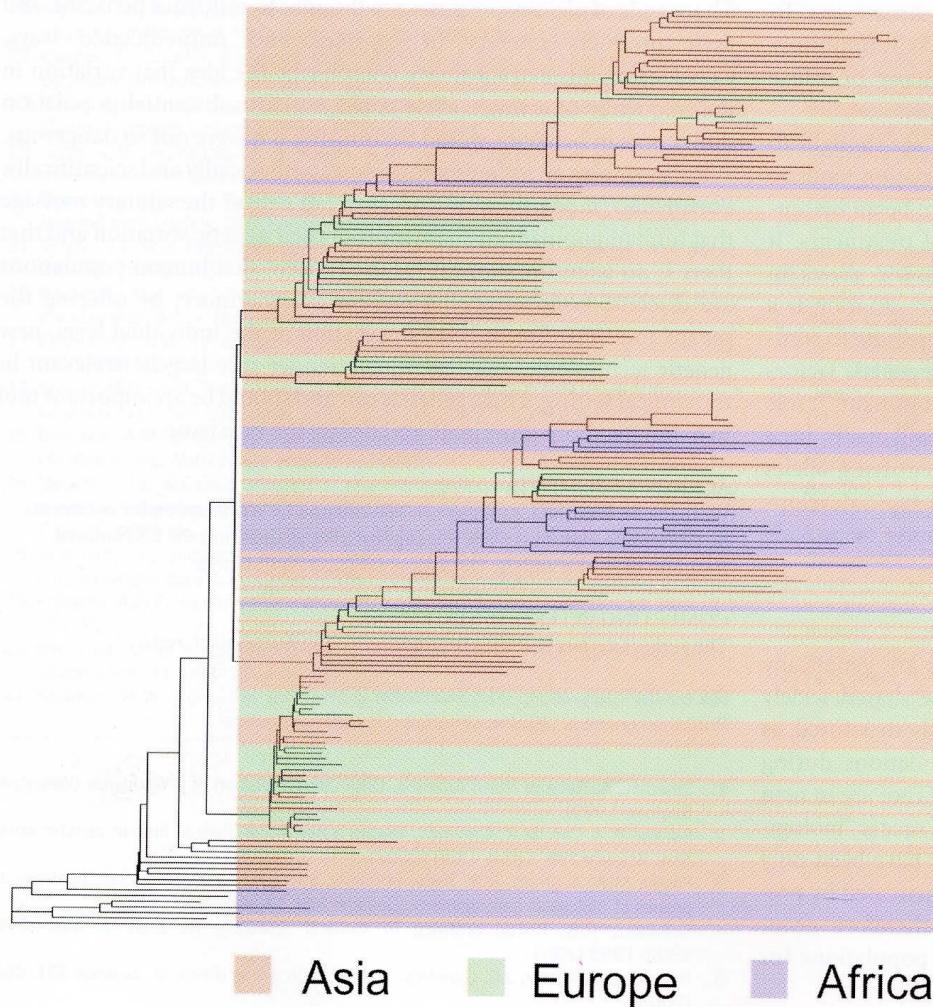


Figure 7.9: Combining data from multiple orthologous loci to resolve the gorilla–chimpanzee–human trichotomy.

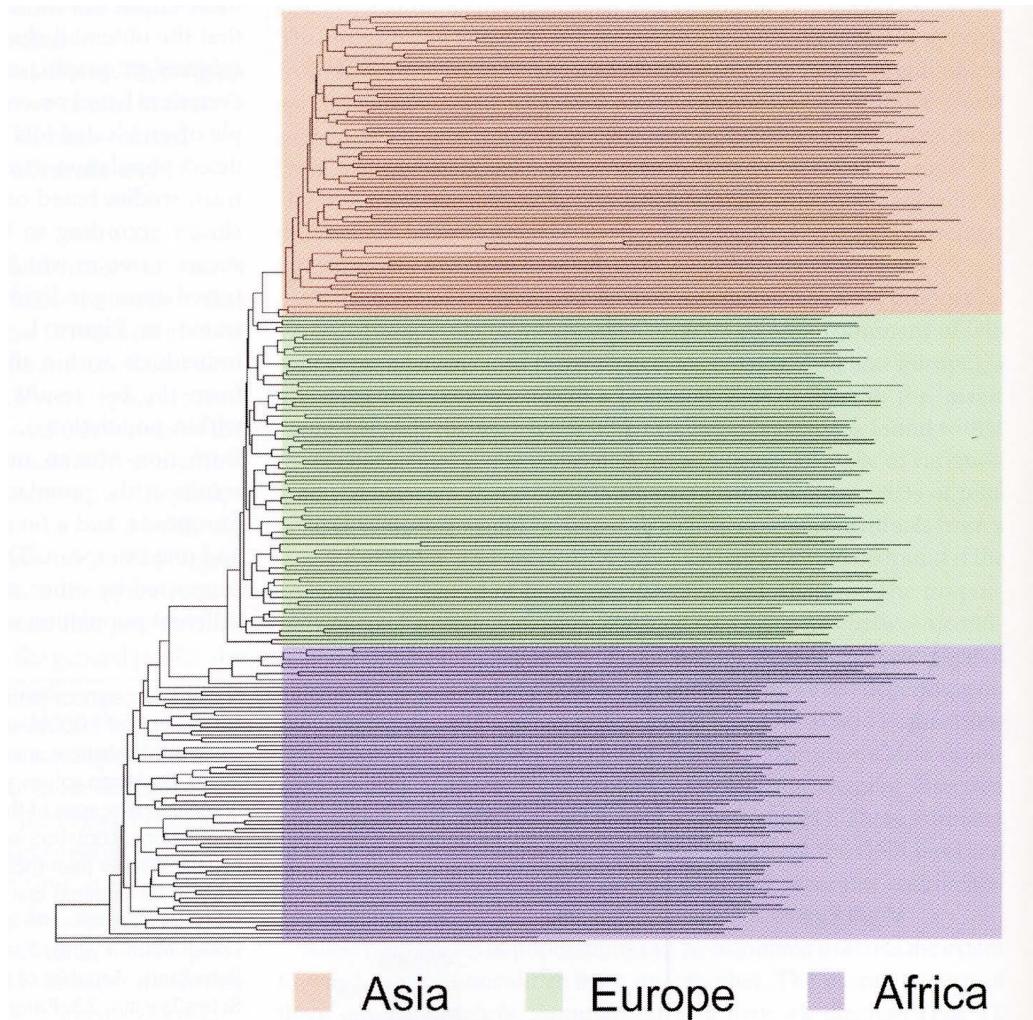
The number (100) next to the human–chimpanzee clade indicates percentage bootstrap support. H, Human; C, chimpanzee; G, gorilla; O, orangutan. Data are from Chen and Li (2001).

常染色体の1つの遺伝子の系統樹



集団ごとにクラスタリングしない！
各遺伝子は同一集団内で近いとは限らない

190遺伝子領域による個体間のクラスタリング



集団ごとにクラスタリング！
ゲノム全体の遺伝距離は同一集団内で近い

従来のヒト集団研究

集団間の系統関係を明らかにするため

- mtDNA
 - Y chromosome
- が多く解析されてきた。

利点

- 組換えが起きない→突然変異が蓄積
- 突然変異率が高い (mtDNA)
- 有効集団サイズが小さいため (常染色体の1/4) 、遺伝的浮動の効果が大きく、集団間に違いが生じやすい
- 女系・男系の祖先を遡れる

問題点

- 組換えが起きない
→それぞれ1つの領域に過ぎない
必ずしも集団の系統を反映しない

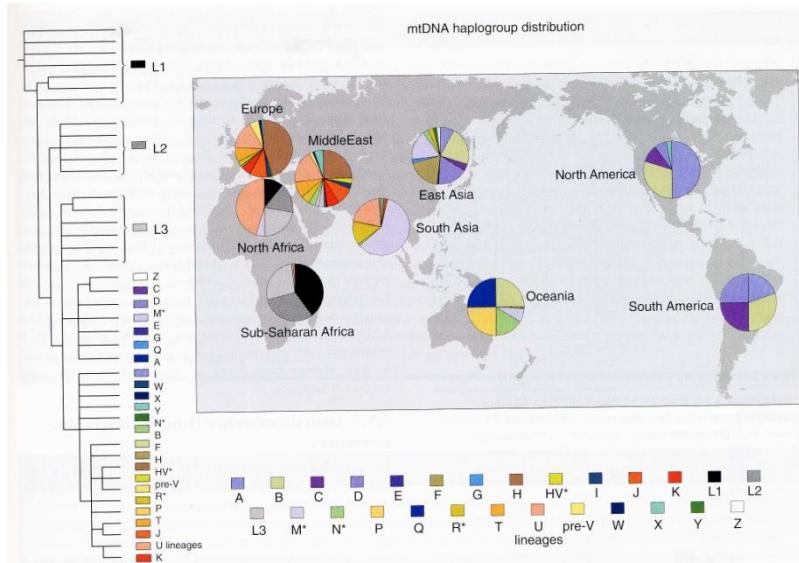


Figure 9.16: Geographical distribution of the major mtDNA clades.

Each major clade (haplogroup) is assigned a color reflecting its position in the phylogeny (left), and its frequency in population samples from broad geographical regions is shown in the pie charts. Note the most basal haplogroups (L) in Africa and the contrast between the frequencies of M (and subclades) compared to N (and subclades) in southern Asia compared with more northern parts of Asia and Europe. Color figure kindly sponsored by Oxford Ancestors Ltd.

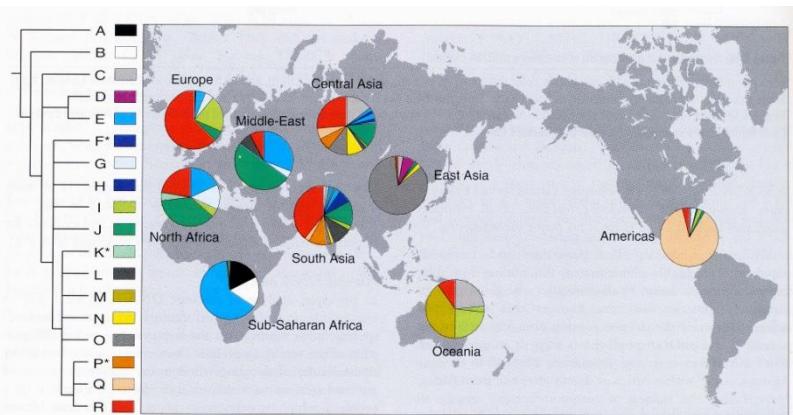
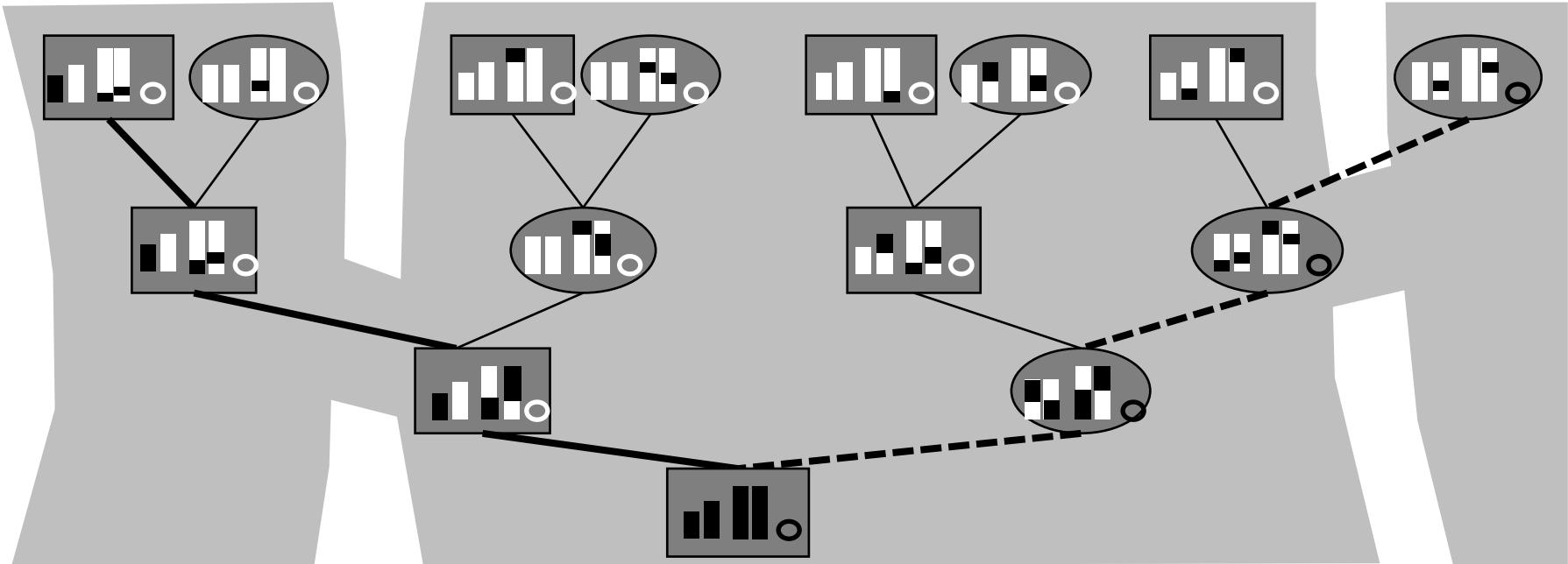


Figure 9.18: Geographical distribution of the major Y-chromosomal DNA clades.

Each major clade (haplogroup) is assigned a color reflecting its position in the phylogeny (left), and its frequency in population samples from broad geographical regions is shown in the pie charts. The basal haplogroups A and B are largely confined to Africa, and there is a broad distinction between East Asia and the rest of Asia and Europe. Data kindly supplied by Peter Underhill, and based on Underhill *et al.* (2001). Colour figure kindly sponsored by Oxford Ancestors Ltd.

一個体への遺伝子の伝達



- ・ひとりの現代人は多くの祖先から遺伝子を受け継いでいる：
 t 世代遡ると、祖先の重複が無ければその世代に祖先は 2^t 人
(伝達経路が 2^t 通り)
- ・父系 (Y染色体) の伝達経路 (太線) や母系 (ミトコンドリアDNA) の伝達経路 (点線) だけでなく、多くの経路を調べることが重要となる。

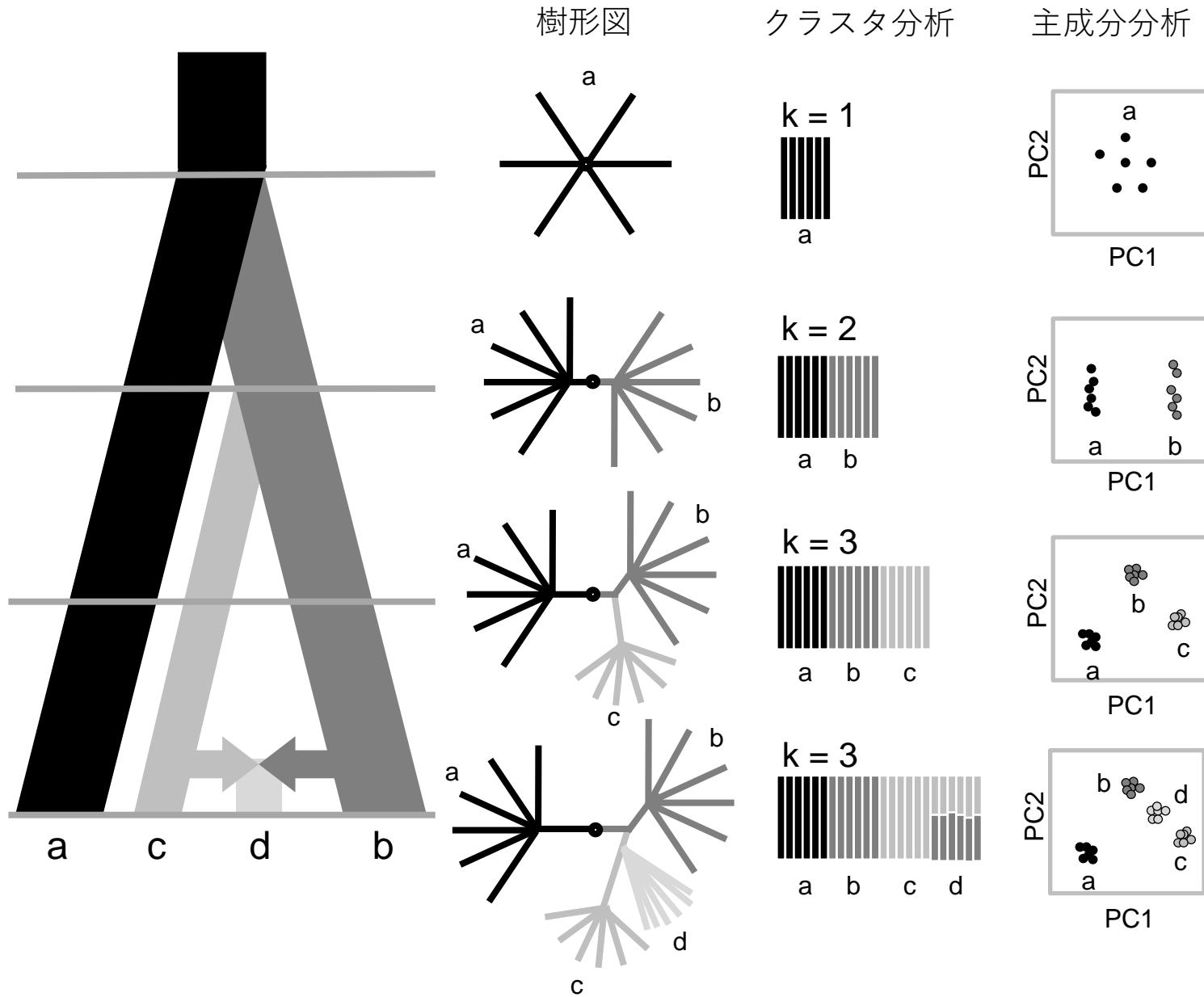
MENU

1. 集団遺伝学入門
2. ゲノム多様性を表現する
3. 琉球人集団の形成
4. ゲノム多様性と表現型多様性

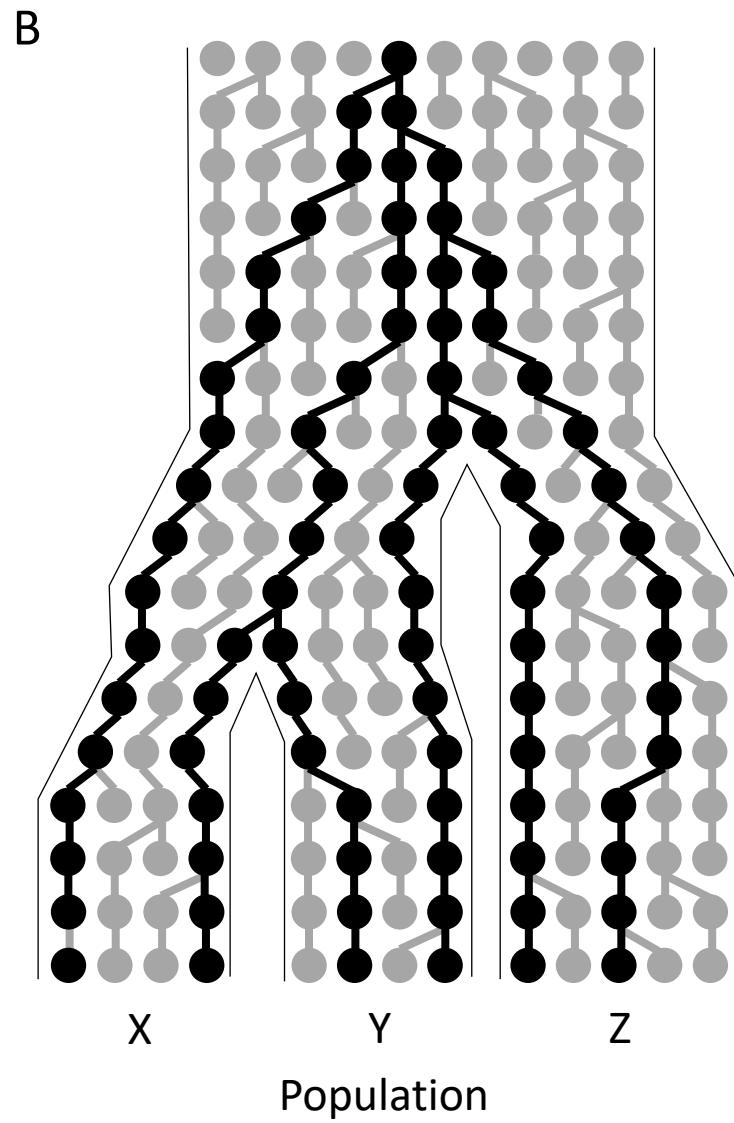
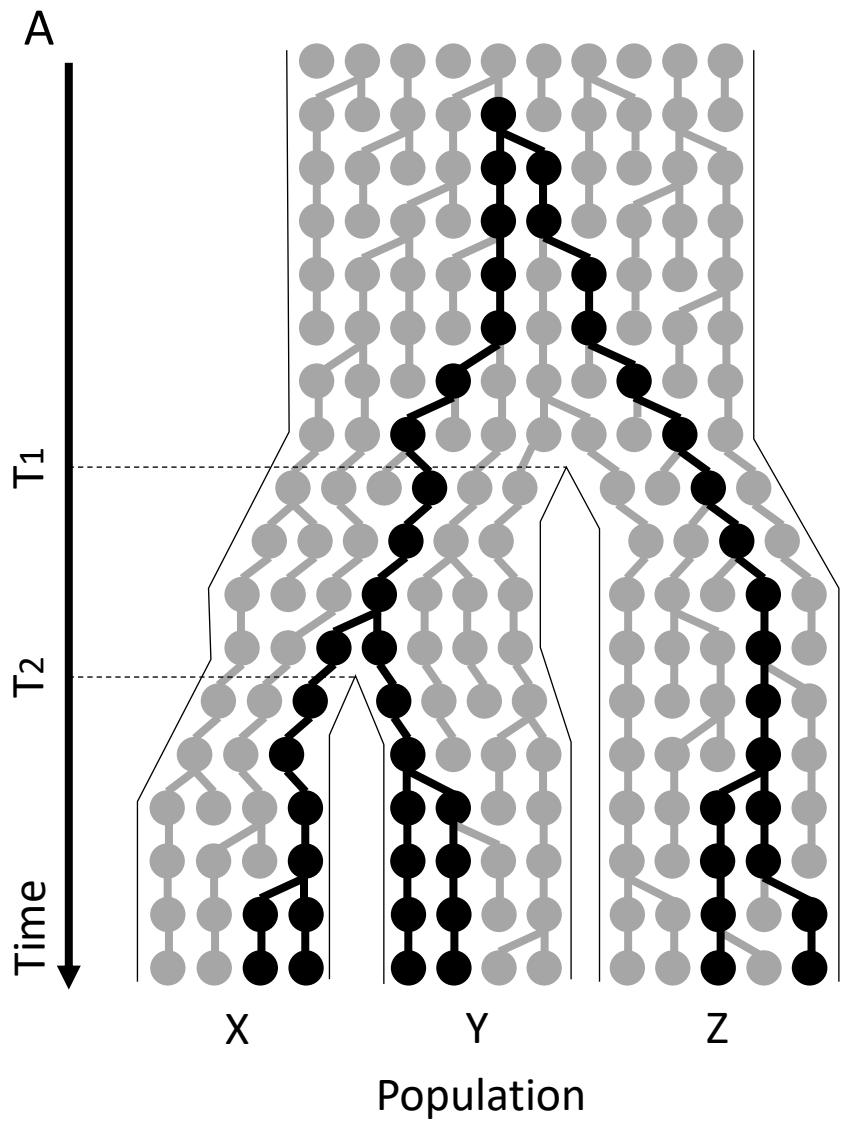
集団ゲノム解析で用いられる多変量解析

1. 遺伝距離マトリックス Genetic distance matrix
⇒系統樹解析 Phylogenetic analysis
⇒多次元尺度構成法 Multidimensional scaling (MDS)
2. 主成分分析 Principal component analysis (PCA)
3. クラスタ分析 Clustering analysis
(STRUCTURE, ADMIXTURE, FRAPPE)

集団動態と様々な解析の結果



1 遺伝子座の配列間の遺伝距離



個体間の遺伝距離

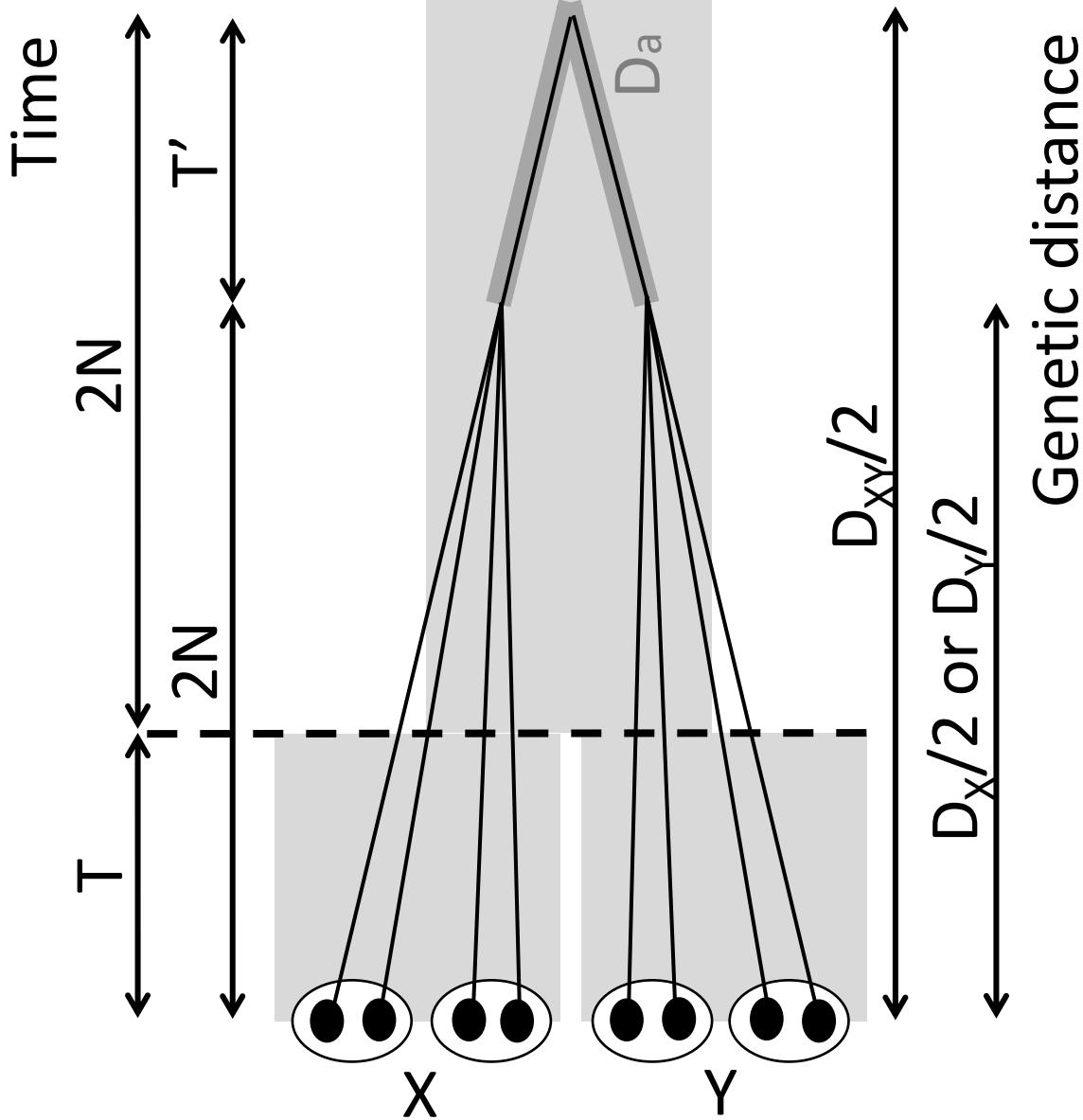
		個体iの遺伝子型		
		AA	AB	BB
個体jの 遺伝子型	AA	0	0.5	1
	AB	0.5	0.5	0.5
	BB	1	0.5	0

個体間のサイトあたりの遺伝距離 d_{ij}

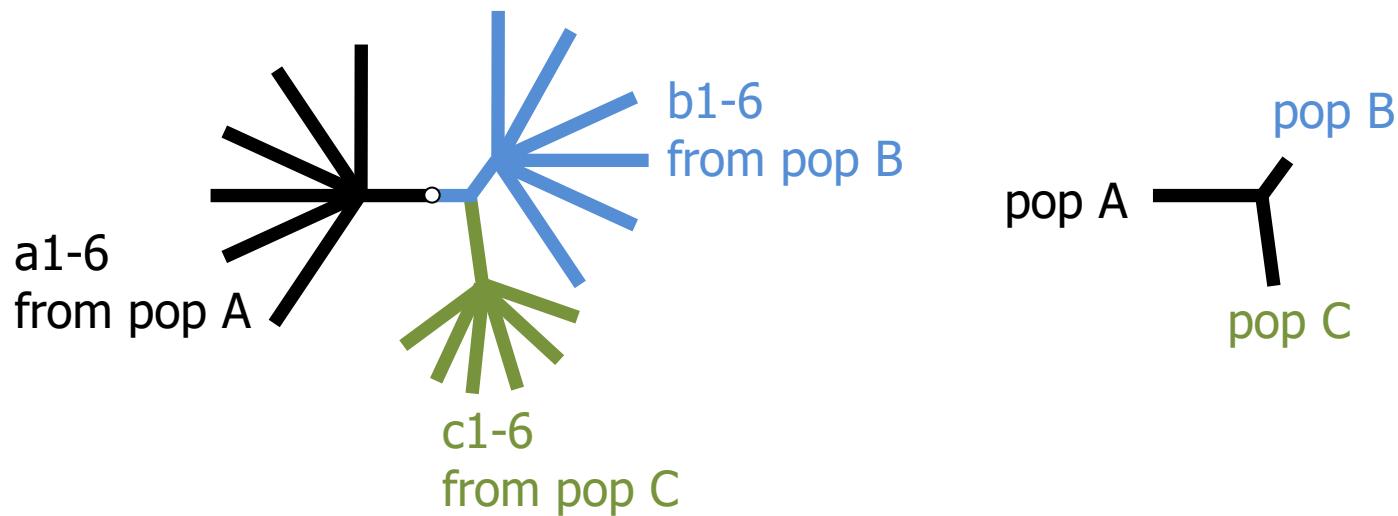
個体間のゲノム全体の遺伝距離 D_{ij}
$$D_{ij} = \frac{1}{L} \sum_l d_{ijl}$$

IBS距離 : $d(AB\text{vs}AB) = 0$

個体間の遺伝距離（集団内、集団間）



個体間樹形図と集団間樹形図の関係



Tree for individuals
from D_{ij} matrix
(i and j : individuals)

Tree for populations
from D_{mXY} matrix
(X and Y : populations)

◆Nei's minimum distance

$$D_{mXY} = D_{XY} - \frac{(D_X + D_Y)}{2} = \frac{1}{L} \sum_{i=1}^L (p_{Xi} - p_{Yi})^2$$

遺伝距離のゲノム平均などに基づいた集団動態に関するパラメータ推定を
moment-based estimationと呼ぶ

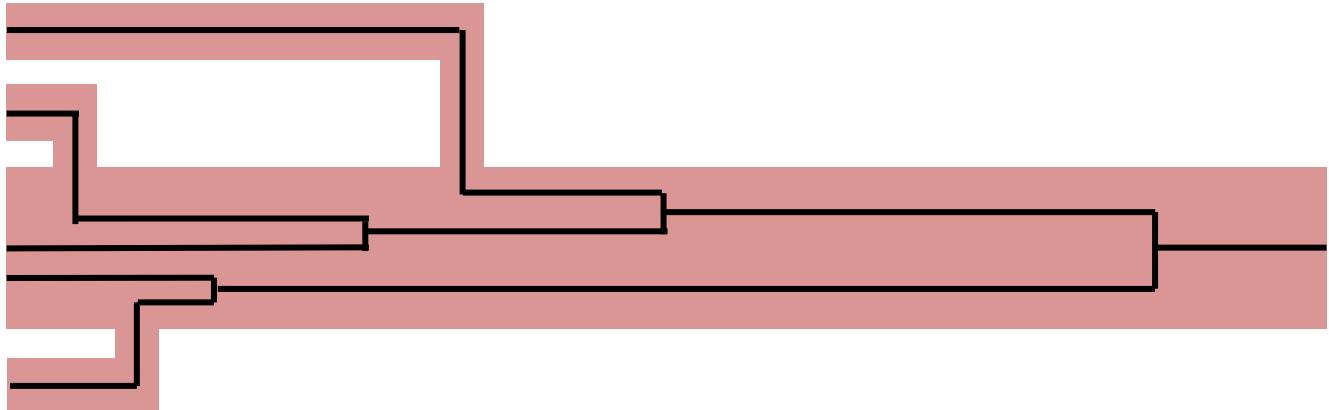
階層的クラスタリング

- 距離行列(総当たりの個体間距離)にもとづいた階層的クラスタリング
 - ・非加重結合法(Unweighted Pair Group Method with Arithmetic mean: **UPGMA**)
すべての系統が同じ速度で進化することを仮定
 - ・近隣結合法(neighbor-joining method)
すべての系統が同じ速度で進化することを仮定しない
- 階層的クラスタリングによる樹形図が、時間や分岐を表すとみなせるとき「系統樹」と呼ぶ
 - 1) ひとつの遺伝子座のデータで描いた樹形図 → 系統樹と呼べる
 - 2) 複数の遺伝子座のデータをまとめて描いた樹形図
 - 2-1) すべての遺伝子座が同じ系統関係(トポロジー) → 系統樹と呼べる
 - 2-2) 遺伝子座ごとに異なる系統関係(トポロジー) → 系統樹と呼べない
- 配列情報にもとづいた系統樹作成法(最尤法、ベイズ法)は、上記2-2のようなケースでは用いることはできない。→ 距離行列にもとづいた階層的クラスタリングを用いる

旧人と新人における遺伝子と集団の樹形

A ひとつの遺伝子座

ネアンデルタル



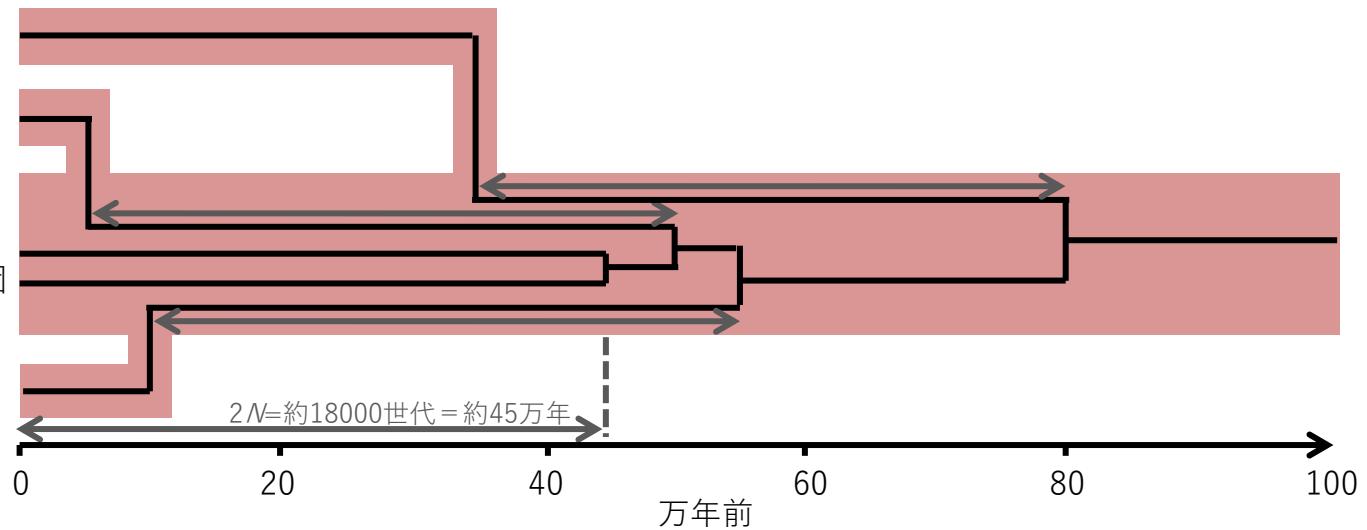
B ゲノム全体での平均

ネアンデルタル

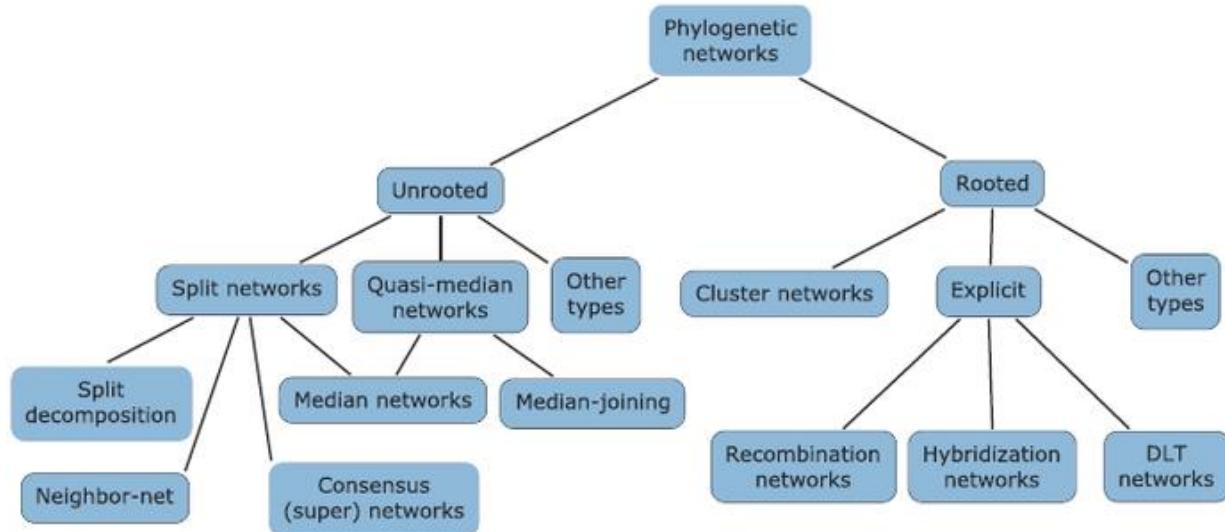
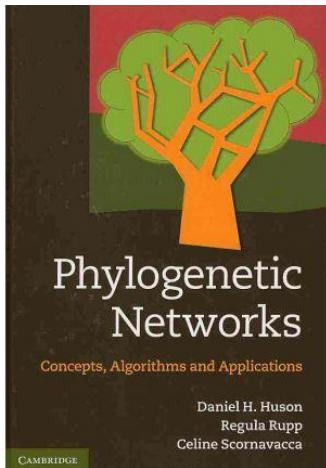
サピエンス
ユーラシア集団

中央・西アフリカ集団

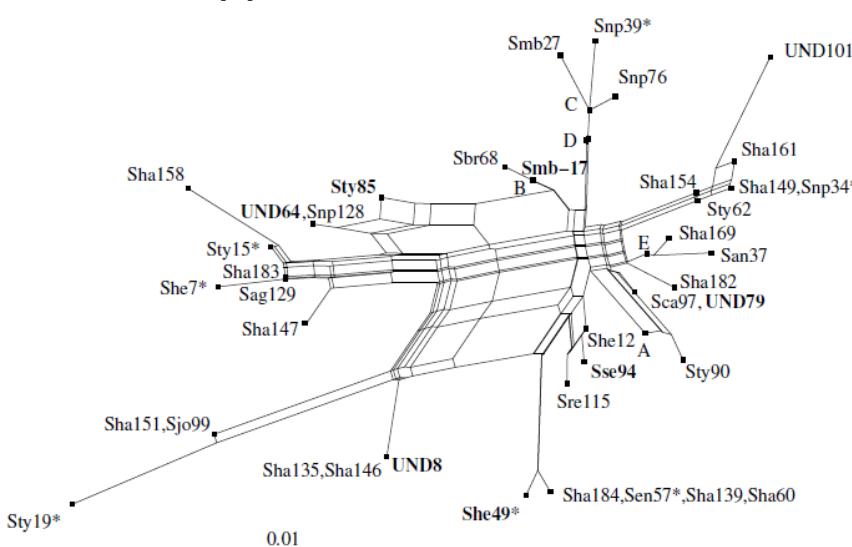
南アフリカ集団



系統ネットワーク

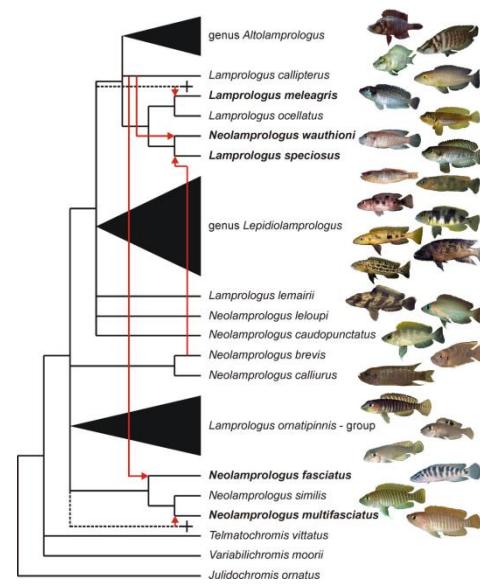


Huson, Rupp, Scornavacca 2010



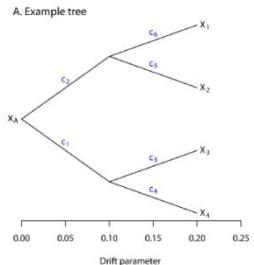
Neighbor-net

Bryant & Moulton 2003



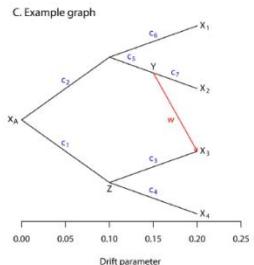
Hybridization network Koblmüller et al (2007)

Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data

Joseph K. Pickrell^{1*}, Jonathan K. Pritchard^{1,2*}¹ Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America, ² Howard Hughes Medical Institute, University of Chicago, Chicago, Illinois, United States of America

B. Covariance matrix for tree in A.

X ₁	c ₂ + c ₆	c ₂	0	0
X ₂	c ₂	c ₂ + c ₅	0	0
X ₃	0	0	c ₁ + c ₃	c ₁
X ₄	0	0	c ₁	c ₁ + c ₄



D. Covariance matrix for graph in C.

X ₁	c ₂ + c ₆	c ₂	w(c ₂)	0
X ₂	c ₂	c ₂ + c ₅ + c ₇	w(c ₂ + c ₅)	0
X ₃	w(c ₂)	w(c ₂ + c ₅)	w(c ₂ + c ₅) + (1-w)c ₁	(1-w)c ₁
X ₄	0	0	(1-w)c ₁	c ₁ + c ₄

Figure 1. Simple examples. A: An example tree. B: The covariance matrix implied by the tree structure in A. Note that the covariance here is with respect to the allele frequency at the root, and that each entry has been divided by $x_i(1-x_i)$ to simplify the presentation. C: An example graph. The migration edge is colored red. Parental populations for population 3 are labeled Y and Z; see the main text for details. D: The covariance matrix implied by the graph in C again, each entry has been divided by $x_a(1-x_a)$. The migration terms are in red, and the non-migration terms are in blue.

doi:10.1371/journal.pgen.1002967.g001

系統樹 + 混血

TreeMix

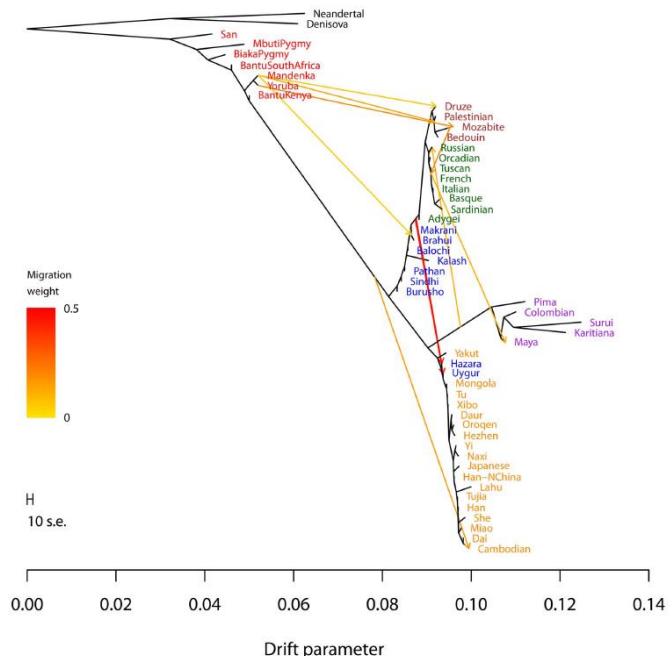
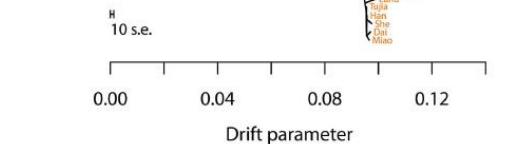
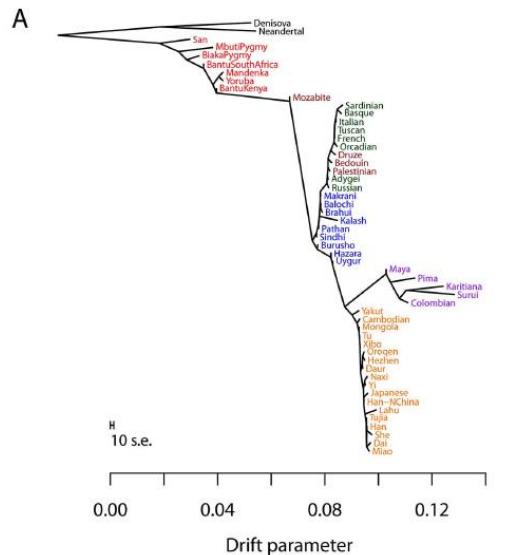
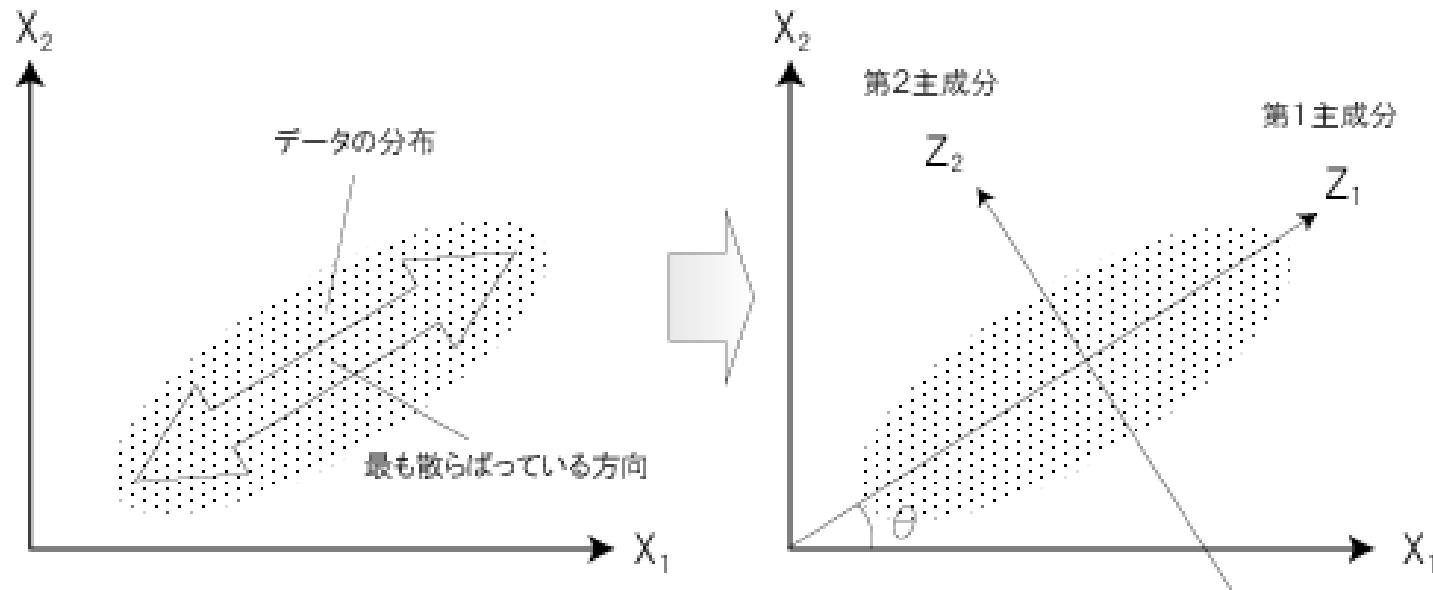


Figure 4. Inferred human tree with mixture events. Plotted is the structure of the graph inferred by TreeMix for human populations, allowing ten migration events. Migration arrows are colored according to their weight. Horizontal branch lengths are proportional to the amount of genetic drift that has occurred on the branch. The scale bar shows ten times the average standard error of the entries in the sample covariance matrix (\mathbf{W}). The residual fit from this graph is shown in Figure S9. Admixture from Neandertals to non-African populations is only apparent when considering subsets of the data (see Discussion and Figure S15).

doi:10.1371/journal.pgen.1002967.g004

主成分分析 Principal component analysis

- ・ 多数項目(多次元)のデータを、互いに無相関な主成分として低次元に縮約する統計手法。
- ・ ばらつき(分散)の大きい方向から第一主成分が生成され、最終的には項目数(あるいはサンプル数)と同じだけ主成分が生成される。
- ・ 上位の主成分のみを扱うことで、データが低次元になる。



PCA: 分散共分散行列または相関行列から計算

MDS: 任意の距離行列から計算

主成分分析 Principal component analysis

成分の数： [項目数]または[サンプル数 - 1]の小さい方

固有ベクトルeigenvector： 成分の方向を表すベクトル

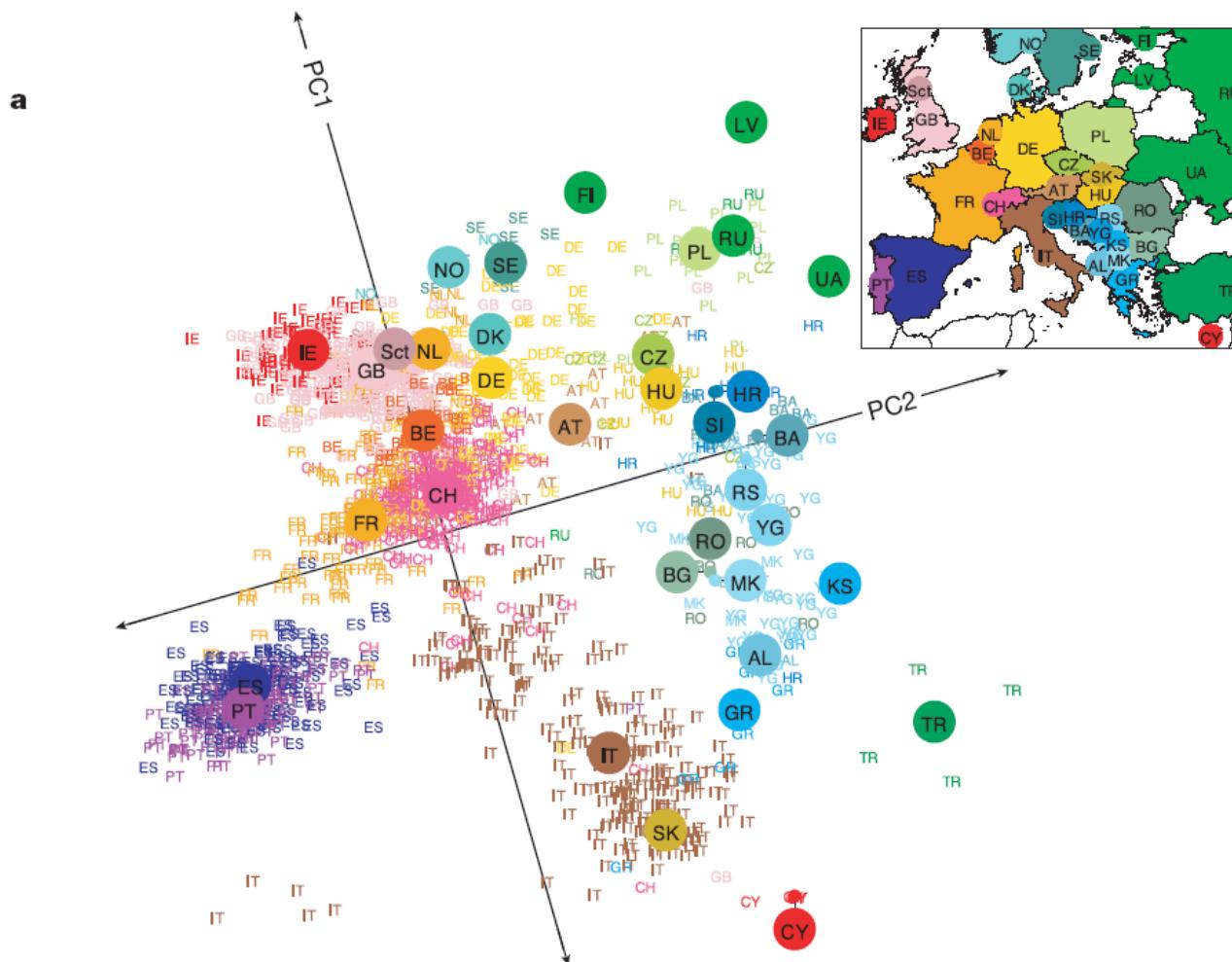
固有値eigenvalue： 成分の分散の大きさを表すスカラー値
全部足すと成分数と一致する
つまり、1より大きいとき、情報が濃縮されている

寄与率： 固有値の和を100%としたときのそれぞれの%

累積寄与率： 主成分1から足し合わせていったときの寄与率の累積

主成分得点(スコア)： 各サンプルにおける当該主成分における値

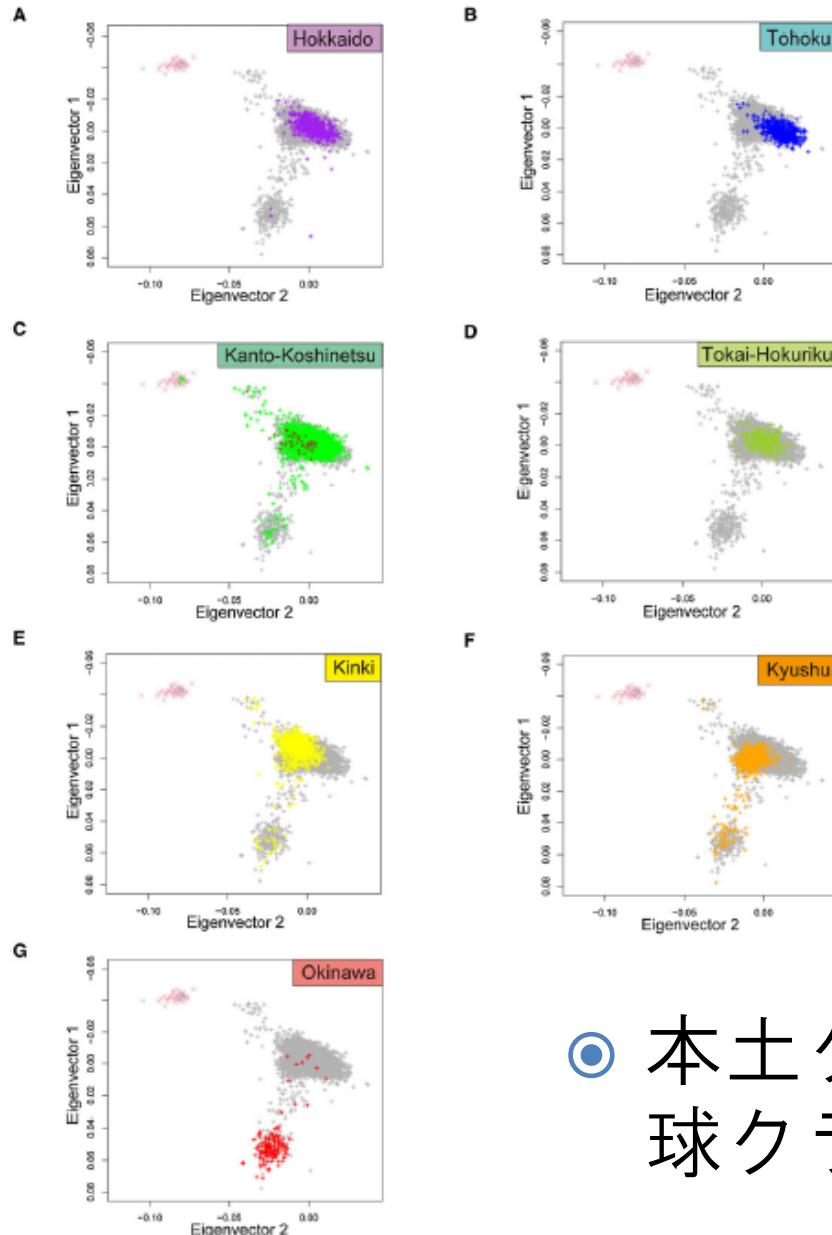
ヨーロッパ人のゲノム構造： 地理的分布と酷似



500K SNPによる主成分分析

Novembre et al. 2008

ゲノムデータが示す日本人の二重構造



約14万SNP

- 本土クラスターと分化した琉球クラスターが存在

次元削減 (<http://univprof.com/archives/16-06-10-3823706.html>)

主成分分析 (Principal Component Analysis, PCA)

説明変数から、なるべく情報量の大きくなるように潜在変数(主成分)を計算しています。

具体的には、潜在変数空間(主成分軸)におけるデータセットの分散が最大になるように主成分軸を決定しています。

独立成分分析 (Independent Component Analysis, ICA)

説明変数から、なるべく潜在変数の間の関係が独立になるように潜在変数(独立成分)を計算しています。

具体的には、たとえば潜在変数空間(独立成分の軸)におけるデータセットの尖度(kurtosis, 分布の尖り具合のこと)が最大もしくは最小になるように独立成分を決定しています。『独立』は『無相関』と似ているものの、『独立』ならば『無相関』である一方で、『無相関』だからといって『独立』とはいえません。ただまず変数の間を『無相関』にするため、ICAの前処理としてPCAが行われます。

部分的最小二乗法 (Partial Least Squares, PLS)

説明変数から、なるべく説明変数の情報をもち、かつ目的変数との関係が大きいように潜在変数を計算しています。

具体的には、目的変数との共分散が最大になるように潜在変数を決定しています。

正準相関解析 (Canonical Correlation Analysis, CCA)

説明変数から、なるべく目的変数との相関関係が大きいように潜在変数を計算しています。

PLSとともによく似ています。PLSと異なるのは、PLSでは目的変数との共分散が最大になるように潜在変数を決定しているのに対し、CCAでは目的変数との相関係数が最大になるように潜在変数を決定しているところです。これだけです。いいかえるとCCAでは、目的変数がオースケーリングされているとき、共分散を潜在変数の分散で割ったものを最大化していることになります。PLSと比べて、潜在変数の分散は小さい一方で目的変数と相関の高い潜在変数が計算できることになります。

線形判別分析 (Linear Discriminant Analysis, LDA)

説明変数から、なるべくクラスを分けられるように潜在変数を計算しています。

具体的には、潜在変数空間におけるクラス間のばらつきを、同じく潜在変数空間におけるクラス内でのばらつきで割ったものが最大になるように潜在変数を決定しています。

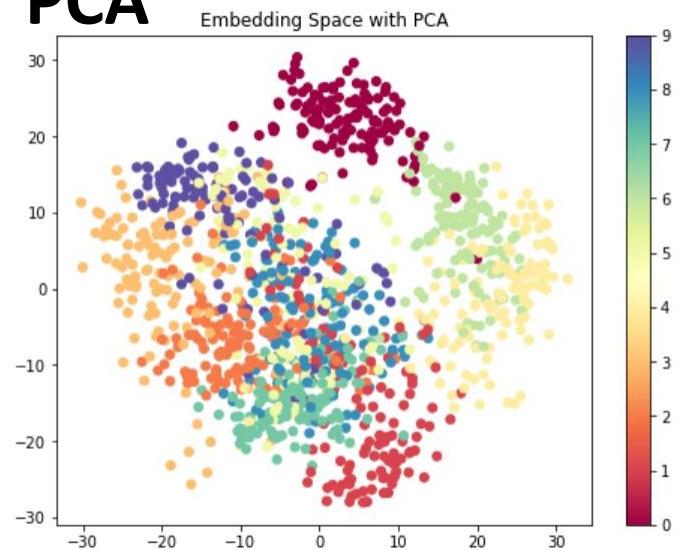
t分布型確率的近傍埋め込み法(T-distributed Stochastic Neighbor Embedding, t-SNE)

Laurens van der Maatenとジェフリー・ヒントンにより開発された可視化のための機械学習アルゴリズムである。^[1]これは、高次元データの可視化のため2次元または3次元の低次元空間へ埋め込みに最適な非線形次元削減手法である。具体的には、高次元のデータ集合を2次元または3次元へ配置する際に、高い確率で類似した集合が近傍に、異なる集合が遠方となるように対応付ける。

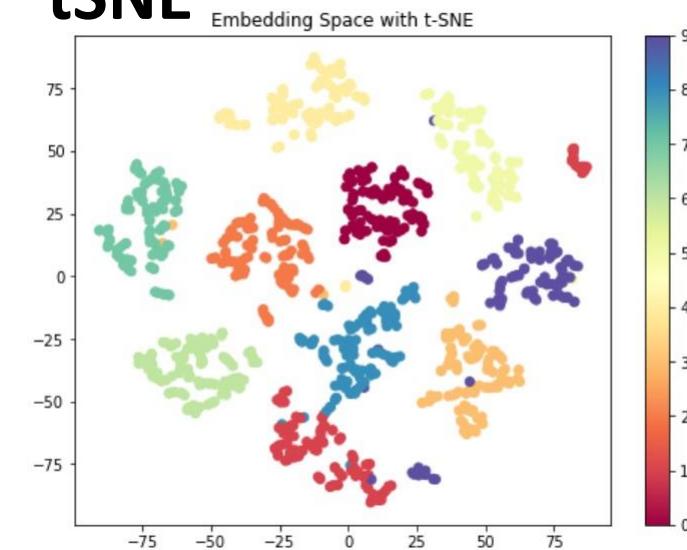
均一マニホールド近似投影 (uniform manifold approximation and projection, UMAP)

最も新しい機械学習による非線形次元削減手法の一つ。リーマン幾何学と代数トポロジーに基づき、高次元空間のデータ構造を保ち、トポロジー間のクロス・エントロピーを最小にしながら低次元のデータに変換する手法。計算速度がt-SNEよりも高速であるのも特徴。

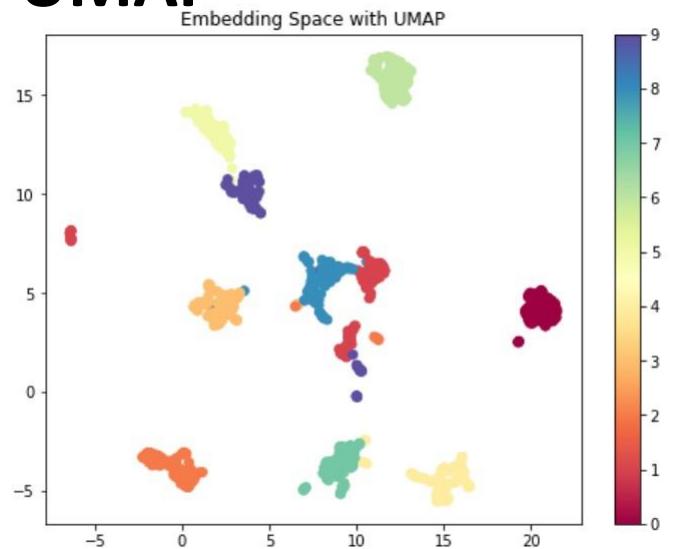
PCA



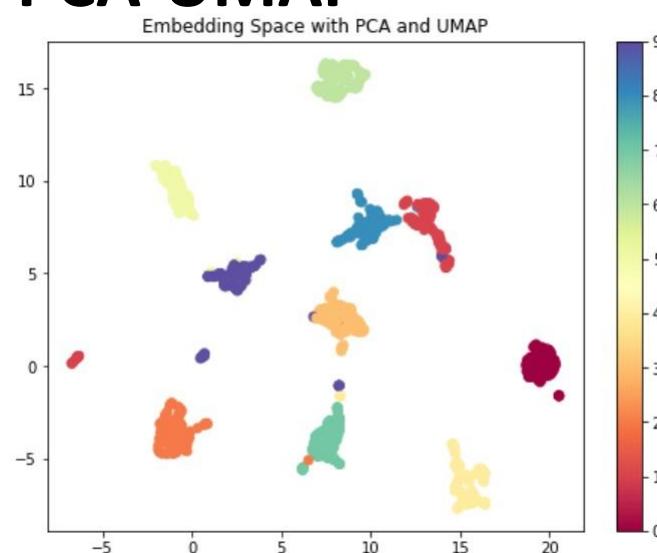
tSNE



UMAP



PCA-UMAP



クラスタ分析

代表的なクラスタ分析法

k-means法 : <http://tech.nitoyon.com/ja/blog/2013/11/07/k-means/>

STRUCTURE解析

類似のクラスタ分析法

FRAPPE, ADMIXTURE

遺伝学的なモデルに基づいたクラスタ分析

Inference of Population Structure Using Multilocus Genotype Data

Jonathan K. Pritchard, Matthew Stephens and Peter Donnelly

Department of Statistics, University of Oxford, Oxford OX1 3TG, United Kingdom

Manuscript received September 23, 1999

Accepted for publication February 18, 2000

MCMC algorithm (with admixture): The following algorithm may be used to sample from $\Pr(Z, P, Q|X)$.

ALGORITHM 2: Starting with initial values $Z^{(0)}$ for Z (by drawing $Z^{(0)}$ at random using (3) for example), iterate the following steps for $m = 1, 2, \dots$

Step 1. Sample $P^{(m)}, Q^{(m)}$ from $\Pr(P, Q|X, Z^{(m-1)})$.

Step 2. Sample $Z^{(m)}$ from $\Pr(Z|X, P^{(m)}, Q^{(m)})$.

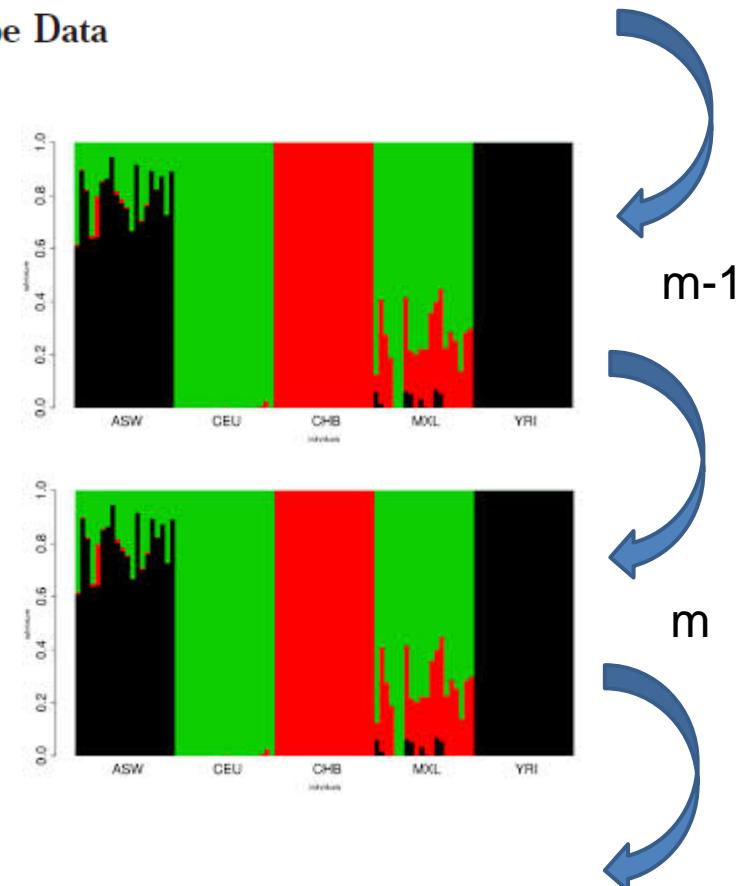
Step 3. Update α using a Metropolis-Hastings step.

Z: populations of origin of the individuals

P: allele frequencies in all populations

Q: admixture proportions for each individuals

X: observed genotypes



Human Genome Diversity Panel (HGDP)の650K SNP解析

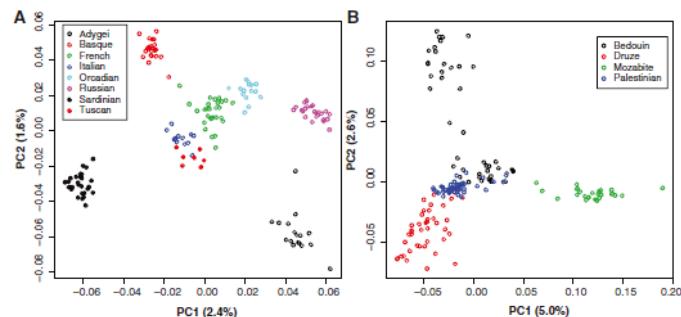


Fig. 2. Fine-scale population structure principal component analyses in two geographic regions, using all autosomal SNPs. (A) Europe. (B) The Middle East.

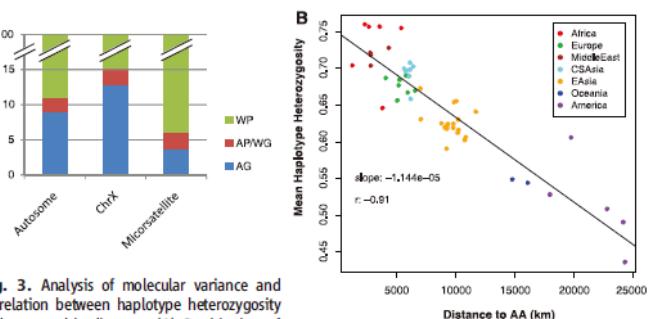


Fig. 3. Analysis of molecular variance and correlation between haplotype heterozygosity and geographic distance. (A) Partitioning of genetic variance into three components (28); Within-Population (WP), Among-Population-Within-Region (AP/WG), and Among-Region (AG), by using autosomal SNPs, microsatellite markers, and ChrX SNPs, respectively. (B) SNP haplotype heterozygosity versus geographic distances from Addis Ababa (AA), Ethiopia. The linear regression slope is indicated along with the Pearson correlation r .

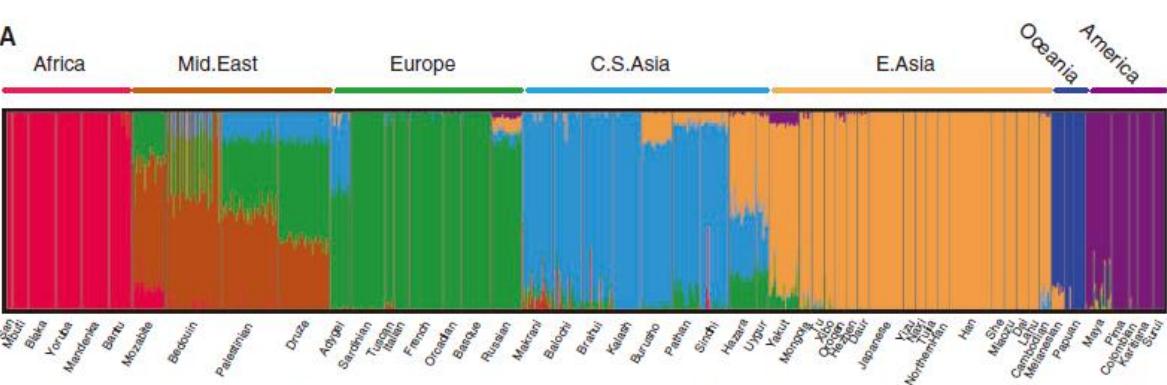
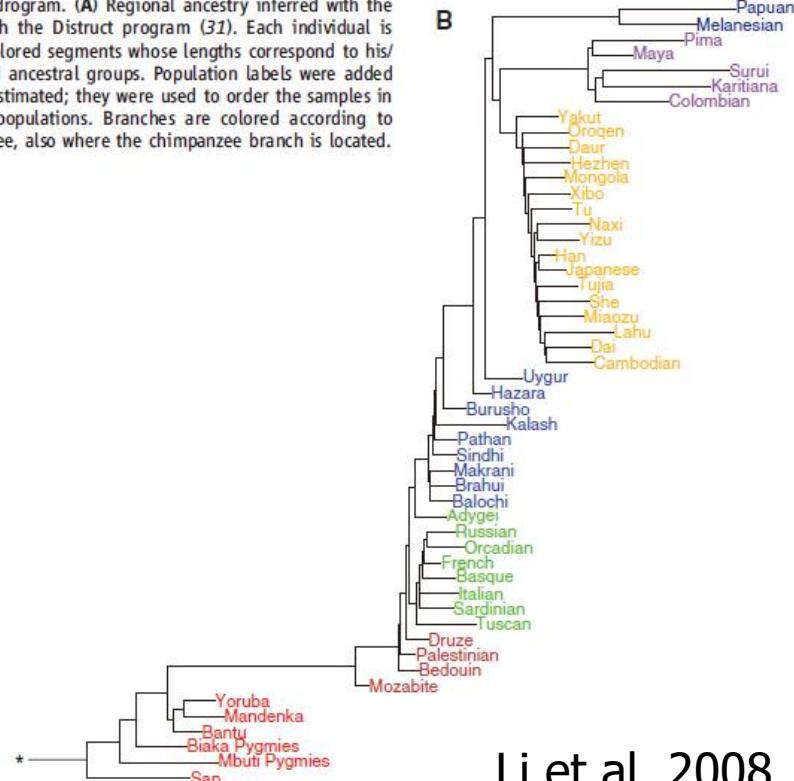


Fig. 1. Individual ancestry and population dendrogram. (A) Regional ancestry inferred with the *rappe* program at $K = 7$ (13) and plotted with the *Distruct* program (31). Each individual is represented by a vertical line partitioned into colored segments whose lengths correspond to his/her ancestry coefficients in up to seven inferred ancestral groups. Population labels were added only after each individual's ancestry had been estimated; they were used to order the samples in plotting. (B) Maximum likelihood tree of 51 populations. Branches are colored according to continents/regions. * indicates the root of the tree, also where the chimpanzee branch is located.



Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation

Jun Z. Li,^{1,2,*†} Devin M. Absher,^{1,2,*} Hua Tang,¹ Audrey M. Southwick,^{1,2} Amanda M. Casto,¹ Sohini Ramachandran,⁴ Howard M. Cann,⁵ Gregory S. Barsh,^{1,3} Marcus Feldman,^{4,‡} Luigi L. Cavalli-Sforza,^{1,‡} Richard M. Myers^{1,2,‡}

Li et al. 2008

50K SNPを用いたアジア諸国の遺伝的多様性

Mapping Human Genetic Diversity in Asia

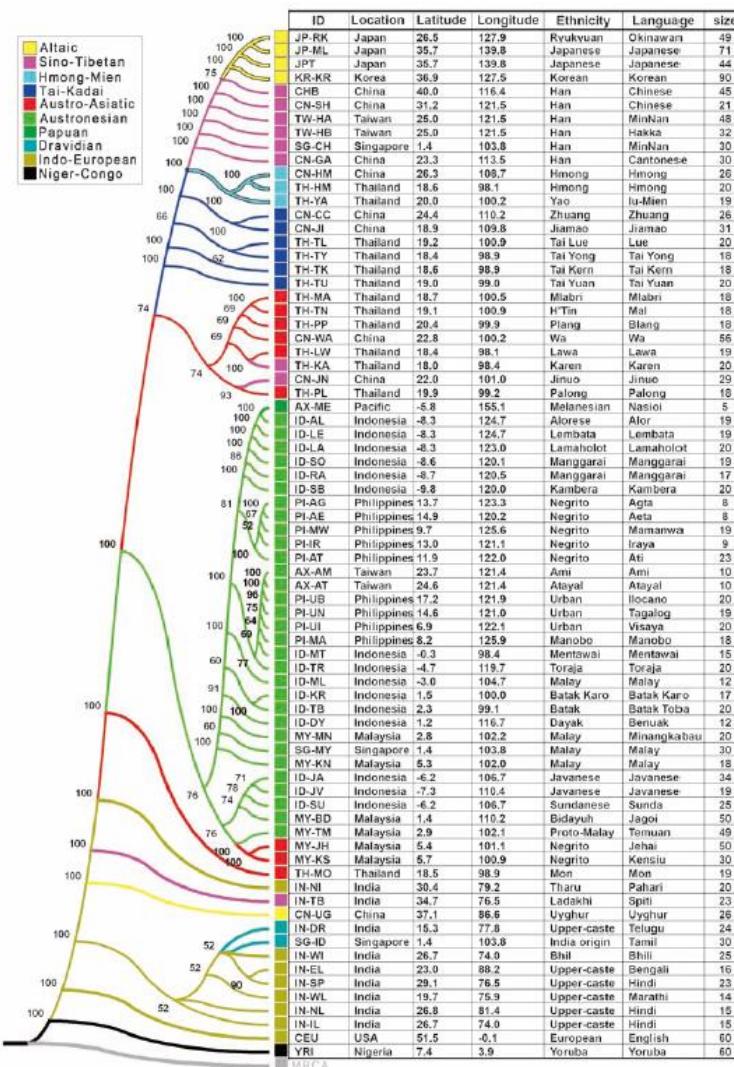
The HUGO Pan-Asian SNP Consortium*†

Asia harbors substantial cultural and linguistic diversity, but the geographic structure of genetic variation across the continent remains enigmatic. Here we report a large-scale survey of autosomal variation from a broad geographic sample of Asian human populations. Our results show that genetic ancestry is strongly correlated with linguistic affiliations as well as geography. Most populations show relatedness within ethnic/linguistic groups, despite prevalent gene flow among populations. More than 90% of East Asian (EA) haplotypes could be found in either Southeast Asian (SEA) or Central-South Asian (CSA) populations and show clinal structure with haplotype diversity decreasing from south to north. Furthermore, 50% of EA haplotypes were found in SEA only and 5% were found in CSA only, indicating that SEA was a major geographic source of EA populations.

HUGO Pan-Asian SNP Consortium 2009

- China
- Indonesia
- India
- Japan
- Korea
- Malaysia
- Philippines
- Singapore
- Thailand
- Taiwan

遺伝距離
STRUCTURE解析



K = 14

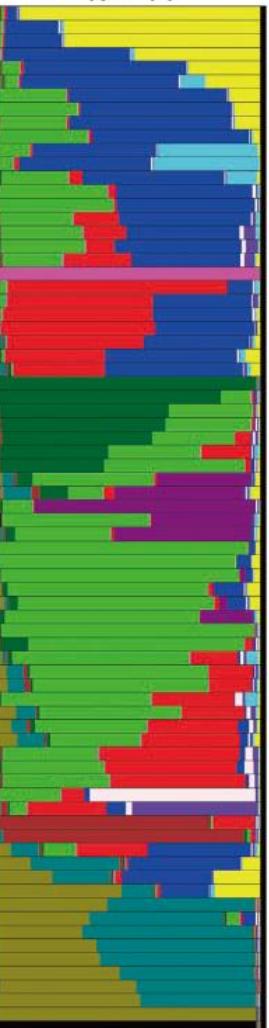


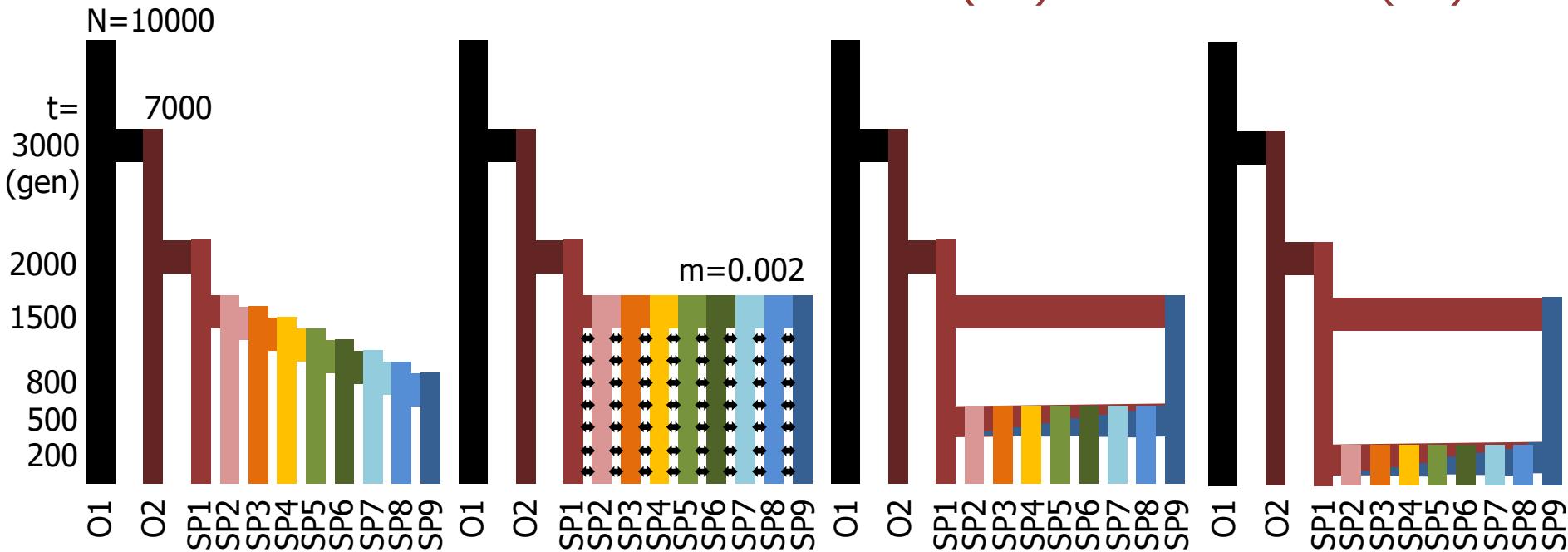
Fig. 1. Maximum-likelihood tree of 75 populations. A hypothetical most-recent common ancestor (MRCA) composed of ancestral alleles as inferred from the genotypes of one gorilla and 21 chimpanzees was used to root the tree. Branches with bootstrap values less than 50% were condensed. Population identification numbers (IDs), sample collection locations with latitudes and longitudes, ethnicities, language spoken, and size of population samples are shown in the table adjacent to each branch in the tree. Linguistic groups are indicated with colors as shown in the legend. All

population IDs except the four HapMap samples are denoted by four characters. The first two letters indicate the country where the samples were collected or (in the case of Affymetrix) genotyped, according to the following convention: AX, Affymetrix; CN, China; ID, Indonesia; IN, India; JP, Japan; KR, Korea; MY, Malaysia; PI, the Philippines; SG, Singapore; TH, Thailand; and TW, Taiwan. The last two letters are unique IDs for the population samples. To the right of the table, an averaged graph of results from STRUCTURE is shown for K = 14.

Simulation study

Coalescent simulation (*msms*)

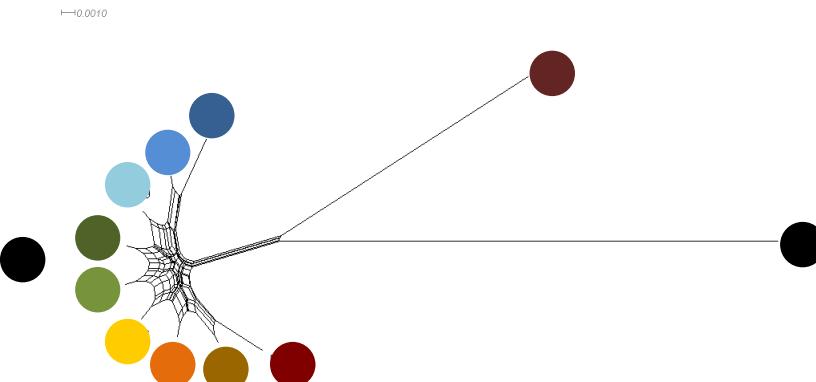
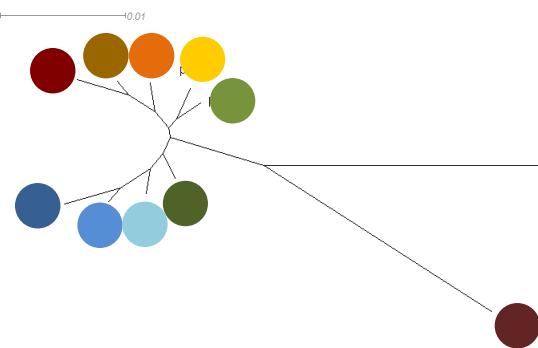
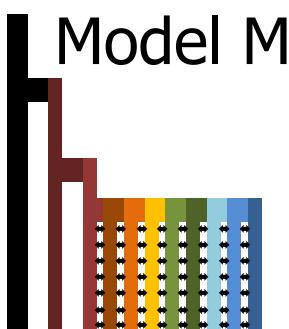
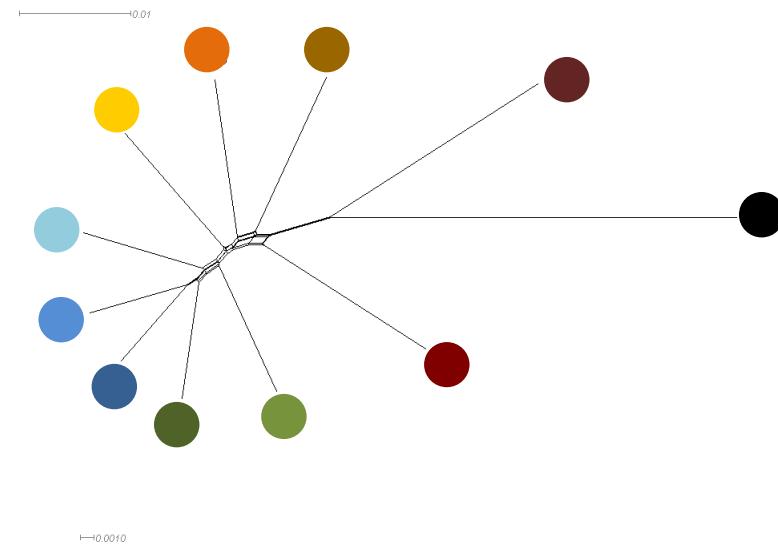
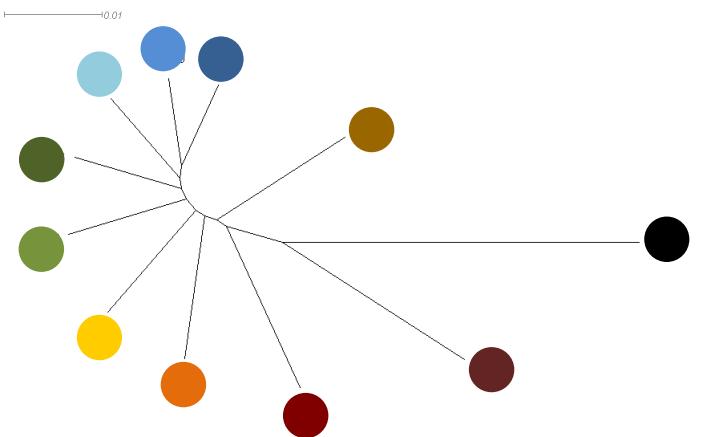
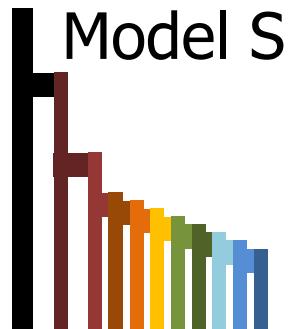
- ◆ Cascading splits(S)
- ◆ Migrations (M)
- ◆ Admixtures-Isolation:
 - Old (OA)
 - Recent(RA)



- ◆ 10,000 independent SNPs with MAF>10%
- ◆ 10 diploids/population

O: outgroup, SP: subpopulation

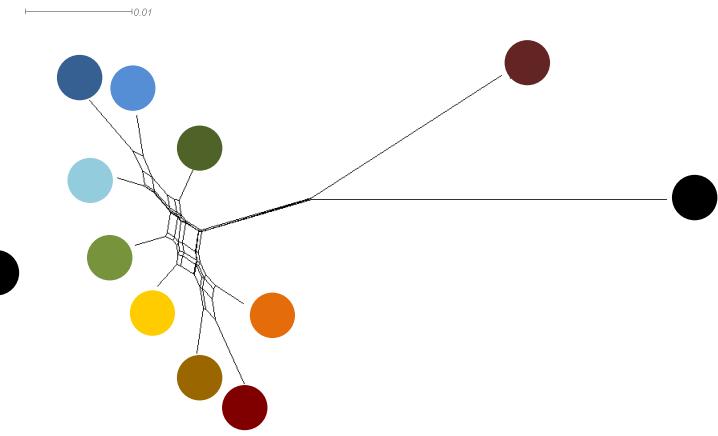
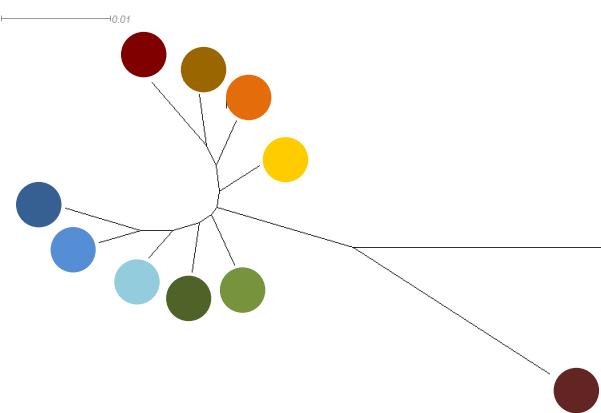
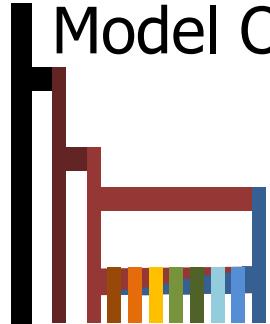
Phylogenetic analysis (Splits Tree4)



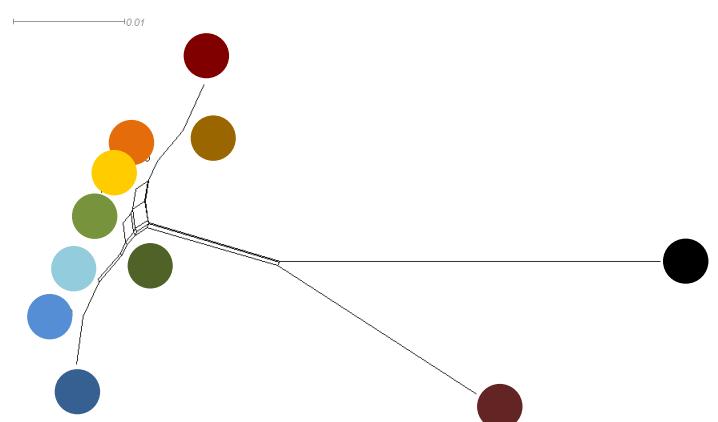
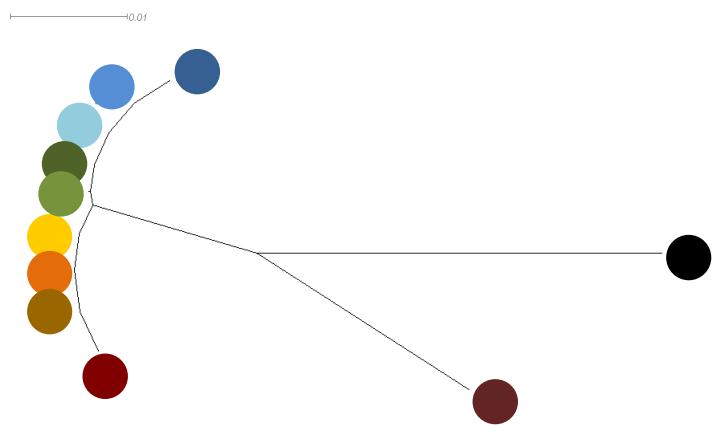
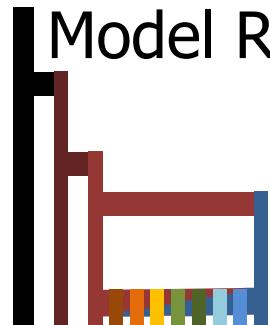
Neighbor-joining tree

Neighbor net

Model OA



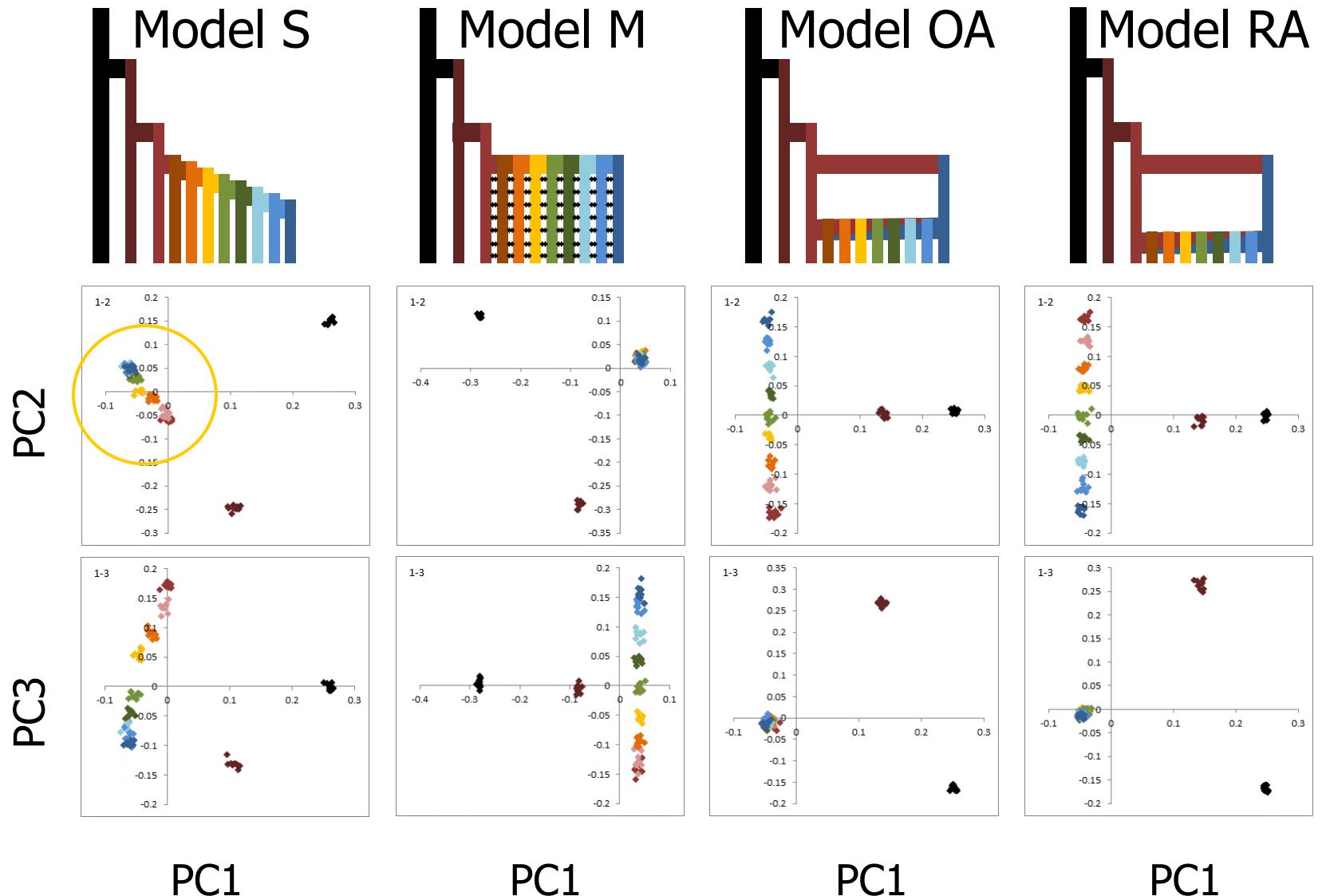
Model RA



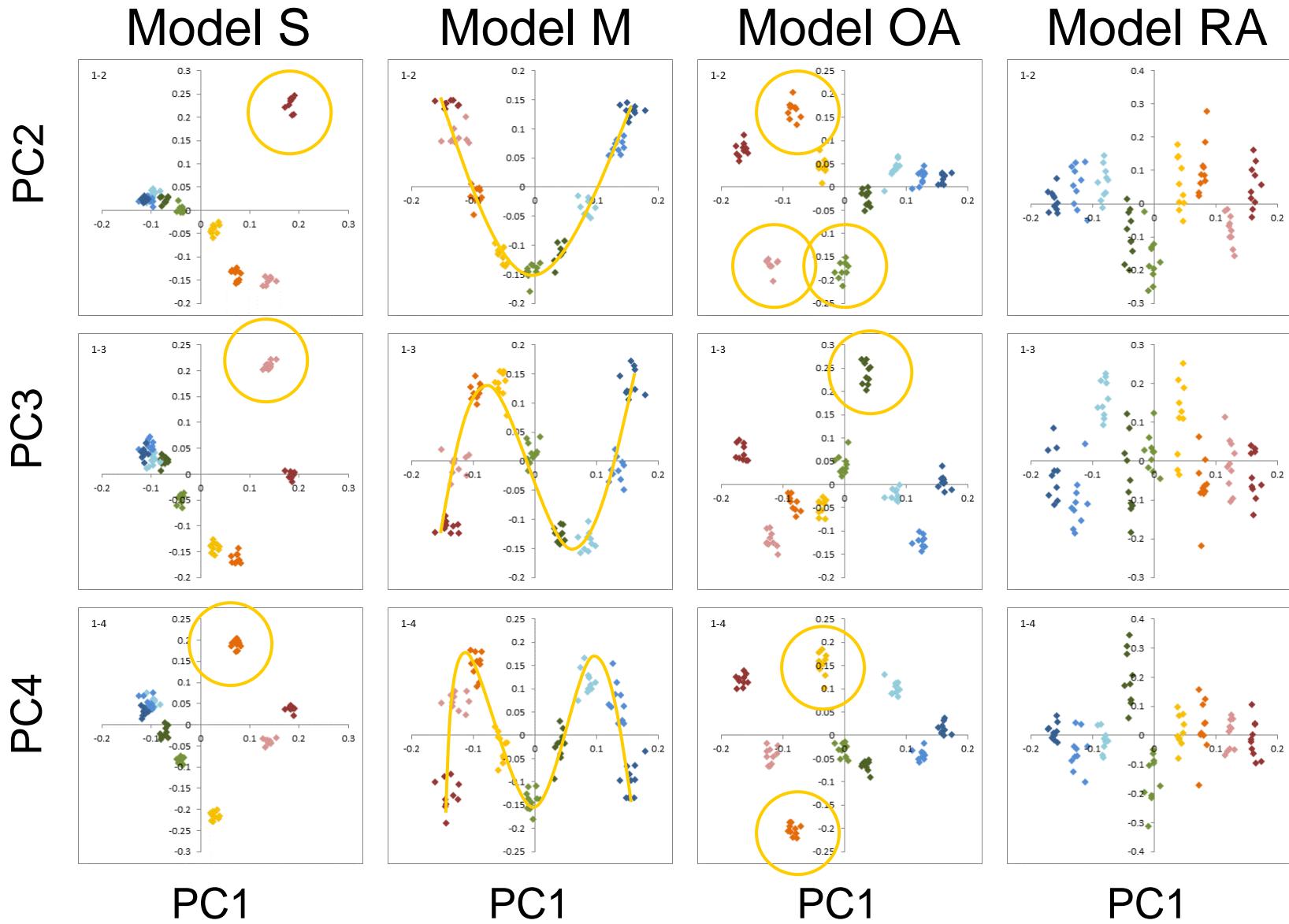
Neighbor-joining tree

Neighbor net

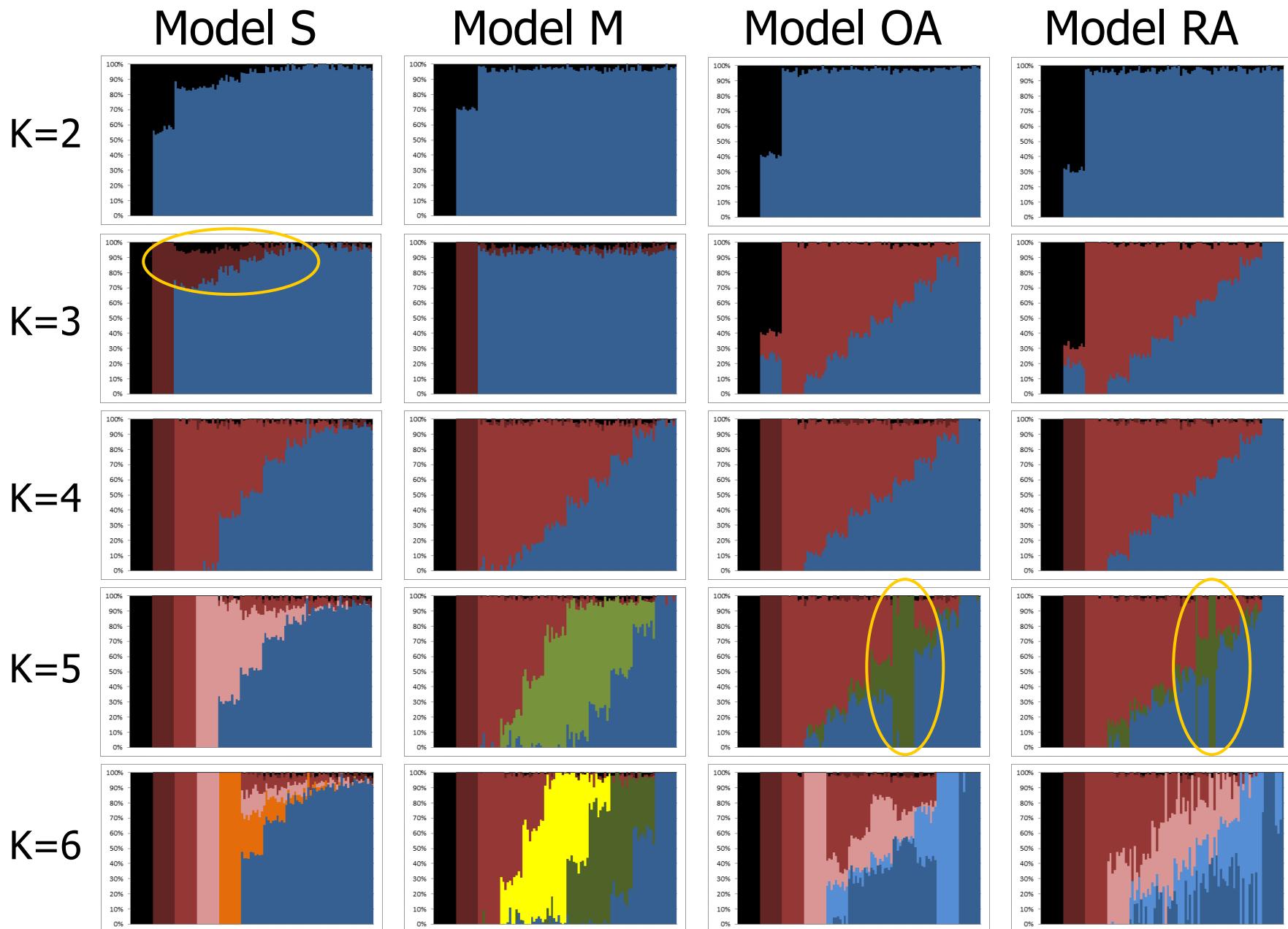
Principal component analysis (EIGENSTRAT)



Without outgroups



Clustering method (ADMIIXTURE)



MENU

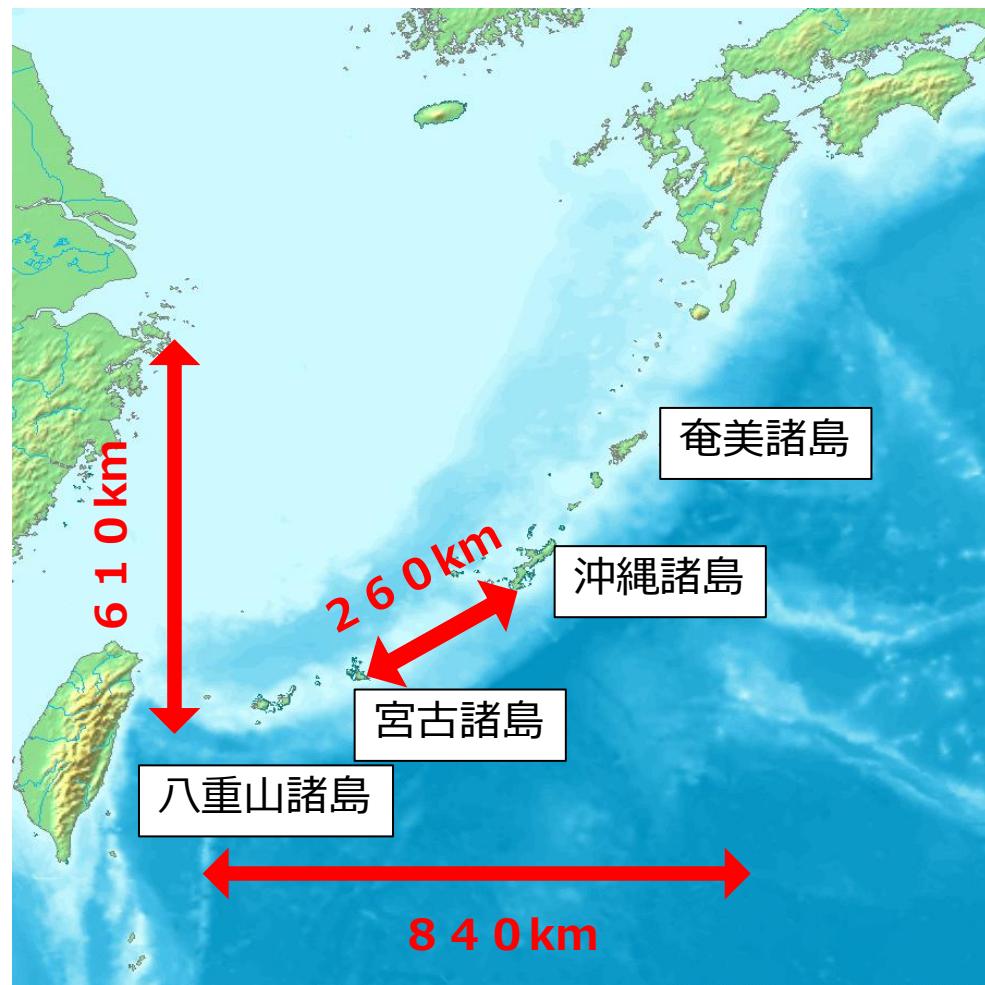
1. 集団遺伝学入門
2. ゲノム多様性を表現する
3. 琉球人集団の形成
4. ゲノム多様性と表現型多様性

琉球列島の地理的特徴

- 広範囲に分布する島々から成る
- 島間は広い海洋によって隔離
- すぐ西には台湾

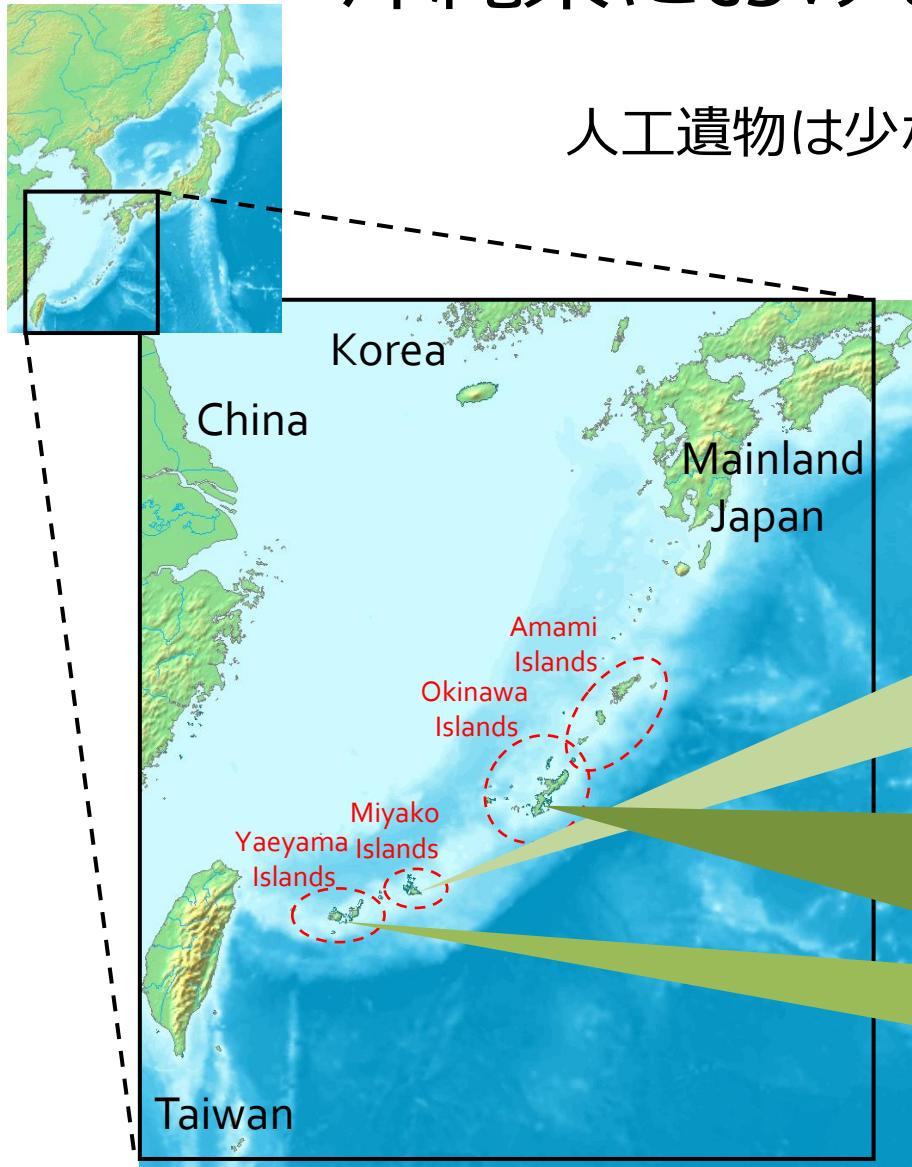


孤立・分集団化が
起こりやすい
地理的環境



沖縄県における更新世人骨

人工遺物は少ないが骨は残存



ピンザアブ



白保竿根田原



港川

石垣島・白保竿根田原遺跡



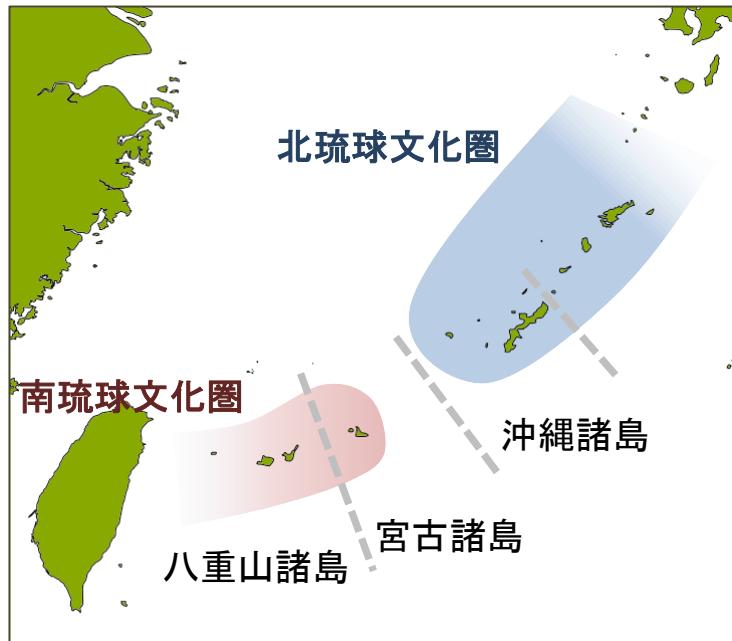
八重山毎日新聞

沖縄島・武芸洞・サキタリ洞遺跡

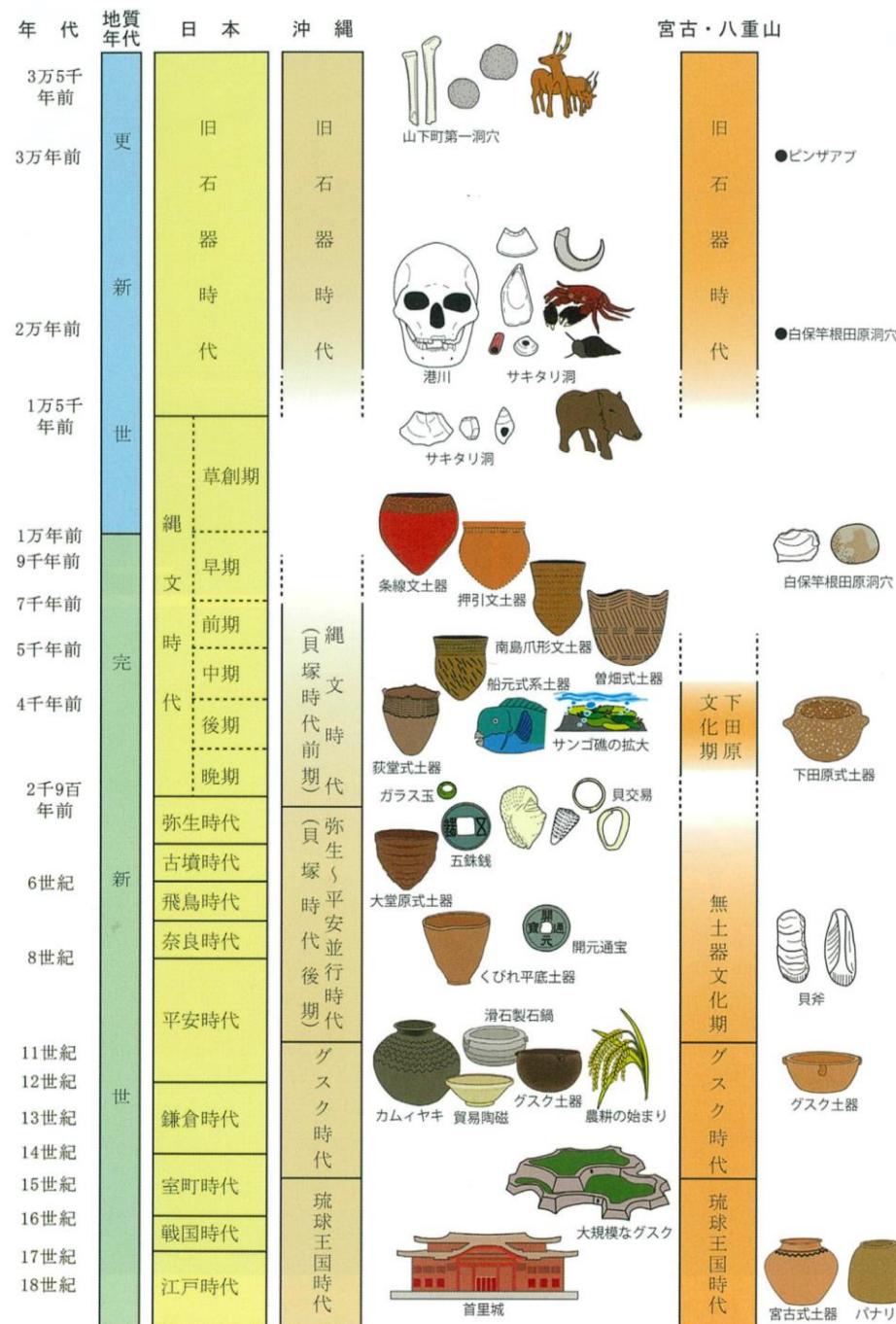


- 沖縄の先史考古学が非常にHOT

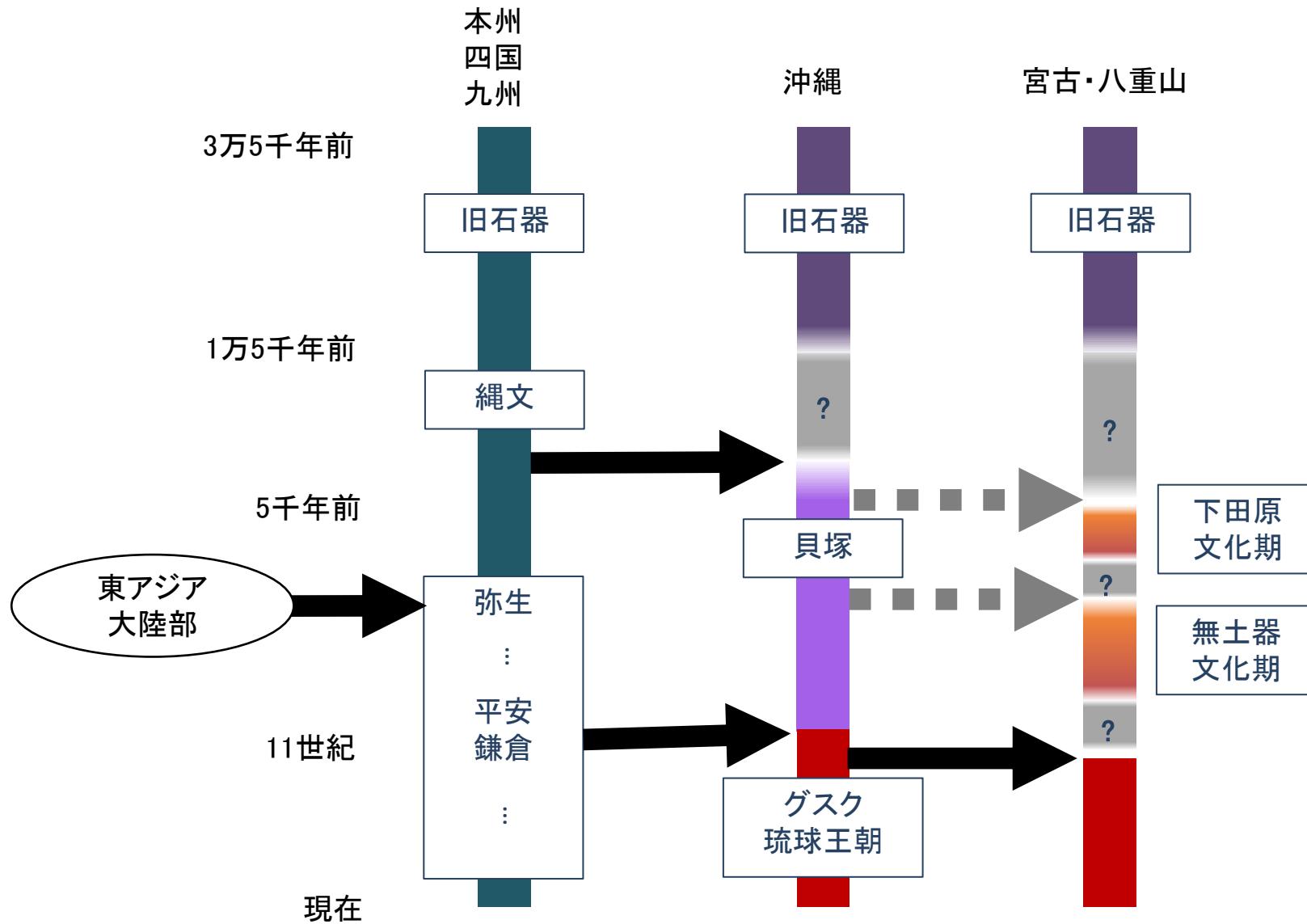
琉球列島の 文化と歴史



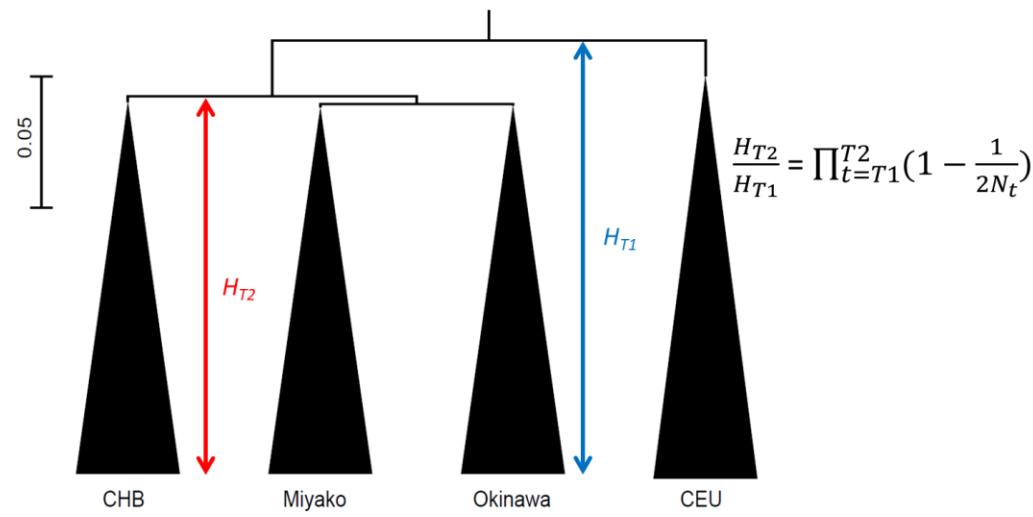
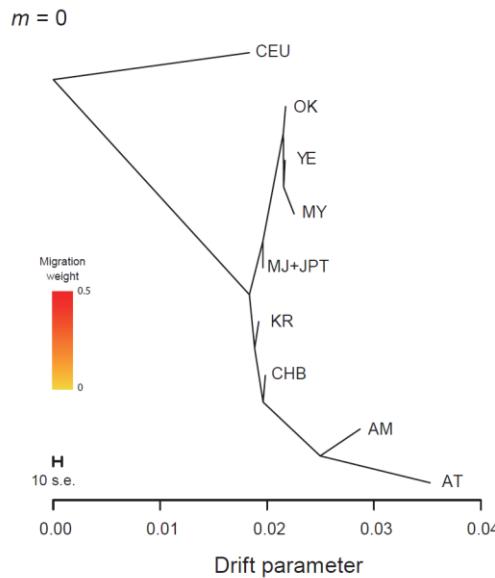
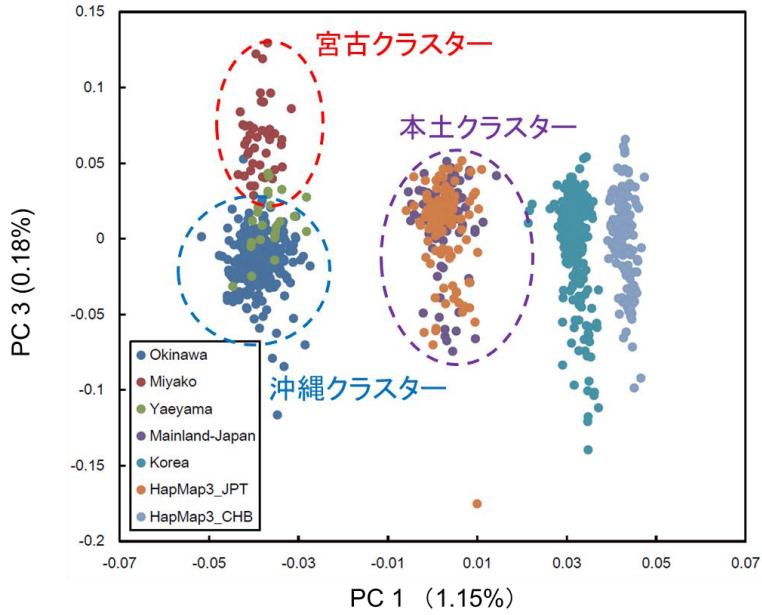
(安里 2003 を改変)



琉球列島への大規模なヒトの移動



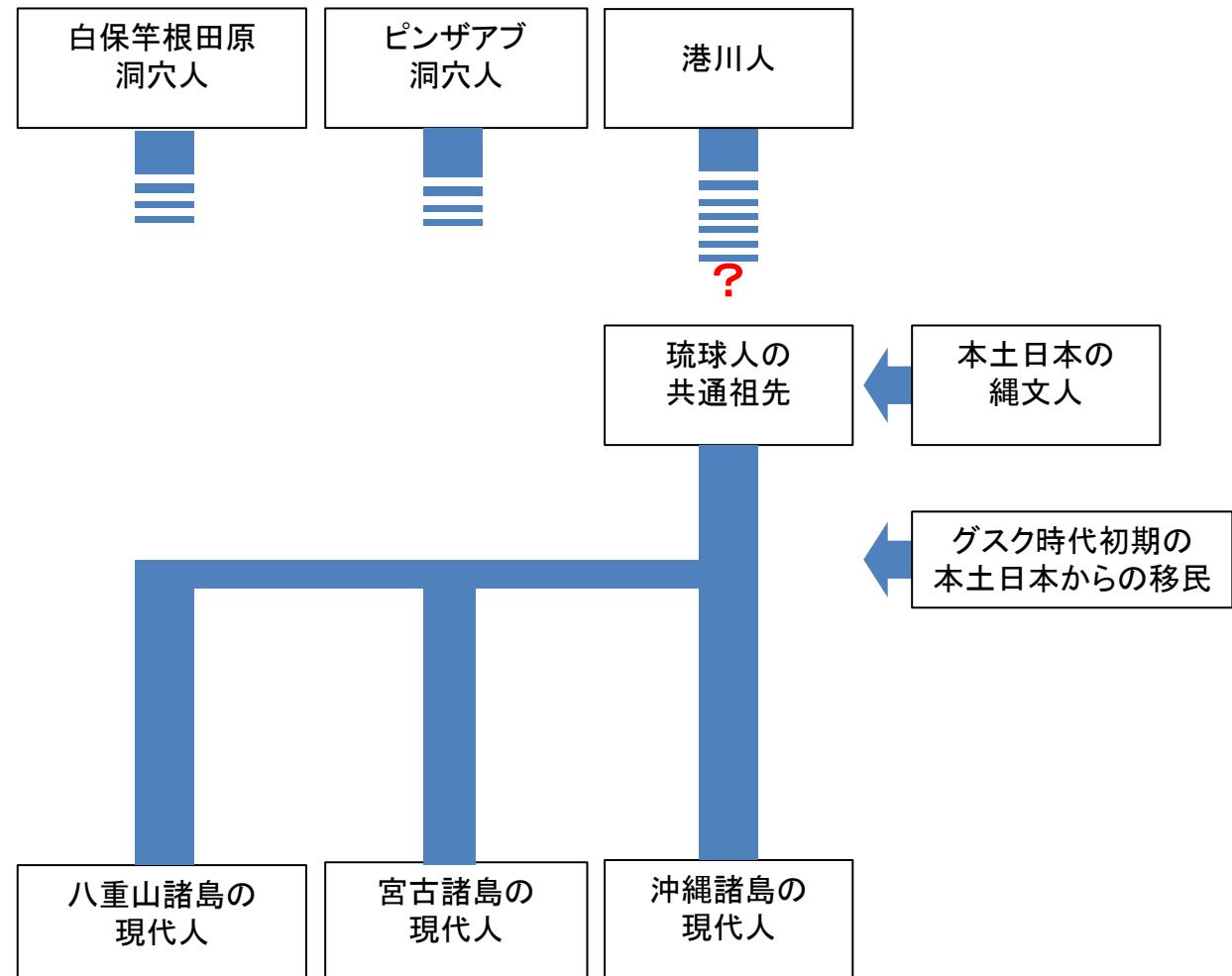
ゲノムワイドSNPデータでみる現代琉球列島人



Sato et al. 2014

宮古諸島人が沖縄本島人と遺伝的に分化していることを見出し、集団の年代を推定

現在の琉球人と更新世時代人の関係



主要な祖先とは考えにくい



沖縄バイオインフォメーションバンク

沖縄バイオインフォメーションバンク 研究へのご協力のお願い

沖縄県民の健康長寿復活を目指し、
ゲノム・遺伝子解析を行い、最適な
医療を構築することを目的としています。



●お願いしたいことは次の通りです。(同意は任意です)

- 1** 健診の採血の時に余分に 7cc の血液をいただきます。
- 2** 健診の問診、検査データなどの情報を提供していただきます。
- 3** 今後の健診情報を医療情報システムを使って提供していただきます。

※ご協力いただいてもすぐにあなたの治療に役立つものではありません。
ご協力いただかなくても、健診は通常通り行われ、あなたの不利益になることはありません。



琉球大学医学部



先進ゲノム検査医学講座
前田士郎 教授
今村美菜子 准教授

主に生活習慣病関連遺伝子同定のために検体を収集

宮古諸島住民1240人を用いた集団解析

Fine-Scale Genetic Structure and Demographic History in the Miyako Islands of the Ryukyu Archipelago

Masatoshi Matsunami,¹ Kae Koganebuchi,² Minako Imamura,^{1,3} Hajime Ishida,⁴ Ryosuke Kimura ,⁴ and Shiro Maeda ,^{1,3}

¹Department of Advanced Genomic and Laboratory Medicine, Graduate School of Medicine, University of the Ryukyus, Nishihara-Cho, Japan

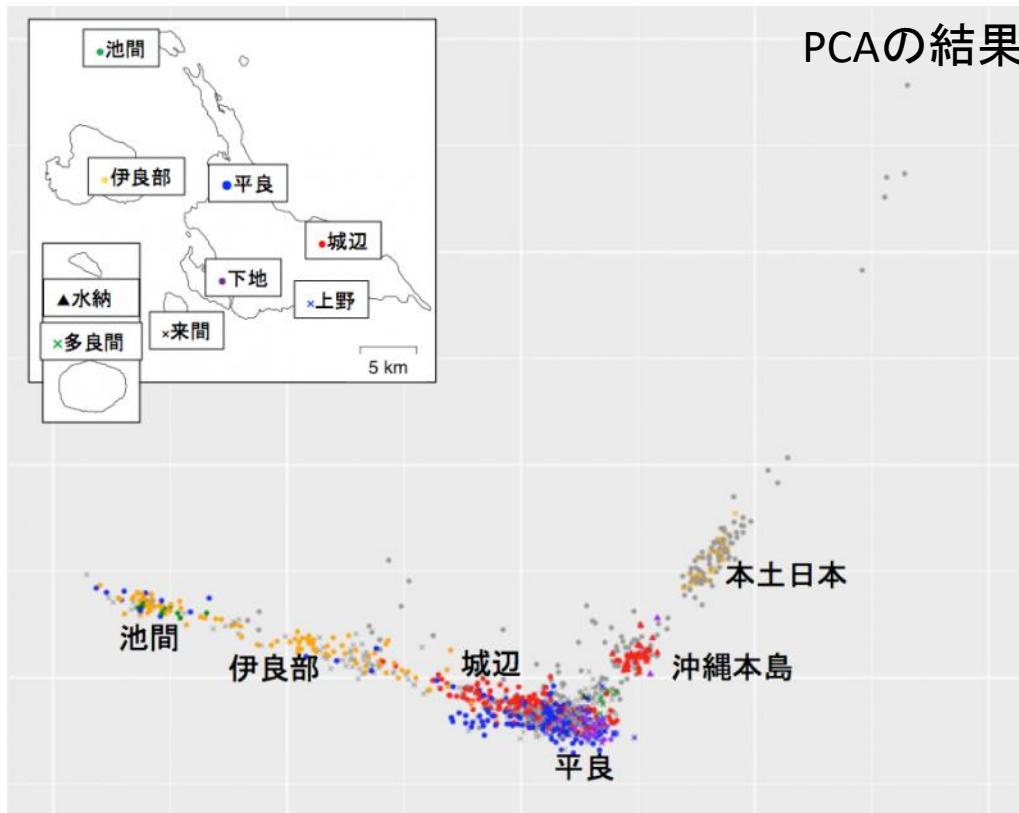
²Advanced Medical Research Center, Faculty of Medicine, University of the Ryukyus, Nishihara-Cho, Japan

³Division of Clinical Laboratory and Blood Transfusion, University of the Ryukyus Hospital, Nishihara-Cho, Japan

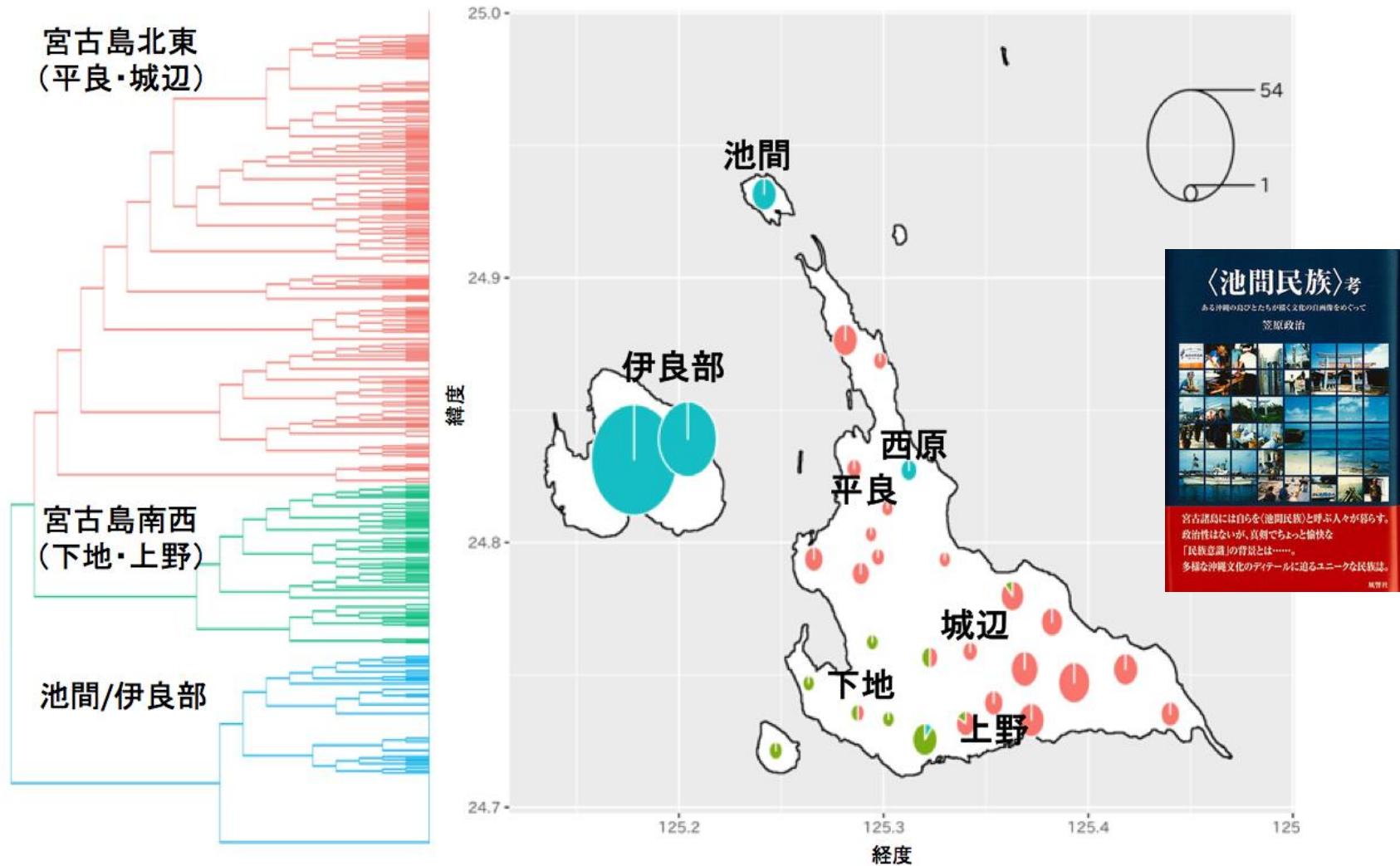
⁴Department of Human Biology and Anatomy, Graduate School of Medicine, University of the Ryukyus, Nishihara-Cho, Japan

*Corresponding author: E-mail: smaeda@med.u-ryukyu.ac.jp.

Associate editor: Kelley Harris



fineSTRUCTURE解析の結果



地域と一致する3つのクラスターを観察

fastsimcoal2を用いた集団史のモデリング

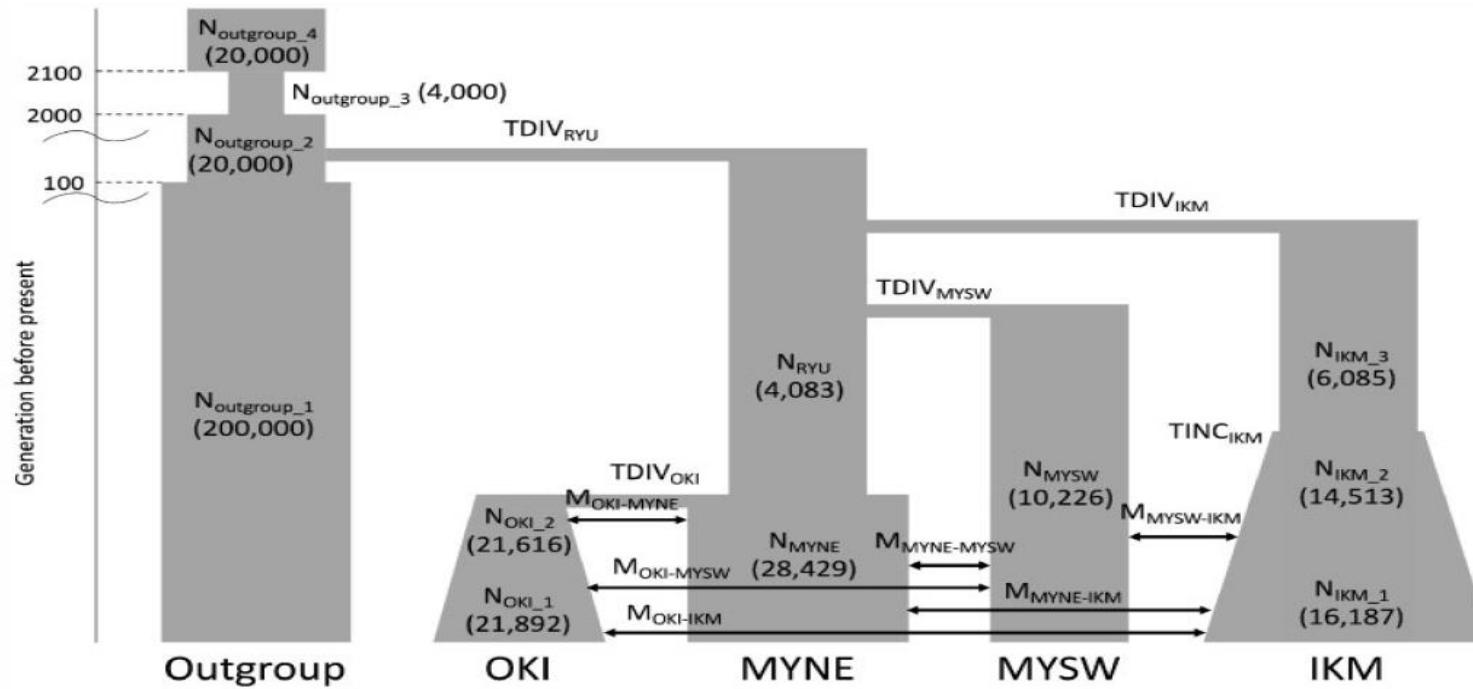


Table 2. Estimated Demographic Parameters with 95% CIs by fastsimcoal2.

Parameter	Point Estimation	95% CI
TDIV_{OKI}	7	2–18
$\text{TDIV}_{\text{MYSW}}$	16	10–27
TDIV_{IKM}	41	38–57
TDIV_{RYU}	109	104–125
TINC_{IKM}	10	1–19
$N_{\text{OKI_1}}$	21,892	20,271–27,770
$N_{\text{OKI_2}}$	21,616	15,435–27,415
N_{MYNE}	28,429	20,840–37,602
N_{MYSW}	10,226	3,156–15,804
$N_{\text{IKM_1}}$	16,187	5,168–16,750
$N_{\text{IKM_2}}$	14,513	3,775–17,496
$N_{\text{IKM_3}}$	6,085	4,277–15,420
N_{RYU}	4,083	3,708–4,939
$M_{\text{OKI-MYNE}}$	3.44×10^{-4}	$6.64 \times 10^{-9}–9.44 \times 10^{-3}$
$M_{\text{OKI-MYSW}}$	8.95×10^{-5}	$5.22 \times 10^{-9}–3.63 \times 10^{-2}$
$M_{\text{OKI-IKM}}$	7.09×10^{-8}	$2.58 \times 10^{-10}–1.05 \times 10^{-2}$
$M_{\text{MYNE-MYSW}}$	4.22×10^{-6}	$2.96 \times 10^{-9}–5.78 \times 10^{-2}$
$M_{\text{MYNE-IKM}}$	1.47×10^{-6}	$2.06 \times 10^{-9}–3.47 \times 10^{-3}$
$M_{\text{MYSW-IKM}}$	1.64×10^{-9}	$5.55 \times 10^{-10}–1.05 \times 10^{-3}$

SNP情報から推定された 沖縄本島から宮古諸島への移動



宮古島・無土器時代の古代ゲノム解析

Article

Triangulation supports agricultural spread of the Transeurasian languages

<https://doi.org/10.1038/s41566-021-04108-8>

Received: 18 February 2021

Accepted: 7 October 2021

Published online: 10 November 2021

Open access

Check for updates

Martijn Robbeets^{1,2*}, Remco Bouckaert^{1,3}, Matthew Conti⁴, Alexander Savel'ev^{5,6}, Tao Li^{1,3,8}, Deog-Im An⁷, Ken-ichi Shinoda⁸, Yingqi Cui^{9,10}, Takamune Kawashima¹¹, Geonyoung Kim¹², Junzo Uchiyama^{12,13}, Joanna Dolinska¹⁴, Sofia Oskolskaya¹⁴, Ken-Yoju Yamano¹⁵, Noriko Seguchi^{16,17}, Hirotaka Tomita^{18,19}, Hiroto Takamatsu¹⁹, Hideki Kanazawa-Kiriyama¹⁹, Hiroki Oota¹⁹, Hajimu Ishida²⁰, Ryosuke Kimura²⁰, Takehiko Sato²⁰, Jae-Hyun Kim²¹, Bingcong Deng²², Rasmus Bjørn²³, Seongha Rhee²⁴, Kyu-Dong Ahn²⁵, Ilya Gruntov²⁶, Olga Mazo²⁶, John R. Bentley²⁷, Ricardo Fernandes^{28,29}, Patrick Roberts¹, Itona R. Bausch^{12,30,31}, Linda Giltazeau¹, Minoru Yoneda³², Mitsugu Kuga³³, Raffaela A. Bianco¹, Fan Zhang¹, Marie Himmel¹, Mark J. Hudson^{1,24} & Chao Ning^{2,30,31}

宮古島・長墓遺跡 無土器時代の古人骨ゲノム

→ 繩文人と同様



無土器文化の南方（フィリピン・台湾）由来说を否定

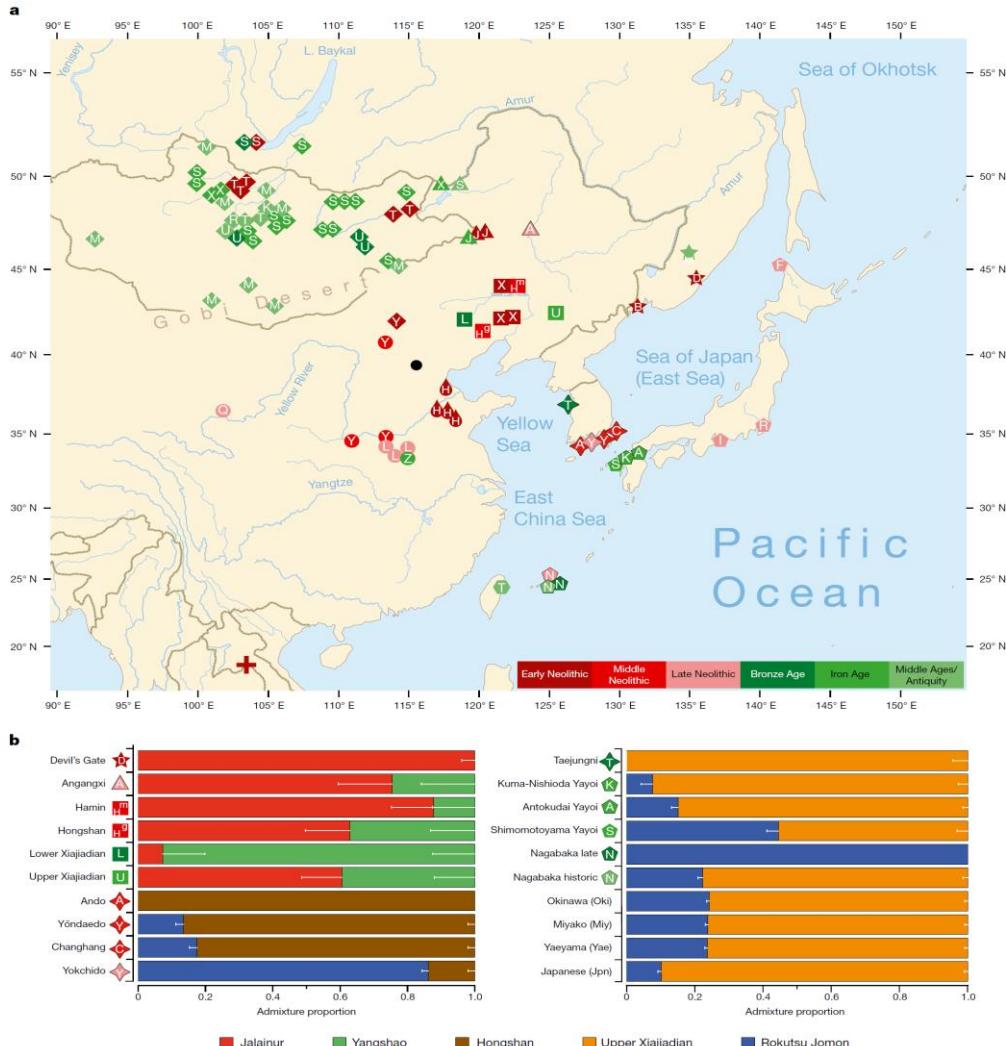


Fig. 3 | Spatiotemporal distribution and admixture of ancient genomes.
a, Ancient genomes located in time and space. For detailed legend, see Extended Data Fig. 4. b, QpAdm proximal admixture modelling of 20 key

ancient populations from this study. The x axis shows ancestry proportion estimates for the target populations in the y axis; the error bars represent ± 1 s.e.m. range, estimated by 5-cM block jackknifing.

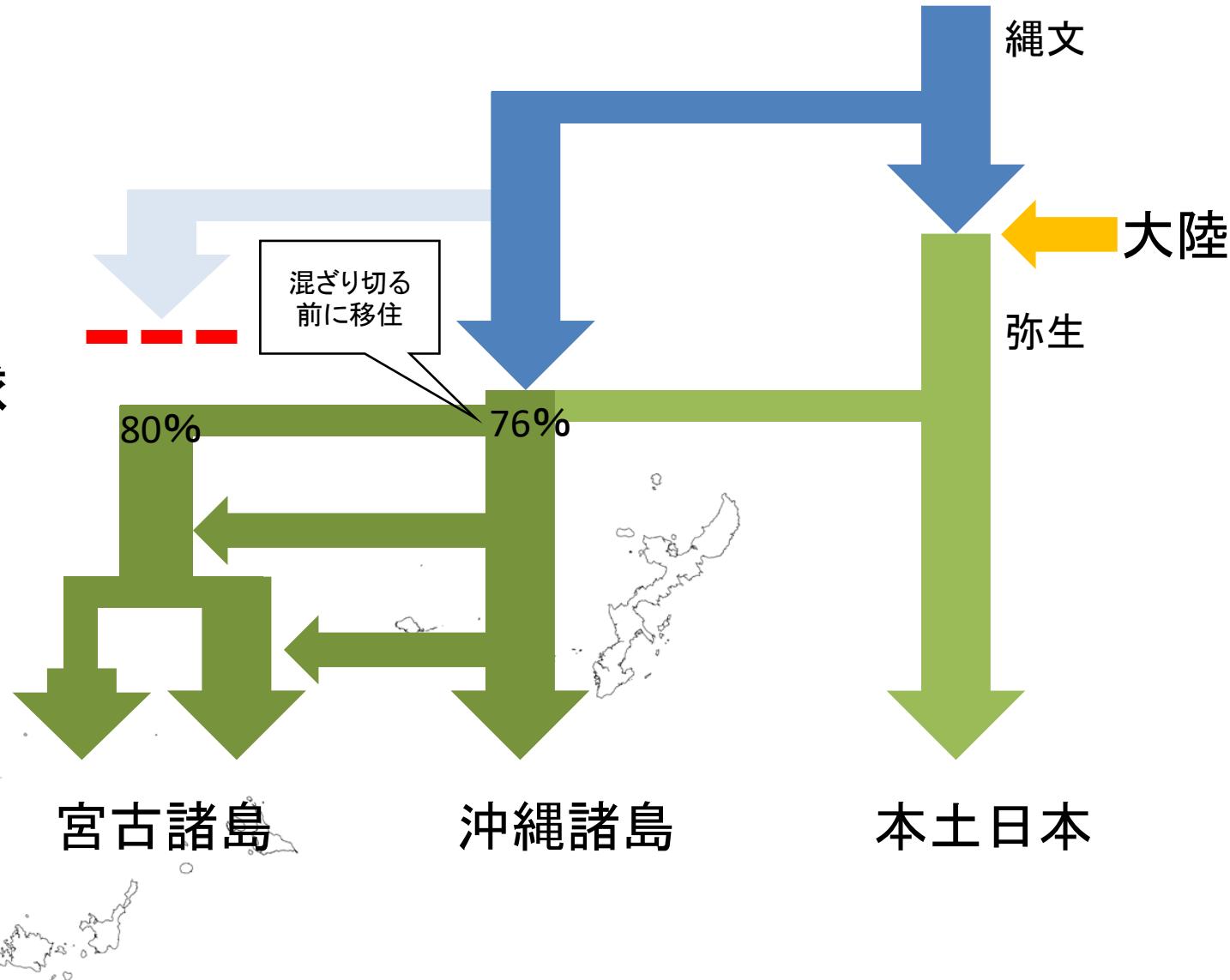
Robbeets et al. 2021

沖縄および宮古諸島集団の形成モデル

縄文・貝塚・
無土器時代

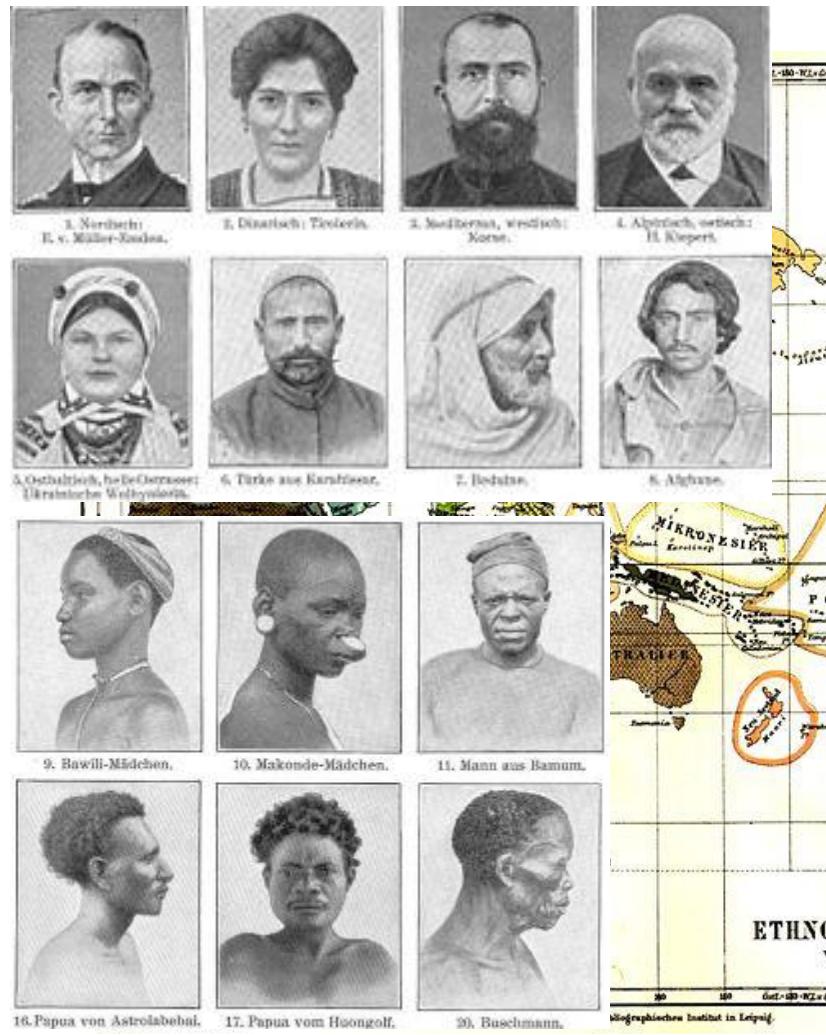
グスク・琉球
王朝時代

近代



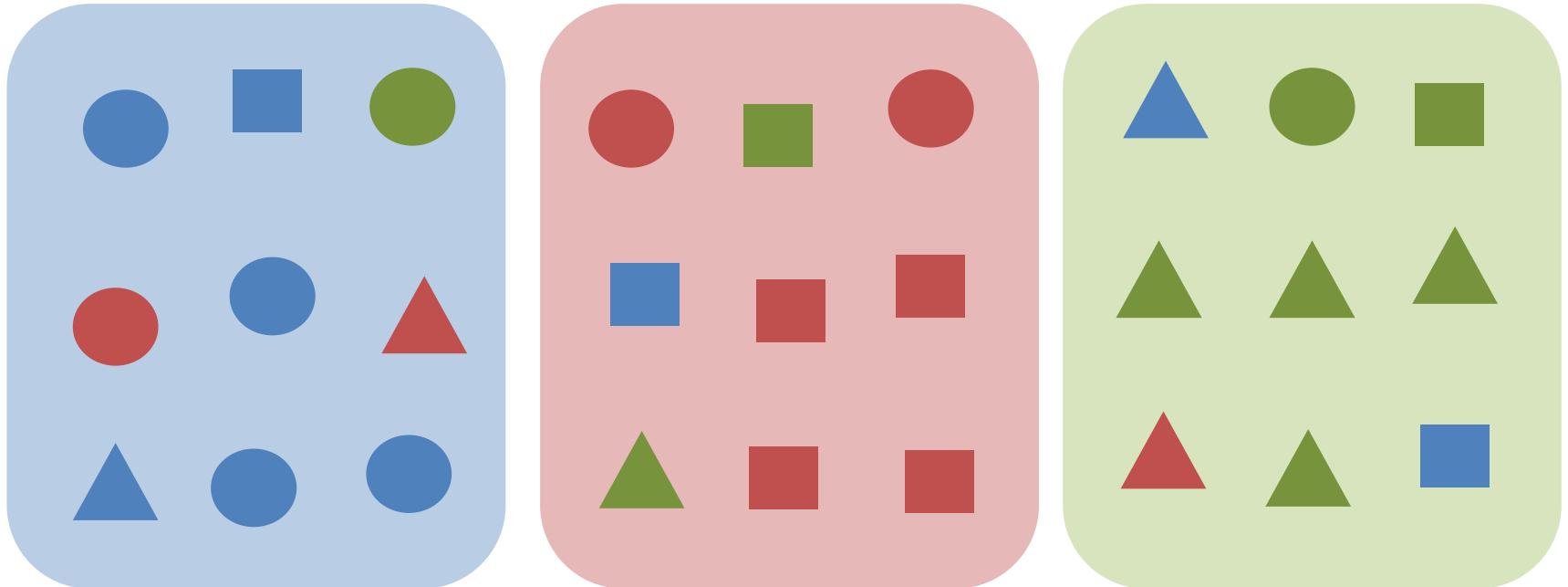
MENU

1. 集団遺伝学入門
2. ゲノム多様性を表現する
3. 琉球人集団の形成
4. ゲノム多様性と表現型多様性



ヒト集団間の形質のちがいを どう理解すべきか

形質の個体差と集団間差



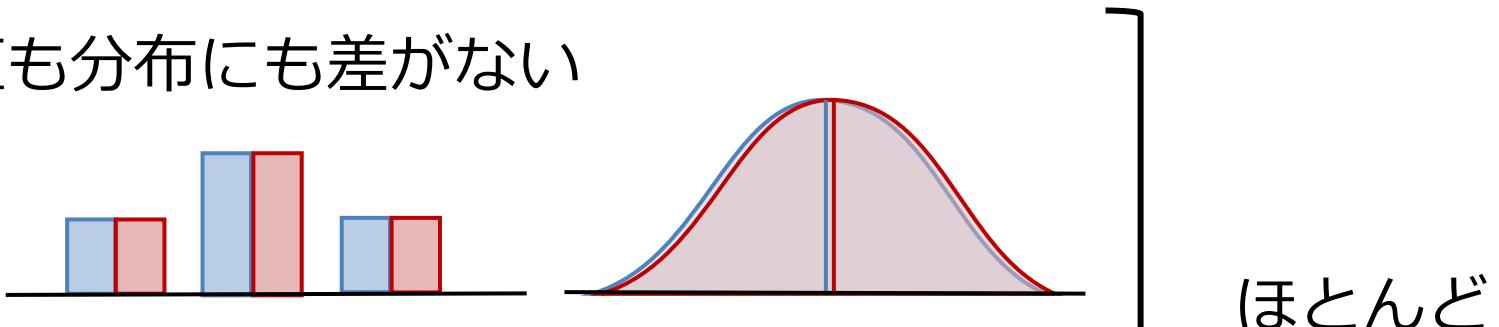
集団における“ステレオタイプ”



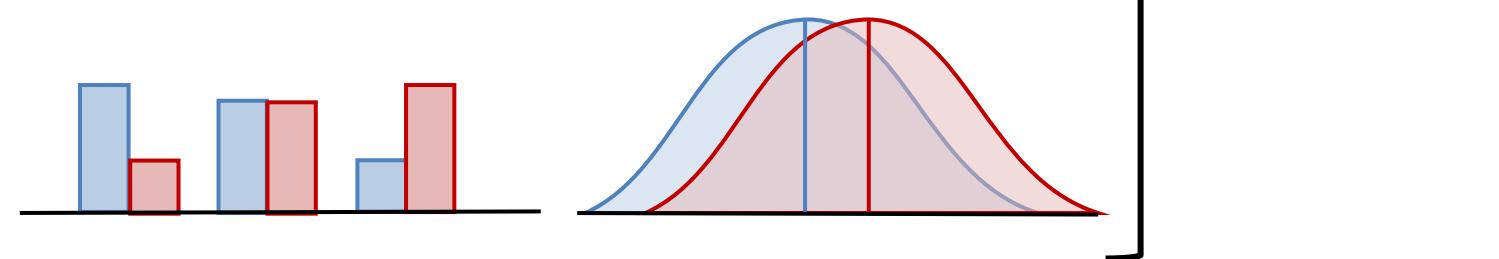
しかし、実際は集団内でも多型的なことがほとんど

形質の集団内および集団間多様性

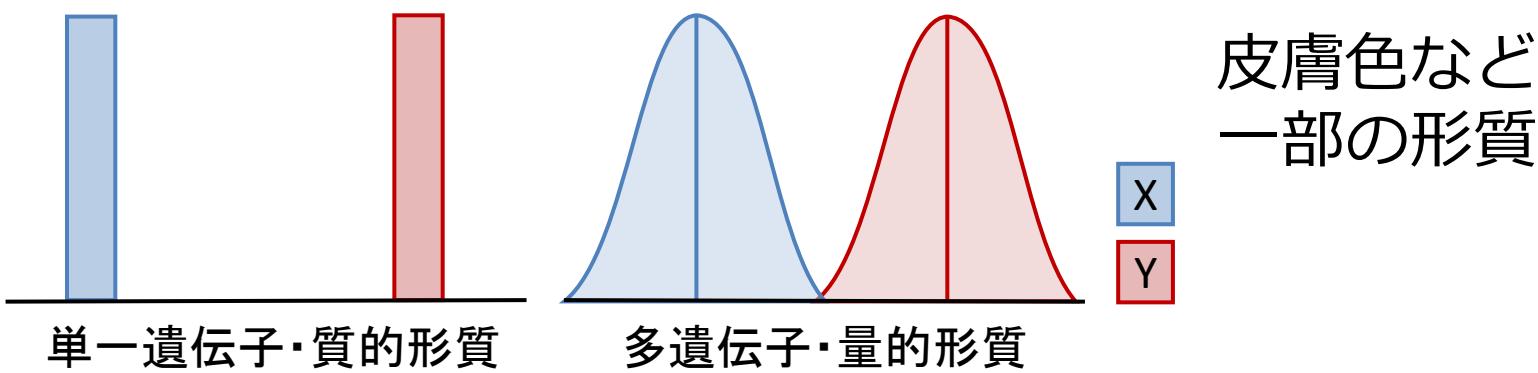
- ① 平均値も分布にも差がない



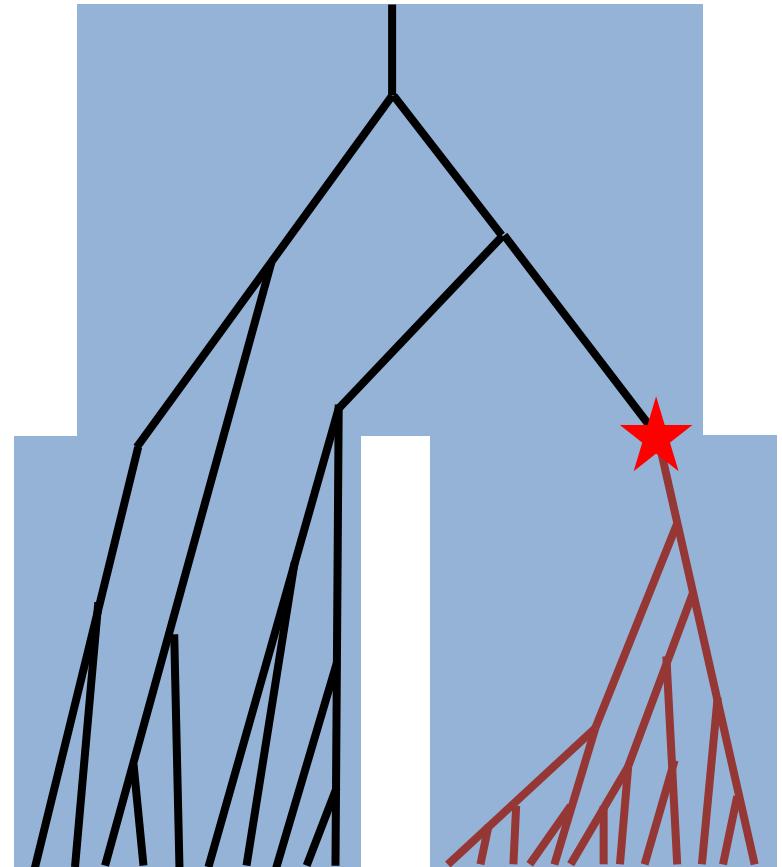
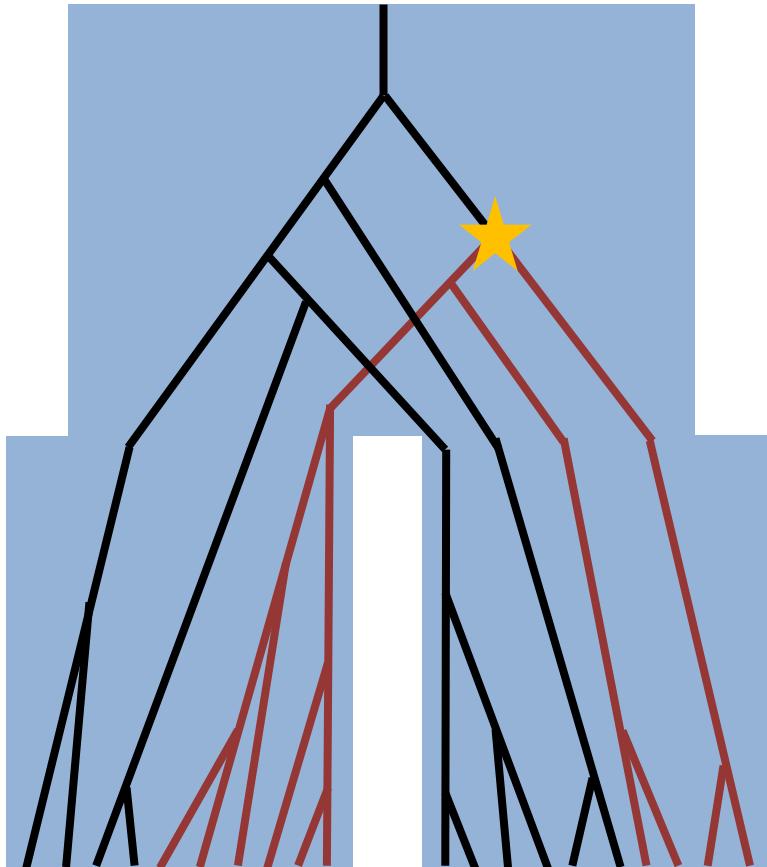
- ② 平均値に差、分布に重なり



- ③ 分布に重なりがほとんどない

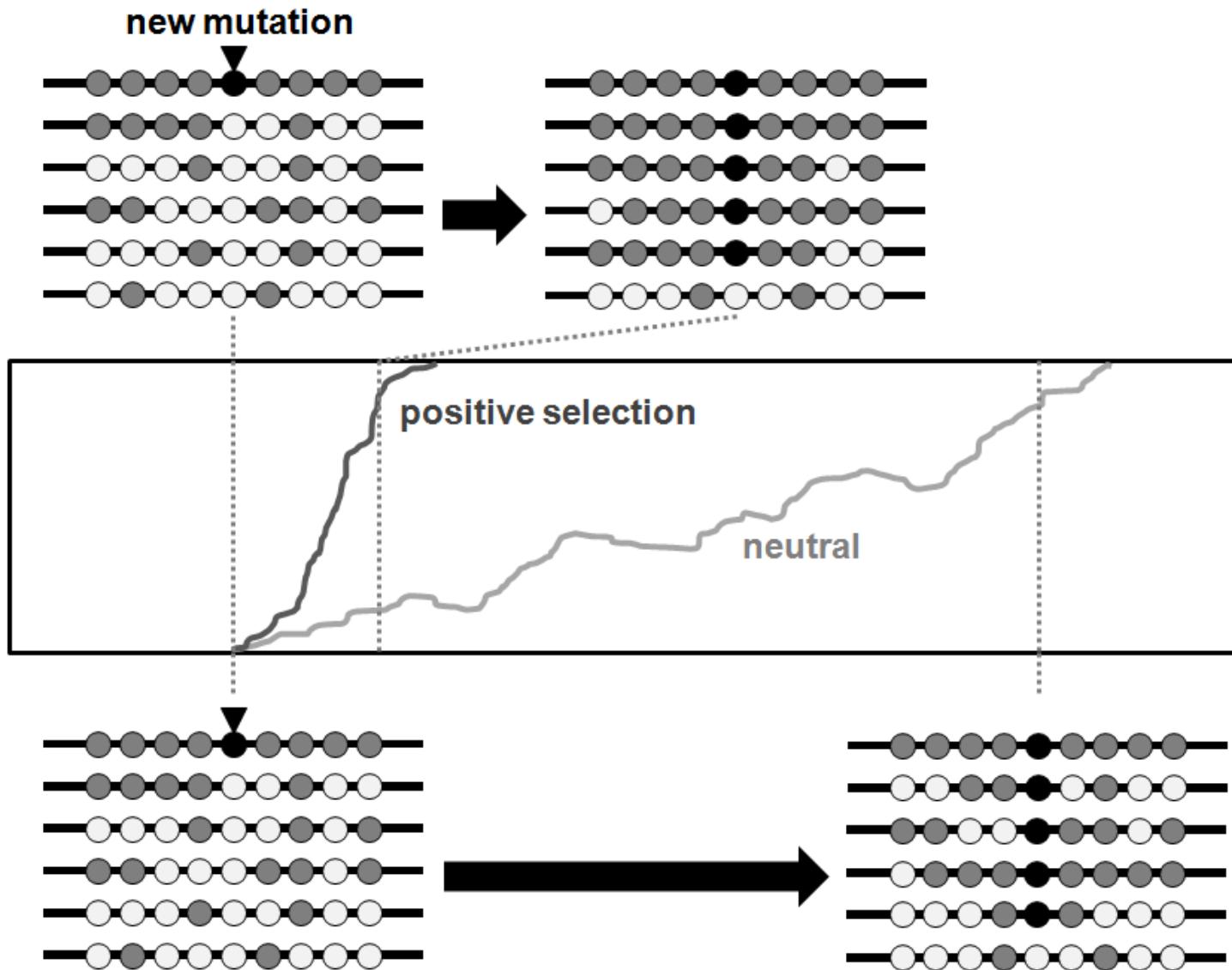


表現型を担う遺伝子の分化

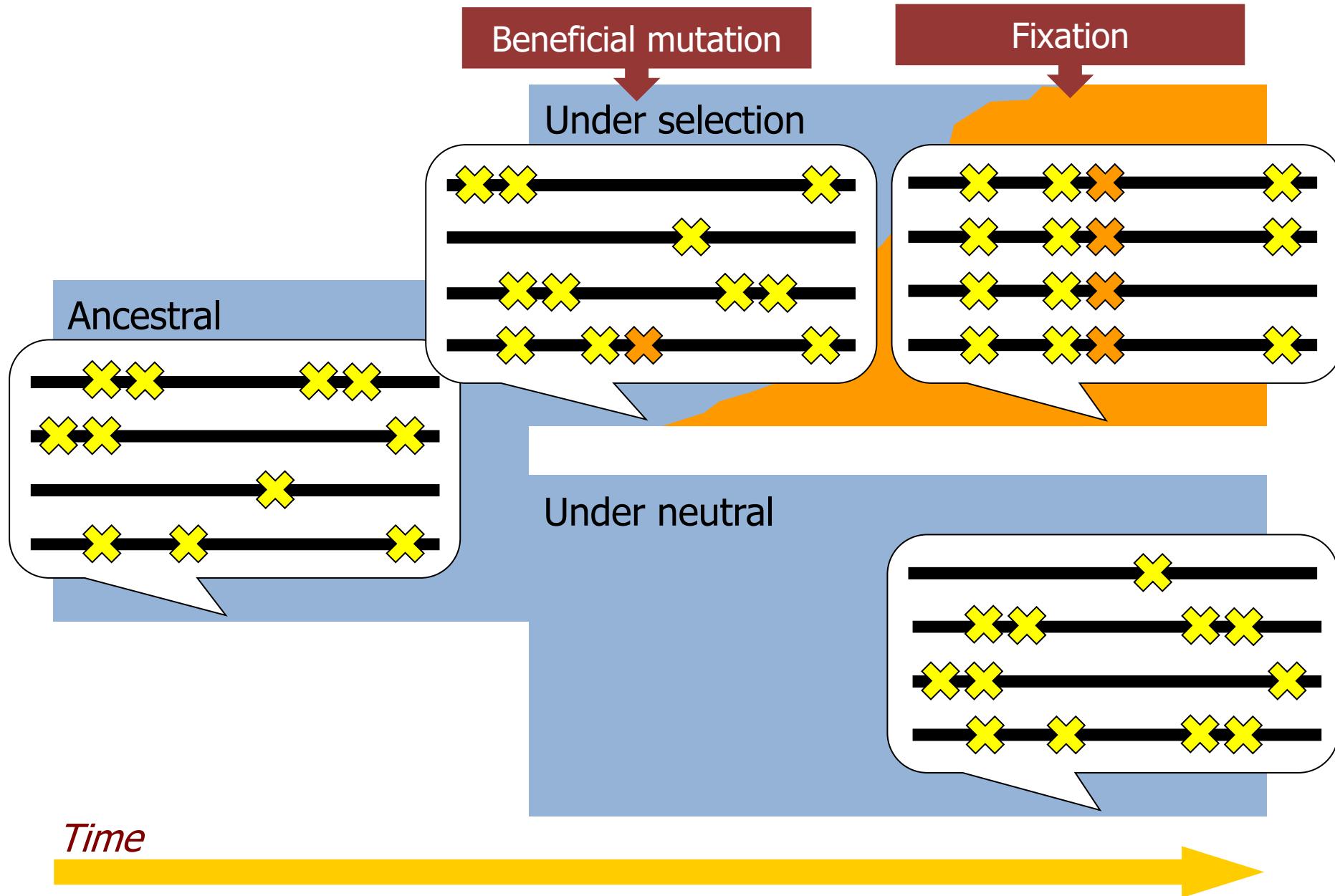


- ・分岐が浅いヒト集団間では、完全に分化した多型は非常に稀
- ・つまり、偶然の作用（遺伝的浮動）だけでは大きな分化は起こりにくい
- ・強い自然選択が働けば、大きな分化が起こる

ゲノム中にみられる自然選択の痕跡 selective sweep



集団特異的なselective sweep



集団特異的な強い自然選択の痕跡の具体例

Population	Gene	Chr	Function
Out-of-Africa	<i>KITLG</i>	12	色素形成
Europe	<i>SLC24A5</i>	15	色素形成
	<i>SLC45A2</i>	5	色素形成
	<i>LCT</i>	2	乳糖分解酵素
East Asia	<i>ABCC11</i>	16	アポクリン腺分泌
	<i>ADH1B</i>	4	アルコール代謝
	<i>EDAR</i>	2	外胚葉由来付属器 (毛髪・歯など) の発生

OPEN ACCESS Freely available online

PLOS BIOLOGY

A Map of Recent Positive Selection in the Human Genome

Benjamin F. Voight¹, Sridhar Kudaravalli¹, Xiaoquan Wen, Jonathan K. Pritchard^{*}

Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America

OPEN ACCESS Freely available online

PLOS ONE

A Practical Genome Scan for Population-Specific Strong Selective Sweeps That Have Reached Fixation

Ryosuke Kimura^{1,2*}, Akiohiro Fujimoto¹, Katsushi Tokunaga¹, Jun Ohashi¹

¹ Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan, ² Japan Society for the Promotion of Science, Tokyo, Japan

OPEN ACCESS Freely available online

PLOS BIOLOGY

A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome

Kun Tang¹, Kevin R. Thornton², Mark Stoneking¹

¹ Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, ² Department of Ecology and Evolutionary Biology, University of California Irvine, Irvine, California, United States of America

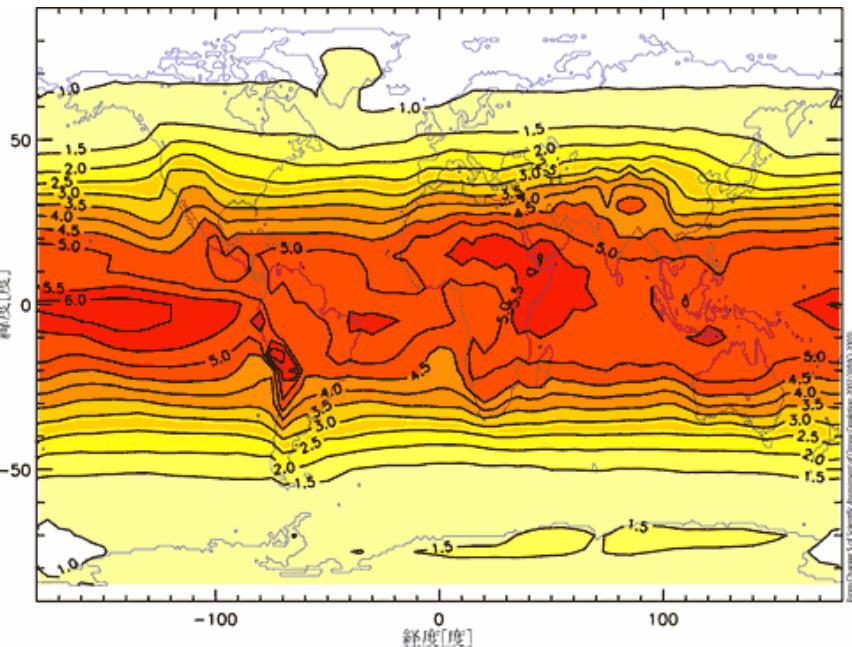
Genome-wide detection and characterization of positive selection in human populations

Paridis C. Sabeti^{1,2*}, Patrick Varilly^{1*}, Ben Fry¹, Jason Lohmueller¹, Elizabeth Hostetter¹, Chris Cotsapas^{1,2}, Xiaohui Xie¹, Elizabeth H. Byrne¹, Steven A. McCarroll^{1,2}, Rachelle Gaudet¹, Stephen F. Schaffner¹, Eric S. Lander^{1,4,5,6}& The International HapMap Consortium¹

Voight et al. 2006
Kimura et al. 2007
Tang et al. 2007
Sabeti et al. 2007

紫外線量と色素形成能

紫外線量



皮膚色分布

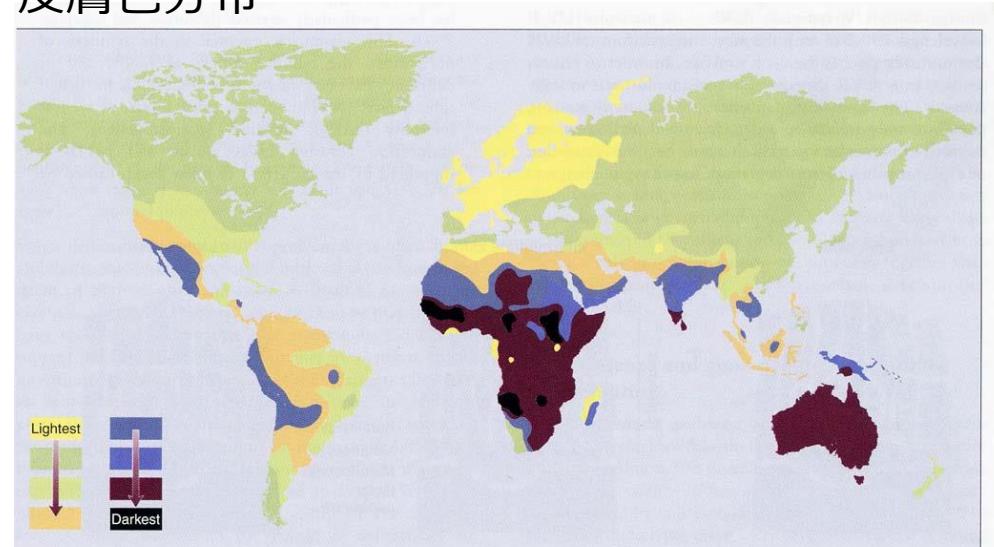


Figure 13.2: Distribution of different skin colors in indigenous populations of the world.

Redrawn from Jurmain *et al.* (2000), after Biasutti (1959).

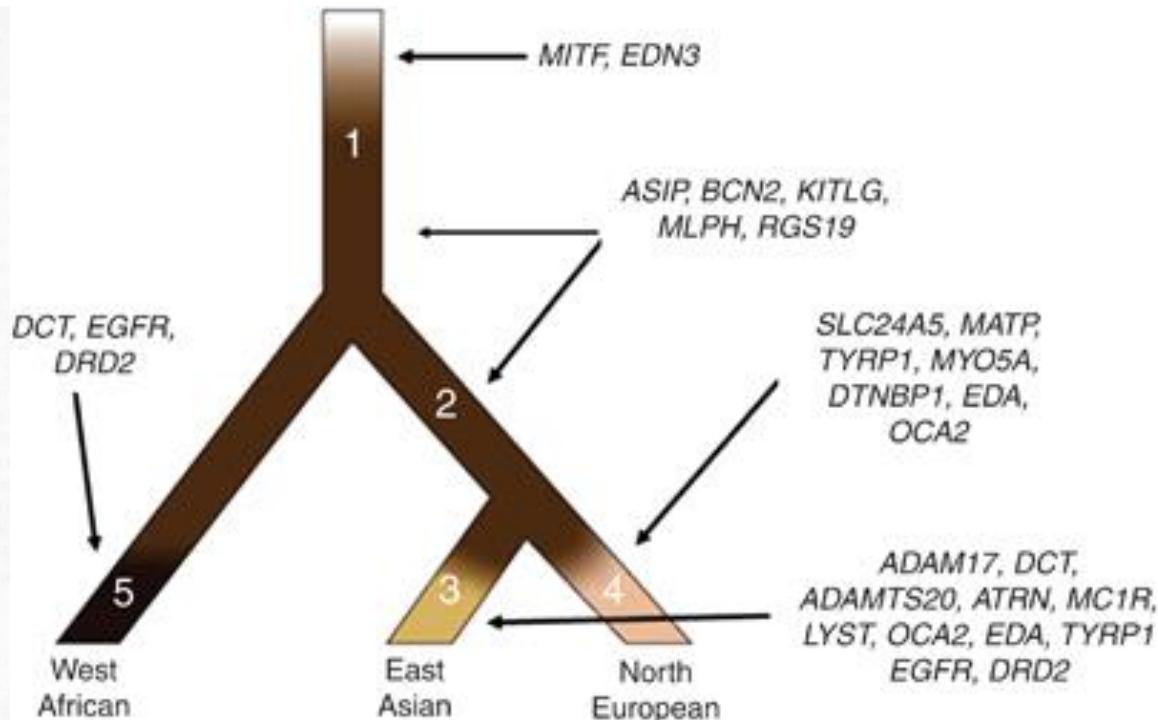
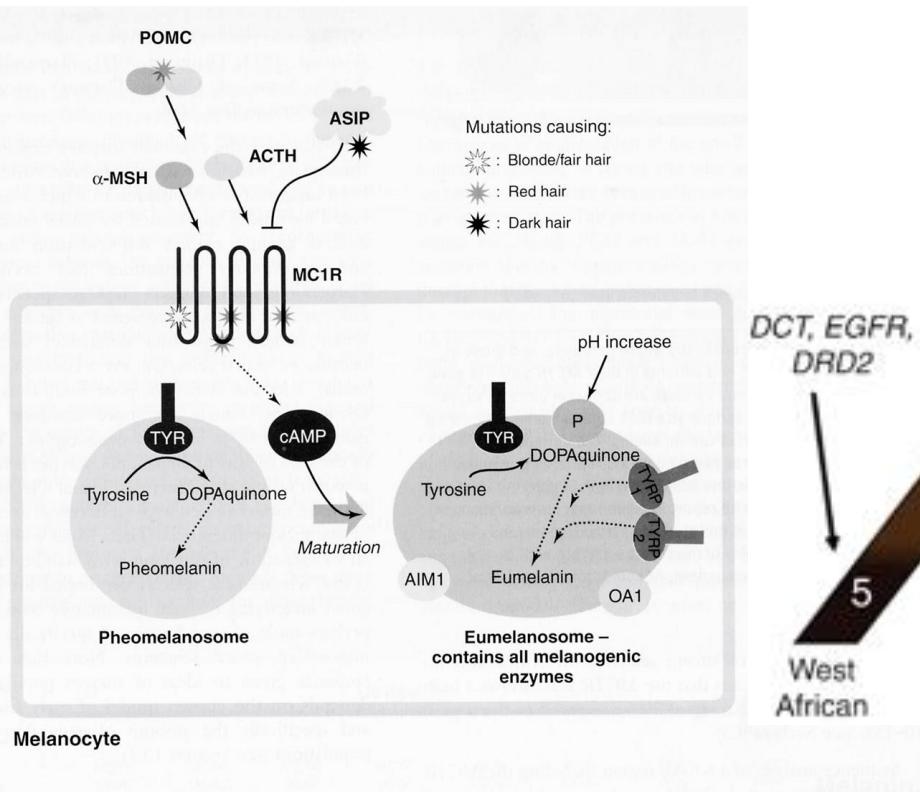


Photography by FENTY BEAUTY



<http://www.vsh.vision/blog/eye-color-trivia/>

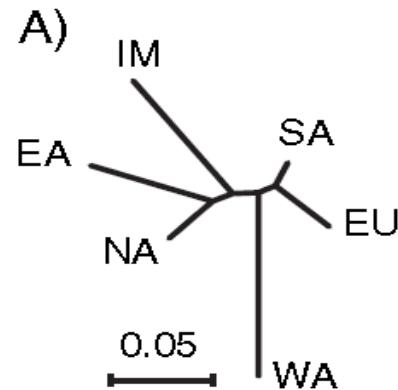
色素形成関連遺伝子変異が生じた時期



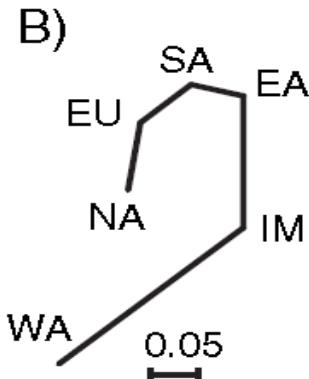
- 低緯度地域では強い紫外線による日焼けや発がんのリスクを低減させるため、色素が濃い (and/or 紫外線による葉酸の破壊：葉酸仮説)
- 高緯度地域では強い紫外線による制約がはずれるだけでなく、色素が薄くなるように強い選択圧が働いた (ビタミンD仮説・性選択仮説)

色素形成関連遺伝子多型頻度による近縁性

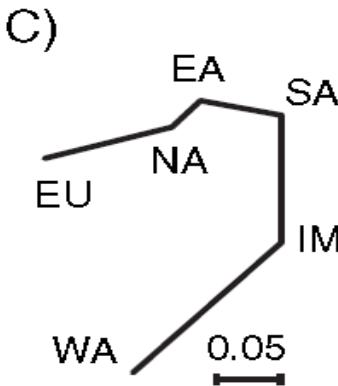
Genome Ave



ASIP



OCA2



WA: West African

SA: South Asian

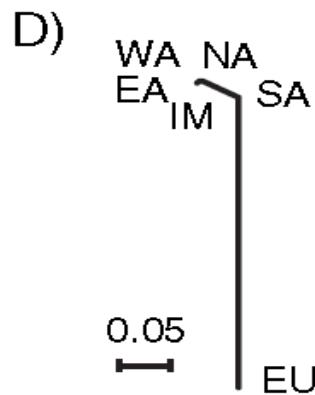
NA: Native American

EU: European

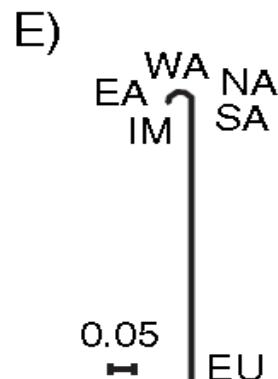
EA: East Asian

IM: Island Melanesian

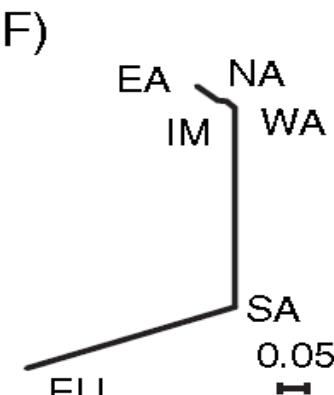
Norton et al. 2007



TYR



MATP
(*SLC45A2*)



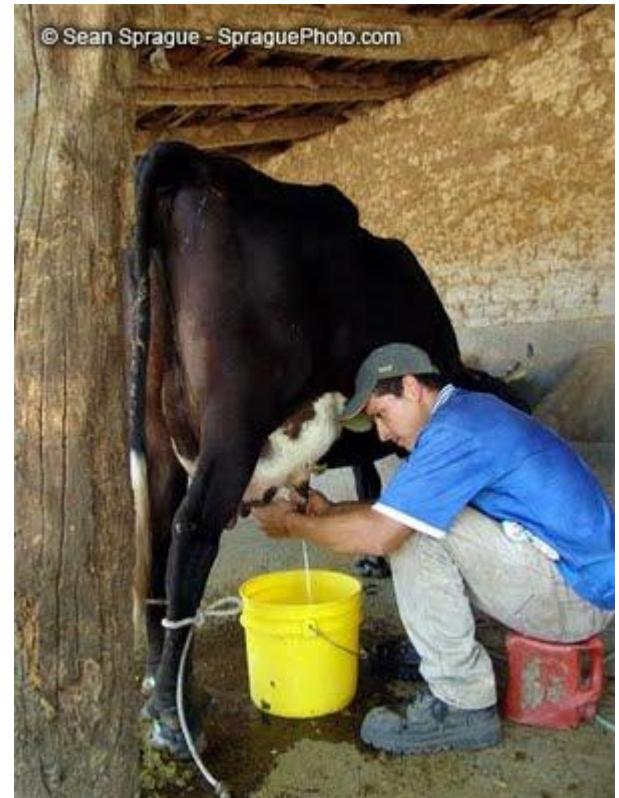
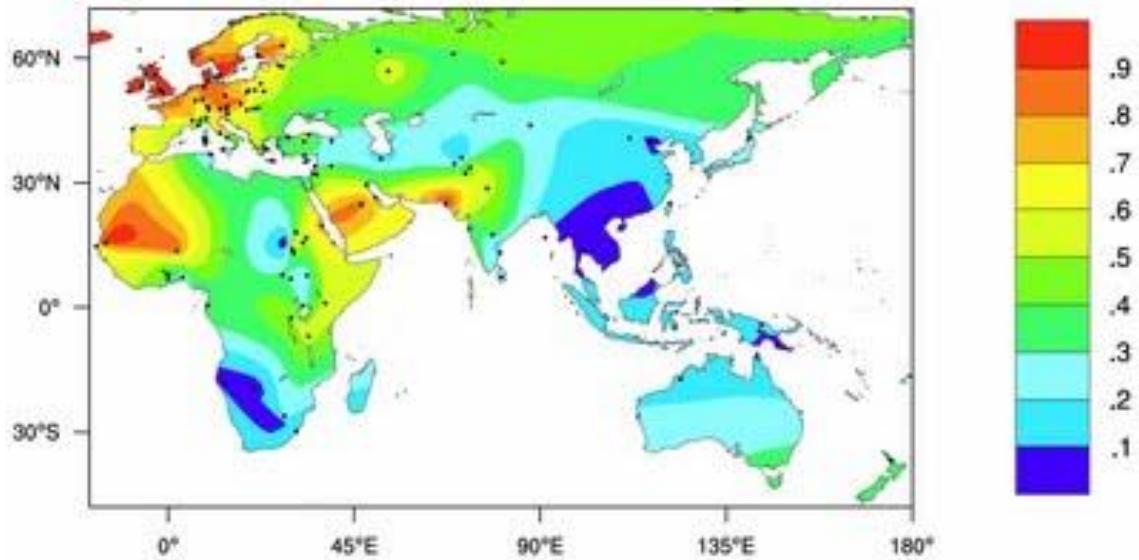
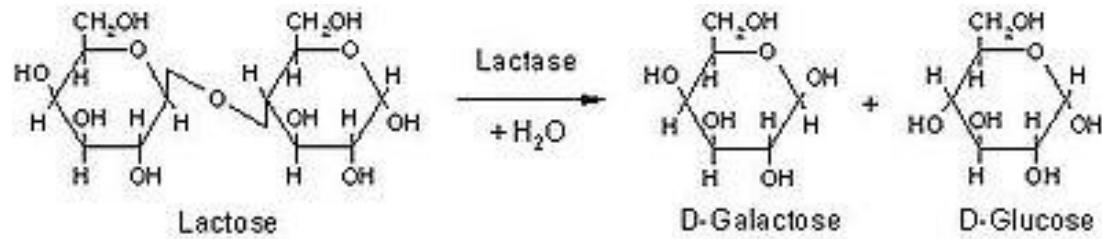
SLC24A5

- ・ 色素形成関連遺伝子多型の分化は表現型の分化をよく説明する
(集団間の遺伝距離のゲノム平均とは大きく異なる)

乳糖耐性

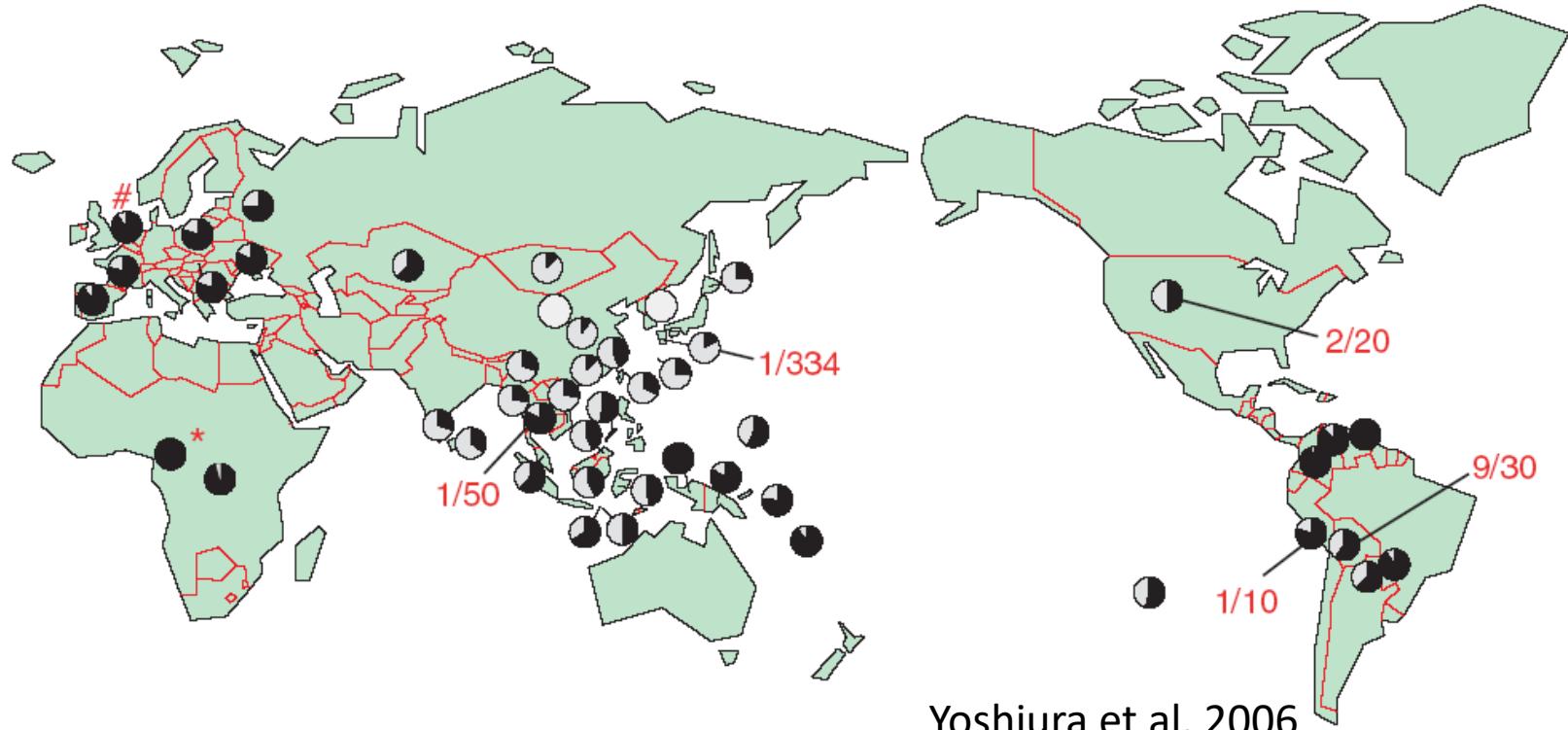
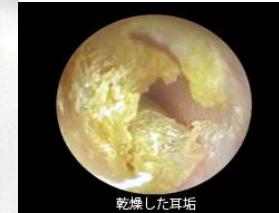
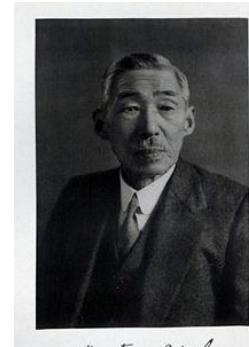
通常、哺乳類は乳児期を過ぎると乳糖分解酵素（Lactase）の活性が低下する。ヨーロッパやアフリカにおける酪農が盛んな地域では、成長しても活性が落ちないタイプの変異が生じ、選択されてきた（LCT -13910C/T）

遺伝子と文化の共進化

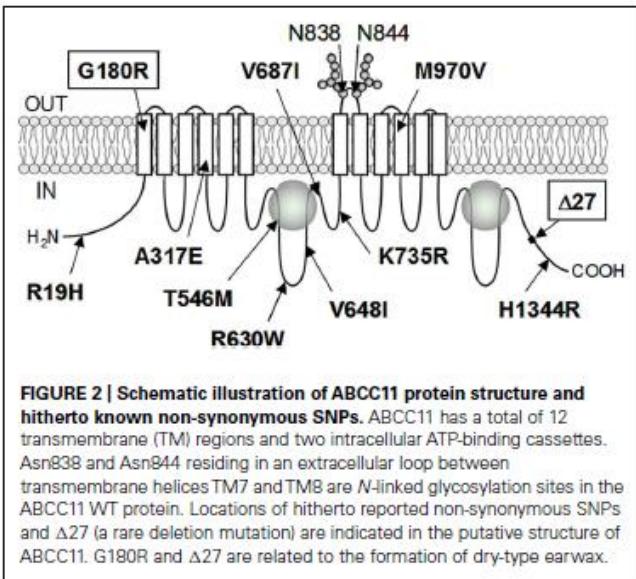


耳垢 Earwax (cerumen)

- ・耳垢の乾湿は単一遺伝形質で湿型が優性
- ・東アジアでは乾型の頻度が高い
- ・アポクリン腺分泌にかかわっており、体臭とも関連
- ・*ABCC11* Gly180Argが主な原因



ABCC11の機能とGly180Arg多型



Ishikawa et al. 2013

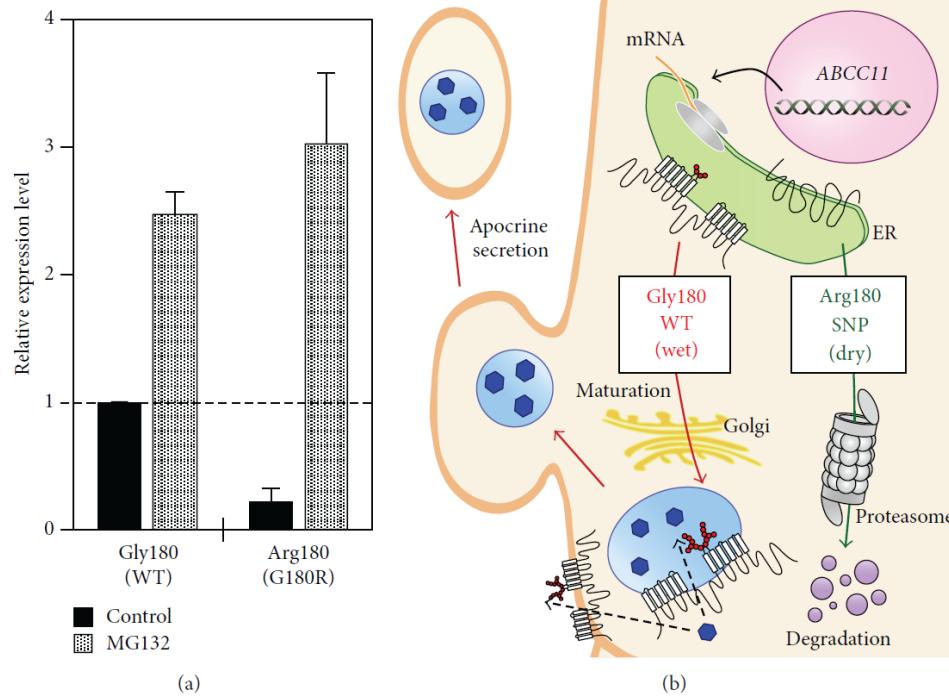


FIGURE 3: Effects of the SNP variant (Arg180) on the protein level and intracellular degradation of ABCC11. (a) To assess the effect of the Arg180 variant on the protein level of ABCC11, Flp-In-293 cells expressing the WT or Arg180 variant of ABCC11 were cultivated in the presence of MG132, a proteasome inhibitor, for 24 h. ABCC11 WT and Arg180 variant proteins were analyzed by immunoblotting with ABCC11-specific antibody after treatment with PNGase F, a glycosidase. The signal intensity ratio (ABCC11/GAPDH, internal control) was normalized to the control and expressed as mean ± SD. (b) Schematic illustration of the posttranslational modification of ABCC11 WT and proteasomal degradation of the Arg180 variant.

アポクリン腺分泌で働く

Toyoda et al. 2016

ABCC11の基質：匂い物質の前駆体

Body odor precursors

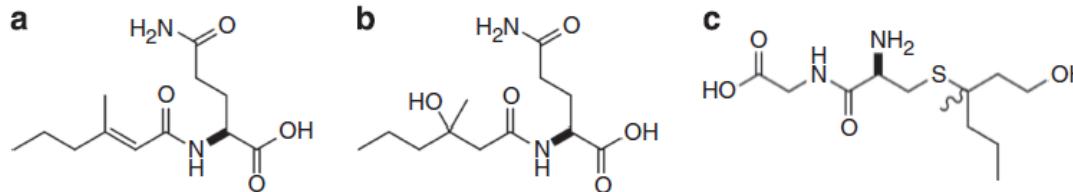


Figure 1. Structures of the key known human body odor precursors. (a) N^{α} -3-methyl-2-hexenyl-glutamine (3M2H-Gln). (b) N^{α} -3-methyl-3-hydroxy-hexanoyl-glutamine (HMHA-Gln). (c) Cys-Gly conjugate of 3-methyl-3-sulfanyl-hexanol (3M3SH).

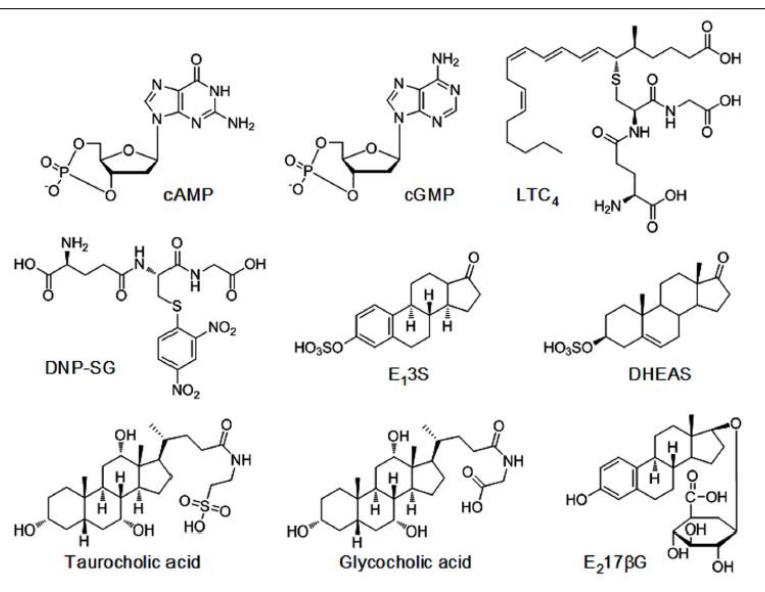


FIGURE A1 | Chemical structures of ABCC11 substrates. Cyclic adenosine monophosphate (cAMP), cyclic guanosine monophosphate (cGMP), leukotriene C₄ (LTC₄), S-(2,4-dinitrophenyl)-glutathione (DNP-SG), estrone 3-sulfate (E_{13S}), dehydroepiandrosterone 3-sulfate (DHEAS), and E_{217βG} (estradiol 17-β-D-glucuronide).

Table 2. Enzyme-released odorants in sweat samples of panelists with different genotypes, determined by GC × GC-ToF MS

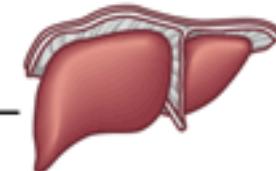
Odorants

Identifier	Structure	Name	$\mu\text{g analyte}/2 \text{ pads}$ (median per genotype)					M-W U test (P-values) GG and AG versus AA
			AA	AG	GG	GG and AG		
A2		3-hydroxy-3-methylhexanoic acid methyl ester	0.14	44.70	45.82	45.26		0.0000
B3		3-hydroxy-4-methyloctanoic acid methyl ester	0.00	1.84	0.69	0.75		0.0000
E1		(E)-3-methyl-2-hexenoic acid methyl ester	0.02	>>2.8 ¹	>>2.8	>>2.8		0.0000
C1		3-hydroxyhexanoic acid methyl ester	0.03	0.08	0.13	0.12		0.035
C3		3-hydroxyoctanoic acid methyl ester	0.09	1.29	1.98	1.77		0.0001

GC × GC-ToF MS, two-dimensional GC coupled with time-of-flight mass spectrometry.

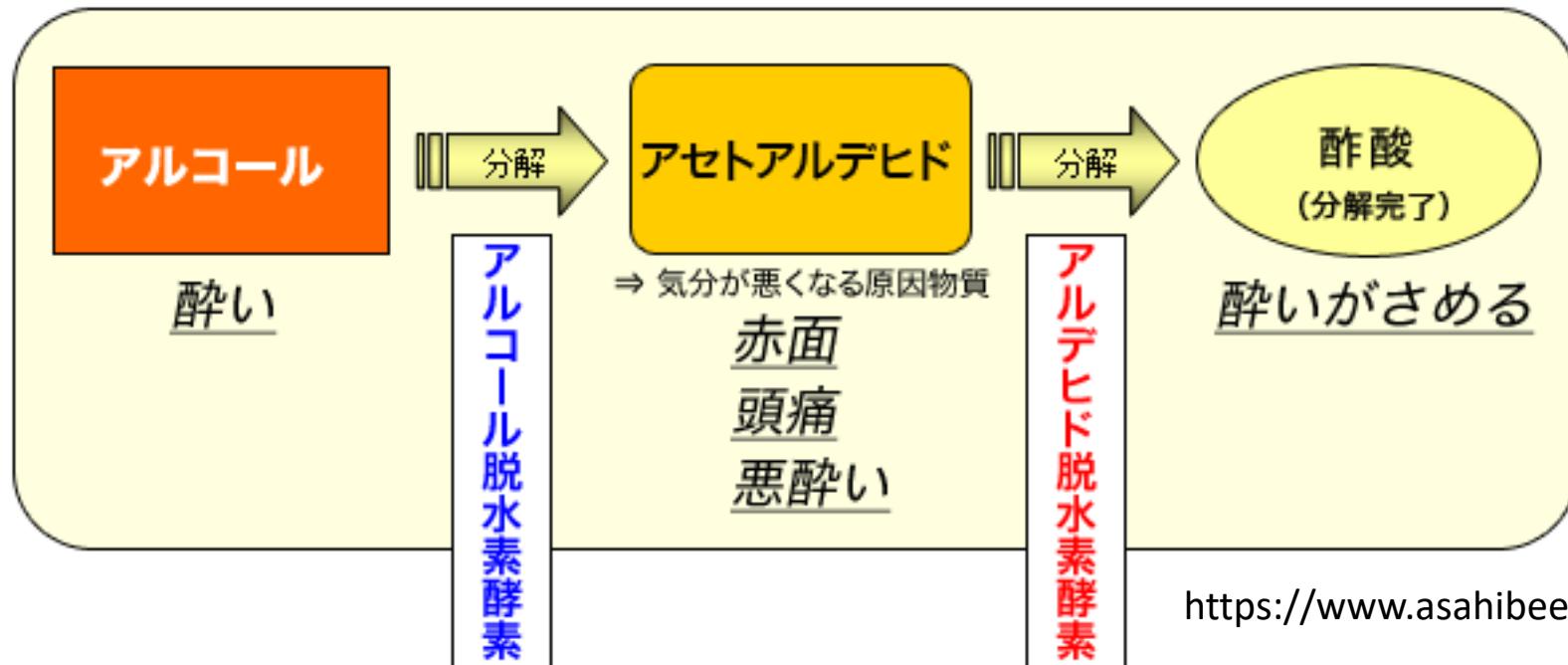
¹This analyte was above the linear range for most of the AG and GG panelists.

アルコール不耐性



アルコールは 主に肝臓で 2段階に分解されます。

「アルコール脱水素酵素」と「アルデヒド脱水素酵素」の2つが作用。



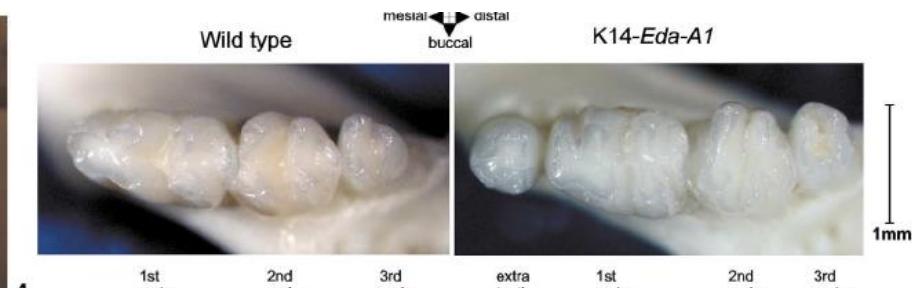
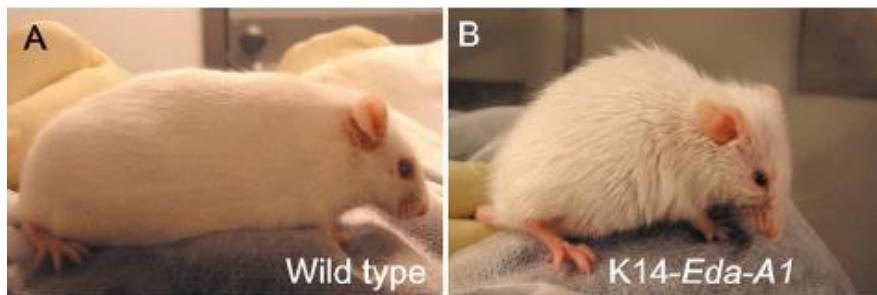
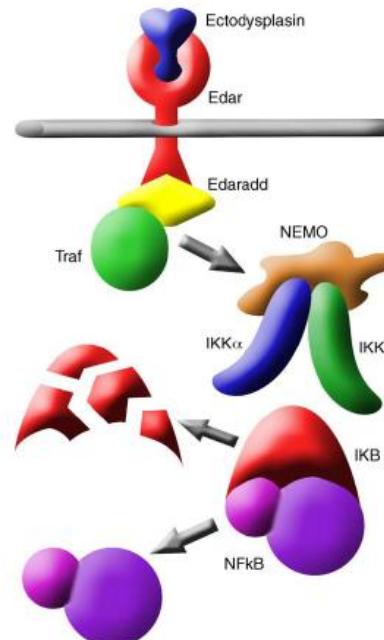
ADH1B His48Arg
祖先型/派生型（アジア特異的）
低活性型/高活性型
日本での頻度約20%/約80%

ALDH2 Glu504Lys
祖先型/派生型（アジア特異的）
活性型/不活性型
日本での頻度約80%/約20%

選択圧は未だ不明。アセトアルデヒド濃度上昇 ⇒ 感染症抵抗性？

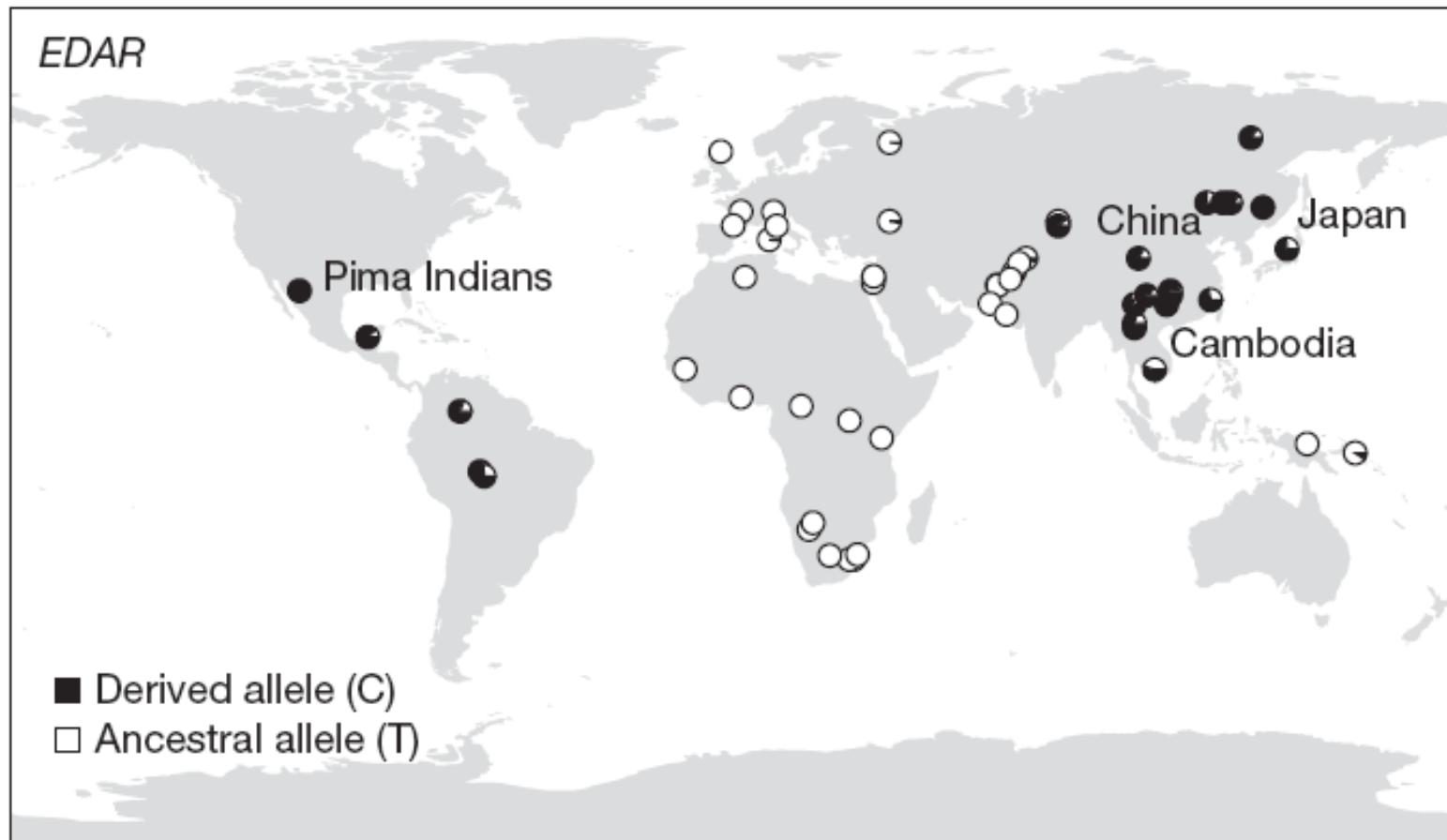
Ectodysplasin A receptor (EDAR)

- Ectodysplasin A (EDA) の受容体
- 外胚葉由来器官の正常な発生に必要
- ヒト*EDAR*：
Ectodermal dysplasia
(外胚葉異形成症)
の原因遺伝子のひとつ
- EDA, EDAR過剰発現マウス：
毛髪、歯、爪、汗腺、皮脂腺形成異常



Mustonen et al. 2003

EDAR 1540T/C (370Val/Ala)の分布



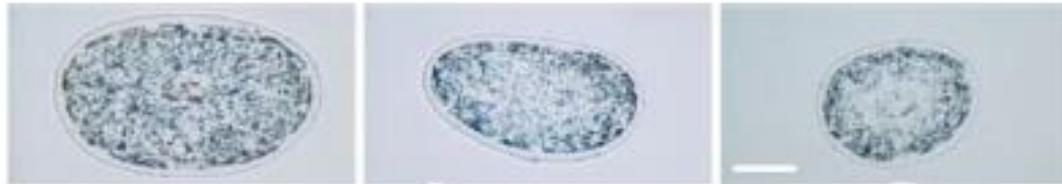
Sabeti et al. 2007

非同義変異370Alaは東ユーラシアおよびアメリカに特異的に観察される

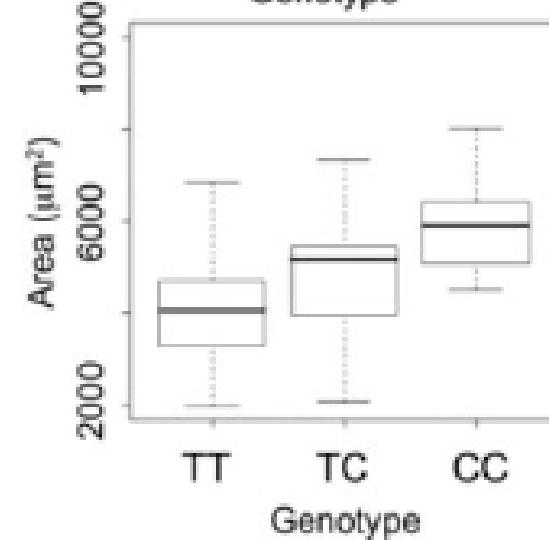
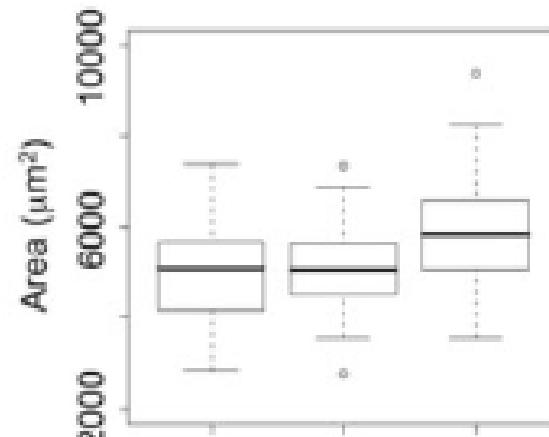
EDAR 多型と毛髪の太さとの関連

A scan for genetic determinants of human hair morphology: *EDAR* is associated with Asian hair thickness

Akihiro Fujimoto¹, Ryosuke Kimura^{5,*}, Jun Ohashi¹, Kazuya Omi¹, Rika Yuliwulandari^{1,6}, Lilian Batubara⁶, Mohammad Syamsul Mustofa⁷, Urai Samakkarn⁸, Wannapa Settheetham-Ishida⁹, Takafumi Ishida², Yasuyuki Morishita³, Takuro Furusawa⁴, Minato Nakazawa¹⁰, Ryutaro Ohtsuka¹¹ and Katsushi Tokunaga¹

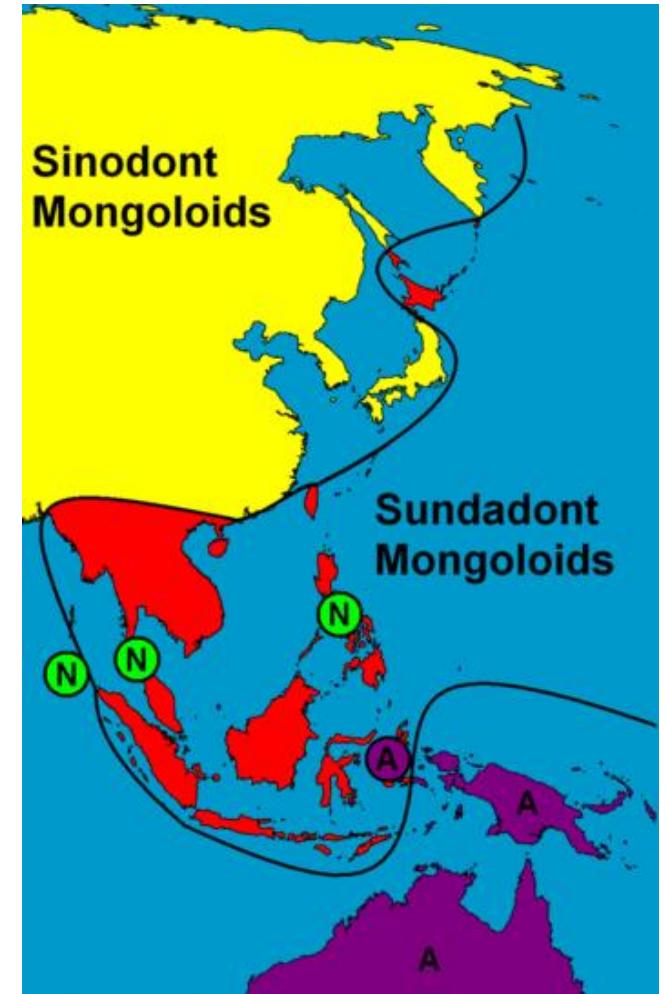
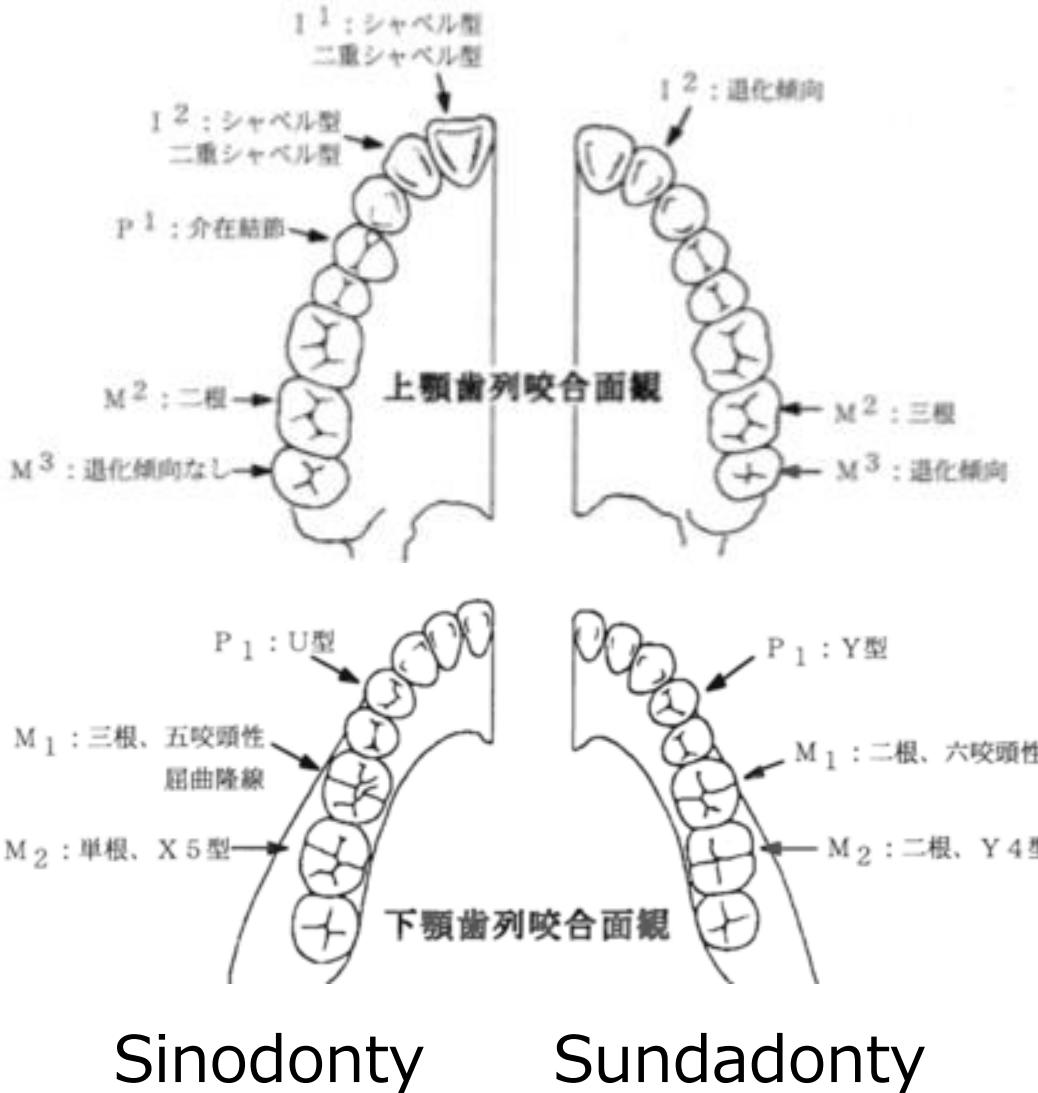


アジア型アリルは太い毛髪と関連



Fujimoto et al. 2008

アジアにおける歯冠形態の多様性 -Sinodonty and Sundadonty-



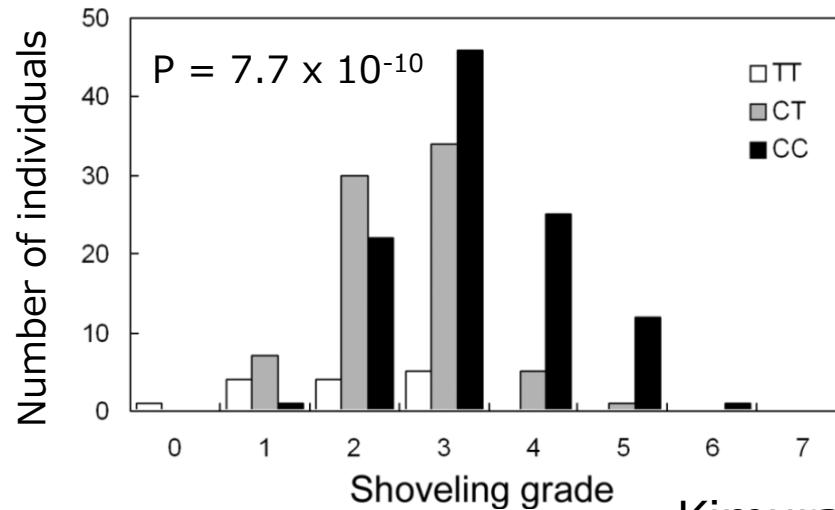
Sinodonty

Sundadonty

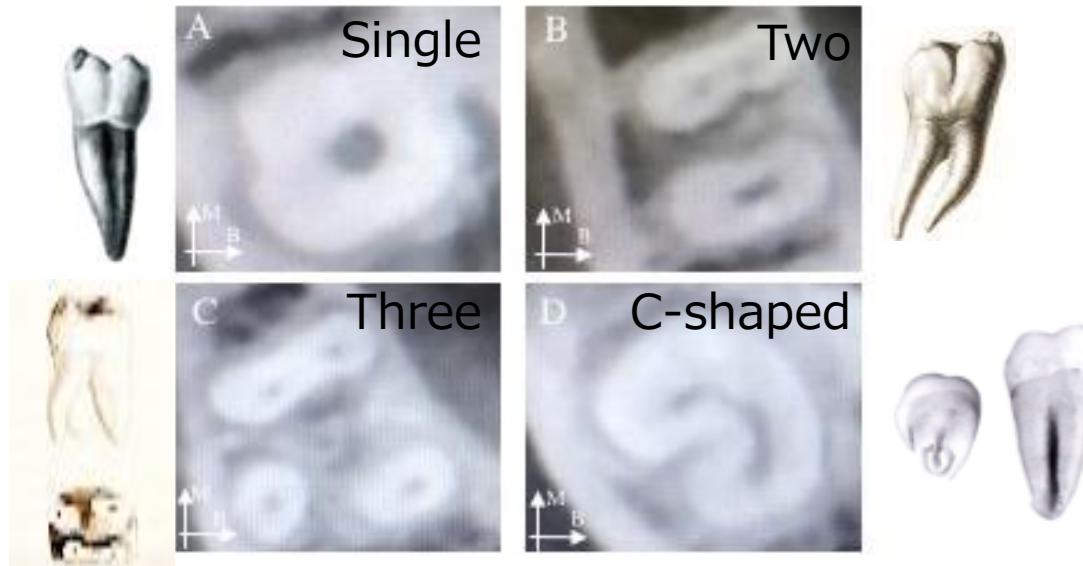
Howells (1997)

EDAR 多型と歯冠および歯根形態との関連

上顎中切歯

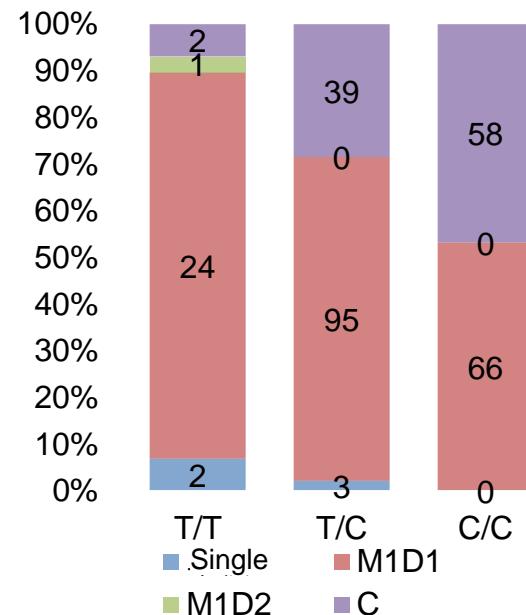


下顎第二大臼歯



アジア型アリルはシノドントと関連

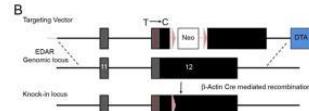
Kimura et al. 2009



Kataoka et al. in prep.

ヒトEDARアリルをマウスにノックイン： アジア型アリルは高密度な汗腺や乳腺のとも関連

A
 Human RMLSSTYNESEKAVVKTWRHLAESGLKRDEIGGMDT
 Mouse RMLSSTYNESEKAVVKTWRHLAESGLKRDEIGGMDT
 Human GMQLFDRISTAGYSIPELLTKLVQIERDAVESLCADIL
 Mouse GMQLFDRISTAGYSIPELLTKLVQIERDAVESLCADIL



C
 A G C G G T C G T G A G C G G T C G T G A G C G G T C G T G

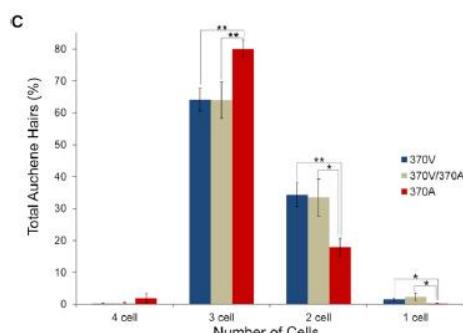
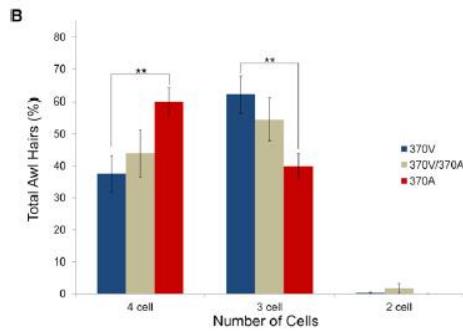
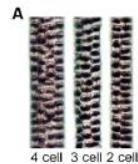


Figure 3. 370A Allele Increases Hair Size in the Mouse Coat

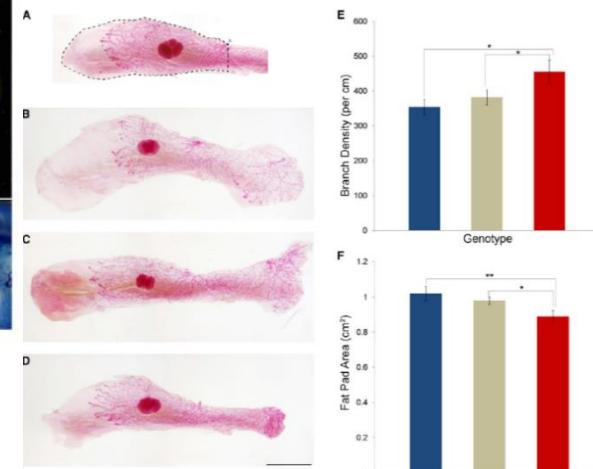
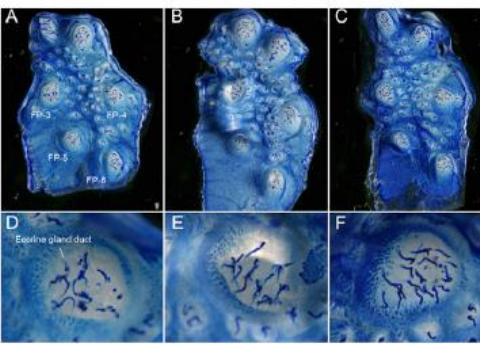


Figure 5. 370A Reduces the Size of the Mammary Fat Pad and Increases Mammary Gland Branch Density

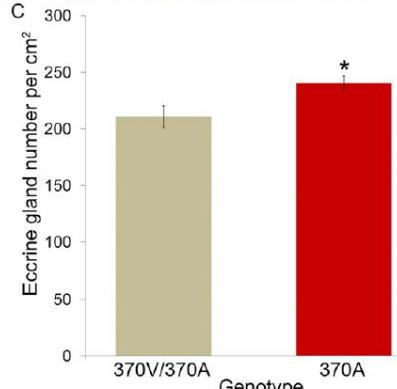
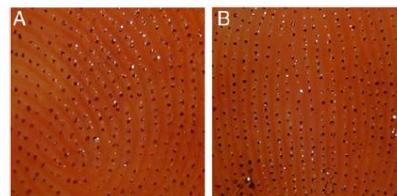
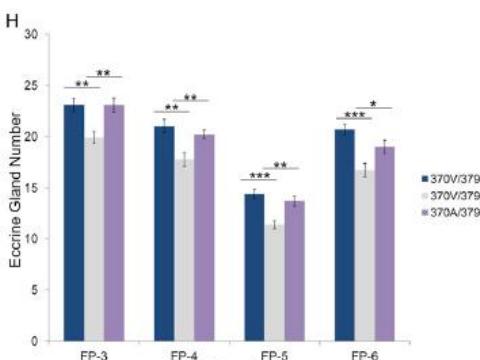
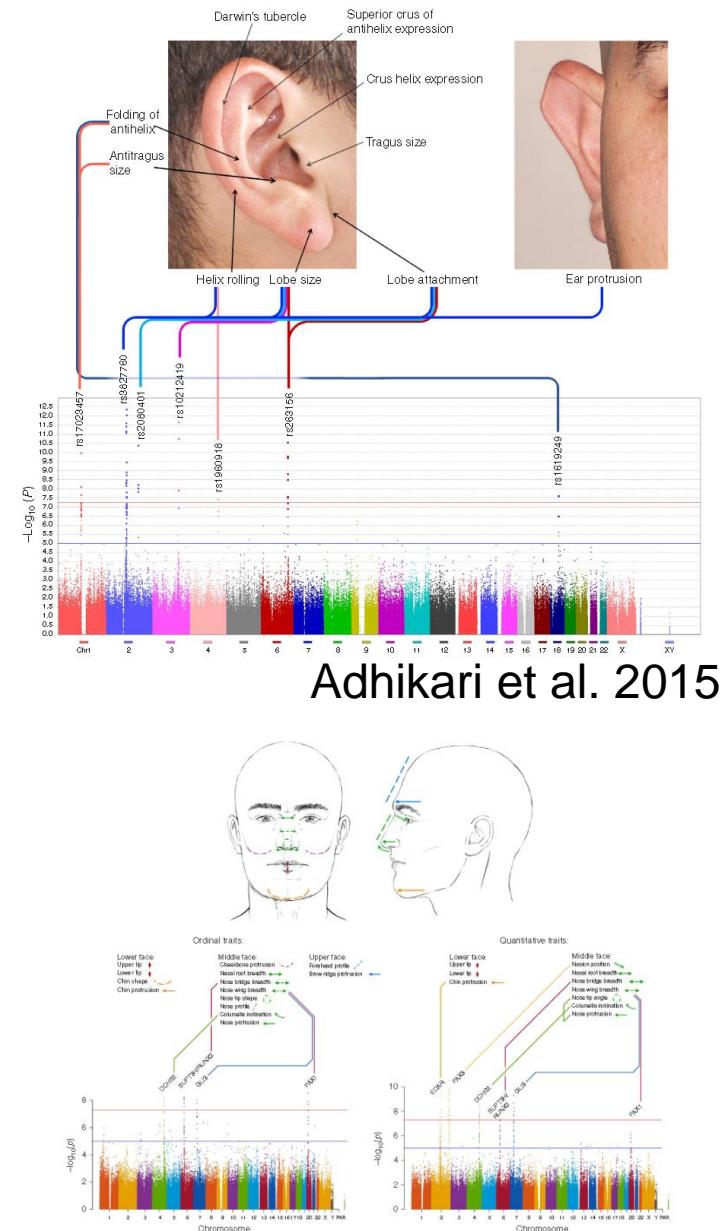
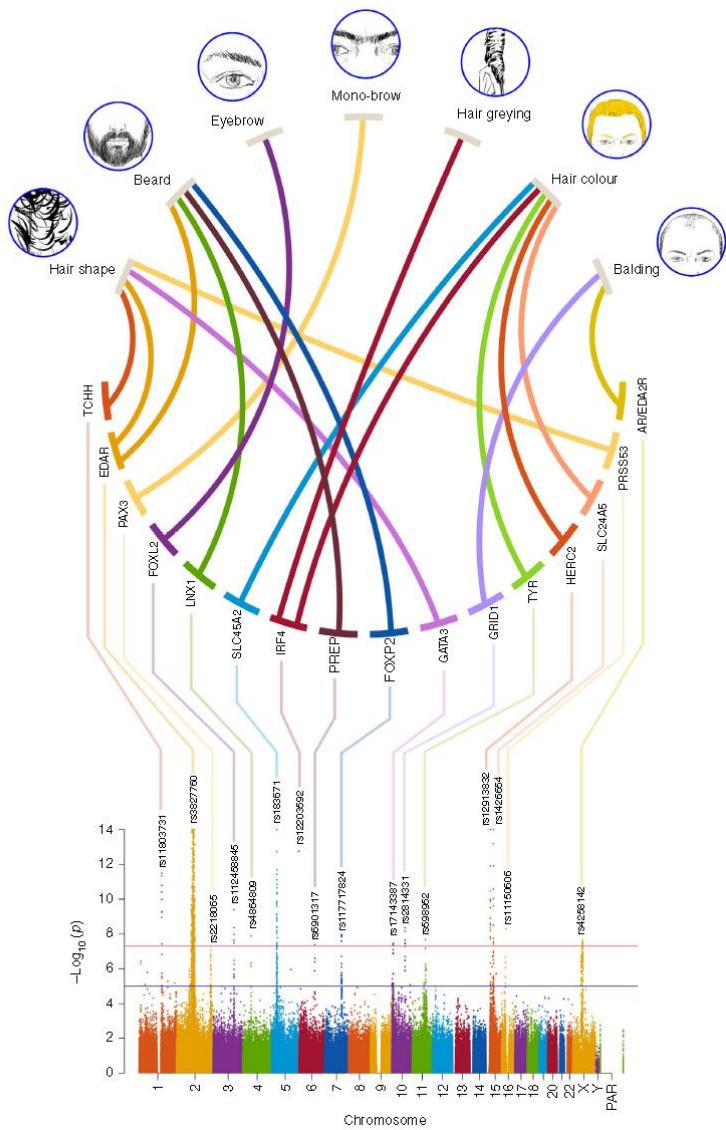


Figure 6. 370A Increases the Number of Eccrine Sweat Glands in Mice

Kamberov et al. 2013

Figure 7. 370A Is Associated with Increased Eccrine Sweat Gland Density in Humans

アジア型EDARは薄い鬚・大きい耳たぶ・下顎の後退とも関連

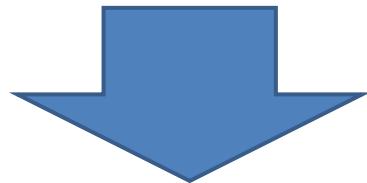
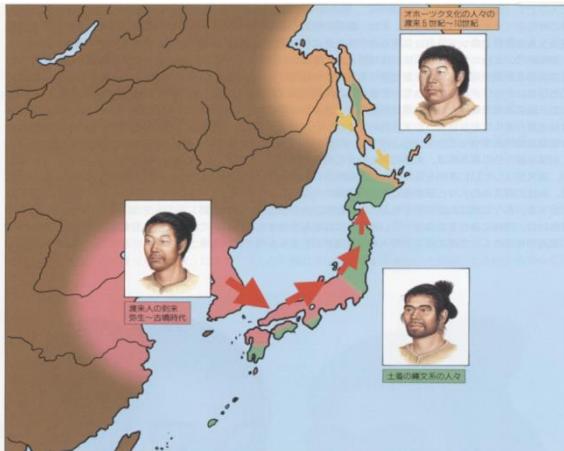


東アジア特異的に自然選択が働いた遺伝子

Population	Gene	Chr	Association
East Asia	<i>ABCC11</i>	16	耳垢、体臭
	<i>ADH1B</i>	4	アルコール不耐性
	<i>EDAR</i>	2	毛髪・歯形態など

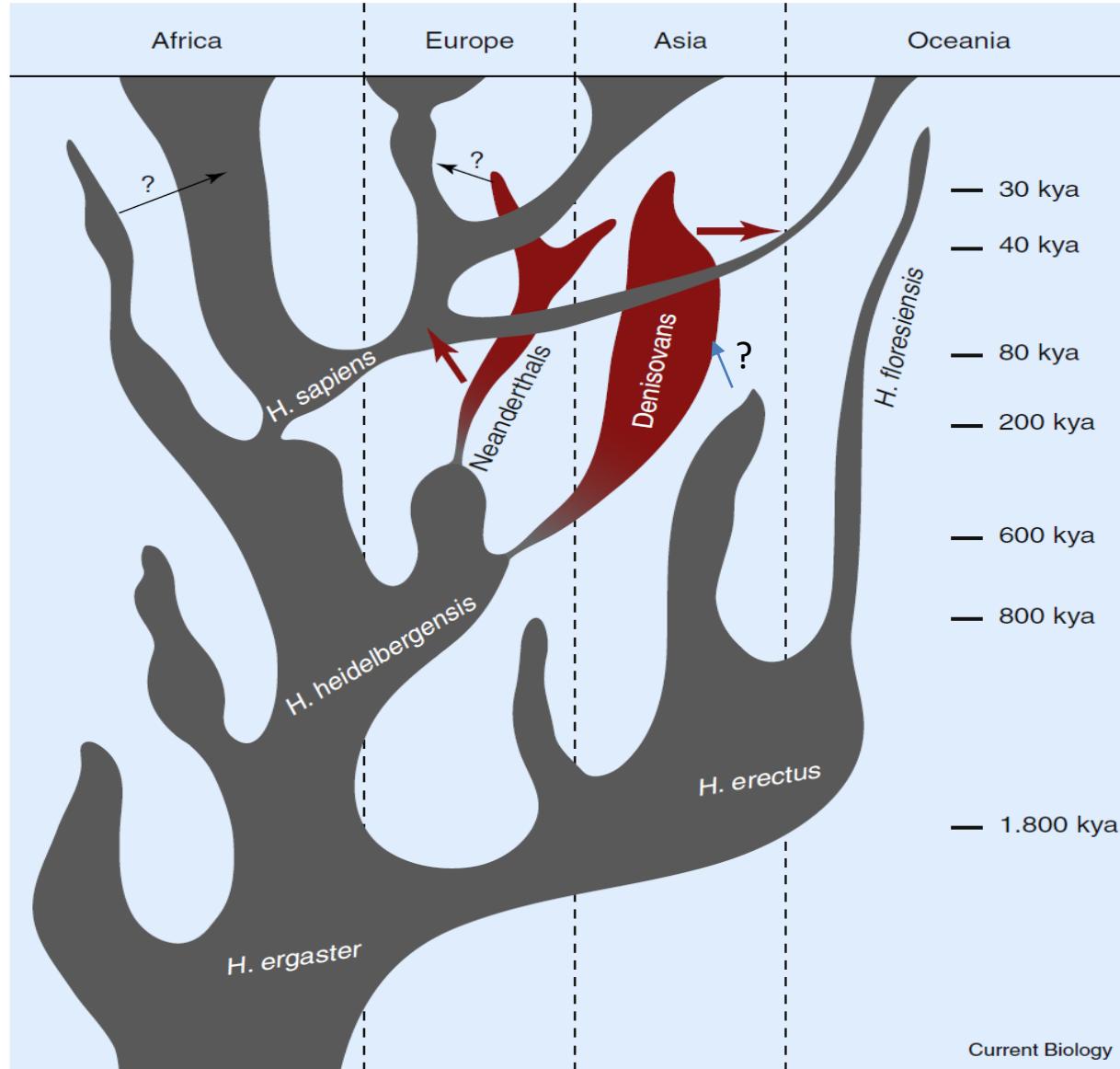
■ 東アジア大陸集団において変異型が選択

■ 縄文人や琉球人においては祖先型が比較的高頻度



琉球－本土間の表現型分化

*Homo*属の進化の歴史



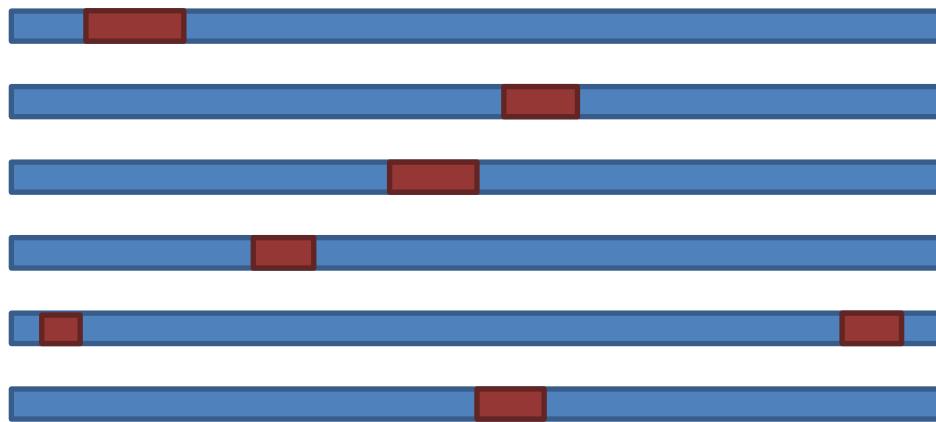
Lalueza-Fox and Gilbert, Current Biology 2011

ネアンデルタール人からの遺伝子流入

Neanderthal



Modern human



Every non-African modern human have 1-4% of genome sequences derived from the Neanderthals.

Adaptive introgression

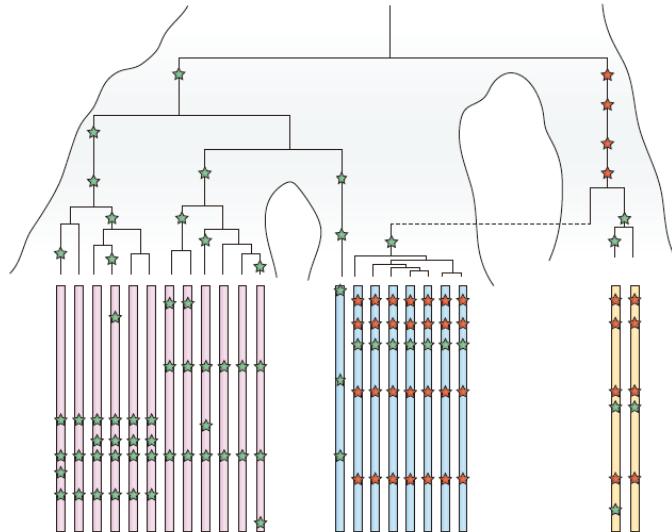


Figure 2 | Example coalescent genealogy of uniquely shared mutations. Several DNA fragments from two modern populations (pink and blue chromosomes) are sequenced. A diploid sequence is also obtained from an extinct archaic population (yellow chromosome) that split from the population tree more anciently than the two modern populations split from each other. Uniquely shared mutations (red stars) occur in the archaic population but are passed on to the ancestors of the blue modern population via admixture (dashed line). These are then swept to high frequency by selection, producing a shallow local coalescent genealogy. This process results in sites with high-frequency derived alleles in the blue samples that are present in the archaic sample but not in the pink samples from the other modern population. Mutations in the genealogy that are not uniquely shared are shown as green stars.

Racimo et al. 2015

Table 1 | Candidates for adaptive introgression

Introgressed haplotype	Closest putative archaic source population	Modern populations showing evidence of introgression	Evidence for positive selection and selection tests performed	Most likely population in which selection occurred
HLA-A, HLA-B and HLA-C	Neanderthal and Denisovan	Europeans, East Asians and Melanesians	Extreme allelic and haplotypic diversity in the HLA region, which is indicative of balancing selection ⁵⁸	Europeans, East Asians and Melanesians
HLA-DPB1	Neanderthal (?), but this was questioned in REF. 73	Europeans (?)	No formal test of neutrality performed ⁷² , but a phylogenetic analysis suggests that the haplotype is not introgressed ⁷³	–
STAT2 (haplotype N)	Neanderthal	Non-Africans	One-tailed test for elevated frequency of diagnostic SNP in introgressed haplotype based on empirical distribution of SNP frequencies ⁴¹	Melanesians
STAT2 (haplotype D)	Denisovan (?)	Melanesians	No formal test of neutrality performed, but haplotype is only present in Papuans at a low frequency ⁴¹	–
OAS1	Denisovan, but extremely ancient coalescence with human reference suggests that the direct source was a different archaic group	Melanesians	<ul style="list-style-type: none"> • No evidence for positive selection in REF. 75 • Evidence for positive selection in REF. 44 • Suggestive signal of balancing selection based on genetic differentiation of haplotypes across continents⁷⁴ 	–
OAS gene cluster	Neanderthal	Non-Africans		
HYAL2 (3p21.31) ⁶⁰	Neanderthal	East Asians	iHS ⁴⁶ , EHH ⁴⁷ and CMS ⁵¹	East Asians
MC1R ⁷⁸	Neanderthal	Non-Africans	Tajima's D ⁴⁹ , Fu and Li's test ⁵² , and iHS ⁴⁶	Taiwanese (?)
SLC16A11 and SLC16A13	Neanderthal	Native Americans	Genotype–phenotype association ⁹⁰	Native Americans (?)
DMD	Neanderthal	Non-Africans	No tests performed ^{91,92}	–
EPAS1 (REF. 57)	Denisovan	East Asians or Tibetans only	<ul style="list-style-type: none"> • High population differentiation⁵⁵ • Genotype–phenotype association⁵⁵ • High archaic haplotype frequency⁵⁷ 	Tibetans
Various regions identified via S ⁴ that contain genes involved in the integumentary system	Neanderthal	Europeans and East Asians	<ul style="list-style-type: none"> • High F_{ST} between Europeans and Asians in variants identified to be introgressed (local adaptation)¹⁹ • Introgressed haplotypes with frequencies in both Europeans and Asians that are higher than expected under neutrality (shared adaptation)¹⁹ 	Europeans and East Asians
Various regions identified via CRF that contain genes involved in keratin filament, sugar metabolism, muscle contraction and oocyte meiosis	Neanderthal	Europeans and East Asians	Windows in which the population frequency of the Neanderthal genetic material is too high to be explained by neutral drift ⁴⁴	Europeans and East Asians
Various regions identified via HMM	Neanderthal and Denisovan	Europeans, East Asians and Melanesians	No claims about selection and no formal tests of neutrality performed ^{43,41}	–

CMS, composite of multiple signals; CRF, conditional random field; DMD, dystrophin; EHH, extended haplotype homozygosity; EPAS1, endothelial PAS domain protein 1; HLA, human leukocyte antigen; HMM, hidden Markov model; HYAL2, hyaluronoglucosaminidase 2; iHS, integrated haplotype score; MC1R, melanocortin 1 receptor; OAS, 2'-5'-oligoadenylate synthetase; SLC, solute carrier; SNP, single-nucleotide polymorphism; STAT2, signal transducer and activator of transcription 2. The table provides a summary of recent studies reporting adaptive introgression with a comparison of the evidence provided in support of positive selection acting on the introgressed regions. A question mark denotes that the source population, the receiving population or the population in which selection occurred remains unresolved and/or has conflicting reports from different studies.

ネアンデルタル人由来のゲノム領域

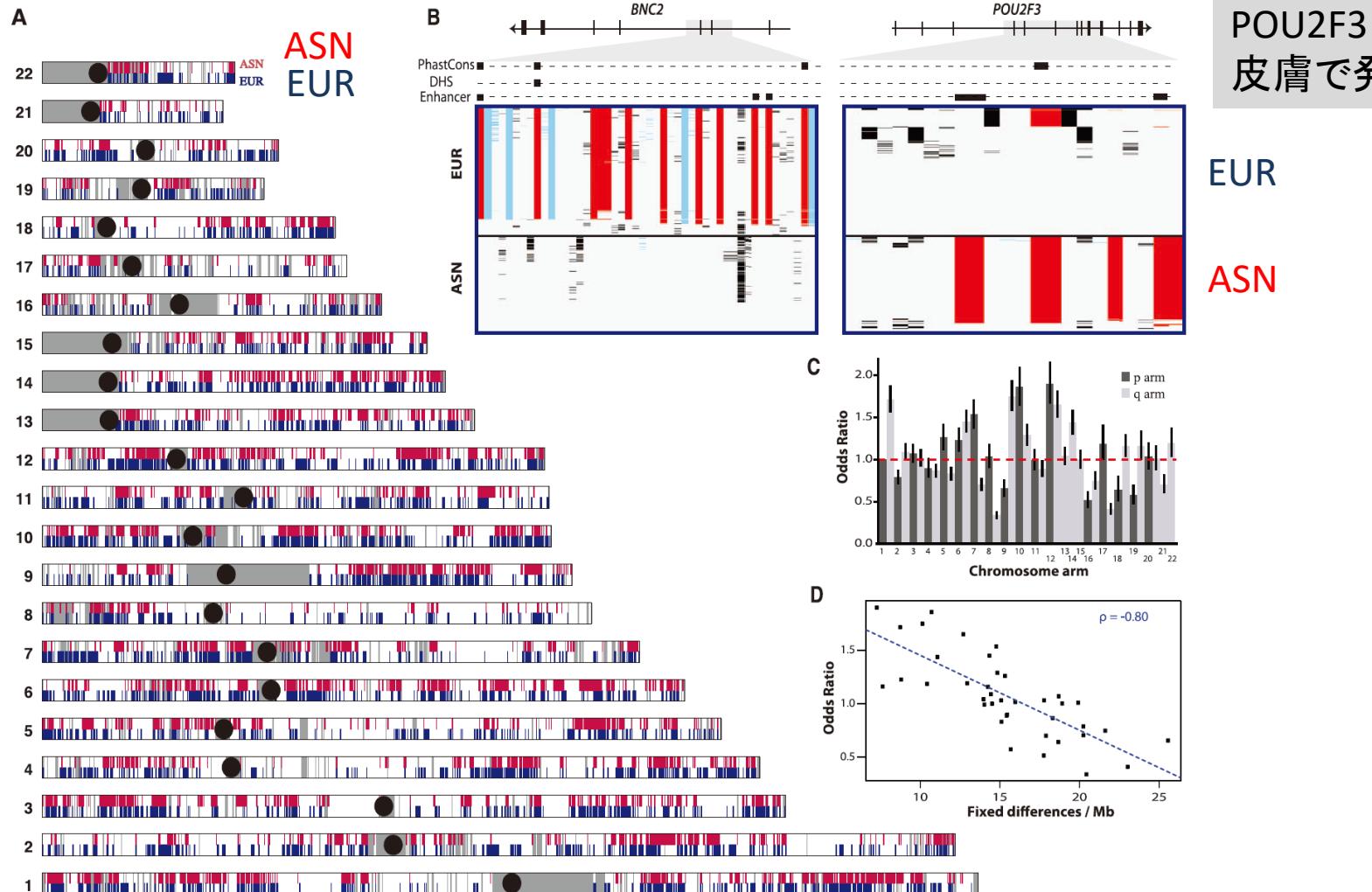


Fig. 2. Genomic distribution of surviving Neandertal lineages. (A) Neandertal lineages identified in East Asians (ASN, red) and Europeans (EUR, blue). Gray shading denotes regions that did not pass filtering criteria (10); black circles represent centromeres. (B) Visual genotype illustrations of introgressed sequences identified in the *BNC2* and *POU2F3* genes. Rows denote individuals, columns indicate variant sites, and rectangles are colored according to genotype (red, predicted Neandertal variant that matches the allele present in the Neandertal reference genome; blue, predicted Neandertal variant that

does not match the allele present in the Neandertal reference genome; black, other variants). Introgressed variants that overlap a PhastCons conserved element, DNaseI hypersensitive site (DHS), or putative enhancer elements are shown as boxes (10). (C) Odds of finding an introgressed lineage on each chromosomal arm calculated from a logistic regression model (10). Odds ratios (ORs) are expressed using chromosome 1p as the baseline level. Horizontal bars represent 95% CIs. (D) Relation between the OR and the number of fixed differences per megabase between humans and Neandertals. ρ , Spearman's rank correlation coefficient.

Vernot et al. 2014

POU2F3:
皮膚で発現

EUR
ASN

チベット人の高地適応 デニソワ人に由来

Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude

Xin Yi,^{1,2*} Yu Liang,^{1,2*} Emilia Huerta-Sanchez,^{3*} Xin Jin,^{1,4*} Zha Xi Ping Cuo,^{2,5*} John E. Pool,^{3,6*} Xun Xu,¹ Hui Jiang,¹ Nicolas Vinckenbosch,³ Thorfinn Sand Korneliussen,⁷ Hancheng Zheng,^{1,4} Tao Liu,¹ Weiming He,^{1,8} Kui Li,^{2,5} Ruibang Luo,^{1,4} Xifang Nie,¹ Honglong Wu,^{1,9} Meiru Zhao,¹ Hongzhi Cao,^{1,9} Jing Zou,¹ Ying Shan,^{1,4} Shuzheng Li,¹ Qi Yang,¹ Asan,^{1,2} Peixiang Ni,¹ Geng Tian,^{1,2} Junming Xu,¹ Xiao Liu,¹ Tao Jiang,^{1,9} Renhua Wu,¹ Guangyu Zhou,¹ Meifang Tang,¹ Junjie Qin,¹ Tong Wang,¹ Shuijian Feng,¹ Guohong Li,¹ Huasang,¹ Jiangbai Luosang,¹ Wei Wang,¹ Fang Chen,¹ Yading Wang,¹ Xiaoguang Zheng,^{1,2} Zhuo Li,¹ Zhuoma Bianba,¹⁰ Ge Yang,¹⁰ Xiping Wang,¹¹ Shuhui Tang,¹¹ Guoyi Gao,¹² Yong Chen,⁵ Zhen Luo,⁵ Lamu Gusang,⁵ Zheng Cao,¹ Qinghui Zhang,¹ Weihan Ouyang,¹ Xiaoli Ren,¹ Huiqing Liang,¹ Huisong Zheng,¹ Yebo Huang,¹ Jingxiang Li,¹ Lars Bolund,¹ Karsten Kristiansen,^{1,7} Yingrui Li,¹ Yong Zhang,¹ Xiuqing Zhang,¹ Ruiqiang Li,^{1,7} Songgang Li,¹ Huanming Yang,¹ Rasmus Nielsen,^{1,3,7†} Jun Wang,^{1,7†} Jian Wang^{1†}

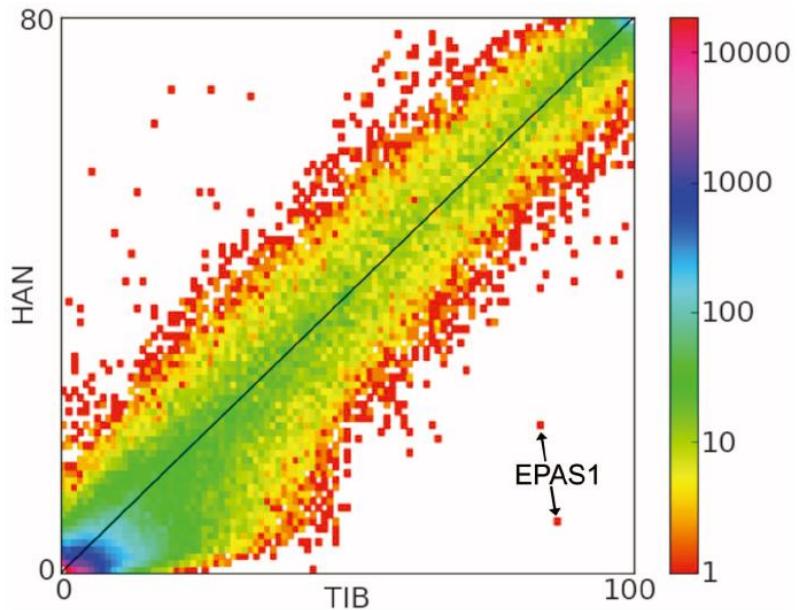
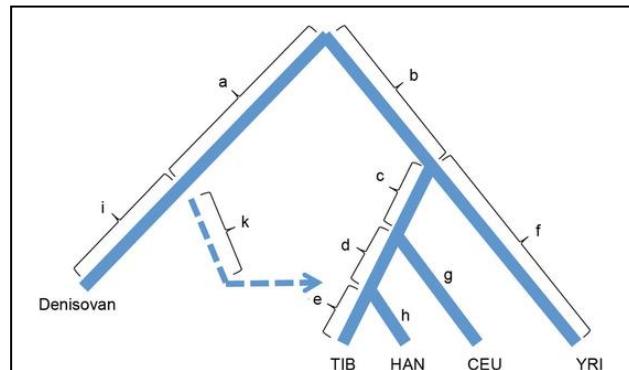
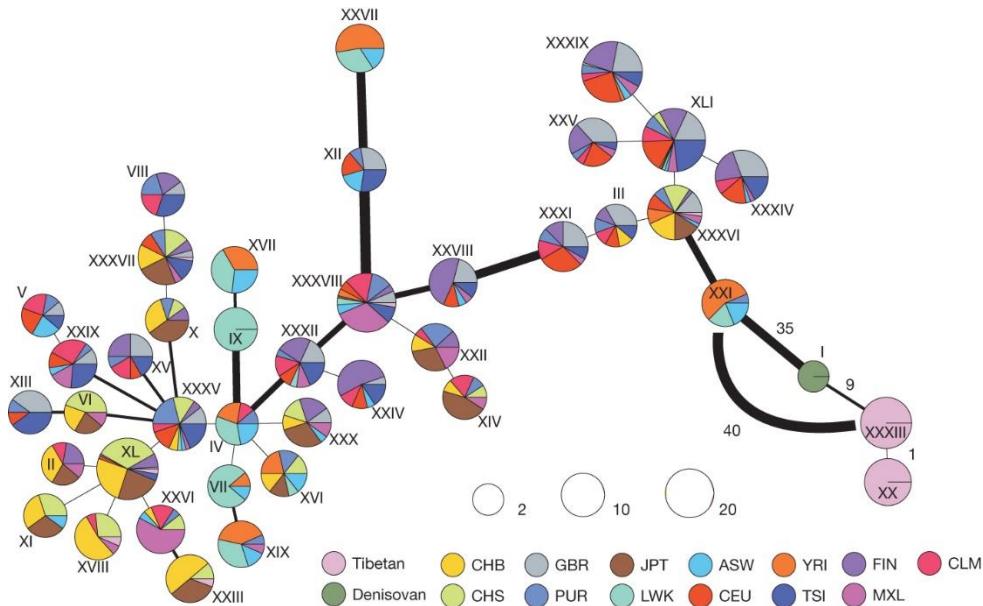


Fig. 1. Two-dimensional unfolded site frequency spectrum for SNPs in Tibetan (x axis) and Han (y axis) population samples. The number of SNPs detected is color-coded according to the logarithmic scale plotted on the right. Arrows indicate a pair of intronic SNPs from the *EPAS1* gene that show strongly elevated derived allele frequencies in the Tibetan sample compared with the Han sample.

Yi et al. 2010

EPAS1 = HIF2A
低酸素誘導性因子



Huerta-Sanchez et al. 2014

古代人類由来ハプロタイプとCOVID-19

Article

The major genetic risk factor for severe COVID-19 is inherited from Neanderthals

<https://doi.org/10.1038/s41586-020-2818-3>

Hugo Zeberg^{1,2} & Svante Pääbo^{1,3}

Received: 3 July 2020

Accepted: 22 September 2020

Published online: 30 September 2020

Check for updates

A recent genetic association study¹ identified a gene cluster on chromosome 3 as a risk locus for respiratory failure after infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). A separate study (COVID-19 Host Genetics Initiative)² comprising 3,199 hospitalized patients with coronavirus disease 2019 (COVID-19) and control individuals showed that this cluster is the major genetic risk factor for severe symptoms after SARS-CoV-2 infection and hospitalization. Here we show that the risk is conferred by a genomic segment of around 50 kilobases in size that is inherited from Neanderthals and is carried by around 50% of people in south Asia and around 16% of people in Europe.

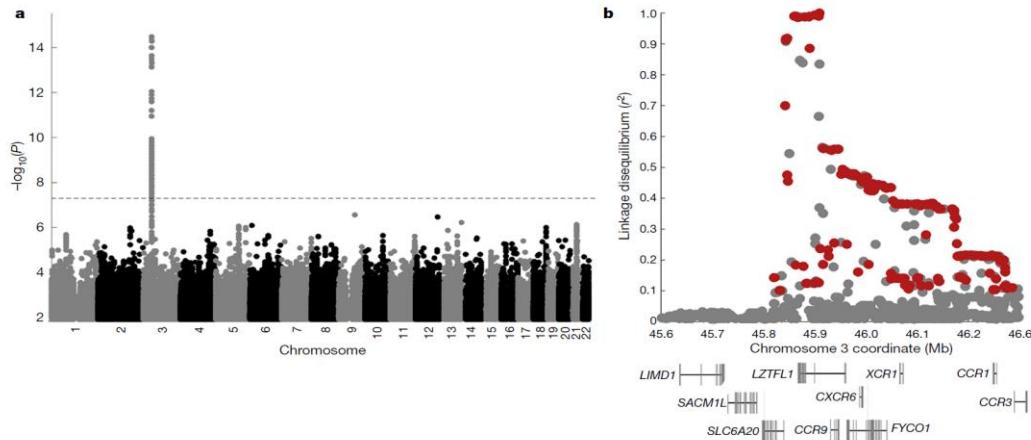


Fig. 1 | Genetic variants associated with severe COVID-19. **a**, Manhattan plot of a genome-wide association study of 3,199 hospitalized patients with COVID-19 and 897,488 population controls. The dashed line indicates genome-wide significance ($P = 5 \times 10^{-8}$). Data were modified from the COVID-19 Host Genetics Initiative² (<https://www.covid19hg.org/>). **b**, Linkage disequilibrium between the index risk variant (rs35044562) and genetic variants in the 1000 Genomes Project. Red circles indicate genetic variants for which the alleles are correlated to the risk variant ($r^2 > 0.1$) and the risk alleles match the Vindija 33.19 Neanderthal genome. The core Neanderthal haplotype ($r^2 > 0.98$) is indicated by a black bar. Some individuals carry longer Neanderthal-like haplotypes. The location of the genes in the region are indicated below using standard gene symbols. The x axis shows hg19 coordinates.

ネアンデルタール由来のハプロタイプは
COVID-19の重症化と関連

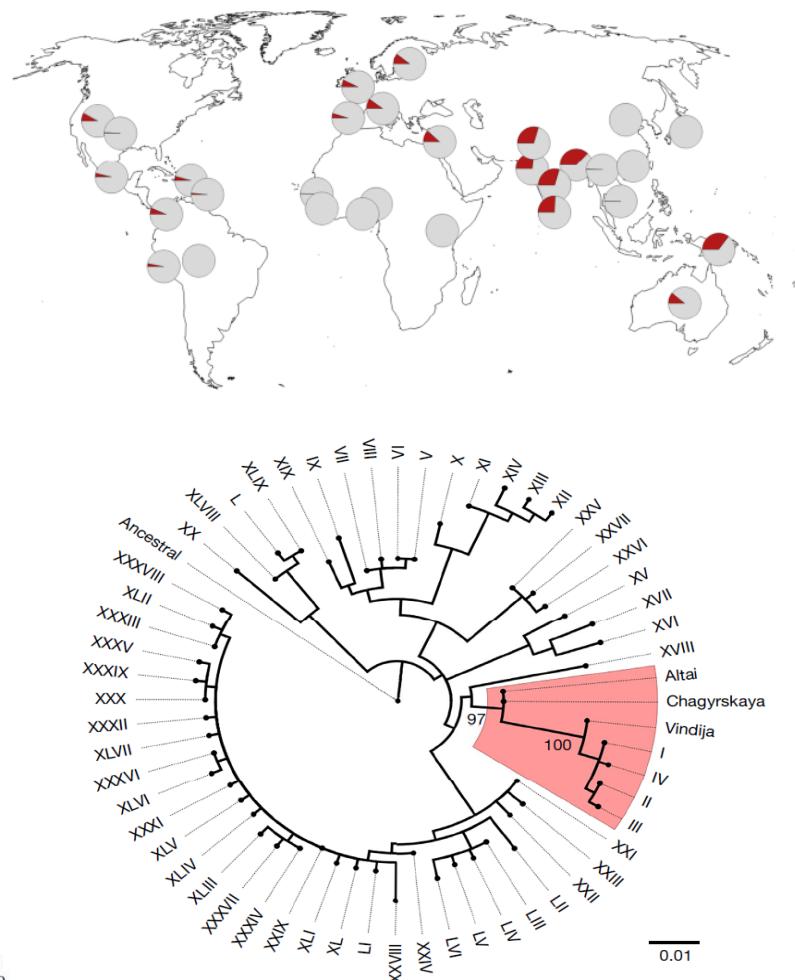


Fig. 2 | Phylogeny relating the DNA sequences that cover the core Neanderthal haplotype in individuals from the 1000 Genomes Project and Neanderthals. The coloured area highlights the haplotypes that carry the risk allele at rs35044562—that is, the risk haplotypes for severe COVID-19. Arabic numbers indicate bootstrap support (100 replicates). The phylogeny is rooted with the inferred ancestral sequence of present-day humans. The three Neanderthal genomes carry no heterozygous positions in this region. Scale bar, number of substitutions per nucleotide position.

Zeberg and Paabo 2020

まとめ

- ほとんどの形質は集団間で分布がオーバーラップ
- 一部の形質だけが集団間で大きく分化
特に皮膚色や毛髪形態など容姿は分化している最たるものつまり、ヒトは集団間で見た目ほどは変わらない
- 選択圧がヒト集団間の形質を大きく分化させる
ヒトのちがいは環境や生業への遺伝適応のたまもの
それぞれの環境において生物として最適化されてきた

- ・ヒトのちがいに対する自然人類学的な理解



- ・不当なステレオタイプ化や差別を解消
- ・ヒトのちがいを尊重する心を育む

當奉納願と賀久面

一三甲麻市國立万松画



Thank you very much
for your attention !