

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 2 日現在

機関番号：82111

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23580055

研究課題名(和文) アレル解析に基づく育種効果の遺伝統計学的検証基盤の開発

研究課題名(英文) Development of genetic statistics verification base in effect of breeding based on a llele analysis

研究代表者

藤井 浩 (FUJII, Hiroshi)

独立行政法人農業・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・主任研究員

研究者番号：00355398

交付決定額(研究期間全体)：(直接経費) 4,400,000円、(間接経費) 1,320,000円

研究成果の概要(和文)：カロテノイド類には健康機能性があり、カンキツ果実での高含有化は重要な育種目標である。カンキツの品種ごとの含有量は多様である一方、主要品種はわずかに14の祖先品種から派生しており、そのアレル数は28に限定される。本研究は、8つのカロテノイド代謝酵素遺伝子とそれを制御する3つの転写因子候補遺伝子について、28のアレルタイプを決定し、高含有化に関連するアレル組合せを明らかにすることを目的とした。

その結果、祖先14品種における、カロテノイド含有量に関わる11遺伝子のゲノム上の遺伝子領域について、塩基配列を決定するとともに、アレル間SNP解析によりアレルタイプを決定することができた。

研究成果の概要(英文)：Carotenoids have healthy functionality. The high content of carotenoids in citrus fruit is an important breeding objective. The content of the carotenoid is different in each cultivar. However the dominant cultivars in Japan originates in the ancestral cultivar of 14. Therefore, the number of allele is limited to 28. The purpose of this study is that the allele type of 28 is shown in eight carotenoid metabolic enzyme genes and three transcription factor candidate genes that control them, and the allele combination related to the high content of carotenoids is clarified.

Consequently, the sequences of 11 genes related to the content of the carotenoid was decided in the ancestral cultivar of 14. The allele type was decided by the SNP analysis.

研究分野：園芸学・造園学

科研費の分科・細目：基盤研究(C)

キーワード：アレル カンキツ カロテノイド 遺伝統計学 SNP 育種 遺伝子 ゲノム

1. 研究開始当初の背景

カンキツには健康機能性を有するさまざまな化合物が知られており、そうした化合物の1つであるカロテノイド(第1図)についても研究が進められてきた。カロテノイドの中でも、 β -Cryptoxanthin (B-Cry)にガン抑制効果があることが証明されたため、カンキツ果実でのB-Cry高含有化は重要な育種目標の1つとなっている。カンキツ果実での各カロテノイド含有量は品種ごとに多様であること、カロテノイド含有量には多くの遺伝的要因が関与していること、カロテノイド代謝酵素遺伝子の発現様式がカロテノイド含有量に関与することがこれまでの研究で知られている。

一方、今井ら(未発表)は、現在栽培されている、あるいは、近年育種されたカンキツの主要品種・系統は、わずかに14祖先品種から派生していることを明らかにした。このことから、各カロテノイド含有量の多様性を支える遺伝的要因は、14祖先品種という遺伝子プールに由来することになる。したがって、各カロテノイド代謝酵素遺伝子(図参照)も含めて、各遺伝子座のアリル数は28に限定される。

申請者らは、カロテノイドのうち特にB-Cry高含有化に関わる分子マーカーを開発するため、カンキツ8品種のカロテノイド生合成経路遺伝子及び転写因子遺伝子ホモログのゲノム領域のシーケンスを行い、一塩基多型(SNP)の検出を行った。

これを基に果樹類では初めてSNPジェノタイプングアレイを開発した(未発表)。本アレイを1枚使用することにより、384SNPについて同時に96サンプルの遺伝子型が決定できるため、ハイスループットな遺伝子型決定が可能となった(未発表)。

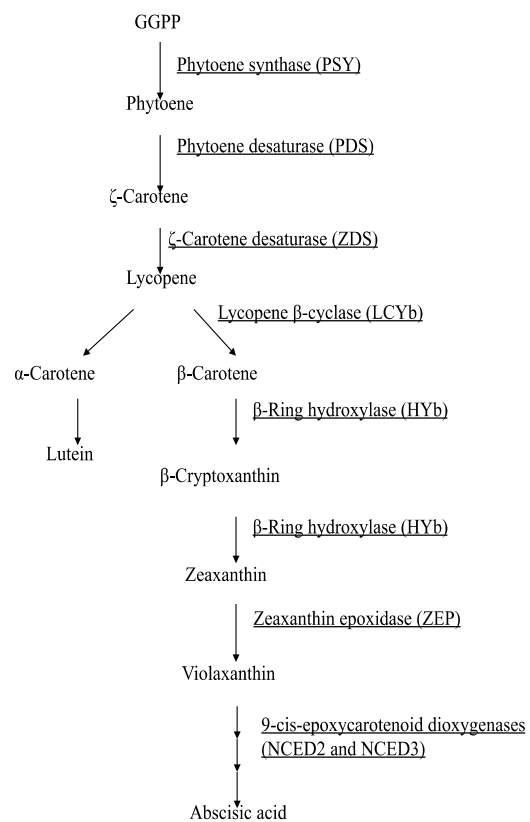
本アレイを1つのカンキツの交雑集団に適用し、申請者が属するグループこれまでに蓄積してきたCleaved Amplified Polymorphic Sequence(CAPS)マーカーの結果(未発表)と合わせて連鎖地図を作成した。この連鎖地図を基にカロテノイド含有量を量的形質とする量的形質遺伝子座(QTL)解析(Sugiyama et al. in press)とカロテノイド生合成遺伝子の発現量を量的形質とするExpression QTL(eQTL)解析を行い以下の結果を得た。

(1) QTL解析では、総カロテノイド含有量に関する13のQTLとそれらと重複する各カロテノイド含有量に関する7つのQTLが検出されたが、これらの領域にカロテノイド代謝酵素遺伝子に由来するマーカーはなく、1つは転写因子ホモログに由来していた(Sugiyama et al. in press)。

(2) eQTL解析では、*Pytoen synthase (PSY)*、 *β -Ring hydroxylase (HYb)*、*Zeaxanthin*

epoxidase (ZEP)(第1図)に関して、遺伝子のアリル構造がシス因子として発現レベルに影響を与えている可能性が示唆された。*Phytoen desaturase (PDS)*と β -*Caroten desaturase (ZDS)*(第1図)では、2つの転写因子由来マーカーが検出され、トランスな要因と考えられた(未発表)。

これらのことから、カロテノイド含有量の多様性は代謝酵素遺伝子のアリル構造(シス要因)や転写因子による制御関係の違い(トランス要因)によって生じていると考えられた(未発表)。しかし、上記の解析は、1つの交雑集団から得られた結果であり、主要カンキツ品種・系統のカロテノイド含有量の制御機構を解明するためには、すべてのアリルタイプを明らかにすることが必要になった。



第1図 植物におけるカロテノイド代謝経路

2. 研究の目的

(1)カロテノイド類には骨粗しょう症予防作用等の健康機能があり、カンキツ果実での高含有化は重要な育種目標である。品種ごとのカロテノイド含有量は多様である一方、主要品種及び育種系統の多くは、わずかに14の祖先品種から派生しており、各遺伝子座のアリル数は28に限定される。

(2) 本研究の目的は、各品種のカロテノイド代謝酵素遺伝子及び制御関係にあると予測

される転写因子遺伝子群について、28に限定された中からアリルタイプを決定し、そのアリル組み合わせとカロテノイド含有量との関係から、カロテノイド高含有化に関連するアリル組み合わせ候補を明らかにすることである。

(3) 本研究では、カンキツ主要品種・系統が14祖先品種由来であるということを利用して、7種類のカロテノイド代謝酵素遺伝子、*PSY*、*PDS*、*ZDS*、*Lycopen -cyclase (LCYb)*、*HYb*、*ZEP*、*9-cis epoxy carotenoid dioxygenases (NCED2,3)* (第1図)及び遺伝統計学的なQTLおよびeQTL解析から関連が予想されている2種類の転写因子候補遺伝子 (*BFC*、*TF271*) に関して、28アリルタイプを明らかにし、カロテノイド含有量とアリルタイプの組合せとの関係から、カロテノイド高含有に關与するアリルタイプを検討することを目指す。

3. 研究の方法

(1) SureSelect 法によるゲノム濃縮用プローブの設計とゲノム濃縮実験および濃縮されたゲノム領域の高速シーケンサーによる解読

本研究で対象とする7種類のカロテノイド代謝酵素遺伝子(第1図)および2種類のカロテノイド代謝酵素遺伝子を制御する転写因子候補遺伝子、および、その他の転写因子を含む、364個の遺伝子領域について、2011年に国際カンキツゲノムコンソーシアムから公開されるクレメンティン半数体の全ゲノム塩基配列 (<http://www.citrusgenomedb.org/>) から、取得した。7種類のカロテノイド代謝酵素遺伝子および2種類の転写因子候補遺伝子については、シス要因の制御機構の存在も予想して、遺伝子領域の上流・下流の2Kbpの配列も取得した。

取得したゲノム領域を SureSelect 法 (Agilent 社) の手順にしたがって、120bpに切断し、配列が完全に重複する断片を排除した後、SureSelect 法のプローブとした。

祖先14品種の葉からゲノムDNAを抽出し、標識をしたプローブとハイブリダイゼーションさせ、プローブとハイブリダイゼーションしたゲノムDNA断片のみを取り出すことによって濃縮した。この操作を品種ごとに行い、品種ごとに異なるタグを接続することによって、品種が識別できるようにした。

濃縮したゲノム断片を次世代シーケンサー-HiSeq2000 (Illumina 社) を用いた解読した。

(2) 次世代シーケンスによって得られたリードのマッピング

解読されたシーケンス(リード)のうち1塩基でもNを含むリードは除去した。タグ等を除去するトリミングを行った後、マッピ

ングツールであるBWA(ver.0.6.2) (<http://bio-bwa.sourceforge.net/>) を用いて、プローブを設計したレファレンス配列にマッピングを行うとともに、コンセンサス配列を作成した。

(4) マッピングされたデータに対する多型解析

BWAによって、レファレンスにマッピングされたデータから、マッピング情報および多型解析ソフトウェアであるSamtools (ver.0.1.18) (<http://samtools.sourceforge.net/>) を用いて、祖先14品種間のSNPおよび挿入・欠失を検出した。

(5) クレメンティン公開ゲノム配列 (version 1.0) における各カロテノイド代謝酵素遺伝子のコピー数

7つのカロテノイド代謝酵素遺伝子とクレメンティンの全ゲノム配列を相同性検索ソフトウェアBLASTNで比較することにより、それぞれの遺伝子のゲノム上のコピーを検出した。

4. 研究成果

(1) SureSelect 法によるゲノム濃縮用プローブの設計

カロテノイド代謝酵素遺伝子である *PSY*、*PDS*、*ZDS*、*LCYb*、*Hyb*、*ZEP*、*NCED* およびカロテノイド転写酵素遺伝子の転写因子候補遺伝子の *BFC*、*TF271* の遺伝子領域のおよびその上流・下流2Kbpをベイト設計の対象としたほか、転写因子を中心に、合計364カンキツ遺伝子領域を120merごとに切断した12472個、総延長約1.5Mbpのベイトを設計した。

(2) ゲノム濃縮および濃縮した領域の次世代シーケンス

濃縮領域の次世代シーケンサーによるシーケンスは、1、2、4の3つのレーンで行った(第1表)。14品種のうち、クレメンティン、ダンシータンジェリン、グレープフルーツ、日向夏、無核紀州、谷川ブタンについては、2つのレーンでシーケンスを行った。最もリード数が少ない地中海マンダリンでも、4,420,845リード(総延長221Mbp)が得られ、90.71%がQ30を超えており、14品種ともに十分なクオリティのリードが得られた(第1表)。

(3) 次世代シーケンスによって得られたリードのアセンブル

次世代シーケンサーで解読したリードをベイトを設計したレファレンス塩基配列(クレメンティン全ゲノム配列 version 0.9等に由来)にマッピングを行った結果、最もマッピングされたリード数が少ない地中海マン

第1表 各品種ごとのシーケンシングの結果

シーケンサー・レーン番号	サンプル名	得られたリードの総延長 (Mbp)	得られたリードの数	それぞれのレーンにおけるリードの割合 (%)	Q0 (PF) 以上のリードの割合	クオリティの平均 (PF)
1	Clm	396	7,914,862	10.37	98.83	39.14
1	DTn	380	7,609,993	9.97	96.95	38.47
1	GrF	657	13,145,949	17.22	96.87	38.42
1	Hgn	378	7,564,269	9.91	96.87	38.41
1	Hsk	524	10,471,311	13.72	96.94	38.46
1	MKs	684	13,671,609	17.91	96.85	38.41
1	TBt	413	8,264,442	10.83	96.87	38.43
2	Clm	182	3,637,361	7.65	98.89	39.15
2	DTn	185	3,704,060	7.8	98.61	39.07
2	GrF	305	6,106,062	12.85	98.82	39.13
2	Hgn	186	3,712,259	7.81	98.62	39.05
2	Hsk	246	4,910,768	10.33	98.83	39.14
2	MKs	339	6,777,081	14.26	98.49	39
2	TBt	196	3,914,838	8.24	98.75	39.11
4	Iyo	397	7,934,159	11.58	90.56	35.87
4	KMd	649	12,986,928	18.95	90.8	35.95
4	MMd	221	4,420,845	6.45	90.71	35.92
4	Mct	371	7,419,953	10.83	90.93	35.99
4	Org	617	12,337,867	18	90.87	35.96
4	Pnk	367	7,345,742	10.72	90.93	35.98
4	Uns	300	5,998,434	8.75	90.86	35.97

Clm: クレメンティン, Dtn: ダンシ-タンジェリン, GrF: グレープフルーツ, Hgn: 日向夏, Hsk: ハッサク, MKs: 無核紀州, TBt: 谷川ブタン: Iyo: 伊予柑, KMd: キングマンダリン, MMd: 地中海マンダリン, Mct: マーコット, Org: オレンジ, Pkn: ポンカン, Uns: ウンシュウミカン

ダリンでも 3,617,540 リードがマップされ、

14 品種ともに十分な depth での SNP 解析が可能な量であった (第2表)。

(4) アセンブルしたデータからのコンセンサス配列の決定

カロテノイド代謝酵素遺伝子 (*PSY*, *PDS*, *ZDS*, *LCYb*, *Hyb*, *ZEP*, *NCED*) および転写因子候補遺伝子 (*BFC*, *TF271*) の遺伝子領域を含む 364 個のカンキツ遺伝子領域について、カンキツ祖先 14 品種のコンセンサス配列を作成した。

(5) リファレンス配列にリードをマッピングで得た品種ごとの多型数

クレメンティン由来のリファレンス配列に各品種のリードをマッピングして、SNP または挿入・欠失の多型を得た (第3表)。日向夏ではクレメンティンのアレル間多型の約 3.4 倍にあたる。37,432 の多型が検出された。このことから、日向夏は他の祖先品種とは異なる遺伝的バックグラウンドを持っていると考えられた。

最も多型数の少ない地中海マンダリンで

第2表 シーケンサーによって得られたリードのレファレンス配列へのマッピング結果

シーケンサー・レーン番号	サンプル名	マッピングされた数	マッピングされなかった数	マッピングされた割合 (%)
1	Clm	5,858,115	833,693	87.54
1	DTn	5,717,017	719,589	88.82
1	GrF	9,377,493	1,766,077	84.15
1	Hgn	5,451,528	951,508	85.14
1	Hsk	7,644,856	1,224,222	86.2
1	MKs	9,870,734	1,711,289	85.22
1	TBt	6,007,771	984,729	85.92
2	Clm	3,114,617	478,659	86.68
2	DTn	3,167,700	429,914	88.05
2	GrF	5,027,247	1,003,458	83.36
2	Hgn	3,049,739	582,461	83.96
2	Hsk	4,147,988	689,059	85.75
2	MKs	5,538,767	1,041,071	84.18
2	TBt	3,280,758	573,625	85.12
4	Iyo	6,403,784	1,489,431	81.13
4	KMd	10,532,775	2,385,670	81.53
4	MMd	3,617,540	779,947	82.26
4	Mct	5,784,886	1,596,173	78.37
4	Org	9,309,100	2,963,891	75.85
4	Pnk	5,917,392	1,390,451	80.97
4	Uns	4,896,555	1,070,724	82.06

サンプル名は第1表を参照

も、7,057 の多型が得られた。これらの多型データを利用して、カロテノイド代謝酵素遺伝子のアレル解析を行った。

第3表 各レファレンス配列 (クレメンティン由来) にリードをマッピングした結果から得られた品種ごとの多型数 (SNP と InDel)

品種名	多型の数 (SNP+InDel)	Clmを100%としたときの多型の割合
Clm	10,224	100%
Dtn	9,761	89%
GrF	20,043	183%
Hgn	37,432	342%
Hsk	19,890	182%
Mks	10,937	100%
TBt	18,951	173%
Iyo	16,510	151%
KMd	13,813	126%
MMd	7,057	65%
Mct	12,288	112%
Org	15,180	139%
Pnk	9,962	91%
Uns	12,543	115%

品種名は第1表を参照

(6) カロテノイド代謝酵素遺伝子のアレル構造の解

14 祖先品種について、レファレンスにマッピングされたリードを解析して、14 祖先品種のアレル構造を解析した。その結果の一例として、*Hyb* のアレル構造を第4表に示した。

第4表 14祖先品種で比較したHyb遺伝子配列におけるSNPの位置

レファレンス配列における位置	280	336	992	1020	1235	1273	2098
CDS番号 / 品種名	CDS.7	CDS.7	CDS.5	CDS.5	CDS.4	CDS.4	CDS.1
Clm	C	C	C	T	A	A	C
Dtn	C	C	AC	T	AG	A	G
GrF	CT	CT	C	CT	A	AC	C
Hgn	CT	CT	AC	T	AG	A	CG
Hsk	CT	CT	AC	T	AG	A	CG
MKs	C	C	C	T	AG	A	CG
TBt	C	C	C	T	AG	A	CG
Iyo	CT	CT	AC	T	AG	A	CG
KMd	C	C	C	T	A	A	CG
MMd	C	C	C	T	A	A	C
Mct	C	C	C	T	AG	A	G
Org	C	C	C	T	A	A	CG
Pnk	C	C	AC	T	AG	A	G
Uns	C	C	C	T	AG	A	CG

レファレンス配列における位置	2169	2200	2201	2204	2222	2233	2301
CDS番号 / 品種名	CDS.1	CDS.1	CDS.1	CDS.1	CDS.1	CDS.1	CDS.1
Clm	T	G	T	C	A	G	T
Dtn	G	A	T	T	A	G	T
GrF	GT	G	T	CT	AT	G	T
Hgn	G	AG	CT	T	AT	AG	CT
Hsk	G	AG	T	T	T	G	T
MKs	GT	AG	T	CT	A	AG	CT
TBt	GT	AG	T	CT	A	G	T
Iyo	G	AG	T	T	AT	G	T
KMd	GT	G	CT	CT	A	G	T
MMd	T	G	T	C	A	G	T
Mct	G	AG	CT	T	A	G	T
Org	GT	G	CT	CT	A	G	T
Pnk	G	A	T	T	A	G	T
Uns	GT	AG	T	CT	A	AG	CT

品種名は第1表を参照

(7) クレメンティン公開ゲノム配列 (version 1.0) における各カロテノイド代謝酵素遺伝子のコピー数

本研究期間中にカンキツにおける最も信頼できるゲノム塩基配列として、クレメンティン半数体のドラフトゲノム配列 Version 1.0 が公開された。この情報を利用して、これまでにウンシュウミカンおよびオレンジで得られてる各カロテノイド酵素遺伝子の塩基配列を質問配列として、相同性検索ソフトウェア BLASTN を用いて、各遺伝子のコピー数とゲノム上の位置を推定した。

その結果、PSY、LCYe、LCYb、Hybは、それぞれ Scaffold 6、1、9、9、に各1コピーずつ存在した。また、PDSは、Scaffold 9に2コピー、ZDSは、Scaffold 5に3コピー存在した(第5表)。

これらのデータは、アレル解析の際の参考にした。

第5表 クレメンティン半数体公開ゲノム(v.1.0)上のカロテノイド代謝酵素遺伝子のコピー数と位置

カロテノイド代謝酵素遺伝子名 (第1図を参照)	Scaffold Id名	Scaffoldでの開始位置	クレメンティン半数体全ゲノム上の遺伝子名
PSY	6	21390477	Ciclev10011841m.g
PDS	9	17778571	Ciclev10007114m.g
	9	17789634	Ciclev10005632m.g
ZDS	5	29890726	Ciclev10002974m.g
	5	30058722	Ciclev10003502m.g
	5	30140475	Ciclev10002967m.g
LCYe	1	10946316	Ciclev10008410m.g
LCYb	9	22728310	Ciclev10004730m.g
Hyb	9	29488658	Ciclev10005481m.g
ZEP	7	3222483	Ciclev10025089m.g
	7	3390375	Ciclev10027323m.g

遺伝子名は第1図を参照

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計2件)

(1) Shimada, T., H. Fujii, T. Endo, T. Ueda, A. Sugiyama, M. Nakano, M. Kita, T. Yoshioka, T. Shimizu, H. Nesumi, Y. Ikoma, T. Moriguchi, M. Omura. Construction of a citrus framework genetic map anchored by 708 gene-based markers. Tree Genetics and Genomes 査読有 2014、DOI 10.1007/s11295-014-0738-9.

(2) Fujii, H., T. Shimada, K. Nonaka, M. Kita, T. Kuniga, T. Endo, Y. Ikoma, M. Omura. High-throughput genotyping in citrus accessions using an SNP genotyping array. Tree Genetics and Genomes、 査読有 . 9 巻、 2013、 145-153 DOI 10.1007/s11295-012-0542-3.

〔学会発表〕(計1件)

杉山愛子、原田健太郎、島田武彦、藤井浩、遠藤朋子、喜多正幸、吉岡照高、生駒吉識、大村三男、eQTLによるUDP-グルコース：リモノイド糖転移酵素遺伝子の発現に関する転写因子候補の探索。日本育種学会第125回講演会、2014年3月22日

6. 研究組織

(1)研究代表者

藤井 浩 (FUJII, Hiroshi)

独立行政法人農業・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・主任研究員

研究者番号：00355398