

第1章 第5節

網羅的遺伝子発現解析のためのカスタムマイクロアレイの開発 — ニホンナシを事例として —

農研機構果樹研究所 品種育成・病害虫研究領域 西谷 千佳子

1) 果樹における網羅的遺伝子発現解析の概況

網羅的遺伝子発現解析とは、或る特定の細胞や組織中でゲノム上の特定の位置から、細胞の状態の維持、成長と分化、外部刺激への応答などの要因により実際に転写されるRNA全体を解析することである。本節では網羅的発現解析手法の一つとして、マイクロアレイを用いた解析手法について紹介する。果樹研究においては、優良品質、省力適性、耐病虫性などを備えた新しい品種の育成や栽培技術の開発の基盤となるような知見を得ることが求められており、重要形質に関連する遺伝子群の同定は重要である。物質の代謝や酵素反応を測定するだけでは説明が困難であった生理現象を、果樹のもつ複雑なゲノム構造に由来する多重遺伝子族 (multigene family) の発現を個別に解析することにより、より正確に把握できるようになるからである。例えば、エチレンは果実成熟やストレス応答に関連する植物ホルモンであることはよく知られているが、生合成経路のキー酵素であるACC合成酵素 (ACC synthase, ACS) については、一つの植物種に対して複数のアイソザイムが存在し、果実の生長やストレスなどに対して特有のACS遺伝子が発現誘導されることが知られている (Yamagami *et al.* 2003, Alexander and Grierson 2002, Tatsuki and Mori 1999)。ニホンナシにおいても、果実由来のACS遺伝子3種類それぞれについて発現解析を行った結果、果実におけるエチレン生産量に関与する2種類の

ACS遺伝子が同定され、マーカー化された (Itai *et al.* 1998, 2003a, 2003b)。果樹におけるこのような事例を増やすためには、より多数の遺伝子群について、発現様式を短時間で解析する網羅的発現解析が非常に有用である。

果樹の全遺伝子配列についての解析が進んだのは、2007年頃の次世代シーケンサの実用化後のことである。次世代シーケンサにより、ブドウ (Jaillon *et al.* 2007)、パパイヤ (Ming *et al.* 2008)、モモ (The International Peach Genome Initiative 2013)、リンゴ (Velasco *et al.* 2010)、イチゴ (Shulaev *et al.* 2011)、チュウゴクナシ (*Pyrus bretschneideri* Rend) (Wu *et al.* 2013)、セイヨウナシ (*Pyrus communis* L.) (Chagné *et al.* 2014) などのゲノム解読が行われた。これらの研究により、果樹の全遺伝子数は樹種ごとの違いはあるものの、predicted genesの数としてチュウゴクナシでは42,812、セイヨウナシでは43,419、イチゴでは34,809、リンゴでは54,921、モモでは27,852との解析結果が発表され、遺伝子配列情報も公開されている。したがって、果樹でも網羅的遺伝子発現解析のツールが整ってきたと言える。

果樹における網羅的遺伝子発現解析に概況を図1-5-1に示す。図1-5-1は、我が国の主要な樹種 (カンキツ、リンゴ、ブドウ、ナシ、モモ) について Gene Expression Omnibus (GEO, URL1-5-1) に登録されている網羅的遺伝子発現解析の報告数を集計したものである。GEOとは、National Center for Biotechnology Information (NCBI) が維持・管理している遺伝子発現情報のデータベースである。網羅的遺伝子発現解析に基づく論文を発表するときは、査読者や編集者から生データ等を公開することを求められることが多く、GEOはその受け皿の一つである。他にはArray Expressなどもある。登録されるデータ数は年々増加傾向にあり、果樹においては今後も網羅的遺伝子発現解析が果樹研究の一つの柱になると考えられる。

2) 網羅的発現解析の手法

網羅的遺伝子発現データ取得の基本的なイメージは、葉と果実、正常果実と生理障害果実、ホルモン等の無処理果実と処理果実、幼果～成熟果実～過熟

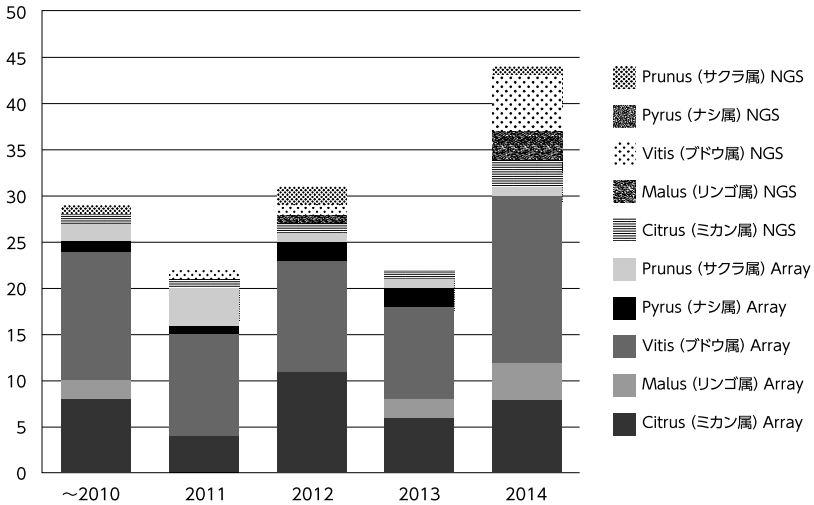


図1-5-1 主要な果樹の網羅的解析のGEO (Gene Expression Omnibus) 登録数
わが国の主要な樹種（カンキツ、リンゴ、ブドウ、ナシ、モモ）についてGEOに登録されている網羅的遺伝子発現解析の報告を集計したものである。

果実といった異なる生育ステージなどの目的のサンプルで発現している全転写産物について、転写産物の種類毎に数値化された発現量情報を得ることである。手法として、マイクロアレイ (microarray) と次世代シーケンサ (next generation sequencer, NGS) を用いたRNA-seqの2つの手法が主流となっている。マイクロアレイについては、初期の論文が1995年に発表された (Schena *et al.* 1995)。次世代シーケンサは2000年代後半に実用化した手法でRNA-Seqも比較的新しい技術である。遺伝子発現の網羅的解析手法として、どちらの方法が良いかはしばしば議論になるが、一般的にはマイクロアレイの方が安価で解析も容易な反面、後述のように事前の配列情報が必要で、マイクロアレイに搭載されていない遺伝子については解析対象とならないのが欠点である。一方で、RNA-Seqは事前の遺伝子配列情報が不要であることが最大のメリットであるが、現時点ではマイクロアレイに比べて高価で解析が困難である。果樹の場合は、現時点ではマイクロアレイを使用した網羅的発現解析の研究報

告の方が多い（図1-5-1）ため、本節ではマイクロアレイを取り上げる。また、果樹の場合は市販のマイクロアレイを使用できる樹種が限定的であるため、マイクロアレイを自作しなければならない場合も多い。そこで、チュウゴクナシ（URL1-1-3）の公開配列情報を利用したカスタムマイクロアレイの作成方法とカスタムマイクロアレイを利用した研究例を紹介する。

3) マイクロアレイを用いた網羅的遺伝子発現解析

(1) マイクロアレイとは

マイクロアレイとは、支持体上に、配列の分かっている数千から数万種類のプローブ（probe）を高密度に整列、固定したスライドガラスである。プローブの本体はcDNAやオリゴヌクレオチド（oligonucleotide）である。Agilent Technologies社の場合はマイクロアレイのプローブは60merのオリゴヌクレオチドである。このオリゴヌクレオチドの5'末端が基盤に結合し、3'側は浮遊している。スライドガラスに対してサンプルから抽出したRNAを逆転写しつつ蛍光でラベリングしたものをハイブリダイゼーションし、それぞれのプローブの蛍光強度を測定することで、そのプローブに対応する遺伝子の発現量を解析できる。マイクロアレイの仕様はメーカーにもよるが、通常はスライドガラス上に数万種類以上のプローブを配置できるため、理論的にはこのプローブ数とほぼ同数の遺伝子の発現様式を一度に解析できる。

マイクロアレイ実験の基本的な流れは、(1) 研究対象の樹種のマイクロアレイを入手または自作する、(2) 発現データを得たい目的のサンプルからRNAを抽出して逆転写酵素でcDNAに変換して蛍光標識する、(3) 同一のマイクロアレイにハイブリダイゼーションさせて、プローブとハイブリしているcDNAの蛍光シグナル強度を専用のスキャナで検出する、である（図1-5-2）。これによって、アレイに搭載された遺伝子の発現量のデータが画像データとして取得され、専用のソフトでテキストデータに変換する。これをGenespring（Agilent Technologies社）やSubio（Subio社）などのソフトウェアに読み込ませてデータ解析を行う。

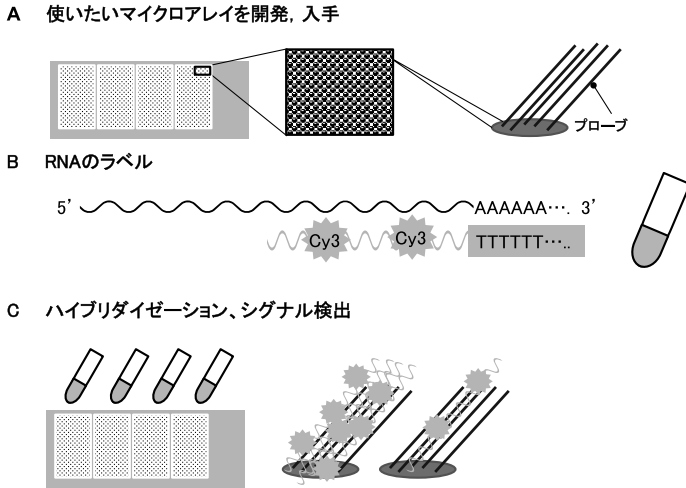


図1-5-2 マイクロアレイ実験の流れ

- A : 研究対象の樹種のマイクロアレイを自作または入手
 B : 発見データを得たい目的のサンプルからRNAを抽出して逆転写酵素でcDNAに変換して蛍光標識
 C : マイクロアレイにハイブリダイゼーションさせて、プローブとハイブリしているcDNAの蛍光シグナル強度を専用のスキャナで検出する。

RNA調整からシグナルの検出までの実験については、メーカーのプロトコルに従って行う。蛍光標識は試薬と一般的なインキュベータおよび遠心機で行えるが、ハイブリダイゼーションとシグナル検出はメーカー指定のハイブリチャンバーと専用スキャナが必要である。これらの装置は高価でメンテナンスも煩雑であるため、著者の研究室ではRNA抽出後、テキストデータの取得までを外注している。各マイクロアレイのメーカーは、外注先としてメーカー公認のサービスプロバイダーを指定している。

(2) カスタムマイクロアレイの開発

カスタムマイクロアレイとは、自作のマイクロアレイの事である。ここでは、Agilent Technologies社のeArray (URL1-5-2) を利用したカスタムアレイの

設計方法について述べる。流れは、(i) マイクロアレイに搭載する遺伝子の配列情報を取得してマルチFASTA形式のファイルを作成する、(ii) プローブデザイン作業として、遺伝子配列から配列特異的な60merの配列を切り出しプローブ候補とし、プローブ配列のTm値などのスコアやクロスハイブリダイゼーションを配慮しながら、プローブを選抜する、(iii) アレイデザインとして選抜したプローブを基板上に配置する、という作業である。(ii)と(iii)の部分をeArrayで行う。ここでは、先述のチュウゴクナシの42,812遺伝子の公開配列情報を用いたマイクロアレイ開発の実際を述べる。

(i) 遺伝子配列情報の取得とマルチFASTAファイルの作成

近年は、各種データベースの充実しており、遺伝子配列の収集には公開データベースの利用が効率的である。

まず、NCBIでデータを収集する方法を述べる。ナシ属(*Pyrus*)の遺伝子情報を収集する場合は、NCBIのAll databaseで「pyrus」と入力して検索することで、「EST」(expressed sequence tag sequences)、「Gene」(collected information about gene loci)などの項目とヒット数が出る(図1-5-3A)。ESTを取得する場合は「EST」を選び、「Send to」のところで「FASTA」を選んでファイルを作成すると(図1-5-3B)マルチFASTA形式のテキストファイルをダウンロードすることができる。この他、Taxonomy Browser (URL1-1-12)を利用すると種レベル、属レベルでの検索が可能である。

NCBI以外にも、作目ごとにデータベースがあり、ナシ属(*Pyrus*)の場合はGenome Database for Rosaceae(GDR, URL1-1-4)が充実している(図1-5-4)。ナシの場合は、このほかにチュウゴクナシ(URL1-1-3)やセイヨウナシ(URL1-5-3)について、解析グループ独自のサーバなどで予測遺伝子配列等を公開している(Wu *et al.* 2013; Chagné *et al.* 2014)。これらのデータベースからダウンロードしたファイルは、多くがgzやtar.gz形式であるため、解凍して用いる。配列数が不足の場合は自分で配列データを取得して追加するか、近縁種のデータを追加する。

以下、eArray (URL1-5-1)でカスタムマイクロアレイを作成する場合に

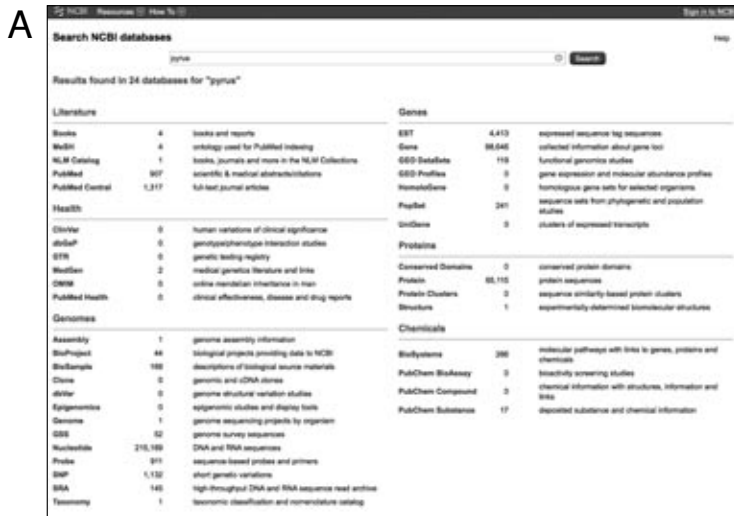


図1-5-3 NCBIでの配列取得

- A : 研究対象の樹種のマイクロアレイを自作または入手
- B : 発現データを得たい目的のサンプルからRNAを抽出して逆転写酵素でcDNAに変換して蛍光標識
- C : マイクロアレイにハイブリダイゼーションさせて、プローブとハイブリしているcDNAの蛍光シグナル強度を専用のスキャナで検出する。



図1-5-4 GDRのトップページ
ナシ属 (*Pyrus*) の場合はGDRのデータベースも利用可能である。

ついて述べる。以下、マルチFASTAファイル (eArrayではターゲットファイルと呼ぶ) 作成時の注意点を述べる。

- ・配列名称は、スペース含めて255文字未満にする。スペースの後の文字はeArrayで出力される結果には出てこない。
- ・シークエンスはセンス方向。
- ・配列内にA, G, T, C以外の文字 (Nなど) を含んではいけない。Nなどを1文字でも含む配列は削除する。または、使用可能な配列が減少しすぎる場合は、Nを境目に配列を切断して、別の配列として扱う。元の配列が同じものが分かるようにしておく。さらに、元の配列内での位置も分かるようにしておく。
- ・マルチFASTA形式のファイルの中に、重複する配列がある場合は予め削除の方が望ましいが、eArrayを用いる場合は、全く同じ配列どうしは配列の重複 (Duplicate) として認識され、どちらか一方からのみプローブが設計されるので、そのままでもよい。一塩基でも異なる配列の場合はDuplicateとは認

識されない。このため、類似した配列を持つ複数の配列から同じ配列のプローブが設計される場合がある。この重複を除くためにデザイン後に「GE probe check」(後述)を行う。また、95%以上相同な配列は同一の遺伝子とみなされ、同じクラスタに分けられる。どの配列がどのクラスタに分けられたのか、はeArrayの結果のファイル(MOST_cluster, 後述)に記載される。同一クラスタ内から同じ配列のプローブが設計された場合は、異なるプローブ名が付いてしまう。この重複もデザイン後の「GE probe check」の処理で回避する。

- ・作成したマルチFASTAファイルのファイル名には日本語、全角文字を入れない。これをZIP化して保存する。保存フォルダにも日本語、全角文字を入れない。

チュウゴクナシの公開配列情報の42,812遺伝子から、Nを含む配列を分断してマルチFASTAファイルを作成した結果、43,753配列が生成され、これをZIP化したものを用いた。

(ii) プローブのデザイン

eArrayを初めて使う場合はアカウント作成が必要であるため、ログインをクリックして「Request for Registration」のサイトで登録する。

登録後、eArrayにログインし、「Probe」タブから「GE Probe Design」に入る。以下、各種設定について順に述べる(図1-5-5A)。

- ・まず、「Base Composition Methodology」、「Tm Matching Methodology」のいずれかを選択する。「Base Composition Methodology」は経験則に基づいて最適化した塩基組成を基準にプローブが選ばれる方法で、真核生物用プローブ設計のスタンダードである。「Tm Matching Methodology」はプローブのTm値を自分で指定するもので、原核生物のプローブを設計するのに適した方法である。ナシ属は真核生物なので、「Base Composition Methodology」を選択した。

- ・「Probe Length」はデザインするプローブの長さの設定である。25から60塩基のプローブを設計できるが、推奨されている60塩基とした。

- ・「Probes per Target」は、個々の遺伝子配列あたりデザインするプローブ数の設定で、1から10を指定できる。後で取捨選択できるので、最大の10を選択

Methodology」が有る。「Best Probe Methodology」は、一つ一つのプローブの塩基構成の善し悪しを優先して選択する方法で、「Best Distribution Methodology」は配列内での位置も考慮して選択する方法である。今回は、「Best Probe Methodology」を選択した。

・「Design with 3' Bias」は、個々の配列の3'の1,000塩基にハイブリするようなプローブを設計するときにチェックを入れる項目である。通常のプロトコルで実験する場合は必ずチェックを入れる。ただし、チェックを入れても1,000塩基よりも5'側に設計されるプローブもある。

・「Target File Details」で、元になる配列ファイルを指定する。「Species」は該当生物種が無い場合は「NA」にする。「Fasta File」を選択してZIP化したマルチFASTAファイルを指定する。ここでは、「Target File Details」を「NA」とし、(i)で作成した43,753配列のマルチFASTAファイルを選択した。

・「Transcriptome Details」には、「Use Target File as Transcriptome」、「Select Agilent-provided Transcriptome」、「Upload Transcriptome File」がある。「Use Target File as Transcriptome」はターゲットファイル内の配列だけで特異性を確認する方法で、ターゲットファイル内にほぼ全ての転写物の情報が有る場合に有効である。「Select Agilent-provided Transcriptome」は、eArrayが用意した特異性確認用データセットで、当該生物種があれば推奨される。「Upload Transcriptome File」は、自分で独自の特異性確認用データセットが有る場合に使う。「Use Target File as Transcriptome」として、(i)で作成した43,753配列のマルチFASTAファイルを指定した。

ここまでの設定が終了したら、「Submit」をクリックする。設計が終了するとメールが届く。数日かかることもある。今回のデザインでは約3時間であったが、eArrayの混み具合などにも依る。この間、ログアウトすることも可能で、「Status」で状況を確認できる。デザイン完了のメールが来たら、再びeArrayにログインし、「Probe」タブから「GE Probe Design」をクリックして該当の「Job Name」の「Actions」から「Download」をクリックするとデザインをダウンロードできる。MOST_sum, MOST_tdt, MOST_fate, MOST_cluster

の4つのテキストファイルが得られる。

MOST_sum (図1-5-5B)：全プローブの長さやTm値などの集計値のファイルである。1-4行目は43,753配列の内41,461配列から一つ以上のプローブがデザインでき、合計414,468個のプローブができたことが分かる。また、14行目から18行目はプローブ配列のスコア毎のプローブ数があり「BC_1」で示される、塩基組成の観点から質が良いとされるプローブが390,719デザインできたことを示す。19行目の「numXHyb」とはそのプローブが元の配列以外の転写産物とハイブリするかどうかの予測で、245,522プローブはクロスハイブリダイゼーションの可能性を示す。詳細はMOST_tdtファイルに記述されている。これらのプローブは他の転写産物とハイブリする恐れがあるため、使わない方が望ましい。

MOST_fate (図1-5-5C)：プローブ設計の元となったターゲット配列の詳細情報のファイルである。ターゲット配列毎に「Status」、「Target length」、「Probes Generated」が記載されている。「Status」を見ると、一つ以上のプローブが設計された配列には「Pass」と表示されている。「Pass」以外はプローブがデザインされなかったもので、理由が記載されている。チュウゴクナシのデータではプローブがデザインのできなかった理由として、「Too short」(60塩基未満)=165個、「Repeat Masked Out」など=22個、「Duplicate sequence」=2,105個となっていた。

MOST_cluster：配列の類似性(95%以上)に基づいてクラスタリング(clustering)した結果を示す。今回は41,483(「Too short」、「Duplicate sequence」を除いたもの)が37,908個のクラスタ(cluster)に分かれていた。

MOST_tdt (図1-5-5D)：デザインされた全プローブの詳細が記載されている。「BPStart」はターゲット配列中のプローブスタート位置で5'側からの方向で記載されている、「End Dist」がターゲット配列の3'末端からのプローブのスタート位置で、「Design with 3' Bias」のチェックを入れたが、414,468プローブ中2,831プローブについては3'末端から1,001塩基以上離れた位置にプローブが設計されていた。「X-HybTarget」のカラムはクロスハイブリの可能性

のある他の遺伝子が有る場合は「X-HybTarget」のカラムにそのIDが記載されている。ただし、95%以上の類似度の配列同士については同じ遺伝子由来と判断されるため、クロスハイブリチェックが行われず、「X-HybTarget」には記載されない。別途「GE Probe Check」でクロスハイブリチェックを行うことにより示される。

次に「GE Probe Check」によってクロスハイブリチェックを行う。その準備として、全414,468プローブ配列のマルチFASTAファイルを作成し、ZIP化した。eArrayに再びログインし、「Application TypeをExpression」にして、「Probe」タグから「GE Probe Design」を選択する。

・「Probe DetailのJob Name」には任意の名前を入力する。「species」は「Na」とし、「Upload Probe File」にチェックをいれて全414,468プローブ配列のマルチFASTAファイルのZIP化したものを指定した。

・「Transcriptome Details」（クロスハイブリチェックをするためのトランスクリプト）については「Upload Transcriptome File」で最初のZIP化した43,753配列のマルチFASTAファイルを選択した。

結果をダウンロードするとMOST_sum, MPST_tdtの2つのテキストファイルが得られる。

MOST_sum (図1-5-5E)：全てのプローブの長さなどの集計値の情報である。20行目は、他のプローブとクロスハイブリする可能性があるプローブ数で、今回は278,175であった。

MOST_tdt (図1-5-5F)：全プローブごとの詳細リストである。「Predicted Target」は、プローブがハイブリダイズすると予測されるターゲットである。元の配列と同じになるのが通常だが、違う物もある。違うものはプローブとしてふさわしくない。チュウゴクナシ公開配列からのプローブデザインでは、同じものが362,311個、違う物が52,157個であった。「X-HybPotential」は、「Predicted Target」以外のトランスクリプトがクロスハイブリする可能性が高いと予測される場合は1、予測されない場合は0と記載される。チュウゴクナシ公開配列からのプローブデザインでは、0が136,293個、1が278,175

個であった。このファイルを用いて、プローブの取捨選択を行うことができる。チュウゴクナシ公開配列からのプローブデザインでは、 $X-Hyb=0$ 、 $BC=1$ 、元の配列のなかで3'に近いこと、PolyXの少ないこと、の順に絞り込んで、供試した元の42,812配列から、それぞれ一つのプローブを選抜し、40,700配列について一つずつのプローブが設定された。

次いで、デザインした配列をマイクロアレイに配置するために「Probe Group」というリストを作成する。「Probe Group」とは一つ以上のプローブで構成されるまとまりで、マイクロアレイのパーツとして利用できる。マイクロアレイにプローブを配置する際に、「Probe Group」ごとに繰り返し搭載数を設定するので、異なる繰り返し数で搭載する場合はProbe Groupを分けておく。最もシンプルなのは、図1-5-5Gのようにプローブ名と60塩基の配列をリストにした「Minimal Format」で、先のMOST_tdtファイルから作成する。eArrayにログイン後、「Probe」タブの「Upload」の画面でアップロードする。

また、Agilent Technologies社のマイクロアレイでは実験後の数値化を適切に行うことを目的として、少なくとも15種類の任意の配列のプローブを5から10回繰り返して搭載することが推奨されている。解析したい生物種から15個のプローブを準備して「Probe Group」を作り同様にアップロードする。今回設計されたプローブから選ぶことも可能である。これは、後述の「Replicate Probe Group」として使用する。

(iii) マイクロアレイへのプローブの配置

マイクロアレイへの配置であるが、近縁種のプローブ、過去に使用したマイクロアレイのプローブなど、他の機会にデザインした「Probe Group」を新たなマイクロアレイに再び搭載することも可能である。またそれぞれの「Probe Group」は複数回繰り返して搭載することも可能である。繰り返すことで、マイクロアレイ実験の結果を解析する際に、同じプローブの蛍光強度の平均値等を算出できる。eArrayの「Home」タブの「Design Wizards」で「Create a Microarray Design from Existing Probe Groups」を選ぶ。次いで「Define Design」で「Microarray Name」を入力後、「Design Format」でフォーマット

トを選ぶ。チュウゴクナシゲノム公開配列からのマイクロアレイ作成では、4x180Kフォーマットとした。4x180Kでは、一枚のスライドで180,880個のプローブを搭載でき、4サンプル分の実験ができる。ただ、180,880個のプローブのうち4,974個はAgilentが独自に搭載するものであり、先述の「Replicate Probe Group」の15プローブを10回繰り返し搭載すると、実際に搭載できるプローブの数は175,756個となる。なお、Agilentが独自に搭載する4,974個のプローブは、マイクロアレイ実験のポジティブコントロール、バックグラウンドシグナルの算出、デザインファイルとのグリッド合わせ、スキャン画像を数値化する際に必要なコントロールとしての目的がある。この数を勘案して、「Biological Expression Probe Groups」としてプローブグループを選択する。これを送信すればアレイデザインが完了し、独自のアレイデザイン番号が割り振られる。この番号はマイクロアレイを購入するときに使用する。アレイデザインの内容はeArrayから11個のファイルとしてダウンロードすることができる。この11ファイルには、ハイブリ実験後に画像をスキャナでの取りこんだ際に、蛍光強度を数値化するのに必要なファイル等が含まれている。スキャナにより必要なファイルは異なるが、アジレント社のFeatureExtractionであれば「ExternalFullGEML」を、Rosetta社のソフトであれば「ExternalFullGEML2」、Imagene社のソフトの場合は「Imagene」のファイルを用いる。マイクロアレイ実験を外注した場合にも外注先からこれらのファイルの提出を求められる。

4) ナシのカスタムマイクロアレイを利用した研究論文

(1) ニホンナシのカスタムマイクロアレイの開発

著者は2007年にニホンナシのマイクロアレイを開発した (Nishitani *et al.* 2010)。当時はナシの遺伝子配列のGenbankでの公開データは862 (NCBI) であったため、ニホンナシ遺伝子の配列情報は著者の研究室で収集したものと、共同研究者からの提供を受けて11,540配列を収集した。

11,540配列のうちから、11,197配列で少なくとも1つ以上のプローブをデザインできた。クロスハイブリチェックなどを経て結果的に9,812の非重複の配

列を用いてプローブを設計した。このうち、9,060配列については複数のプローブが設計できたので5'側と3'側の2カ所のプローブを選抜し、752配列については1カ所のみでプローブが設計された。合計で、18,872プローブを搭載した。

(2) ニホンナシのカスタムマイクロアレイを用いた研究論文

ニホンナシ‘豊水’果実の発達過程における遺伝子発現解析を行った研究論文 (Nishitani *et al.* 2010) を紹介する。サンプルは、満開後105日, 119日, 133日, 140日, 147日 (成熟時) の5ステージの果肉で、生物学的な反復を2から3反復とり、合計14サンプルの解析を行った。データはGEOに登録した。遺伝子発現データ番号は5種類ステージに対応したGSM463985 からGSM463989である。解析にはGeneSpringを用いた。パラメータは初期設定のままとした。

次に、データの信頼性の確認を行った。本マイクロアレイでは9,060遺伝子については2カ所にプローブを設計したので、この9,060組のプローブについて、同様の発現様式であるかどうかを相関係数を算出することで概算した。その結果7,509組のプローブで正の相関を示し、任意の8遺伝子についてRT-PCRを用いて確認したところ、全ての遺伝子でマイクロアレイと同様の発現様式を示した。また、2つのプローブの発現様式に相関が無い又は負のものについては、RT-PCRで確認した。その結果3'側に設計したプローブの発現様式がRT-PCRの発現様式に一致することが確認された。そのため、2つのプローブを設計したものについては、以後の解析は3'側プローブを用いた。これらの結果から、本カスタムマイクロアレイは信頼性の高い結果を得られることが示された。遺伝子発現解析の結果、以下の様なことが分かった。

(a) 成熟初期、成熟期での発現量の違いの大きい遺伝子としては、既知の細胞壁関連遺伝子群の他に、転写因子やGlutathione S-transferaseと機能予測される新規遺伝子群が見られた。

(b) 成熟初期と成熟時で発現量の大きく異なる遺伝子群を抽出したところ、3,062遺伝子が抽出された。QTクラスタリングにより3,062遺伝子を発現様式ごとにクラスタリングした結果、11パターンに分かれた。このパターンの

中に含まれる遺伝子群の機能分類をMunich Information Center for Protein Sequences (MIPS) のFunCat-1.3 annotation (Schoof *et al.* 2004) に基づいて行うと、成熟に向かって発現量の上がる遺伝子群にはタンパク質の修飾や輸送に関与する遺伝子群が多く、逆に発現量の下がる遺伝子群にはタンパク質合成に関連する遺伝子群が多かった。このことから、ナシ果実の成熟は新たなタンパク質の合成よりは既存のタンパク質の修飾などで進行することが示唆された (Nishitani *et al.* 2010)。

カスタムマイクロアレイは、バージョンアップすることも可能である。著者の研究室でも、ニホンナシマイクロアレイについてホルモン処理果実などの新しいサンプルから得た発現遺伝子配列等を追加して何度かバージョンアップしている。バージョンアップしたアレイは、ナシの単為結果性に係わる遺伝子群の解析 (Nishitani *et al.* 2012a)、ナシの休眠に係わる遺伝子群の解析 (Nishitani *et al.* 2012b)、セイヨウナシの果実成熟過程の解析 (Nashima *et al.* 2013) などの研究で活用されている。

5) 遺伝子の網羅的解析研究の今後

網羅的遺伝子発現解析は、着目する現象に関連する遺伝子群を絞り込むための初期スクリーニングの手法である。我が国の主要樹種である、リンゴ、ブドウ、モモ、カンキツについてもマイクロアレイや次世代シーケンサを用いた研究が行われている。それぞれの樹種について、着目される現象における遺伝子発現様式の解析が進展するばかりで無く、これらの情報を統合することで果実成熟に普遍的な現象の抽出、または特定の樹種に特徴的な現象の解析も進むと考えられる。

引用文献

- Alexander, L. and Grierson, D. (2002) Ethylene biosynthesis and action in tomato : a model for climacteric fruit ripening. *Journal of Experimental Botany*. 53 : 2039-2055.
- Chagné, D. *et al.* (2014) The Draft Genome Sequence of European Pear (*Pyrus communis* L. 'Bartlett') . *PLoS ONE*. 9 : e92644.
- Itai, A. *et al.* (1999) Identification of 1-aminocyclopropane-1-carboxylic acid synthase genes controlling the ethylene level of ripening fruit in Japanese pear (*Pyrus pyrifolia* Nakai). *Molecular and General Genetics*. 261 : 42-49.
- Itai, A. *et al.* (2003a) Cloning and characterization of a cDNA encoding 1-aminocyclopropane-1-carboxylate (ACC) synthase (*PPACS3*) from ripening fruit of Japanese pear (*Pyrus pyrifolia* Nakai). *Journal of the Japanese Society for Horticultural Science*. 72 : 99-106.
- Itai, A. *et al.* (2003b) Rapid identification of 1-aminocyclopropane-1-carboxylate (ACC) synthase genotypes in cultivars of Japanese pear (*Pyrus pyrifolia* Nakai) using CAPS markers. *Theoretical and Applied Genetics*. 106 : 1266-1272.
- Jaillon, O. *et al.* (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*. 449 : 463-467.
- Ming, R. *et al.* (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus) . *Nature*. 452 : 991-996.
- Nishitani, C. *et al.* (2010) Oligoarray analysis of gene expression in ripening Japanese pear fruit. *Scientia Horticulturae*. 124 : 195-203.
- Nishitani, C. *et al.* (2012a) Parthenocarpic genetic resources and gene expression related to parthenocarpy among four species in pear (*Pyrus* spp.). *Scientia Horticulturae*. 136 : 101-109.
- Nishitani, C. *et al.* (2012b) Transcriptome analysis of *Pyrus pyrifolia* leaf buds during transition from endodormancy to ecodormancy. *Scientia Horticulturae*. 147 : 49-55.
- Nashima, K. *et al.* (2013) Microarray analysis of gene expression patterns during fruit development in European pear (*Pyrus communis*). *Scientia Horticulturae*. 164 : 466-473.
- Schena, M. *et al.* (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 270 : 467-470.
- Schoof, H. *et al.* (2004) MIPS *Arabidopsis thaliana* Database (MAtdB) : an

- integrated biological knowledge resource for plant genomics. *Nucleic Acids Research*. 32 : D373-376.
- Shulaev, V. *et al.* (2011) The genome of woodland strawberry (*Fragaria vesca*). *Nature Genetics*. 43 : 109-118.
- Tatsuki, M. and Mori, H. (1999) Rapid and transient expression of 1-aminocyclopropane-1-carboxylate synthase isogenes by touch and wound stimuli in tomato. *Plant and Cell Physiology*. 40 : 709-715.
- The International Peach Genome Initiative (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics*. 45 : 487-496.
- Velasco, R. *et al.* (2010) The genome of the domesticated apple (*Malus × domestica* Borkh) . *Nature Genetics*. 42 : 833-841.
- Wu, J. *et al.* (2012) The genome of pear (*Pyrus bretschneideri* Rehd.). *Genome Research* 23 : 396-408.
- Yamagami, T. *et al.* (2003) Biochemical diversity among the 1-amino-cyclopropane-1-carboxylate synthase isozymes encoded by the Arabidopsis gene family. *Journal of Biological Chemistry*. 278 : 49102-49112.

