

Summary of Doctoral Thesis

Studies on development and application of high-throughput genomic and bioinformatics tools for citrus fruit physiology and breeding

Hiroshi FUJII*

Citrus Research Division, Institute of Fruit Tree Science,
National Agriculture and Food Research Organization (NARO)
Okitsuakacho, Shimizu, Shizuoka 424-0292, Japan

Summary

Citrus is one of the most economically important fruit species in the world. The fruit is rich in the second metabolites for beneficial for human health, with diverse colors, fragrances and tastes. In addition, citrus is among the most difficult plants to improve through cross breeding approaches and to analyze physiologically varietal characteristics because of its the polyembryony, male sterility or self-incompatibility. Genome science technologies have advanced rapidly over the last decade, and have been adapted to address the challenges of the citrus breeding and physiological analysis. Expressed sequence tag (EST) analysis, DNA markers, linkage mapping and quantitative trait locus (QTL) analysis have promoted the efficient selection systems and analysis of gene expression. Remarkably, three genome assemblies have been released to the public since 2011. Despite the challenges of working with citrus to understand the important characters of citrus, the result is insufficient. Expression analysis of many genes related to important characters and analysis of genome-wide genotyping among many varieties or the combination of these two analyses is necessary to understand important characters of citrus. This study was performed to provide the basis for comprehensive use of citrus genome information, which has been accumulated quickly.

1. 22K citrus oligoarray analysis of gene expression in mature mandarin fruit

1) Profiling ethylene-responsive genes in mature mandarin fruit using a citrus 22K oligoarray

A comprehensive transcriptome analysis using a citrus 22K oligo-microarray was performed to identify ethylene-responsive genes and gain an understanding of the transcriptional regulation by ethylene in Satsuma mandarin fruit (*Citrus unshiu* Marc.). Seventy-two hours after ethylene treatment of mature fruit, 1,493 genes were identified as ethylene-responsive, with more than 3-fold expression change. Interestingly, more than half of the

ethylene-responsive genes were repressed after ethylene treatment. This might suggest that ethylene inhibits various biological processes and plays an important role in fruit ripening and senescence. Ethylene repressed the transcription of many genes involved in photosynthesis, chloroplast biogenesis, and sugar metabolism, while it induced the transcription of several genes related to resistance, defense, stress, amino acid synthesis, protein degradation, and secondary metabolism.

2) Profiling gibberellin (GA₃)-responsive genes in mature fruit using a citrus 22K oligoarray

Gibberellin3 (GA₃)-responsive genes were investigated

* Corresponding author. E-mail: hfujii@affrc.go.jp

with a citrus 22K oligo-microarray to gain further the understanding of the transcriptional regulation by GA₃ treatment in Satsuma mandarin fruit. 213 GA₃-responsive genes were identified that showed a 3-fold or greater expression change after 72h of GA₃ treatment of mature fruit, compared with expression after 72 h of air treatment. GA₃ treatment induced the expression of pathogenesis-related (PR) proteins and genes that function in photosynthesis, chloroplast biogenesis, resistance, defense and stress. Also, GA₃ treatment reduced the transcription of several ethylene-inducible genes, such as carotenoid metabolic genes, which are associated with fruit ripening. Contrasting effects between GA₃ and ethylene were observed. The endogenous GA₃ level might be important for the endogenous regulation of maturation and senescence in mature citrus fruit.

2. An algorithm and computer program for the identification of minimal sets of discriminating DNA markers for efficient cultivar identification

DNA markers are frequently used to analyze crop varieties, with the coded marker data summarized in a computer-generated table. Such summary tables often provide extraneous data about individual crop genotypes, needlessly complicating and prolonging DNA-based differentiation between crop varieties. At present, it is difficult to identify minimal marker sets—the smallest sets that can distinguish between all crop varieties listed in a large marker–summary table—because of the absence of algorithms capable of such characterization. Here, we describe the development of just such an algorithm based on combinatorial optimization and the computer program,

named MinimalMarker. MinimalMarker is available for use with both dominant and co-dominant markers regardless of the number of alleles, including simple sequence repeat (SSR) markers with numeric notation.

3. High-throughput genotyping in citrus accessions using an SNP genotyping array

A 384 multiplexed single nucleotide polymorphism (SNP) genotyping array, named *CitSGA-1*, for the genotyping of citrus cultivars, was developed, and the performance and reliability of the genotyping were evaluated. SNPs were detected in cultivars representing the genetic diversity of citrus currently bred in Japan. The assay using *CitSGA-1* was applied to a hybrid population of 88 progeny and 103 citrus accessions for breeding in Japan. A total of 351 SNPs could call different genotypes among the DNA samples. To confirm the reliability of SNP genotype calls, parentage analysis was used, which indicated that the numbers of reliable SNPs were 276. Using 7 SNPs that were identified by MinimalMarker, all germplasm accession could be discriminated from each other. The SNP genotyping array reported here will be useful for the efficiently constructing linkage maps, for the detection of markers for marker-assisted breeding and for the identifying of cultivars.

By developing two genomic tools in this study, comprehensive and high-throughput gene expression analysis and genotyping have become accessible. In addition, the bioinformatics tool has developed to use the genomic tools thoughtfully. These tools are available as the research base for detection of regulatory genes which control trait, linkage mapping and maker aided selection.

カンキツ果実の生理学と育種学のためのハイスループットなゲノム及び バイオインフォマティクスツールの開発と応用に関する研究

藤井 浩

国立研究開発法人 農業・食品産業技術総合研究機構
果樹研究所カンキツ研究領域
424-0292 静岡県静岡市

摘 要

カンキツは世界的に経済上、最も重要な果樹の1つであり、果実の色や香り、味、健康機能性を持つ二次代謝成分などの多様性に基づいた新品種の育成が進められている。しかし、カンキツには多胚性や雄性不稔性、開花までの年数が長いといった性質があり、交雑育種や品種特性の生理学的な解析を困難にしている。近年、急速に進展しているゲノム科学の成果を利用して、カンキツの育種や生理学的な解明が進められている。例えば、発現遺伝子配列断片 (EST) 解析や DNA マーカー、連鎖地図、量的形質遺伝子座 (QTL) 解析等が育種の効率化や果実における遺伝子発現の解明に利用されてきている。また、2011年以降に3つのカンキツゲノムの全塩基配列解析結果が公開された。こうした研究にもかかわらず、カンキツの重要形質に関する遺伝子レベルの解明は十分ではない。重要形質の解明のためには、形質に関わる多数の遺伝子の発現解析や多数の品種についてのゲノムワイドな遺伝子型情報解析及び、それらを組み合わせた解析が不可欠である。本研究は、急速に蓄積が進んでいるカンキツゲノム情報を総合的に利用するための基盤を構築するために、一連の研究を行ったもので、以下の課題により構成される。

1 オリゴアレイによるカンキツ成熟果実における遺伝子発現解析

1) オリゴアレイによるカンキツ成熟果実におけるエチレン応答遺伝子のプロファイリング

ウンシュウミカン (*Citrus unshiu* Marc.) の成熟果実におけるエチレン応答遺伝子の特定とその転写制御の解明のために、カンキツの22K オリゴマイクロアレイ (オリゴアレイ) を作成し、遺伝子発現解析を行った。エチレン処理72時間後に3倍以上の発現変化があった1493個の遺伝子をエチレン応答遺伝子と特定した。エチレン応答遺伝子の半分以上はエチレン処理により発現が抑制されていた。このことは、エチレンが多くの生物学的なプロセスを低下させ、果実の成熟と老化に関して重要な役割を果たしていることを示していた。エチレンは光合成や葉緑体生合成、糖代謝等の遺伝子の転写を抑制した一方、病害抵抗性や防御、ストレス、アミノ酸合成、タンパク質分解、2次代謝に関連する遺伝子の転写を誘導した。

2) オリゴアレイによるカンキツ成熟果実におけるジベレリン応答遺伝子のプロファイリング

ウンシュウミカンの成熟果実でのジベレリン3 (GA_3) による転写制御を知るために、オリゴアレイを用いて GA_3 応答遺伝子を調査した。 GA_3 処理72時間後のウンシュウミカン成熟果実と対照区との遺伝子発現を比較したところ、3倍以上の発現変化を示す213個の GA_3 応答遺伝子が特定された。 GA_3 処理は、生体防御タンパク質や光合成、葉緑体生合成、病害抵抗性、防御、ストレスに関連する遺伝子の発現を誘導した。また、果実の成熟に関連するカロテノイド代謝遺伝子等のエチレンに誘導される遺伝子の転写を低下させた。その効果はエチレンと対照的で、内性の GA_3 のレベルが果実の成熟と老化の制御に重要であると推測された。

2 DNA マーカーによる効率的な品種識別のためのソフトウェアの開発

DNA マーカーを様々な品種に適用した結果は、DNA

マーカー型を記号に置き換えた表に整理され、品種識別に利用される。膨大な2次元表から最も少ない数で全ての品種を識別することができる最少マーカーセットを求めることは簡単ではない。そこで、最少マーカーセットを求めるための組み合わせ最適化理論に基づくアルゴリズムとその計算を実現するソフトウェア **MinimalMarker** を開発した。MinimalMarker は共優性マーカーにも優性マーカーにも適用可能であり、アレルの数を問わない。また、単純反復配列 (SSR) マーカーのような数値表記のマーカーにも適用できる。

3 SNP アレイを用いたカンキツ品種のハイスループットなジェノタイピング

カンキツのハイスループットなジェノタイピングを実現するために、*CitSGA-1* と名付けた384個の一塩基多型 (SNP) を搭載する SNP ジェノタイピングアレイ (SNP アレイ) を開発し、その性能と信頼性を評価した。搭載

した SNP はわが国のカンキツ育種の遺伝的な多様性を代表する品種から収集した。*CitSGA-1* を88個体の交雑集団とわが国のカンキツ育種に関連する103種類の品種・系統に適用した結果、351個の SNP で遺伝子型を識別できた。SNP アレイ解析による遺伝子型の信頼性を確認するために親子分析を行った結果、276個の SNP が信頼できると判定された。これらの SNP に MinimalMarker を適用したところ、7つの SNP を用いることにより、すべての品種・系統を識別できることがわかった。

これらの研究を通じて開発したオリゴアレイと SNP アレイの2つのゲノム解析ツールにより、網羅的かつハイスループットな遺伝子発現解析と品種ジェノタイピングを可能になり、これらの情報を使いこなすバイオインフォマティクスツールの開発とあわせて、形質制御遺伝子の探索、連鎖地図の効率的な作成や、マーカー支援育種などの研究基盤としての利用されるものと期待される。