

# 大規模個人ゲノムデータによる ゲノム多様性理解の革新

大学共同利用機関法人 情報・システム研究機構  
統計数理研究所 数理・推論研究系

間野修平

情報・システム研究機構シンポジウム2012

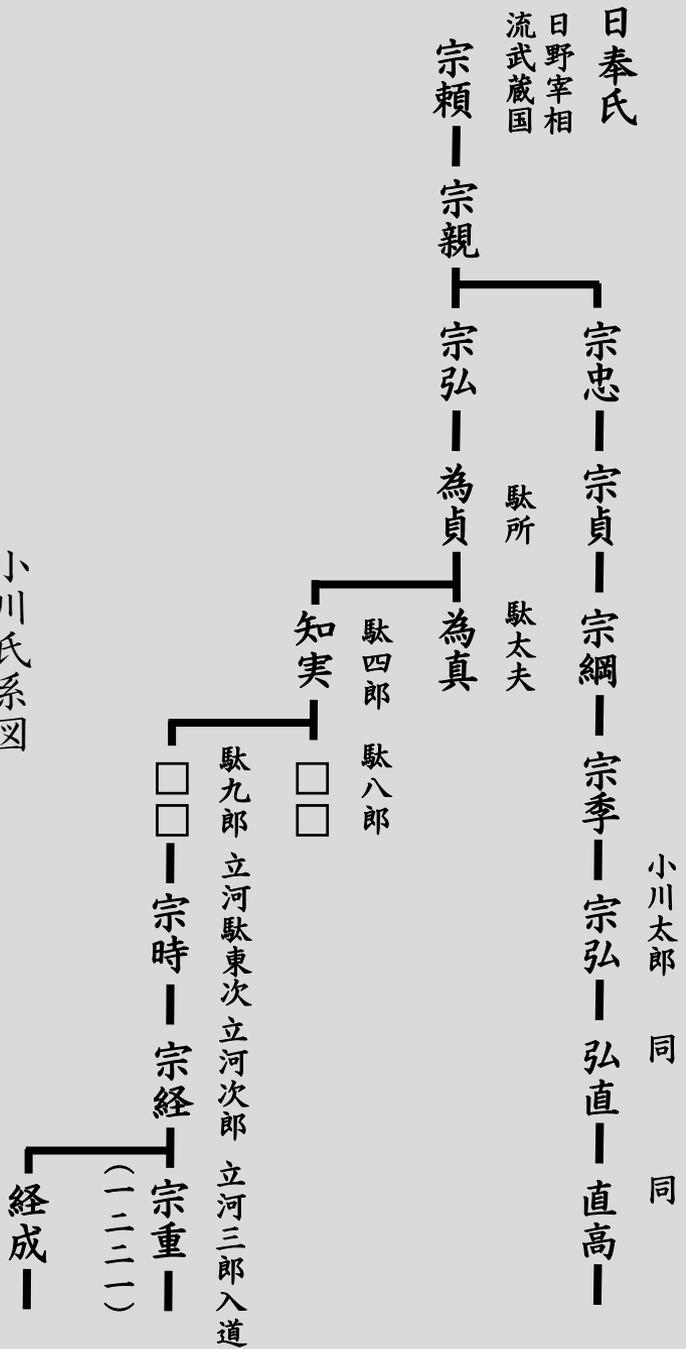
一橋講堂

2012年11月9日



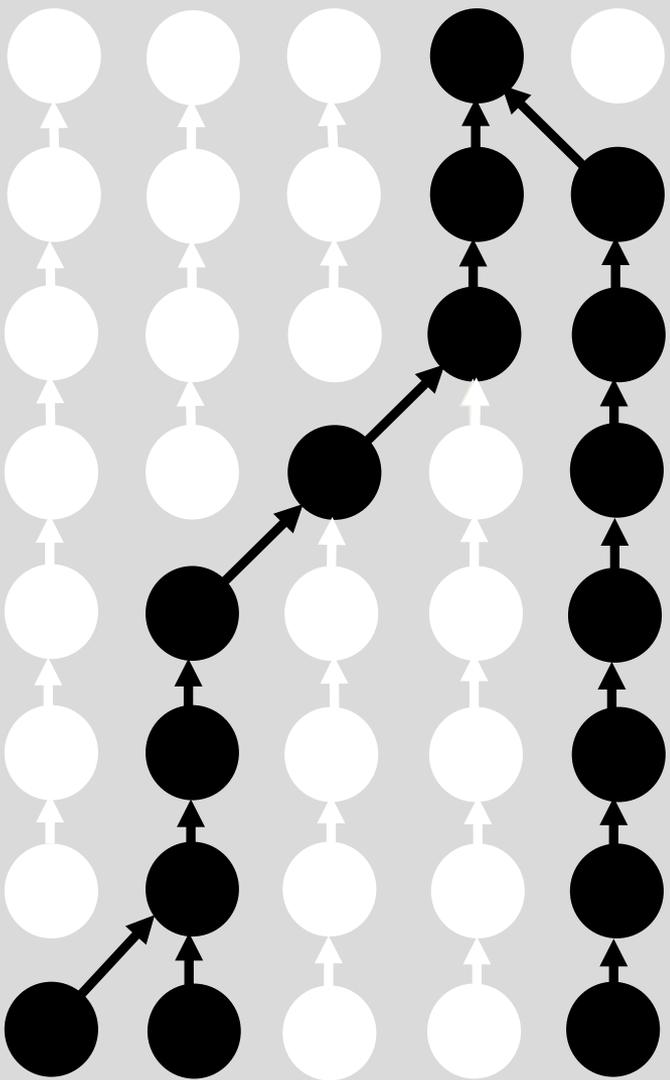
# 立河氏

日奉氏  
日野宰相  
流武蔵国



小川氏系図

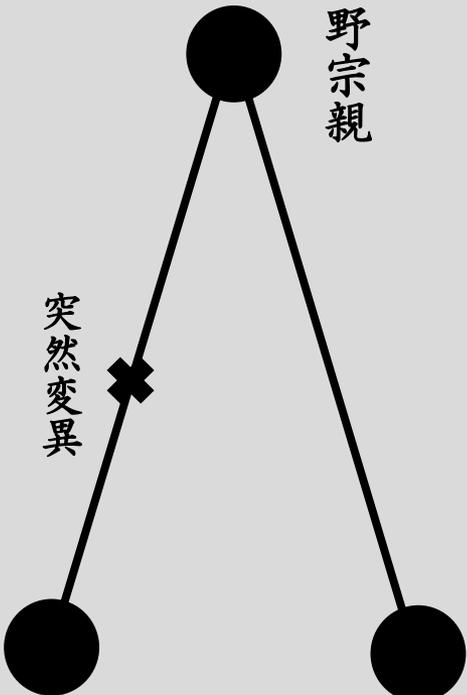
出典 岩間富文著「古文書の語る立河氏」



# 待ち時間

七代 ↑

日野宗親



小川直高

立河三郎入道宗重

ベイズの法則  
事後分布は  
尤度×事前分布  
に比例する



トーマス・ベイズ牧師

(一七〇一—一七六一)

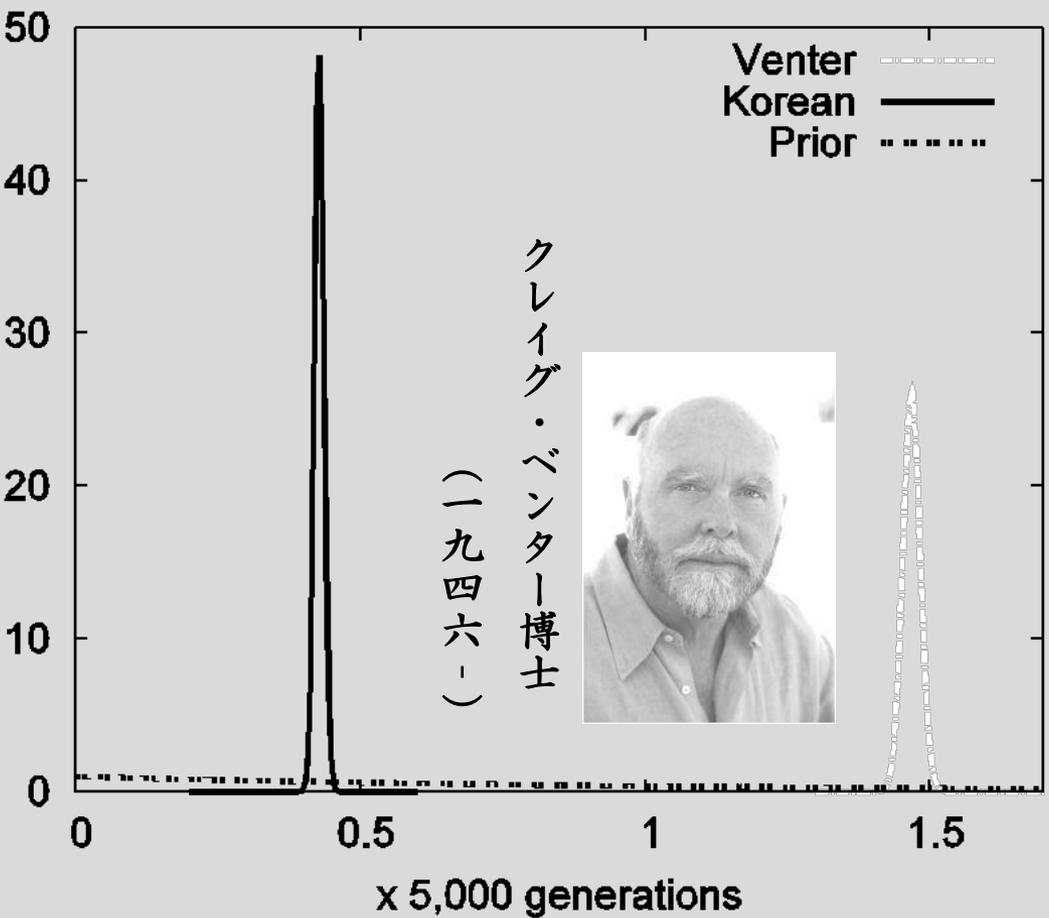
出典 ウィキペディア

事前分布は指数分布、

二人の配列を較べて得られる情報は異なる箇所の数、

その尤度はポアソン分布、事後分布はガンマ分布

# 個人ゲノム配列



匿名の日本人（藤本ら、二〇一〇）、匿名の韓国人（安ら、二〇〇九）、ベンター博士（レビら、二〇〇七）のY染色体の配列データを用いた

# 近似ベイズ計算

母数の事前分布は既知として、複雑なモデルにおいて尤度を陽に求めることは難しいけれども、データが得られたときの事後推定を行いたい

## 棄却法（古典）

- 一、母数を生成し、
- 二、その母数を用いてデータをシミュレートし、
- 三、シミュレートしたデータが実データに等しければ母数を受理せよ

高次元データではほとんど受理できない

## 近似技（タバレら、符と李、一九九七）

次元削減、即ち要約統計量が等しければ受理  
許容、即ち要約統計量の乖離が許容できれば受理

## 回帰法（ビューモントラ、二〇〇二）

シミュレートしたデータを用いて母数を要約統計量に  
回帰する規則を学ぶ

# カーネル法

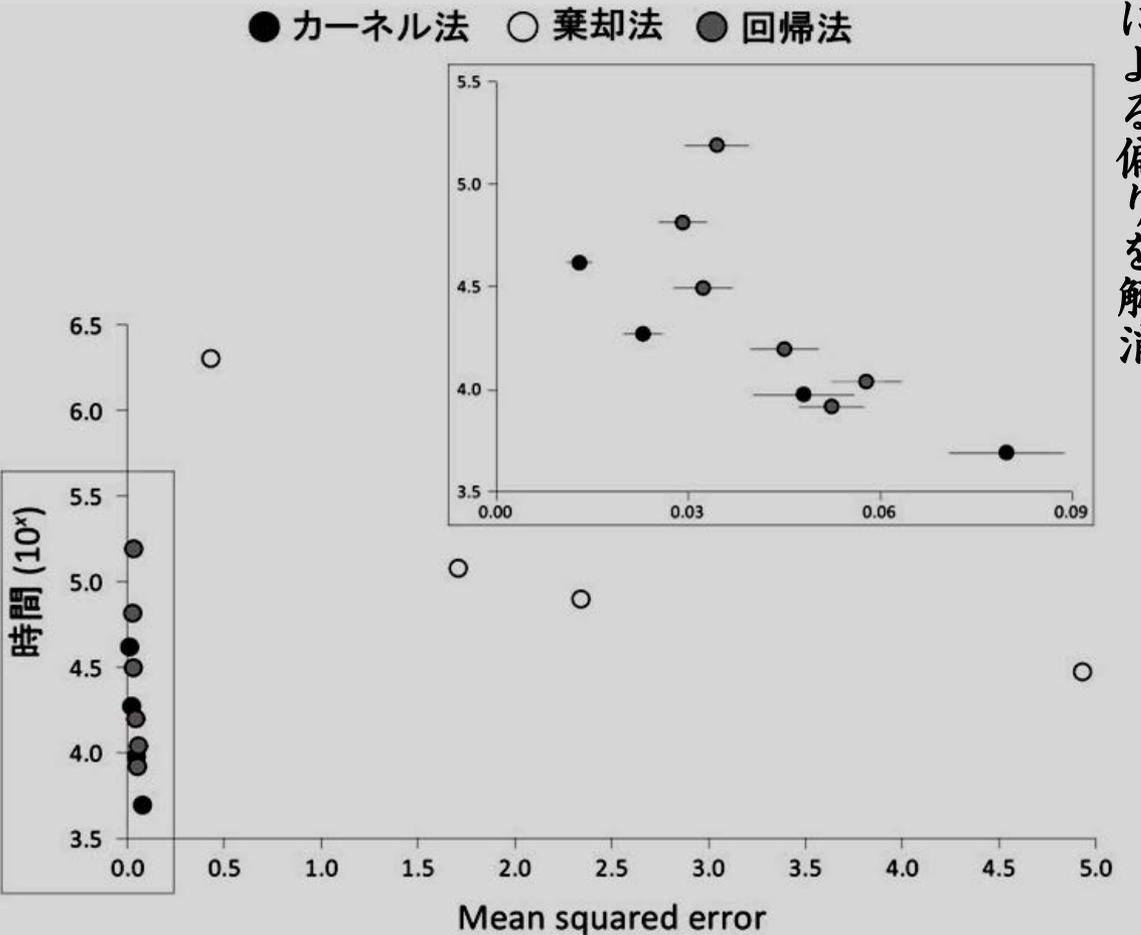
要約統計量を高次元にすれば真の事後分布に近づくが、  
受理が難しく、要約統計量の間に関連が生じる

カーネルリッジ回帰（中込、福水、間野、アーカイブにて

公表、二〇一〇）

次元削減による情報損失を緩和

許容による偏りを解消

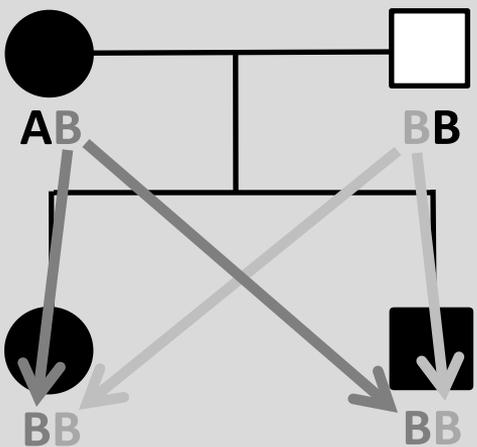


# 疾患遺伝子の探索

ポジショナル・クローニング

一九八九年、嚢胞性線維症の原因遺伝子同定において  
初成功し、ヒトゲノム配列の決定(二〇〇一)により加速

罹患同胞対解析

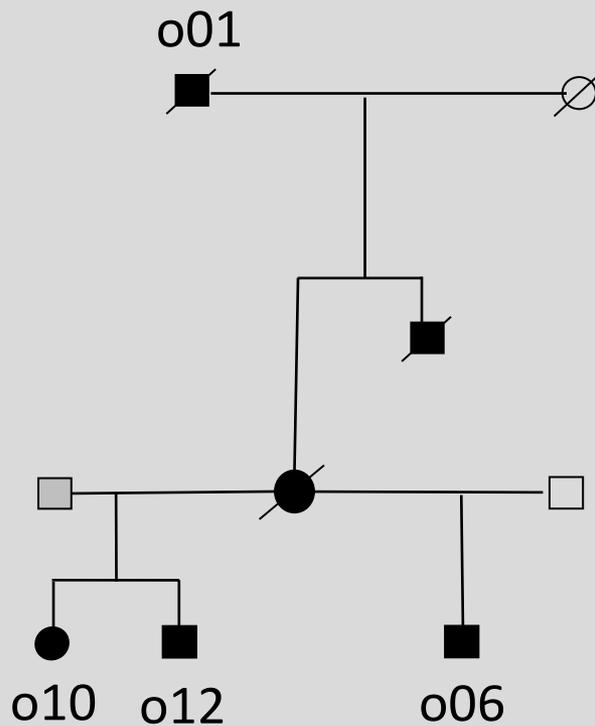


罹患した同胞において、  
先祖を共有する数が多い  
マーカーは、疾患遺伝子の  
近くにあるだろう

全ゲノム連鎖解析

ゲノムの四百近くの位置にマーカーを設定し、それぞれ  
について先祖を共有する数を推定

# 先祖の確定



西野、間野、池尾、岩田ら（準備中）

(図を削除)

# 北極熊



出典 ウィキペディア

分類学的に橪熊に極めて近く、交配が可能で、高々十五万年前の分岐と考えられてきた

今年四月のサイエンス誌において、ドイツの研究グループが、四十五個体のゲノムの十四箇所の配列を解析し、六十万年前に分岐したことを見出したと報告

# データの統合

従来説

新説



全体の共通先祖



樺熊の共通先祖

すべての箇所に有意な違いは検出されなかった

(中込、間野、長谷川、テクニカル・コメントを投稿)

# 結語

一、究極のデータ

一、近似と捨象

一、モデリングの変容

一、推定していた事象を確定

一、統合を誤ると大きな間違い

# 謝辞

国立病院機構 東京医療センター 分子細胞生物学研究部

岩田岳 部長

国立遺伝学研究所 生命情報・DDBJ研究センター

池尾一穂 准教授

統計数理研究所 数理・推論研究系 福水健次 教授

統計数理研究所 統計思考院・中華人民共和国 復旦大学

生命科学学院 長谷川政美 教授

(写真を削除)

日本学術振興会特別研究員(統計数理研究所)

中込滋樹 博士

国立遺伝学研究所 生命情報・DDBJ研究センター

特任研究員 西野穰 博士