

カンキツ EST データベースシステムの開発^{†1}

藤井 浩・島田武彦・遠藤朋子・大村三男

独立行政法人農業技術研究機構

果樹研究所カンキツ研究部興津

424-0292 静岡県清水市

Development of Relational Database System for Citrus ESTs

Hiroshi FUJII, Takehiko SHIMADA, Tomoko ENDO and Mitsuo OMURA

Department of Citrus Research, National Institute of Fruit Science

National Agricultural Research Organization, Shimizu, Shizuoka 424-0292, Japan

Summary

A total of 6,154 ESTs were generated from eight cDNA libraries derived from various tissues and developmental stages of citrus fruits. We have constructed the 'Citrus EST database' using Filemaker Pro (Version 5.5-PC/MacOS) in order to manage EST nucleic acid sequences, putative functions and functional annotations. Concurrently, the effective processes of annotation for ESTs and data registration to the database were developed as a database system. The contents of the database may be accessed through keyword searches from the putative gene name, functional annotation or nucleic acid sequence data. Consequently, we were able to search for the needed cDNA clones from the libraries with ease. Utility of the database has been demonstrated through its use to detect several homologous clones associated with the formation of flower buds and terpenoid synthesis. The database would contribute to the development of linkage maps of DNA markers as well as the construction of a cDNA microarray.

Key words: citrus, database, ESTs, cDNA library, functional annotation

緒 言

現在、多数のゲノム関連のデータベースが WWW (World Wide Web) 上で公開されている (Kanehisa, 2000)。大量の核酸配列データが蓄積されている DDBJ (DNA Database of Japan, <http://www.ddbj.nig.ac.jp/Welcom-j.html>) や Genbank (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>),

EMBL (<http://www.ebi.ac.uk/embl/>) といった公的データベースをはじめ, SWISS-PROT (<http://www.edi.ac.uk/swissprot>), PIR (<http://www-nbrf.georgetown.edu/pirwww/dbinfo/piraln.html>) などのアミノ酸配列データベース, Pfam (<http://www.sanger.ac.uk/SoftWare/Pfam>) のようなアミノ酸ドメインデータベース, KEGG (<http://www.genome.ad.jp/kegg/>) のようなパスウェイデータベースなどが公開され (Baxevanis, 2002), インターネット接続環境さえあれば誰でも簡単に利用することができる。このため, ある生命現象について研究を開始するときには, まず, これらのデータベースを検索するのが生物学研究

^{†1} 果樹研究所業績番号: 1306

(2002 年 11 月 28 日受付・2003 年 3 月 7 日受理)

の出発点となりつつある (Bartley and Ishida, 2002)。また、研究対象とする生物に関するデータを WWW 上の多様なデータベースから抽出し、独自の研究データと合わせて、新たなデータベースを構築するという作業は、ゲノム研究および遺伝子研究を効率的に進める上で不可欠である。

本来遺伝子は単体で機能するものではなく、他の遺伝子と協調して機能するため (五斗, 2000)、多数の遺伝子がかかわる複雑な生命現象の解明には、部品としての個々の遺伝子に関する情報の蓄積と、得られた情報が容易に検索・閲覧できる形で管理されているデータベースの存在が研究の前提条件と言える。

果樹研究所カンキツ研究部遺伝解析研究室では、果実形質に関与する網羅的な遺伝子の収集を目的に、カンキツ果実のさまざまな部位や時期に発現する八つの cDNA (complementary DNA) ライブラリーを作成して EST (Expressed Sequence Tags) の解析を行ってきた (Hisada et al., 1996; Hisada et al., 1997; Moriguchi et al., 1998; Kita et al., 2000; Shimada et al., in press; Fujii et al., 2003)。その結果、6,154 クローンの cDNA に関する EST のカタログ情報を得た。EST に関する情報は、遺伝子の収集だけでなく、マッピングや発現解析、遺伝子推定といったゲノム研究に広く利用可能である (Wolfsberg and Landsman, 2001)。

これまで我々は、EST のカタログ情報をライブラリーごとに表計算ソフトウェアによって管理していたが、データの量的な拡大に伴い、検索・閲覧やデータの共有が困難になった。このため、得られた EST に関する情報を一元的に管理し、研究グループ内で共有するためのシステムが必要になった。大量の EST データを管理・共有する方法として、市販の遺伝子情報管理ソフトを利用する方法と、Linux などの Unix 系 OS の上で動くフ

リーソフトを利用して構築する方法が考えられた。市販の管理ソフトはコンピュータに関する専門知識なしに利用できるが、我々のように一研究室単位でゲノム解析を行っているような小規模な研究グループには高価であり、また、研究内容に合わせたシステムのカスタマイズはメーカーに依頼しなければならない。一方、Linux はパーソナルコンピュータ上で動き、MySQL や PostgreSQL のようなフリーの優れたデータベースソフトウェアが利用できるため、安価にシステムを構築できるが、データベースの作成や利用には Linux に関する専門的な知識が必要である。

そこで、我々は Windows と MacOS 上で動く市販のデータベースソフトウェア FileMakerPro を使用して、すべての情報を統合的に管理する「カンキツ EST データベース」を作成した。FileMakerPro は入手し易く、また、専門的な知識がなくてもデータベースの構築や変更およびデータの共有が容易に行える。このデータベースを中心に、EST に対して機能推定情報や機能注釈情報を効率的に付加する手順および得られた情報を半自動的に登録するための手順などを含めたデータベースシステムを開発したので、その内容を報告する。

材料および方法

1. 使用したデータ

「カンキツ EST データベース」(以下「データベース」と略す。)に収録した EST データは、カンキツ果実に由来する八つの cDNA ライブラリーから得られた。これらのライブラリーの略号とそれぞれのライブラリーが由来する果実の品種・組織・採集時期および cDNA クローン数・塩基配列数・塩基数・DDBJ への登録数を第 1 表に示した。各ライブラリーの作成方法およびシーケンスの

第 1 表 「カンキツ EST データベース」に収録されている EST が得られた cDNA ライブラリーの由来とそれぞれのクローン数、配列数、塩基数

ライブラリー名	品種	組織	時期	クローン数	配列数	塩基数(bp)	DDBJ 登録数
VSS	バレンシア	幼種子	受粉後 1 ヶ月	532	577	328,017	87
FRI	宮川早生	果肉	肥大期	874	1,054	276,690	874
FRM	宮川早生	果肉	成熟期	308	386	138,527	297*
ALM	宮川早生	アルベド	成熟期	442	622	361,752	377
OVA	宮川早生	子房	開花期	590	827	403,038	393*
ALP	宮川早生	アルベド	剥皮開始期	626	941	413,721	0
WFY	宮川早生	全果	幼果期(開花後 1 ヶ月)	1,235	1,688	1,042,826	0
BFC	宮川早生	果皮	着色期	1,547	1,654	925,469	0
合計				6,154	7,749	3,890,040	2,028 (690*)

*2002 年 11 月 1 日現在非公開

方法は, Hisada et al. (1996), Hisada et al. (1997), Moriguchi et al. (1998), Kita et al. (2000), Shimada et al. (in press) および Fujii et al. (2003) に示した。

VSS ライブラリーの一部と FRI, FRM, ALM, OVA の EST データはデータベースの構造やデータ登録処理手順が作成される以前に Excel (Microsoft 社) 形式でカタログ化されていたので (Hisada et al., 1996; Hisada et al., 1997; Moriguchi et al., 1998; Kita et al., 2000), このカタログ情報を登録データとした。VSS の一部と ALP, WFY, BFC については後述のデータ登録処理手順に従って, シーケンサーから出力された塩基配列の処理および相同性検索, 機能推定, 機能注釈を行い, それらの情報を順次登録した。

2. 使用したシステム

データベースの開発に使用したシステムの構成を第2表に示した。データベースの構築には市販のデータベースソフトウェア FileMakerPro (Version 5.5; FileMaker 社) を使用し, 果樹研究所カンキツ研究部興津の LAN に接続された Macintosh G3 (Apple 社) と Inspiron4000 (Dell 社), Dimension4100 (Dell 社) 上で開発した。

Macintosh G3 ではデータベース構造の構築と塩基配列データの登録処理を行った。また, 塩基配列データの登録処理を効率的に行うために, プログラム言語 Realbasic (Version 2.0; ASCII 社) によってツールを開発

した。Inspiron4000 は, シーケンス・アセンブリ・ソフトウェア ATGC (Version 2.0; ソフトウェア開発社) を用いて, シーケンサーから出力された塩基配列からベクターに由来する部分を除去するなどの配列処理に用いた。また, 解析済みのカンキツ EST 塩基配列と新たにシーケンスして得られた EST 塩基配列との重複性を調べるために, プライベート・データベース・ソフトウェア GENETYX-PDB (Version 3.0; ソフトウェア開発社) を用いた。Dimension4100 は, 機能推定情報と機能注釈情報の登録に使用した。

FileMakerPro にはネットワーク上で複数のユーザーが同時にファイル共有を行う機能がある。また, MacOS 用の FileMakerPro で作られたファイルであっても, Windows 用の FileMakerPro でそのまま閲覧・編集することが可能である。これらの機能を利用して, Macintosh G3 をデータベースのサーバーとし, FileMakerPro を導入した他の2台の Windows マシンから LAN を通じてデータベースファイルを呼び出し, データの登録を行った。

結 果

1. データベースの構造

データベースは, EST の塩基配列を管理するテーブル CHAOS と相同性検索結果を管理するテーブル APOLLO,

第2表 システムを構成するハードウェアとソフトウェア

ハードウェア	データベース開発兼サーバー用マシン	Macintosh G3 (Apple 社)	
		CPU	PowerPC G3 450MHz
	塩基配列処理用マシン	内蔵メモリ	256MB
		HDD	8.5GB
		OS	MacOS 8.6 日本語
		Inspiron4000 (DELL 社)	
	機能推定・機能注釈登録用マシン	CPU	Pentium III 800MHz
		内蔵メモリ	256MB
		HDD	18.4GB
		OS	Windows Me
	機能推定・機能注釈登録用マシン	Dimension4100 (DELL 社)	
		CPU	Pentium III 733MHz
		内蔵メモリ	128MB
		HDD	14.2GB
	機能推定・機能注釈登録用マシン	OS	Windows 98SE
ソフトウェア	データベースソフトウェア	Windows 用 FileMaker Pro (Version 5.5; FileMaker 社)	
		MacOS 用 FileMaker Pro (Version 5.5; FileMaker 社)	
		MacOS 用 Realbasic (Version 2.0; ASCII 社)	
		Windows 用 ATGC (Version 2.0; ソフトウェア開発社)	
		Windows 用 GENETYX-PDB (Version 3.0; ソフトウェア開発社)	
	プログラム開発言語	Windows 用 ATGC (Version 2.0; ソフトウェア開発社)	
		Windows 用 GENETYX-PDB (Version 3.0; ソフトウェア開発社)	
		Windows 用 GENETYX-PDB (Version 3.0; ソフトウェア開発社)	
	シーケンスアセンブリソフトウェア	Windows 用 ATGC (Version 2.0; ソフトウェア開発社)	
		Windows 用 GENETYX-PDB (Version 3.0; ソフトウェア開発社)	
		Windows 用 GENETYX-PDB (Version 3.0; ソフトウェア開発社)	
	プライベートデータベースソフトウェア	Windows 用 GENETYX-PDB (Version 3.0; ソフトウェア開発社)	
		Windows 用 GENETYX-PDB (Version 3.0; ソフトウェア開発社)	
		Windows 用 GENETYX-PDB (Version 3.0; ソフトウェア開発社)	

機能推定の結果や機能注釈などの情報を管理するテーブル ZEUS および EST が由来する cDNA クローンの保存情報を管理するためのテーブル GAIA から構成される (第1図)。CHAOS, APOLLO, ZEUS および GAIA はそれぞれのテーブルの愛称である。これらのテーブルは、各 EST に一意的に付された 8 桁の EST コードをプライマリーキーとして結び付けられたリレーショナルデータベースを構成する。

データベースの初期画面を第2図Aに示した。この初期画面には、ライブラリー名を記した2列のボタンが配置してある。上段のボタンをマウスでクリックすると対応するライブラリーの ZEUS 表示画面 (第2図B) へ移動し、下段のボタンをクリックすると対応するライブラリーの APOLLO 表示画面 (第2図C) に移動する。初期画面とそれぞれの表示画面に貼り付けられたボタンには、FileMakerPro に付属するスクリプト言語 (簡易プログラム言語) によって処理手順が関連付けられており、ボタンをクリックすることによって処理が自動的に実行される。ZEUS 表示画面 (第2図B) では、ZEUS のテ-

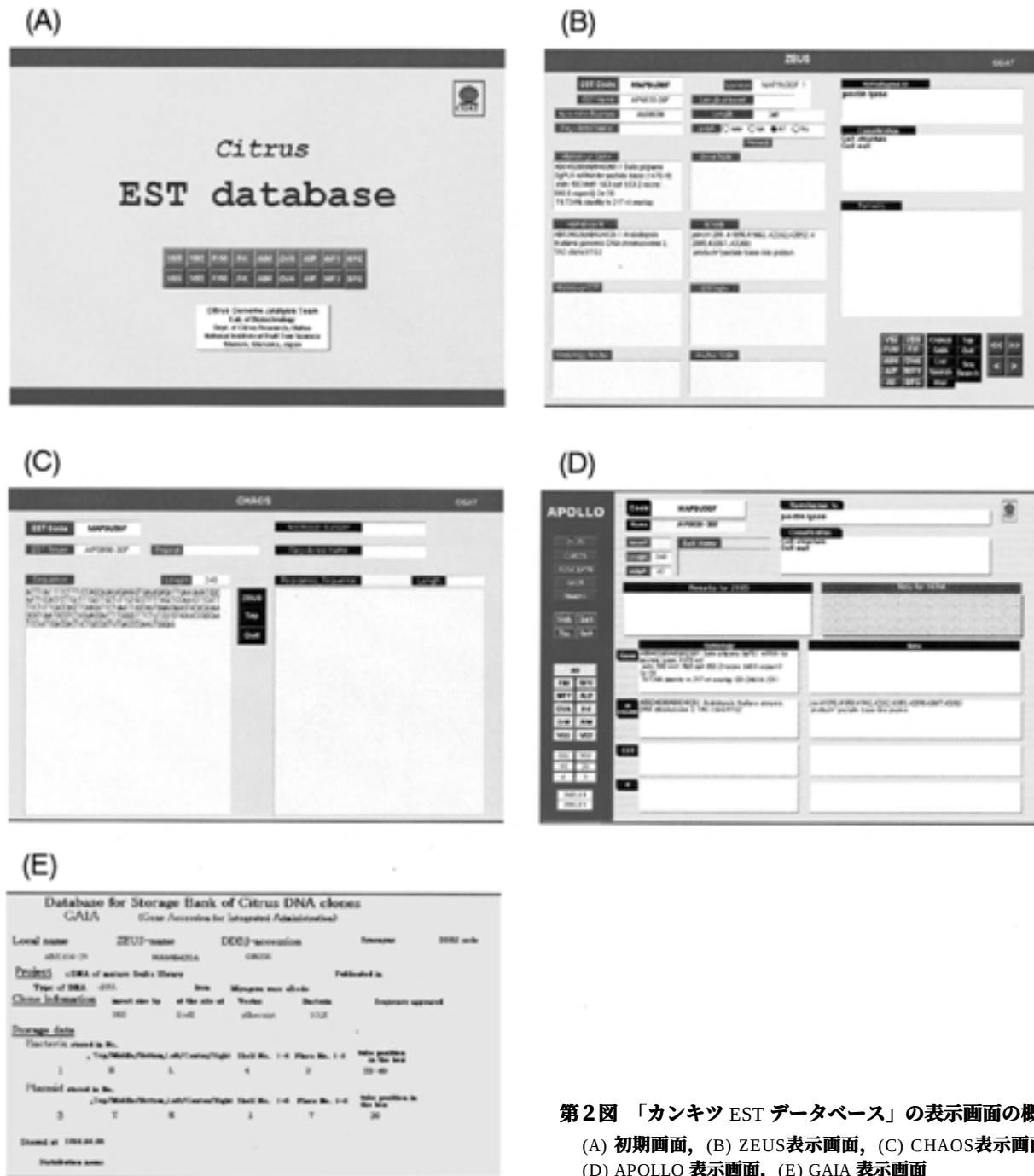
ブルに収録された情報だけでなく、各テーブルからの情報を引用し、EST に関する情報を統合して表示できるように設計した。また、ZEUS 表示画面 (第2図B) の「CHAOS」ボタンをクリックすると、CHAOS 表示画面 (第2図D) に移動し、EST の塩基配列を閲覧できる。同様に、ZEUS の表示画面 (第2図B) の「GAIA」ボタンをクリックすると GAIA の表示画面 (第2図E) に移動し、cDNA クローンの保存情報を閲覧できる。

2. ツールの開発

シーケンサーで解読された塩基配列データは、配列ごとに別々のテキストファイルとして出力される。これらのデータを効率的にデータベースに登録するために、プログラム言語 Realbasic を使用して MacintoshG3 上でツールを開発した。このツールは、ファイル名を EST コードとし、ファイルの内容をその塩基配列とするようなテキストファイル群を1つの CSV (Comma Separated Value) ファイルに変換するツールで、oktCSV と名づけた。また、逆に、データベースから出力された CSV 形式



第1図 「カンキツ EST データベース」のスキーマモデル



のファイルを各行ごとに別々のテキストファイルに分割するツールも作成し、`oktText` と名づけた。

3. GENETYX-PDB による PDB (Private DataBase) の作成

過去のカタログ化で得られた EST の塩基配列と新たに得られた EST との重複性を調べるために GENETYX-PDB を用いた。GENETYX-PDB で相同性検索を行うには、特有のファイル構造を持つプライベート・データベース (PDB) を作成し、過去に得られたすべての EST

第2図 「カンキツ EST データベース」の表示画面の概要

(A) 初期画面, (B) ZEUS表示画面, (C) CHAOS表示画面,
(D) APOLLO 表示画面, (E) GAIA 表示画面

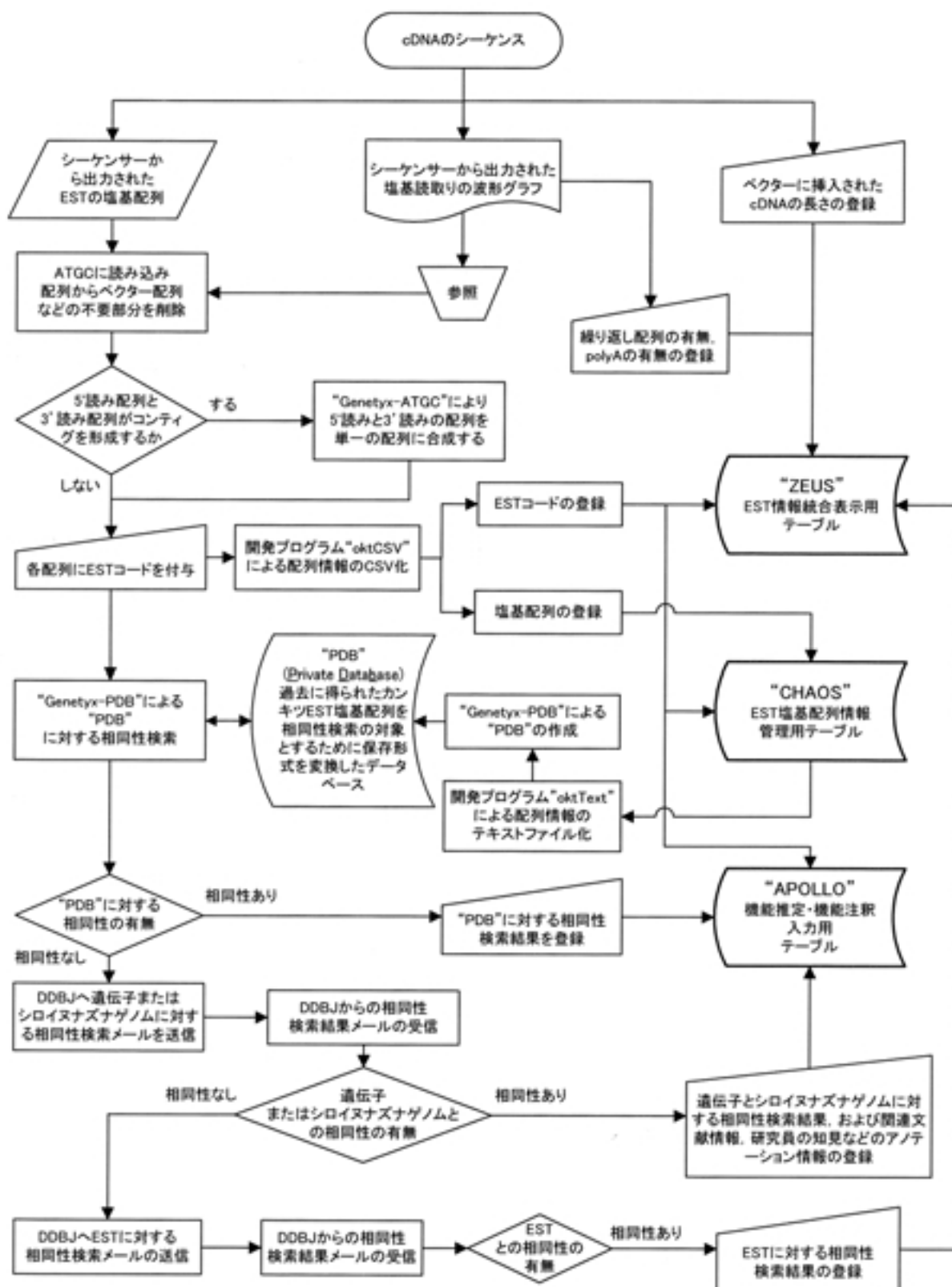
塩基配列を登録しなければならない。PDB の作成手順を以下に示した。

- 1) データベースのレコードの書き出し機能を使って、全レコードの EST コードとその塩基配列を CSV ファイル形式で書き出す。
- 2) oktText を用いて、EST コードをファイル名とし、塩基配列を内容とするテキストファイル群を作成する。
- 3) このテキストファイル群を GENETYX-PDB の機能を用いて、PDB に変換する。

4. データ登録手順

シーケンサーから出力された EST 塩基配列の機能推

定や機能注釈を行い、その情報をデータベースに登録する手順を以下に示した (第3図)。



第3図 「カンキツ EST データベース」 へのデータ登録処理の流れ

- 1) シーケンサーから出力されたテキストファイル形式のESTの塩基配列データをInspiron4000上で動作するソフトウェアATGCに読み込み、プリンタ出力された塩基読み取りの波形グラフを参照しながら、ベクター塩基配列などの不要部分を除去する。
- 2) ATGCの機能を使って、同一クローンの5'側と3'側からの塩基配列のアライメントを取り、コンティグを形成する場合は1つの塩基配列に合成する。
- 3) ESTに一意的な8桁のESTコードを付し、ESTコードをファイル名とし、ファイル内容をその塩基配列とするテキストファイルを作成する。
- 4) 作成したテキストファイル群をMacintoshG3に転送する。oktCSVによって、複数のESTコードと塩基配列を1つのCSVファイルに変換し、その内容をFileMakerProのレコードの取り込み機能を利用して、CHAOSにはESTコードと塩基配列、APOLLOとZEUSにはESTコードを一括して登録する。
- 5) GENETYX-PDBを用いて相同性検索を行い(Lipman-Pearson法)、新しく得られたESTが過去に既に得られているESTと塩基配列が重複しているか否かを判定する。
- 6) ESTが過去に得られたESTと重複していた場合は、APOLLOの表示画面(第2図C)の過去に得られたESTとの相同性検索結果を記入するためのフィールド「Self Homo」にその結果を記入する。
- 7) 過去のESTと重複性がないESTについては、DDBJに対してFASTAを用いた相同性検索メールを送信する。メール送信は、ZEUS表示画面(第2図B)に配置した「Mail」ボタンを押すと自動的に実行される。
- 8) DDBJからの相同性検索結果メールを受信し、OPTの値が400以上で他の生物の遺伝子と相同性が見られた場合にはAPOLLO表示画面(第2図C)の対応するフィールド「Gene」に相同性検索結果メール上のデータを転記する。遺伝子との相同性はないが、シロイヌナズナゲノムあるいは他の生物のESTと相同性が見られた場合には、それぞれの対応するフィールド「At genome」,「EST」に相同性検索結果メール上のデータを転記する。
- 9) 相同性検索結果から、更に文献や他のゲノムデータベースを検索し、追加的な注釈情報をAPOLLO表示画面(第2図C)のフィールド「Note」に記入する。
- 10) 相同性検索結果から推定される遺伝子名をAPOLLO表示画面(第2図C)の「Homologous to」に、遺伝子の機能分類を「Classification」に記入する。また、総合的な注釈を「Remarks」に記入する。

- 11) ベクターに挿入されたcDNAの長さや、シーケンサーから出力された配列の波形グラフから読み取ったSSR(Simple Sequence Repeats)の有無、cDNA末端のpolyAの有無といった付帯情報をZEUS表示画面(第2図B)に記入する。

5. データベースの利用法

FileMakerProにはLAN経由でファイルを共有する機能がある。このため、同一LAN内という制約はあるが、研究グループのメンバー各自のLANに接続されたパソコンからデータベースを利用できる。

収録したデータの検索機能として、ZEUS表示画面(第2図B)の各フィールドに対して、文字列検索を行える。例えば、画面右上の推定された遺伝子名を記入したフィールド「Homologous to」では、遺伝子名によってESTを検索可能であり、それ以外の「Classification」(遺伝子の機能分類)や「Remarks」(総合的な注釈情報)のフィールドでも任意の文字列で検索が可能である。興味のあるESTが検索された場合には、CHAOS表示画面(第2図D)でESTの塩基配列を閲覧できる。

6. データベースのセキュリティ対策

データベースは、ファイアウォールソフト(Norton Personal Firewall, Symantec社)を用いたIPアドレスによるコンピュータ単位でのアクセス制限とFileMakerProのパスワード管理機能によるユーザー単位のアクセス制限によってセキュリティ管理を行っている。

パスワード管理機能は、ユーザー単位だけでなく、レコードおよびフィールド単位で制御することができる。従って、特定のレコードを閲覧不能にする、あるいは、特定のユーザーに対して特定のフィールドのみへの書き込みを許可するなどの設定を行い、誤入力危険を回避している。

考 察

本研究では、パーソナルコンピュータで動くデータベースソフトウェアFileMakerProを用いて、これまでに解析したすべてのEST情報を収録したデータベースを構築した。併せてESTの機能推定情報や、機能注釈情報の効率的な付加手順や半自動的な情報の登録手順を含めた、データベースシステムを開発した。使用したソフトウェアは、すべてWindowsまたはMacOS上で動く市販のものであり、特別な機器や専門的な知識なしに、システムの構築や運営が可能である。このため、情報科学

の専門家のいない研究グループでも、このシステムを用いれば、ESTの大量解析の結果を容易にデータベース化し、LAN経由で情報を共有できる。

データベース化と情報の共有によって、研究グループ内での重複的な実験を省き、効率的な研究遂行が可能となった。例えば、データベースの検索によって既存のcDNAライブラリー内に研究対象としている遺伝子候補クローンの存在が確認できれば、保存場所から対象クローンを取り出すだけで、単離実験なしに目的のクローンを入手できる。これまでに、データベースの利用によって、花芽形成に関連が予想される数種類の遺伝子ホモログを発見した(遠藤ら, 2002)ほか、イソプレノイド代謝系遺伝子の解析においても、テルペノイド合成に関わる多数の遺伝子ホモログを発見した(島田ら, 2001b; 島田ら, 2002b)。また、DNAマーカーおよびcDNAマイクロアレイの作成にも、データベースの情報が利用され、228のDNAマーカーによる803cMのカンキツ連鎖地図の作成(大村ら, 2000; 大村ら, 2001)や、約2,300スポットのcDNAマイクロアレイの作成(島田ら, 2001a, 2002b)に貢献した。

データベースの存在意義は、上記のような研究を支援するためだけではない。最近、ゲノム研究においてバイオインフォマティクスの重要性が増しており、果樹のゲノム研究においても、シロイヌナズナやイネとの*in silico*(生命現象・機能・構造をコンピュータ上で解析する研究手法)でのゲノム比較研究から多くの有益な情報が得られると期待されている。*in silico*での実験には、収集した情報から研究の目的に沿った情報を自在に選択して利用できる環境が必要であり、データが一元的に管理されたデータベースの存在は不可欠である。この点においても、今後データベースはカンキツゲノム研究および分子生物学的解析の基盤的な情報となり得る。

データベースの利用は、既にいくつかの成果をあげているが、残された課題もある。カンキツと同じ顕花植物に属するシロイヌナズナ(*Arabidopsis thaliana*)では、全ゲノム塩基配列の解説が完了し、ゲノム上に25,498個の遺伝子が存在することが推定されている(The Arabidopsis Genome Initiative, 2000)。一方、我々が収集・登録したcDNAクローンの数は、リダンダントなものも含めて6,154個にすぎない。今後も新たなcDNAライブラリーの作成とシーケンスを行い、EST情報を追加していく必要がある。

また、公的データベースに対して相同性検索を用いたESTの機能推定のシステムも充分とはいえない。我々の機能推定はDDBJの核酸配列データベースに対して

FASTAを用いて相同性検索を行った結果のみを用いているが、マウスの完全長cDNAデータベースFANTOM(<http://www.gsc.riken.go.jp/e/FANTOM>)では、核酸配列やタンパク質配列、タンパク質モチーフ、タンパク質ドメイン、タンパク質ファミリーに関する情報を蓄積した14のデータベースに対して、BLAST, FASTA, Wise2を用いた相同性検索の結果から機能推定を行っている(Kawai et al., 2001)。我々も、より高度な機能推定を行うため、複数のデータベースに対して複数の相同性検索ソフトウェアを用いた検索を行うシステムを開発する必要がある。

カンキツ特有の新規な機能を遺伝子cDNAクローンについては、上記のようなESTと既存の遺伝子データとの相同性検索だけでは機能を推定することは困難である。このような遺伝子機能の解明にはcDNAマイクロアレイによる発現情報の解析が有効であるため、引き続きデータベースからcDNAマイクロアレイ作成に最適なクローンを抽出して提供するとともに、今後の発現情報解析の進展に備えて、cDNAマイクロアレイの発現情報をデータベースと連携させる方法を確立することが課題である。

摘 要

カンキツ果実に由来する八つのcDNAライブラリーから得られた6,154のcDNAクローンについてESTを得た。これらのESTに関する情報を一元的に管理するために、パーソナルコンピュータで動くデータベースソフトウェアFileMakerPro (Version5.5; FileMaker社)を使用して「カンキツESTデータベース」を構築した。併せてESTの機能推定や機能注釈を効率的に行うための手順や、得られた情報を半自動的に登録するための手順を含めたデータベースシステムを開発した。本システムで使用したソフトウェアは、すべてWindowsまたはMacOSで動く市販のものであり、専門的な知識なしにシステムの変更や運営、データの共有が可能である。データベースには、推定された遺伝子名や機能注釈情報に対してキーワード検索を行う機能があり、研究に必要なcDNAクローンをライブラリーから容易に検索することができる。データベースの利用によって、花芽形成に関連が予想される遺伝子ホモログやテルペノイド合成に関わる多数の遺伝子ホモログを発見することができた。また、DNAマーカーの開発やcDNAマイクロアレイ作成に貢献した。このような実験系へのデータの提供のほかに、今後重要性を増すと予想されるバイオインフォマティク

スの手法を用いた *in silico* での研究にも利用可能である。

引用文献

- 1) Bartley, G. E. and B. K. Ishida. 2002. Digital fruit ripening: data mining in the TIGR tomato gene index. *Plant Mol. Biol. Repr.* 20: 115-130.
- 2) Baxevanis, A. D. 2002. Molecular biology database collection: 2002 update. *Nucl. Acids. Res.* 30: 1-12.
- 3) 遠藤朋子・島田武彦・藤井 浩・大村三男. 2002. カンキツ由来 FT/TFL 相同性クローンの発現解析. *園学雑.* 71(別 2): 257.
- 4) Fujii, H., M. Kita, T. Shimada, T. Endo and M. Omura. 2003. Expressed sequence tags from citrus albedo at the initiation stage of rind peeling. *Bull. Natl. Inst. Fruit Tree Sci.* 2: 127-144.
- 5) 五斗 進. 2000. パスウェイからのゲノム機能解析. p.84-99. : 高木利久・富田勝(編) *ゲノム情報生物学*. 中山書店. 東京.
- 6) Hisada, S., T. Moriguchi, T. Hidaka, A. M. Koltunow, T. Akihama and M. Omura. 1996. Random sequencing of sweet orange (*Citrus sinensis* Osbeck) cDNA library derived from young seeds. *J. Japan. Soc. Hort. Sci.* 65: 487-495.
- 7) Hisada, S., T. Akihama, T. Endo, T. Moriguchi and M. Omura. 1997. Expressed sequence tags of citrus fruit during rapid cell development phase. *J. Amer. Soc. Hort. Sci.* 122: 808-812.
- 8) Kanehisa, M. 2000. Post-genome informatics. Oxford University Press, New York. p.24-29.
- 9) Kawai, J., A. Shinagawa, K. Shibata, M. Yoshino, M. Itoh, Y. Ishii, T. Arakawa, A. Hara, Y. Fukunishi, H. Konno, J. Adachi, S. Fukuda, K. Aizawa, M. Izawa, K. Nishi, H. Kiyosawa, S. Kondo, I. Yamanaka and T. Saito. 2001. Functional annotation of a full-length cDNA collection. *Nature* 409: 685-690.
- 10) Kita, M., T. Endo, T. Moriguchi and M. Omura. 2000. cDNA catalogs expressed in albedo of citrus fruit: a comparative analysis of cDNA libraries from pulp and albedo of Satsuma mandarin (*Citrus unshiu* Marc.). *Acta Hort.* 521: 179-183.
- 11) Moriguchi, T., M. Kita, S. Hisada, T. Endo and M. Omura. 1998. Characterization of gene repertoires at mature stage of citrus fruits through random sequencing and analysis of redundant metallothionein-like expressed during fruit development. *Gene* 211: 221-227.
- 12) 大村三男・上田高則・島田武彦・遠藤朋子・藤井 浩・喜多正幸・小松晃・根角博久・吉田俊雄. 2000. カンキツ EST のマッピング (その 2). 2000. *育種学研究* 2(別 2): 22.
- 13) 大村三男・上田高則・島田武彦・遠藤朋子・喜多正幸・藤井 浩・小松 晃・根角博久・吉田俊雄. 2001. カンキツ EST に基づく CAPS マーカーによるグラフ連鎖地図の作成. *園学雑.* 70(別 1): 214.
- 14) 島田武彦・遠藤朋子・喜多正幸・藤井 浩・上田高則・大村三男. 2001a. カンキツでの cDNA マイクロアレーを利用した遺伝子発現プロファイリング. その 1 アレー解析の確立と有効性の検討. *園学雑.* 70(別 1): 75.
- 15) 島田武彦・遠藤朋子・藤井 浩・上田高則・大村三男. 2001b. カンキツセスキテルペノイド合成遺伝子 (CitSTS1) ホモログの単離と発現解析. *育種学研究* 3(別 2): 149.
- 16) 島田武彦・遠藤朋子・藤井 浩・上田高則・久保達也・矢崎潤史・中村桂子・藤井文子・真保佳納子・山本公子・坂田克己・佐々木卓治・岸本直己・菊池尚志・大村三男. 2002a. カンキツでの cDNA マイクロアレーを利用した遺伝子発現プロファイリング. その 2 幼植物・花・幼果での発現プロファイリング. *園学雑.* 71(別 1): 59.
- 17) 島田武彦・遠藤朋子・藤井 浩・原 正和・上田高則・久保達也・大村三男. 2002b. ウンシュウミカンの香気に関与する γ -テルピネン合成酵素遺伝子の単離と解析. *育種学研究* 4(別 2): 145.
- 18) Shimada, T., M. Kita, T. Endo, H. Fujii, T. Ueda, T. Moriguchi and M. Omura. Expressed Sequence Tags of ovary tissue cDNA library in *Citrus unshiu* Marc. *Plant Sci.* in press.
- 19) The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plants *Arabidopsis thaliana*. *Nature* 408: 796-815.
- 20) Wolfsberg, T. G. and D. Landsman. 2001. Expressed sequence tags (ESTs) In: Baxevanis, A. D. and B. F. F. Queller (eds.) *Bioinformatics* 2nd ed. p.283-302. Wiley-Interscience. New York.