

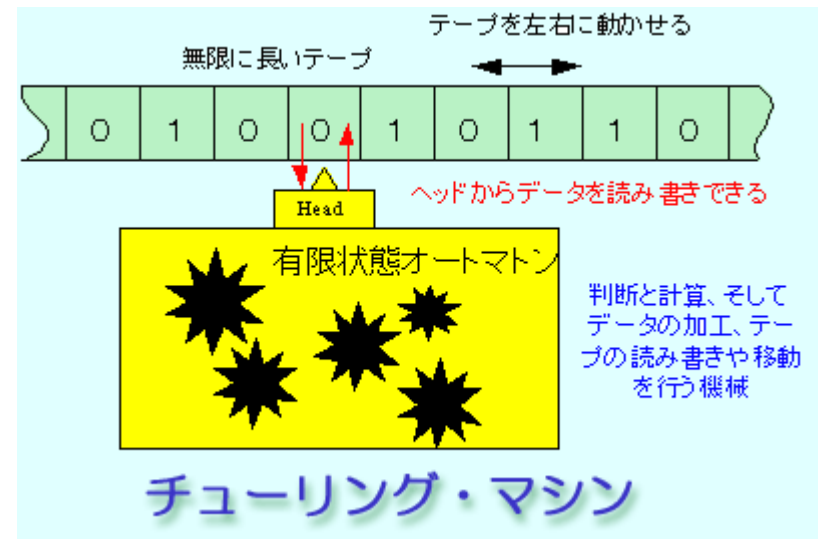
統計遺伝学 講義

- **Sep 26** (Mon, 13:00-15:00) Overview of this lecture series in Japanese with basics on population genetics. (a) What does statistical genetics handle besides disease-gene mapping? (b) Introduction to “Jurg’s introduction” to statistical thinking in Japanese. (c) Disease-gene study. Mapping and others. (d) Japanese-to-English glossary for statistical genetics. (e) Topics in population genetics.
- **Sep 27** (Tue, 13:00-15:00) (a) Statistical thinking. Random variables. The concept of a statistic. Statistical estimation and test procedures. Likelihood and Bayesian approaches. Multiple testing. Bonferroni correction. False discovery rate. Independent and dependent tests. *Note:* This is all very basic, so don’t get frightened! (b) Mendelian inheritance. Quantitative and qualitative phenotypes. Genetic determinism, heritability, Risch’s I_s statistic.
- **Sep 27** (Tue, 15:00-17:00) Genetic variations (1): Genetic linkage analysis. Pedigree analysis. Allele sharing statistics. Sibpair analysis, concordant and discordant sibpairs. Multiple affected offspring. Equivalence with recessive inheritance. Whittemore statistic. Binary traits and QTLs. The Haseman and Elston approach. Age at disease onset.
- **Sep 28** (Wed, 13:00-15:00) Genetic variations (2): Genetic association analysis (linkage disequilibrium). Case-control and family-based control (TDT, FBAT) designs. Genetic variants causing phenotypic variations. Haplotype (= allele pattern) and genotype pattern frequencies. “Common trait, common variant” hypothesis. Phylogeny of haplotypes (just briefly).
- **Sep 28** (Wed, 15:00-17:00) New methods and ideas to dissect genetic variations and their roles. Effect of environment. Replication. Proportion of true results in published disease gene associations. Role of scientist vis-à-vis public health.

統計遺伝学とバイオインフォマティクス

- 統計遺伝学
 - ベイズ
 - メンデル
 - フィッシャー
- バイオインフォマティクス(計算機科学・情報利用)
 - フォン ノイマン
 - チューリング

$$\Pr\{B|A\} = \frac{\Pr\{A \cap B\}}{\Pr\{A\}} = \frac{\Pr\{B\} \times \Pr\{A|B\}}{\Pr\{B\} \times \Pr\{A|B\} + \Pr\{\bar{B}\} \times \Pr\{A|\bar{B}\}}$$



本日の構成

- 統計遺伝学～その取り扱い対象
- 統計遺伝学的なアプローチ
- 疾患感受性遺伝子の探索と統計遺伝学
- 疾患感受性遺伝子探索において統計遺伝学的に取り扱われる生物学的要素・現象
- 統計遺伝学用語集

統計遺伝学的の守備範囲に共通することは

- 統計的有意差
- 確率的な考え方
 - ベイズ(事前確率と事後確率)
 - 尤度
- 仮説とモデルとその「ありやすさ」と適合度
 - 帰無仮説
 - その他のモデルとモデルで表現した仮説
- なぜ、遺伝学が、統計の祖なのか
 - 不確実なこと、確率的に起きる現象を対象にしたから
 - 「遺伝する」ということが「○だったら、×、×だったら」というように、ベイズ的な発想が当てはまる現象だったから

統計遺伝学

その取り扱い対象

1. 配列・構造・発現関係
2. 進化遺伝学
3. 繁殖
4. 集団遺伝学
5. 遺伝疫学・・・疾患関連遺伝子の探索を含む

統計遺伝学

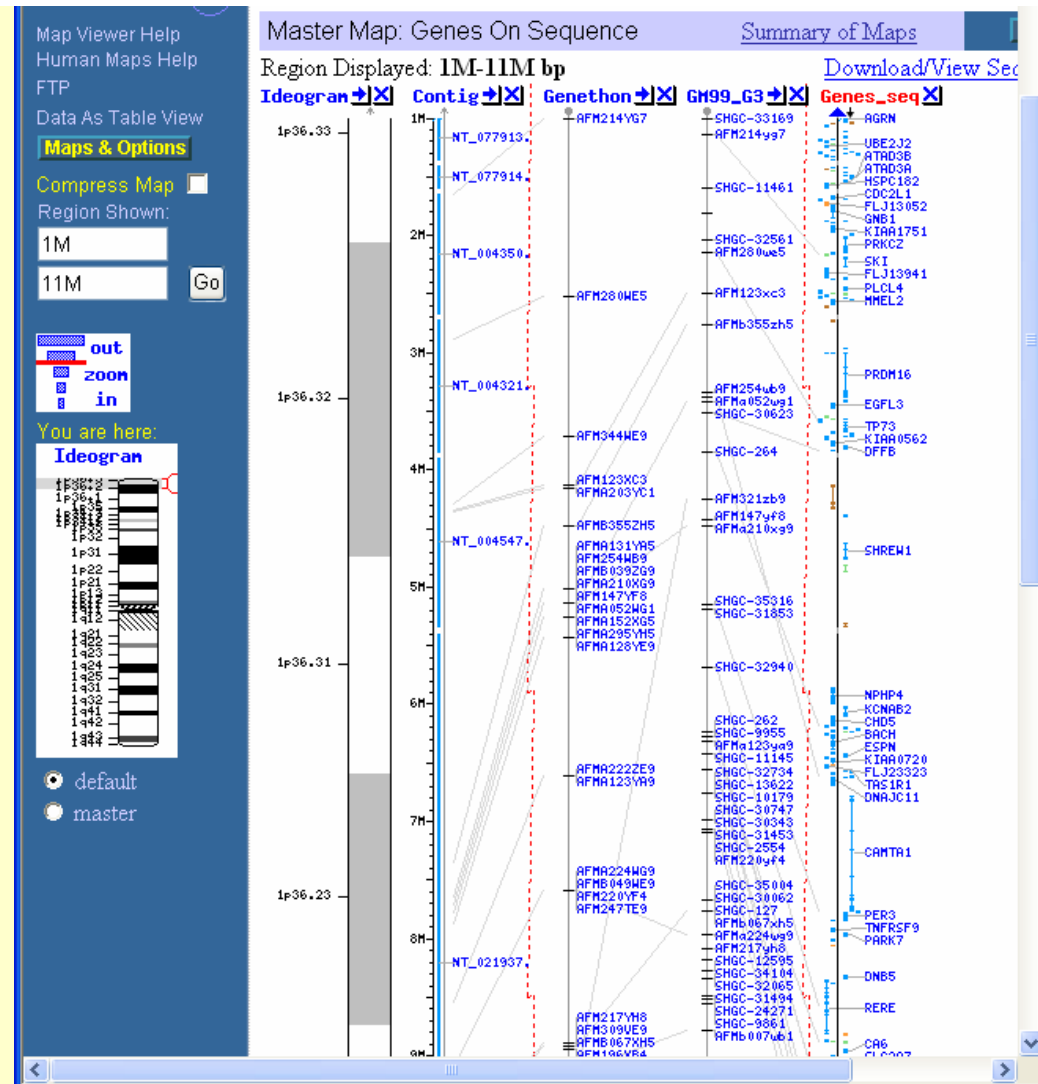
その取り扱い対象

1. 配列・構造・発現関係

- 遺伝地図の作成
- 配列比較
- 遺伝子構造予測
- 比較ゲノム学
- 体系的発現解析データの解釈
- タンパク質進化の確率モデル

統計遺伝学 その取り扱い対象

- 地図の作成
 - － 形態学上の地図
 - 染色バンド
 - － 組換え基準の地図
 - 遺伝地図
 - Radiation Hybrid Map
 - － DNA分子構造上の地図
 - 物理マップ
 - － 塩基配列塩基番号
- マーカーをそれぞれの地図上に配置
 - － マイクロサテライト
 - － SNP



統計遺伝学

その取り扱い対象

Expect = $4e-180$

配列一致度の**確率**表現

- 配列比較
 - 配列的分子とその情報
 - 核酸
 - タンパク質
 - 配列検索
 - Blastその他
 - 分子配列ホモロジー

```
>ref|NM_016233.1| UEG Homo sapiens peptidyl arginine deiminase, type III (PADI3), mRNA
Length=3183

Score = 633 bits (329), Expect = 4e-180
Identities = 827/1057 (78%), Gaps = 75/1057 (7%)
Strand=Plus/Plus

Query 1002 CTGCCATGAAAGCCAG-TGCAAGCTGACCATCTGCCCTGAGGAGGAGAACATG-GATG 1059
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 1017 CTGCCAGGAAGGCC-GGCTGCAAGCTGACCATCTGCCACAGCCGAGAAC-CGCAACG 1074

Query 1060 ACCAG-TGGATGCAGGATGAAATGGAGATCGGCTACATCCAAGCCCCACAAAAACGCTG 1118
          ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 1075 ACC-GCTGGATCCAGGATGAGATGGAGCTGGGCTACGTTACGGCGCCGCAAGACCCCT- 1132

Query 1119 CCC-GTGGTCTTCGACTCTCCAAGGAACAGAGG-CCTGAAGGAGTTTCCCAT-CAACGA 1175
          ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 1133 CCCGGTGGTCTTTGACTCCCCAAGGAA-TGGGGAAGTGCAGGA-TTCCCTTACAAAAGA 1190

Query 1176 GTGATGGGTCCAGATTTTGGCTATGTAAGTACGAGG-GCC-C-CAACAGGGGGTATCAGT 1232
          ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 1191 ATCCTGGGTCCAGATTTTGGTTACGTGACTCG-GGAACCACGC-GACA-GGTCTGTGAGT 1247

Query 1233 GGACTGGACTCCTTTGGGAACCTGGAAGTGAGCCCCCAGTCACAGTC-AGGGGCAAGGA 1291
          || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 1248 GGCTGGACTCCTTTGGGAACCTGGAGGTCAGCCCTCAGTGGTGGCCAATGGG-AAAGA 1306

Query 1292 ATACCCGCTGGGCAGGATTCTC-TTCGGGGAC-AGCTG-TTATCCCAGCAATGACAGCCG 1348
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 1307 GTACCCCTGGGGAGGATCCTCATT-GGGGCAACCTGCCTGGGTCA--AGTG---GCC- 1359

Query 1349 GCA-GATGCA-CCAGGCCCTGCAGGACTTCCT-CAGTGCCAGCAGGTGCAGGCCCTGT 1405
          ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 1360 GCAGGGT-CACCCAGGTGGTCCGGGACTTCCTCCA-TGCCCAGAGGTGCAGCCCCCGT 1417

Query 1406 GAAGCTCTATTCT-GACTGGCTGTCCGTGGGCCAGGTGGACGAGTTCTGAGCTTTGTGC 1464
          ||||| ||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 1418 GGAGCTCT-TTGTGGACTGTTGGCCGTGGCCATGTGGATGAGTTTCTGAGCTTTGTCC 1476

Query 1465 CAG-CACCCGACAGGAAGGGCTTCCGGCTGCTCCTGGCCAGCCCCAGGTCTGCTACAAA 1523
          || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 1477 CTGTC-CCCGATGGGAAGGGCTTCCGGATGCTCCTGGCCAGCCCTGGGGCTGCTTCAAG 1535

Query 1524 CTGTTCCAGGAGCAGCAGAA-TGAGGGCCACGGGGA-GGCCCTGCTGTTTCA-AGGGATC 1580
          ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 1524 CTGTTCCAGGAGCAGCAGAA-TGAGGGCCACGGGGA-GGCCCTGCTGTTTCA-AGGGATC 1580
```


統計遺伝学

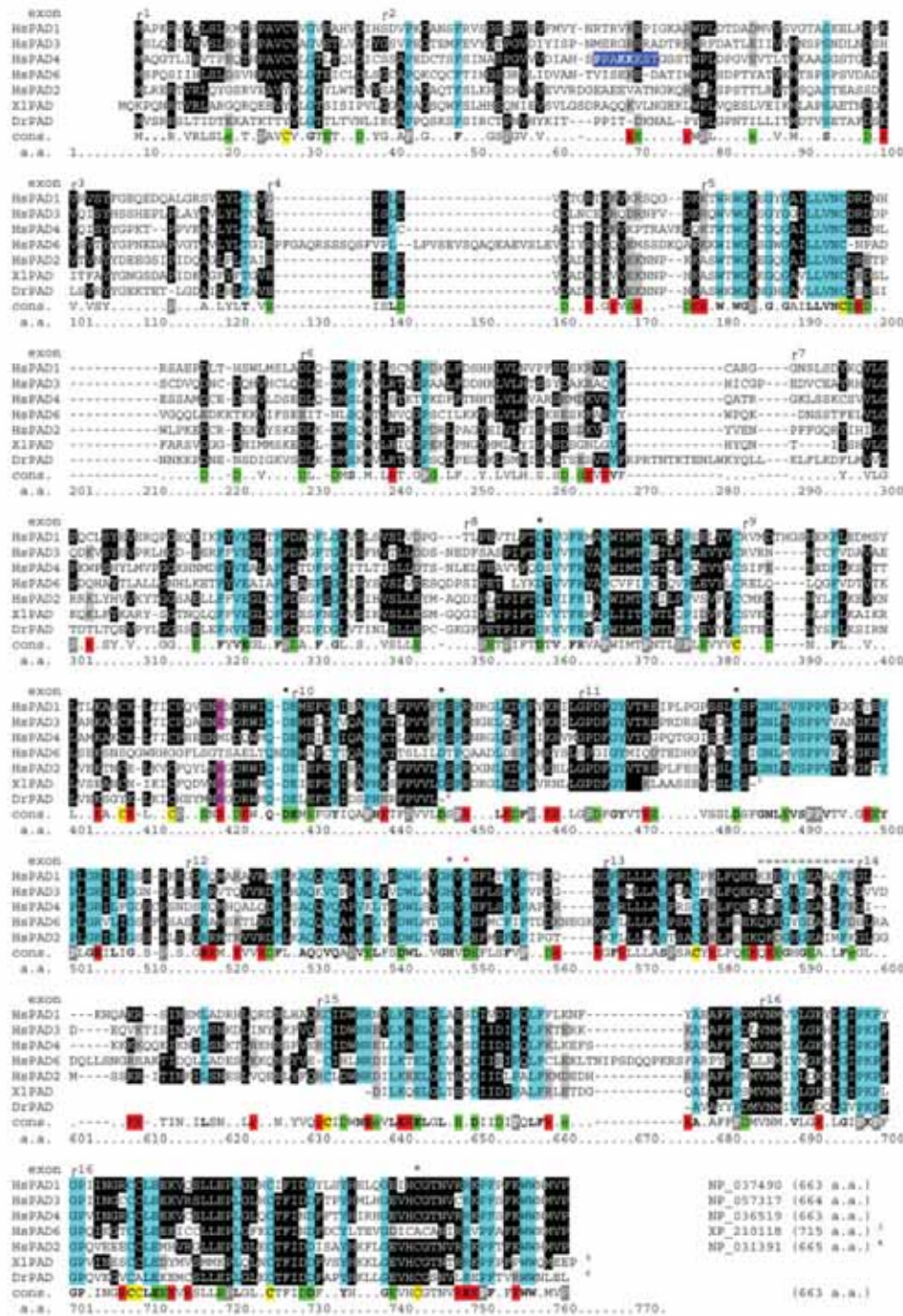
その取り扱い対象

- 遺伝子構造予測
 - エクソン
 - エクソン-イントロン境界
 - 調節配列
- それぞれの配列の特徴を抽出し、その特徴にマッチする構造を検出
 - …このプロセスで**確率**的手法を使用

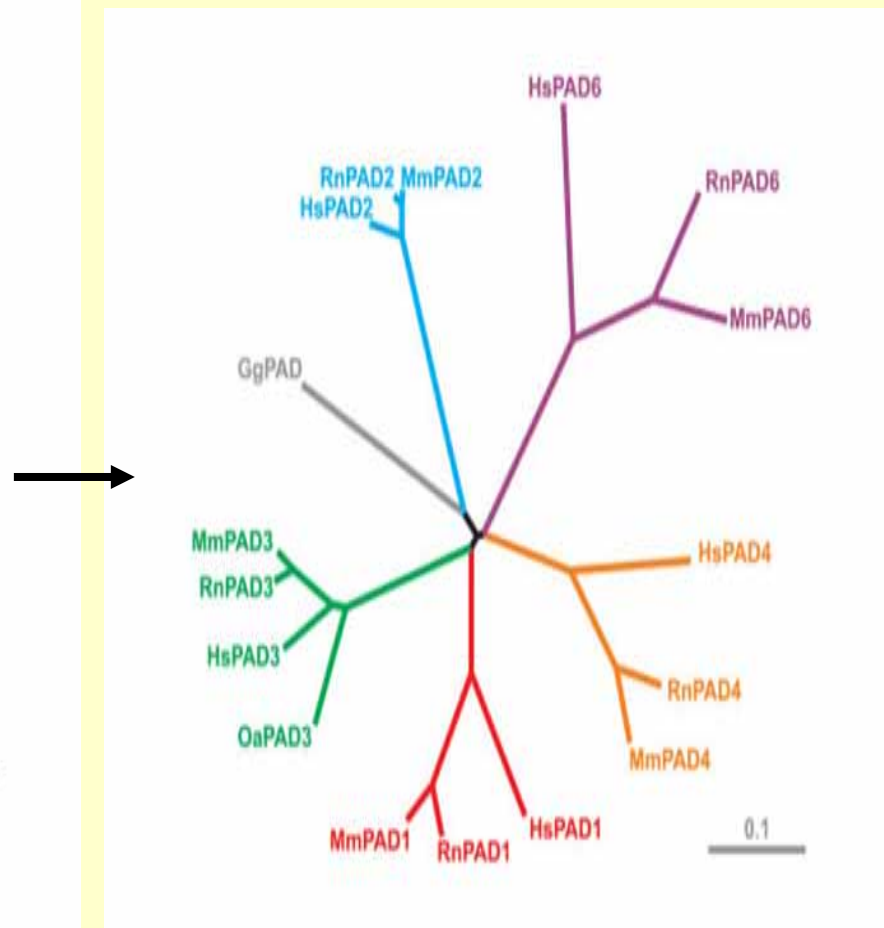
統計遺伝学

その取り扱い対象

- タンパク質進化の確率モデル
 - 1次元構造(ペプチド鎖)比較によるホモロジー
 - アミノ酸同士の化学的近似度の評価と組み込み
 - 2次元構造比較によるホモロジー
- ホモロジーの数値化
 - … 機能上距離の推定



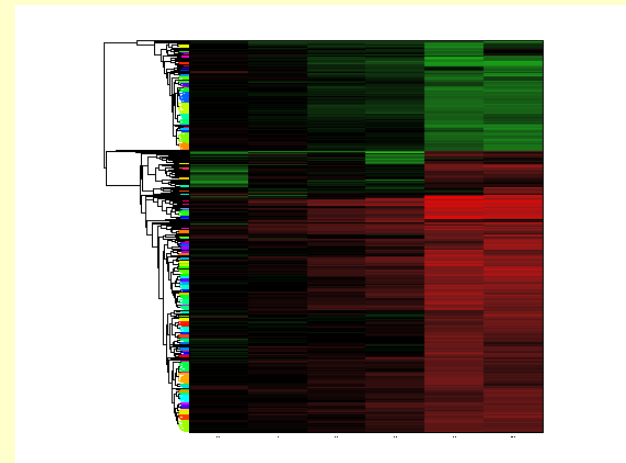
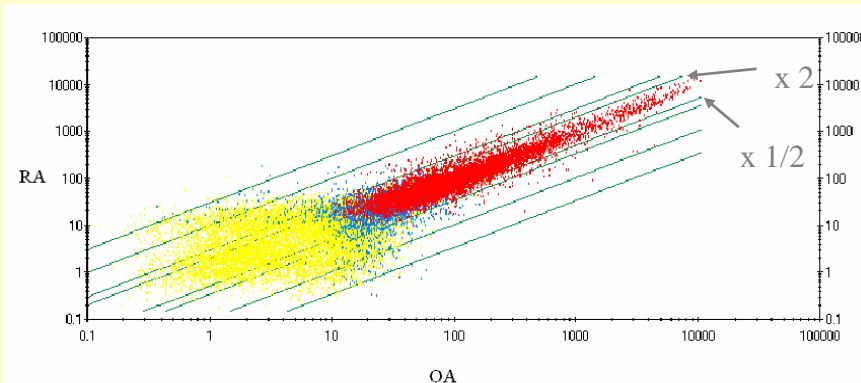
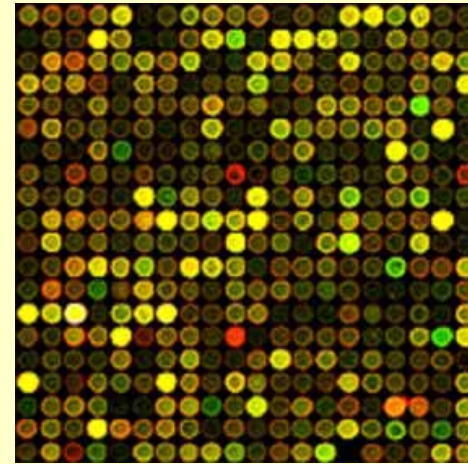
PADI酵素アミノ酸配列の配列比較と系統樹(ヒト・ラット・マウス)



統計遺伝学

その取り扱い対象

- 体系的発現解析データの解釈
 - データクオリティ評価・データ補正
 - 有意差検定・マルチプルテスト補正
 - パターン抽出・クラスタリング



統計遺伝学

その取り扱い対象

2. 進化

- 歴史の解析
- 進化に関する研究はすべて推定

統計遺伝学

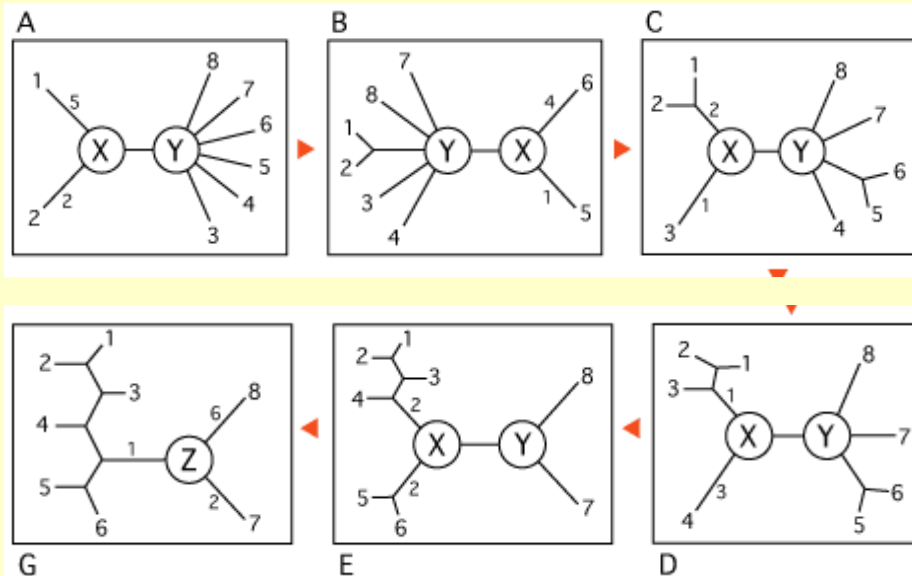
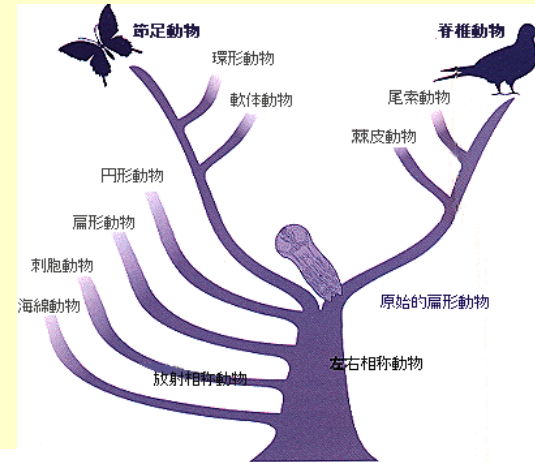
その取り扱い対象

- 進化遺伝学
 - 形態進化学
 - 分子進化学
 - ゲノム進化学
 - 系統樹学
 - 尤度
 - 距離の解釈

統計遺伝学

その取り扱い対象

- 比較ゲノム学
 - － 遺伝的距離の決定
 - － 距離の図示
 - 系統樹描図アルゴリズム



距離行列を樹に変換する方法の1つNJ法

距離行列を扱う点でクラスター解析と同様

統計遺伝学

その取り扱い対象

3. 繁殖

- 遺伝学の実学的側面
- 「進化学」における「過去推定」の反対で「実験科学」である
- 条件設定とそれに従った介入実験、その結果の観察が可能
- 動植物・微生物
- 交配
- 形質と形質関連遺伝子の探索

統計遺伝学

その取り扱い対象

- 繁殖における統計遺伝学の役割
 - 推定するのは何か
 - 交配という介入は既知
 - 推定されるべき「未知」は
 - 組換え現象
 - » 遺伝子組み換え技術の成立前は、組換え現象は介入不能だった



統計遺伝学

その取り扱い対象

4. 集団遺伝学

- 個から集団へ

- 個の研究

- 配列・発現の比較・解析では、個々の**確定**している配列・遺伝子産物とその差異を対象とした
- 進化学では、「種」を取り扱った。「種」は、集団から構成されるが、その集団は、「種」の名の下に、**均一**な存在である
- 繁殖学では、個体のかけあわせにより、個体間の差異を対象とするが、個々の個体は、**確定**した存在であった
- 均一・確定した存在の解釈にも確率・推定が必要である

統計遺伝学

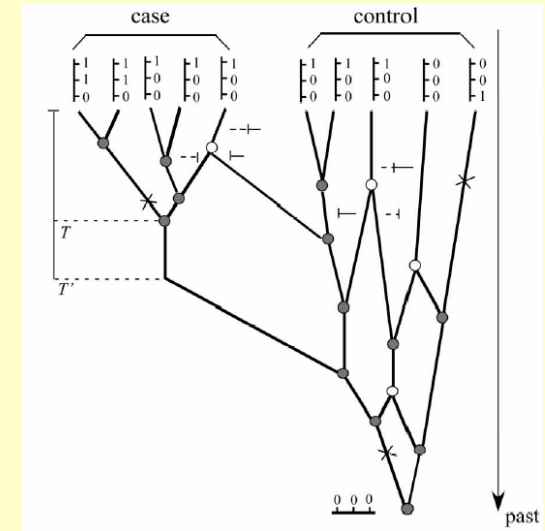
その取り扱い対象

- 個から集団へ
 - － 集団の研究
 - 集団の構成要素が**多様**であるとき、その集団について研究するときには、集団の特性を**確率**的に捉えることとなる
 - 種としてのヒトの集団を対象とするとき、集団内の**多様性**を解析する
 - － 多様集団の内部構造
 - － 多様集団の空間分布
 - － 多様集団の歴史的背景
 - － 多様集団の疫学
 - － 多様・不均一な対象の解釈には、さらに確率・推定を必要とする

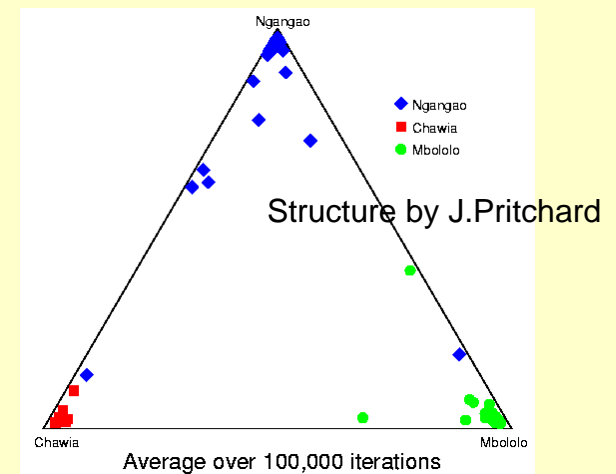
統計遺伝学

その取り扱い対象

- 集団遺伝学
 - 数学モデル
 - CoalescenceとRecombination
 - 連鎖不平衡
 - 集団比較遺伝学
 - ヒトの遺伝的歴史



Ying Wang and Bruce Rannala
Am. J. Hum. Genet. 76:000–000, 2005



統計遺伝学

その取り扱い対象

5. 遺伝疫学

- 疾患遺伝子解析と遺伝疫学
 - 個体を対象とした疾患遺伝子解析と集団を対象とした疾患遺伝子解析
 - 個体を対象
 - 家系サンプリング
 - 集団を対象
 - 集団サンプリング
 - 個体・集団のリスク評価
 - 個人に「なにか」が起こる確率の推定
 - 集団に「なにか」が起こる確率の推定
 - 「確率」を決める因子の探索
 - 遺伝因子
 - 環境因子
 - 浸透率
 - 解析手法
 - 連鎖解析
 - ノン-パラメトリック連鎖解析
 - 関連解析
 - TDT
 - これらを遂行するためのツール群にも確率的手法を用いる
 - 薬理遺伝学

統計遺伝学的なアプローチ

- 確率的な考え方

- 今、黒と白のボールが入った袋がありました。ボールは全部で2個あり、黒・白のうちわけは不明です。黒いボールが入っている個数を予想してください。

- まず、予想してみてください。
 - 袋から1個を取り出したら、それが白いボールだと、小学1年生の女の子が教えてくれました。2個のうち黒いボールは、0個、1個、2個のどれだと「賭け」ますか？
 - 袋から1個出しては戻して、を繰り返したところ、○○○ と3回続けて白ボールが出たとのことです。さて、どうしますか？
 - ○○○ となりました。さて、どうします？
 - ○○○ ○○○○○○○○ となりました。さて、どうします？
 - ○○○ ○ ○○ ○○ だったら、どうでしょうか？



トチオトメ



サチノカ○

• 確率的な考え方

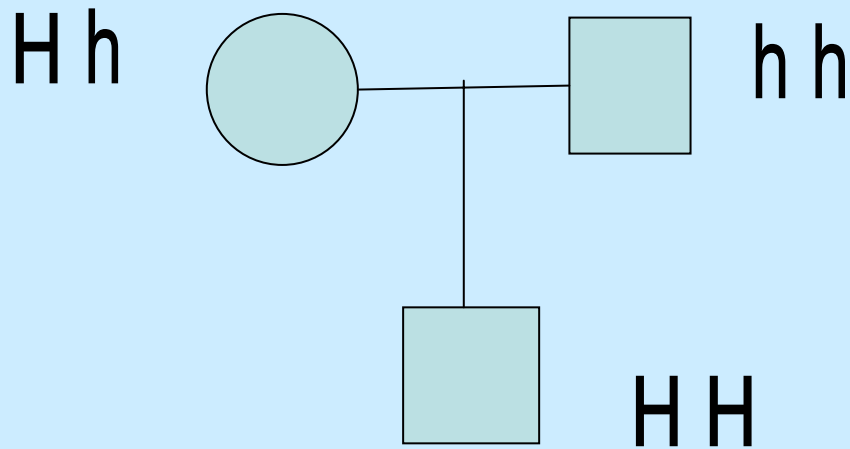
- 今、2種類のイチゴが入った袋がありました。イチゴは全部で2種類あり、トチオトメ とサチノカ○です。トチオトメとサチノカのうちの割合は不明です。トチオトメが入っている個数を予想してください。
 - まず、予想してみてください。
 - 袋から1個を取り出したら、それがサチノカ○だったとのことでした。2個のうちトチオトメ は、0個、1個、2個のどれだと「賭け」ますか？
 - 袋から1個出しては戻して、を繰り返したところ、○○○ と3回続けてサチノカ○が出たとのことでした。さて、どうしますか？
 - ○○○ となりました。さて、どうします？
 - ○○○ ○○○○○○○○ となりました。さて、どうします？
 - ○○○ ○ ○○ ○○ だったら、どうでしょうか？
 - トチオトメとサチノカのどちらを取り出したのかを教えてくれる人が、**栃木県のイチゴ栽培農家の方**だったら、何かかわるでしょうか？

- 確率的な考え方

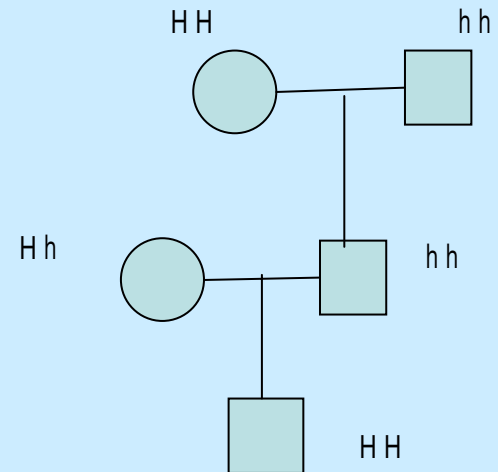
- 今、黒と白のボールが入った袋がありました。ボールは全部で1000個あり、黒・白のうちのわけは不明です。黒いボールと白いボールが500個ずつ入っているか、数に違いがあるかを知りたいとしたら、どうしますか？
 1. 全部数える
 2. どちらかが501個になるまで数える
 3. 一部を抜き取って数える
 - …不確かなことしか言えない
 - …不確かなりに、信じられればそれでよい
 - » 不確かなりに信じるための条件は何？
- 全部で1,000,000個だったら…

家系・血縁関係での

- 確率的な考え方



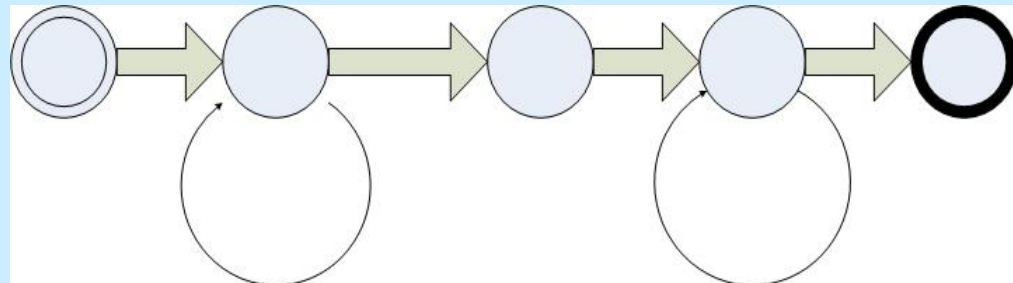
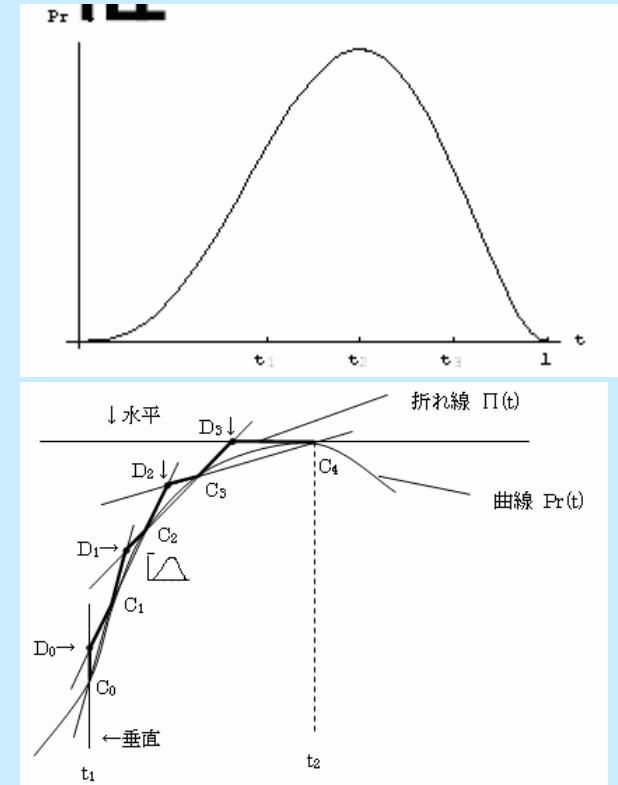
なにかがおかしい



なにかがおかしい、
けど、それは状況
によりけり

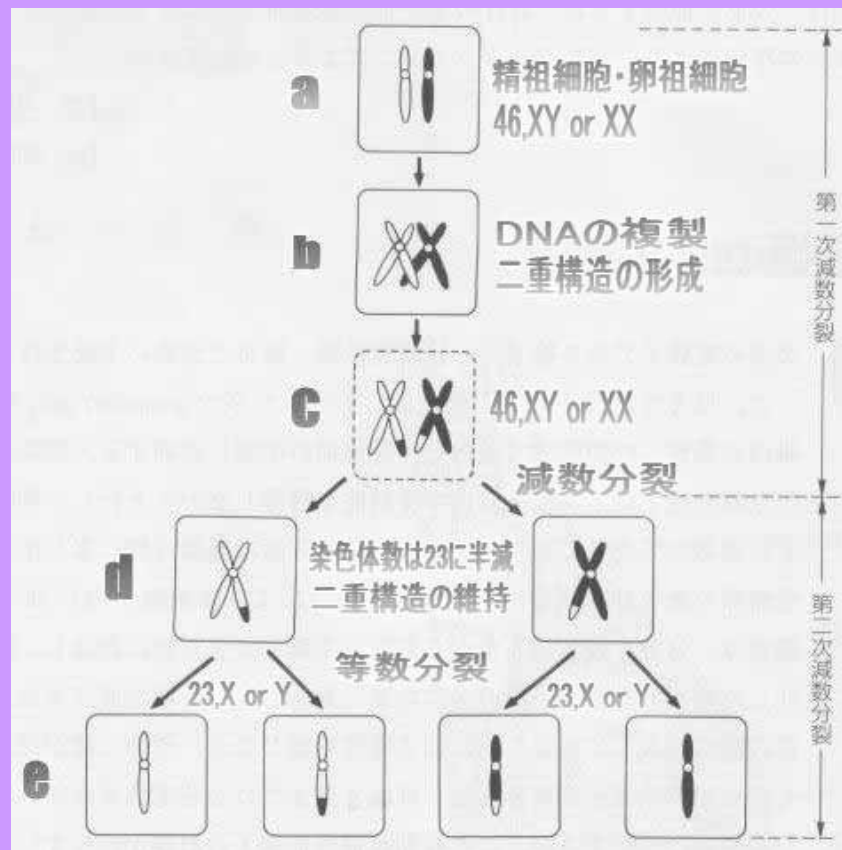
統計遺伝学的なアプローチ

- 統計的有意差
- 確率的な考え方
 - ベイズ(事前確率と事後確率)
 - 尤度
- 仮説とモデルとその「ありやすさ」と適合度
 - 帰無仮説
 - その他のモデルとモデルで表現した仮説
- 方法
 - シミュレーション (Monte-Carlo Method)
 - 最尤法
 - 隠れマルコフモデル
 - マルコフ連鎖モンテカルロ(MCMC)法
 - Metropolis-Hastings アルゴリズム
 - Gibbs Sampler



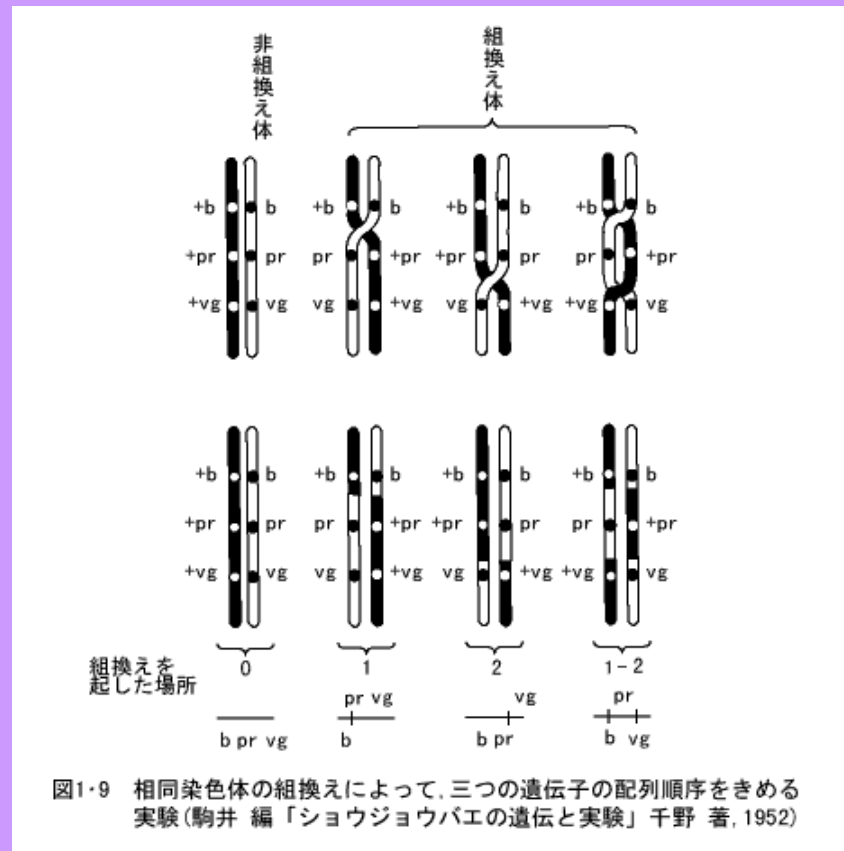
統計遺伝学用語集

減数分裂 Miosis



組換え Recombination

- 組換え



アレル Allele

- 対立遺伝子

- 染色体ごとに異なるタイプのこと
- SNPなら、AかTか
- Short tandem repeatなら5回か6回か7回か・・・
- アルコールデヒドロゲナーゼなら、高活性型か低活性型か
- SNPが作るハプロタイプなら、GCATTかCATGCかGCAGCか・・・

ハプロタイプ Haplotype

- 個々の染色体に存在するアレルの組み合わせ
 - A/T-G/T-GCの3 SNPが作るハプロタイプは
 - AGG,AGC,ATG,ATC...
 - A/TのSNPと5,6,7回リピートが作るハプロタイプは
 - A-5,A-6,A-7,T-5,T-6,T-7...

ジェノタイプ Genotype

- 遺伝子型
- 2対の染色体を持つ生物の場合には、それぞれの染色体が持つハプロタイプを合わせたものになる
 - SNPならAA AG GG
 - リピートなら5-5,5-7,6-7...
 - ハプロタイプならGCATT/ GCATT、 GCATT/GCAGC、...

ホモとヘテロ

Homozygotes and Heterozygotes

- 2倍体が、2つの同一ハプロタイプを持つ場合がホモ、異なるハプロタイプを1つずつ持つ場合がヘテロ

フェノタイプ Phenotype

- 形質・表現型

- 構造・機能上の分類

- 生物の特徴はすべてPhenotype

- 疾患、その亜分類

- 高い・低い・大きい・小さい・強い・弱い

- 質的形質と量的形質

- メンデル遺伝形質

- 量的形質(Quantitative traits)

- » QTLの解析には独自のアプローチをする

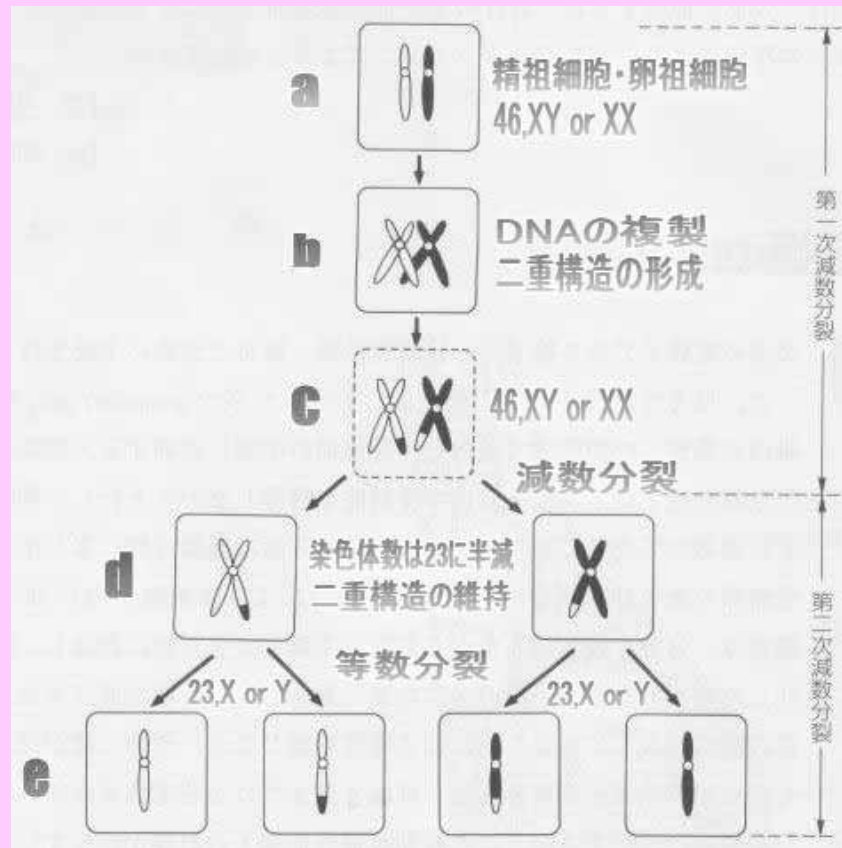
Hardy-Weinberg 平衡

Hardy-Weinberg Equilibrium

- ある集団の2倍体が持つハプロタイプの組み合わせの分布が、その集団のハプロタイプ頻度からのランダムな組み合わせにしたがっていること
 - 集団が均一で、ランダムメーティングがなされている場合に成立する

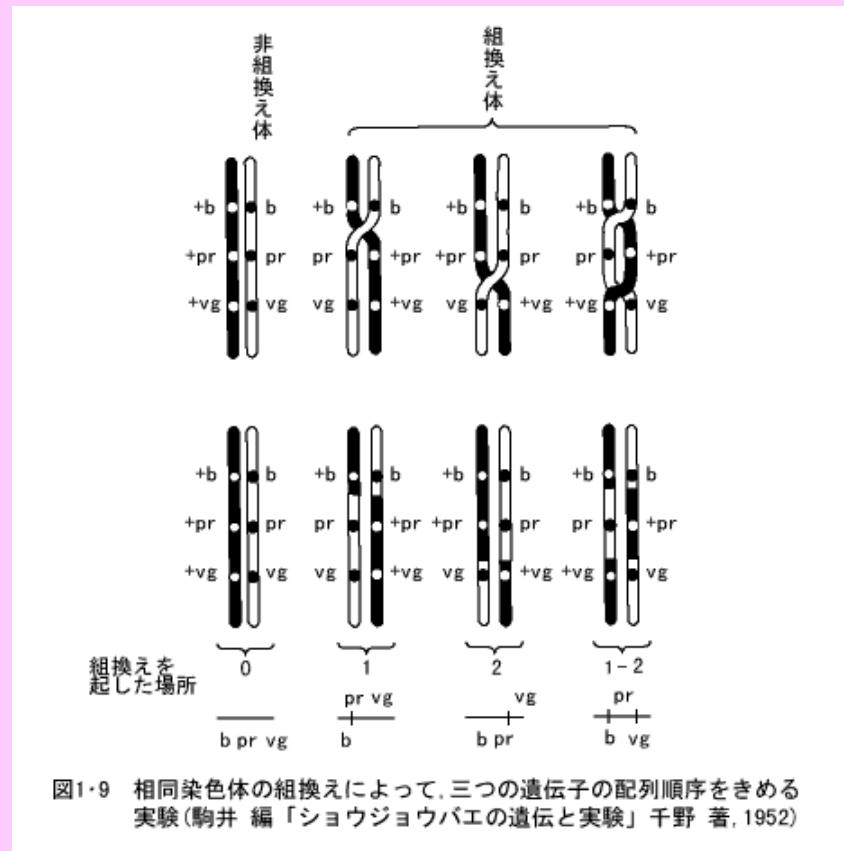
統計遺伝学用語集

減数分裂 Meiosis



組換え Recombination

- 組換え



アレル Allele

- 対立遺伝子

- 染色体ごとに異なるタイプのこと
- SNPなら、AかTか
- Short tandem repeatなら5回か6回か7回か・・・
- アルコールデヒドロゲナーゼなら、高活性型か低活性型か
- SNPが作るハプロタイプなら、GCATTかCATGCかGCAGCか・・・

ハプロタイプ Haplotype

- 個々の染色体に存在するアレルの組み合わせ
 - A/T-G/T-GCの3 SNPが作るハプロタイプは
 - AGG,AGC,ATG,ATC...
 - A/TのSNPと5,6,7回リピートが作るハプロタイプは
 - A-5,A-6,A-7,T-5,T-6,T-7...

ジェノタイプ Genotype

- 遺伝子型
- 2対の染色体を持つ生物の場合には、それぞれの染色体が持つハプロタイプを合わせたものになる
 - SNPならAA AG GG
 - リピートなら5-5,5-7,6-7...
 - ハプロタイプならGCATT/ GCATT、 GCATT/GCAGC、...

ホモとヘテロ

Homozygotes and Heterozygotes

- 2倍体が、2つの同一ハプロタイプを持つ場合がホモ、異なるハプロタイプを1つずつ持つ場合がヘテロ

フェノタイプ Phenotype

- 形質・表現型

- 構造・機能上の分類

- 生物の特徴はすべてPhenotype

- 疾患、その亜分類

- 高い・低い・大きい・小さい・強い・弱い

- 質的形質と量的形質

- メンデル遺伝形質

- 量的形質(Quantitative traits)

- » QTLの解析には独自のアプローチをする

Hardy-Weinberg 平衡

Hardy-Weinberg Equilibrium

- ある集団の2倍体が持つハプロタイプの組み合わせの分布が、その集団のハプロタイプ頻度からのランダムな組み合わせにしたがっていること
 - 集団が均一で、ランダムメイティングがなされている場合に成立する

連鎖 Linkage

- 同一染色体上に存在する2つのローカスは親から子へ伝わるときに、一緒に伝わる。これは、2つのローカスが異なる染色体上にあるときには見られない現象である。
- このような2遺伝子が遺伝上、挙動を共にするときに、2遺伝子は連鎖していると言う。
- したがって、ある遺伝子とある遺伝子は、連鎖の関係にあるか、独立の関係にあるかのどちらかに属する。

連鎖不平衡

- 相互に連鎖の関係にある2つの多型の間に存在しうる関係。
- ある2つの多型が同一の染色体上にあるとき、この2つの多型は「連鎖」している。
- この2多型間で組換えが起きなければ、2多型のアレルの組み合わせは1つしかなく、片方の多型のアレルがわかればもう片方のアレルも自動的に決まる。
- この2つの多型が発生して後、時間が経過するにしたがって、2点間には、組換えが繰り返し起きる。
- 「十分に多数回」組換えが起きると、片方のアレルがわかっててももう片方のアレルは全く予想できない。
- 片方のアレルからもう片方のアレルの予想がつく関係を「連鎖不平衡」の関係にある、といい、予想が全くつかない関係を「連鎖平衡」の関係にある、という。
- 連鎖不平衡は、「連鎖」関係にある多型において、その間におきた組換えの歴史を反映したものとなる。また、「組換えの歴史」は個々の染色体が持つものではなく、集団が持つものであるから、「連鎖不平衡」はある「集団」の「2つの多型」の間に存在するものである。

連鎖マーカーのフェーズ

Phasing markers in linkage

- 相互に連鎖の関係にある多型マーカーのアレルは、ある組み合わせで2親から伝達されてきている。
- この組み合わせ決めること。
- ハプロタイプを実験的に決めることもできるが、個々のマーカーのジェノタイプを実験で決定し、その上で、伝達されたアレルの組み合わせを決めることの方が労力が少ない。
 - 親子のジェノタイプが得られればより高率に組み合わせが確定する。確定しない場合には、推定することになる。
 - 親子の情報がない場合には、推定することになる。

多型 Polymorphism

- 同一種内で、染色体ごとに異なるDNA塩基配列
- 近い用語 Variant, Mutation

LODスコア LOD score

- LOD: Logarithm of Odds
- オッズ比の対数値
 - 連鎖解析に限らず、オッズ比の対数値をとることは可能である。
 - 対数なので、LOD score=3ということは、 $P=0.001(10^{-3})$ に相当する。
 - 連鎖解析において、あるマーカーが、疾患の原因ローカスと連鎖していれば、特定のアレルが伝達していることと、発病との間に関係が存在し、連鎖していなければ、関係ないはずである。連鎖しているという仮定と、連鎖していないという仮定のどちらがより確からしいかを観測データに照らして計算し、そのオッズ比を取り、それを対数化して評価することで、疾患ローカスの検出を行うのが連鎖解析である。

モルガン・センチモルガン

M, cM

- 遺伝的距離の単位

- 遺伝子解析にあたっては、距離を物理的距離(塩基数)で知ることでも大事だが、組換えが頻繁に起きるか稀にしか起きないかも重要である。特に連鎖解析・連鎖不平衡解析においては、連鎖していること・連鎖不平衡関係にあることが、ローカス検出・ローカス解析の基礎であり、組換えベースの距離が必要である。
- 1Mとは減数分裂1回あたり、1回の交差を起こす距離であり、1cMとは、減数分裂100回あたり、1回の交差を起こす距離に相当する。
- ヒトでは、男女・ゲノム全体での平均が1.22cM/Mb

- (The Sequence of the Human Genome by J. Craig Venter et al., Science)

罹患同胞対 Affected-sib pair

- ある疾患に兄弟姉妹(sib)で複数名の発症をしているようなsibのペアのこと
- ペネトランスの低い疾患においては、大家系は得られにくく、それよりは、罹患同胞対を数多く集め、そのペアを用いて、疾患発症に連鎖関係にあるマーカーを検出することで疾患ローカスを同定する手法に多用される

Transmission Disequilibrium Test

- 伝達不平衡テスト

- 疾患ローカスと連鎖しているマーカーは、親から罹患者へと高確率で伝達されているはずであり、その原理を利用した解析。
- 子の保有アレルと親のそれとを比較し、罹患者に多いアレルが疾患関連アレルとなる。
- 親をコントロールとして用いることで、「内部コントロール」を用いていることになり、適切な一般集団コントロールが得にくい場合(集団が不均一な場合など)でも実施が可能な利点がある。

マルチプルテストとP-値の補正

Multiple Testing and Correction of P-value

- 1回試すのと、100回試すのとでは、意味が違
う

疾患感受性遺伝子の探索

- 疾患形質と関係のある遺伝子の探索とその生物学的根拠の解明
 - 遺伝コード情報の差異の探索
 - 個体・状態における遺伝子の働きの差異の探索

疾患関連遺伝子とは

全遺伝子

疾患感受性遺伝子
(DNAに刻まれたもの)

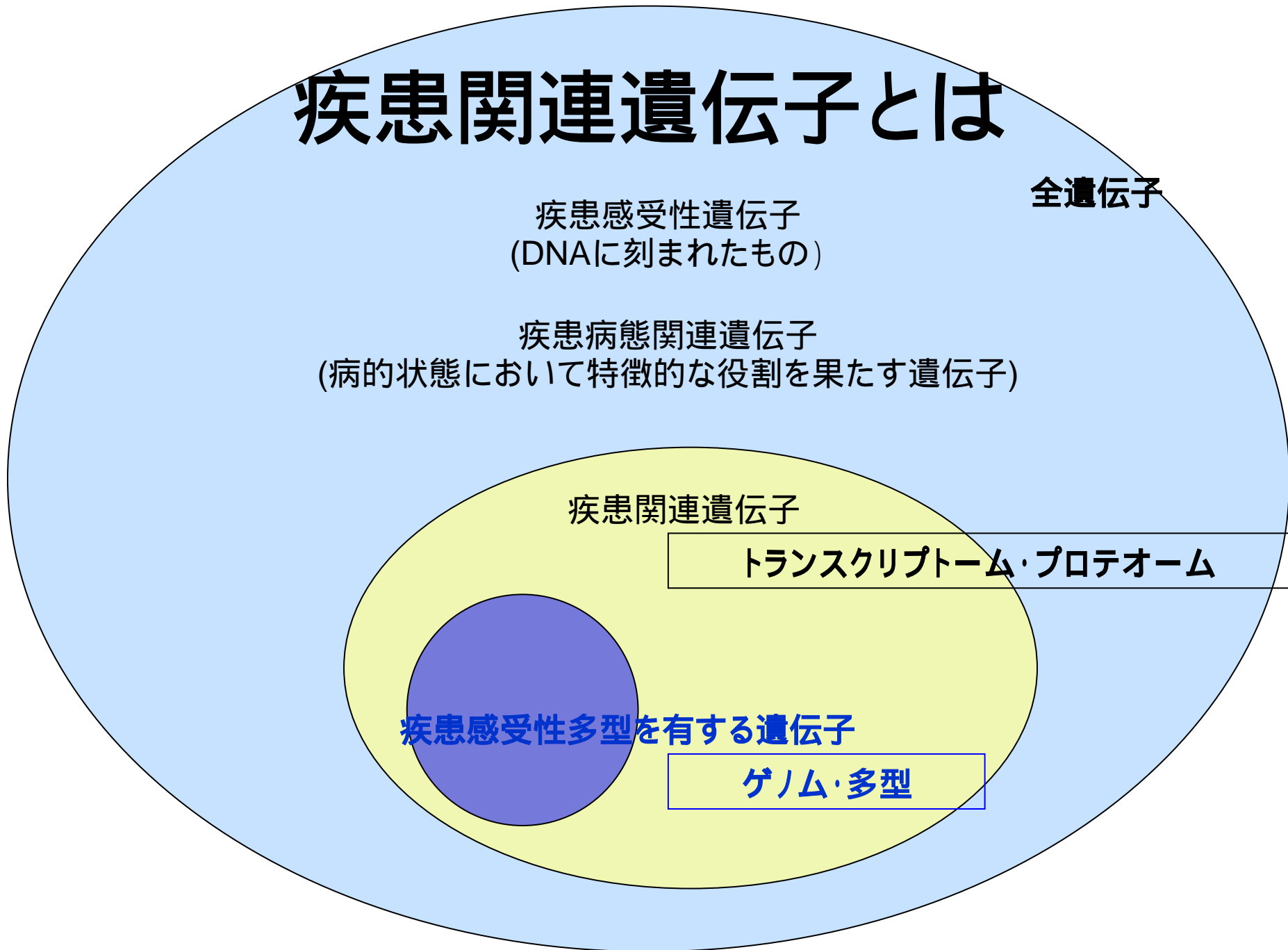
疾患病態関連遺伝子
(病的状態において特徴的な役割を果たす遺伝子)

疾患関連遺伝子

トランスクリプトーム・プロテオーム

疾患感受性多型を有する遺伝子

ゲノム・多型



疾患遺伝子マッピング

- ゲノム探索をするための、地図の作成
- 地図には、多型マーカーを配置
- 広範囲スクリーニングデータからの候補領域同定
- 領域のFine mapping
- 標的領域の原因バリエーション探索
- 複数多型の複合効果
 - 同一遺伝子内複数多型
 - 遠位ロカス間の複合効果

連鎖不平衡マッピングとは

遺伝子多型を用いた、
ケース・コントロール解析と、
ケース内フェノタイプ別関連解析は、
すべて
「連鎖不平衡マッピング」
の一部です

連鎖不平衡マッピングとは

集団 と 集団 との比較

遺伝子多型を用いた、
ケース・コントロール解析と、
ケース内フェノタイプ別関連解析は、
すべて
「連鎖不平衡マッピング」
の一部です

ゲノムDNAの解析

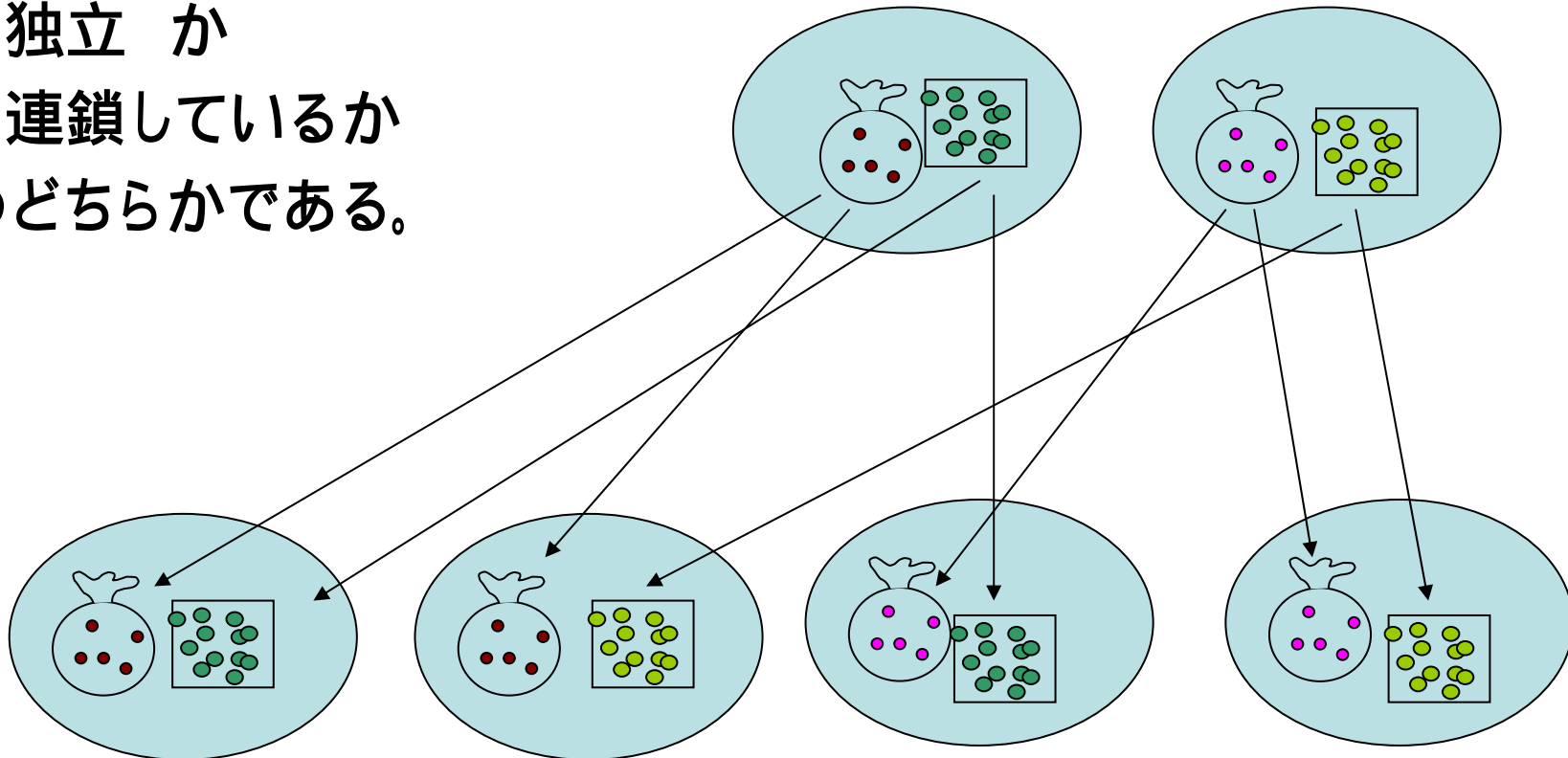
内容

- 連鎖不平衡マッピングの基礎
 - 連鎖と連鎖不平衡
 - 連鎖不平衡マッピングの方法
 - ゲノム地図
 - ゲノムスキャンと候補領域同定
 - 候補領域のFine mapping
 - 標的領域の原因バリエーション探索
 - 複数多型の複合効果

連鎖と独立



- 形質は遺伝する
- 形質Aと形質Bの伝わり方は、
 - 独立 か
 - 連鎖しているかのどちらかである。



連鎖 Linkage



- 同一染色体上に存在する2つのローカスは親から子へ伝わるときに、一緒に伝わる。これは、2つのローカスが異なる染色体上にあるときには見られない現象である。
- このような2遺伝子が遺伝上、挙動を共にするときに、2遺伝子は連鎖していると言う。
- したがって、ある遺伝子とある遺伝子は、連鎖の関係にあるか、独立の関係にあるかのどちらかに属する。

連鎖の束の単位

- 「多数ある遺伝子は、それぞれが“ばらばら”に存在しているわけではなく、(ヒトならば)23個に束ねられている」



連鎖 + 組換え



- 「多数ある遺伝子は、それぞれが“ばらばら”に存在しているわけではなく、(ヒトならば)23個に束ねられている」
- 「束」に変化がなければ、遺伝解析は23個を相手にすればよいのだが...

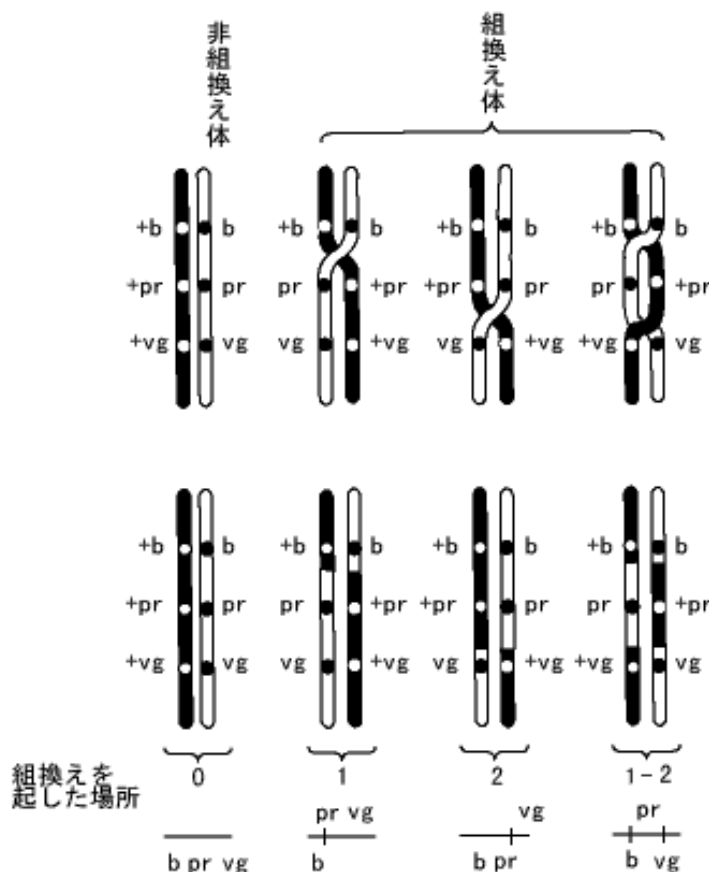
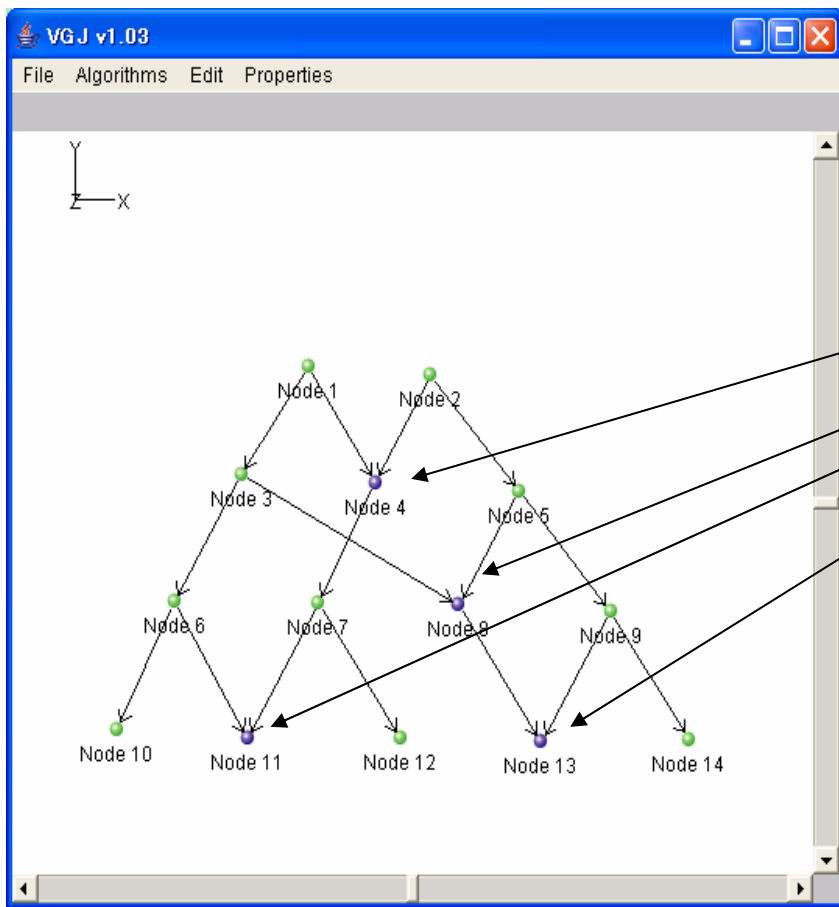


図1・9 相同染色体の組換えによって、三つの遺伝子の配列順序をきめる実験(駒井 編「ショウジョウバエの遺伝と実験」千野 著, 1952)

組換えは起きたり起きなかったり



組換え体

組換え

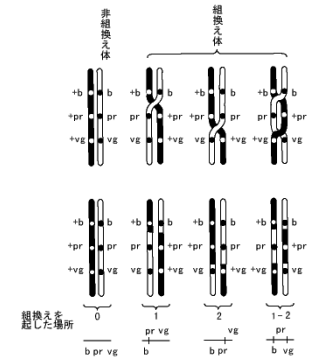


図1-9 相同染色体の組換えによって、三つの遺伝子の配列順序をきめる実験(駒井 編「ショウジョウバエの遺伝と実験」千野 著, 1952)

- ヒトでは、男女・ゲノム全体での平均が
1.22cM/Mb
 - (The Sequence of the Human Genome by J. Craig Venter et al., Science)
- ヒトゲノム30億塩基対(3000Mb)
- 3000Mb ->36.6 組換え/1減数分裂
 - $1.22 \times 10^{-2} \times 3000 = 36.6$

連鎖不平衡とは

「連鎖」が
「平衡」に
達していない

いつかは「昔」からの連鎖は崩れて、「平衡」に
達する

世代を積み重ねると連鎖が途切れる

- 組換えがおきることにより、同一染色体上の遺伝子は、「必ず」一緒に伝達されるわけではない。

「必ず」一緒に伝達される遺伝子セットは？

組換えが起きても分断されないこと

歴史の結果

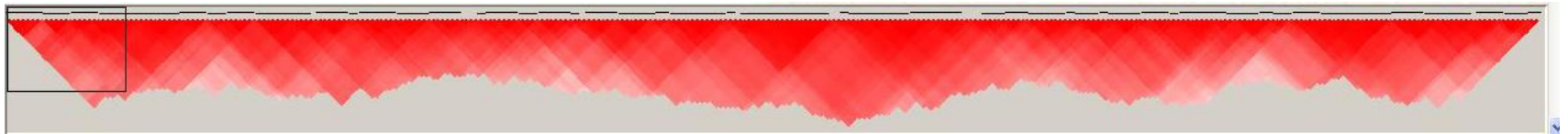
…連鎖不平衡

連鎖不平衡

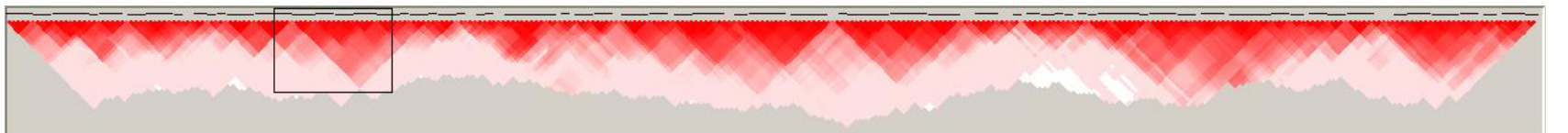
- 家系(2-5代)ではなく、集団(>数十代)を対象にするときには、連鎖不平衡ブロックが「伝達」追跡の単位

スタート2人、交差率 $1/10^8$ bp回、1000SNP、定常5000人(10000hap)、出力1000人、変異率 $1/10^6$ bp回

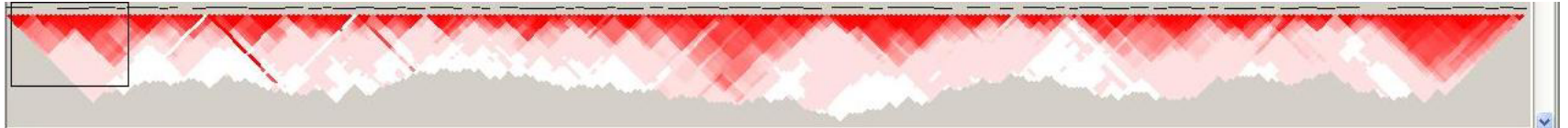
100世代



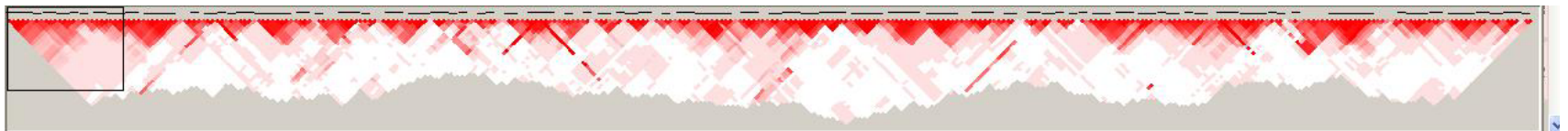
300世代



500世代



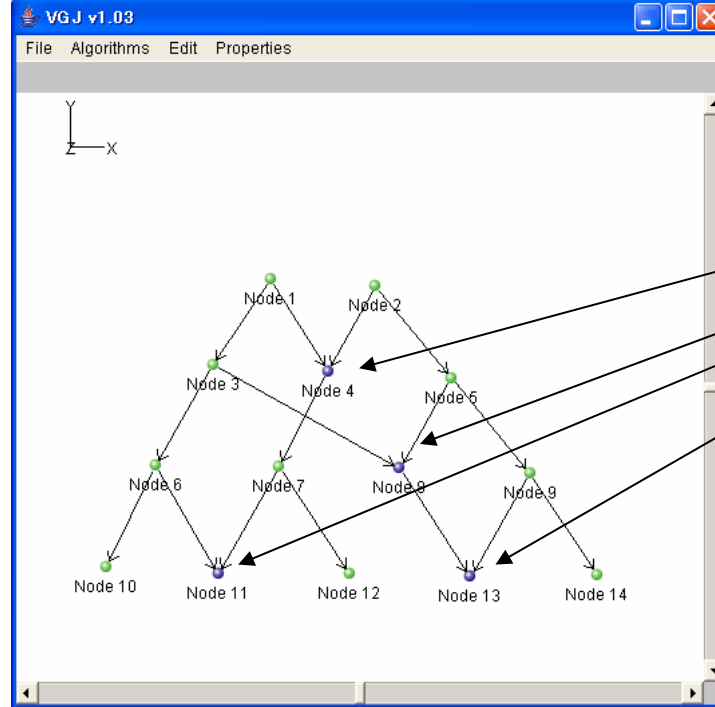
1000世代



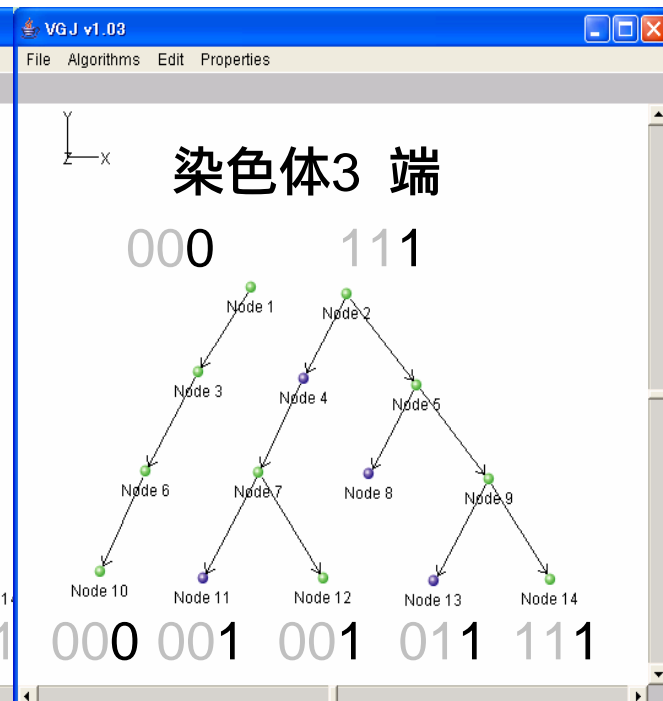
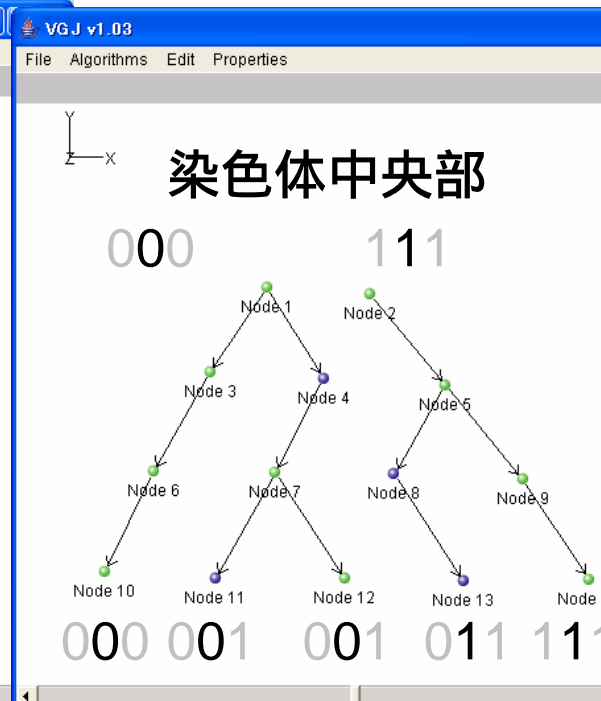
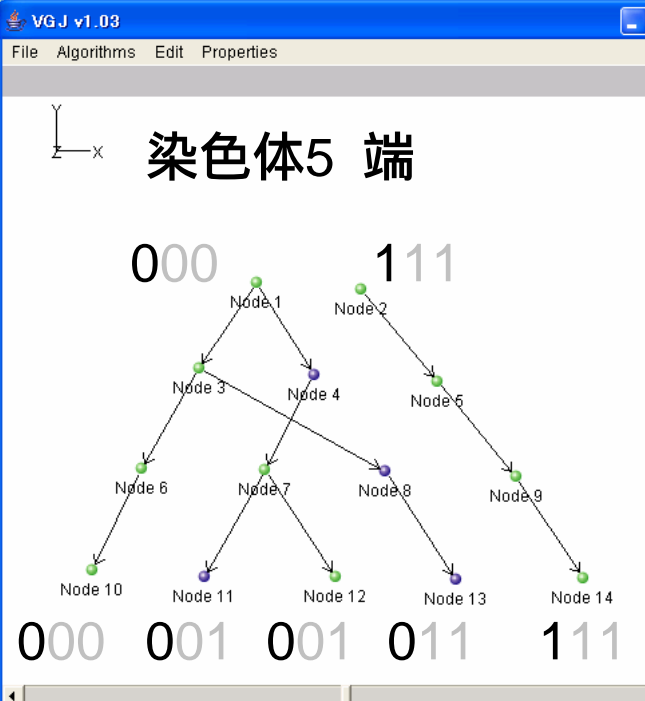
Mat>5%のみ

遺伝子の伝達の単位は、染色体ではなく、
それが分割された「連鎖不平衡ブロック」

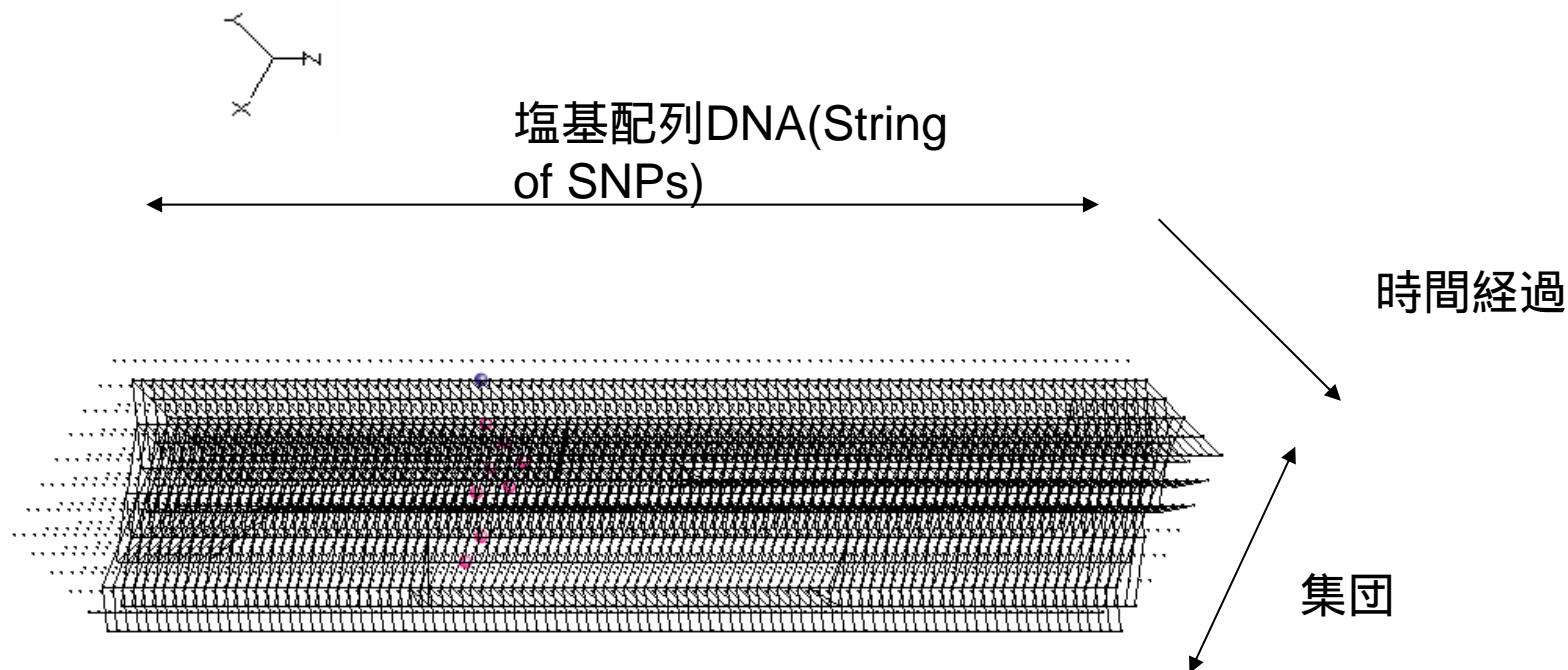
染色体全体で
考えると、親子
関係は複雑
(複数の祖先
がいる)だが、
部分的には、
単純(祖先は1
つ)



組換え体







100 SNPs

5 generations

1->2->4->7->5 chromosomes

$(0+1+1+4)=6$ recombinations

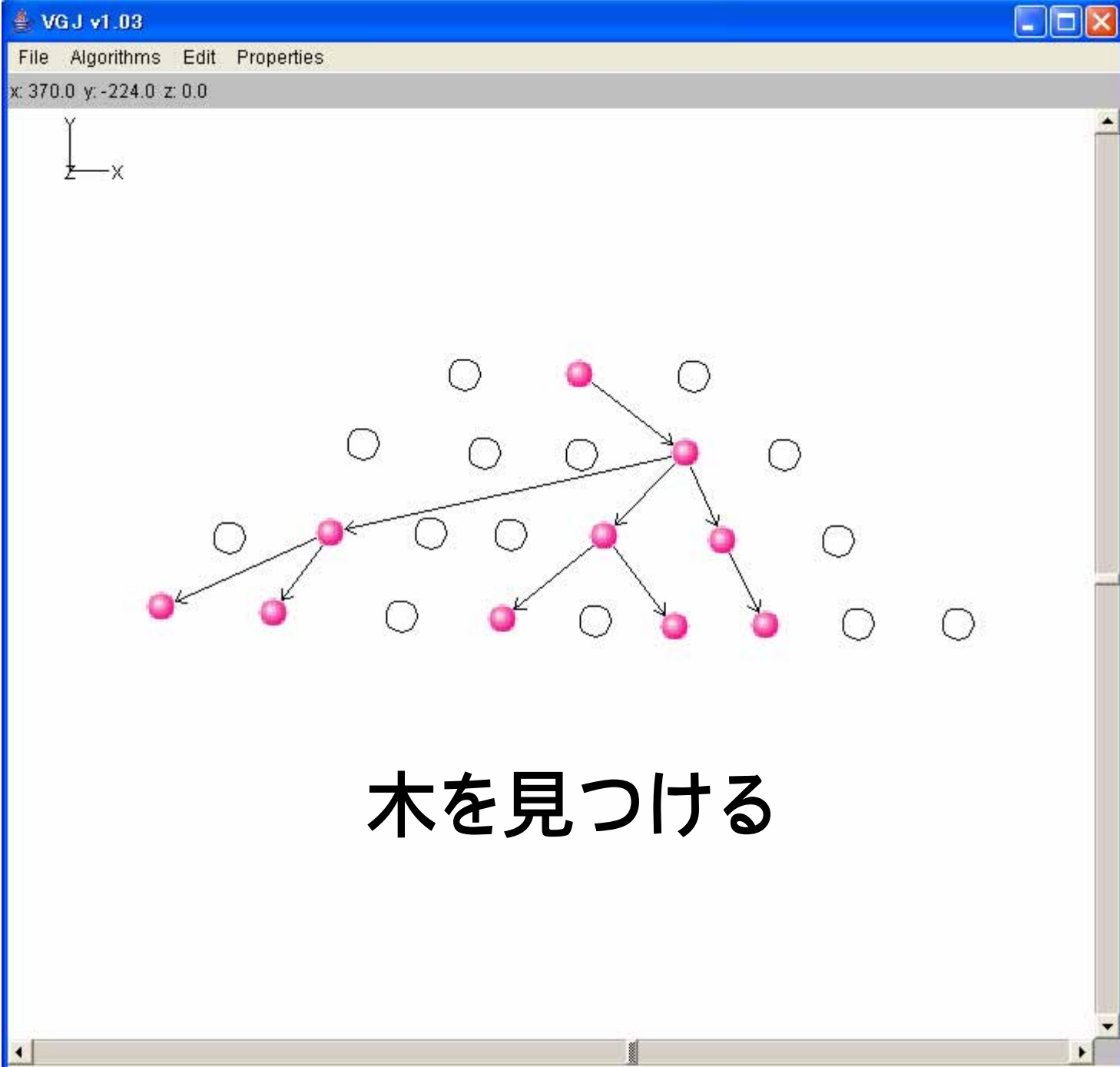
青 赤 にて感受性変異が生じた
ものとし、赤 は伝達された感受性
変異を表す

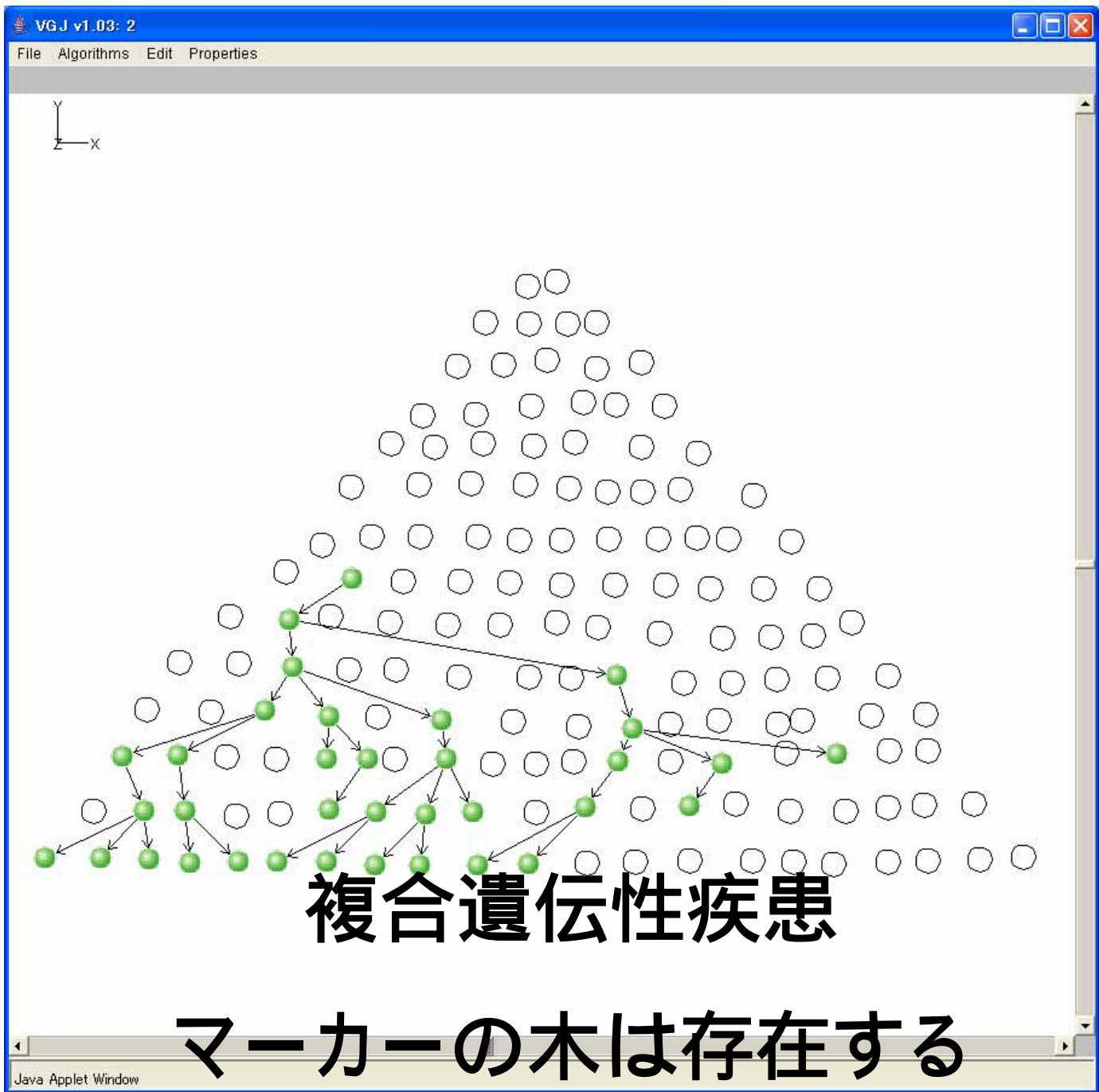


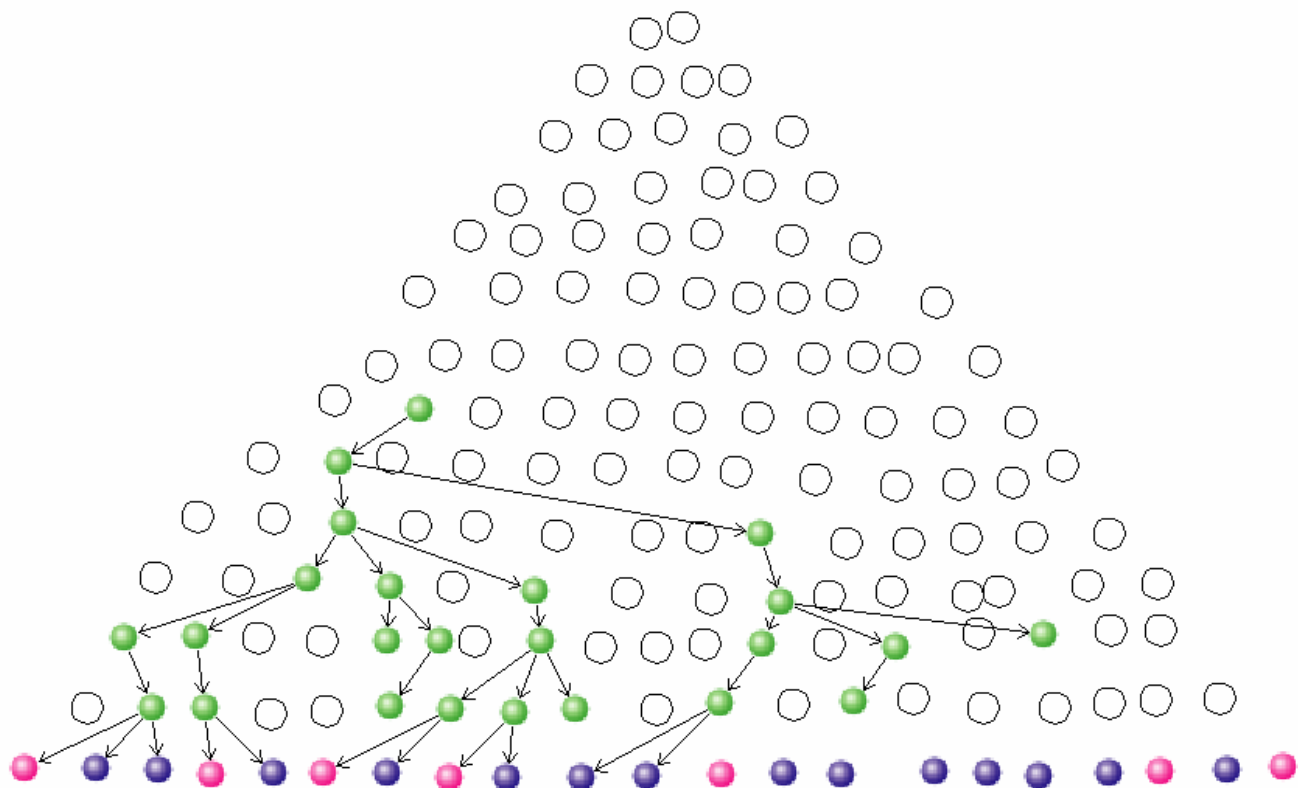
疾患遺伝子解析は・・・

- 集団の遺伝子伝達は、
 - 祖先から現代人への伝達の「木」
 - ゲノムは「木」が並んだ「林」
- 疾患関連遺伝子検出は、
 - 疾患リスクを上げる「木」を「林」から見つけ出すこと



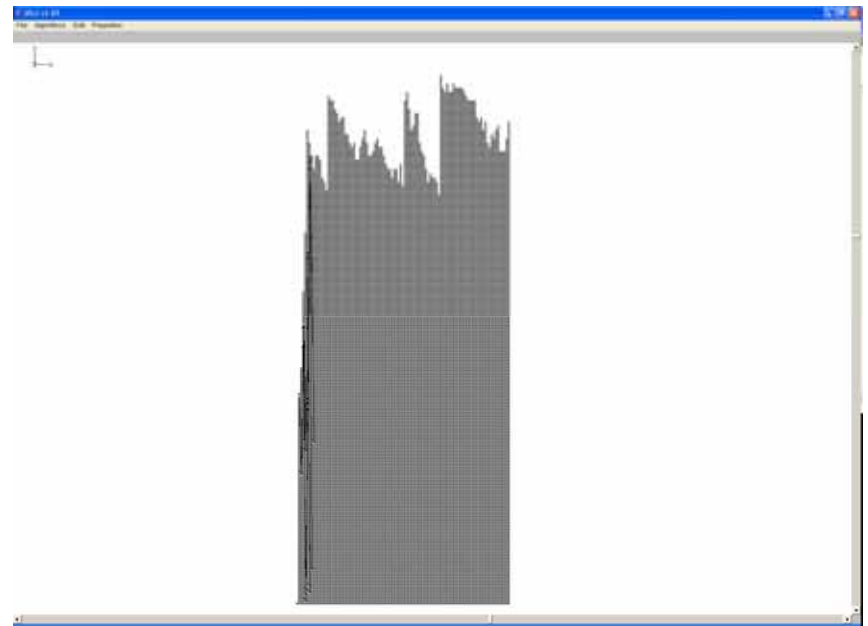
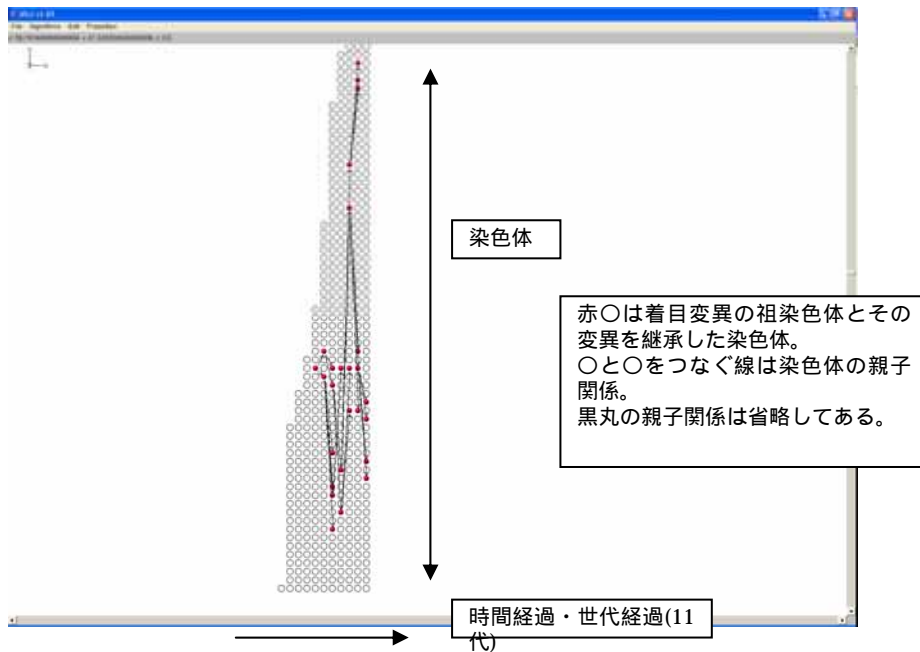






- 木の葉の部分が不確実
 - ペネトレーションが低い
 - Phenocopyが存在する
 - …別の木が邪魔をする
 - …民族が代わると、森の枝ぶりが変わる

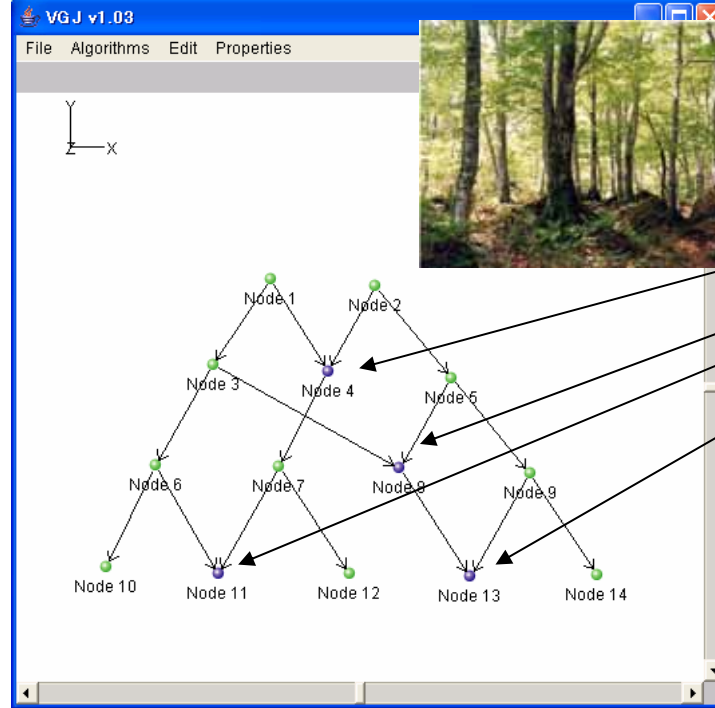
歴史の経緯によっては、 「多型」は断絶する



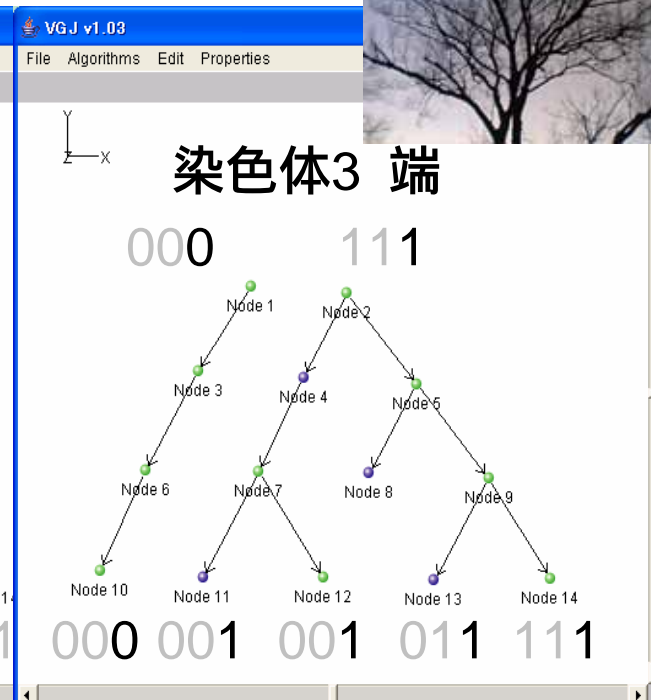
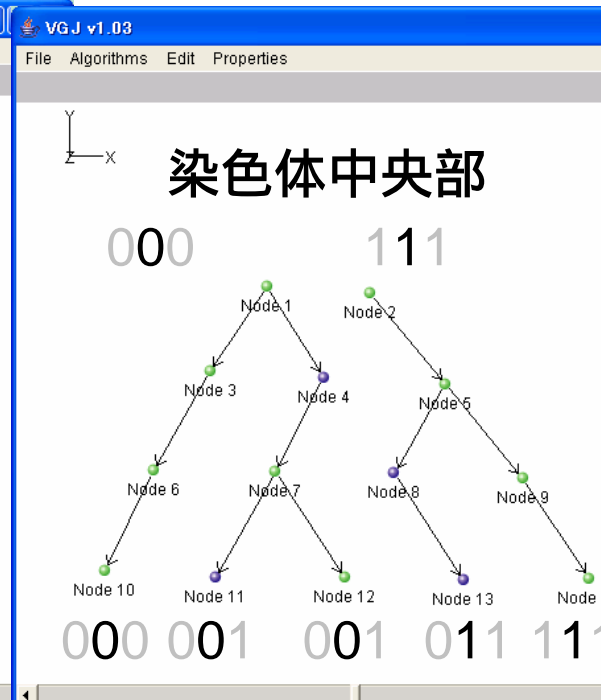
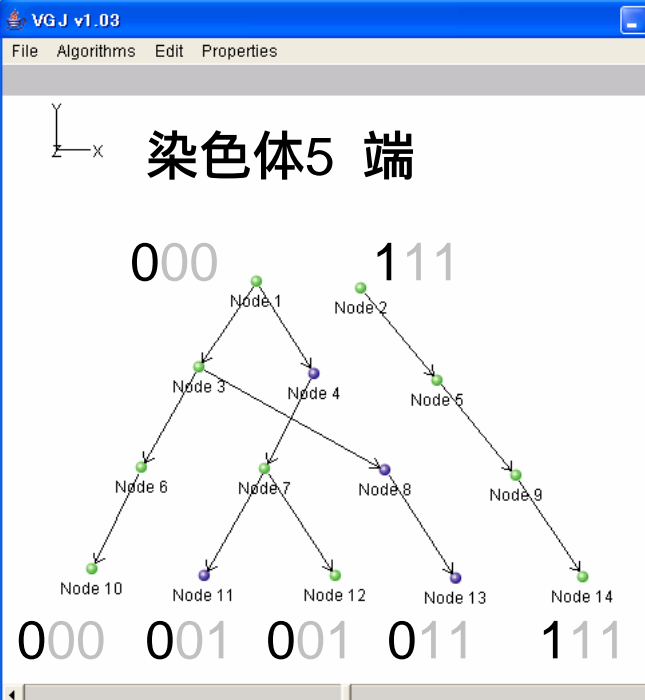
連鎖不平衡

- 相互に連鎖の関係にある2つの多型の間に存在しうる関係。
- ある2つの多型が同一の染色体上にあるとき、この2つの多型は「連鎖」している。
- この2多型間で組換えが起きなければ、2多型のアレルの組み合わせは1つしかなく、片方の多型のアレルがわかればもう片方のアレルも自動的に決まる。
- この2つの多型が発生して後、時間が経過するにしたがって、2点間には、組換えが繰り返し起きる。
- 「十分に多数回」組換えが起きると、片方のアレルがわかっててももう片方のアレルは全く予想できない。
- 片方のアレルからもう片方のアレルの予想がつく関係を「連鎖不平衡」の関係にある、といい、予想が全くつかない関係を「連鎖平衡」の関係にある、という。
- 連鎖不平衡は、「連鎖」関係にある多型において、その間におきた組換えの歴史を反映したものとなる。また、「組換えの歴史」は個々の染色体が持つものではなく、集団が持つものであるから、「連鎖不平衡」はある「集団」の「2つの多型」の間に存在するものである。

染色体全体で
考えると、親子
関係は複雑
(複数の祖先
がいる)だが、
部分的には、
単純(祖先は1
つ)



組換え体



疾患遺伝子マッピング

- ゲノム探索をするための、地図の作成
- 地図には、多型マーカーを配置
- 広範囲スクリーニングデータからの候補領域同定
- 領域のFine mapping
- 標的領域の原因バリエーション探索
- 複数多型の複合効果
 - 同一遺伝子内複数多型
 - 遠位ロカス間の複合効果



連鎖不平衡マッピング その理論的な話し データ解析の観点から

- 遺伝性のある形質の遺伝性の評価
 - その遺伝性の評価方法
 - 形質全体の遺伝性
- 形質の遺伝性を構成する、ローカス単位の(伝達する単位での)遺伝性の評価
- 個々のローカスを発見するための方法の確認
- ローカスの発見とその細部の検討
 - 責任バリエーションの特定
 - 複数の責任バリエーションとその相互関係
- 複数ローカスの相対的評価・相互関係評価
- 遺伝的バックグラウンドとの関係の評価・環境バックグラウンドとの関係の評価



