

集団遺伝学的手法による 琉球諸語の解析

シンポジウム「日琉諸方言系統論の展望」
2020年12月19-20日

木村 亮介
和智 仲是
狩俣 繁久
(琉球大学)



言語学と集団遺伝学とのコラボレーション

1. 集団遺伝学的手法の言語データへの適用
2. ヒトゲノム解析と言語解析の比較研究

言語の類縁性を解析するためのデータ

1. 文法
要素の有り無し(1・0)データ
2. 語彙
有り無し(1・0)データ
3. 音韻体系
音素記号のmono/di/tri-gram解析(登場回数)
4. 同源語の音韻
音素のアラインメント

など

遺伝子と言語のアナロジー

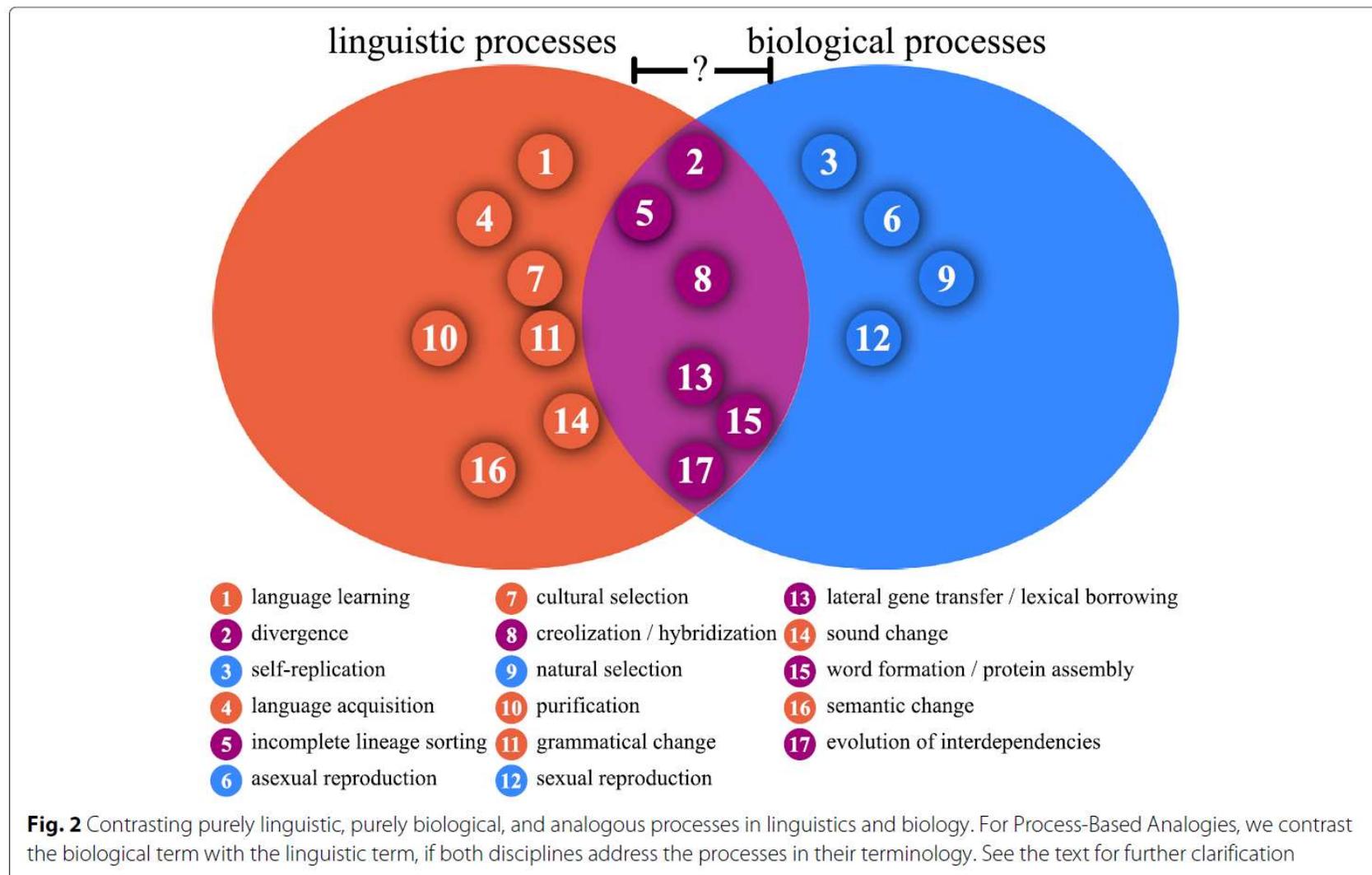
遺伝子	—	単語
相同遺伝子	—	同源語
塩基配列	—	音素配列



Unity and disunity in evolutionary sciences: process-based analogies open common research avenues for biology and linguistics

List et al. (2016)

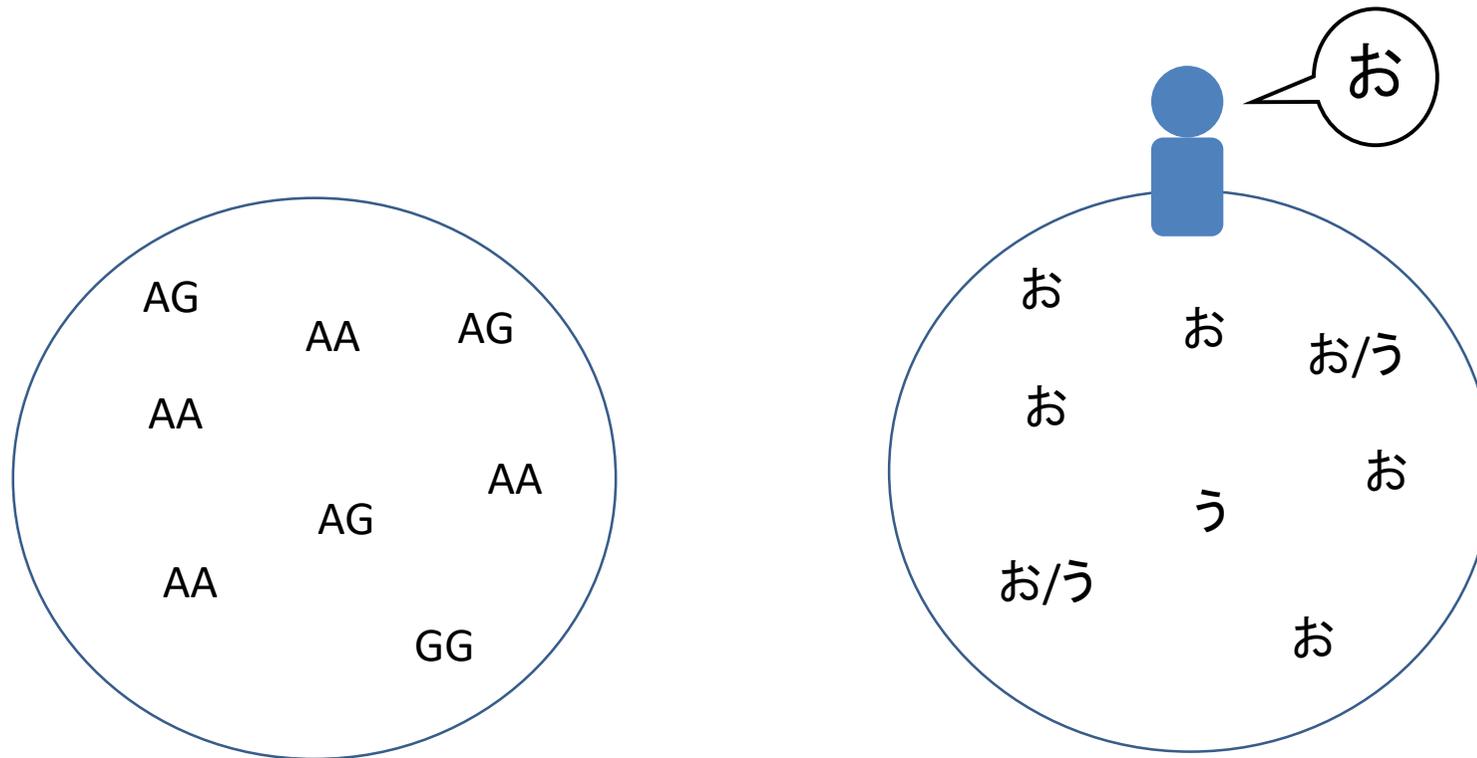
Johann-Mattis List^{1,2*}, Jananan Sylvestre Pathmanathan², Philippe Lopez² and Eric Baptiste²



遺伝子と言語の違い①ーデータの取り方ー

遺伝子:
個体レベルでデータを取る

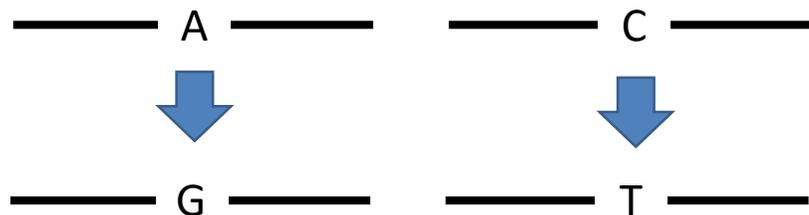
言語:
集団を代表するデータを取る



遺伝子データと言語データは同様に扱える部分と扱えない部分があり、
集団遺伝学の理論を適用するときには注意が必要！

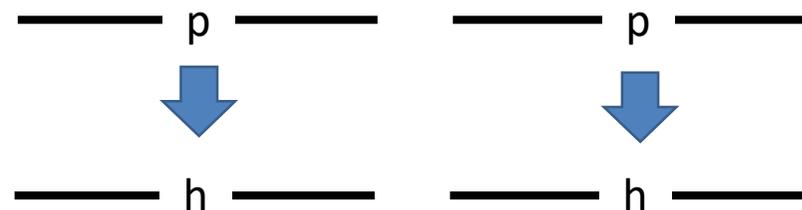
遺伝子と言語の違い②—規則変化—

遺伝子:
変異はランダムで互いに独立



ランダムで独立

言語:
規則的な変化 (音韻体系の変化)
がある

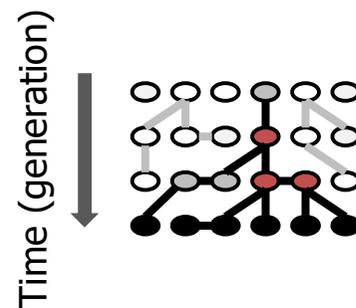
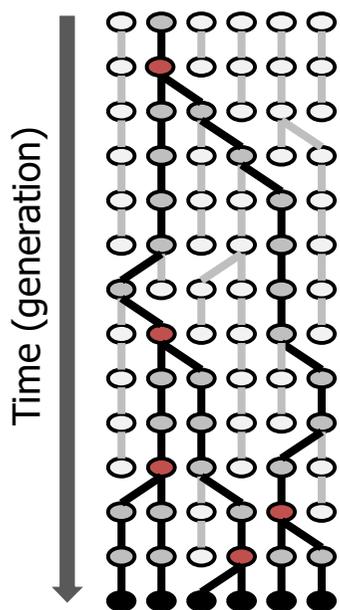
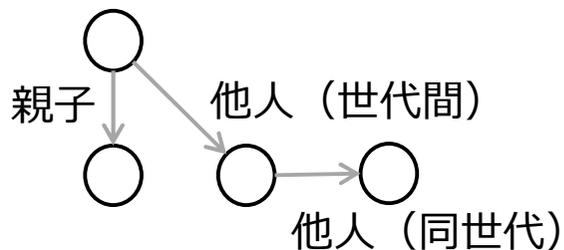
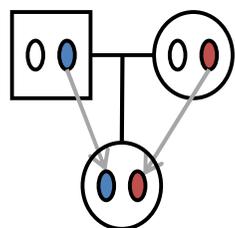


全てのpがhに変わる

遺伝子と言語の違い③—伝達様式—

遺伝子:
親子の垂直伝達

言語:
垂直・斜行・水平伝達



集団全体に
固定する時間
(多型的な状態)

長い
($4N$ 世代)

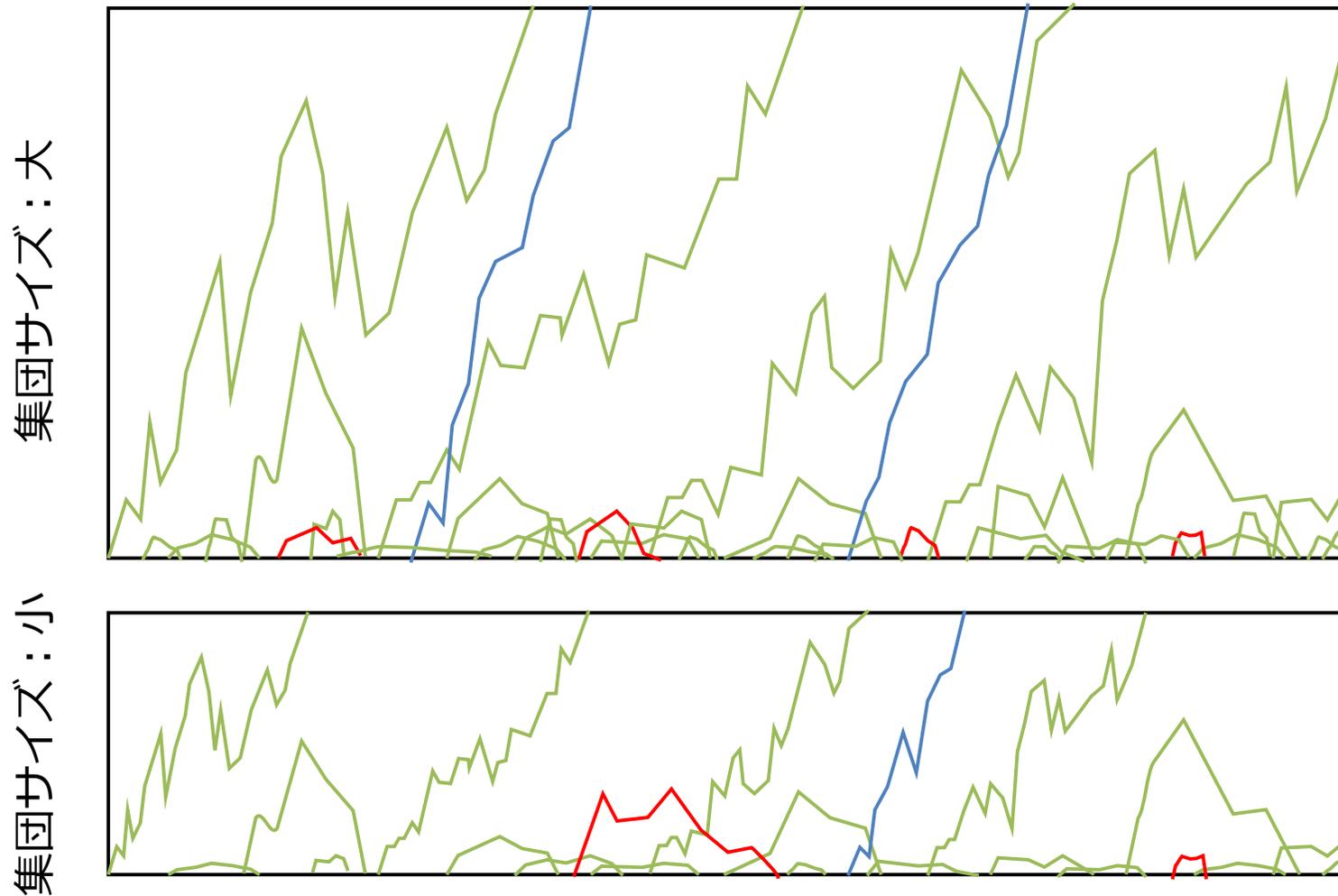
短い

遺伝子と言語の違い④ー進化にかかるバイアスー

	遺伝子	言語
変異を起こす人	ランダム	ランダム or 特定の人 (権力者・教師・若者・移民)
変異の伝達	垂直伝達	垂直・斜行・水平伝達 ランダム or 1対多
選択圧	◎中立 正の選択・負の選択 頻度依存選択	中立 正の選択・負の選択 頻度依存選択

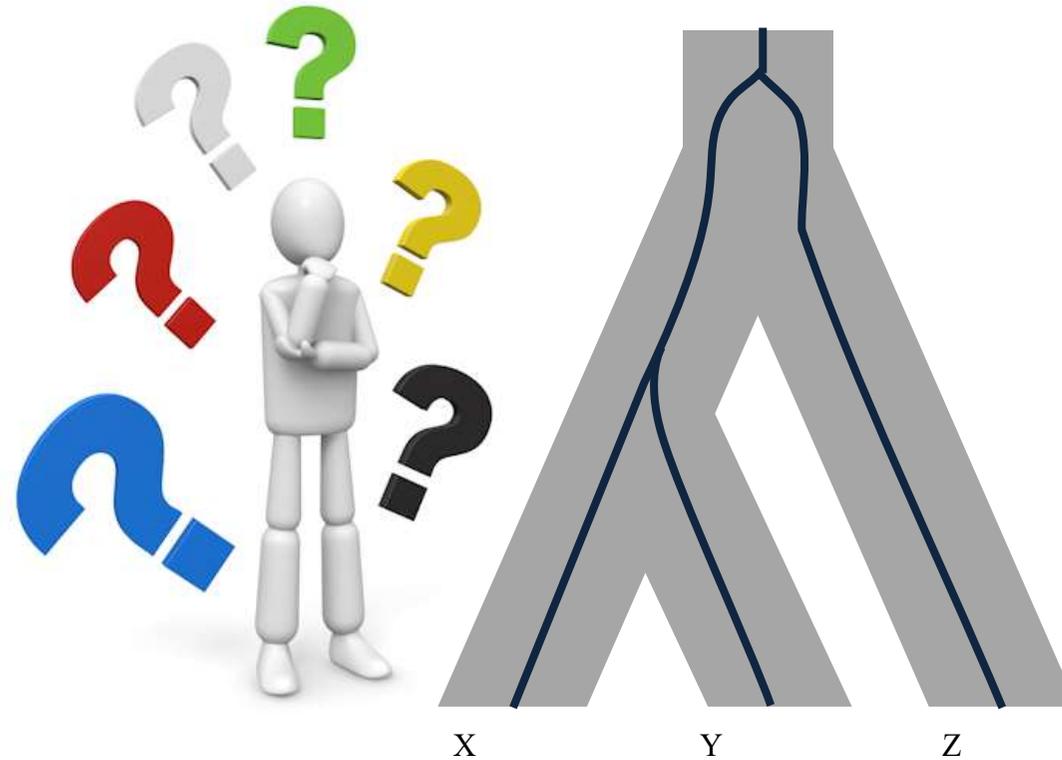
これらに加え、集団サイズ・分集団構造・移住などのデモグラフィが、進化速度を決める

遺伝子の進化速度：遺伝子変異の確率論的振る舞い



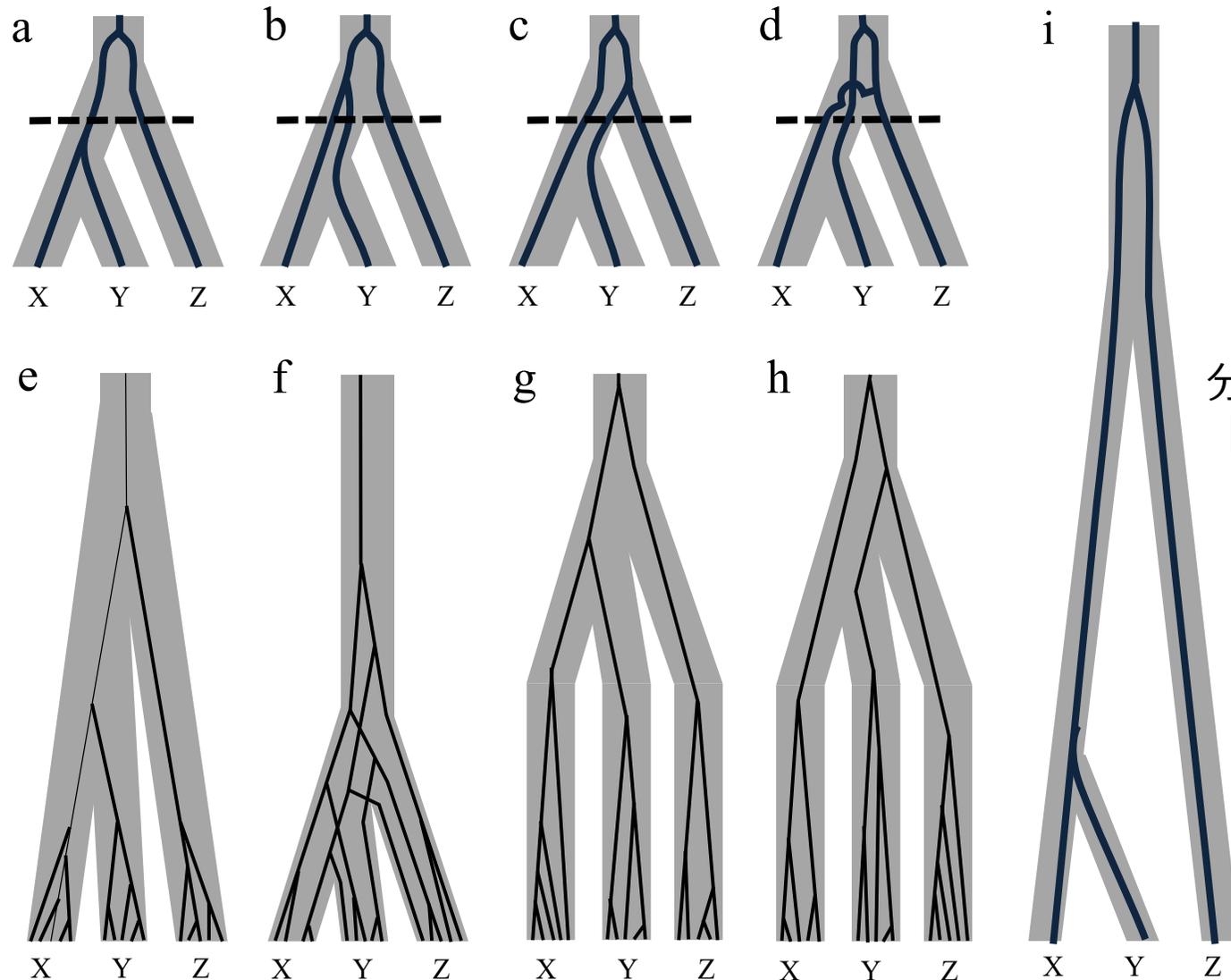
- 突然変異の数 : 集団サイズに比例 ($2N\mu$)、ただしそれぞれの初期頻度 ($1/2N$)
- 中立変異 : 集団サイズに寄らず進化速度一定 ($v = \mu$: 固定確率: $1/2N$)
- 有利な変異 : 集団サイズが大きいほど進化速度が速い
- 弱有害変異 : 集団サイズが小さいと除去されにくい (中立的にふるまう)

集団の系統関係と遺伝子（単語）の系図



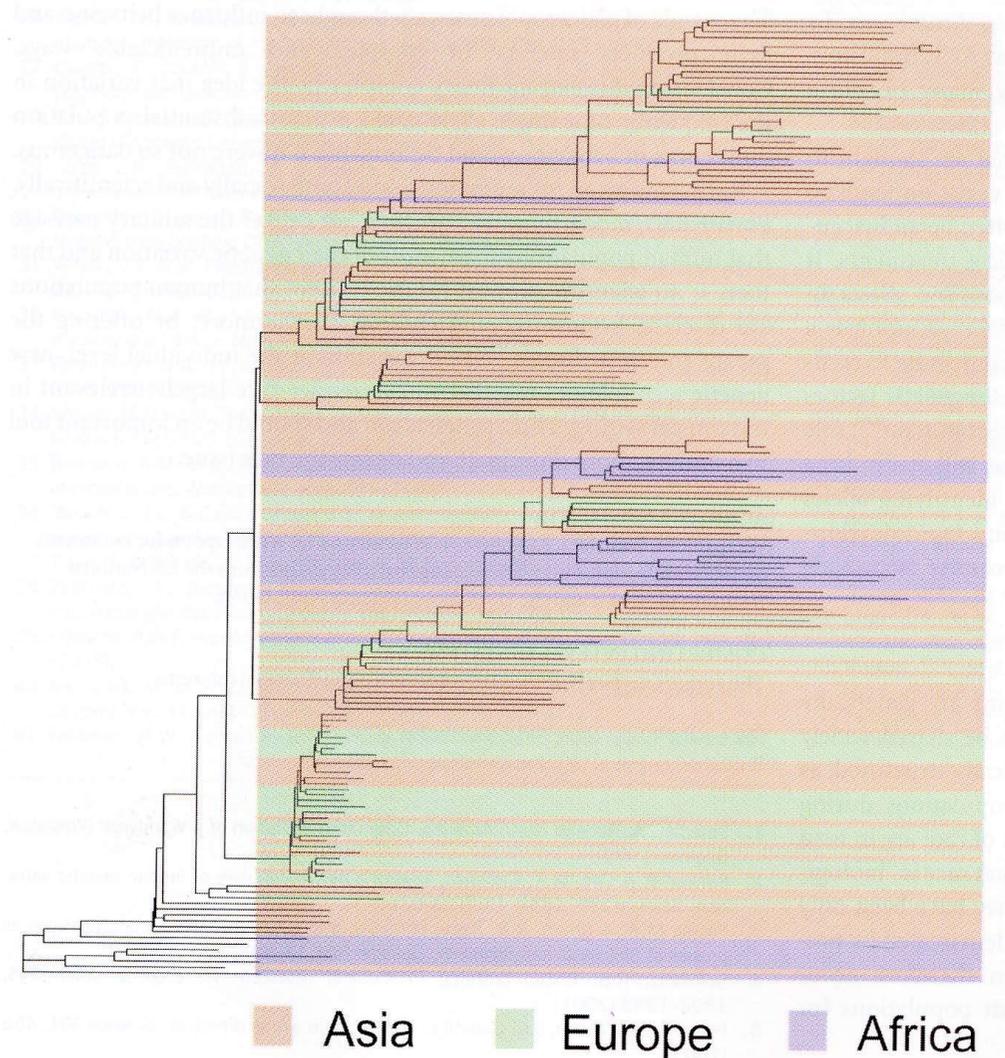
Q. 集団の系統関係が分岐のみで表されるとき、集団の系統関係と遺伝子（単語）の系図は必ず一致する？

Incomplete lineage sorting (ILS) 不完全な系統仕分け



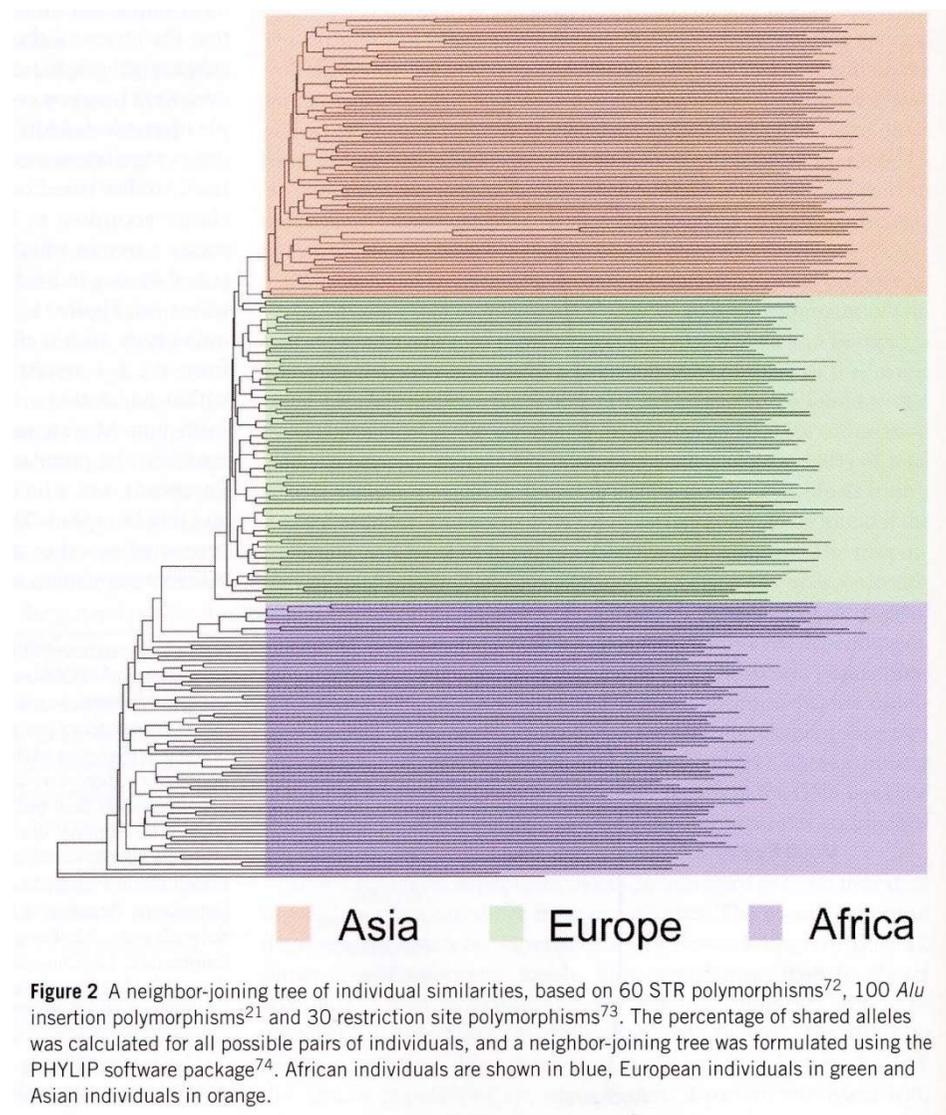
個々の遺伝子の樹形は集団の樹形と必ずしも一致しない

常染色体の1つの遺伝子領域による個体間の樹形図



ILSが起きる状況のとき、1つあるいは少数の遺伝子座（単語）で描いた系統樹は集団の関係を必ずしも反映しない。

多数の遺伝子領域（190遺伝子座）による個体間の樹形図



例えばILSがあったとしても、多数の遺伝子座（単語）を解析することで、正しい集団の系統関係は復元可能（ILSを取り除く必要はない）

Q. 集団の系統関係が分岐のみで表されるとき、集団の系統関係と遺伝子（単語）の系図は必ず一致する？

A. 必ずしも一致しない

ただし、

- 1) 分岐から（分岐と分岐の間）の時間が長ければ、
- 2) 多型的な期間が短ければ

ILSは起きにくい。

遺伝子と言語の違い⑤—ILS—

遺伝子：
起きる

言語：
比較的起きにくい

斜行・水平伝達する言語は、遺伝子と比べるとILSは起きにくいと考えられるが、全くないとは言い切れない。

系統樹作成の注意点

Distance-based methods (距離行列)

- **UPGMA**
- **Neighbor-joining (NJ)**

Character (sequence)-based methods (配列情報)

- **Maximum likelihood (ML)**

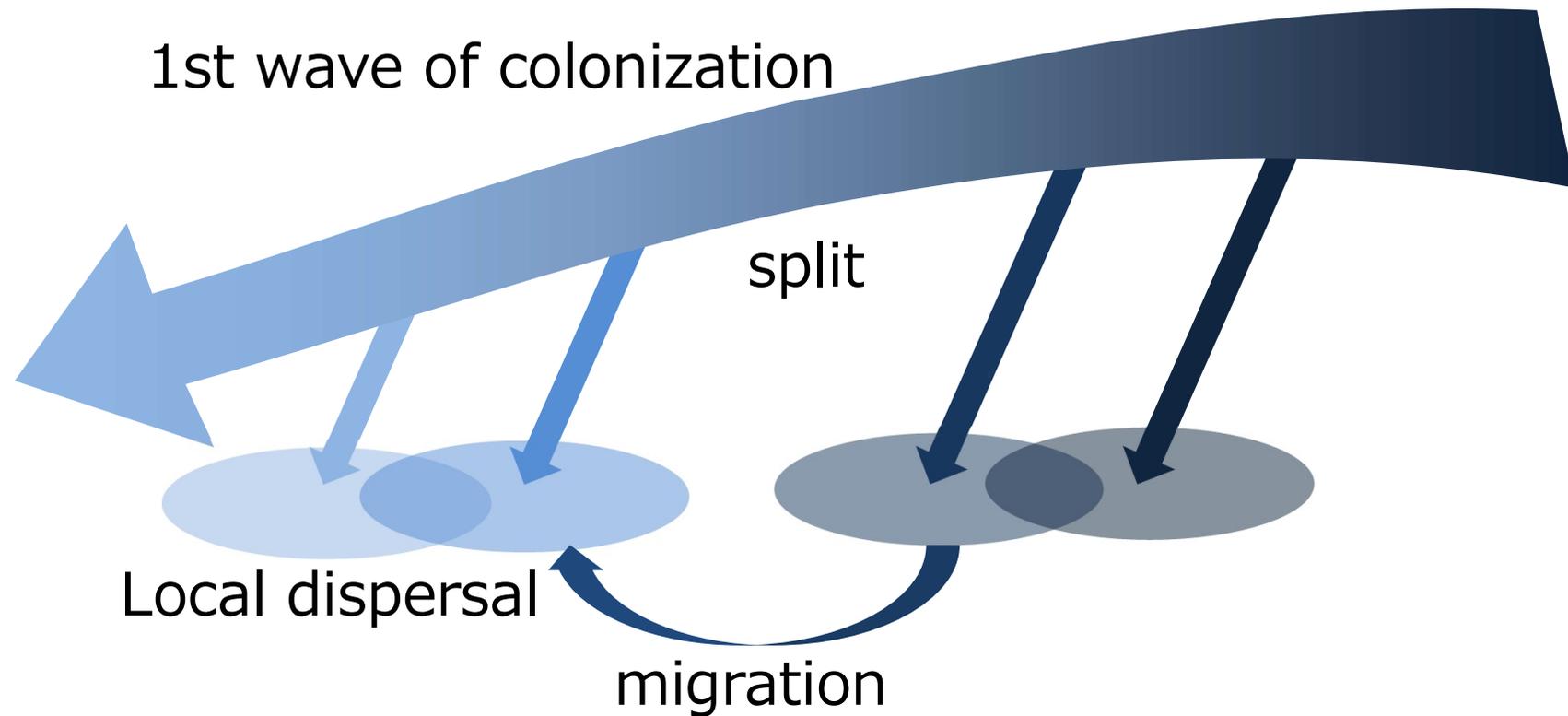
個々の単語の系統樹

⇒ Distance-based○、Character-based◎
ただし1単語あたりの情報量は少なすぎる

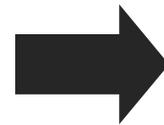
複数単語のconcatenated dataの系統樹

すべての単語が同じトポロジーで分岐 ⇒ どちらも○
異なるトポロジーの単語が混在 ⇒ Distance-based

実際のヒトのゲノムや言語の類縁関係

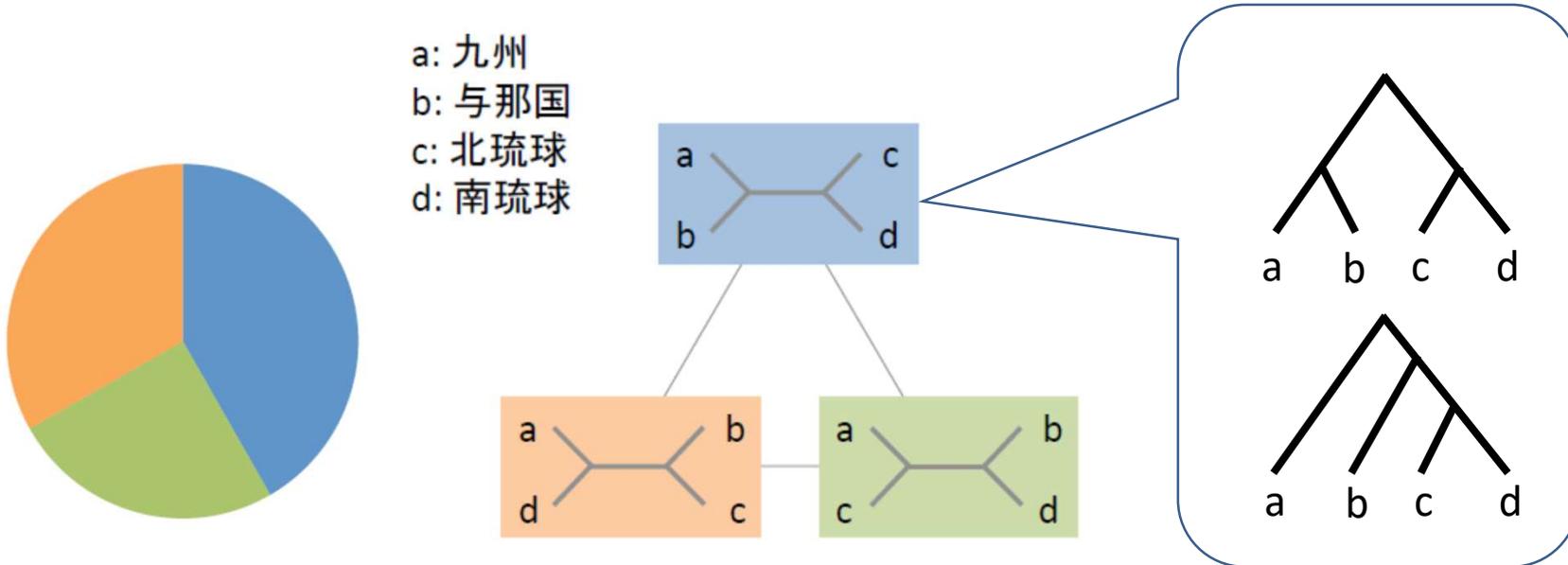


系統樹では表すことができない
(表せたとしても間違っている)



系統ネットワーク
主成分分析 (PCA)
クラスタ分析
などを用いて図示・解釈

単語毎の系統関係の食い違い

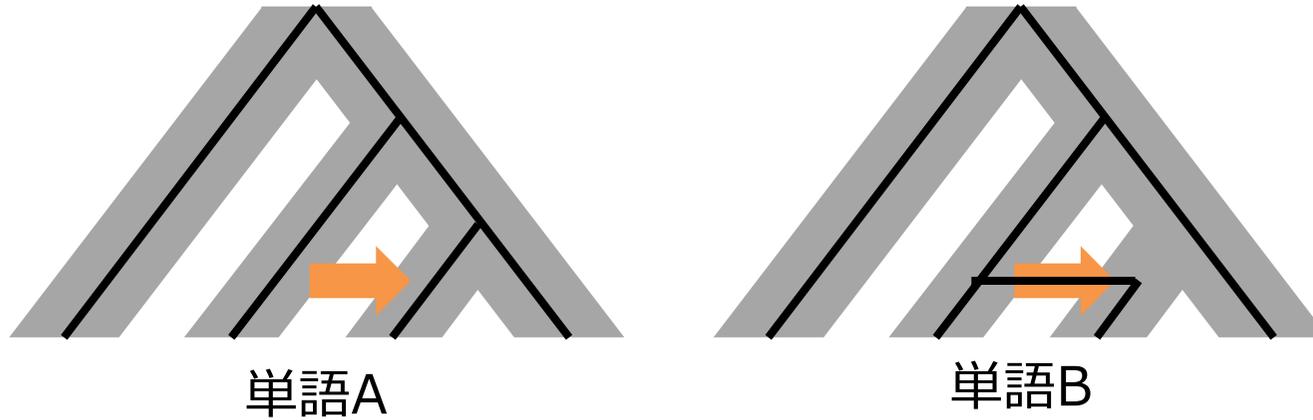


無根系統樹を有根にする場合、
根の取り方によって樹形が変わることに注意
(必ずしもaとbが近いことを示していない)

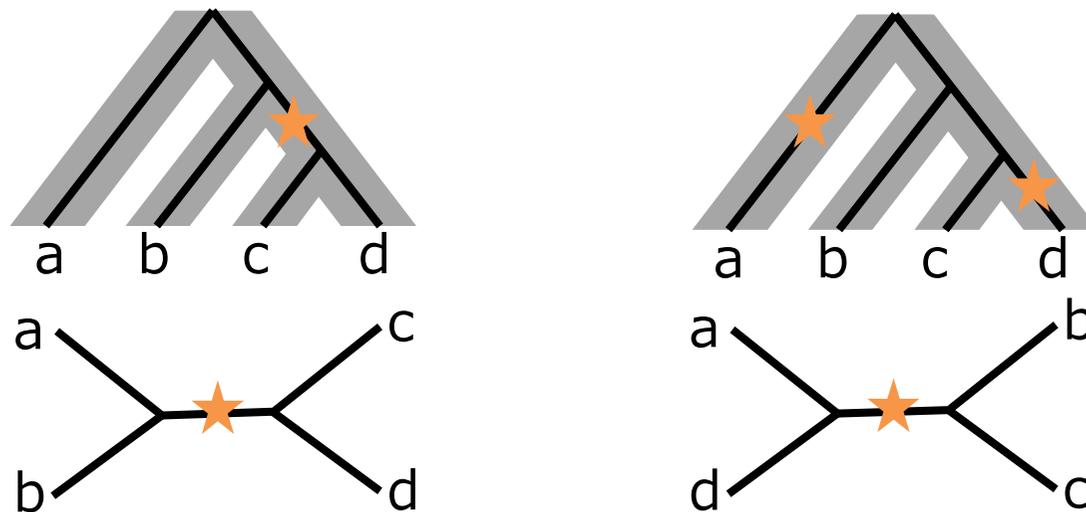
単語毎の系統関係が異なることの解釈

- ① incomplete lineage sorting
- ② 流入 (借用)
- ③ 別々に同じ変化が生じた

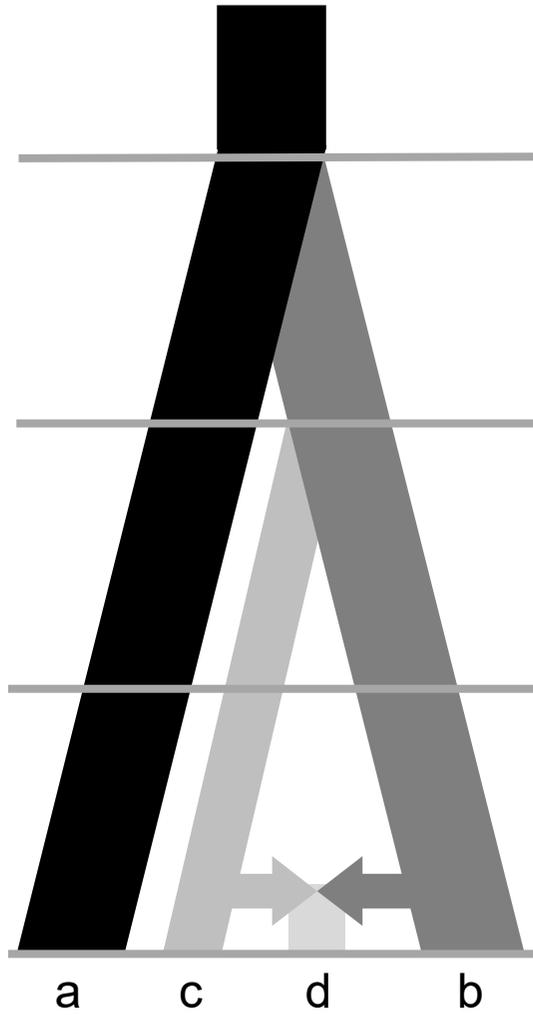
② 流入 (借用)



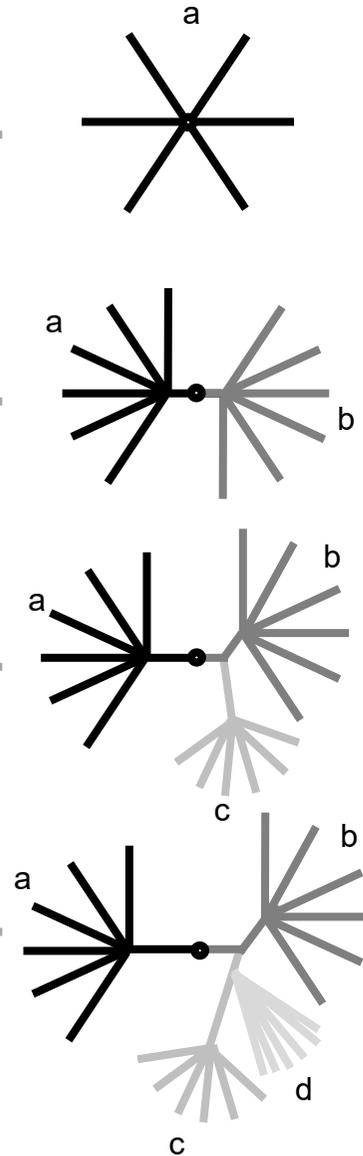
③ 別々に同じ変化が生じた (見かけ上の系統関係)



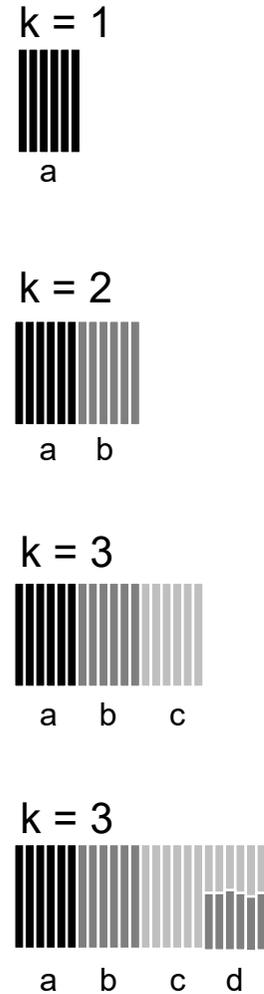
集団動態と様々な解析の結果



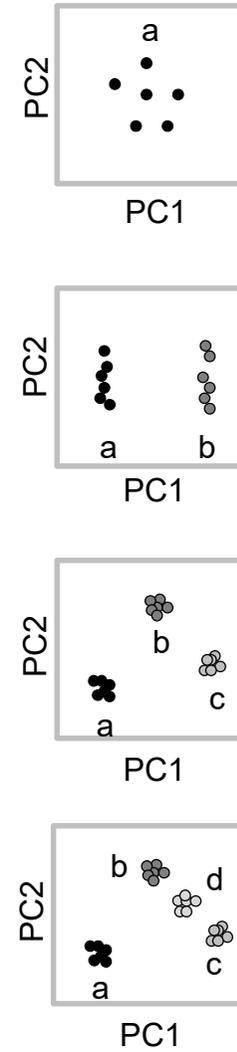
系統樹



クラスタ分析



主成分分析



基盤研究（S） 「言語系統樹を用いた琉球語の比較・歴史言語学的研究」（代表：狩俣繁久）

琉球方言における以下のデータをデータベース化

- 1) 609集落の328語 および
- 2) 115集落の1061語の基礎語彙の手書き資料
- 3) 1万語から1万8千語の語彙を収録した14冊の方言辞典：

『沖縄語辞典』

『奄美方言分類辞典』

『奄美龍郷方言辞典』

『与論方言辞典』

『沖縄今帰仁方言辞典』

『伊是名島方言辞典』

『伊江島方言辞典』

『宮古伊良部方言辞典』

『石垣方言辞典』

『竹富方言辞典』

『与那国ことば辞典』

『種子島方言辞典』

『鹿児島県方言辞典』

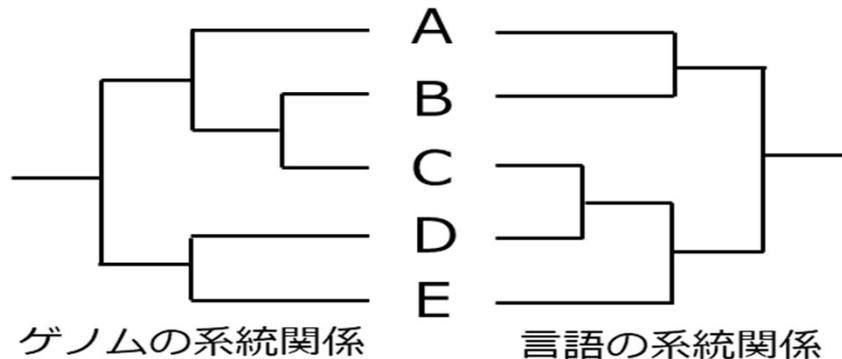
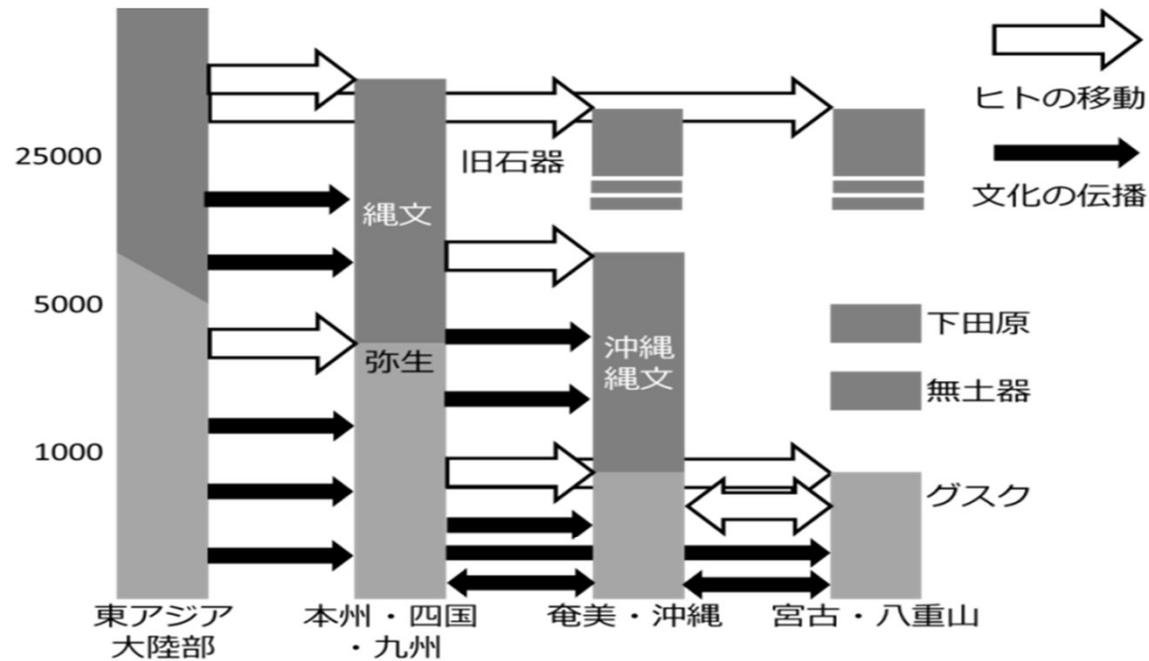
『種子島方言辞典』

言語進化および伝播にまつわる展望

1. 言語進化は中立か？変異および伝達にバイアスがあるか？
(新しい流行り言葉は採用されやすい？)
2. 言語進化速度と集団サイズの関係は？
3. 言語変化における政治・権力の影響は？
(中央と地方とでの伝播の方向、変化速度の違い)
4. 音韻体系、音韻の不規則変化、語彙、文法で変化・伝播の起こり方に違いはあるか？
5. 農業や漁業などの生業・文化の伝播と関連単語の流入

言語進化および伝播にまつわる展望

6. 考古学・歴史学的知見およびゲノム解析の結果と合わせて言語データを解釈





**Thank you very much
for your attention !**