第1章　第4節

バイオインフォマティクス技術を活用した DNAマーカー開発

農研機構果樹研究所 品種育成・病害虫研究領域　　奈島 賢児

1）DNAマーカー

DNAマーカーは，品種・系統・個体間で観察されるゲノムDNA配列上での多型性を利用した分子マーカーである．DNAマーカーを用いることで，塩基配列パターンが異なる品種・系統・個体を識別できる．犯罪捜査の場面では，従来の指紋や血液鑑定に代わる新しい個人識別の方法としてDNAマーカーを用いたDNA鑑定が利用されている．DNA鑑定は人の犯罪捜査に限られず，果樹においても品種の識別，親子関係の判定，遺伝的な類縁関係の解明などに用いられる．果樹の育種を行う際，新品種に導入したい形質によって適切な交雑親を選定する必要がある．交雑親となる品種・系統は，有する形質や来歴情報を参考に選定するが，偶発実生に由来する品種ではその来歴は不明であり，また文献等に来歴が記載されている品種においても，その交雑組合せでは説明できない表現型（phenotype）や遺伝子型（genotype）が報告される例がある（Yamamoto *et al.* 2003; Sawamura *et al.* 2008）．DNAマーカーを用いた品種識別や親子鑑定を行うことで，交雑品種の親子関係を特定でき，また枝変わり品種（bud mutant），同名異品種や異名同品種の判定ができる．

DNAマーカーは有用形質に対する育種選抜においても利用される．有用形質を支配する遺伝子座に連鎖するDNAマーカーを用いたDNAマーカー選抜（marker assisted selection，MAS）では，幼苗の段階でDNAマーカーの遺伝

Kenji NASHIMA　✉nashima.kenji@gmail.com
DNA marker development using bioinformatics technology

子型による選抜・淘汰を行う．果樹は個体サイズが大きく，播種から開花・結実まで数年の期間を要するため，栽培には労力と広い圃場面積を必要とする．そのため幼苗段階で不要な個体を淘汰し，有望個体のみを圃場に定植できるMASは，個体の維持管理に必要な労力・圃場・費用の大幅な削減ができるという利点がある．MASに利用可能な，有用形質を支配する遺伝子座に連鎖したDNAマーカーの選定には連鎖地図（linkage map）の作成が有効である．形質が分離する交雑集団についてDNAマーカー解析を行うことで，DNAマーカー間の組換え価（recombination value）に基づいた連鎖地図の作成が可能であり，各DNAマーカーの位置関係を明らかにすることができる．さらに量的形質遺伝子座（quantitative trait loci, QTL）解析の適用により，量的形質を支配する遺伝子座の位置と，QTLに連鎖するDNAマーカーを推定できる．QTL解析では，各個体のDNAマーカーの遺伝子型および調査対象形質の表現型値を合わせて解析する．

果樹の品種識別や，MASを実施するためにはDNAマーカーが必要である．DNAマーカーは後述する単純反復配列（simple sequence repeat，SSR）や，レトロトランスポゾン（retrotransposon）の挿入多型，一塩基多型（single nucleotide polymorphism，SNP）など，品種・系統・個体間で多型のある塩基配列領域について設計するため，配列データからこれらの領域を見出す必要がある．本項においては，公開ゲノム配列や，次世代シークエンサ（next generation sequencer, NGS）解析データなど，大規模配列データからのDNAマーカー設計法について紹介する．

2）DNAマーカー設計を行う配列の取得

DNAマーカー設計を行うためには基になる塩基配列が必要である．塩基配列取得の方法として，公共データベースに登録されている配列の利用およびシークエンス解析の実施の二通りが挙げられる．

a．公共データベースからの配列取得

個別の配列情報については公共データベースであるNational Center for Biotechnology Information（NCBI, URL1-4-1）やDNA Data Bank of Japan（DDBJ）（URL1-4-2），European Molecular Biology Laboratory（EMBL）（URL1-4-3）に登録されている．これらのデータベース間ではデータの同期が行われているのでどのデータベースを利用しても良い．

目的の樹種について，ゲノム配列が公開されている場合には塩基配列をダウンロードして利用できる．また樹種によっては，塩基配列情報に加えて転写領域や反復配列など各種配列アノテーション（annotation）をウェブブラウザ上で閲覧できるGeneric Genome Browser（GBrowse, URL1-4-4, Stein *et al.* 2002）が整備されている．果樹ではリンゴ（*Malus domestica* Borkh.）・モモ（*Prunus persica*（L.）Batsch），セイヨウナシ（*Pyrus communis* L.）のゲノム配列がGenome Database for Rosaceae（GDR, URL1-1-4-4, Jung *et al.* 2014）において，ヨーロッパブドウ（*Vitis vinifera* L.）のゲノム配列が*Vitis vinifera* Genome Data Base（VvGDB, URL1-4-5, Duvick *et al.* 2008）において，カンキツのスイートオレンジ（*Citrus sinensis*（L.）Osbeck）およびクレメンティン（*Citrus clementina* hort. ex Tannaka）のゲノム配列がCitrus Genome Database（URL1-4-6）においてGBrowse上で参照可能である．さらにGBrowse上では目的とする遺伝子やその近傍領域の塩基配列をFASTA形式のテキストデータファイルとして出力することができる（図1-4-1）．出力した塩基配列についてDNAマーカー設計を行うことで，目的遺伝子や領域と強く連鎖したDNAマーカーが設計できる．また，GBrowseは整備されていない場合でも，ウェブサイト内で配列が公開されている樹種もある．チュウゴクナシにおいてはPear Genome Project（URL1-1-3）内で配列が公開されている．このような場合，相同性検索プログラムであるbasic local alignment search tool（BLAST, Altschul *et al.* 1990）検索を用いて目的とする遺伝子が座乗している配列を見出し，その配列についてDNAマーカー設計を行うことも有効である．

NGS解析で得られたデータについても公共データベース内に登録されてい

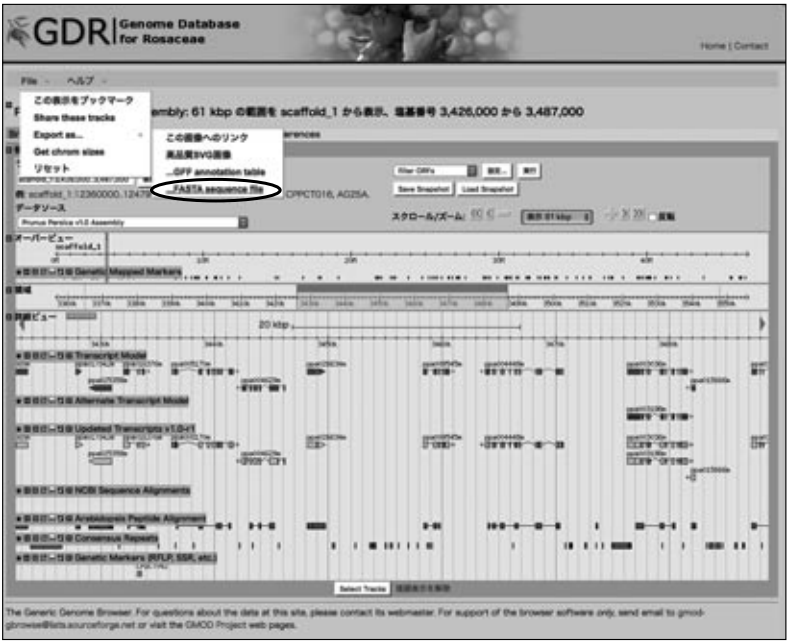


図 1-4-1　GDR における GBrowse

FILE > Export as… > …FASTA sequence file で任意の領域を FASTA 形式で出力することが可能（GDR から引用，https://www.rosaceae.org/, D）

る．NGS解析データは，NCBI内のSequence Read Archive（SRA）データベース（URL1-4-7）や，DDBJ内のDDBJ Sequence Read Archive（DRA）データベース（URL1-4-8），European Nucleotide Archive（ENA）データベース（URL1-4-9）のいずれかから入手できる．これらのデータベース間ではデータの同期が行われているのでどのデータベースを利用しても良いが，DRAでは日本語で操作できる利点がある．NGS解析は，塩基配列あたりのコストは低いものの，一度の解析で数十万円以上と多額の費用がかかるため，利用可能な解析データがあれば大幅なコスト削減となる．

b．次世代シークエンス解析の実施による配列取得

樹種によっては，先行研究が乏しく公共データベースに目的とする配列が登

録されていないケースもあり得る．その場合にはDNAマーカー設計の基となる配列を得るため，シークエンス解析の実施が必要となる．近年発達したNGSでは一度に大量のシークエンスを得ることができるため，DNAマーカー設計に適している．表1-4-1に代表的なNGSの特徴をまとめたが，使用する機種により得られる配列のデータ量，リード長，リード数が異なるので，目的に適した機種を選択する必要がある．

表 1-4-1 代表的な NGS の機種．数値は各社のウェブサイトに記載されていた代表値を記載

機器名	データ量	リード長	リード数
PacBio	550 Mb	10 kb	5万
GS FLX +	1 Gb	1 kb	100万
GS Junior	50 Mb	500 b	10万
Ion Proton	10 Gb	200 b	8000万
Ion PGM	2 Gb	400 b	500万
HiSeq	120 Gb	100 b x 2	60億
Miseq	15 Gb	300 b x 2	2500万

GS FLX+/GS Junior（Roche社）およびPacBio（Pacific biosciences社）は，得られるリード長が500 b以上と，他のシークエンサに比較して長いことが利点である．得られるリード長が長いため，短い配列断片から元の長い塩基配列を再構築する作業であるアセンブル（assemble）を行わなくてもSSRマーカー（詳細は後述する）やレトロトランスポゾン挿入多型マーカー（詳細は後述する）が設計できる利点がある．しかし，1リードのみでマーカーを作成すると，シークエンシングエラーを評価できないことが不利な点として挙げられる．

HiSeq/MiSeq（Illumina社）は,データ量がHiSeqで120 Gb，MiSeqで15 Gbと他のシークエンサと比較して多いことが特徴である．データ量が多く得られるため，参照配列（reference）へ個別のリードを対応させることで一つのアラインメント（alignment）を作成するマッピングを行うのに適している．アラインメントを構成するリードの厚み（depth）が大きいとシークエンシングエラーを評価できる利点がある．またマーカー設計時にマッピングデータが必要なSNPマーカーの設計に適している（詳細は後述する）．

Ion Proton/Ion PGM（Thermo fisher社）は，リード長はGS FLX+/GS

Junior，リード数はHiSeq/MiSeqには及ばないものの，ランニングコストが低く，ラン時間が短いことが利点である．Ion PGMはリード長が長いため，SSRマーカー設計やレトロトランスポゾン挿入多型マーカーの設計に適しており，Ion Protonではリード数が多いためSNPマーカーの設計に適している．

3）バイオインフォマティクス技術を活用したDNAマーカー開発

NGS解析データや公開ゲノム配列からDNAマーカーを設計する際，扱う配列数が数十万個以上になる場合や，塩基長が数百Mbに達する場合など，データ量が膨大となることがある．そのためDNAマーカー設計を行う際には，一度に大量のデータを扱うことのできる各種のソフトウェアを用いる必要がある．本項ではSSRマーカー，レトロトランスポゾン挿入多型マーカーおよびSNPマーカー開発について紹介する．

(1)　SSRマーカー

a．SSRおよびSSRマーカー

SSRとは単純反復配列のことで，特に数塩基の単位配列の反復からなる反復配列である．SSRは「(AG) n」のように表記され，( ) 内に反復の単位配列であるモチーフが，モチーフの後ろのnにはモチーフの反復回数が記載される．SSRはゲノム中の他の領域と比較して変異速度が増大しており，反復回数の変異が起こりやすい（Kalia *et al.* 2011）．SSRマーカーはSSRの多型をDNAマーカー化したもので，近傍に設計したPCRプライマーを用いてSSRを含む領域を増幅し，SSRの反復回数の違いをPCR産物長の違いとして検出する共優性（co-dominant）マーカーである（図1-4-2）．SSRマーカーは一塩基から十数塩基程度の僅かな増幅産物長の違いを検出する必要があるため，DNAシークエンサを用いてPCR産物長を検出するフラグメント解析（fragment analysis）が行われる（図1-4-3）．SSRマーカーは連鎖地図の作成や，品種識別に広く利用されている（Kalia *et al.* 2011）．

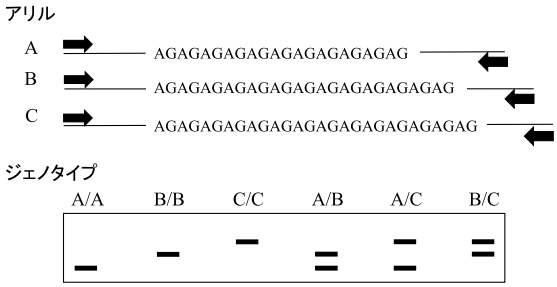


図 1-4-2　SSR マーカーによる多型検出法の模式図

SSR（上記の場合は AG モチーフの反復）を挟むように PCR プライマー（矢印）を設計し PCR を行う事で，反復回数の異なるアリルから長さが異なる増幅産物が得られる．増幅産物の長さを検出し，ジェノタイピングを行う．

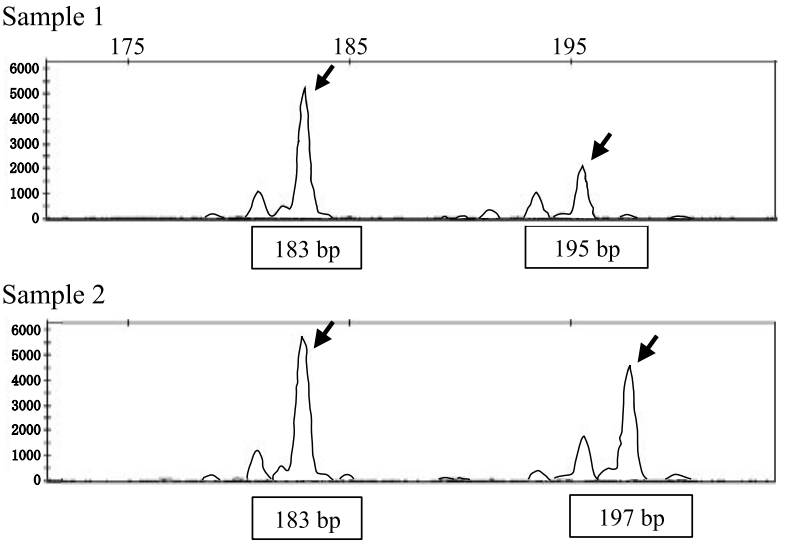


図 1-4-3　DNA シークエンサを用いた SSR マーカーのフラグメント解析例

横軸はフラグメントサイズ，縦軸はシグナル強度を表す．SSR の反復回数の違いが，矢印で示したピークのフラグメントサイズの違いとして検出される．

b．SSR の同定および SSR マーカー設計

SSRの同定およびSSRマーカー設計は，農業生物資源研究所において公開されているGalaxy/NIAS（URL1-1-26）を用いて行う（図1-4-4）．Galaxy/NIASの基本的な使用方法については，Galaxy/NIASのウェブサイトで紹介されている．Galaxy（URL1-1-25; Goecks *et al.* 2010）はゲノムなどのデータに対して様々なバイオインフォマティクス解析を行うことができるウェブベースのアプリケーションである．SSR同定およびSSRマーカー設計はGalaxy/NIAS内に導入されているソフトウェアを用いて行い，SSR同定にはmicrosatellite identification tool（MISA，URL1-4-10）を，SSRマーカー設計にはPrimer3（Untergrasser *et al.* 2012）を用いる．

図 1-4-4　Galaxy/NIAS の操作画面

画面左部には各種バイオインフォマティクス解析用ツールが，画面右部には解析の履歴や閲覧可能なデータが表示されている．SSR マーカー設計作業で用いる，画面左部の「Get Data」，画面上部メニューの「Workflow」および「Shared Data」について矢印で示した．（Galaxy/NIAS より引用，http://galaxy.dna.affrc.go.jp/, S）

必要なファイルは二つであり，一つはマーカー設計対象配列のマルチFASTAファイル，もう一つはSSRプライマーが目的とする領域のみ結合するかどうかを調べるための参照配列（全ゲノム配列等）のマルチFASTAファイルである（図1-4-5a）．マーカー設計対象配列のマルチFASTAファイルと参照配列のマルチFASTAファイルは同一でも実行できる．

Galaxy/NIASにおけるSSRマーカー設計の手順を以下に示した．

① 画面左部メニュー内の「NGS DATA ANALYSIS」-「Get Data」-「Upload File from your computer」を選択し，マーカー設計対象配列のマルチFASTAファイルおよび参照配列のマルチFASTAファイルを選択後，Executeボタンを選択しファイルをアップロードする．

② 画面上部メニュー内の「Shared Data-Published Workflows」を選択する．続けて「SSR MISA+Primer 3 」を選択し，「Import」を選択する．

③ 画面上部メニュー内の「Workflow」を選択し，「imported : SSR MISA+Primer 3 」を選択し，さらに「Edit」を選択する．「SSR MISA+Primer 3」のworkflowが表示される（図1-4-5b）．

④ 「Parse BLAST」ボックス内の「merge_out（tabular）」の右部にある印を選択する．選択した項目については，解析後に解析結果を記載したファイルが作成される．必要に応じて他の項目，「MISA」ボックス内の「misa_statistics（tabular）」（配列内に検出されたSSR数をモチーフ・反復回数ごとにカウントしたデータ）等を選択する（図1-4-5b）．

⑤ 「Workflow Canvas」右上の設定ボタンをクリックし，「Save」を選択する．続けて設定ボタンの中から「RUN」を選択する．

⑥ 各項目の設定を行う．「Step 1 :Input dataset」の「Input Dataset」ではマーカー設計対象配列のマルチFASTAファイルを，「Step 2 :Input dataset」の「Reference for BLAST」では参照配列を選択する．

⑦ 「Step 3 :MISA」では，検出するSSRのモチーフの塩基数（definition（unit_size））と，その最低反復回数（definition（min_repeats））を定義する．

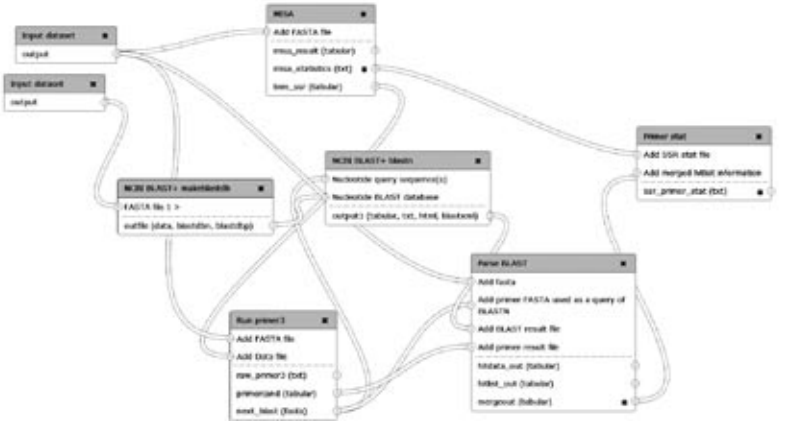
⑧ 「Step 5 :Run primer 3 」では，プライマー設計のパラメータ（PCR産物長，プライマー長，Tm値）を設定する．

⑨ 最下部の「Run workflow」をクリックし，解析を始める．
解析終了後，画面右部のHistory内に，設計されたプライマー配列が記載された「Merged hitlist information」が生成される（図1-4-5c）ので，データをダウンロードする．ダウンロードしたデータはテキストエディタやExcelで閲覧・編集が可能である．

A

```
>100057-
AATCCCATTTATTAAGGAAATAAAATTGATTTTCTAAATAATTCAATTCTATTTCTGTTTTCAGTAACTGTCATTTCAACAGTTAGTTCT
CCATTTTGACCAAAATTTACGTAAACTAACTACTTACTCATTGTTGACGTAAGACCAGAGAAATTGATTGGAATTTGGAATAAAGAGC
TTAATTTCAAAATTGAAAGTGGAAAATGGGTCCATCTTTATTTTGAAGAACCCTAATTTCCTTTACTAGCCAACCAAATCCTAATCCAT
CATCATTAACGCCACCAACCATTTTCCATTCATCTAGAAGTCCAATTGTGCGTGGCATCGTGTCCGGAAAATCAAAATCTCAATAATA
TAATAAAGGCGGAAAGACTTCTTCCCCCACCCATCTGCACCGTGAGAGATAAGGACATTGTCTTTATTGTGTAGTCATTGTAGAGAG
AGAGAGAGAGAGAGATCTCATCGAATGGTATTCGTCAATGAAAAAGAAAAGAGAGATAATTATCAAATTTCAACGTC-
>101877-
AGTTGAGGATGAAGATAAATATATCTTTGATTTATCTCATCTCTGATTCATTTCCTCTCTTACATGTTTCTTAAACTCTTATATATTGAA
TTACCTTTTAAAATTTTGGGTTATTACACACTTATGTATATTATTATATAATTGAATATTACTTTATCTTTAATTTAAAATTATTCAATTGT
ATAATGACACATATTAAATATGTACTTATTTATGAACTTAAAATTGGTATATATAATTTTATTACAGTTAAAACATATTCTATTATTTTTCT
AGTAAATATAGGGAGAGGAAGGGGAGAAATATTTTTCCCATGGGAAAGGACAAAGATTTTCCTGTTATTTTCGTGTTGAGGACATGG
TCGGGATGGGGATAAATGTCTCTCTCTCTCTCTCTCTCTCATGGAGCACATGCAAATTAATTAATTTGAGTGCAGAGTTATTTTCTT
```

B



C

Sequence ID	SSR length	SSR start position	SSR end position	Left Primer position	Right Primer position	SSR Sequence
100057_437	20	437	456	320	498	AGAGAGAGAGAGAGAGAGAG
101877_380	24	380	403	356	598	TCTCTCTCTCTCTCTCTCTCTCTC

図 1-4-5　Galaxy/NIAS 内の「MISA+Primer 3」のワークフロー（https://galaxy.dna.affrc.go.jp/nias/）を用いた SSR マーカー設計

A：SSR マーカーを設計する対象配列．ファイルはマルチ FASTA 形式でアップロードする．

B：「MISA + Primer 3」のワークフロー．ファイルとして出力したい場合は＊にチェックを入れる．図中では，「MISA」の「misa statistics（txt）」，「Parse BLAST」の「mergeout（tabular）」および「Primer stat」の「ssr_primer_stat（txt）」にチェックを入れている．

C：「MISA + Primer 3」ワークフロー内「Parse BLAST」の「mergeout（tabular）」より出力される結果の一部．「mergeout（tabular）」には SSR の位置，モチーフおよび反復回数，プライマーの配列等が記載されている．（Galaxy/NIAS より引用，http://galaxy.dna.affrc.go.jp/）（Galaxy/NIAS より引用，http://galaxy.dna.affrc.go.jp/）

設計されたSSRマーカーから，目的に沿ったマーカーを選定することでより望ましい結果が得られると期待される．連鎖地図を作成する目的ではアリル数や多型程度が高い特徴がある（Merritt *et al.* 2015）2塩基モチーフが適している．また2塩基モチーフの中でも（AT）nモチーフや（GC）nモチーフはGC含量の極端な偏りがあるためPCRでの増幅安定性に欠けるため，（AG）nおよび（AC）nモチーフを優先的に選定すると良い．一方，品種識別を行う目的では，スタッターバンドが発生しにくく（Guichoux *et al.* 2011）誤判定が起こり難い3塩基以上の多塩基モチーフが適している．

c．EST－SSRの特徴

SSRマーカーは設計する基となる配列により，トランスクリプトームシークエンスを基にしたEST-SSRマーカーとゲノムシークエンスを基にしたgenomic-SSRマーカーに大別される．EST-SSRは遺伝子の転写領域に存在するSSRであり，genomic-SSRは全ゲノムDNA上に存在する全てのSSRに由来する．EST-SSRはSSR全体のうち5%以内程度である．SSR全体では2塩基モチーフが最も多く観察されるが，転写領域上に存在するSSRでは3塩基モチーフが最も多く見られる．転写領域はコーディング領域（coding region）と非翻訳領域（untranslated region, UTR）とに分けられるが，コーディング領域上に存在するSSRは大半が3塩基モチーフである（Varshney *et al.* 2005）．コーディング領域に3塩基モチーフが多い理由として，反復回数に変化が起こってもフレームシフト（frameshift）を起こさないためと考えられている（Metzgar *et al.*, 2000）．一方，UTR上に存在するSSRは大半が2塩基モチーフである．

一般に転写領域，特にエキソンはその他の領域に比較して配列保存性が高いことが知られているが，その配列保存性の高さから，ある生物種のEST-SSRマーカーが近縁種において利用できることがある．例えばアンズ（*Prunus armeniaca*）のESTより開発されたEST-SSRマーカーが，同科異属であるナシやリンゴにおいて利用されている（Decroocq *et al.* 2003）．そのため近縁種における利用を目的とする場合にはコーディング領域上のEST-SSRを選択するのが適している．一方で，その配列保存性の高さから，他の領域上のSSRに比

較して多型程度が低い（Varshney *et al.* 2005）点には注意が必要である．一方，UTR上のSSRはコーディング領域上のSSRよりも多型が多い傾向がある（Scott *et al.* 2000）．そのため，連鎖地図作成を目的とする場合にはUTR上のSSRを選択するのが適していると考えられる．

d．果樹で実施されたNGS解析データからのSSRマーカー設計

NGS解析データ中からのSSRマーカー設計例は多く，果樹においてはマンゴー（*Mangifera indica* L.）（Ravishankar et al. 2015），バナナ（*Musa acuminate* Colla）（Passos *et al.* 2013），ウメ（*Prunus mume* Sieb. et Zucc.）（San *et al.* 2013）などが報告されている．

(2) レトロトランスポゾン挿入多型マーカー

a．レトロトランスポゾン

レトロトランスポゾンは，真核生物（eukaryote）のゲノム内に存在し，複製型転移（replicative transposition）を行う可動遺伝因子（movable genetic element）である．自身をRNAに転写した後，逆転写酵素（reverse transcriptase）によりDNAに逆転写され，その後インテグラーゼ（integrase）の働きにより異なるゲノムDNA領域へ組み込まれる，すなわちコピー アンド ペースト型の複製を行う．レトロトランスポゾンは5′および3′の両末端に100 bp程度から数千bp程度の長さのlong terminal repeat（LTR）と呼ばれる反復配列を有するLTR型レトロトランスポゾンと，LTRを有さない非LTR型レトロトランスポゾンに類別される．LTR型レトロトランスポゾンにおいては，両末端のLTRは互いに高い配列相同性を示し，またコピー アンド ペーストにより増えた，由来の同じレトロトランスポゾン間では，別々のコピーであってもLTRは高い配列相同性を示す．しかし由来の異なるレトロトランスポゾン間においては，LTRの配列相同性は低い性質がある．LTR型レトロトランスポゾンの特徴として，5′LTRの下流にはtRNAの相補配列であるprimer binding site（PBS）と呼ばれる転写開始点が存在し，3′LTRの上流にはpolypurine tract（PPT）と呼ばれる，逆転写反応の開始点が存在するこ

とが知られている（図1-4-6）．LTR型レトロトランスポゾンはPBSからPPTの間に位置する配列により，さらにTy1-copia型，Ty3-gypsy型，Trim型，LARD型に分類される（Kumar *et al.* 1999）（図1-4-6）．

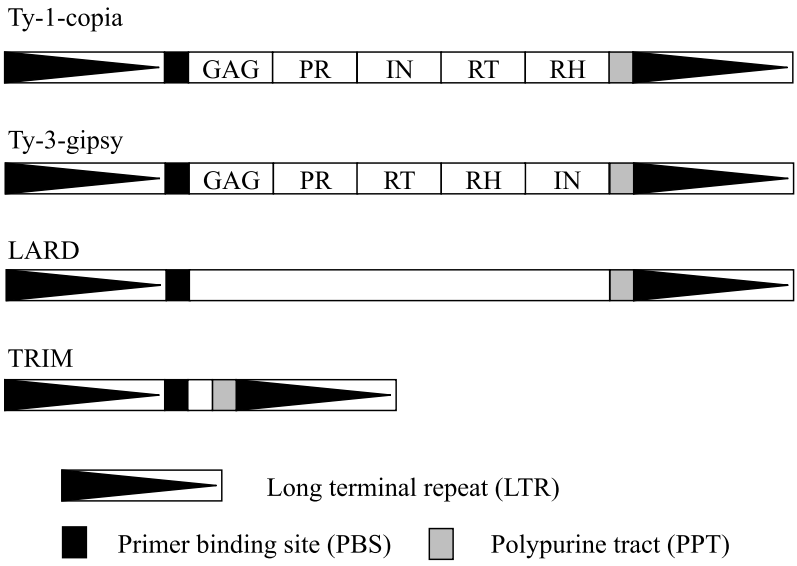


図 1-4-6　LTR 型レトロトランスポゾンの種類とその構造

LTR 型レトロトランスポゾンは両端に相同性の高い配列である long terminal repeat（LTR）を有しており，5′側 LTR の下流には転写開始点となる primer binding site（PBS）が，3′側 LTR の上流に polypurine tract（PPT）が存在する．Ty1−copia，Ty−3−gypsy，TRIM および LARD は，PBS から PPT の間の構造により分類される．Ty−1−copia および Ty−3−gypsy は内部に capsid proteins（GAG），protease（PR），integrase（IN），reverse transcriptase（RT），RNase H（RH）をコードしている．LARD はコード領域が長鎖の非コード領域に置き換わっており，TRIM ではほとんどの部分を失っている．

b．レトロトランスポゾン挿入多型マーカーの種類

レトロトランスポゾンの挿入多型を利用したDNAマーカーはいくつか知られており，二つのレトロトランスポゾンとその挿入間を増幅するinter-retrotransposon amplification polymorphism（IRAP），レトロトランスポゾンとその近傍に存在するSSRとの間を増幅する retrotransposon microsatellite amplification polymorphism（REMAP），レトロトランスポゾンとその近傍に

存在する制限酵素（restriction enzyme）認識配列との間を増幅するsequence-specific amplified polymorphism（S-SAP）などが知られている（Kalendar *et al.* 2006, Schulman *et al* 2007）．これらの手法では，PCRプライマーの片側をLTR配列上に設計し，もう片側についてLTR，SSR，制限酵素サイトなどゲノム上に多数存在する配列上に設計する．PCRを行うことで数十から数百のレトロトランスポゾン挿入－近傍配列間が増幅されるので，増幅産物のパターンをアガロースゲル電気泳動等で検出する．

単一の増幅産物が得られ，またランダムショットガンシークエンスや公開ゲノム配列から設計可能なDNAマーカーとして，レトロトランスポゾン挿入多型（retrotransposon based insertion polymorphism, RBIP）マーカーが挙げられる．RBIPマーカーは，LTRおよび隣接するゲノム領域にプライマーを設計し，PCR増幅することで，レトロトランスポゾン挿入アリルを検出する（図1-4-7, 8）．本項ではRBIPマーカー設計について解説する．

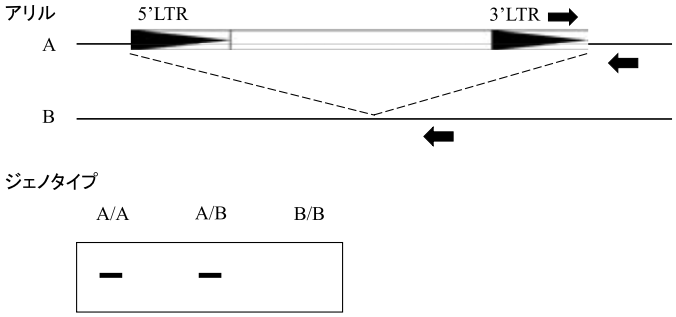


図 1-4-7　RBIP マーカーの模式図

LTR と隣接するゲノム領域にプライマー（矢印）を設計することで，レトロトランスポゾンが挿入されている場合アリル特異的に増幅される．挿入がない場合は増幅しない．このマーカーは優性マーカーであるため，レトロトランスポゾン挿入アリルをホモで有しているかヘテロで有しているかは判定できない．

c．LTR_FINDER を用いた LTR の同定

RBIPマーカーの設計には，レトロトランスポゾン挿入のされたゲノム領

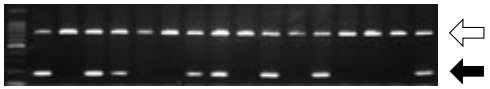


図 1-4-8 RBIP マーカーの電気泳動図

白い矢印はポジティブコントロール（クロロプラストゲノムコード遺伝子の rbcL），黒い矢印は RBIP マーカーによる増幅産物．

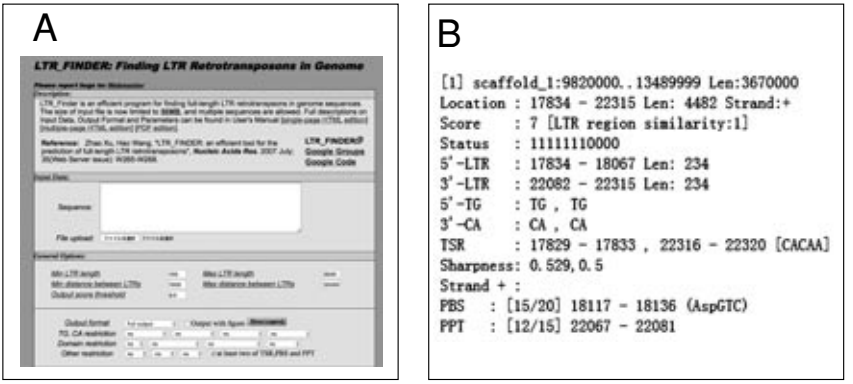


図 1-4-9 LTR_FINDER

A：LTR_Finder のウェブサイト．上記インターフェイスから，50 Mb までの塩基配列を入力することができる．また，検索する LTR 長，LTR 間の距離等を設定できる．
B：LTR_FINDER 処理で出力されるファイル．配列中における LTR 型レトロトランスポゾン全長，5′LTR，3′LTR，PBS および PPT の位置と長さが記載されている．（LTR_FINDER より引用，http://tlife.fudan.edu.cn/ltr_finder/, S）

域およびLTR配列の同定が必要である．ゲノム配列が公開されている，あるいはゲノムDNAの高精度なアセンブルデータを所持しているのであれば，LTR_FINDER（URL1-4-11, Xu *et al.* 2007, 図1-4-9a）を用いることでLTRを含む全長レトロトランスポゾン配列が同定できる．LTR_FINDERは5′LTRと 3′LTRが高い相同性を示す性質を利用し，指定した長さのLTR配列（初期値:100～3500 bp）のペアが，指定した長さ（初期値:1000～20000 bp）の間に存在するかどうかを検索することで，5′LTRおよび 3′LTR配列の長さや位置を同定する（図1-4-9b）．LTR配列およびその位置が同定されることで，同時に隣接するゲノム領域が判明する．LTRとゲノム領域にプライマーを設計することでRBIPマーカーとなる．5′LTRから 3′LTRまで，レトロトランスポゾ

ンの全長が一つの配列内に存在するような，長い配列を用いることができる場合にLTR_FINDERは有用であるが，NGS解析データをそのまま使用するなど対象配列が充分に長くない場合は，LTR_FINDERによるLTRの同定は困難である．

d．ランダムショットガンシークエンスからの LTR の同定

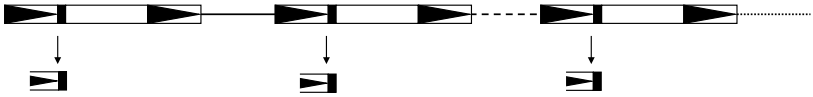
LTR_FINDERによるLTR同定には高精度なアセンブル済みの配列データや，公開ゲノム配列が必要であるが，ゲノムDNAのランダムショットガンシークエンスデータがあればLTRが同定できる．この方法は，1．各レトロトランスポゾンの5′LTRと3′LTRは高い配列相同性を示す．2．レトロトランスポゾンの5′LTRの下流にはPBSが存在する．3．異なるレトロトランスポゾンのLTR配列間では配列相同性が低い．4．コピー＆ペーストにより増えた由来の同じレトロトランスポゾンでは，別々のコピーであってもLTRは高い配列相同性を示す，という性質を利用したものである（図1-4-10）．下記方法により，奈島ら（2015）はパインアップルのRBIPマーカー設計を実施し，約700個のRBIPマーカー候補配列を得た．

① 全シークエンス中よりPBS配列を有するシークエンスを抽出し，さらにPBSより上流の配列についてマルチFASTA形式の配列リストを作成する（Excel，Perlプログラム等）．作成した配列リストはPBSの上流に存在する5′LTR配列のリストと推定する．

＊植物のレトロトランスポゾンでは，PBS配列として開始メチオニンのtRNAの相補配列（5′-TGGTATCAGAGC-3′）が広く用いられている（Monden *et al.* 2014, 2015）

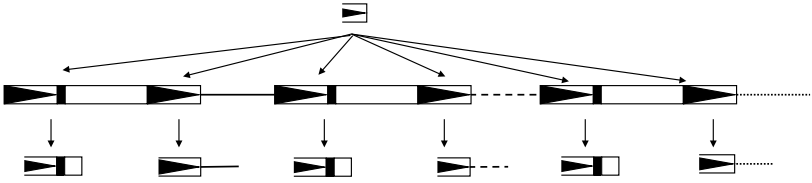
② ①で作成した5′LTR配列リストについて，BLASTclust（URL1-4-12）を用いて，相同性の高い配列を集約する処理であるクラスタリングを行う．コピー数が多いレトロトランスポゾンのLTRである場合には，クラスターを構成する配列数が多くなる．BLASTclust処理により，各レトロトランスポゾンファミリーのLTR配列が得られる．

①ショットガンシークエンスから、PBS配列およびその上流配列を抽出する．

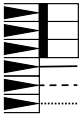


②抽出された配列についてクラスタリングを実施する。PBSの上流を5′LTRと推定する。

③推定された5′LTRをqueryに、全シークエンス配列をデータベースに指定してBLAST検索を行う。5′LTRおよび3′LTRを含む周辺配列が得られる．



④ヒットしたシークエンスについて、アライメントを作成する。3′LTRの下流にはPBSがなく、3′LTRの下流では配列間の相同性がない．配列相同性のない領域をゲノム領域と推定する．



⑤推定されたゲノム領域とLTR領域でプライマーを設計し、RBIPマーカーとする。

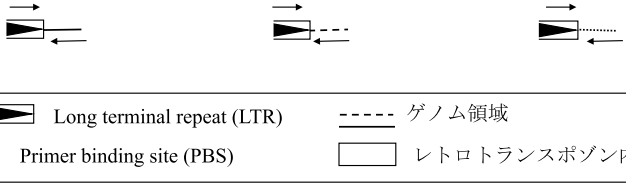


図 1-4-10　ゲノム DNA のショットガンシークエンスからの RBIP マーカー設計の模式図

③　各クラスターを代表する配列をクエリに，全シークエンス配列をデータベースに指定してBLAST検索を行う．ヒットした配列には，3′LTRと 5′LTRの両方の 3′末端部が含まれる．

＊任意の配列をデータベースとしたBLAST検索は，Galaxy/NIASやCLC MainWorkbench（Qiagen社）などで実行できる．また大学共同利用機関法人 情報・システム研究機構のライフサイエンス統合データベースセンターが作成しているウェブサイトである統合TV（URL1-1-18）では自己のコンピュータ上にBLAST実行環境を構築し，実行する方法が解説されている．

④ ヒットした配列について，CLC MainWorkbench等を用いてアラインメントを作成する．作成されるアラインメントは2種類に大別される．5′LTRを含む配列はPBS配列およびその下流のレトロトランスポゾンのコード領域を含むため，配列間の相同性が極めて高い（図1-4-11A）．一方3′LTRを含む配列では，LTR配列は配列間相同性が高いが，LTR配列の下流では配列間相同性が見られなくなる（図1-4-11B）．配列間相同性が見られない領域は，レトロトランスポゾンに隣接するゲノム領域であると考えられる．

⑤ 得られたレトロトランスポゾン挿入箇所の配列について，BLASTclustを用

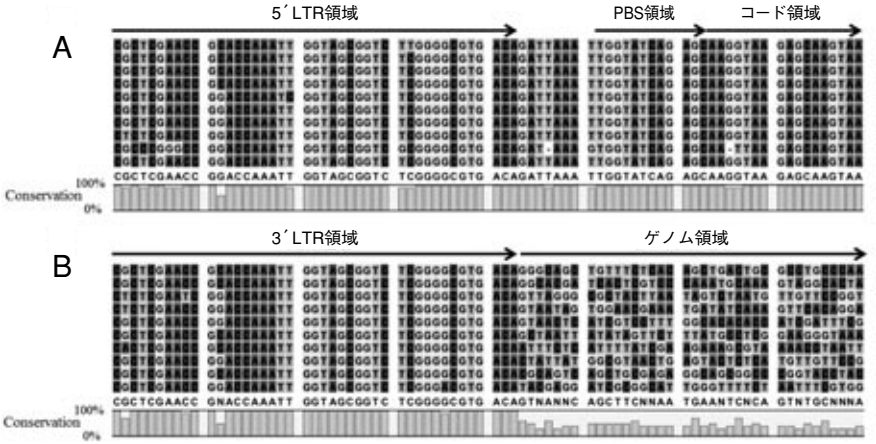


図1-4-11　5′LTRおよび3′LTRのアラインメント（奈島ら，2015から引用）
A：5′LTR由来のシークエンスアラインメント．
B：3′LTR由来のシークエンスアラインメント．

いてクラスタリングを行い，重複する配列を削除する．得られたユニークなゲノム挿入領域と，LTR領域でプライマーを設計し，RBIPマーカーとする．

e．果樹で実施された NGS 解析データからの RBIP マーカー設計

これまでに果樹においての実施例は少なく，ニホンナシ（*Pyrus pyrifolia* Nakai）（Kim *et al.* 2012），およびパインアップル（*Ananas comosus*（L.）Merr.）（奈島ら 2015）において設計されている．

(3) SNP マーカー

a．SNP および SNP マーカー

一塩基多型（single nucleotide polymorphism, SNP）はアリル間でDNA配列上のひとつの塩基が異なっている多型を指す．SNPはゲノムDNA上に最も多く存在する多型であり，数百塩基に一つ程度存在する．SSRやレトロトランスポゾン挿入と比較して数が多いため，SNPをマーカー化することで高密度な連鎖地図が作成できる．

個別のSNPについてDNAマーカー化する方法は複数開発されている．アガロースゲルで検出可能な方法としては，cleaved amplified polymorphic sequence（CAPS）法（Konieczny *et al.* 1993），derived amplified polymorphic sequence（dCAPS）法（Neff *et al.* 1998），allele-specific PCR法などが知られている．CAPS法はSNPを含む領域についてPCRを行い，その後特定の制限酵素サイトによる切断の可否でアリルを区別する手法である（図1-4-12）．CAPS法ではSNPが制限酵素の認識配列上に存在する必要がある．dCAPS法はSNPが制限酵素の認識配列に含まれない場合に用い，片方のプライマーをSNPに隣接させて制限酵素の認識配列に組み込む手法である．Allele-specific PCR法は，対象アリルのみ増幅するよう 3′末端部にSNPが含まれるPCRプライマーを設計し，PCRを行い識別する方法である．

キャピラリー電気泳動によるフラグメント解析により識別する方法として，SNaPshot（Applied biosystems社）が知られている．SNapShot法はプライマーをSNP直前に設計し，一塩基のみ伸長させて取り込んだ塩基の種類を判別す

ることでSNPを識別する手法である．

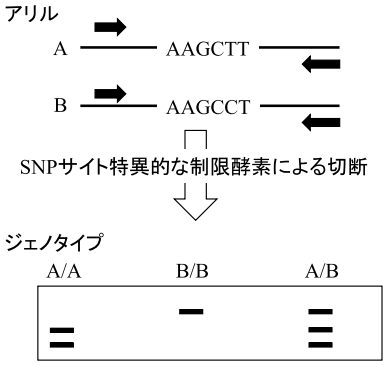


図 1-4-12　CAPS 法の模式図

SNP サイトを挟む PCR プライマーを設計し，PCR を行った後 SNP サイトに特異的な制限酵素で切断する．本図の場合では A アリルに存在する AAGCTT を認識する制限酵素 HindIII 処理を行い，HindIII による切断の有無でジェノタイピングを行う．

リアルタイムPCRを利用する方法として，TaqMan（Applied biosystems 社），Tm-Shift PCR（Garmer *et al.* 1999），HybProbe（Roche 社），High Resolution Melt（HRM）解析（Vossen *et al.* 2009）などが知られている．TaqManは，SNPの存在する領域に，相補的に結合する，すなわちハイブリダイズ（hybridize）するプローブを用いる．このプローブにはレポーター色素と消光分子（quencher, クエンチャ）が結合しており，SNPにプローブがハイブリダイズしていた場合にはPCRの進行時にプローブが分解を受け，レポーター色素が放出される．放出された蛍光を検出することで，SNPアリルが判定される．Tm-Shift PCRでは，それぞれのアリル特異的プライマーに異なる長さの配列を付加して増幅産物の融解温度（Tm）の差異を拡大し，リアルタイムPCR装置を用いて遺伝子型を判定する手法である（Garmer *et al.* 1999）．HybProbeでは，2本の配列特異的なプローブが目的配列にハイブリダイズした際に，それぞれのプローブに結合されたドナー色素とアクセプター色素が近接し，蛍光共鳴エネルギー転移（fluorescence resonance energy transfer,

FRET)が起こる．その際発生する蛍光を検出することでSNPアリルを識別する．HRM解析は，温度上昇にともなう2本鎖DNAの1本鎖DNAへの解離度をモニタリングする融解曲線分析を行い，その解離パターンの違いからSNPアリルを識別する．

遺伝子型を決定する作業であるジェノタイピング（genotyping）を多数のSNPについて行うには，SNPアレイやMassARRAY（Agena bioscience社），genotyping by sequencing（GBS）法を利用する．Infinium（Illumina社），Axiom（Affymetrix社）などのSNPアレイにおいては，各アリルに特異的にハイブリダイズするオリゴDNAを用い，ハイブリダイズした場合のみ蛍光が検出される手法が用いられている．SNPアレイはInfiniumで3,000以上，Axiomで1,500以上のSNPを一度に解析するシステムであり，数十から数百SNP程度の中規模のSNPジェノタイピングには不向きであること，多くの果樹ではSNPアレイが設計されていないため，カスタムSNPアレイの設計を行う必要がある点は注意が必要である．MassARRAYにおいては，SNPサイトの1塩基手前までのプライマーを用い，一塩基の伸長反応を行った後，伸長された塩基をMALDI-TOF-MSを用いて検出・特定する手法が用いられている．SNPアレイと比較して小規模の解析（数個～数十個のSNP，48以上のサンプル）を行うことができることが利点である．GBSは，シークエンシングによるジェノタイピング手法であり，NGSを用いたSNP検出を各個体について行うことで直接ジェノタイピングを行う（Elshire *et al.* 2011）．GBSはシークエンシングライブラリの調製方法によって検出するSNP数を調節することができることが利点として挙げられる．

b．SNPの検出

SNPの検出では全ゲノム配列やアセンブル後の配列などの参照配列に対してNGSなどから得られた配列のマッピングを実施した後，マッピングデータ中からSNPが存在している箇所を探索する．Galaxy/NIASのワークフローを利用することでSNPが検出できる．ワークフロー内では，マッピングにBWA（URL1-4-13, Li *et al.* 2009a），SNP検出にSAMtools（URL1-4-14, Li *et al.* 2009b）

を使用している．

A

```
@HWI-ST1387:154:HABLAADXX:1:1101:12540:2416 2:N:0:GGCTAC
CATTTACGCATTTTACTCTATGCTTTATTGTATTTAATGGCTGGTCATGTTTATTGACTCGCTTTACATGTCAT
+
CCCFFFFFHGHHHJJJJJJJJJJJJJJJJHIJJJJJJJJJJJBGHIGFHI?DHAFH@F@A@GG;CGGDEG>C
```

B

```
chr1  gbrowse  CDS  28371966  28372044  .  +  0  ID=MDP0000174606_exon_1;Parent=MDP0000174606;CONTIG=MDC003205.158
```

C

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO
lcl|LG01	2198607	.	A	G	28	.	DP=6;VDB=7.040000e-02;RPB=0.000000e+00;A
lcl|LG01	23128120	.	G	A	21	.	DP=6;VDB=7.680000e-02;RPB=0.000000e+00;A
lcl|LG02	10618118	.	A	G	49	.	DP=6;VDB=5.365061e-02;RPB=1.579671e+00;A
lcl|LG03	2846760	.	C	T	33	.	DP=6;VDB=7.840000e-02;RPB=1.111223e+00;A
lcl|LG04	965090	.	G	A	34	.	DP=6;VDB=3.680000e-02;RPB=8.071455e-01;A
lcl|LG04	2193737	.	G	A	57	.	DP=6;VDB=4.493050e-02;RPB=-1.579671e+00;

図 1-4-13　Galaxy/NIAS 内「BWA mapping Illumina」,「SNP/INDEL Calling」および「SNP Marker」ワークフローを用いた SNP 検出に必要なファイルと出力されるファイル

A : NGS 解析で得られた FASTQ 形式のファイル．1 行目は @ から始まる配列名，2 行目に塩基配列，3 行目は文字「+」，4 行目に 2 行目に記載した配列のクオリティ値が記載されている．

B : gff 形式の（GDR より引用）．左から,「配列名」「配列のタイプを決めたプログラム」「配列のタイプ」「開始位置」「終了位置」「スコア」「ストランド」「フレーム」「グループ」が記載されている．

C :「SNP marker」ワークフロー内「Marker selection」の「output_vcf（vcf）」の vcf 形式のファイル．左から「CHROM（染色体番号）」「POS（位置）」「ID（ID）」「REF（リファレンス塩基）」「ALT（代替塩基）」「QUAL（クオリティ値）」「FILTER（フィルター）」「INFO（付加情報）」が記載されている．（Galaxy/NIAS より引用 、http://galaxy.dna.affrc.go.jp/, S）HindIII による切断の有無でジェノタイピングを行う．

① NGS解析で得られたシークエンスおよび参照配列について「BWA mapping Illumina」を実施する．「BWA mapping Illumina」ワークフローを「Published workflow」からインポートする．「BWA mapping Illumina」についてEditを選択後,「Mark duplicate read」中の「out_file (bam)」にチェックを入れ,「Save」する．マルチFASTA形式の参照配列と，FASTQ形式（図1-4-13a）のNGSシークエンスを指定し，実行する．実行後，マッピング情報が記載されたBAMファイルが作成される．

② ①で得られたBAMファイルについて,「SNP/INDEL Calling」を実施する．「Published workflow」からインポート，Editを選択後に,「Unity VCF of GATK-SAMtools」の「com_vcf_file (vcf)」にチェックを入れ,

セーブする．実行時に①で得られたBAMファイルと①で使用したリファレンスを指定後，実行する．SNP情報が記載されたvcfファイルが作成される．

③ ②で得られたvcfファイルについて「SNP Marker」ワークフローを実施する．「SNP Marker」ワークフローでは，SNPマーカーの設計位置として適しているかどうか（1．検出されたSNPの周辺に別のSNPがない．2．検出されたSNPの確からしさを示す指標であるvariant quality scoreが一定値（default=20）以上である．3．SNPサイトにおける配列の厚みが一定値（default=5）以上である 4．参照配列中に1箇所のみ存在する,ユニークな配列である.）が判定される.手順としては,「Published workflow」から「SNP Marker」ワークフローをインポート，Editを選択後に,「Marker selection」の「output_vcf（vcf）」にチェックを入れる．本ワークフロー中では，リファレンスに対応したアノテーション情報が記載されたGFFファイル（図1-4-13b）が必要であるので，GFFファイルがある場合にはアップロードする．GFFファイルがない場合,「Search for GFF」を削除し,「Marker BLAST」内の「megablastinfo_out（vcf）」と「Marker selection」内の「Add marker format VCF」を繋げ，セーブする．実行時に②で得られたvcfファイルと,リファレンスを指定する．またGFFファイルがある場合には「Step12:Search for GFF」で指定したfeatureを有する領域をSNPマーカー選抜のターゲットとして指定することができる．解析後，作成されるvcfファイルにSNPマーカーとして適しているSNPが記載されている（図1-4-13c）．

なおSNP検出では，見出されるSNPの確からしさの保証のためには配列の厚みの確保が重要である．参照配列へNGSシークエンスをマッピングした際に，マップされた配列の厚みが少ない領域では，仮にSNP候補が見出されたとしても，それがシークエンスエラーなのか，実際にSNPなのかを判断をすることが困難である（図1-4-14）．そのためGalaxy/NIASの「SNP marker」ワークフローにおいては，SNPサイトにおける配列の厚みが5以上あることが，SNP

マーカー設計に適したSNPであると判断される条件の一つとなっている.また，Illumina社の公表しているテクニカルノート「Calling Sequencing SNPs」においては,推定したSNPの99%以上が実際のSNPであることを保証するためには配列の厚みが30x以上，95%以上ならば約20x以上が必要であるとされている．

SNPは数百塩基対に一つ程度とゲノム中に多数存在するため，全SNPを対象としてSNPマーカーの設計やジェノタイピングをするのは過剰である場合がある．そのような場合にはシークエンスを行う領域を絞ってSNPを探索する手法が用いられる．代表的な手法として，restriction site associated DNA sequence（RAD-seq）（Davey *et al.* 2010）が挙げられる．RAD-seqは制限酵素処理を利用し，制限酵素認識配列近傍のごく一部の配列のみをNGSで解読する手法である．ゲノム上の限られた領域に絞ってシークエンスを行うことができるので，全ゲノムでの解析に比較して低コストで一定量の配列の厚みが確保できることが利点ある．

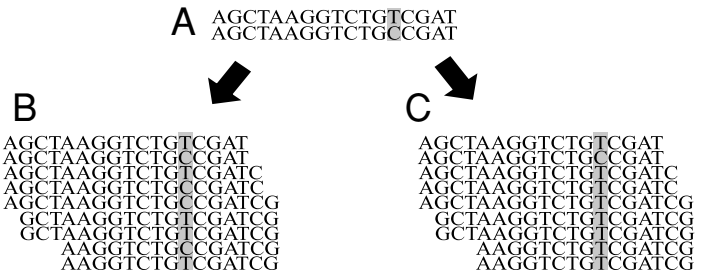


図 1-4-14　SNP 検出における配列の厚みの重要性

A：配列の厚みが少ない場合．塩基に違いがあるが，配列の厚みが少なく，シークエンスエラーなのか SNP なのか判別が困難である．
B：配列の厚みが充分ある SNP 領域．T/C の存在比（コールレート）から T/C の SNP がある領域だと推定される．
C：配列の厚みが充分あるシークエンスエラー領域．T/C のコールレートから T から C へのシークエンスエラーであると推定される．

c．果樹において実施された SNP マーカー解析

ニホンナシ（Terakami *et al.* 2014），モモ（Martínez-García *et al.* 2013）およびリンゴ（Bianco *et al.* 2014）においてはNGS解析データからSNPアレイ開発，さらにジェノタイピングによる連鎖地図作成が実施されている．ま

たブドウにおいてはMyles *et al.*（2010）が，カンキツにおいてはFujii *et al.*（2013）がSNPアレイを作成している．またナツメ（*Ziziphus* Mill.）（Zhao *et al.* 2014），カンキツ（Guo *et al.* 2015），モモ（Bielenberg *et al.* 2015）ブドウ（Chen *et al.* 2015）においては，Rad-seqによるGBSが実施されている．

4）今後の展望

かつてDNAマーカーの開発に要する費用と手間が大きく，DNAマーカーが十分に開発されていた植物種は限られていた．現在はNGSの出現およびバイオインフォマティクス技術の発展により，DNAマーカー開発は格段に安価かつ簡易になり，多くの植物種でDNAマーカー開発が実施されている．今後，各樹種においてDNAマーカー開発が進み，連鎖地図作成やQTL解析が実施されることが期待される．

引用文献

Altschul, S. F. *et al.*（1990）Basic local alignment search tool. Journal of Molecular Biology. 215:403-410.

Bianco, L. *et al.*（2014）Development and validation of a 20k single nucleotide polymorphism（SNP）whole genome genotyping array for apple（*Malus x domestica* Borkh）. Plos One. 9:e110377.

Chen, J. *et al.*（2015）Construction of a high-density genetic map and QTLs mapping for sugars and acids in grape berries. BMC Plant Biology. 15:28.

Davey, J. W. *et al.*（2010）RADSeq:next-generation population genetics. Briefings in Functional Genomics. 9:416-423.

Decroocq, V. *et al.*（2003）Development and transferability of apricot and grape EST microsatellite markers across taxa. Theoretical and Applied Genetics. 106:912-922.

Duvick, J. *et al.*（2008）PlantGDB:a resource for comparative plant genomics. Nucleic Acids Research. 36:D959-D965.

Elshire, R. J. *et al.*（2011）A robust, simple genotyping-by-sequencing（GBS）

approach for high diversity species. Plos One. 6:e19379.
Fujii, H. *et al.* (2013). High-throughput genotyping in citrus accessions using an SNP genotyping array. Tree Genetics and Genomes. 9:145-153.
Garmer, S. *et al.* (1999) Single-tube genotyping without oligonucleotide probes. Genome Research. 9:72-78.
Goecks, J. *et al.* (2010) Galaxy:a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biology. 11:R86.
Guichoux, E. *et al.* (2011) Current trends in microsatellite genotyping. Molecular Ecology Resources. 11, 591-611.
Guo, F. *et al.* (2014) Construction of a SNP-based high-density genetic map for pummelo using RAD sequencing. Tree Genetics and Genomes. 11:2.
Jung, S. *et al.* (2014) The Genome Database for Rosaceae (GDR):year 10 update. Nucleic Acids Research. 42:D1237-1244.
Kalendar, R. *et al.* (2006) IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. Nature Protocols. 1:2478-2484.
Kalia, R. K. *et al.* (2011) Microsatellite markers:an overview of the recent progress in plants. Euphytica. 177:309-334.
Kim, H. *et al.* (2012) Development of cultivar-specific DNA markers based on retrotransposon-based insertional polymorphism in Japanese pear. Breeding Science. 62:53-62.
Konieczny, A. *et al.* (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant Journal. 4:403-410.
Li, H. *et al.* (2009a) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25:1754-1760.
Li, H. *et al.* (2009b) The Sequence Alignment/Map (SAM) format and SAMtools. Bioinformatics. 25:2078-2079.
Martínez-García, P. J. *et al.* (2013) High density SNP mapping and QTL analysis for fruit quality characteristics in peach (*Prunus persica* L.). Tree Genetics and Genomes. 9:19-36.
Merritt, B. J. *et al.* (2015) An empirical review:Characteristics of plant microsatellite markers that confer higher levels of genetic variation. Applications in Plant Sciences. 3:1500025.
Metzgar, D. *et al.* (2000) Selection against frameshift mutations limits microsatellite

expansion in coding DNA. Genome Research. 10:72-80.

Myakishev, M. M. *et al.* (2001) High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer-labeled primers. Genome Research. 11:163-169.

Myles, S. *et al.* (2010) Rapid genomic characterization of the genus *Vitis*. Plos One. 5:e8219.

Monden, Y. *et al.* (2014) Efficient screening of long terminal repeat retrotransposons that show high insertion polymorphism via high-throughput sequencing of the primer binding site. Genome. 57:245-252.

Monden, Y. *et al.* (2015) Plant transposable elements and their application to genetic analysis via high-throughput sequencing platform. The Horticulture Journal. 84:283-294.

奈島賢児ら (2015) パインアップルにおけるレトロトランスポゾン挿入多型マーカー開発と品種識別への適用. DNA多型. 23:29-33.

Neff, M. M. *et al.* (1998) dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. Plant Journal. 14:387-392.

Passos, M. A. N. *et al.* (2013) Analysis of the leaf transcriptome of *Musa acuminata* during interaction with *Mycosphaerella musicola*:gene assembly, annotation and marker development. BMC Genomics. 14:78.

Ravishankar, K. V. *et al.* (2015) Development and characterization of microsatellite markers in mango (*Mangifera indica*) using next-generation sequencing technology and their transferability across species. Molecular Breeding. 35:93.

San, L. *et al.* (2013) Genome-wide characterization and linkage mapping of simple sequence repeats in mei (*Prunus mume* Sieb. et Zucc.). Plos One. 8:e59562.

Sawamura, Y. *et al.* (2008) Identification of parent-offspring relationships in 55 Japanese pear cultivars using S-RNase allele and SSR markers. Journal of the Japanese Society for Horticultural Science. 77:364-373.

Schulman, A. H. (2007) Molecular markers to assess genetic diversity. Euphytica. 158:313-321.

Scott, K. D. *et al.* (2000) Analysis of SSRs derived from grape ESTs. Theoretical and Applied Genetics. 100:723-726.

Stein, L. D. *et al.* (2002) The Generic Genome Browser:A building block for a model organism system database. Genome Research. 12:1599-1610.

Terakami, S. *et al.*（2014）Transcriptome-based single nucleotide polymorphism markers for genome mapping in Japanese pear（*Pyrus pyrifolia* Nakai）. Tree Genetics and Genomes. 10:853-863.

Untergrasser, A. *et al.*（2012）Primer3-new capabilities and interfaces. Nucleic Acids Research. 40:e115.

Urasaki, N. *et al.*（2015）Leaf margin phenotype-specific restriction-site-associated DNA markers for pineapple（*Ananas comosus* L.）. Breeding Science. 65:276-284.

Vossen, R. H. *et al.*（2009）High-resolution melting analysis（HRMA）:more than just sequence variant screening. Human Mutation. 30:860-866.

Xu, Z. *et al.*（2007）LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Research. 35:W265-W268.

Yamamoto, T. *et al.*（2003）Parentage analysis in Japanese peach using SSR markers. Breeding Science. 53: 35-40.

Zhao, J. *et al.*（2014）Rapid SNP discovery and a RAD-based high-density linkage map in jujube（*Ziziphus* Mill.）. Plos One. 9:e1098.