

**MOLECULAR ECOLOGY  
RESOURCES**

Molecular Ecology Resources (2011) 11, 591-611

doi: 10.1111/j.1755-0998.2011.03014.x

## Current trends in microsatellite genotyping

E Guichoux, L Lagache, S Wagner, P Chaumeil, P. Leger, O. Lepais, C. Lepoittevin, T Malausa, E Revardel, F Salin, RJ Petit.

### Abstract

- マイクロサテライトマーカーは分子生態学で最もよく使われる DNA マーカー
- Multiplex PCR の改良と次世代シーケンスという技術革新が利用を加速させている
- 次世代シーケンスで多配列→いい遺伝子座を選べる→Multiplex PCR の効率上昇
- 一方で実際の利用にはラグがある：多くの研究で効果的な Multiplexing なし
- 技術革新をうまく利用するための様々な戦略についてアウトラインを提示し議論する

### Introduction

#### マイクロサテライトとは？

- 分子生態学の多くの問いの答えるには？…マイクロサテライトマーカーは有効
- マイクロサテライトは Simple Sequence Repeats (SSRs)、Short Tandem Repeats (STRs)とも
- 1-6 塩基の単純な塩基のモチーフの繰り返しで突然変異率が極めて高く、超多型
- SSRs は分子生態学の研究で最もよく使われる (Fig. 1)
- 80 年代後半から親子解析、遺伝構造、ゲノムマッピングなどに応用されてきた
- ゲノム中のダイナミクス、実務上の有効性についてはレビューがたくさん
- ライバルとしては一塩基多型 (Single Nucleotide Polymorphisms, SNPs) マーカー

SSR マーカー：単純繰り返しの違い

SNP マーカー：塩基置換

個体 A	<pre> ...AGCACACACACACTCC... ...AGCACACACACACACTCC... </pre>	<pre> ...ATGCTGTAATGCTAGTTGC... ...ATGCTGTAATACTAGTTGC... </pre>
個体 B	<pre> ...AGCACACACACACACTCC... ...AGCACACACACACACTCC... </pre>	<pre> ...ATGCTGTAATGCTAGTTGC... ...ATGCTGTAATGCTAGTTGC... </pre>

- SSR の利点の活用と欠点の克服が、これから SSR が生き残るカギ

### Box 1 SSRs vs. SNPs

#### SSRs と SNPs の 3 つの大きな違い

- ① ゲノム中の頻度は SSRs よりも SNPs の方が多い  
ヒトの場合、SSRs は 2-30 kb に 1 つ、SNPs は 100-300 bp に 1 つ
- ② 突然変異率は SSRs の方が SNPs よりも高い  
1 世代あたりの突然変異率：SSRs は  $10^{-3} \sim 10^{-4}$  で、SNPs は  $10^{-9}$
- ③ 1 遺伝子座あたりのアレルの数は SSRs の方が SNPs よりも多い  
SNPs は普通 diallelic (アレルが二つ)：ヒトで 3 以上アレルがある SNPs は 0.1% 以下  
SSRs は 1 遺伝子座あたり 10 以上のアレルを持つこともしばしば

#### SNPs に対する SSRs の利点

- アレル多様度が高い (遺伝子座あたりのアレル数が多い)
- 近縁種への転用が簡単 (SNPs は最大で 50%)
- 混合構造を見つけるのに優れている
- 家系解析 (メンデルの分離の法則) でマーカーの精度の評価が容易  
SNPs はアレルが二つなので家系解析では検出できないエラーがたくさん
- 突然変異率が高いため最近の集団拡大を検出しやすい。
- 同じ情報量を得るのに SSRs の方が少ない遺伝子座で済む (Table 1)

#### SNPs に対する SSRs の欠点

- アレルが多いので、正確なアレル頻度の推定にたくさんのサンプルが必要
- 突然変異率が高いということは親子解析を難しくする場合がある
- 復帰突然変異 (もとのアレルに戻る) 率が高く、集団の長い歴史の推定には向かない
- SSRs の多様性は潜在的なゲノムの多様性を正確に反映しないかもという報告がある
- SSRs の標準的な解析方法は蛍光検出のキャピラリーゲル電気泳動のみ  
SNPs はマイクロチップをはじめ様々な解析方法がある
- 研究室間や解析機器間での再現性が低い
- PCR 産物長が SSRs の方が長いため、糞などの劣化した DNA の解析が難しい

#### まとめ

- SSRs の一番の利点：高いアレル多様度→生物学的課題を解決するパワーが大きい  
⇨欠点は技術革新でカバーできることが多い

#### マイクロサテライト解析における 2 つの技術革新

- SSRs 解析に関わる 2 つの技術革新→SSRs はこれからも生き残ることを示唆
- ① 次世代シーケンス技術の台頭：安くて早いマーカー開発 (最初の報告は 2009 年)

- ② Multiplexing (複数遺伝子座を同時に解析すること) がより簡便化されてきた  
 True Multiplexing : 複数遺伝子座を同時に PCR して解析←以下、"Multiplexing" はこっち  
 Pseudo Multiplexing, Poolplexing : 別々に PCR した産物を同じレーンでフラグメント解析
- SSR 解析のおおまかな手順
    - ① マイクロサテライトを含む領域を増幅できるプライマーを設計
    - ② 検出用の蛍光がついたプライマーを用いた PCR で特定の SSR 領域を増幅
    - ③ 電気泳動したサンプルの蛍光を機械が読み取って長さの違いを測定
  - アリルのレンジが被らない遺伝子座は同じ蛍光で、被る遺伝子座は違う蛍光でラベル
  - Multiplexing によって遺伝解析の経済的・時間的コストを大幅に削減 (Box 2)

### Box 2 Multiplex SSR typing の費用対効果

#### で、実際どれくらい違うの？



円高なのに試薬の  
値段が下がらない!!

- SSR ジェノタイピングのコストが Multiplexing によってどれくらい異なるのか試算
- 目標は 2,500 サンプルを 24 遺伝子座の SSRs でジェノタイピング
- 5 つのモデル : simplex、2-plex、4-plex、8-plex、12-plex (plex は遺伝子座数)
- コストは消耗品・試薬類・実験助手の給料 (フランス価格) を含む
- Multiplexing は明らかに費用対効果が高い (Fig. 2)  
 12-plex は約 1,875 ユーロ (195,500 円)、simplex は約 23,750 ユーロ (2,477,000 円)
- Multiplex の開発 (有効なプライマーの組み合わせの検証) コストも十分に回収可能  
 SSR の開発時に Multiplexing を意識してプライマー設計→ほとんど開発コストなし
- 8-plex であってもコストは約 4 分の 1 (8-plex であれば現実的な Multiplexing)

※1 ユーロ=104.28 円換算

#### コスト削減のための他のアプローチ

- Multiplexing の最も一般的な商品は QIAGEN の Multiplex PCR kit (うちも使ってます)
- このキットはサンプルあたりのコストは高いが最終液量を 5 $\mu$ L にできる (推奨は 50 $\mu$ L)  
 サンプルあたり 1.88 ユーロ (196.05 円) が 0.13 ユーロ (13.56 円) に
- 96 プレートから 384 プレートに移行すればコストダウンできる
- M13 tailed プライマーなど小技を駆使 (詳細は本文で)

#### この論文で議論すること

- 文献調査から SSR 研究の動向と効率的 Multiplexing はまだ一般的でないことを示す
- SSR マーカーと Multiplex の開発の最新動向を紹介  
 マイクロサテライトの探索から、プライマーデザイン、データ収集まで
- ジェノタイピングの精度とその改善について
- SSR 開発・Multiplexing それぞれのレビューは多数存在するが、SSR ジェノタイピングの完全なプロセスとワークフローをレビューしたものは本報告がはじめて

### A review of current practices

- 文献調査によって SSR ジェノタイピングの最近の動向を評価  
2009-2010 年の間に Molecular Ecology 誌に出版された SSRs を使った論文 100 本を調査  
Supplementary file 1 : 調査した論文のリスト (End note のライブラリーファイル)
- 100 本中 69 が集団構造解析、31 が親子・兄弟解析
- すべてが 2 倍体生物で、脊椎動物、無脊椎動物、植物など様々 (Table 2)
- 平均で 11.6 遺伝子座、564 個体を解析、座数・個体数に分類群間で大きな違いはなし
- Multiplexing を使った研究は 42% と少ない (Table 2)
- 平均の Multiplexing あたりの遺伝子座数は 5.0 遺伝子座
- Multiplexing を使っている研究は平均で 13.9 遺伝子座を使用 (全体平均より多い)
- 8-plex 以上の Multiplexing は 11 の研究のみだった
- 20 座以上の Multiplexing の報告もあり、技術的にはもっとたくさんできるはず
- 8-plex 程度であればハードルはそれほど高くないがなぜ使われないのか？
  - ・ Multiplexing は複雑で開発の経済的・時間的コストが膨大？  
→これは昔の話で、随分ハードルは下がった
  - ・ Multiplexing による結果は正確性にかける？  
→これについても相当改善している (後ほど詳しく言及)

### SSR selection

#### シーケンスデータ情報

- SSR の探索にはシーケンスデータ (DNA 塩基配列の情報) が必須
- 現在までには 2 つの方法：ゲノミック DNA から単離 or EST ライブラリー
- EST とは？ : Expression Sequence Tag のことで遺伝子転写産物の一部に当る短い配列  
転写産物の「目印」となるため、様々な生物の EST がデータベースに登録されている
- EST-SSRs の利点と欠点 (表 1)
 

一般にゲノミック SSRs よりも多型が少ない	}	欠点
PCR 産物のフラグメントサイズが想定と異なることが多い		
配列を得るまでのコストがゼロ	}	利点
コード領域の近くなのでマーカーと形質の関連が見つかる可能性が高い		
コード領域の近くなので保存的であり、近縁種に転用できる確率が高い		
- クローニングやサンガーシーケンスなど従来の実験室作業はコストと時間がかかる
- ブレイクスルーとして、次世代シーケンス技術による SSR 開発
- 現在までの成功例 (15 例) では従来法よりも 2-5 倍コストダウン、劇的な時間短縮  
Abdelkrim et al. 2009 : Roche 454 の 1/16 run を使って 1,500 US \$ (115,575 円) 以下

Bule Duck について 13 座の多型マーカーができた

Castoe et al. 2010 : Roche 454 の 1/8 run を使って約 2,000 US \$ (154,100 円)

SSR マーカーを設計できる配列を 4,564 得るのに 9 日間

※1\$ドル= 77.05 円換算

### SSR を見つけるためのゲノムショットガンシーケンシング (次ページの図)

- 次世代シーケンスで SSR 開発をする際の注意点  
ゲノムサイズ、SSR モチーフの種類、coverage (全ゲノムの何%をカバーできているか)
- ゲノムの大きな生物 (例えば針葉樹) は SSR を組めない配列をたくさん持っている可能性が高いのでいきなりショットガンシーケンスをするのはリスクが高い  
→cDNA ライブラリー (転写産物) を作ってそれを読むなどで対処可能
- SSR の頻度が低い生物の場合、SSR 濃縮ライブラリーを使う
- しかし、本研究で調べた 15 の論文のうち 12 ではいきなりゲノムショットガンで開発  
※Supplementary file 2 : 次世代 SSR 開発の 15 論文の End note ライブラリーファイル

### Read (次世代シーケンスで読んだ断片) の長さ

- これまで出版された 15 の論文はすべて Roche の 454 を使用
- Roche の 454 は次世代シーケンサーのうちで 1 read の長さが最長  
各社 1 read の長さ…454 (Roche) : 400bp、GA (Illumina) : 75bp、SOLiD (life tech.) : 50bp
- Roche 454 なら、1 read の配列の中で SSR プライマーが設計できる  
→シーケンスの assembly をパスできるので Bioinformatics での消費が小さい
- 454 の配列データから SSRs を探す便利なソフト (MSAT-COMMANDE や QDD)
- Roche 454 の 1 read の長さはそのうち 400 bp を越えるので、かなり効率が上がるはず

### 単純な費用対効果以外の次世代シーケンスの利点

- うまくいけば数千の SSR 遺伝子座が 1 回の run データから得られる
- Coverage が十分なら、ゲノム中の重複配列もあらかじめ除去 (非特異的増幅の予防)
- 2 個体以上の DNA をプールして run すれば、あらかじめ多型情報も得られるかも  
この方法はまだ 1 例で、しかもそれほど効果的ではなかった (Parchman et al. 2010)
- 現状では coverage の高い解析はまだないが、今後の技術発展に期待

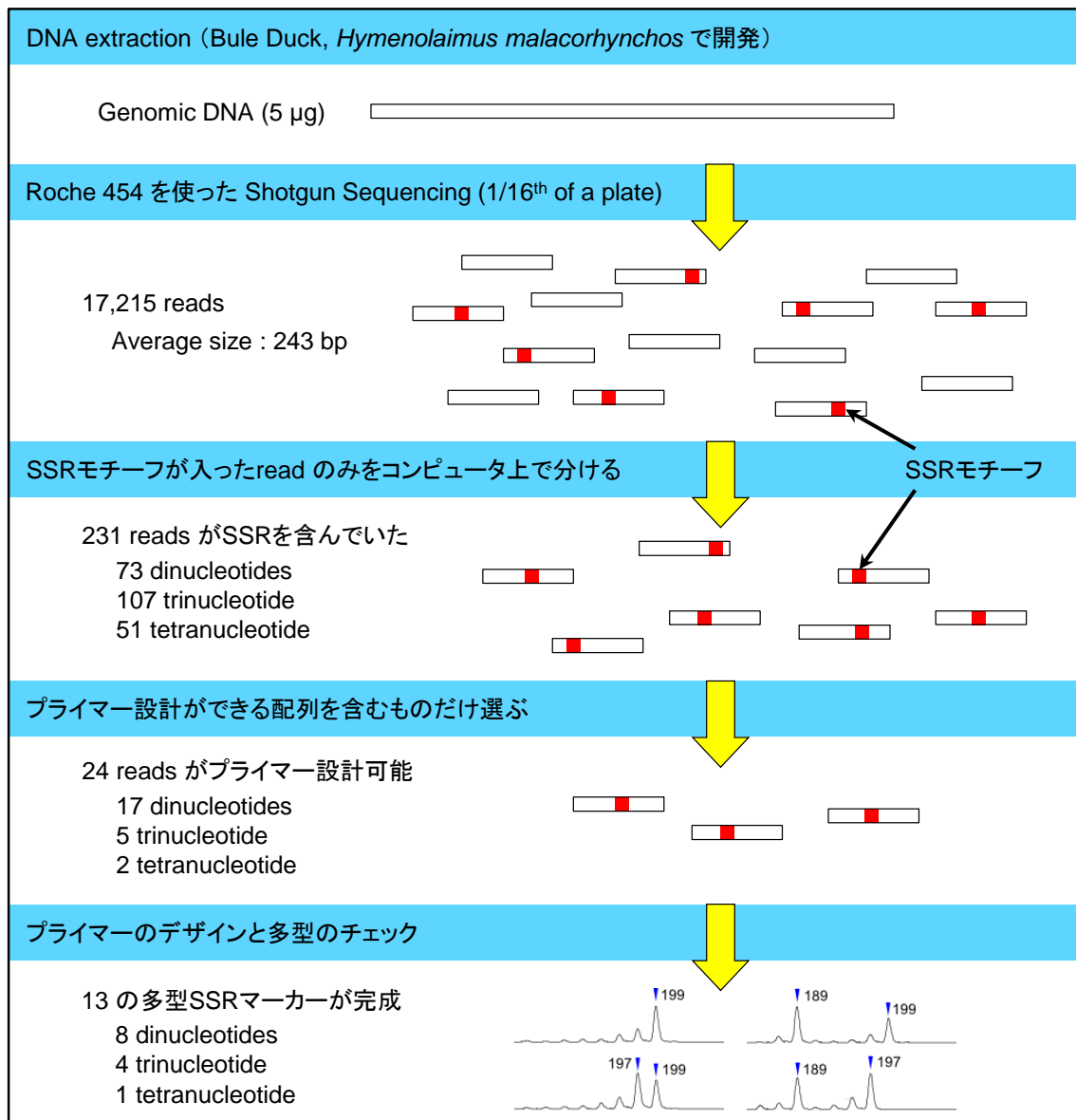


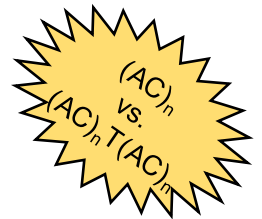
図1. 次世代シーケンス技術を使ったSSRマーカー開発のフロー.

マーカーのスクリーニング過程はAbdelkrim et al (2009) のデータを用いた.

## SSR のタイプの選び方

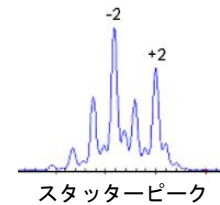
Perfect repeats か imperfect repeats

- Perfect repeat :  $(AC)_n \Leftrightarrow$  Imperfect repeat :  $(AC)_nT(AC)_n$
- Perfect repeat の方が Stepwise Mutation Model に合うことが多い→後々解析に便利
- 文献調査した SSRs のうち 26% は Imperfect repeat だった (Table 2)
- Perfect repeat が優先だが、Imperfect repeat でも大きな問題はない



リピートユニットのサイズ (何塩基繰り返しの SSR がいいのか?)

- 短いモチーフ (mono- or di-) はアリルのサイズレンジが狭い  
→多くの遺伝子座を使って multiplexing が可能
- ⇒ Mononucleotide SSRs はピークの読み取りが難しいのでダメ
- ⇒ Dinucleotide SSRs は一番使われるが比較的スタッターが出やすい (後述)
- 長いモチーフ (tri-, tetra- or pentanucleotide) はスタッターが少なく、読み取りが容易  
→法医学や親子解析などピークの読み取りミスが許されない分野で広く使われる



リピートユニットの数 (モチーフの繰り返しがどれくらいあればいいのか?)

- 繰り返し数は突然変異の挙動に大きな影響を持つ→繰り返しが多いと高突然変異率
- 繰り返しが多い配列は多型性が高いことが多い
- ⇒ 繰り返しが多いとアリルドロップアウトやスタッターピークの増加などの問題  
※アリルドロップアウト…ヘテロのうち、片方 (主に長い方) がうまく増えない、うまく検出されないなどの問題で見かけ上ホモになってしまう現象
- アリルの数が多いとサイズレンジが大きい→multiplexing で遺伝子座の数を増やせない
- 中間的な繰り返し数が妥協点  
→van Asch et al. (2010)では tetranucleotide で 11 から 16 繰り返しの間を推奨



**Box 3 SSRs の PCR 増幅で生じる様々な問題 早期発見・早期対処**

- PCR に関係する問題によって、アリの決定が難しくなる (Fig. 3, 4)
  - これらの問題は早期発見・早期対処が望ましい
- ① ヘテロ接合のピークの高さが不均一 (Fig 3b)
    - ・PCR の競合による (低い方が増えにくい)、解決策は null アリルと同じ (後述)
  - ② スタッターピーク (Fig 3c) stutter : 【自動】言葉に詰まりながら話す
    - ・スリップにより数繰り返しかけ異なる PCR 産物が増幅するケース
    - ・これはマイクロサテライトあるあるだが、酷い場合はピークが取れない
    - ・後述する M13 tailed PCR 法ではスタッターが出にくい→蛍光ラベルを買う場合は注意
    - 【解決策】変性温度を 83°C まで下げる、次世代酵素を使う (fusion 酵素)
    - そもそもスタッターの少ない遺伝子座を選ぶ
  - ③ ピークが分裂する (Fig 4d)
    - ・Taq polymerase が PCR 産物に 1 塩基付加 (アデニン付加) するプロセスが中途半端
    - ・図のようにピークをとる手もあるが、2 塩基違いのヘテロが困難な場合がある
    - 【解決策】PIG-tail などリバースプライマーの 5' 末端に配列を足す
    - Tailed プライマーは PCR の効率が悪いのでサイクル数に注意
    - DNA 濃度 and/or プライマー濃度を下げる、Taq を増やす、Taq を変える
  - ④ Null アリル (Fig 4a, b)
    - ・特定のアリルが増幅されないことで、ヘテロがホモに、ホモが非増幅として検出
    - ・プライマーがくつつくサイトが突然変異している場合に起こる
    - ・データ解析の際に null アリルを考慮する方法があるが、事前に解決したい
    - ・null アリルに関する報告がある論文は文献調査 100 本のうち 40 本のみだった。
    - 【解決策】多型のある領域にプライマーを組まない
    - ・事前にシーケンス変異を調べる、メンデルの分離実験などで null のチェックが可能
    - ・著者らのラボでは、母親 1 個体、子供 7 個体を 1 セットとして、12 か 24 のセットをジェノタイピングすることで null アリルを推定している (図 2)
    - ・家系チェックが大変な場合は Hardy-Weinberg 平衡からのズレを検定しておく
  - ⑤ プライマーダイマーなどの偽ピーク (Fig. 4c) と三本ピーク (Fig. 4d)
    - ・プライマー同士の増幅により短いサイズのピークが出る (プライマーダイマー)
    - ・ゲノム中の 2ヶ所以上増やすなどの問題で 3 本以上のピークが出ることも
    - 【解決策】明らかな偽ピークは無視する、プライマーを組み直す



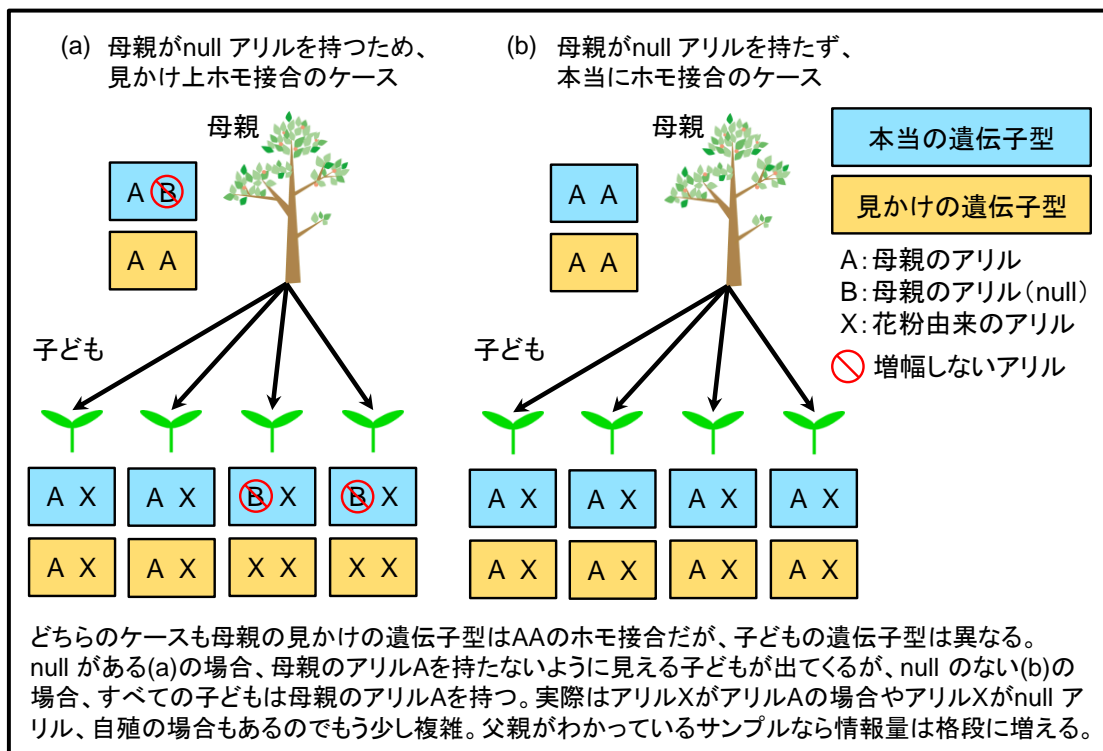


図2. 家系解析によってnull アリルを推定する方法の概略

### Primer design

- Multiplexing をするために適切なプライマー設計が重要で厳しい選択が必要
- 同じ dye (蛍光) で PCR 産物のサイズレンジが被らないように慎重に設計
- Multiplexing を考慮して自動的にプライマー設計してくれるプログラムも多数  
Multiplex Manager や NetPrimer など (Appendix 1)
- 当然、Multiplexing する遺伝子座同士のアニーリング温度は同じ程度に (58-60°C)
- すでに記載されているが Multiplexing を考慮していないプライマーを使う場合は、再設計するか、短い塩基配列を足してアニーリング温度を調整

### Primer validation in simplex

- 開発の早い段階で、プライマーの質を完全にチェックしておく
- この段階では simplex (1 遺伝子座で PCR) でテスト
- 蛍光プライマーを買って、使えなかったら勿体ないので M13-tailed などを利用 (図 3)  
蛍光ラベルプライマーはノンラベルプライマーの 10 倍ぐらい高い!!
- M13 を使った PCR と標準的な PCR は反応条件 (特にサイクル数) が異なるので注意
- プライマーチェック用のサンプルは可能な限り変異が出そうなサンプルで  
→チェック用の 8 サンプルや 16 サンプルに異なる集団由来の個体を入れる  
→早い段階でアリルレンジを把握しておくとも Multiplexing を開発しやすい

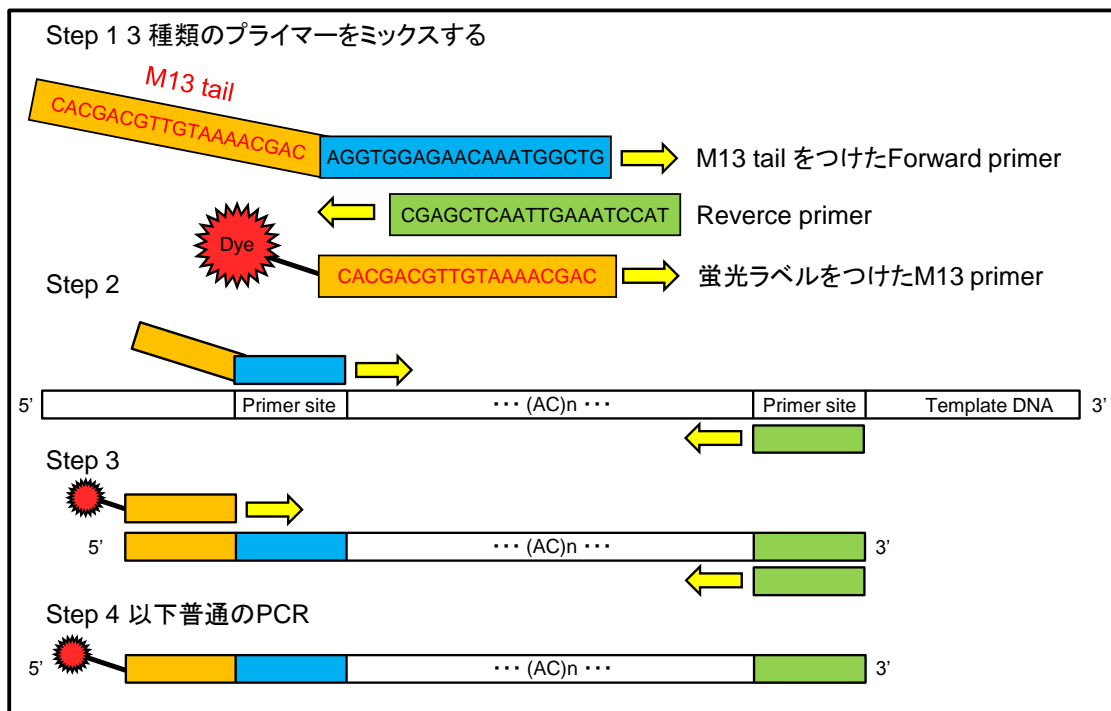


図3. M13 tailed プライマーを用いたPCR法 (Schuelke (2000) Nature Biotechnologyを改編)

- 遺伝子座のアリルレンジを簡便に把握する方法として DNA pooling 法がある (図 5)

### The multiplexing phase :

#### スクリーニングの過程は Fig 5

- Multiplexing は経済的・時間的コストを大幅に削減する
- PCR の回数が減るのでサンプルあたりの DNA が少なくて済む
- すべてのマーカーを使って、最小反応数になる組み合わせを考える
- Multiplex Manager は最適な Multiplexing の組み合わせを作ってくれる (Fig. 6)
- Multiplex PCR は繊細な技術なので標準化をしっかりと行う (特に DNA 濃度)  
薄すぎる：遺伝子座間で増幅にムラが出やすく、ドロップアウトを起こす  
濃すぎる：偽ピーク、遺伝子座間の増幅ムラ、強いスタッター (←濃い方が×)
- 現存する商用のキットでは QIAGEN Multiplex PCR kit が一番優秀
- タッチダウン PCR (サイクルごとに  $T_m$  を下げる) で増幅ムラを抑えられることも
- Simplex PCR で厳しくチェックした場合でも Multiplex phase でいろいろ問題が起こる  
ex. 遺伝子座間での増幅のムラやピークの高さが不均一になる  
→とにかくアニーリング温度を揃える  
→ピークが低いプライマー濃度を上げる or ピークが高いプライマー濃度を下げる
- 同じサンプル・遺伝子座でも Dye を変えるとアリルサイズが変わるので注意 (図 6)
- Dye (蛍光ラベル) の質を高く保つ  
→融解・冷凍は最小限に

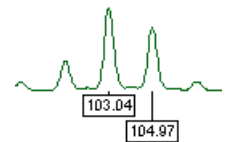
- 中間的な Multiplexing (8-plex 以下) では Fig 5 の手順を踏んでいればそれほど神経質にならなくていい  
↑ 試薬の質が向上したから
- 一昔前は、プライマーと DNA の比率、dNTP と MgCl<sub>2</sub> のバランス、buffer 濃度 etc... あらゆることを最適化して大変な作業だった

### Size precision

- アナライザーによるフラグメントサイズの測定：同じ機械では再現性を確保する
- Run ごとにばらつくようだと要改善 (1 bp あたり 0.2 bp までのばらつきなら許せる)
- ばらつきの要因：PCR、キャピラリー長、電圧、検出時間、室温、ポリマー、Dye、etc...

### Allele calling and binning

- 機械が測定したフラグメントサイズから遺伝子型を読む
- 二つのステップ
  - Allele calling：正確なフラグメントサイズの決定 (ピークの確定)
  - Allele binning：フラグメントサイズの実測値の整数化
- Allele calling の様々な障害：ピークの振り切り、ノイズ、スパイク、スタッター etc...  
アナライザー付属のソフトはこれらを考慮して自動でピークを決めてくれるが、遺伝子座の特徴によっては手動で行わないといけない
- Allele calling の自動化には挙動のしっかりしたマーカーと適切な PCR が重要
- Allele binning はかなり重要なステップ：洗練された整数化ルールを作る  
ジェノタイピングエラー全体の 21-40% が Allele binning ステップに起因するらしい
- 自動化するにしてもマニュアルチェックはとても大事
- 生のフラグメントサイズを取り出して解析するのはマーカーの質を知る上で重要  
エクセルなどでフラグメントサイズの頻度分布を作ってみる  
基本的には想定されるアリルサイズ周辺に点がプロットされるはず  
想定から外れるもの…本当にずれているアリル？ジェノタイピングエラー？
- フラグメントサイズの頻度分布の視覚化用のソフトもある  
MsatAllele は STRand というフリーのピーク解読プログラムのデータを R 上で解析
- Autobin は著者らが作ったエクセルマクロで商用ソフトの生データを解析する  
自動で遺伝子座数やサンプル数を認識・データソートし、頻度分布を作る (Fig. 7)  
マニュアル+自動で整数化し、GENEPOP、STRUCTURE 形式のフォーマットを作る



解析したピーク  
サイズの実測値が出る

### Measuring and reporting error rates

- ジェノタイピングエラー率は丁寧に調べるべきである
- 文献調査では 26% の研究しかジェノタイピングエラーの報告がない

- ジェノタイピングエラーのチェック方法
  - ① 親子ペアでメンデル遺伝との不一致をカウント
  - ② 同じはずの遺伝子型間で不一致をカウント（同じサンプルを複数回タイピング）
- これらのチェックは万能ではないため組み合わせて使うのが望ましい  
例えば②だと突然変異や null アリルはわからない
- 同サンプルの再ジェノタイプが不一致だった場合…エラーの本質の見極めは難しい  
例えばアリルドロップアウトと単純な PCR エラーは区別できない

### Data management

- 膨大なデータをコントロールするのは本当に難しい
- 特に過去のサンプルとの比較するケースは十分起こり得るのでデータ管理は重要
- ジェノタイピングデータは整数化したものだけではなく、生データも保管
- データ管理システムもいくつかリリースされている
- （ちゃんとデータ整理して引き継いで卒業してください!! 遺伝解析に限りません!!）

### Conclusions and perspective

- 現在の技術で可能なことと、現実の遺伝解析にはラグがある（特に Multiplexing）
- 次世代による大量の配列によってただ早く・安くだけでなくいいマーカーが作れる
- 近い将来、フラグメント解析ではなく SSR 配列を直接決めるようになるかも  
数百のサンプルを数千遺伝子座について配列ベースでタイピングするのも夢ではない

おまけ：Mol Ecol Res の 2011 年 9 号と 11 号に次世代・SSR 絡みの論文が出ています。

Glenn TC (2011) Field guide to next-generation DNA sequencers Mol Ecol Res 11(5):759-769 Sep

Gardner MG, Fitch AJ, Bertozzi T, Lowe AJ (2011) Rise of the machines - recommendations for ecologists when using next generation sequencing for microsatellite development

Mol Ecol Res 11(6):759-769 Nov