第1章　第1節

果樹研究のためのゲノム情報

農研機構果樹研究所 カンキツ研究領域　　藤井　浩

カンキツやリンゴ，ブドウ，モモといった主要な果樹を含めて，ウェブ上のサーバから果樹研究に利用可能なドラフトゲノム情報が公開されている（表1-1-1）．これら情報を利用することで，研究対象とする遺伝子ホモログ（homolog）のゲノム上の個数や位置情報，上流域・下流域の塩基配列やエキソン－イントロンの構造を簡単に入手できる．また，全ての予測遺伝子の塩基配列もマルチFASTA形式のファイルとしてサーバに保存されているので，ダウンロードして独自のマイクロアレイの設計に活用できる．実験計画をたてる際に，既知のゲノム情報を参照することで，研究の効率化が期待できる．本節では，ウエットな実験を中心に研究を進めている研究者を対象に，果樹研究にとって基本的なゲノム情報をウェブブラウザを使って入手する方法と，入手したゲノム情報をおもにWindows環境で閲覧する方法を解説する．紹介するソフトウエアには，有償のものを含め，できるかぎりグラフィカルなインターフェースやサポートを備えているものを選んだ．

本格的なLinuxを利用したバイオインフォマティクス用の環境の構築については，巻末に参考図書を掲げるにとどめて，本節では述べない．次世代シークエンサに関するバイオインフォマティクスツールに関しては次節で，ゲノムアノテーション（genome annotation）に関するツールは本章第3節で，ゲノムデータからDNA多型（polymorphism）を検出するためのツールについては本章第4節で，マイクロアレイ設計に関するツールは本章第5節で詳述する．

Hiroshi FUJII　✉ hfujii@affrc.go.jp
Genome information for fruit tree science

1）Phytozome

ゲノム情報は，多くの場合，それぞれのゲノム解読プロジェクトのサイトから公開されているが（表1-1-1），米国エネルギー省の研究所であるJoint Genome Institute（JGI）が開設しているゲノムサイトPhytozome（URL1-1-1, Neupane, R. *et al.* 2012）には，2016年1月時点で，57種の生物種に関する63のゲノムデータベースが収録されている（図1-1-1）．この中には，クレメンティン（*Citrus clementina* v1.1），スイートオレンジ（*Citrus sinensis* v1.1），リンゴ（*Malus domestica* v1.0），ヨーロッパブドウ（*Vitis vinifera* Genoscope 12X），モモ（*Prunus persica* v1.0, v2.1）など主要な果樹とモデル植物であるシロイヌナズナ（*Arabidopsis thaliana* TAIR10）などのゲノム情報が収録されている．（学名の後ろの番号はそれぞれのゲノムアセンブリのバージョンを示している）．Phytozomeの利点は，共通の操作方法でゲノム情報が利用可能で，複数種を指定した横断検索もできるところにある．なお，果樹の公開ゲノム情報のうち，中国のグループが発表しているスイートオレンジ（*Citrus sinensis* Annotation project, URL 1-1-2）やチュウゴクナシ（*Pyrus bretschneideri* : Pear Genome Project, URL1-1-3）のゲノム情報はphytozomeに収録されていない．また，バラ科ゲノムデータベース（Genome Database for Rosaceae, GDR, URL1-1-4）に公開されているモモのマーカー情報（図1-1-2）のように，各ゲノムプロジェクト独自の解析については，収録されていない．こうした独自の内容の解析データが追加・更新されることもあるので，各ゲノムプロジェクトのオリジナルサイトも巡回する必要がある．

（1）遺伝子検索

研究対象とする遺伝子ホモログをゲノム情報から網羅的に取得する場合，遺伝子名によるキーワード検索と配列の類似性によるBLAST（URL1-1-5, Altschul *et al.* 1997）検索の双方の検索を必ず行うべきである．キーワード検索のみの場合，配列が類似しているホモログを収集できない可能性があり，

表1-1-1　温帯果樹のゲノムデータベース

名称・学名・系統	ゲノムデータベース	URL番号 (URL目次を参照)	アセンブリバージョン	論文
クレメンティン *Citrus clementina* Clemenules (haploid)	Citrus Genome Database Phytozome	URL1-4-6 URL1-1-1	v0.9, v1.0 v1.0	Wu *et al.* 2014
スイートオレンジ *Citrus sinensis* Ridge Pineapple	Citrus Genome Database Phytozome	URL1-4-6 URL1-1-1	v1.0 v.1.1	Wu *et al.* 2014
スイートオレンジ *Citrus sinensis* Valencia	Citrus sinensis Annotation Project	URL1-1-2	v1.0, v2.0, v2.1	Xu *et al.* 2013
リンゴ *Malus domestica* Golden Delicious	Genome Database for Rosaceae (GDR) Phytozome	URL1-1-4 URL1-1-1	v1.0, v1.0p, v2.0, v3.0.a1 v1.0	Velasco *et al.* 2010
ヨーロッパブドウ *Vitis vinifera* PN40024	Grape Genome Browser (Genoscope) Phytozome VvGDB	URL2-4-1 URL1-1-1 URL1-4-5	8X, 12X 12X 12X	Jaillon *et al.* 2007
モモ *Prunus persica* Lovell (double haploid)	Genome Database for Rosaceae (GDR) Phytozome	URL1-1-4 URL1-1-1	v1.0, v2.0.a1 v1.0, v2.1	The International Peach Genome Initiative 2013
セイヨウナシ *Pyrus communis* Bartlett	Genome Database for Rosaceae (GDR)	URL1-1-4	v1.0	Chagné *et al.* 2014
チュウゴクナシ *Pyrus bretschneideri* Dangshansuli	Pear Genome Project	URL1-1-3	v1.0	Wang *et al.* 2013

Amaranthus hypochondriacus v1.0	*Eucalyptus grandis v2.0*	*Physcomitrella patens v3.3*
Amborella trichopoda v1.0	*Eutrema salsugineum v1.0*	*Populus trichocarpa v3.0*
Ananas comosus v3	*Fragaria vesca v1.1*	*Prunus persica v1.0*
Aquilegia coerulea v1.1	*Glycine max Wm82.a2.v1*	*Prunus persica v2.1*
Arabidopsis halleri v1.1	*Gossypium raimondii v2.1*	*Ricinus communis v0.1*
Arabidopsis lyrata v1.0	*Kalanchoe marnieriana v1.0*	*Salix purpurea v1.0*
Arabidopsis thaliana TAIR10	*Linum usitatissimum v1.0*	*Selaginella moellendorffii v1.0*
Boechera stricta v1.2	*Malus domestica v1.0*	*Setaria italica v2.1*
Brachypodium distachyon v2.1	*Manihot esculenta v4.1*	*Setaria italica v2.2*
Brachypodium distachyon v3.1	*Manihot esculenta v6.1*	*Setaria viridis v1.1*
Brachypodium stacei v1.1	*Medicago truncatula Mt4.0v1*	*Solanum lycopersicum ITAG2.3*
Brassica rapa FPsc v1.3	*Micromonas pusilla CCMP1545 v3.0*	*Solanum tuberosum v3.4*
Capsella grandiflora v1.1	*Micromonas sp. RCC299 v3.0*	*Sorghum bicolor v2.1*
Capsella rubella v1.0	*Mimulus guttatus v2.0*	*Sorghum bicolor v3.1*
Carica papaya ASGPBv0.4	*Musa acuminata v1*	*Sphagnum fallax v0.5*
Chlamydomonas reinhardtii v5.5	*Oryza sativa v7.0*	*Spirodela polyrhiza v2*
Citrus clementina v1.0	*Ostreococcus lucimarinus v2.0*	*Theobroma cacao v1.1*
Citrus sinensis v1.1	*Panicum hallii v2.0*	*Triticum aestivum v2.2*
Coccomyxa subellipsoidea C-169 v2.0	*Panicum virgatum v1.1*	*Vitis vinifera Genoscope.12X*
Cucumis sativus v1.0	*Phaseolus vulgaris v1.0*	*Volvox carteri v2.0*
Eucalyptus grandis v1.1	*Physcomitrella patens v3.0*	*Zea mays 6a*

図1-1-1　Phytozomeから提供されているゲノムデータ
57種の生物について63種類のゲノムデータが収録されている

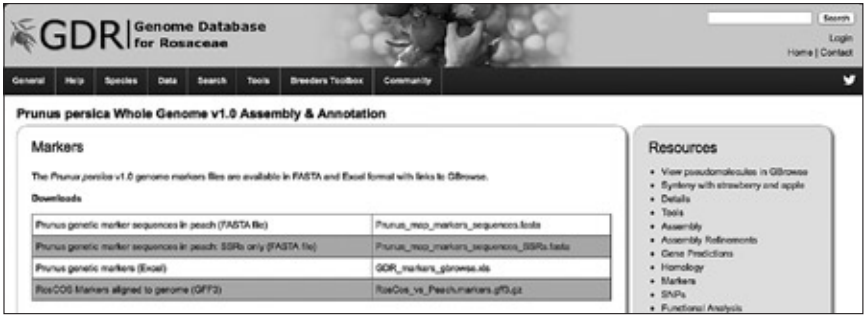

図1-1-2　GDRのモモの遺伝マーカーのダウンロード画面

BLAST検索のみだと配列が類似していないが機能が類似しているホモログを収集できないからである．Phytozomeのトップページ上部のメニューバーの左から２つ目の「Tools」を選択すると（図1-1-3），プルダウンメニューから「Keyword search」，「BLAST」，「BLAT（URL1-1-6, Kent 2002）」などのツールが選べる．例としてクレメンティンを対象に「MADS」という名称の遺伝子を検索してみる．「Keyword search」を選択して，「*Citrus clementine* v1.0」を選択したのち、キーワード入力欄に「MADS」と記入して（図1-1-4）検索を実行すると，図1-1-5のように遺伝子の注釈（アノテーション，annotation）に「MADS」という文字列が含まれている遺伝子が列挙される．目的のMADS遺伝子の行の左端の白抜きの「B」を選ぶとゲノムブラウザが起動して，当該遺伝子のゲノムにおける位置や周辺の遺伝子の状況が画像として表示され（図1-1-6），「G」を選ぶとPANTHER（Protein ANalysis THrough Evolutionary Relationships, URL1-1-7, Thomas *et al.* 2003）やPfam（URL1-1-8），KOG（EuKaryotic Orthologous Groups, URL1-1-9），「GO」（遺伝子オントロジー，gene ontology, URL1-1-10, Ashburner *et al.* 2000）といったデータベースへのリンクリストが表示され（図1-1-7），詳しい遺伝子機能情報を入手できる．

（2）遺伝子構造の閲覧と上流域塩基配列の取得

検索した遺伝子の遺伝子構造を閲覧するための方法を説明する．図1-1-7の中段付近にある「Sequence」タブを選択し，白抜きで現れる「Genomic sequence」ボタンを選択すると，口絵1-1-1のように，5′UTRが緑色に，CDSが青色に，3′UTRが桃色に，構造別に塗り分けられて表示される．また，この遺伝子の上流域3,000塩基を取得するには，口絵1-1-1の塩基配列の上部にある「upstream」の表示の右隣の入力欄に「3000」と記入して，右側の「Submit」ボタンを選択すると，図1-1-8のように上流域の3,000塩基の塩基配列が表示される．

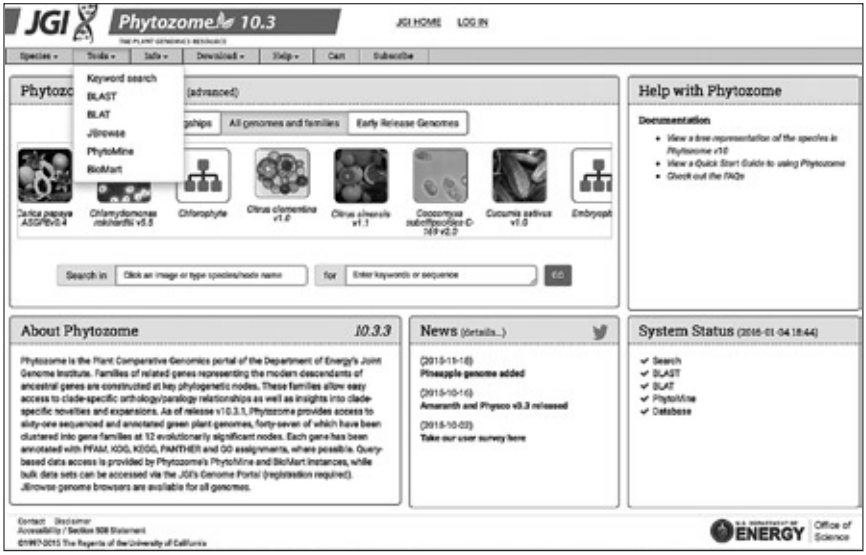

図1-1-3　Phytozomeのトップページ

Toolsメニューを開いた状態．Toolsメニューに「KeyWord search」,「BLAST」,「BLAT」などがある．「BLAT」は質問配列がゲノム配列中のどの部分にヒットするかゲノムランディング（genome landing）探索するのに用いる．

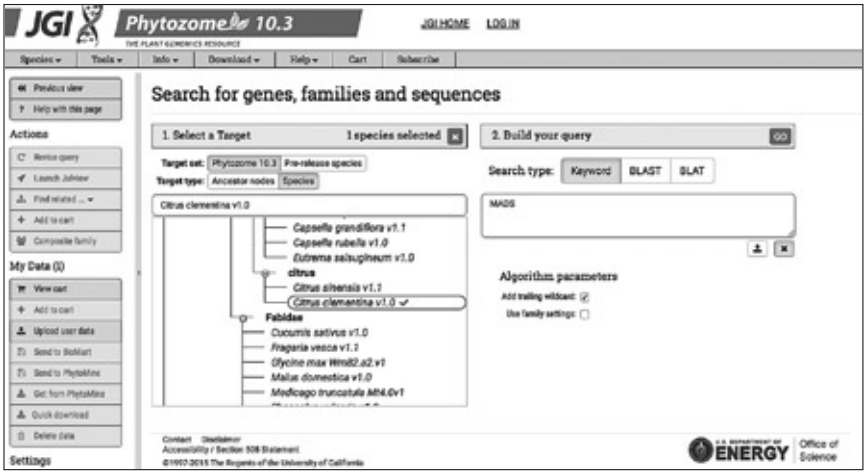

図1-1-4　Phytozomeにおけるキーワード検索画面

ここでは，*Citrus clementina*が選ばれ，「MADS」をキーワードとしている．

図1-1-5 PhytozomeにおけるキーワードMADS検索結果表示
アノテーションに「MADS」を含む遺伝子が表示される．

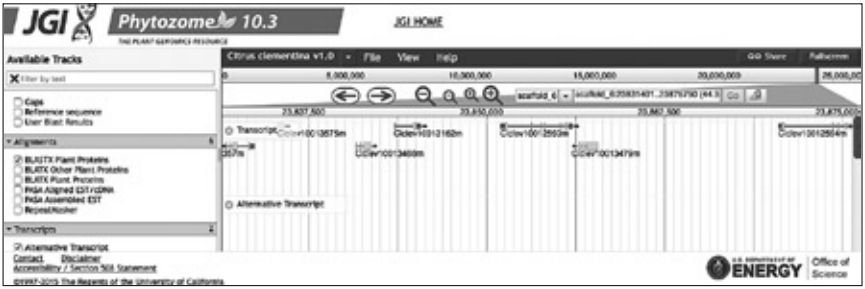

図1-1-6 Phyozomeのゲノムブラウザ表示
ゲノム上での遺伝子の位置関係がわかる．拡大縮小が可能である．

(3) ゲノム情報のダウンロード

ゲノムデータを取得するには，Phytozomeのアカウントを取得してログインする必要がある．ログインすると，検索結果を保存することもできる．一般に，ゲノムデータベースを使用する際は，アカウントを取得して使用すると利便性が高まる．ログイン後，図1-1-3のメニューバーの「Downlaod」を選択する．Phytozomeのバージョンを選択するプルダウンメニュー表示されるので，とくに理由がなければ最新のバージョンを選択する．すると，「Genome Portal」のページが表示される（図1-1-9）．種名が表示されているフォルダの中から，

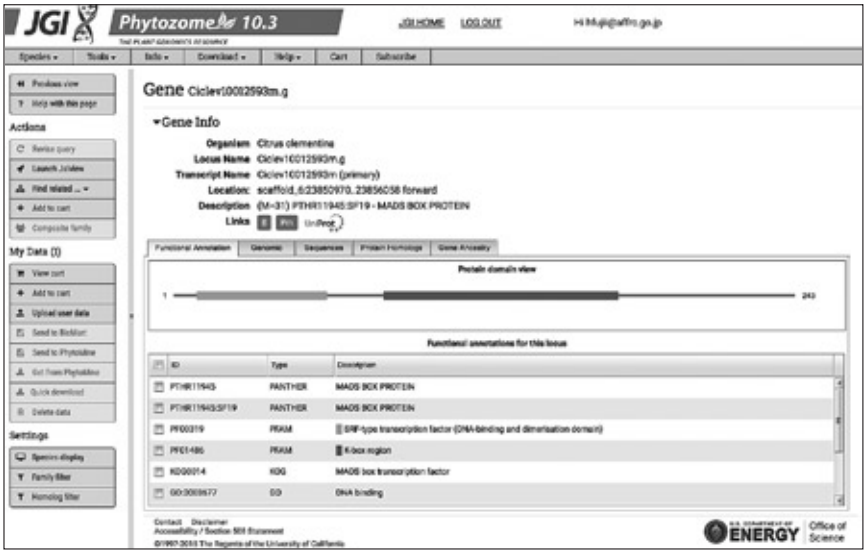

図1-1-7　Phytozomeの遺伝子アノテーション表示
画面下部に示されているようにPANTHERやPfam, GOのアノテーションを閲覧できる．

例えば「Cclementina」のフォルダアイコンを開くと，図1-1-10のようにダウンロード可能なファイルが表示される．ダウンロードしたいファイルにチェックを入れて緑色のボタン「Download Selected Files」（図1-1-9）を選択するとダウンロードが開始される．

図1-1-10に示されている「annotation」フォルダには，coding sequence（CDS）配列やタンパク質配列のファイルがマルチFASTA形式で納められており,「assembly」フォルダにはゲノム配列としてのスキャフォルド（scaffold）がマルチFASTA形式で納められている．ファイル名の拡張子のうち，faはFASTAファイルを示し，gzはUNIX互換OSでよく用いられるファイル圧縮ソフトGZIPによって圧縮されたファイルであることを示す．Windows環境でもLhaplusやLhacaなどのフリーソフトウエアを用いてgz形式で圧縮されたファイルを解凍することができる．また拡張子がgffで示されるファイルはgeneral feature format version3（GFF3）の意味で遺伝子構造情報が格納

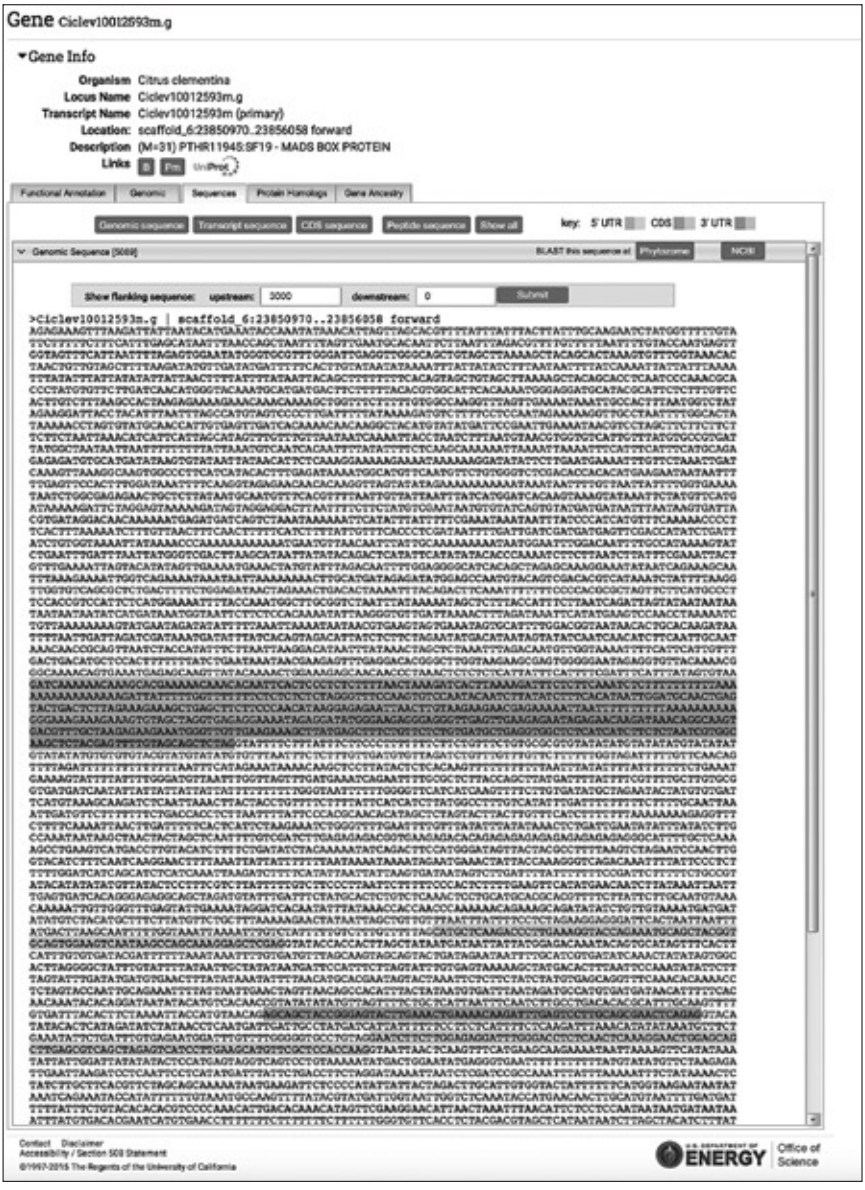

図1-1-8　Phytozomeの遺伝子上流域の表示
「upsteam」の欄に「30000」を入れると上流域の3,000塩基が表示される

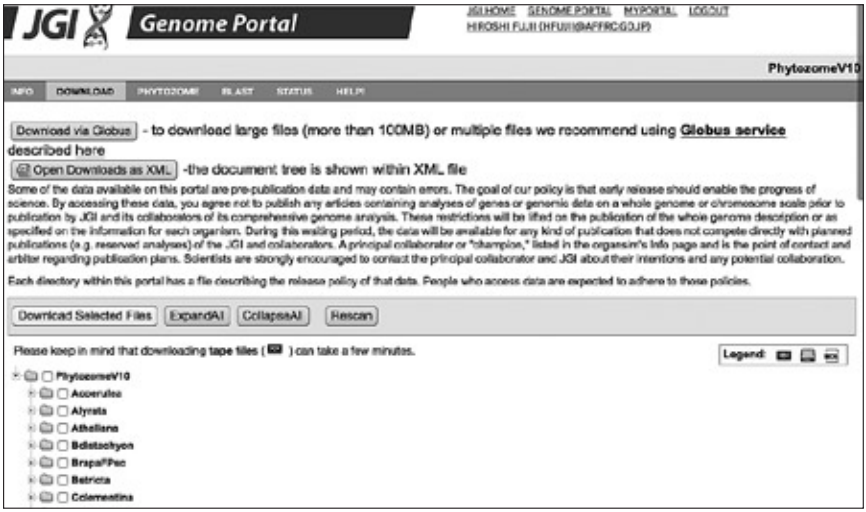

図1-1-9　Phytozomeにおけるゲノムデータのダウンロード画面
下部のフォルダーに生物種ごとのゲノムデータが格納されている

図1-1-10　Phytozomeのダウンロードファイルの例
クレメンティンのゲノムデータファイルを示している．

されていることを示す．図1-1-6に示したゲノムブラウザの画像は，GFF3ファイルのデータを基づいてグラフィカルに遺伝子構造を表示した結果である．テキストデータとしてGFF3ファイルを閲覧する方法については後述する．repeatmaskedなどmaskedが付されたファイルは，繰り返し配列がNでマスクされた配列であることを示す．繰り返し配列があるとアラインメントや類似性検索が正しく行えないことがある．それぞれの詳しいファイルの内容につい

ては，同じくダウンロードサイトにあるCclementina_182_v1.0. readme. txt やData_Release_Policy. htmに記述されている．

2）TAIR

研究対象とする遺伝子ホモログについて，ゲノム情報を取得する方法の一つとして，モデル植物のシロイヌナズナ（*Arabidopsis thaliana*）のゲノムアノテーションデータベースであるTAIR（The Arabidopsis Information Resource, URL1-1-11, Huala *et al.* 2001 ）の利用がある．TAIRでは，各遺伝子にGOをはじめ，発現情報や文献情報など豊富なアノテーションが付されているので，当該遺伝子に関する既往の情報を広範に得ることができる．Phytozomeの場合と同様に，研究対象とする遺伝子ホモログを網羅的に取得する場合，遺伝子名によるキーワード検索と配列の類似性によるBLAST検索の双方の検索を必ず行うべきである．

3）Taxonomy Browser

アメリカ国立衛生研究所（National Institutes of Health；NIH）の所属機関である国立生物工学情報センター（National Center for Biotechnology Information, NCBI）が管理・運営している国際塩基配列データベースの一つであるTaxonomy Browser（URL1-1-12）を利用すると，研究対象とする果樹の属や種に関して，NCBIの統合データベースEntrez（URL1-1-13）に登録されている塩基配列情報や遺伝子発現情報などを網羅的に取得することができる．例として，Taxonomy Browserの検索欄に「citrus」をキーワードにして検索すると，カンキツ属（*Citrus*）の種ごとに整理されたEntrezのポータルページへのリンクが表示される（図1-1-11）.「Citrus clementina」を選択すると，右端の表に*Citrus clementina*に関するEntrezのすべての登録情報の数が表示される（図1-1-12）．右端の表を見ると，この時点での*Citrus clementina*の「Nucleotide」の登録数が43,615であることがわかる．数字「43,615」に付さいるリンクを選択すると，「Nucleotide」のリストが表示される（図1-1-

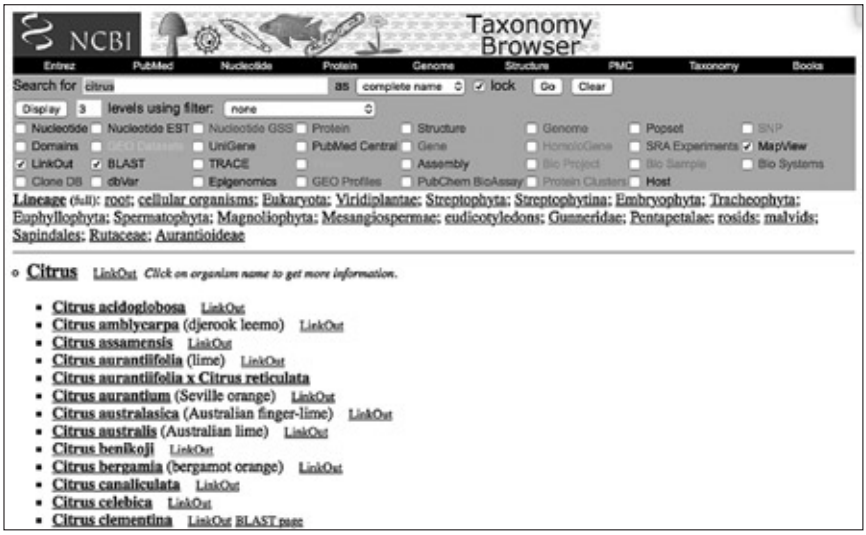

図1-1-11　NCBI Taxonomy Browserでの検索結果

キーワード「Citrus」で検索した．上部の「Citrus」をクリックすると，*Citrus*属全体のEntrezのデータを観ることが出来る．

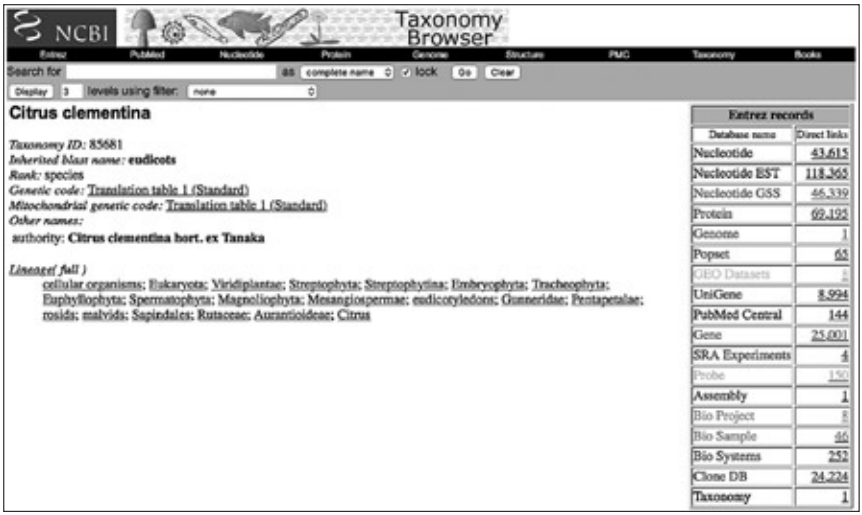

図1-1-12　NCBI Taxonomy Browserでの検索結果表示

右側の表にクレメンティンに関してEntrezに収録されている各種データの数が表示される．

13)．43,615の「Nucleotide」をFASTA形式のファイルでダウンロードするには，図1-1-13の右上方の「send to」を選択して現れたプルダウンメニューの「Choose Destination」で「File」を選択し，「Format」で「FASTA」を選択して，「Create File」をクリックすれると，43,615の「Nucleotide」の塩基配列がFASTA形式のファイルとしてユーザのPCにダウンロードされる．ダウンロードした塩基配列は，自身のPC内（ローカル, local）で実行するBLASTのデータベース作成用などに利用できる．ローカル環境でのBLASTの実行方法や，ダウンロードした塩基配列をBLASTのデータベースにする方法は後述する．

また，図1-1-14に示したように，左上方の「Summary」を選択すると「Format」を選択するプルダウンメニューが現れ，各遺伝子の画面表示のフォーマットをGenbank形式などに変更することもできる．

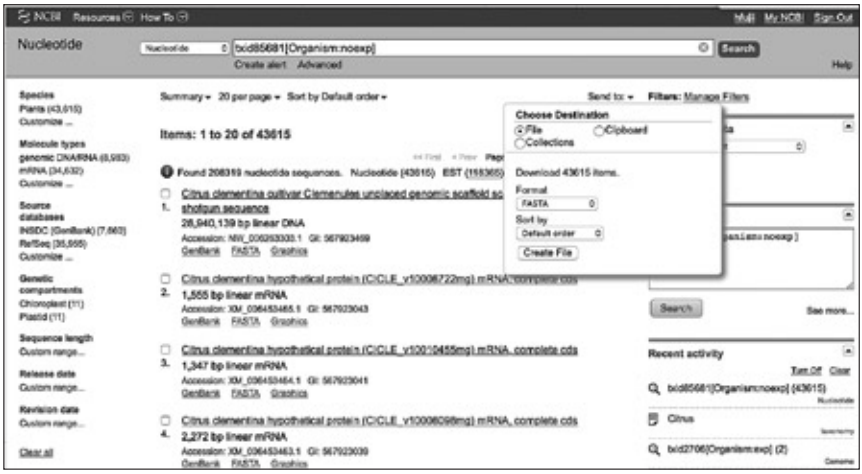

図1-1-13　NCBI Nucleotideの検索結果表示1
クレメンティンの43,615の登録NucleotideがSummary形式で表示されている．右上の「Send to」を選択するとダウンロードの方法とファイル形式のプルダウンメニューが表示される．

図1-1-14　図1-1-13NCBI Nucleotideの検索結果表示2
右上のSummaryを選択すると，表示形式をプルダウンメニューから選ぶことができる．

4）データ閲覧用のソフトウエア

次世代シークエンサのデータ解析に用いるような本格的なバイオインフォマティクスツールついては，第1章第2節で紹介する．本節では，上記で得たゲノムデータを利用するための汎用的ソフトウエアを紹介する．多くの場合，ゲノムデータのファイルサイズは大きいので，Microsoft社のWordやExcelでの閲覧は難しく，下記に紹介するようなソフトウエアが実用的である．

(1) テキストエディタ

Mery（URL1-1-14）やMKeditor for Widows（URL1-1-15），秀丸エディタ（URL1-1-16）といったテキストエディタの多くは，本来はコンピュータプログラムの記述を目的としたツールであるが，大量の塩基配列データのようなファイルサイズの大きいテキストデータを閲覧するのに適している．また，テキストエディタの多くは，正規表現による文字以外の制御記号の編集が可能である．例えば，タブ記号を改行記号に置換がすることができるので，ファイ

ル形式の変換にも役立つ．grepという機能を用いると特定の文字列を含んだ行をすべて調べることができるので，複数のマルチFASTAファイルにおける特定のアノテーションが含まれる行の書き出しが容易である．テキストエディタを選択する場合，ソフトウエア流通サイトのVector（URL1-1-17）を利用すると，多数のフリーソフトウエアやシェアウエアの中から好みのソフトウエアを探すことができる．巨大なテキストファイルを閲覧することができるGiga Text Viewer（URL1-1-18）や前述のファイルを圧縮・解凍するためのアーカイブ・ユーティリティなども紹介されている．

(2) データベースソフトウエア

前述のGFF3ファイルを閲覧する場合は，有償のデータベース作成ソフトウエアであるFileMaker Pro（FileMaker社, URL1-1-19）の利用が便利である．GFF3ファイルのサイズは非常に大きいことが多いが，GFF3ファイルをFileMakerのショートカットアイコンにドラッグ・アンド・ドロップするだけで，GFF3ファイルの内容が表示され，簡易なデータベースとして利用できる．同様に，（commma separated value, CSV）ファイルやExcelファイルもドラッグアンドドロップだけでFileMaker形式のファイルに変換できる．また，FileMakerから，CSVやExcel形式のファイルに出力することも容易である．簡易なリレーショナル・データベースを構築できるので，複数のデータセット間のリレーションを取って，任意のデータの組み合わせでの表示や出力が可能である．FileMakerは人気のあるソフトウエアなので，わかり易い参考書も複数出版されていて，自学も容易である．巻末の参考図書に，その一部を掲げた．

5）ローカルBLASTおよびファイル形式変換のためのソフトウエア

有償ソフトウエアのGENETYXパッケージ（株式会社ゼネティクス，URL1-1-20）やCLC Main Workbench（QIAGEN社，URL1-1-21）などを利用すると，自分用のBLASTデータベースを作成して，ローカル環境でBLASTを実行できる．例えば，自身が解読したゲノム塩基配列などをデータベース側とし

て，BLASTが実行できるようになる．また，上記3）のTaxonmy Browserからダウンロードした対象樹種の遺伝子のタンパク質配列をデータベースにすることで，ノイズの少ないBLASTの結果が得られる．

また，これらの塩基配列・タンパク質配列処理ソフトウエアはファイル形式の変換にも利用できる．バイオインフォマティクス分野では表1-1-2のような多くのファイル形式が使用されるが，CLC Main Workbenchは，これらのファイル形式の入出力ができるので，例えば，マルチFASTA形式の塩基配列を入力して，CSV形式で出力することも可能である．CLC Main Workbenchは第1章第2節で解説するCLC Genomics Workbenchから，次世代シークエンサデータのアセンブリ機能を除いたソフトウエアである．

6）もうすこし先のバイオインフォマティクス

バイオインフォマティクス手法を利用するには，Apple社のMacは比較的有用である．MacのOSであるMac OSXはUNIX系のOSであるので，ターミナルというアプリケーションを使うとUNIXコマンドを使用できる．UNIXコマンドを用いると，例えば，多数のファイルの結合や名前の付け替え，ファイルサイズの大きなテキストファイルの閲覧も容易である．Macによるバイオインフォマティクス解析の方法を解説した書籍を巻末の参考図書に掲げた．なお，Macに備えられているBoot Campというソフトウエア用いるとMac OSX環境とWindows環境を併存させることができるので，MacとしてもWindowsマシンとしても使用できる．また，VMware Fusion（VMware社，URL1-1-22）を使用すると，Mac上でWindowsを実行することもできる．

日常使用しているWindowsやMacでは困難な大規模計算を行う場合，一つの方法として，フリーの統計解析ソフトR（URL1-1-23）の利用がある．RにはBioconductor（URL1-1-24）というゲノムデータ解析用のパッケージがあり，プログラミングをする必要は少ない．Rは，研究機関や大学の共用の計算用サーバに導入されていることが多いので，こうしたサーバのアカウントを入手できれば，手許のPCから計算用サーバのRを操作して，サーバに大規模な計

表1-1-2 バイオインフォマティクスで用いられるファイル形式

ファイル名	拡張子	ファイルの種類
ACE	.ace	コンティグ
Phylip Alignment	.phy	アラインメント
GCG Alignment	.msf	アラインメント
Clustal Alignment	.aln	アラインメント
Newick	.nwk	ツリー記述
FASTA	.fsa/.fasta	配列
GenBank	.gbk/.gb/.gp	配列
GCG	.gcg	配列
PIR	.pir	配列
Staden	.sdn	配列
DNAstrider	.str/.strider	配列
Swiss-Prot	.swp	タンパク質配列
Lasergene	.pro	タンパク質配列
Lasergene	.seq	塩基配列
Embl	.embl	塩基配列
Nexus	.nxs/.nexus	配列, アラインメント, ツリー記述,他
CLC	.clc	配列, アラインメント, ツリー記述,他
Text	.txt	テキスト
CSV	.csv	カンマ区切りテキスト
ABI	.abi	塩基配列トレースファイル
AB1	.ab1	塩基配列トレースファイル
SCF2	.scf	塩基配列トレースファイル
SCF3	.scf	塩基配列トレースファイル
Phred	.phd	塩基配列トレースファイル
mmCIF	.cif	構造
PDB	.pdb	構造
BLAST Database	.phr/.nhr	BLASTデータベース
Vector NTI Database	–	配列
Vector NTI	.ma4/.pa4/.oa4	配列
Gene Construction Kit	.gcc	配列

算を行わせることは,それほど難しくない．とくに繰り返し計算を行う場合に,ループ機能が使えるので便利である．研究機関や大学によっては，グラフィカルなインターフェースで計算用サーバのRを使用する環境であるR studioを備えている．RはWindowsでもMacでもインストールできるので，ローカルな環境で利用することもできる．Rに関する解説書を巻末の参考図書に掲げた．このほか，外部のサーバでインフォマティクス解析を行う方法として，Galaxy

（URL1-1-25, Giardine, B. *et al.* 2005）やGalxy/NIAS（URL1-1-26），DDBJ Annotation Pipeline（URL1-1-27, Kaminuma, E. *et al.* 2010）などがある．これらについては，第1章第2節や第4節でも紹介する．

バイオインフォマティクス関連のツールやデータベースは，すべてではないが，バイオサイエンスデータベースセンター（National Bioscience Database Center, NBDC）のポータル（図1-1-15, URL1-1-28）にまとめられている．この中のゲノム解析ツールリンク集では，ツールが種類ごとに整理され，簡単ではあるが日本語の解説が付けられている．また，第1章第4節でも解説する統合TV（URL1-1-29，図1-1-16）は，バイオインフォマティクス関連のデータベースやツールの使い方を動画で解説していてわかりやすい．例えば，上記のTaxonomy Browserの使い方に関する動画もある．

バイオインフォマティクス関連のツールやデータベースは変遷の速度が早いので，最新の状況を知るには書籍よりも，講習会などに参加すると現状が把握しやすいし，講師に直接質問することで，自分にとって有益なツールやデータベースを効率的に知ることができる．定期的な講習会としては，DDBJが開催するDDBJing（URL1-1-30）やNBDCの統合データベース講習会（URL1-1-31）がある．また，農業分野としては，育種学会講演会で開催されるバイオインフォマティクス講習や農業生物資源研究所が主催する次世代シークエンサデータ処理のためのデータ解析実習があり，果樹研究者にとって非常に有益である．農研機構果樹研究所主催の果樹インフォマティクス・キャンプの隔年に1回程度開催される．

図1-1-15　バイオサイエンスデータベースセンター（NBDC）のトップページ
統合TVやゲノム解析ツールリンク集へのリンクがはられている.

図1-1-16　統合TVのトップページ

引用文献

Altschul, S. F. *et al.*（1997）Gapped BLAST and PSI-BLAST : a new generation of protein database search programs. Nucleic Acids Research. 25 : 3389-3402.

Ashburner, M. *et al.*（2000）Gene ontology: tool for the unification of biology. Nature Genetics. 25 : 25-29.

Chagné, D. *et al.*（2014）The Draft Genome Sequence of European Pear（*Pyrus communis* L. 'Bartlett'）. PloS one. 9 : e92644.

Giardine, B. *et al.*（2005）A Galaxy : a platform for interactive large-scale genome analysis. Genome Research. 15 : 1451-1455.

Huala, E. *et al.*（2001）The arabidopsis information resource（TAIR）: a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Research. 29 : 102-105.

Jaillon, O. *et al.*（2007）The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature. 449 : 463-467.

Kaminuma, E. *et al.*（2010）DDBJ launches a new archive database with analytical tools for next-generation sequence data. Nucleic Acids Research. 38 : D33-38.

Kent, W. J.（2002）BLAT-The BLAST-Like Alignment Tool. Genome Research. 12 : 656-664.

Neupane, R. *et al.*（2012）Phytozome : a comparative platform for green plant genomics. Nucleic Acids Research. 40（D1）: D1178-D1186.

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 : 487-494.

Thomas, D. P. *et al.*（2003）PANTHER : A Library of Protein Families and Subfamilies Indexed by Function. Genome Research. 13 : 2129-2141.

Velasco, R. *et al.*（2010）The genome of the domesticated apple（*Malus* × *domestica* Borkh.）. Nature Genetics. 42 : 833-839.

Wang, W. J. *et al.*（2013）The genome of the pear（*Pyrus bretschneideri* Rehd.）. Genome Research. 23 : 396-408.

Wu, G. A. *et al.*（2014）Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nature biotechnology. 32 : 656-662.

Xu, Q *et al.*（2013）The draft genome of sweet orange（*Citrus sinensis*）. Nature Genetics. 45 : 59-66.