

ГЕНЕТИКА И СЕМЕНОВОДСТВО ПОЛЕВЫХ КУЛЬТУР

Известия ТГУ, выпуск 1, 2000 год

УДК 631,523:575.18:681.3.016

СОЗДАНИЕ УПРОЩЕННОЙ КОМПЬЮТЕРНОЙ БАЗЫ ГЕНЕТИЧЕСКИХ ДАННЫХ

Л.В. СМИРЯЕВ, К.А. ЛЕБЕДЕВ, А.В. БОЛТУНОВА

(Кафедра генетики)

Изложен подход кумулятивного развития информационного компьютерного обеспечения научных исследований "на примере работы с генетической коллекцией. Этот подход основан на признании неоптимальности разработки новых банков данных в отличие от создания и встраивания минимального информационного «фильтра» — упрощенной базы данных по коллекции в давно существующие и постоянно пополняемые банки. В результате вместо прямого поиска фрагментов полезных сведений в банке данных возникает возможность автоматизированного получения сечении информации, сопряженной с данными об образцах коллекции. Создана упрощенная компьютерная база данных Mutant, содержащая минимальные сведения об образцах идентифицированной маркерной коллекции томата. Эта база данных в рамках единого банка данных объединена с доступной через m^tNet базой SolGenes, в которой с 1993 г. накапливаются генетические данные по некоторым видам культурных растений семейства пасленовых. Приведены примеры запросов, «фильтрующих» данные банка по загруженной информации о коллекции.

Объем информации, получаемой в процессе научных исследований, с каждым годом все более возрастает, а поиск и обобщение необходимых ретроспективных данных требуют все больше времени. Выход состоит в создании и использовании компьютерных

банкой данных. К настоящему времени созданы такие системы управления базами данных (СУБД), как MS Access, FoxPro, являющиеся фактически стандартными. Но создание на их основе новых банков данных, например, по генетике культурных растений,

требует длительного времени и больших затрат.

Так, на кафедре генетики МСХА и по ВНИИССОК ведется научная работа по генетике томата, причем в экспериментальных исследованиях широко используется идентифицированная маркерная коллекция образцов томата Mo, сформированная А. А. Жученко к Молдавии [1,2]. Коллекция состоит из образцов, несущих по одному или несколько мутантных морфофизиологических генов с известной локализацией большинства из них. Компьютерный банк данных по генетике томата, состыкованный со сведениями об образцах этой коллекции, мог бы значительно повысить эффективность экспериментальной работы и теоретических исследований. Однако в названных организациях нет возможностей для создания и поддержания собственного банка данных.

С другой стороны, многие годы за рубежом совершенствуется, пополняется и корректируется достаточное число специальных генетических банков данных по отдельным видам, семействам и культурам. В настоящее время в электронной сети InterNct представлено более 40 таких банков, из которых по растениям минимум 19 [8]. Однако не простой задачей является прямой поиск в этих банках фрагментов информации, полезной для конкретной исследовательской работы, например, с образцами коллекции томата. Для подобного целенаправленного использования накопленной информации достаточно состыковать сведения о новой коллекции с соответствующими

банком данных посредством организации упрощенного информационного «фильтра», в которой будут отражены минимально необходимые сведения об образцах коллекции. Подобный «фильтр» упростил получение из банка информационных сечений — конкретных сведений, сопряженных с новой коллекцией. Приемы реализации такого подхода изложены далее.

Был проведен поиск в InlerNet с целью сравнения программного обеспечения, позволяющего организовать подобный «фильтр», а также обзор существующих банков данных генетической направленности. Выяснилось, что почти все они организованы с помощью СУБД ACEDB, которая была разработана для хранения генетических данных по *C. elegans* в 1991 г. Эта СУБД бесплатная, свободно распространяема, соответствует задачам работы с образцами коллекции и позволяет наладить обмен информацией между различными банками данных.

Одна из баз данных (БД) SolGems объединяет различимо биологическую и прежде всего генетическую информацию по основным культурам семейства пасленовых (Paul et al., 1994). Она является частью программы исследований генома растений, которая проводится U.S. Department of Agriculture (USDA), и регулярно пополняется через InterNct. Руководит работой по ведению БД SolGenes доктор С.Тенкелей из Корнельского университета [5]. Используя ACEDB, структурированную организацию и содержимое этой БД, возможно при минимальных затра-

тах внести в банк данных основную, ограниченную информацию о новой коллекции томата Мо, которая и будет служить информационным «фильтром».

Сопоставление приемов внесения информации о новой коллекции

Организацию «фильтра» можно осуществить, пользуясь следующими приемами:

1) создать упрощенную БД, содержащую минимальные данные о коллекции и состыкуемую с БД SolGcnec;

2) создать специальный файл (АСЕ-файл) с этими данными для добавления в уже существующие разделы БД SolGcnec.

Можно также решить задачу, частично объединив оба приема.

Создание отдельной БД имеет свои плюсы и минусы. К плюсам следует отнести то, что, во-первых, эта БД является самодостаточной: ее можно использовать в СУБД ACEDB не только совместно с БД SolGcnec, получая из последней требуемые информационные сечения по новой коллекции, но и автономно — для решения узких задач. Во-вторых, минимальную БД можно расширять, не заботясь о полном соответствии структуре БД SolGcnec. Например, в случае с коллекцией томата можно внести информацию о наличии семенного материала по ее образцам, что не предусмотрено в структуре БД SolGcnec. К минусам - в случае, если предполагается сопоставить новую информацию с другими, ранее загруженными коллекциями, то информационный запрос составить, как правило, будет сложнее.

Причина в том, что и минимальной БД содержится не вся информации об образцах часть приходится получать через ссылки к БД SolGcnec.

Второй прием добавление специального нового файла в старом формате данных о всех образцах коллекций (в БД SolGcnec существует специальный раздел Gcrlnplasm) - имеет следующие недостатки. Во-первых, нельзя ввести непредусмотренные данные, не потеряв совместимость с БД SolGcnec; во-вторых, после состыковки с новой коллекцией для выделения ее образцов из всех требуется некоторое усложнение запросов по той же причине, что при реализации 1-го приема. В то же время с помощью специально написанной программы данные, приготовленные к Загрузке в минимальную БД, могут быть легко адаптированы для загрузки в раздел Gcrlnplasm БД SolGcnec. В нашей работе были исследованы оба приема, по для получения запросов в основном использовалась созданная отдельно БД Mutant, содержащая минимальные сведения о новой коллекции и объединенная в единый банк данных с БД SolGcnec на основе СУБД ACEDB.

Благодаря внесению информационного «филтра», во-первых, через сведения, загруженные в новую БД (идентифицированные мутации каждого образца коллекции, известные образцы-аналоги других коллекций т 'омата и т.п.), становится легко доступной вся сопряженная информация, накопленная в БД SolGcnec ранее. В частности, классифицированные и структурированные лите-

ратурные данные за многие годы. Во-вторых, появляется возможность сравнить разнообразно схожих по мутациям образцов из разных коллекций, легко выявятся уникальные образцы, возможные новые мутации коллекции томата, а также спорные данные. В-третьих, несложные информационные запросы в банк данных позволят сопоставлять сведения о новой коллекции с другими коллекциями и наметить оптимальные пути пополнения первой. В-четвертых, сами образцы новой коллекции становятся доступны для автоматизированного анализа, сопоставления. Появляется компьютерная основа решения таких частных задач, как оптимизации схем подбора пар этих образцов для скрепления и отбора с целью создания многомаркерных мутантов. В-пятых, новую БД легко сделать доступной через InterNet для всех пользователей БД SolGenes.

Аналогичная минимальная БД может быть встроена и в другие банки данных, предельно расширяя доступную информацию и упрощая доступ к ней. Такой подход при минимальных затратах создаст кумулятивный эффект автоматического привлечения всесторонних научных сведений для информационного обеспечения конкретных генетических исследований.

Материалы и методы

В качестве СУБД использовано программное обеспечение ACEDB for Microsoft Windows версии 4.6.b (Durbin, Thierrv-Mieg, 1991, 1999; Bruskiwicz, 1995, 1999). СУБД и база данных Sol-

Genes доступны через InterNet. [8,12]. В новую БД Mutant занесена минимальная информация об образцах идентифицированной маркерной коллекции томата, полученная из описания образцов [1]. На выполнение подобной работы требуется 2—3 дня. БД Mutant объединена с БД SolGenes, размещенной на IBM-compatible PC на базе Intel Pentium 133 под управлением операционной системы Windows NT Workstation 4.00.1381/SP4. Для составления информационных запросов использован язык запросов, реализованный в СУБД ACEDB: ACEDB Query Language.

Содержание банка данных SolGenes с встроенной БД Mutant

БД Mutant включает минимальные данные о каждом из 372 образцов идентифицированной маркерной коллекции томата Mo: название образца (содержит две буквы — «Mo» и номер — две или три цифры); название известного аналога из другой коллекции, если неизвестен видовой принадлежность образца с перечислением генов, которые он несет (достаточно общепринятых аббревиатур, т.к. подробные описания есть в SolGenes). Дополнительно внесены сведения о местонахождении образца коллекции (есть только в МСХА, во ВНИИССОК или в двух местах). Кроме того, для каждого образца, находящегося в МСХА, занесены данные об имеющемся объеме семенного материала, полученного по каждому из семи последних лет. Следует отметить, что данные БД Mutant о семенном материале

предоставляют расширение возможностей для поддержания коллекции по сравнению с исходной БД SolGenes.

Данные БД Mutant загружались с помощью СУБД ACEDB в соответствии со структурными элементами БД SolGenes. Благодаря этому сведения о мутантных генах новых образцов полностью сопоставлены с основными сведениями БД SolGenes, т.е. с данными о томатах и о других культурных растениях семейства пасленовых, полученными за многие годы в основных научных генетических центрах по этим культурам. В результате возникла возможность при составлении запросов в банк данных исходить из сведений по новой БД Mutant, отраженной в главном меню банка как отдельный раздел (рис. 1). Причем эти сведения не нужно прямо называть в запросе. Достаточно назвать образец (группу образцов, всю коллекцию), сведения по которому являются исходными для поиска. Через систему перекрестных ссылок, автоматически расширяющуюся при подстыковке БД Mutant, стали легко доступны все сопряженные с коллекцией данные остальных разделов банка. При этом остается возможность доступа к любой информации (в том числе к БД Mutant) исходя из остальных разделов главного меню SolGenes, кратко описанных далее.

2 point data. Раздел содержит описание парного расположения локусов на отдельных хромосомах различных видов. На основе подобных данных ранее были созданы интегрированные карты хромосом. Они связывают распо-

ложение комплементарных ДНК (зондов), геномных ДНК, RAPD- и RFLP- маркеров с классическими морфологическими маркерами. В настоящее время в разделе находится 304 описания.

Аstonum. Это раздел, в котором описаны 22 аббревиатуры, широко использующиеся в банке данных. (Например, расшифрованы типы маркеров из раздела **a 2 point data**).

Allele. В разделе описаны основные аллели, представленные в банке. Описание, кроме вида, аббревиатуры и назначения аллеля и локуса, содержит также шифры образцов коллекции, в которых он находится (в том числе и коллекция Mo), характер фенотипического проявления аллеля. Для большинства аллелей даны ссылки на литературные источники. В настоящее время в разделе находится 441 описание.

Author. Этот раздел содержит алфавитный указатель 11032 авторов статей, ссылки на которые представлены в банке.

Collection. В данном разделе кратко описаны 10 различных коллекций, в том числе Mo. В описание каждой коллекции входят: ее название, координаты и фамилия куратора и, как правило, краткая история и сведения о составе.

Colleague. В разделе представлены 1744 записи: имена специалистов и названия организации, где занимается генетикой пасленовых, и координаты, по которым с ними можно установить связь;

DNA. Нуклеотидную структуру ДНК определяют путем секвенирования. Сейчас для этого используются два метода: Сейгера и Максема-Гилберта [2].

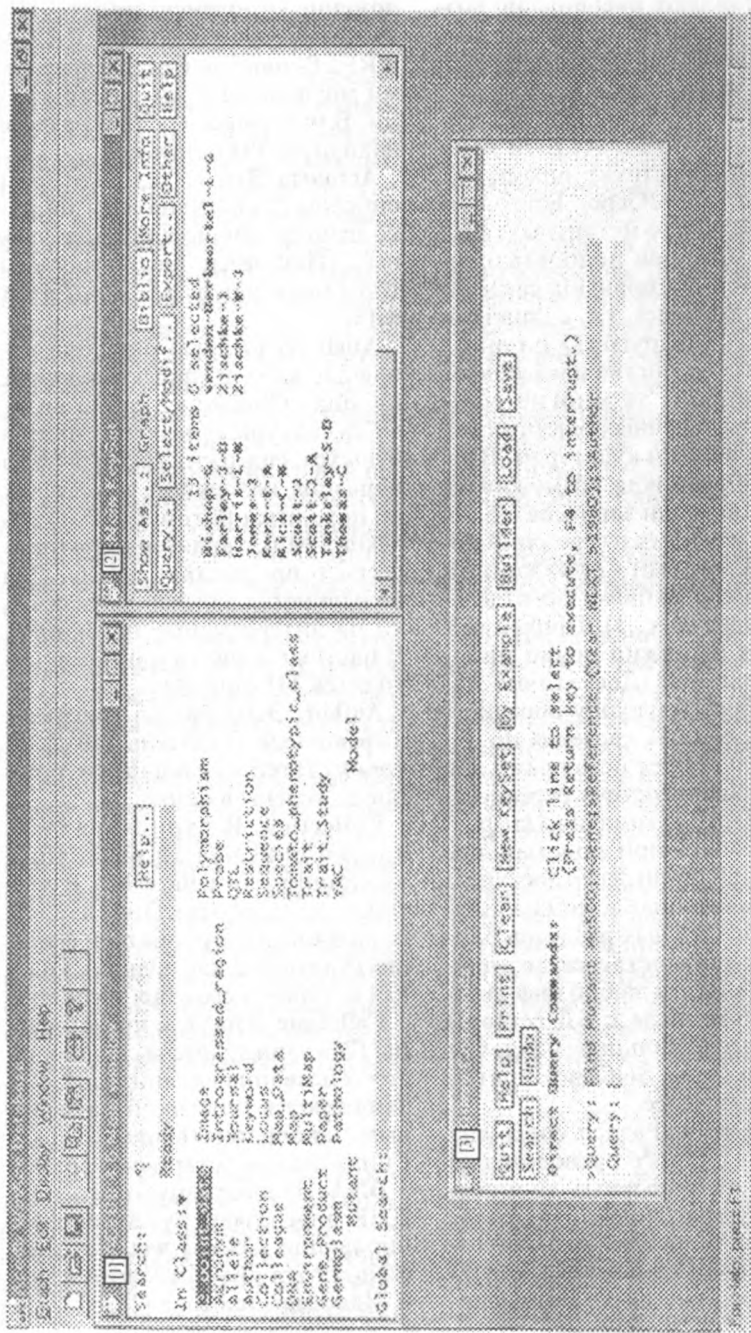


Рис. 1. Основные окна объединенного банка данных: [1] – окно, содержащее категории поиска (разделы банка данных); [2] – окно полученных результатов поиска любых запросов, [3] – окно запроса, которое раскрывается только при использовании запросов на языке Query Language.

Раздел содержит 4785 описаний секвенированных последовательностей ДНК, которые обозначаются шифрами, например, LEAC51A1. По каждой из них можно получить следующую ин-

формацию: последовательность нуклеотидов, местонахождение в геноме вида, иногда название гена и т.д. Информация носит документально графический характер (рис. 2).

[5]:LEAC51A1

Page A1

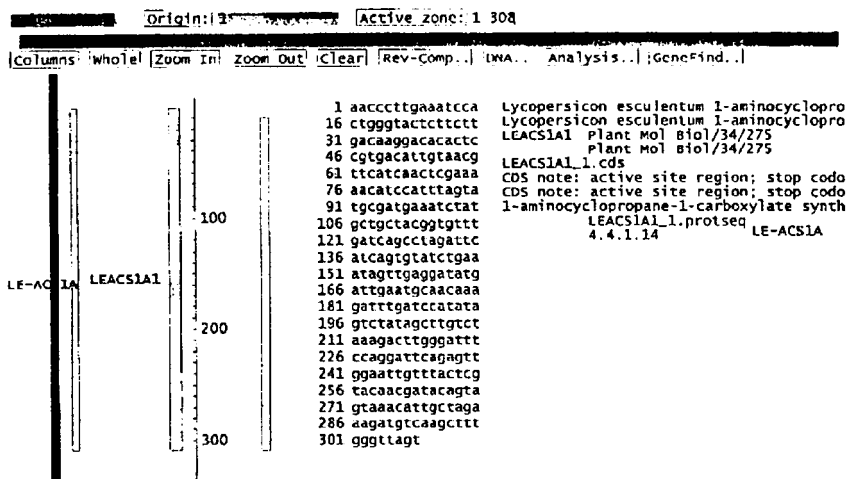


Рис. 2. Пример нуклеотидной последовательности LEAC51A1 с ее графическим изображением и пояснениями (Oetiker et al., 1997).

Environment. Данный раздел содержит описание экспериментов с различными образцами пасленовых, проведенных в 15 научных центрах. В описание входит материал и методика опыта, перечислены изучаемые признаки, имеется ссылка к библиографическим данным.

Gene Product. Представлено описание 1553 клеточных продуктов, которые кодируются участками ДНК. Каждый продукт определяется типом, например, т-РНК, м-РНК, р-РНК, или белок с указанием размера кодирующей области, кодирующей цепи и ело-

лесного описания продукта. Например, (1-) - beta - glucanase - генетический продукт, кодируемый геном Sp41a, которому соответствует нуклеотидная последовательность NTSR41AGN, и EC_number 3.2.1.39- европейский номер кодирования белка [11].

Gene plasm. Раздел содержит сведения о 9134 образцах различных культур семейства пасленовых, находящихся в 10 коллекциях, представленных в банке. Описание образца содержит в себе: название (шифр), ссылки к коллекциям, в которых представлен образец, в т.ч. на БД Mutant,

идентифицированные гены, вид, иногда — чем вызвана мутация и др. Некоторые сведения об образце могут быть опущены.

Image. В разделе находятся фотографии, схемы и рисунки, всего 322 изображения. Например, результаты электрофореза исследуемых последовательностей у образцов томата (рис. 3).

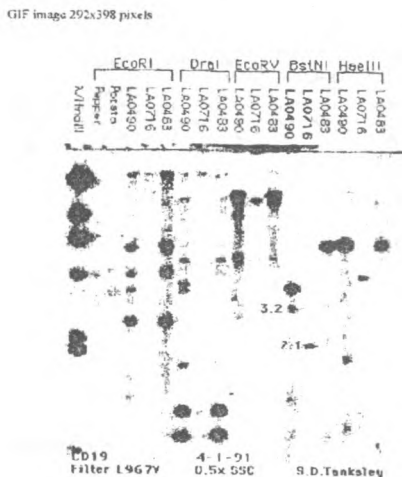


Рис. 3. Пример результатов электрофореза трех образцов томата, по одному — перца и картофеля с перечислением рестриктаз, обозначением зонда, фильтра, буфера (Tanksley, 1992).

Introgressed region. В разделе описаны 50 обобщенных карт отдельных хромосом томата с указанием на каждой нескольких локализаций интрогрессий — участков ДНК разных видов томата (см. **QTL**). Каждая интрогрессия маркирована молекулярными зондами, позволяющими ее идентифицировать.

Journal. В данном разделе представлена информация о 556 различных журналах, и которых опубликовались научные статьи, ссылки па которые даны и БД SolGenes.

Keyword. В разделе представлено 5400 ключевых слов и словосочетаний, которые выделялись в статьях, использованных при наполнении БД SolGenes.

Locus. В данном разделе представлены сведения о 6319 локусах, их типах (AFLP, RFLP, морфологический и др.), ссылки к графическому представлению локуса па хромосомной карте и т.п. Для морфологических описание фенотипического проявления аллелей, в некоторых случаях представлены фотографии или рисунки проявления (ссылки к раздел **Image**).

Map Data. В разделе представлены данные о 38 картированных видах пасленовых, локусах со ссылками к генетическим картам (раздел **Map**), скрещиваниях, в которых был изучен тот или иной ген, и другие сведения с указанием статей, где была опубликована эта информация.

Map. Заложена информация о генетических и физических картах с обозначением исследуемых локусов и расстояниями между ними. Генетические карты видов получают традиционным генетическим картированием. Оно состоит в определении положения гена на карте, которое базируется на оценке частоты рекомбинации между маркерными генами. Физические (молекулярные) карты — это схемы взаимного расположения последовательностей

ДНК или генов на хромосоме, полученные с помощью физических методов: рестрикционного, электронной микроскопии [4] и некоторых вариантов электрофореза для оценки межгенных расстояний в парах нуклеотидов (рис. 4). В частности, рестрикционные карты основаны на определении положения гена с помощью рестриктаз. Этот метод картирования состоит в получении фрагментов анализируемой последовательности, вырезанных рестриктазами и электрофоретически разогнанных с последующим сопоставлением их размеров и определением расстояний между последовательностями.

Представлена 291 карта по отдельным хромосомам различных культур. Например, генетическая карта 2 хромосомы томата с расположением маркеров и удобным выходом на дополнительную информацию по каждому локусу.

MultiMap. Содержится 40 схем — сопоставлений карт отдельных хромосом для двух и более видов пасленовых. На схемах результаты сопоставления представлены в виде документально-графического описания, гомологически сходные участки на генетических картах соединены линиями с обозначением локусов.

Paper. Раздел, в котором содержатся сведения о 7131 статье. В данный раздел занесены авторы статей, названия и другая информация. Особенно полезны для поиска ключевые слова, а также краткое содержание статей.

Pathology. В разделе перечислены некоторые бактерии, вирусы

и другие патогены. Данный раздел пока один из самых малообъемных: рассмотрено только 23 патогена. Однако это полочный раздел с точки зрения селекции и иммунитета, так как в нем содержатся сведения об устойчивых линиях и о генах устойчивости.

Polymorphism. Приведено 1111 описаний - характеристик генетического полиморфизма 3 сортов томата. Каждое описание получено методом электрофореза продуктов RFLP с использованием одной из пяти рестриктаз (см. Restriction) и одного зонда (см. Probe). Для сравнения каждый зонд, по сути одна рестриктаза EcoRI применены к одному образцу картофеля и одному сорту. Каждое описание имеет ссылку к изображению результатов электрофореза (пример на рис. 3), обобщающего результаты применения одного из зондов ко всем образцам. В качестве стандарта на изображениях используется результат электрофореза фрагмента ДНК фага, расщепленного рестриктазой Hind III.

Probe. Дана информация о 1703 зондах, которые используются для выявления определенных последовательностей ДНК или РНК. Зонды способны связываться с комплементарными цепочками нуклеотидов ДНК или РНК с образованием гибридных молекул [4]. Используя нужный зонд и рестриктазу, которая режет ДНК по определенному сегменту, например AAGCTT, с помощью электрофореза получают отображение исследуемого участка (см. рис. 3).

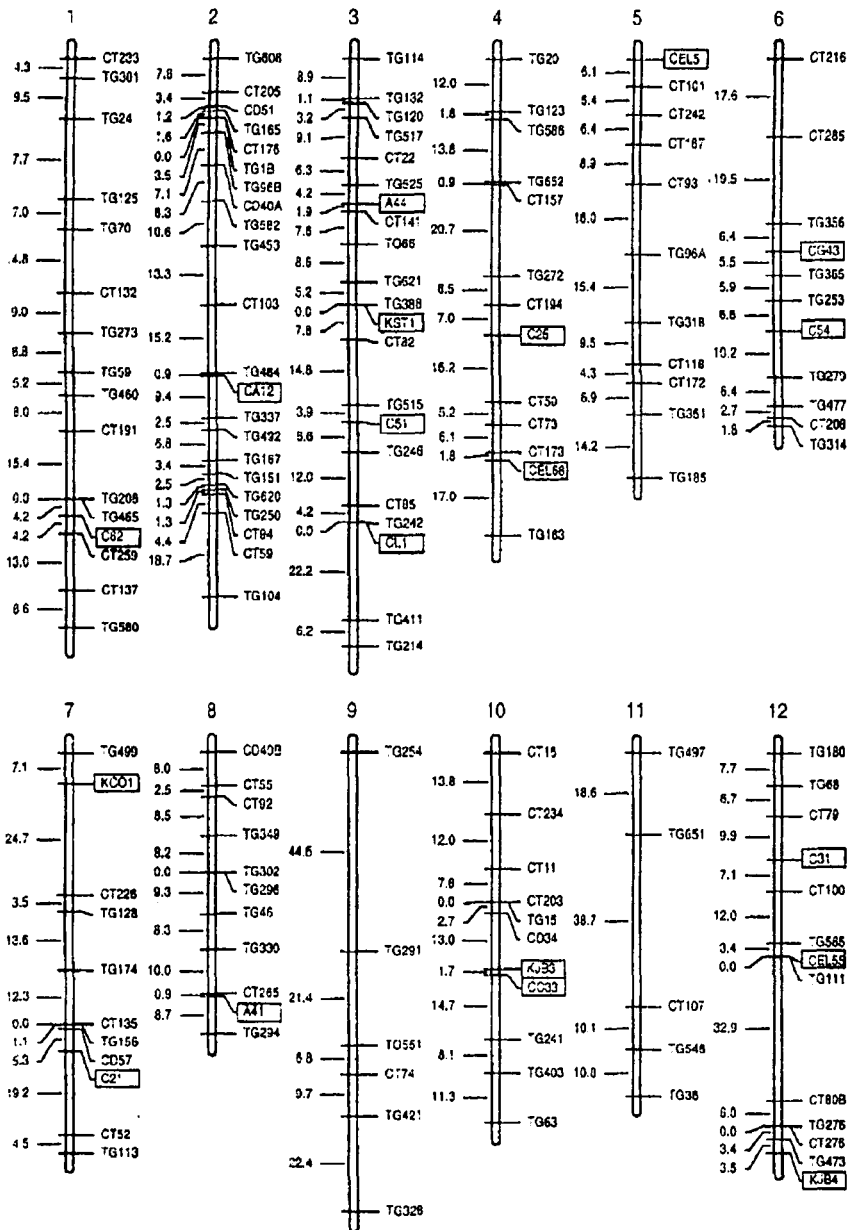


Рис. 4. Пример изображения физической карты хромосом с обозначением последовательностей и расстояниями между ними (Chen, Foolad, 1999).

QTL. Раздел содержит сведения о 603 локусах (точнее, участках хромосом), существенно влияющих на изменчивость количественных признаков (обычно оценка влияния ведется по результатам межвидового скрещивания). В частности, укачай вид, локализация и длина участка, сопутствующие маркеры, аддитивный эффект, контролируемый количественный признак, краткое описание методики опытов по локализации, ссылка на литературу.

Restriction. Категория содержит информацию о 341 рестриктазе: ёе шифр, сайт рестрикции, нуклеотидное основание место разреза. Рестриктаза (эндоуклеаза рестрикции) э- ■ то фермент, который узнает в молекулярной ДНК определенный сайт '----- нуклеотидную последовательность из 4, 5 или 6 остатков и разрезает (расщепляет) ДНК по ней. В результате рестриктаза расщепляет любую ДНК на сравнительно небольшое число строго определенных фрагментов [3], благодаря чему, в частности, возможна оценка генетического полиморфизма образцов (раздел **Polymorphism**).

Sequence. Каждая последовательность ДНК (см. рис.2) в этом разделе имеет дополнительную информацию. В частности, вид, положение последовательности в его геноме, способ ее выделения, карта у 'час' тка хромосомы, включающая данную последовательность, и более подробные ли театурные сведения о последовательности. В разделе приведена информация о 10919 последовательностях ДНК.

Species. В разделе находится краткая информация о 148 видах, отдельно о 76 пластидах, в частности, о хлоропластах томата, и об одной систематической конструкции. В разделе по большинству видов имеются прямые ссылки к данным нескольких разделов банка.

Tomato phenotypic class. В разделе приводится классификация фенотепических признаков тома 'та на 26 групп. Даны ссылки к генам, определяющими изменчивость признаков каждой группы.

Trait. Перечислены 70 признаков, доя каждого у казаны сопряженные с ним признаки, а ' также дано несколько ссылок к разделу '**Trait Study** с указанием видов, па которых признак изучался.

Trait Study. Приведены детали изу чения признака па конкретном виде: материал, у 'словия проведения и методики опыта, способы измерения (оценки) признака, методы обработки экепериментальных данных. Еми. ссылка к литературному источнику. ' Всего в разделе 128 описаний.

УАС — дрожжевые искусственные хромосомы используются для создания геномных библиотек различных видов. Кроме центромерных и теломерных областей дрожжевая хромосома содержит сайт, пригодный для клонирования чужеродного участка ДНК (кДНК). После встраивания последний сохраняется и удваивается один раз за клеточный цикл дрожжей.

Описание библиотеки дрожжевых искусственных хромосом содержит информацию о 27 образцах: размеры кДНК в kb (тысяч

пар нуклеотидов), источник кДНК — вид пасленовых, маркерные гены образца, методика выделения и клонирования, библиографические ссылки.

Model. Описание внутренней структуры базы данных, которое необходимо для работы с языком запросов. В разделе перечислены возможные категории, поля данными и их типы, наличие информационных ссылок.

Работа с банком данных

После запуска программы на экране открываются одно или два окна, нумерованные в левом верхнем углу: [1] ACEDB 4.5 Main window и [2] Main KeySet (см. рис. 1). Первое окно запросов, содержащее категории поиска (разделы банка данных), второе — окно полученных результатов поиска.

В программном обеспечении ACEDB существуют различные способы поиска информации. Наиболее простой состоит в прямом просмотре и поиске нужных единиц информации в открывающихся окнах и щелчке мышью на нужном объекте. При этом обычно раскрывается новое окно, в котором можно проводить дальнейший поиск объектов из различных разделов как они связаны ссылками. Различные варианты автоматического поиска с помощью запросов перечислены ниже.

1. Простой поиск (Search) среди элементов внутри выбранного раздела банка данных. Может быть использован для быстрого нахождения нужного элемента внутри раздела. Пример запроса,

задаваемого и окне [1] — все журналы, начинающиеся с Izv внутри заранее выделенного раздела Journal: Izv*. Ответ в новом окне [2]: Izv Timiriazevsk S kh Akad — «Известия ТСХА».

2. Глобальный поиск (Global search) среди некоторых полей внутри всех разделов банка. Полечен, например, для поиска по ключевым словам в статьях и для общего обзора информации по какой-то теме. Пример запроса по сочетанию слов, задаваемых в окне [1]: genetic transformation. Ответ в новом окне [2]: 289 статей, содержащих это сочетание ключевых слов.

3. Поиск с помощью 3 упрощенных вариантов языка запросов ACEDB Query Language: Query by example, Query builder и Table maker. Последний, в частности, позволяет получать результаты запросов в виде таблиц.

4. Поиск с помощью языка запросов ACEDB Query Language. Такие запросы набирают в окне [3] (см. рис.1). Далее приведены примеры с использованием этого языка, в которых БД Mutant используется как информационный «фильтр».

А. Найти авторов, имеющих публикации после 1980 г., связанные с генами, содержащимися в образце Mo500.

Find Mutant Mo500; >Gene > Reference (Year NXFXT > 1980); >.Author.

В результате выполнения запроса в окне [2] (см. рис. 1) получается список из 13 авторов. Можно продолжить поиск в окне с помощью более простых методов, перечне-

лепных в начале раздела, например, выйти на статьи этих авторов.

Б. Найти уникальные образцы из коллекции Мо (т.е. не имеющие аналогов в SolGenes), в которых есть гены 'sf и 'bls':

Find Mutant Gene = sf AND HERE - his.

В результате выполнения запроса в окне [2] получаем образцы Мо505 и Мо626.

В. Найти образцы Мо, имеющие аналоги в других коллекциях БД SolGenes:

Find Mutant General_collection.

Получаем список из 280 образцов.

Г. Найти уникальные образцы коллекции Мо, содержащие хлорофильные мутации:

Find Tomato phenotypic_class IS «Chi»; >Gene; >Mutant.*

В результате из 372 образцов Мо выделен список, в котором содержится 40 образцов.

Д. Аналогичный запрос по образцам Мо, имеющим аналоги:

Find Tomato phenotypic_class IS «Chi»; >Gene; >Germplasm; > > Mutant|coUeclion.*

В данном случае - 108 образцов.

Е. Перечислить гены, содержащиеся в образцах Мо, которые находятся в ТСХА, имеют аналоги в других коллекциях и по каждому из которых есть более 50 семян, полученных после 1996 г.:

Find Mutant (Seeds AND NEXT > >1996AND NEXT >50) AND Collection and ID = «MUTANT. TSHA »; > >General_collection: >Gene.

В результате выполнения запроса получаем список из 15 генов.

Ж. Какие гены содержатся в мутантах Мо, не имеющих аналогов и содержащих ген sf:

Hind Mutant Gene sf; Gene.

Получаем список из 7 генов: alb, bis, Ln, inua, m, sl'. sy.

З. Найти статьи, где упоминаются гены, названия которых начинаются на букву a и содержатся в образцах Мо, не имеющих аналогов:

Hind Mutant; >Gene a; >Reference.*

Получаем список из 168 статей.

Следует подчеркнуть, что без информационного «фильтра» (БД Mutant) подобные запросы, сопряженные с коллекцией Мо, либо вообще нельзя составить, либо поиск подобной информации в БД SolGenes каждый раз требовал бы подготовительной работы со сведениями о коллекции Мо. Причем для большинства запросов такая работа весьма сложна и продолжительна.

ЛИТЕРАТУРА

1. Бочарникова П.П., Козлова В.М. Мутантные формы томата». Кишинев, Штиница, 1992. 59 с. 2. Иванова С.В., Кучковская К.В. Биологические особенности гаплоидных растений томата. — Изв. ТСХА, 1999, вып. 3, с. 58-71. — 3. Мальшев С. В., Картель Н.А. Молекулярные маркеры в генетическом картировании растений. — Молекулярная биология, 1997, т.31, N2, с. 197 - 208. — 4. Миронов А.А., Миронов В.А., Комиссарчик Я.Ю. Методы электронной микроскопии в биологии и медицине. М.: Медицина, 1994. — 5. Beer S. TGC Report, 1977, vol. 47, p. 5. — 6. Chen F.Q., Foolad M.R. Genome, 1999, vol. 42, N 1, p. 94 — 103. 7. Databases at the National Agricultural Lib-

- raru InterNet, <http://probe.nalusda.gov:80000/all dbs.html> . — 8. Durbin R., Thierry-Mieg J., 1991. A C. elegans database. InterNet, <ftp://ftp.sanger.ac.uk/nub/acedb> . — 9. **Oeliker J.H., Olson D.C., Shin O. Y., Yang S.F.** — Plant Mol. Biol., 1997, vol.34, N 2, p. 275—286. —10. **Paul E., Goto M., Tanksley S.D.** — Euphytica, 1994, vol. 79, p. 181—186. - -
11. **Sesxa G., Fluhr R.** — Plant Mol. Biol, 1995, vol. 29, p. 969 - 982. —
12. SolGcnes — a Solanaceae Genome Database. InterNet, <http://nrobc.nalusda.gov:80000/plant/aboutso lnenc s.lilml> . — 12. **Tanksley S. D., Ganai M. I P r i n c e J.P. and e.l** — Genetics, 1992, vol.132, N6, p. 1141 - 1160.

*Статья поступила 18 октября
1999 г.*

SUMMARY

Approach of cumulative development of informational software of scientific investigations is described, the work with genetic collection taken as an illustration. This approach is based on recognizing non-optimum to develop new data banks in contrast to setting up and building in minimum information filter-simplified data base 011 the collection in existing for a long time and constantly filled up banks. As a result instead of direct searching fragments of useful information in data bank it becomes possible to automatically obtain sections of information connected with data on samples of the collection. A simplified computer data base Mutant has been developed which contains minimum information about the samples of identified marker collection of tomato and within common data bank it was united with SolGenes base, accessible through InterNet, where genetical data on some species of cultivated plants of nightshade family have been accumulated since 1993. The examples of requests-filtrating the bank data on the loaded information about the collection are presented.