

Подсистема «Трансгенез»: планирование генно-инженерных экспериментов на растениях с целью создания организмов с качественно новыми или улучшенными свойствами

Структура документа (оглавление)

1. Цель и задачи подсистемы «Трансгенез».....	1
2. Использование методов и подходов биоинформатики в генной инженерии растений: структура подсистемы «Трансгенез» и детальное руководство по ее применению.....	2
2.1. Информационные компоненты подсистемы «Трансгенез».....	2
2.1.1.1. Графический редактор генных сетей GenEd.....	4
2.1.1.2. Интерфейс редактора.....	4
2.1.1.3. Отображение объектов генной сети.....	5
2.1.1.5. Поиск компонентов в базе данных с помощью редактора.....	6
2.1.1.6. Ввод информации в базу данных с помощью редактора.....	6
2.1.1.7. Фильтры данных.....	7
2.1.2.1. Описание работы с БД TGP.....	11
2.1.3.1. Описание работы с БД TRSIG.....	18
2.2. Программные компоненты подсистемы «Трансгенез».....	24
2.2.2.1. Предсказание эффективности сигналов инициации и терминации трансляции (InTerOpt).....	28
2.2.2.2. Предсказание эффективности трансляции мРНК трансгена на основе полного анализа контекстных характеристик (Full_mRNA).....	31
2.2.2.3. Предсказание трансляционной активности 5'-НТП мРНК трансгена на основе анализа структурных характеристик (Leader_hairpin).....	32
2.2.2.4. Предсказание синонимических кодонов, повышающих эффективность элонгации в клетках растения-реципиента.....	35
2.2.3.1. Оптимизации кодонного состава трансгена.....	38
2.2.3.2. Элиминация потенциальных сайтов сплайсинга и полиаденилирования.....	42
3. Полезные ссылки.....	45

1. Цель и задачи подсистемы «Трансгенез»

Генная инженерия растений

Генная инженерия является одной из наиболее активно развивающихся областей современной биотехнологии. Генетически модифицированные организмы (ГМО), не так давно вошедшие в практику, широко применяются в сельском хозяйстве и их удельный вес в производстве сои, хлопчатника и многих других культур постоянно увеличивается. В настоящее время культивируются генетически модифицированные растения, несущие один или два трансгена, и их преимущества уже достаточно очевидны. По-видимому, перенос больших комплексов генов не только расширит возможности улучшения существующих сортов, но и позволит изменять характеристики растений качественно – вплоть до получения новых форм с заранее определенными свойствами. Кроме того, исследование экспрессии генов растений с использованием в качестве лабораторных моделей трансгенных растений – один из наиболее широко применяемых методов современной молекулярной генетики. Таким образом, трансгенез является не только одним из основных подходов современной биотехнологии, но и типовым методом лабораторных исследований, использование которого совершенно необходимо для проведения успешных исследований во многих областях фундаментальной науки.

Проблемы генной инженерии

В настоящее время существует ряд нерешенных проблем, затрудняющих получение ГМО. Для эффективного трансгенеза необходимо обеспечивать адекватное управление экспрессией перенесенного гена. Чужеродные гены часто содержат комбинации нуклеотидов, которые могут распознаваться в клетках организма-реципиента как сигналы экспрессии. Даже при ограниченном использовании чужеродной ДНК – когда переносится только белок-кодирующая часть – в ней могут содержаться различные ложные сигналы, например, сайты полиаденилирования (Diehn et al., 1998; Grec et al., 2000), сайты сплайсинга (Haseloff et al., 1997; Rouwendal et al., 1997) или неоптимальные (редкие) кодоны (van Hoof and Green, 1997; Komar et al., 1999) и т.д. Для того чтобы экспрессия трансгена осуществлялась по заданному паттерну, необходимо уметь выявлять негативные элементы и вносить в нуклеотидную последовательность соответствующие модификации (Larrick and Thomas, 2001). Это особенно существенно в тех случаях, когда планируется перенести несколько трансгенов, паттерны экспрессии которых должны быть точно сопряжены, или экспрессия трансгенов должна коррелировать с определенными метаболическими состояниями клеток или фазами роста (Gallie, 1998; DellaPenna, 2001). Очевидно, что здесь возможности генной инженерии принципиально ограничены уровнем знаний в молекулярной генетике и молекулярной биологии, поскольку охарактеризована только некоторая часть сигналов экспрессии. Однако, использование методов и ресурсов биоинформатики способно существенно увеличить возможности планирования генно-инженерных экспериментов и – в определенных случаях – отчасти заменить отсутствие точных знаний проверяемыми гипотезами.

2. Использование методов и подходов биоинформатики в генной инженерии растений: структура подсистемы «Трансгенез» и детальное руководство по ее применению

Подсистема «Трансгенез» представляет собой информационно-программный комплекс, предназначенный для планирования генно-инженерного эксперимента за счет решения следующих задач:

- Подбор сигналов экспрессии, обеспечивающий адекватный паттерн экспрессии трансгена в растениях. Планирование генно-инженерных модификаций на основе информации о генных и метаболических сетях растений (Дизайн ГМО).
- Предсказание эффективности экспрессии нативного варианта трансгена в растениях
- Модификация структуры трансгена с целью повышения эффективности его экспрессии в растениях.

Ниже будут рассмотрены компоненты подсистемы «Трансгенез» и приведено детальное (пошаговое) описание использования этих компонентов.

2.1. Информационные компоненты подсистемы «Трансгенез»

Дизайн генно-инженерных модификаций

Генно-инженерные модификации (ГМО) могут быть использованы для решения широкого круга биотехнологических и селекционно-генетических задач. По-видимому, возможности изменения самых разных морфофизиологических или биохимических характеристик растений и животных ограничены только уровнем наших знаний о межгенных взаимодействиях и генных сетях. Планирование генно-инженерного эксперимента должно в

первую очередь заключаться в построении схемы модификаций генома, которые должны привести к желаемому эффекту на уровне фенотипа. Это сложная задача, для решения которой необходимо проанализировать большой объем информации в различных областях биологии. Фактически, необходимо научиться определять, какие генно-инженерные модификации – будь то перенос трансгенов, инактивация или модуляция паттерна экспрессии генов растения-реципиента или комбинации различных воздействий приведут к желаемому результату.

Одним из перспективных подходов, позволяющих решать проблему планирования генно-инженерного эксперимента, являются базы данных генных сетей, сигнальных и метаболических путей (Schacherer et al., 2001; Kanehisa et al., 2002; Mueller et al., 2003), в которых накапливаются и систематизируются сведения о регуляции экспрессии генов, путей передачи сигналов и об участии различных ферментов и субстанций в процессе синтеза того или иного вещества в организме. Генная сеть - это группа координировано функционирующих генов, обеспечивающих формирование определенного фенотипического признака организма (молекулярного, биохимического, физиологического, морфологического, поведенческого и т. д.). Очевидно, что БД генных сетей могут быть использованы в качестве источника базовой информации при планировании ГМО, поскольку содержат данные, необходимые для выбора направления генетической модификации. Огромные возможности представляет экспериментаторам изменение паттернов экспрессии генов организма-реципиента с помощью механизмов генетического сайленсинга или косупрессии. Перенос чужеродных генов, кодирующих отсутствующие у организма ферменты, позволяет модифицировать метаболические цепи, изменять спектр и соотношение липидов, углеводов, различных вторичных метаболитов. Таким образом, БД генных сетей – по мере их появления и развития – будут использоваться в качестве важного компонента информационного обеспечения генно-инженерного эксперимента, поскольку позволяют эффективно осуществлять выбор генов-мишеней организма-хозяина для модификации экспрессии или трансгенов, которые должны быть перенесены в геном модифицируемого организма.

В составе подсистемы «Трансгенез» содержится генная сеть пролина - одного из ключевых метаболитов растений, отвечающих за устойчивость к широкому спектру абиотических стрессов. В составе генной сети приведены гены, кодирующие ферменты синтеза и деградации пролина, а также обозначены регуляторные контуры и лимитирующие звенья. Специалист в области генной инженерии растений может использовать эту информацию для планирования эксперимента, направленного на модификацию метаболизма пролина и стрессоустойчивости. Например, увеличение содержания пролина может быть следствием как усиления синтеза (с помощью переноса в геном растений дополнительной копии гена пирролин-5-карбоксилатсинтазы или транскрипционного фактора, усиливающего работу этого фермента); или же аналогичный метаболический эффект будет иметь снижение активности фермента, обеспечивающего деградацию пролина (с помощью переноса антисмыслового или dsRNA-супрессора гена пролиндегидрогеназы). Кроме этого, комплексные (мультикомпонентные) генные сети позволяют визуализировать и принимать в расчет взаимодействие генов и взаимовлияние метаболических путей, что необходимо учитывать в случае генно-инженерных модификаций.

2.1.1. База данных генных сетей растений PlantGeneNet.¹

Для создания трансгенных растений с заданными свойствами, необходимо выявить ключевые гены, модификация экспрессии которых приведет к желаемому результату. Для решения этой задачи разработана база данных PlantGeneNet – содержащая информацию о

¹ Подробнее о работе с базой данных PlantGeneNet смотри в Руководстве пользователя АСНИ-01, раздел 4.2.7

генных сетях растений, включающих биотехнологически-значимые донорные гены — объекты трансгеноза (на примере генной сети метаболизма пролина). Из литературы известно, что повышение концентрации осмопротектантов, в частности пролина у растений, приводит к адаптации растений к условиям засухи и засоления.

2.1.1.1. Графический редактор генных сетей GenEd

Графический редактор GenEd (Loktev et al., 2002) предназначен для создания и редактирования графических представлений генных и метаболических сетей, а также ввода и изменения информации в базе данных. С помощью этой программы пользователи могут вводить данные в GeneNet, оперируя графическими объектами и естественными понятиями молекулярной биологии, связанными с экспрессией генов и их регуляцией, ферментативными реакциями и т.д. Редактор осуществляет автоматическую трансляцию введенной информации в XML формат и в таком виде передает ее на сервер для записи в базу данных (Ananko et al., 2005).

Редактор GenEd позволяет получить общее представление о структуре сети и функциональных взаимоотношениях внутри нее, изучать строение двумерных графов генных сетей, выделять ключевые регуляторные компоненты, проводить поиск замкнутых регуляторных циклов и путей, связывающих два заданных объекта.

Редактор GenEd устанавливается локально на компьютере пользователя. Существует две разновидности редактора: для работы первой необходим сетевой доступ к базе данных и серверу GeneNet. Другая разновидность работает полностью автономно с локальной базой данных, устанавливаемой одновременно с редактором.

2.1.1.2. Интерфейс редактора

Сетевая версия редактора требует идентификации пользователя. Для локальной версии такая идентификация не обязательна.

Прежде, чем попасть в рабочее окно редактора, пользователь должен выбрать базу данных (эукариотическую или прокариотическую), а затем генную сеть, с которой он будет работать (из списка), либо создать новую сеть.

Окно редактора содержит панель меню, рабочее поле, в котором отображается схема генной сети, и "деревя" иерархии элементов схемы (Рис. 1). Атрибуты объектов отображаются и редактируются в отдельном окне, которое может быть вызвано или закрыто по желанию пользователя.

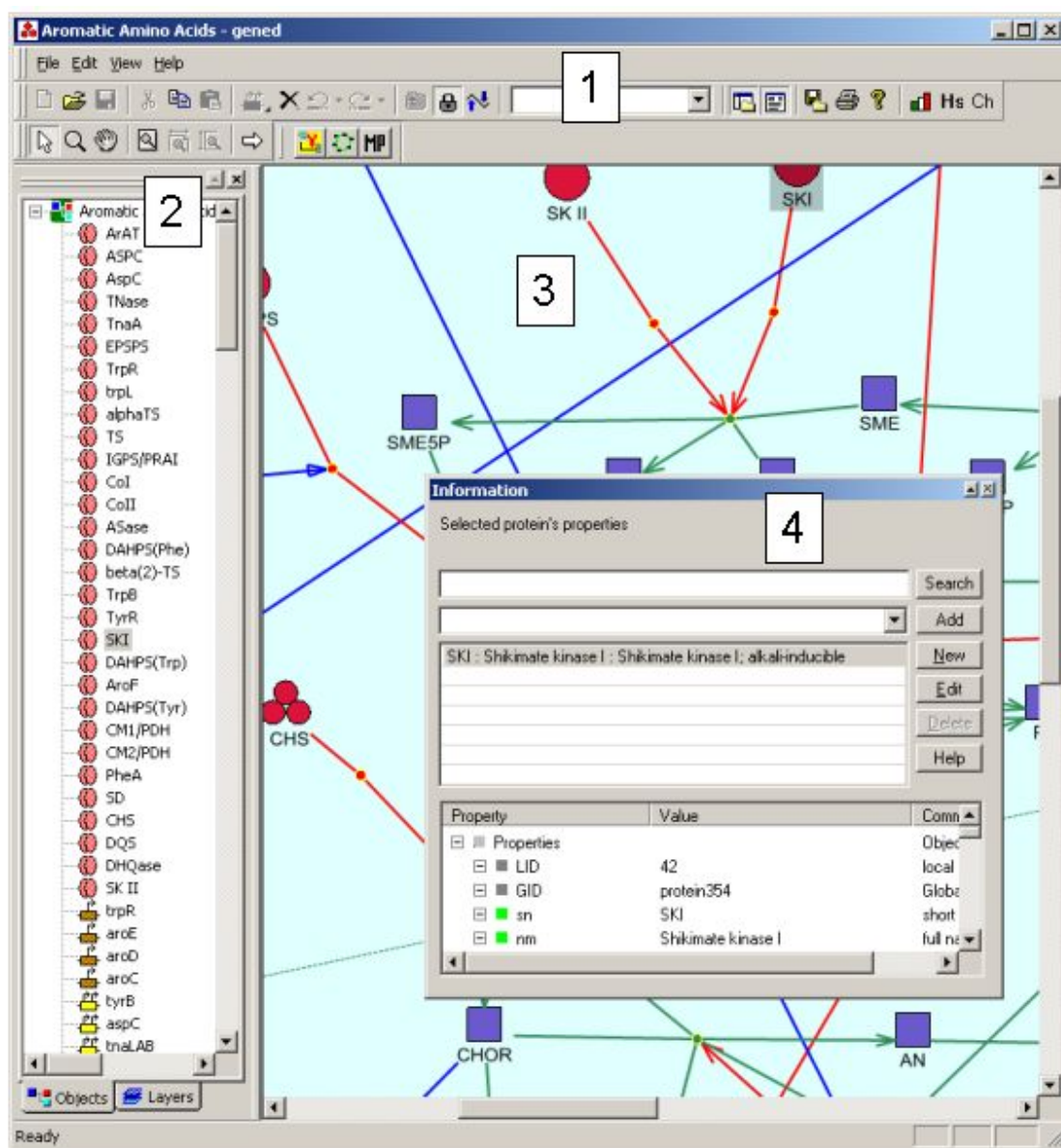


Рисунок 1. Окно редактора GenEd: 1 - панель меню; 2 - "дерево" иерархии элементов схемы; 3 - рабочее поле со схемой сети; 4 - отдельное окно отображения атрибутов объекта.

2.1.1.3. Отображение объектов генной сети

Каждый компонент генной сети имеет свое изображение, отражающее его особенности. Например, форма изображения белка отражает степень его мультимеризации, цвет – функциональное состояние. Мономерность белка, либо отсутствие данных о его мультимерном состоянии отображается на схеме в виде одиночного кружочка. Неактивные белки на схемах всегда зеленого цвета, активные - красного. Отображение объекта строится программой на основе значений некоторых атрибутов этого объекта. При изменении соответствующего атрибута изменяется и изображение на экране. Подробное описание отображений различных типов объектов и взаимодействий в сети, а также способы отображения некоторых атрибутов объектов даны в Руководстве пользователя АСНИ-01, раздел 4.2.7.2.

2.1.1.5. Поиск компонентов в базе данных с помощью редактора

Для того чтобы найти в базе данных нужный объект, необходимо кликнуть мышью по рабочей области схемы и выбрать тип объекта с помощью соответствующей кнопки панели меню. Откроется окно для поиска и выбора объекта (Рис. 2). Сверху в окне написан класс объектов, в котором проводится поиск (protein в данном примере). Ниже область ввода термина для поиска по базе данных. Следует отметить, что поиск проводится по всем названиям объектов данного типа, в том числе и синонимам, которые внесены в базу данных. Найденные объекты отображаются в раскрывающемся списке ниже строки поиска. Атрибуты выбранной записи базы данных отображаются в нижней части окна.

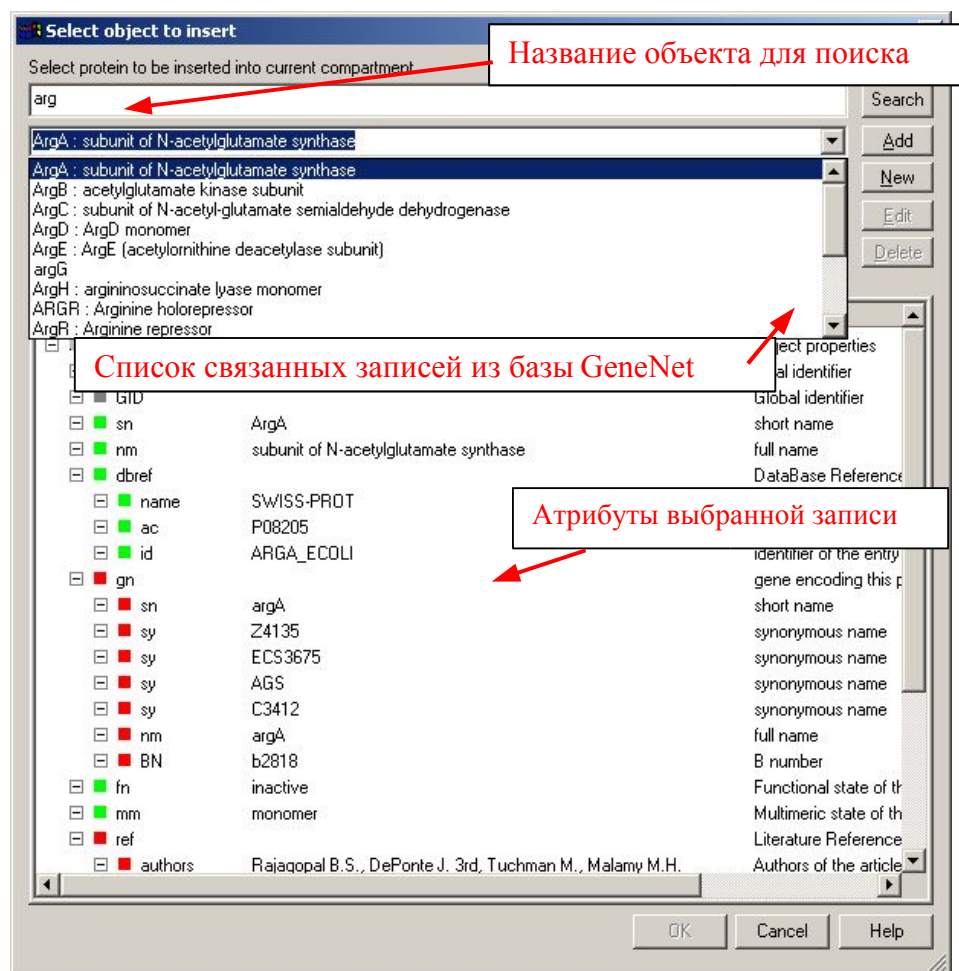


Рисунок 2. Окно поиска и выбора объекта для добавления на схему генной сети.

2.1.1.6. Ввод информации в базу данных с помощью редактора

Для редактирования записи базы данных существует два пути:

1. Если объект, соответствующий данной записи, отображен на схеме сети, нужно кликнуть по нему мышью и в окне отображения атрибутов объекта (номер 4 на Рис. 1) нажать кнопку 'Edit', после чего объект откроется на редактирование.
2. Если объект не отображен на схеме, его нужно найти в базе (см. предыдущий раздел), кликнуть по соответствующей строке мышью чтобы строка изменила цвет на синий, затем нажать кнопку 'Add' (Рис. 2), после чего активируется кнопка 'Edit' и можно будет открыть окно редактирования свойств объекта.

2.1.1.7. Фильтры данных

Система фильтров данных, позволяет выделять подсистемы по заданным параметрам их состояния, например, по тканеспецифичности, по клеткам, тканям или органам, по индукторам и репрессорам, внешним условиям, замкнутым циклам и т.д. Выделение подсистем основано на использовании тематических слоев, интерпретируемых в биологических терминах. Множество объектов, определяющее тематический слой, описывается ограничениями на атрибуты объектов и задается в виде запроса к базе данных. При наложении фильтра на данные объекты на схеме не меняют своего положения, что упрощает визуальное сравнение при использовании фильтров. Авторизованный пользователь выбирает интересующие его генные сети Plant:salt stress из общего списка и нажимает кнопку «OPEN» (Рис.3).

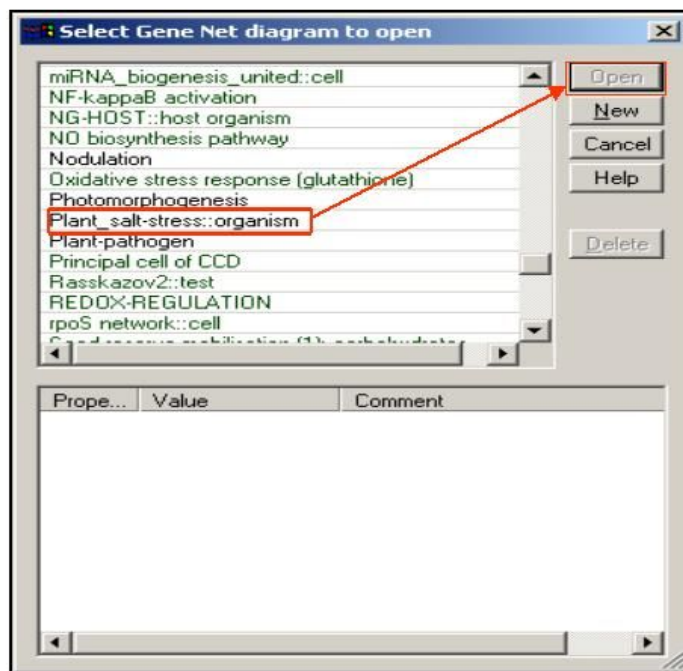


Рисунок 3. Выбор генной сети.

Перед пользователем появляется схема, на которой изображены сигнальные пути и регуляция экспрессии генов при солевом стрессе, а также субклеточная локализация участников процесса: хлоропласты, митохондрии, ядро (Рис. 4).

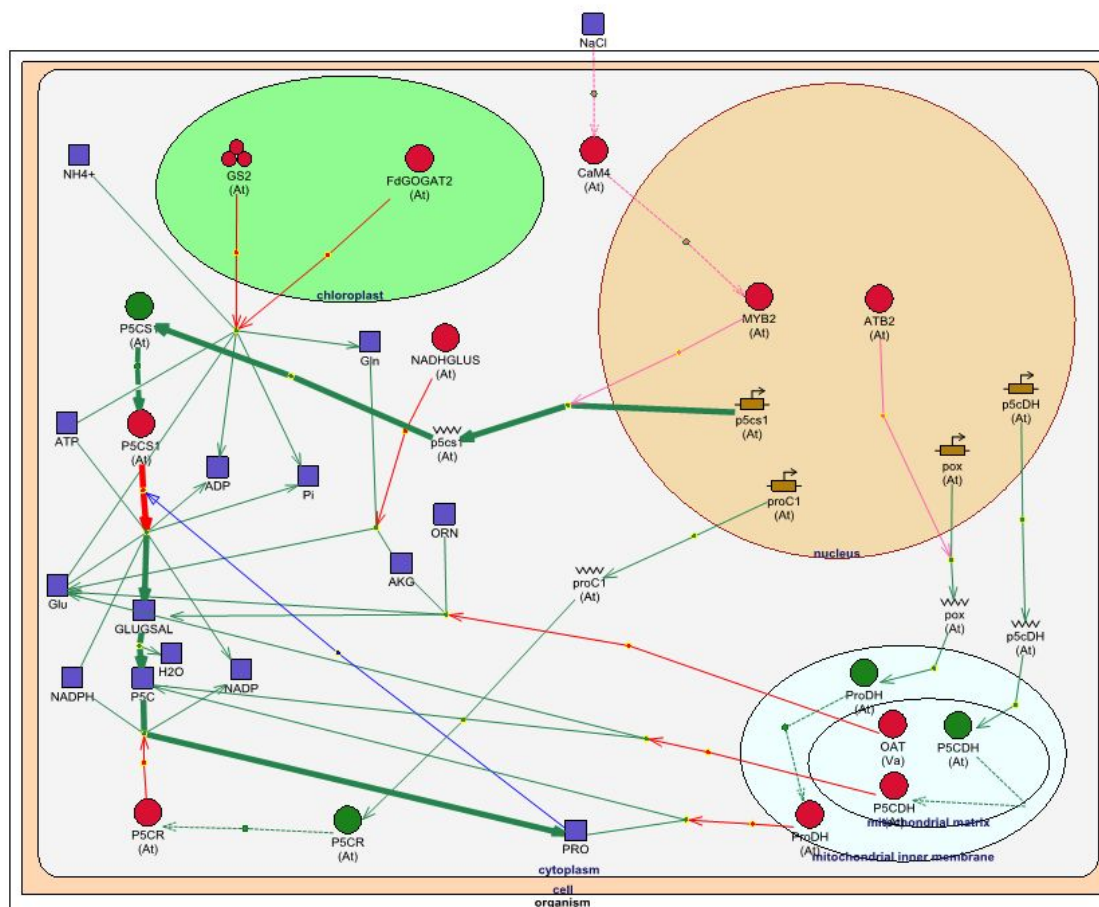


Рисунок 4. Интерфейс графической схемы Plant:salt stress.

Используя систему фильтров (proline), имеющихся на панели инструментов можно, получить информацию об участниках биосинтеза пролина у растений (генах и их продуктах). Появляется зеленая стрелочка, последовательно соединяющая участников процесса (Рис.2).

Далее пользователь может получить информацию об интересующем его гене, выделяя объект на схеме (Рис. 5). Информация включает название вида растения, синонимичные названия гена, ссылки на внешние базы данных (EMBL/GenBank), хромосомную локализацию гена, ссылки на литературу. Аналогичную информацию можно получить для ферментов, осуществляющих данный метаболический путь, и метаболитов участвующих в ферментативных реакциях.

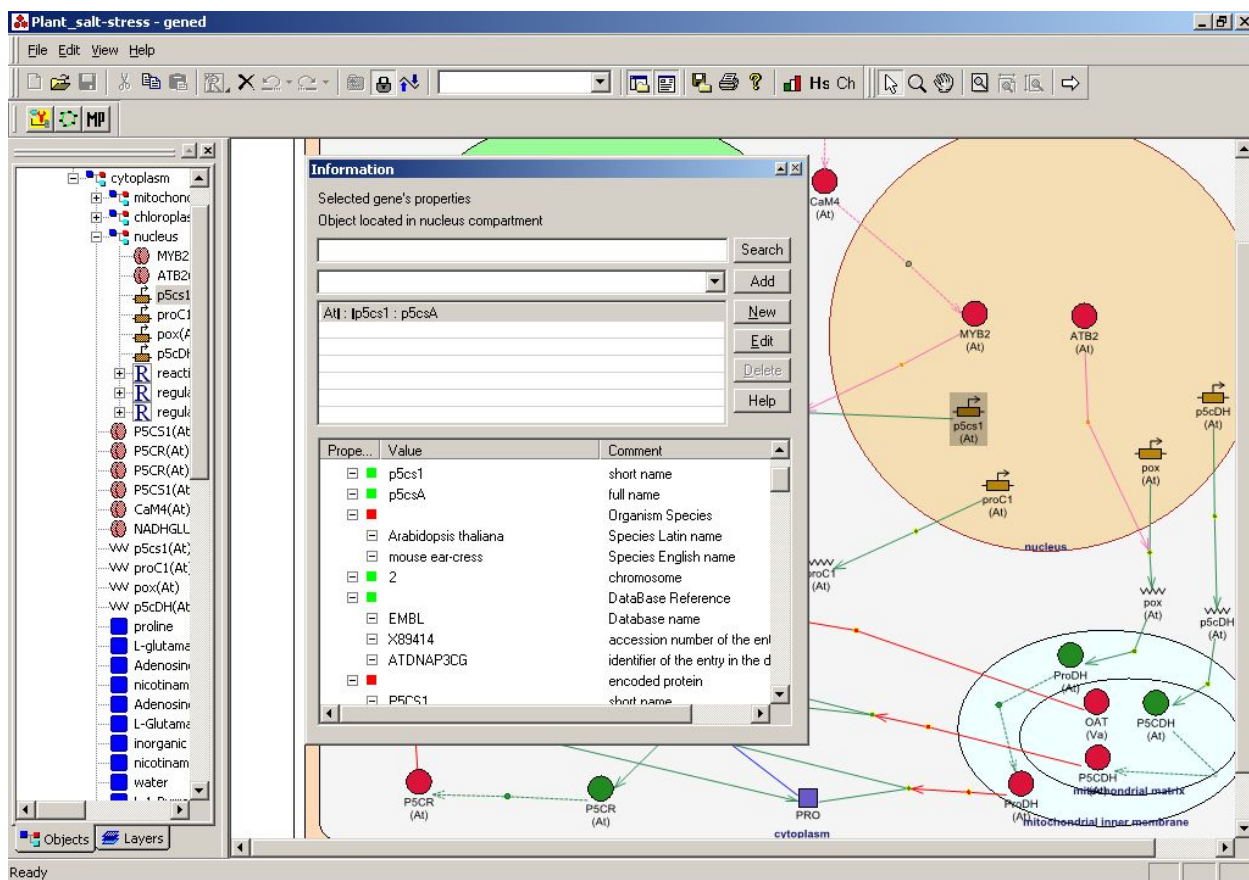


Рисунок 5. Таблица описания гена *p5cs1* на схеме.

Пользователь может получить информацию из графической схемы Plant:salt stress о влиянии транскрипционных факторов, метаболитов на исследуемый процесс: усиление регуляторных воздействий — розовым цветом, ослабление — синим, включение процесса с нулевого уровня — красным и полное выключение процесса черным (Рис. 6).

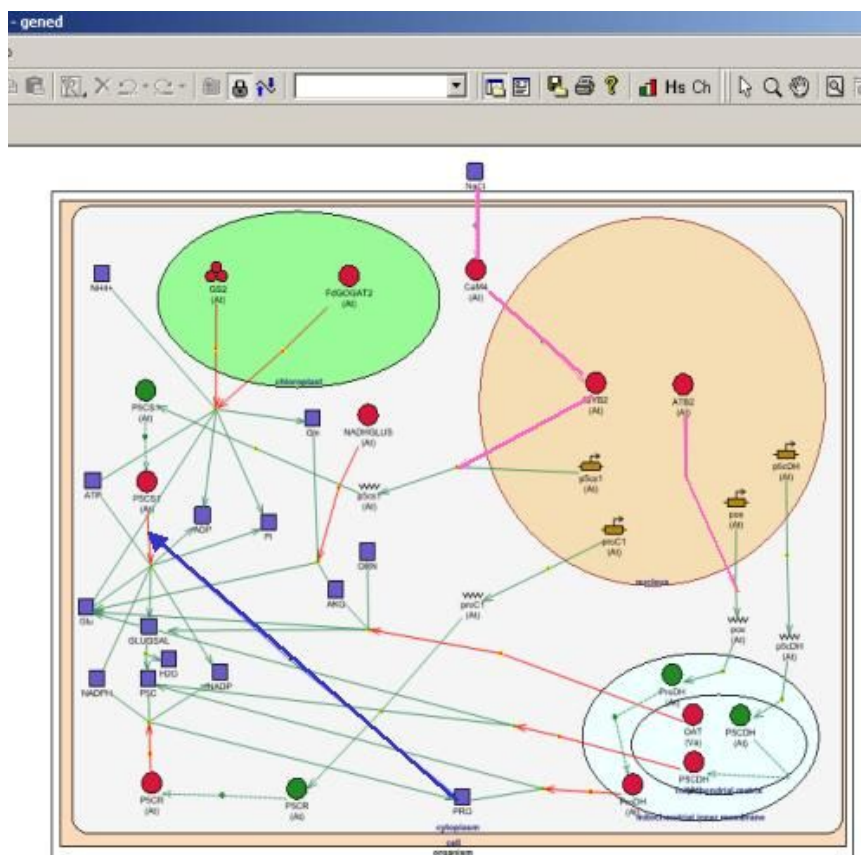


Рисунок 6. Интерфейс графической схемы Plant:salt stress, демонстрирующий характер взаимодействия между участниками процесса.

Полученные в результате экспертные данные могут быть использованы для планирования сложных генетических модификаций (например, одновременного переноса нескольких корегулируемых генов или генов, выполняющих функции в пределах одной метаболической цепи). Дополнительную информацию о целевых генах (регуляторных районах гена, паттернах его экспрессии) участниках процесса можно получить из базы TransGenePromoter.

После выполнения запланированных процедур пользователь имеет возможность сохранить результаты работы программы с помощью стандартных средств интернет-браузера в виде текстового или html-файла.

2.1.2. База данных промоторов для трансгенеза БД TGP

Для обеспечения необходимого паттерна экспрессии перенесенного гена необходимо использовать адекватные регуляторные последовательности и сигналы экспрессии. В настоящее время активно используется ограниченный набор промоторов, что, очевидно, недостаточно для решения большинства сложных биотехнологических задач. В научной литературе опубликовано огромное количество экспериментальных данных о промоторах и сигналах транскрипционного контроля различных генов. Следует отметить, что в статьях приводятся данные о транскрипционной активности производных промоторов – например, делеционных мутантах, промоторах с измененным порядком следования элементов и т.д., паттерн транскрипции которых может отличаться от исходного варианта. Используя эту информацию можно с большой вероятностью подобрать промотор с нужным паттерном

транскрипции даже в тех случаях, когда необходимо обеспечить высокую специфичность экспрессии трансгена. Для подбора подходящего промотора можно использовать некоторые базы данных – например, TRRD (Kolchanov et al., 2002), PLACE (Higo et al., 1999), Plant CARE (Lescot et al., 2002), TRANSFAC (Wingender et al., 2001), однако все они нацелены на детальное описание сайтов связывания транскрипционных факторов и при аннотации промоторов часть ценной с генно-инженерной точки зрения информации не аннотируется. Нами была создана специализированная БД Transgene Promoters (TGP), содержащая детальное описание не только нативных вариантов промоторов, охарактеризованных в генно-инженерных экспериментах, но и данные о сравнительной активности их дериватов.

Отличительной особенностью базы данных TGP является то, что она содержит только экспериментально подтвержденную информацию. Данные заносятся в TGP на основе аннотации статей, содержащих описание результатов экспериментов различных типов. Базовым программным средством, обеспечивающим поиск и навигацию по TGP, а также интеграцию TGP с внешними информационными и программными ресурсами является Sequence Retrieval System (SRS).

2.1.2.1 Описание работы с БД TGP

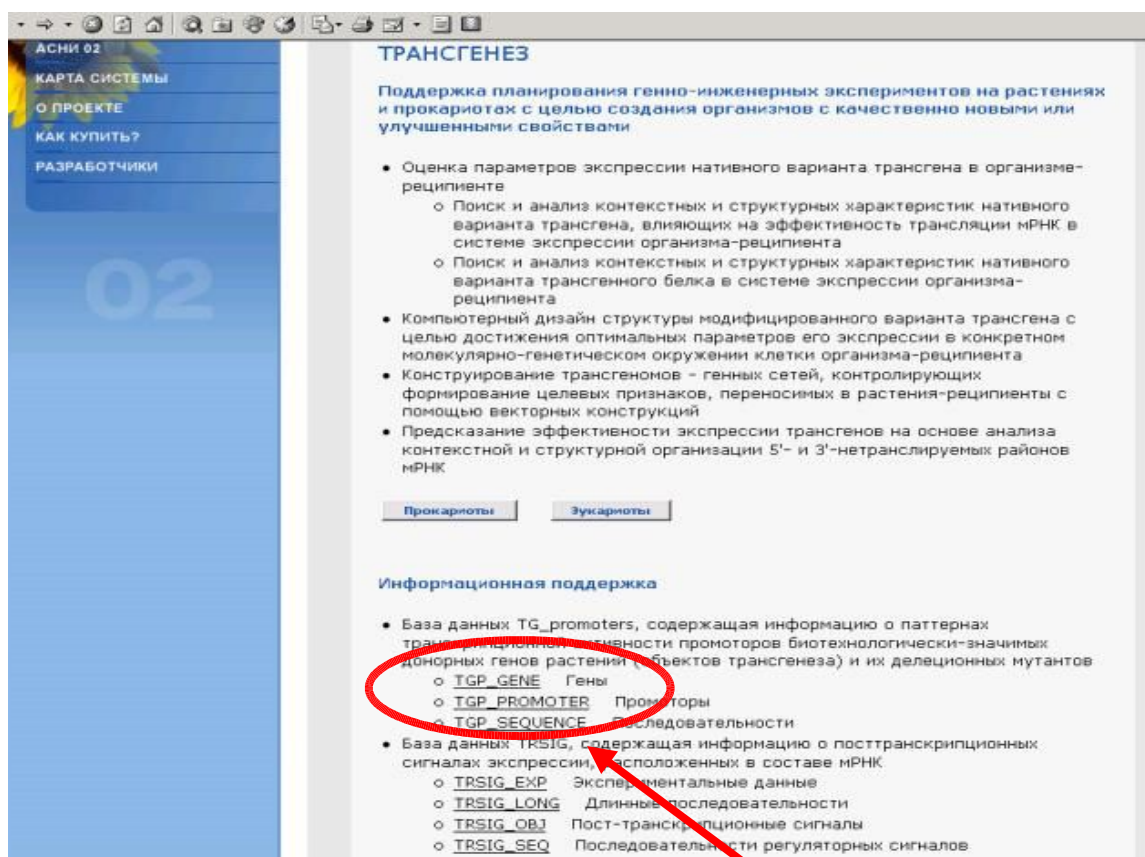


Рисунок 7.

На первом этапе работы с интерфейсом п/с «Трансгенез» необходимо осуществить выбор гиперссылки TGP_Promoter, расположенного непосредственно на входной странице подсистемы, после чего загружается интерфейс системы SRS.

Name TGP_PROMOTER

Status The current release has 77 entries and was indexed 26-Jul-2006.

Description This database accumulates information on functionally active promoters described in annotated scientific sources.

Data-fields in SRS

Name	Short Name	Type	No of Keys	No of Entry References	Indexing Date	Status
PromoterID	pri	id	77	77	26-Jul-2006	ok
GeneID	gnu	index	34	77	26-Jul-2006	ok
Gene	gen	show	0	0		not indexed
Localization	loc	show	0	0		not indexed
Description	desc	show	0	0		not indexed
SequenceID	seq	index	73	74	26-Jul-2006	ok
Reporter	rep	show	0	0		not indexed
Specificity	spc	index	686	2789	26-Jul-2006	ok
Inducer	ind	index	38	105	26-Jul-2006	ok
Comment	spc	index	686	2789	26-Jul-2006	ok
Comment	cc	index	686	2789	26-Jul-2006	ok
Reference	ref	index	686	2789	26-Jul-2006	ok
Pubmed	pmb	show	0	0		not indexed

Links

From Databank	Entries Linked	To Databank	Entries Linked	Links Total	Indexing Date
TGP_GENE	32	TGP_PROMOTER	67	70	26-Jul-2006
TGP_PROMOTER	67	TGP_GENE	32	70	26-Jul-2006
TGP_PROMOTER	70	TGP_SEQUENCE	69	70	26-Jul-2006

SRS Description [Structure](#) ([ssgen.i](#), [ssdb.i](#), [href.i](#))
[Syntax](#)
[Information](#)

SRS 6.1.3.11 | [feedback](#)

Рисунок 8.

На появившейся странице с типовым описанием структуры полей карточки TGP_promoter нажимается кнопка Search.

QUERY

Reset search **TGP_PROMOTER** [Info](#) about field **PromoterID**

Submit Query

append wildcards to words ☒

combine searches with **AND**

Number of entries to display per page **30**

Extended query form

separate multiple values by & (and), | (or), ! (and not)

PromoterID

PromoterID

PromoterID

PromoterID

retrieve entries of type **Entry**

Use view *** Names only ***

Create your own view

Select fields to display:

PromoterID
GeneID
Gene
Localization
Description
SequenceID
Reporter

SRS 6.1.3.11 | [feedback](#)

Рисунок 9.

В левой части таблицы можно выбрать одно из 8 поисковых полей, например, «Inducer».

The screenshot shows the SRS 6.1.3.11 web interface. The top navigation bar includes links for TOP PAGE, QUERY, RESULTS, PROJECTS, VIEWS, DATABANKS, and HELP. The main search area has a 'Reset' button and a search box containing 'TGP PROMOTER'. Below the search box, there are several options: 'append wildcards to words' (checked), 'combine searches with' (set to AND), and 'Number of entries to display per page' (set to 30). A 'Submit Query' button is present. On the right, there is a section for 'separate multiple values by & (and), | (or), ! (and not)'. This section contains a list of search fields: Inducer, PromoterID, GeneID, SequenceID, Specificity, Inducer (highlighted with a red arrow), Comment, Comment, Reference, and *all entries*. Below this list, there is a 'retrieve entries of type' dropdown set to 'Entry'. At the bottom, there is a 'Use view' dropdown set to '*Names only*' and a 'Create your own view' button. The footer shows 'SRS 6.1.3.11 | feedback'.

Рисунок 10.

После этого в правой части поисковой таблицы набирается название желаемого индуктора, например, «elictor».

The screenshot shows the SRS 6.1.3.11 web interface after the search criteria have been updated. The 'Inducer' field in the search criteria section is now filled with the text 'elictor', and a red arrow points to it. The 'PromoterID' field is also visible. The 'Number of entries to display per page' is set to 30. The 'Extended query form' button is visible. The 'retrieve entries of type' dropdown is still set to 'Entry'. The 'Use view' dropdown is still set to '*Names only*'. The 'Create your own view' button is still present. The footer shows 'SRS 6.1.3.11 | feedback'.

Рисунок 11.

В результате поиска («**Submit Query**») появляется новая страница с идентификаторами промоторов, которые индуцируются заданным индуктором, в данном случае элиситором (elicitor).

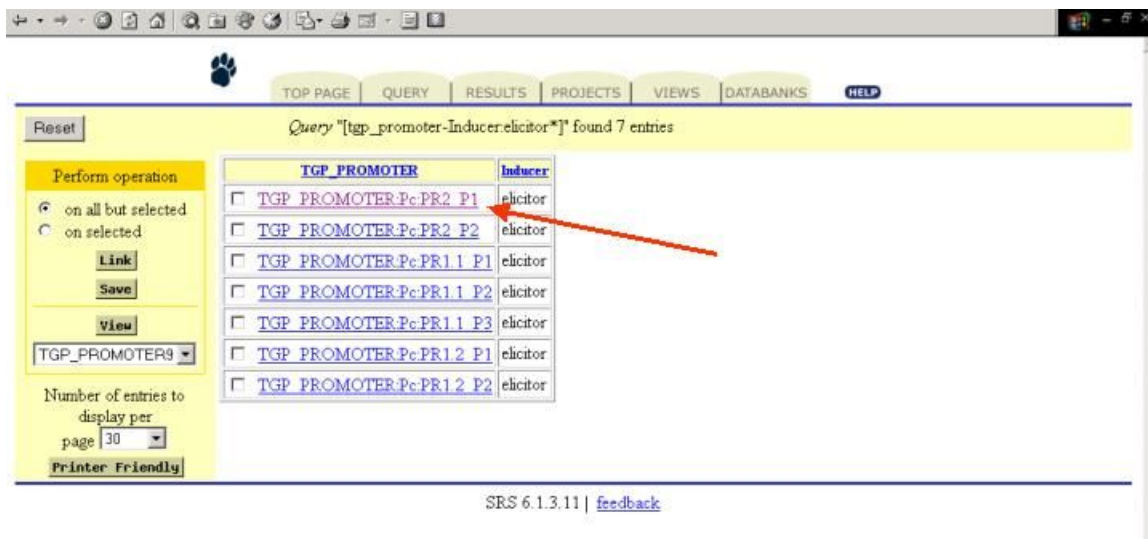


Рисунок 12.

После нажатия левой клавишей мыши на идентификатор промотора появляется новое окно, в котором представлена карточка промотора.

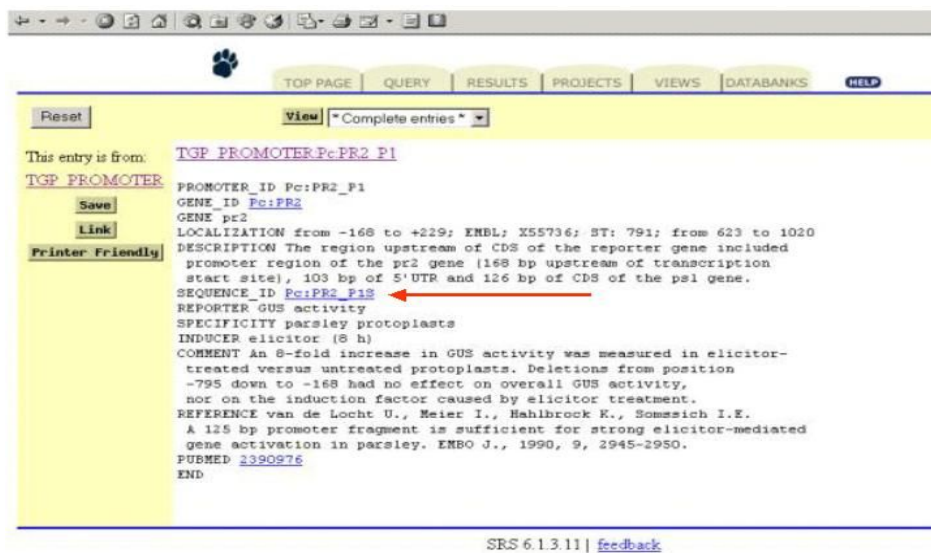


Рисунок 13.

По гиперссылкам в полях идентификатора гена (GENE_ID) или идентификатора последовательности промотора (SEQUENCE_ID) можно перейти на описание соответствующего гена или последовательности данного промотора.



Рисунок 14.

Список индукторов, по которым можно проводить поиск в базе TGP можно получить, если в поисковом окне выставить «**Inducer**» и нажать кнопку «**Info**».

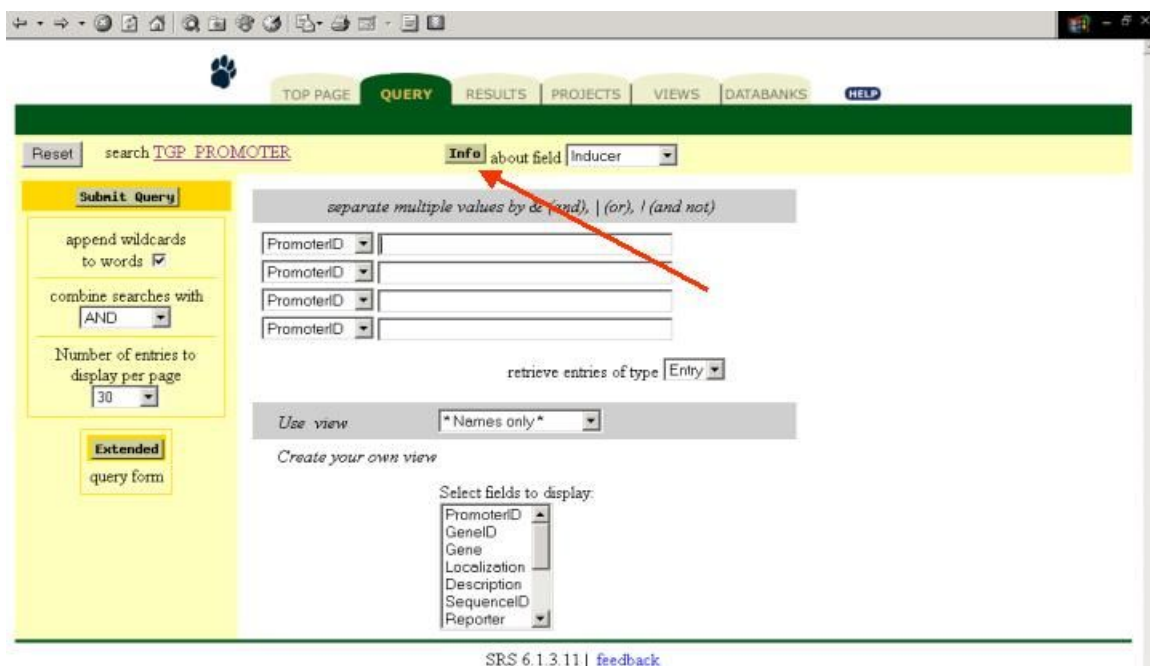


Рисунок 15.

После выполнения данной операции появляется окно, в котором нужно нажать кнопку «List Values».

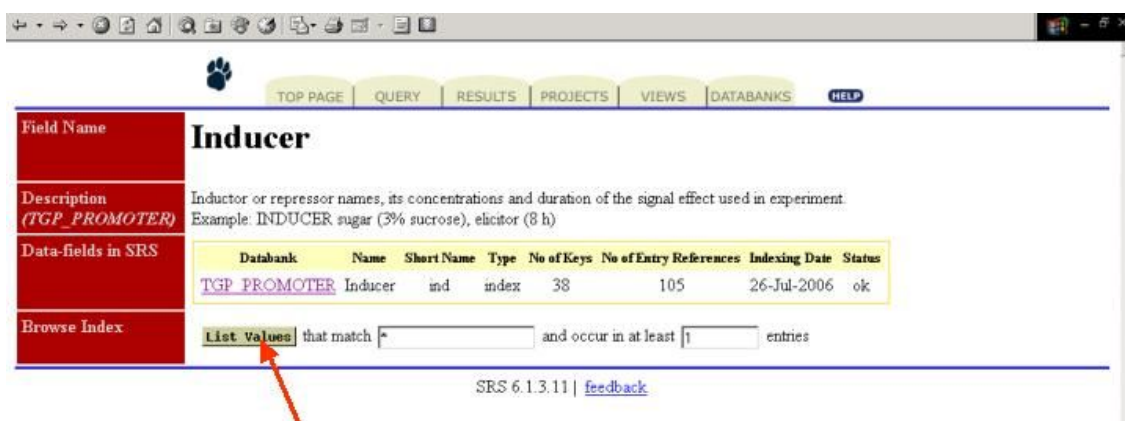


Рисунок 16.

После выполнения этой операции появится список индукторов, представленных в базе.

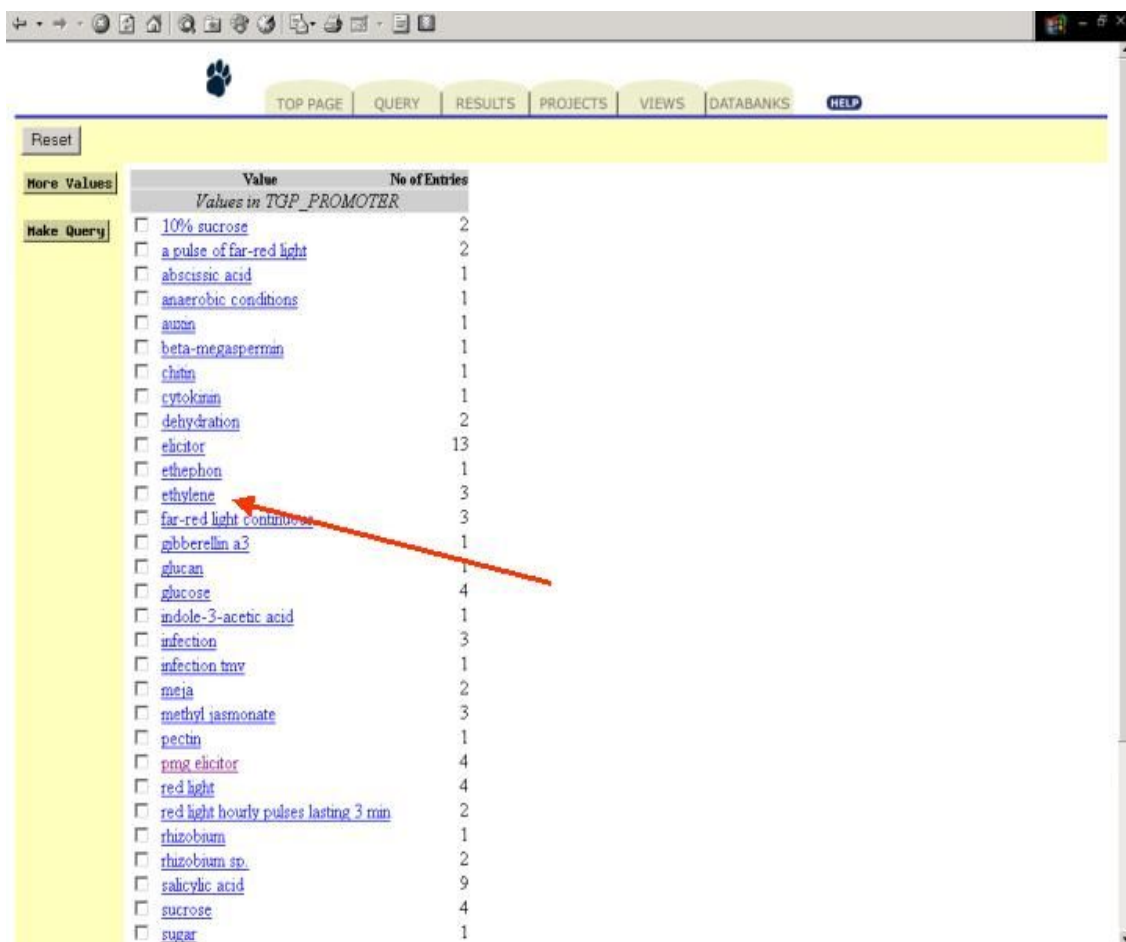


Рисунок 17.

Выбрав любой из индукторов, например, «ethylene», можно выйти в карточки промоторов, индуцируемых этиленом.

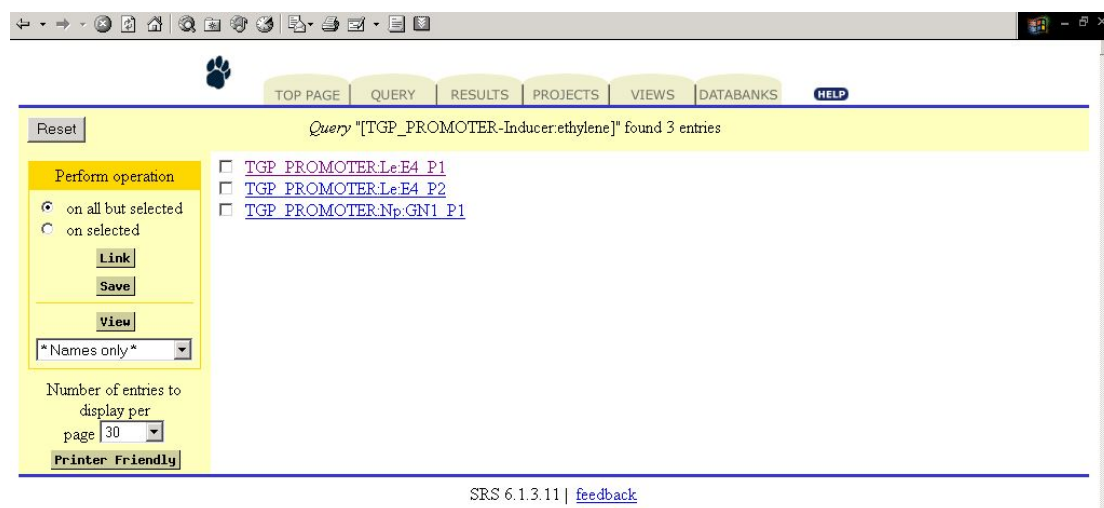


Рисунок 18.

Используя основную поисковую таблицу, можно вести поиск сразу по нескольким полям. Помимо поля «Inducer», в таблице **TGP_promoter** можно вести информационный поиск по полям «Comment» и «Specificity». Примеры типовых запросов для таблицы **TGP_promoter**: найти промоторы, которые активируются конкретным индуктором (поиск по полям «Inducer» и «Comment»); найти тканеспецифические промоторы (поиск по полю «Specificity»); найти промоторы, активность которых изучалась у определенного вида трансгенных растений, например, у табака (поиск по полю «Specificity»).

После выполнения запланированных процедур пользователь имеет возможность подобрать карточку (карточки) БД TGP, содержащие информацию о промоторе с нужным паттерном транскрипции (если такая информация содержится в БД). Эта информация может быть сохранена на компьютере пользователя с помощью стандартных функций интернет-браузера в виде текстового или html-файла и использована непосредственно для планирования эксперимента.

2.1.3. База данных посттранскрипционных сигналов экспрессии БД TRSIG

Помимо сигналов транскрипционного контроля, важное значение может иметь оптимизация экспрессии трансгена на посттранскрипционном уровне. В составе эукариотических мРНК часто содержатся сигналы, контролирующие эффективность трансляции или цитоплазматическую стабильность матрицы (как общую, так и стадии-, ткане- или стресс-специфическую). Применение таких сигналов в дизайне трансгена может существенно увеличить эффективность трансгенеза (Gallie, 1998; Biemelt et al., 2003). Кроме того, в составе чужеродной ДНК могут содержаться элементы, сходные с посттранскрипционными сигналами экспрессии растительных генов: обнаружение и – при необходимости – элиминация таких сигналов может оказаться необходимым условием для достижения нужного паттерна экспрессии трансгена.

В настоящее время существует несколько БД, содержащих информацию об эукариотических посттранскрипционных сигналах экспрессии. Например, структура БД TransTerm (Jacobs et al., 2002) и UTRsite (Pesole et al., 2002) позволяет проводить поиск нескольких десятков сигналов в мРНК, интересующих пользователя. Однако, в настоящее время известна полная структура очень небольшого числа таких сигналов, что существенно

ограничивает применимость таких информационных ресурсов. В подсистему «Трансгенез» включена БД TRSIG (Kochetov et al., 2001), созданная на основе другого подхода. В этой БД собрана информация о потенциальных сигналах – участках мРНК, которые, согласно экспериментальным данным, способны влиять на эффективность трансляции или на цитоплазматическую стабильность матрицы. Таким образом, в TRSIG содержатся не только сигналы, структура которых была досконально изучена, но и нуклеотидные последовательности – регуляторы трансляции, активные элементы которых еще не определены.

БД TRSIG реализована на платформе SRS и состоит из четырех основных баз данных: (1) БД объектов (содержит общую информацию о типе трансляционного сигнала), (2) БД экспериментов (описывает условия эксперимента и активность экспериментальных последовательностей), (3) БД экспериментальных нуклеотидных последовательностей (фрагменты мРНК, активность которых изучалась в экспериментах), (4) БД полноразмерных экспериментальных мРНК (предназначена для реконструкции протяженных вторичных структур). Карточки этих БД содержат перекрестные ссылки, что – при использовании возможностей SRS – позволяет находить нужные сведения. В TRSIG также предусмотрена возможность поиска гомологичных участков между фрагментами мРНК, содержащими пост-транскрипционные сигналы экспрессии (БД экспериментальных нуклеотидных последовательностей), трансляционная активность которых была определена экспериментально (и описана в БД экспериментов), и мРНК, интересующей пользователя. С этой целью использован классический BLAST алгоритм. Наличие локальной гомологии с участком, содержащим пост-транскрипционный сигнал, можно интерпретировать исходя из имеющихся в TRSIG совокупности данных об относительной активности мутационных вариантов этой последовательности.

2.1.3.1. Описание работы с БД TRSIG

На первом этапе работы с интерфейсом п/с «Трансгенез» необходимо осуществить выбор гиперссылки **TRSIG_OBJ**, расположенной непосредственно на входной странице подсистемы.

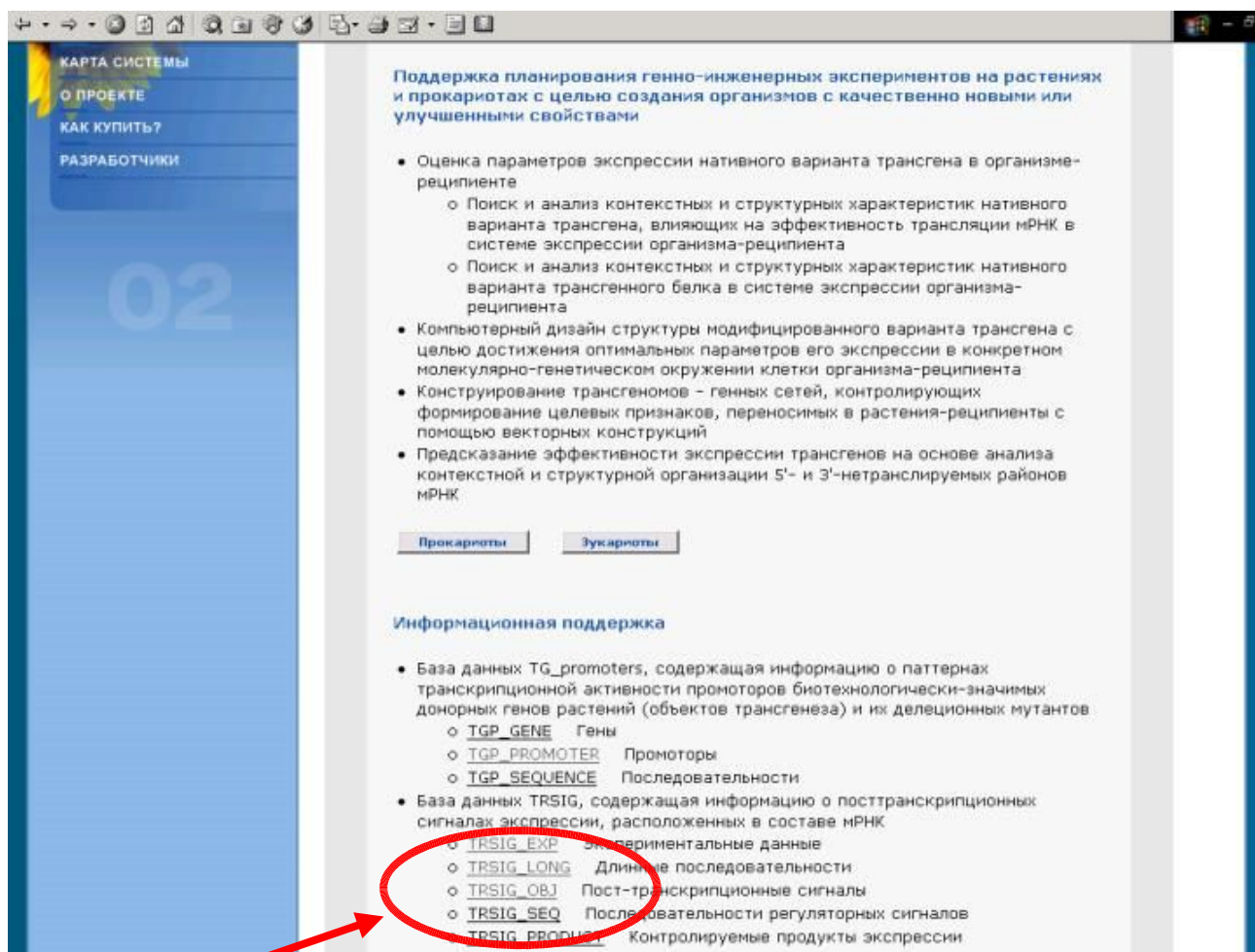


Рисунок 19.

На появившейся странице с типовым описанием структуры полей карточки **TRSIG_OBJ** нажимается кнопка «Search».

[TOP PAGE](#)
[QUERY](#)
[RESULTS](#)
[PROJECTS](#)
[VIEWS](#)
[DATABANKS](#)
[HELP](#)

Name

Status

Description

Data-fields in SRS

SRS Description

TRSIG_OBJ

The current release has 38 entries and was indexed 05-Oct-2006.

Description of post-transcriptional signals

Name	Short Name	Type	No of Keys	No of Entry References	Indexing Date	Status
ID	id	id	38	38	05-Oct-2006	ok
Date	dat	num	15	38	05-Oct-2006	ok
Author	au	index	3	40	05-Oct-2006	ok
Location	loc	index	8	47	05-Oct-2006	ok
Type	typ	index	21	92	05-Oct-2006	ok
OC	oc	index	74	369	05-Oct-2006	ok
OS	os	index	42	103	05-Oct-2006	ok
GeneName	gen	index	69	134	05-Oct-2006	ok
Cap	cap	index	8	53	05-Oct-2006	ok
PolyA	pla	index	5	61	05-Oct-2006	ok
Sequence	sq	show	0	0		not indexed
CommentSeq	csq	index	154	365	05-Oct-2006	ok
Keywords	key	index	43	123	05-Oct-2006	ok
Comment	cc	index	676	2777	05-Oct-2006	ok
Link	lnk	index	32	37	05-Oct-2006	ok
Product	pro	index	32	35	05-Oct-2006	ok
Experiment	exp	index	106	127	05-Oct-2006	ok

Structure ([srsgen.i](#), [srsdb.i](#), [href.i](#))
[Syntax](#)
[Information](#)

SRS 6.1.3.11 | [feedback](#)

Рисунок 20.

База данных объектов **TRSIG_OBJ** включает информацию об основных характеристиках трансляционного сигнала (Табл. 1).

Имя поля	Описание
ID	Идентификатор записи.
DATE	Дата посленного обновления
AUTHOR	Автор записи (аннотатор)
LOCATION	Локализация потенциального сигнала
TYPE	Тип потенциального сигнала
OC	Таксономическая принадлежность организма
OS	Название организма
GENE	Название гена, в котором локализован сигнал
CAP	Наличие или отсутствие кепа на 5'-конце мРНК
POLYA	Наличие или отсутствие поли(А)-хвоста у мРНК
SQ	Нативный вариант нуклеотидной последовательности сигнала
COMMENTSEQ	Комментарий к полю SQ
KEYWORD	Ключевые слова
COMMENT	Краткий комментарий – описание сигнала
LINK	Идентификаторы БД, связанных с TRSIG
EXPERIMENT	Идентификаторы соответствующих карточек БД TRSIG_EXP

Таблица 1. Список полей TRSIG_OBJ и их краткое описание.

Пример типового запроса для таблицы TRSIG_OBJ: найти сигналы, локализованные в 5'-НТП, 3'-НТП и CDS (поиск по полю LOCATION); найти тканеспецифические сигналы (поиск по полю TYPE, COMMENT).

The screenshot displays the TRSIG database query interface. At the top, there is a navigation bar with tabs: TOP PAGE, QUERY (selected), RESULTS, PROJECTS, VIEWS, DATABANKS, and HELP. Below the navigation bar, a search bar contains the text "search TRSIG_OBJ" and an "Info" button. To the right of the search bar is a dropdown menu labeled "about field" with "ID" selected. Below the search bar, there is a "Submit Query" button. To the left of the main search area, there are several options: "append wildcards to words" (checked), "combine searches with" (set to "AND"), and "Number of entries to display per page" (set to "30"). Below these options is an "Extended query form" button. The main search area contains a section titled "separate multiple values by & (and), | (or), ! (and not)". Below this title are four rows of input fields: "Location" with the value "utr", and three rows for "ID". To the right of these fields is a dropdown menu labeled "retrieve entries of type" with "Entry" selected. Below this is a "Use view" dropdown menu with "*Names only*" selected. Below the "Use view" dropdown is a section titled "Create your own view" with a "Select fields to display:" label and a list of fields: ID, Date, Author, Location, Type, OC, and OS. At the bottom of the page, there is a footer with the text "SRS 6.1.3.11 | [feedback](#)".

Рисунок 21.

Если найденный сигнал удовлетворяет требованиям пользователя, нуклеотидная последовательность из поля **Sequence** соответствующей карточки **TRSIG_OBJ** может в дальнейшем использоваться как специфический сигнал экспрессии при дизайне регуляторных элементов трансгена.

На следующем этапе работы с интерфейсом п/с «Трансгенез» необходимо осуществить выбор гиперссылки **TRSIG_EXP** расположенной непосредственно на входной странице подсистемы,. Из списка доступных баз данных выбирается **TRSIG_EXP** (Рис. 22).

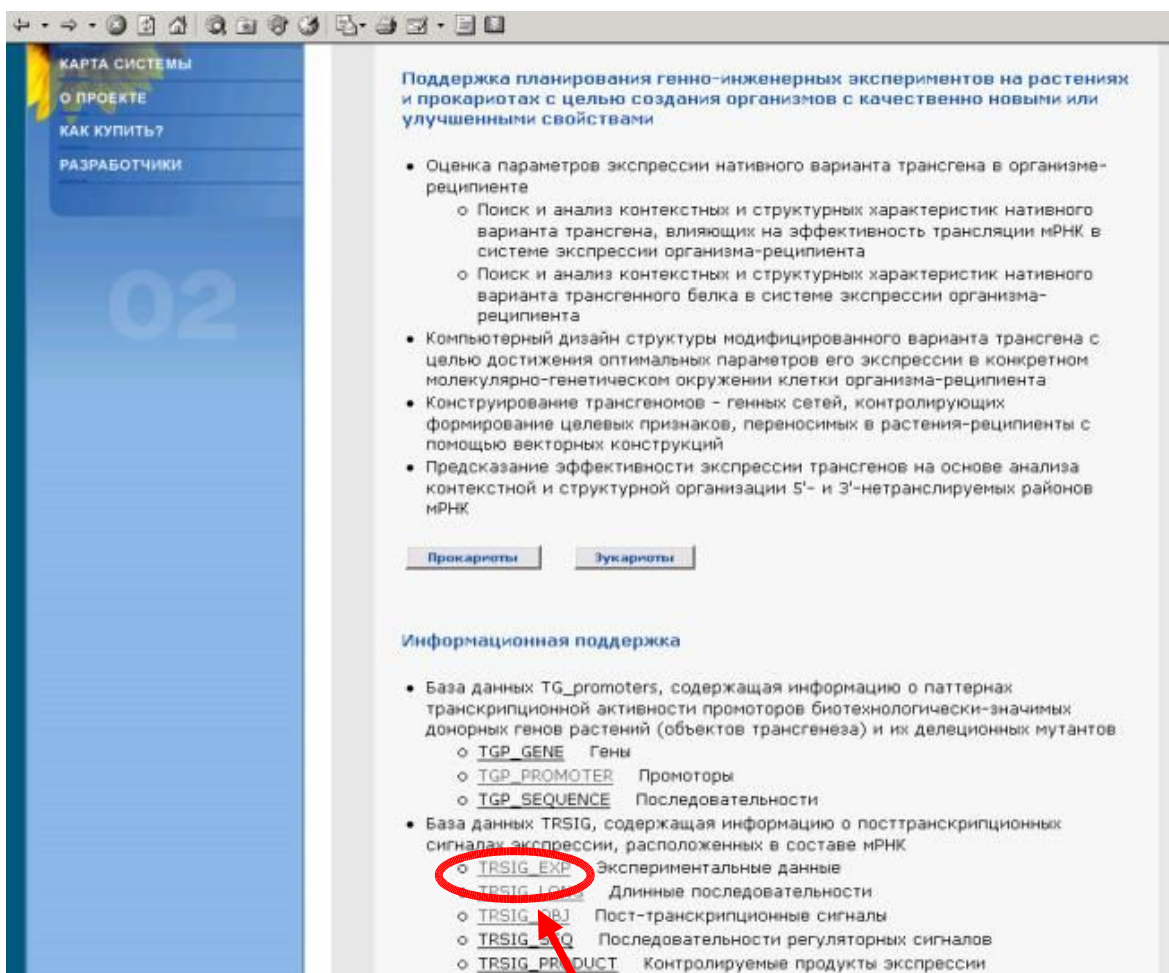


Рисунок 22.

На появившейся странице с типовым описанием структуры полей карточки **TRSIG_EXP** нажимается кнопка «Search».

TOP PAGE

QUERY

RESULTS

PROJECTS

VIEWS

DATABANKS

HELP

Name

Status

Description

Data-fields in SRS

SRS Description

TRSIG_EXP

The current release has 106 entries and was indexed 05-Oct-2006.

Database on experimental data on mRNA-located post-transcriptional signals

Name	Short Name	Type	No of Keys	No of Entry References	Indexing Date	Status
ID	id	id	106	106	05-Oct-2006	ok
ObjID	oi	index	37	120	05-Oct-2006	ok
Reference	ref	index	523	2599	05-Oct-2006	ok
Type	typ	index	112	648	05-Oct-2006	ok
Cell	cel	index	87	304	05-Oct-2006	ok
Comment	cc	index	625	4631	05-Oct-2006	ok
Activity	act	index	225	398	05-Oct-2006	ok

Structure ([srsgeni](#), [srssdbi](#), [hrefi](#))

[Syntax](#)

[Information](#)

Search

SRS 6.1.3.11 | [feedback](#)

Рисунок 23.

БД включает краткое описание эксперимента, в котором была продемонстрирована активность потенциального сигнала (Табл. 2).

Имя поля	Описание
ID	Идентификатор записи
OBJID	Идентификатор записи в БД TRSIG_OBJ
REFERENCE	Ссылка на публикацию
TYPE	Тип эксперимента
CELL	Клетки, в которых велся эксперимент
COMMENT	Краткое описание эксперимента
ACTIVITY	Численные активности экспериментальных последовательностей

Таблица 2. Список полей TRSIG_EXP и их краткое описание.

Пример типового запроса для этой таблицы: найти сигналы, экспрессия которых была исследована в определенных тканях растений (поиск по полям **CELL** и **COMMENT**). Если найденный сигнал удовлетворяет требованиям пользователя, нуклеотидная последовательность из поля **SQ** карточки **TRSIG_SEQ**, ссылки на которые содержатся в карточке **TRSIG_EXP**, может в дальнейшем использоваться как специфический сигнал экспрессии при дизайне регуляторных элементов транскена.

Альтернативно, на входной странице подсистемы «Трансгенез» пользователь может выбрать одну из баз данных **TRSIG_SEQ**, **TRSIG_PRODUCT**, **TRSIG_LONG**. Допускается использование сложных запросов с булевскими операторами. Также возможно составление запросов одновременно к нескольким из этих таблиц базы даггых.

База данных **TRSIG_SEQ** аккумулирует информацию о сигнал-содержащих экспериментальных нуклеотидных последовательностях (Табл. 3).

Имя поля	Описание
ID	Идентификатор записи.
OBJID	Идентификатор записи в БД TRSIG_OBJ
SQ	Нуклеотидная последовательность сигнала
LONG	Ссылка на наличие полноразмерной версии нуклеотидной последовательности
COMMENT	Краткое описание экспериментальной последовательности
EXPR	Численные значения экспериментальных активностей

Таблица 3. Список полей TRSIG_SEQ и их краткое описание.

База данных **TRSIG_PRODUCT** представляет собой описание белкового продукта, кодируемого геном, в котором расположен потенциальный сигнал, и поставляется в формате базы **SwissProt** либо (при отсутствии информации в **SwissProt**) – в формате **GenBank**.

База данных **TRSIG_LONG** аккумулирует информацию о сигнал-содержащих полноразмерных экспериментальных нуклеотидных последовательностях (Табл. 4).

Имя поля	Описание
ID	Идентификатор записи.
COMMENT	Краткое описание нуклеотидной последовательности
START	Позиция сайта инициации трансляции
STOP	Позиция сайта терминации трансляции
SQ	Полноразмерная версия нуклеотидной последовательности

Таблица 4. Список полей TRSIG_LONG и их краткое описание.

После выполнения запланированных процедур пользователь имеет возможность подобрать карточку (или карточки) БД **TRSIG**, содержащие информацию о посттранскрипционном сигнале с нужным паттерном транскрипции (если такая информация содержится в БД). Эта информация может быть сохранена на компьютере пользователя с помощью стандартных функций интернет браузера в виде текстового или html-файла и использована непосредственно для планирования эксперимента.

2.2. Программные компоненты подсистемы «Трансгенез»

Помимо выбора подходящих регуляторных элементов (промотора и сигналов посттранскрипционного контроля), для достижения эффективного взаимодействия с аппаратом экспрессии организма-реципиента часто бывает необходимо модифицировать структуру трансгена. Обычно в трансгенезе используют только белок-кодирующую часть (БКП). Однако, БКП чужеродного гена также может обладать негативными характеристиками, снижающими эффективность экспрессии. Можно выделить два вида таких характеристик: (1) параметры мРНК, влияющие на эффективность взаимодействия с аппаратом трансляции организма-реципиента, (2) элементы нуклеотидной последовательности, сходные с сигналами экспрессии организма-реципиента и неправильно воспринимаемые как такие сигналы (например, ложные сайты сплайсинга или полиаденилирования).

Известно, что некоторые параметры растительных мРНК влияют на трансляционную активность (Кочетов и Шумный, 1998). Показано, что 5'НТП регулирует взаимодействие аппарата инициации трансляции и рибосомы (Kozak, 2002). Лидерный район может содержать как негативные (стабильные шпильки, AUG триплеты и рамки считывания), так и позитивные (трансляционные энхансеры) сигналы. К числу известных сигналов общего характера относят контекст стартового кодона трансляции, влияющий на его распознавание

рибосомами (Lukaszewicz et al., 2000; Sawant et al., 2001). В ряде случаев 5'-НТП содержит сигналы, опосредующие инициацию трансляции по механизмам внутренней инициации трансляции (Pestova et al., 2001). По-видимому, лидерная последовательность может также влиять на цитоплазматическую стабильность мРНК (Gutierrez et al., 1999; Biemelt et al., 2003). В настоящее время критерии, позволяющие определять функциональную активность лидерной последовательности, исследованы недостаточно.

Для обеспечения эффективной инициации трансляции мРНК транскрипта часто в качестве 5'-НТП используют трансляционные энхансеры (информация о них представлена в TRSIG). Кроме этого, необходимо оптимизировать контекст стартового кодона трансляции (Lukaszewicz et al., 2000; Sawant et al., 2001) и оценить стабильность вторичной структуры 5'-концевого района мРНК (включающего 5'-НТП и начало CDS). Ранее нами было показано, что характеристики мРНК высокоэкспрессирующихся генов оптимизированы для обеспечения эффективной трансляции (Kochetov et al., 1998). На основе выявленных закономерностей была создана компьютерная система Leader_RNA, позволяющая оценивать трансляционную активность 5'-НТП мРНК растений и млекопитающих (Kochetov et al., 1999). Эта система может быть также использована для оценки функциональной активности 5'-НТП мРНК транскрипта в клетках двудольных и однодольных растений. Эти результаты были использованы в качестве базовых при разработке подсистемы «Трансгенез».

Ниже перечислены основные программные компоненты подсистемы «Трансгенез» и приведена краткая информация об их назначении и детальное описание правил пользования.

2.2.1. Операции для ввода данных в подсистему «Трансгенез»

Программные компоненты подсистемы «Трансгенез» интегрированы в общую систему ввода данных. Нуклеотидная последовательность варианта транскрипта вводится пользователем в виде последовательностей трех функциональных районов (5'-НТП, белок-кодирующей части и 3'-НТП), что позволяет определить разметку районов мРНК. Эта последовательность проверяется на адекватность (использование канонических символов нуклеотидов, целостность рамки считывания, правильное использование кодонов инициации и терминации трансляции) и вводится в рабочую (буферную) область подсистемы. В дальнейшем при работе с программами, допускающими оптимизацию нуклеотидной последовательности, у пользователя появляется возможность выбора: оставить «буферный» вариант без изменений, или заменить его на модифицированный вариант. Таким образом, в рамках подсистемы «Трансгенез» есть возможность внесения последовательных модификаций в мРНК транскрипта согласно рекомендациям, предложенным различными программными компонентами.

Корректность вводимой информации (нуклеотидных последовательностей функциональных районов мРНК (5'-НТП, CDS, 3'-НТП); требования к вводимым данным могут различаться для некоторых программных компонентов, задействованных в реализации функций п/с «Трансгенез» и в таком случае будут оговорены в при описании каждого конкретного случая отдельно. В общем случае система ввода данных проверяет использование неканонических символов (каноническими являются A,T,U,G,C,a,t,u,g,c), а также параметры CDS (должна начинаться с ATG или atg и заканчиваться TAA (taa), TGA (tga) или TAG (tag); этих трех сигналов также не должно быть в рамке считывания CDS, что является стандартным правилом при обработке нуклеотидных последовательностей этого класса). При несоблюдении этих требований система ввода п/с «Трансгенез» выдает сообщение об ошибке и прерывает выполнение дальнейших процедур.

Для работы с программными компонентами необходимо нажать кнопку **Эукариоты** на входной странице подсистемы «Трансгенез» (рис. 24).

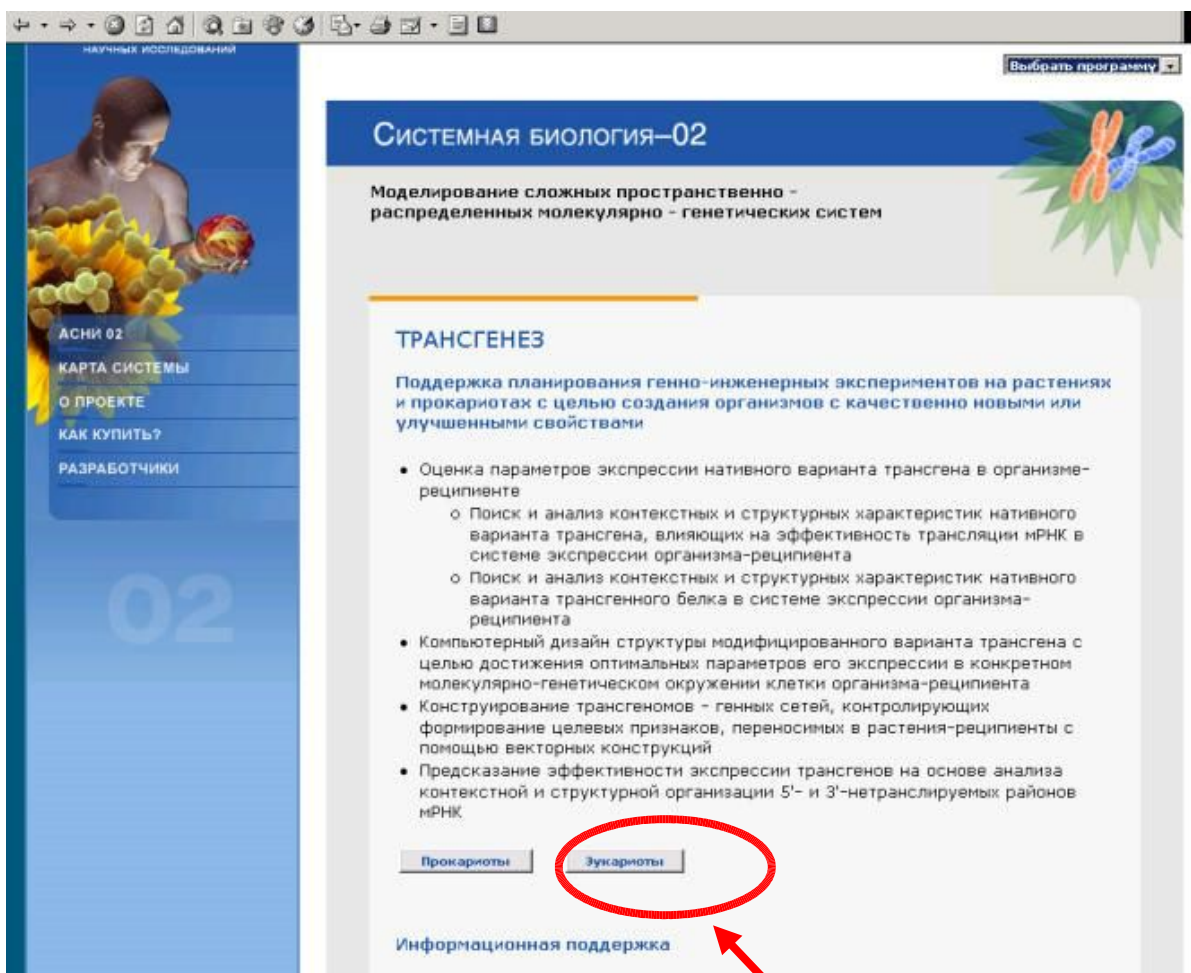


Рисунок 24.

После этого осуществится переход на страницу программных компонентов (рис. 24) (здесь и далее нажатие на кнопки и ссылки производится левой кнопкой мыши). На этой странице расположены окна для ввода функциональных районов мРНК (5'-НТП – 5'-нетранслируемый или лидерный район мРНК; CDS – белок-кодирующая часть мРНК; 3'-НТП – 3'-нетранслируемый или трейлерный район мРНК). Возможен ввод нуклеотидных последовательностей либо в виде нуклеотидных последовательностей без всяких дополнительных сигнификаторов, либо из файлов с локального компьютера пользователя.

Введите нуклеотидную последовательность 5'-НТП (в нижерасположенное окно или из файла):

Введите белок-кодирующую последовательность мРНК (в нижерасположенное окно или из файла):

Введите нуклеотидную последовательность 3'-НТП (в нижерасположенное окно или из файла):

Введите нуклеотидную последовательность мРНК трансгена: 5'-НТП – с 5'-конца мРНК до стартового кодона AUG (AUG кодон не включать). Белок-кодирующая последовательность: от стартового кодона трансляции (AUG) до стоп-кодона (UAA, UGA, UAG). 3'-НТП: нуклеотидную последовательность, расположенную сразу после стоп-кодона и до начала поли(А)-хвоста. Интрон-содержащие геномные последовательности использовать нельзя, можно использовать только процессоризованные (зрелые) мРНК.

Ввести мРНК Очистить

Рисунок 25.

Следует учесть, что специфика работы п/с «Трансгенез» в большинстве случаев требует работы с индивидуальной мРНК, поэтому возможность массового анализа выборок мРНК не предусмотрена. После ввода нуклеотидных последовательностей в соответствующие области интерфейса, необходимо нажать на кнопку **Ввести мРНК**. После этого система ввода подсистемы проведет анализ корректности введенных данных и загрузит их в систему для выполнения последующих операций. В новом окне появится информация о введенной последовательности (размеры функциональных районов), а также будет зарезервирована область для отображения информации о модифицированном варианте последовательности, который будет получен в ходе работы подсистемы (Рис.26).



Рисунок 26.

На этом этапе ввод данных в систему считается завершенным и пользователь может осуществлять работу с введенной нуклеотидной последовательностью мРНК трансгена с помощью имеющихся в п/с «Трансгенез» программных компонент. После выполнения запланированных процедур пользователь имеет возможность визуализировать различия между исходным вариантом нуклеотидной последовательности трансгена и конечным вариантом модифицированной последовательности (нажатием кнопки **Показать различия (Show Differences)**), результат показан на рис. 27), а также пользователь может сохранить исходный и модифицированный варианты нуклеотидной последовательности трансгена в файле с помощью кнопки **Сохранить (Save all)**.

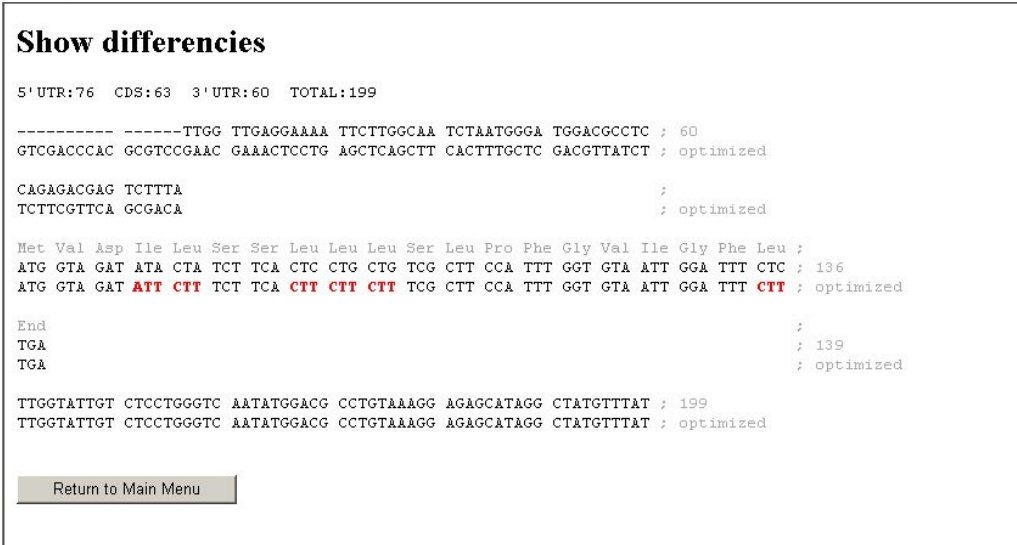


Рисунок 27.

2.2.2. Программные компоненты, позволяющие предсказать эффективность экспрессии трансгена в клетках растения-реципиента.

Известно, что некоторые чужеродные гены экспрессировались в растениях на уровне, достаточном для решения многих экспериментальных задач, не связанных с суперпродукцией трансгенного белка. Однако, в ряде случаев заметная экспрессия трансгена может быть достигнута только после реструктуризации ДНК трансгена и приближения ее характеристик к генам растений (например, в случае *Bt*-токсина). Таким образом, предсказание эффективности экспрессии трансгена в клетках растений может быть важно и для решения фундаментальных задач молекулярной генетики растений с помощью экспериментов на трансгенных растениях, и для целей практической биотехнологии, поскольку модификация структуры трансгена является достаточно дорогостоящей процедурой.

Следует отметить, что использованные компьютерные технологии предназначены не только для учета известных трансляционно-значимых характеристик, потому что только с их помощью предсказать эффективность трансляции мРНК нельзя. Эти методы также учитывают спектры характеристики, характерные для видоспецифических генов высокого уровня экспрессии, биологический смысл которых пока не известен. Компьютерный анализ выборки генов высокого и низкого уровней экспрессии проводился с помощью разных подходов (позиционные частоты олигонуклеотидов, весовые матрицы, усредненное распознавание), реализованный в разных программных компонентах.

2.2.2.1. Предсказание эффективности сигналов инициации и терминации трансляции (InTerOpt).

Программная компонента *InTerOpt* предназначена для оценки сходства сигналов инициации и терминации трансляции с сигналами, характерными для высокоэкспрессирующихся генов растения-реципиента трансгена. Сигналы инициации и терминации трансляции являются одним из ключевых регуляторных элементов в структуре мРНК. Известно, что рибосомы в клетках растений движутся по мРНК в направлении 3'-конца матрицы в поиске стартового кодона трансляции (в большинстве случаев, AUG). Эффективность распознавания AUG в качестве стартового кодона зависит от его нуклеотидного окружения (контекста): считается, что пурины в -3 и гуанин в +4 положениях вокруг AUG обеспечивают его эффективное распознавание рибосомами (Lukaszewicz et al., 2000; Kozak, 2002). Однако, в настоящее время известны далеко не все сигналы, определяющие параметры сайта инициации трансляции (например, Sawant et al., 2001). Кроме того, возможно существование видоспецифических сигналов. Принцип организации программной компоненты *InTerOpt* основан на сравнении контекстных характеристик мРНК (олигонуклеотидных словарей) в районе сигналов инициации и терминации трансляции мРНК высокого и низкого уровней экспрессии, причем этот анализ был проведен для нескольких видов растений, на которых генно-инженерные эксперименты проводятся наиболее часто. В результате этого анализа были определены характеристики, позволяющие дискриминировать сигналы генов высокого и низкого уровней экспрессии, а также относить сигналы в составе мРНК чужеродного гена к одной из этих групп. Следует учесть, что функциональный смысл многих выявленных параметров (позиционных частот олигонуклеотидов) пока не известен. Предложенный подход позволяет обойти это ограничение.

2.2.2.1.1. Описание работы InTerOpt

При работе программы *InTerOp* осуществляется оценка эффективности сигналов инициации и терминации трансляции в определенных генетически-модифицированных организмах (выбор производится из списка: *Zea mays*, *Lycopersicon esculentum*, *Nicotiana tabacum*, *Hordeum vulgare*, *Triticum aestivum*, *Pisum sativum*, *Oryza sativa*, *Arabidopsis thaliana*;

также опционально доступны *Homo sapiens* и *Mus musculus*). Для предсказания необходимо ввести нуклеотидные последовательности мРНК трангена размерами не менее: 5'-НТП - 45, CDS - 150, 3'-НТП – 70 нуклеотидов.

Пользователю необходимо загрузить html-страницу с интерфейсом п/с «Трансгенез», нажать кнопку «**Эукариоты**» и загрузить нуклеотидную последовательность таким образом, как это описано в пункте «Операции для ввода данных в подсистему «Трансгенез»».

На странице выбора программ выбирается «**InTerOpt**: предсказание эффективности и оптимизация сигналов инициации и терминции трансляции (InTerOpt: prediction efficiency and optimization of translation start/termination sites)». Пользователь также имеет возможность выбора организма, в котором будет экспрессироваться транген (выбор осуществляется из списка *Zea mays*, *Lycopersicon esculentum*, *Nicotiana tabacum*, *Hordeum vulgare*, *Triticum aestivum*, *Pisum sativum*, *Oryza sativa*, *Arabidopsis thaliana*; также доступны *Homo sapiens* и *Mus musculus*). Кроме этого пользователь выбирает район, в котором осуществляется оптимизация (в данном случае **сигнал инициации трансляции**) и нажимает кнопку «Запустить».



Рисунок 28.

В результате генерируется html-страница, на которой: (а) в случае, если расположенный в составе мРНК сигнал инициации или терминции трансляции предсказан как эффективный, выдается только предсказание; (б) в случае, если сигнал инициации трансляции предсказан как неэффективный, отображаются предсказание и два варианта нуклеотидных последовательностей мРНК трангена (нативная и оптимизированная), выровненные друг относительно друга в районе CDS.

InTerOpt: prediction efficiency and optimization of translation start/termination sites

5'UTR:76 CDS:150 3'UTR:56 TOTAL:282

Sequence of initiation signal is optimized, seq_weight=6.03 >= border=2.0

[Return to Main Menu](#)

Рисунок 29.

InTerOpt: prediction efficiency and optimization of translation start/termination sites

```
5'UTR:56 CDS:150 3'UTR:56 TOTAL:262

Sequence of initiation signal is not optimized, seq_weight=-41.070001 < border=2.0

----- ATGAAAATCA ACCCCGGACT CGAGCCGTAC ACCAATGCTA ; 56
GTCGACCCAC GCGTCCGAAC GAAACTCCTG AGCTCAGCTT CACTTTGCTC GACGTTATCT ; optimized

GCATTGGACA GAATGA ;
TCTTCGTTCA GCGACA ; optimized

Met Leu Trp Thr Asp Cys Leu Thr Arg Leu Arg Gln Glu Leu Ser Asp Asn Val Phe Ala ;
ATG CTT TGG ACA GAC TGC TTA ACT CGC TTG CGA CAA GAG CTA TCT GAT AAC GTC TTT GCG ; 116
ATG CTT TGG ACA GAC TGC TTA ACT CGC TTG CGA CAA GAG CTA TCT GAT AAC GTC TTT GCG ; optimized

Met Trp Ile Arg Pro Leu Val Ala Glu Glu Thr Thr Asp Ser Leu Arg Leu Tyr Ala Pro ;
ATG TGG ATT CGT CCT TTA GTA GCT GAA GAG ACG ACA GAT AGT CTA CGC CTG TAT GCG CCG ; 176
ATG TGG ATT CGT CCT TTA GTA GCT GAA GAG ACG ACA GAT AGT CTA CGC CTG TAT GCG CCG ; optimized

Asn Pro Tyr Trp Thr Arg Tyr Ile Gln End ;
AAC CCT TAC TGG ACG CGT TAT ATT CAG TGA ; 206
AAC CCT TAC TGG ACG CGT TAT ATT CAG TGA ; optimized

ATGAAAATCA ACCCCGGACT CGAGCCGTAC ACCAATGCTA GCATTGGACA GAATGA ; 262
ATGAAAATCA ACCCCGGACT CGAGCCGTAC ACCAATGCTA GCATTGGACA GAATGA ; optimized

Sequence of initiation signal is not optimized, seq_weight=-41.070001 < border=2.0

Save & Main Menu Drop & Main Menu
```

Рисунок 30.

Далее пользователь имеет возможность либо сохранить модифицированный вариант нуклеотидной последовательности (нажатием «**Save & Main Menu**», дальнейший анализ будет осуществляться с этой уже модифицированной нуклеотидной последовательностью трансгена), либо отказаться от модификаций (нажатие «**Drop & Main Menu**», дальнейший анализ будет осуществляться с использованием нативного варианта трансгена).

После выполнения запланированных процедур пользователь имеет возможность визуализировать различия между исходным вариантом нуклеотидной последовательности трансгена и конечным вариантом модифицированной последовательности (нажатие **Show Differences - Показать различия**), а также пользователь может сохранить исходный и модифицированный варианты нуклеотидной последовательности трансгена в файле с помощью кнопки «**Save all – Сохранить**».

Show differences

```
5'UTR:76 CDS:150 3'UTR:56 TOTAL:282

----- ATGAAAATCA ACCCCGGACT CGAGCCGTAC ACCAATGCTA ; 56
GTCGACCCAC GCGTCCGAAC GAAACTCCTG AGCTCAGCTT CACTTTGCTC GACGTTATCT ; optimized

GCATTGGACA GAATGA ;
TCTTCGTTCA GCGACA ; optimized

Met Leu Trp Thr Asp Cys Leu Thr Arg Leu Arg Gln Glu Leu Ser Asp Asn Val Phe Ala ;
ATG CTT TGG ACA GAC TGC TTA ACT CGC TTG CGA CAA GAG CTA TCT GAT AAC GTC TTT GCG ; 136
ATG CTT TGG ACA GAC TGC TTA ACT CGC TTG CGA CAA GAG CTA TCT GAT AAC GTC TTT GCG ; optimized

Met Trp Ile Arg Pro Leu Val Ala Glu Glu Thr Thr Asp Ser Leu Arg Leu Tyr Ala Pro ;
ATG TGG ATT CGT CCT TTA GTA GCT GAA GAG ACG ACA GAT AGT CTA CGC CTG TAT GCG CCG ; 196
ATG TGG ATT CGT CCT TTA GTA GCT GAA GAG ACG ACA GAT AGT CTA CGC CTG TAT GCG CCG ; optimized

Asn Pro Tyr Trp Thr Arg Tyr Ile Gln End ;
AAC CCT TAC TGG ACG CGT TAT ATT CAG TGA ; 226
AAC CCT TAC TGG ACG CGT TAT ATT CAG TGA ; optimized

ATGAAAATCA ACCCCGGACT CGAGCCGTAC ACCAATGCTA GCATTGGACA GAATGA ; 282
ATGAAAATCA ACCCCGGACT CGAGCCGTAC ACCAATGCTA GCATTGGACA GAATGA ; optimized

Return to Main Menu
```

Рисунок 31.

2.2.2.2. Предсказание эффективности трансляции мРНК трансгена на основе полного анализа контекстных характеристик (Full_mRNA)

Программа **Full_mRNA** представляет собой новую (расширенную) реализацию предложенного нами ранее подхода (программы **Leader_RNA**, основанной на сравнительном анализе контекстных характеристик лидерных районов мРНК генов эукариот (Kochetov et al., 1999)). Известно, что различные функциональные районы мРНК генов растений обладают различной контекстной организацией, что может быть непосредственно связано с выполняемыми ими функциями. Ранее нами было показано, что параметры мРНК генов высокого и низкого уровней экспрессии эукариот достоверно различаются по ряду характеристик (Kochetov et al., 1998). Программа анализирует 5'-НТП, белок-кодирующую часть и 3'-НТП мРНК трансгена и сравнивает их с характеристиками, достоверно различающимися у генов высокого и низкого уровней экспрессии растений. Общая оценка эффективности трансляции нативного варианта мРНК трансгена в клетках растений делается на основе метода «усредненного распознавания» отдельно для каждого из этих функциональных районов и, кроме того, делается общий анализ для полноразмерной нуклеотидной последовательности мРНК. Такое предсказание может применяться в двух целях:

Во-первых, для предсказания уровня экспрессии собственных генов растений. Возможны ситуации, в которых гены характеризуются высоким уровнем экспрессии на отдельной стадии развития или в тканеспецифично. Если нет соответствующей экспериментальной информации (например, данных по профилированию транскриптов) для данной ткани на данной стадии развития, то уровень экспрессии гена определить нельзя. Однако, если программа **Full_mRNA** предскажет высокое сходство гена с другими высокоэкспрессирующимися генами растений, это может быть использовано в качестве весомого аргумента в пользу адаптации структуры гена для поддержания высокого уровня экспрессии и может быть далее использовано в планировании исследования паттерна экспрессии данного гена.

Во-вторых, модификация нуклеотидной последовательности трансгена для повышения уровня экспрессии в клетках растений представляет собой дорогостоящую процедуру. Если программа **Full_mRNA** предскажет сходство контекстной организации мРНК трансгена, имеет смысл экспериментально проверить эффективность экспрессии нативного варианта трансгена в растениях (например, с помощью транзитной экспрессии в листьях методом агроинфильтрации) и уже после этого принимать решение о необходимости внесения модификации или об использовании немодифицированного варианта.

2.2.2.2.1 Описание работы *Full_mRNA*

Программа **Full_mRNA** предназначена для предсказания эффективности трансляции мРНК трансгена в клетках организма – реципиента, основанного на сравнительном анализе мРНК генов высокого и низкого уровней экспрессии. Для работы программы необходимо ввести нуклеотидные последовательности мРНК трансгена размерами не менее: 5'-НТП - 35, CDS - 60, 3'-НТП – 30 нуклеотидов.

После загрузки html-страницы с интерфейсом п/с «Трансгенез», нажать кнопку «**Эукариоты**» и загрузить нуклеотидную последовательность таким образом, как это описано в пункте «Операции для ввода данных в подсистему «Трансгенез»».

На странице выбора программ выбирается «Full_mRNA»: предсказание трансляционной активности мРНК растений, основанное на анализе контекстных особенностей функциональных районов (Full_mRNA: prediction of translation efficiency of plant mRNAs based on the analysis of contextual features of functional regions)».



Рисунок 32.

Пользователь определяет, в каких растениях (двудольных или однодольных) будет экспрессироваться данный трансген и нажимает кнопку RUN. В результате генерируется html-страница, на которой представлены результаты предсказания трансляционной активности мРНК в целом (суммарное предсказание по 5'-НТП, CDS и 3'-НТП), а также предсказания для каждого из этих функциональных районов по отдельности.

Full_mRNA: prediction of translation efficiency of plant mRNAs based on the analysis of contextual features of functional regions

Full_mRNA makes a prediction of mRNA translational efficiency in plant cells. For this purpose, the program evaluates the similarity of a user defined transgene mRNA (5'-UTR, CDS, 3'-UTR) with plant mRNAs encoded by either highly or lowly expressed genes. The discriminative features were revealed through a comparative analysis of high and low expression samples of dicot and monocot genes. Three regions (5'-UTR, CDS, and 3'-UTR) were analyzed and the features characteristic for high and low expression gene sets were recorded. The numeric values characteristic for high expression genes were assigned as +1, and for low expression genes, as -1. The corresponding features of mRNA under evaluation are approximated on the segment from -1 to +1. Total evaluation of the translational efficiency is made on the basis of mean recognition approach (described in detail in: Kochetov A.V., Ponomarenko M.P., Frolov A.S., Kisselev L.L., Kolchanov N.A. Prediction of eukaryotic mRNA translational properties. Bioinformatics. 1999. 15. 704-712) and may vary from -1 (typical low expression mRNA) to +1 (typical high expression mRNA). If the resulting value >0 but <+1 this means that higher number of discriminating features resembles those characteristic for high expression genes, and vice versa. The program output presents four values (weights): average for full mRNA, and for its functional regions (5'-UTR, CDS, 3'UTR). This output allows user to select the most unfavorable region for optimization.

Information on the mRNA inputted: 5'UTR:76 CDS:150 3'UTR:56 TOTAL:282

Average weight for full mRNA:	-0.287592
Average weight for 5'-UTR:	0.267230
Average weight for CDS:	-0.369988
Average weight for 3'-UTR:	-0.760018

[Return to Main Menu](#)

Рисунок 33.

Красным цветом выделены значения, соответствующие предсказанным эффективным функциональным районам, синим цветом – предсказанным неэффективным. Последнее может быть использовано пользователем в качестве экспертной оценки, рекомендующей реконструировать нуклеотидную последовательность соответствующих функциональных районов трансгена для повышения уровня его экспрессии. После получения информации нужно нажать кнопку «Return to Main Menu».

После выполнения запланированных процедур пользователь имеет возможность сохранить результаты работы программы с помощью стандартных средств интернет-браузера в виде текстового или html-файла.

2.2.2.3. Предсказание трансляционной активности 5'-НТП мРНК трансгена на основе анализа структурных характеристик (Leader_hairpin)

Хорошо известно, что мРНК в цитоплазме способна формировать различные вторичные структуры за счет комплементарных взаимодействий. Предсказание этих структур (в

особенности с учетом РНК-белковых взаимодействия) – это сложная задача, не решенная до конца в настоящее время. Тем не менее, особенности вторичной структуры в лидерном районе мРНК, влияющие на эффективность инициации трансляции, могут быть оценены с помощью методов компьютерного анализа. Экспериментально показано, что вторичная структура в 5'-НТП негативно сказывается на трансляционной активности мРНК (Kozak, 2002). При этом особенно негативное влияние (в аспекте взаимосвязи ингибирующего эффекта и стабильности вторичной структуры) проявляли шпильки, расположенные на небольшом расстоянии от 5'-конца мРНК, что, по-видимому, может быть связано с ингибированием взаимодействия кеп-и кеп-связывающего комплекса трансляционных факторов. Следует отметить, что в клетках разных организмов этот эффект может варьировать, что связано с особенностью организации аппарата инициации трансляции (например, активность геликаз). Так, мРНК млекопитающих и однодольных растений характеризуются значительно более высоким содержанием G+C и, соответственно, более высоким потенциалом формирования вторичной структуры, чем мРНК двудольных растений. Таким образом, структурные характеристики мРНК трансгенов могут быть неоптимальными и высокий уровень экспрессии достигнут не будет.

Программа **Leader_hairpin** предназначена для оценки структурных параметров 5'-НТП мРНК трансгена. Программа оценивает как полную вторичную структуру, так и ее элементы, расположенные в 5'-концевой части лидерной последовательности. Если 5'-концевая часть неструктурирована, это может означать высокую эффективность взаимодействия мРНК и кеп-связывающего комплекса и, как следствие, высокую эффективность этой стадии процесса инициации трансляции.

Программа может применяться как для предсказания эффективности трансляции мРНК собственных генов растений, так и для оценки трансляционной активности мРНК трансгена в клетках растений.

2.2.2.3.1. Описание работы *Leader_hairpin*

Предсказание эффективности инициации трансляции трансгена программой **Leader_hairpin** осуществляется на основе анализа вторичной структуры мРНК в районе 5'-НТП.

Для анализа необходимо ввести нуклеотидные последовательности мРНК трансгена размерами не менее: 5'-НТП – 30 нуклеотидов, 5'-концевой фрагмент CDS – не менее 70 нуклеотидов. На странице выбора программ выбирается **Leader_hairpin**: предсказание трансляционной активности 5'-НТП (**Leader_Hairpin**: prediction of 5'-UTR translational properties). Пользователь также может выбрать размер 5'-концевого сегмента (по умолчанию принят равным 15 нуклеотидам), вторичная структура в котором может наиболее сильно ингибировать инициацию трансляции.



Рисунок 34.

Нажимается кнопка «Запустить». В результате генерируется html-страница, на которой представлены результаты предсказания стабильности вторичной структуры протяженного 5'-концевого участка мРНК, а также вторичной структуры, локализованной в 5'-терминальном участке, границы которого определены пользователем.

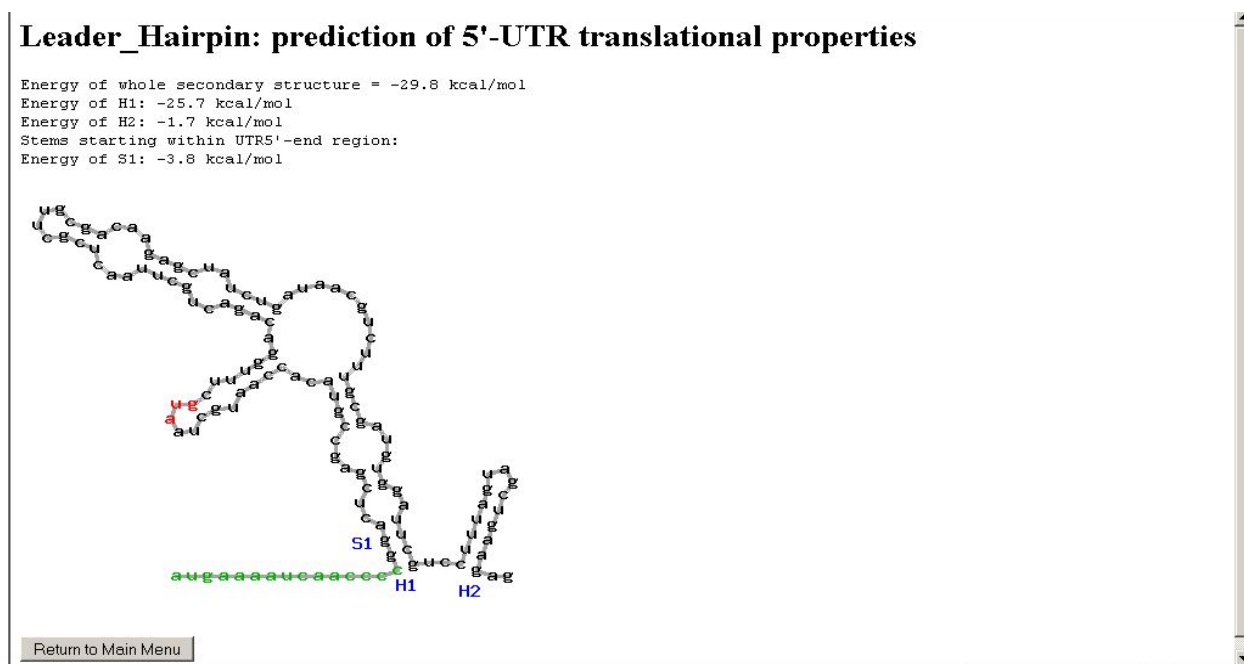


Рисунок 35.

Эта экспертная информация позволяет пользователю оценить вклад структурных характеристик 5'-НТП в эффективность инициации трансляции. Также приводится изображение предсказанной вторичной структуры с обозначенным стартовым кодоном трансляции (выделен красным цветом) и 5'-терминальным участком (выделен зеленым цветом). После получения информации нужно нажать кнопку **Return to Main Menu**.

2.2.2.3.2 .Описание работы AUG_hairpin

На странице выбора программ выбирается **AUG_hairpin**: предсказание эффективности распознавания стартового AUG кодона (**AUG_Hairpin**: prediction of recognition efficiency of start AUG codon). Эта программа предназначена для предсказания вторичной структуры, влияющей на распознавание стартового кодона трансляции, расположенного в слабом контексте.



Рисунок 36.

Такая структура должна начинаться в границах определенного «критического участка», расположенного на определенном расстоянии ниже стартового кодона AUG. Пользователь имеет возможность изменить эти границы в пределах от 1 до 30-го нуклеотида, по умолчанию этот промежуток расположен между 12 и 18 нуклеотидами («5'-граница

интервала, в котором будет анализироваться вторичная структура» или «**3'-граница** интервала, в котором будет анализироваться вторичная структура», соответственно). Пользователь также имеет возможность определить, будет ли программа учитывать только совершенные шпильки (опция «**Принимать в расчет только совершенные шпильки**»; по умолчанию выключена); в противном случае пользователь имеет возможность ограничить максимально возможный размер внутренних и боковых петель (по умолчанию = 3). После выбора этих опций, нажать кнопку «**RUN**». В результате генерируется html-страница, на которой представлены результаты предсказания стабильности вторичной структуры в 5'-концевом участке мРНК.

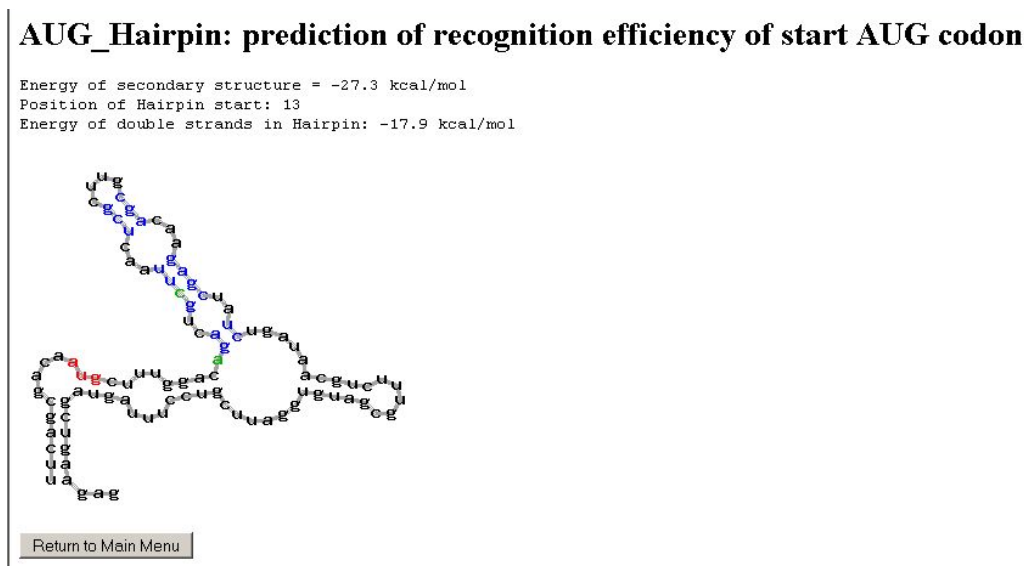


Рисунок 37.

Эта экспертная информация позволяет пользователю оценить вклад структурных характеристик мРНК в распознавании «слабого» стартового кодона трансляции. Также приводится изображение предсказанной вторичной структуры с обозначенным стартовым кодоном трансляции (выделен красным цветом). Если вторичная структура начинается в границах «критического участка» (выделены зеленым), стебли выделены синим цветом. Также приведена предсказанная стабильность вторичной структуры 5'-концевого участка, позиция начала стабильной шпильки (если таковая есть) и ее энергия. После получения информации нажать кнопку **Return to Main Menu**.

После выполнения запланированных процедур пользователь имеет возможность сохранить результаты работы программы с помощью стандартных средств интернет-браузера в виде html-файла. Изображение вторичной структуры генерируется в виде файла в формате jpeg, который также может быть сохранен независимо.

2.2.2.4. Предсказание синонимических кодонов, повышающих эффективность элонгации в клетках растения-реципиента

Известно, что кодонный состав влияет на скорость элонгации трансляции у многих прокариот и дрожжей (McCarthy, 1998). Однако, для мРНК растений и млекопитающих роль этих факторов в определении скорости элонгации трансляции до настоящего времени не изучена. Получены некоторые данные о взаимосвязях между кодонным составом генов растений и их функцией или уровнем экспрессии (Chiapello et al., 1998; Duret and Mouchiroud, 1999), но в целом эта проблема также исследована недостаточно и выявленные

закономерности достаточно слабы. По-видимому, существенным фактором может оказаться взаимосвязь между локализацией редких кодонов и структурой белка (Komar et al., 1999) или стабильностью мРНК (van Hoof and Green, 1997).

Несмотря на отсутствие доказательств взаимосвязи между кодонным составом и эффективностью экспрессии, оптимизацию кодонного состава активно используют для повышения уровня экспрессии трансгенов (Khanna et al., 2002; Biemelt et al., 2003). При этом в качестве оптимальных кодонов используют наиболее часто встречающиеся синонимические кодоны генов организма-реципиента (таблицы средних частот кодонов доступны, например, в БД TransTerm (Jacobs et al., 2002) или CUTG). Считается, что такие модификации могут повысить эффективность экспрессии трансгена. Однако, реорганизация контекстных характеристик БКП может влиять и на другие параметры экспрессии трансгена: например, при этом могут элиминироваться ложные сигналы сплайсинга и полиаденилирования. Это весьма вероятно, поскольку чужеродные гены часто существенно отличаются от генов организма-реципиента по нуклеотидному составу. В целом, оптимизация кодонного состава, которая обычно осуществляется с помощью синтеза нуклеотидной последовательности из олигонуклеотидов, является наиболее дорогостоящей процедурой оптимизации экспрессии и применяется либо для максимизации уровня экспрессии трансгена (суперпродукция белка, «живые вакцины» и т.п.), либо для разрешения особенно сложных случаев (Bt-токсин, зеленый флюоресцирующий белок и т.п.).

Программа **RSCU_comparer** предназначена для сравнительного анализа частот синонимических кодонов в белок-кодирующих частях мРНК генов растений. Программа сравнивает гены высокого и низкого уровней экспрессии и определяет кодоны, частоты которых достоверно различаются в этих выборках, а также общую достоверность различий в паттерне использования синонимических кодонов. В результате работы программы пользователь получает информацию о наиболее и наименее оптимальных синонимических кодонах, которая может быть использована в качестве исходных данных для работы программы оптимизации SynCodeOpt.

2.2.2.4.1. Описание работы RSCU_comparer

Для работы с программой пользователю необходимо иметь выборки белок-кодирующие последовательностей мРНК генов высокого и низкого уровней экспрессии организма-реципиента трансгенеза. Выборки формируются пользователем самостоятельно. Пользователь должен загрузить html-страницу с интерфейсом п/с «Трансгенез», нажать кнопку «**Эукариоты**» и выбрать программу: «**RSCU_comparer**»: сравнение генов высокого и низкого уровней экспрессии для выявления оптимальных и неоптимальных синонимических кодонов (**RSCU_comparer**: comparison of high and low expression genes to reveal optimal/suboptimal synonymous codons)», после чего нажать кнопку «**Запустить**».



Рисунок 38.

Открывается окно программы, в котором есть окна для ввода выборок белок-кодирующих последовательностей генов высокого и низкого уровней экспрессии организма-реципиента в fasta-формате.

RSCU_comparer: сравнительный анализ генов высокого и низкого уровней экспрессии для выявления оптимальных/неоптимальных синонимических кодонов

Программа сравнивает паттерн использования синонимических кодонов в выборках генов организма реципиента-трансгена, характеризующихся высоким и низким уровнями экспрессии (выборки готовит пользователь на основе информации об организме-реципиенте, доступной из литературы). Достоверные различия в частотах синонимических кодонов будут означать, что существует взаимосвязь между эффективностью экспрессии гена и интенсивностью элонгации трансляции. Программа RSCU_comparer может быть использована для целей практической биотехнологии: выбор оптимальных кодонов является существенным элементом при модификации нуклеотидной последовательности трансгена для максимизации его экспрессии.

Введите нуклеотидную последовательность белок-кодирующей части мРНК (в окно на экране или из файла):

Гены высокого уровня экспрессии:

Гены низкого уровня экспрессии:

Pmax, уровень значимости различий:

Qsub, порог для селекции неоптимальных кодонов:

Qopt, порог для селекции оптимальных кодонов:

Рисунок 39.

Также есть возможность для ввода этих выборок из файлов. Помимо этого пользователь имеет возможность выбора параметров программы: уровня достоверности различий в использовании синонимических кодонов в анализируемых выборках (по умолчанию = 0.05), границы групп оптимальных (более характерных для высокоэкспрессирующихся генов) и субоптимальных (более характерных для низкоэкспрессирующихся генов) синонимических кодонов (по умолчанию = 0.1). Границы параметров могут меняться в зависимости от размера выборок High и Low; рекомендованный диапазон значений в данном случае – от 0.01 до 0.25. Пользователь вводит нуклеотидные последовательности и – при необходимости — изменяет значения этих параметров, после чего нажимает кнопку «Запустить». Генерируется html-страница, на которой приведены следующие данные:

- оценка нормальности распределений значений индекса RSCU для проанализированных выборок высоко- и низкоэкспрессирующихся генов, а также разности между этим значениями;
- оценка достоверности различий между выборками;
- списки эффективных и неэффективных синонимических кодонов.

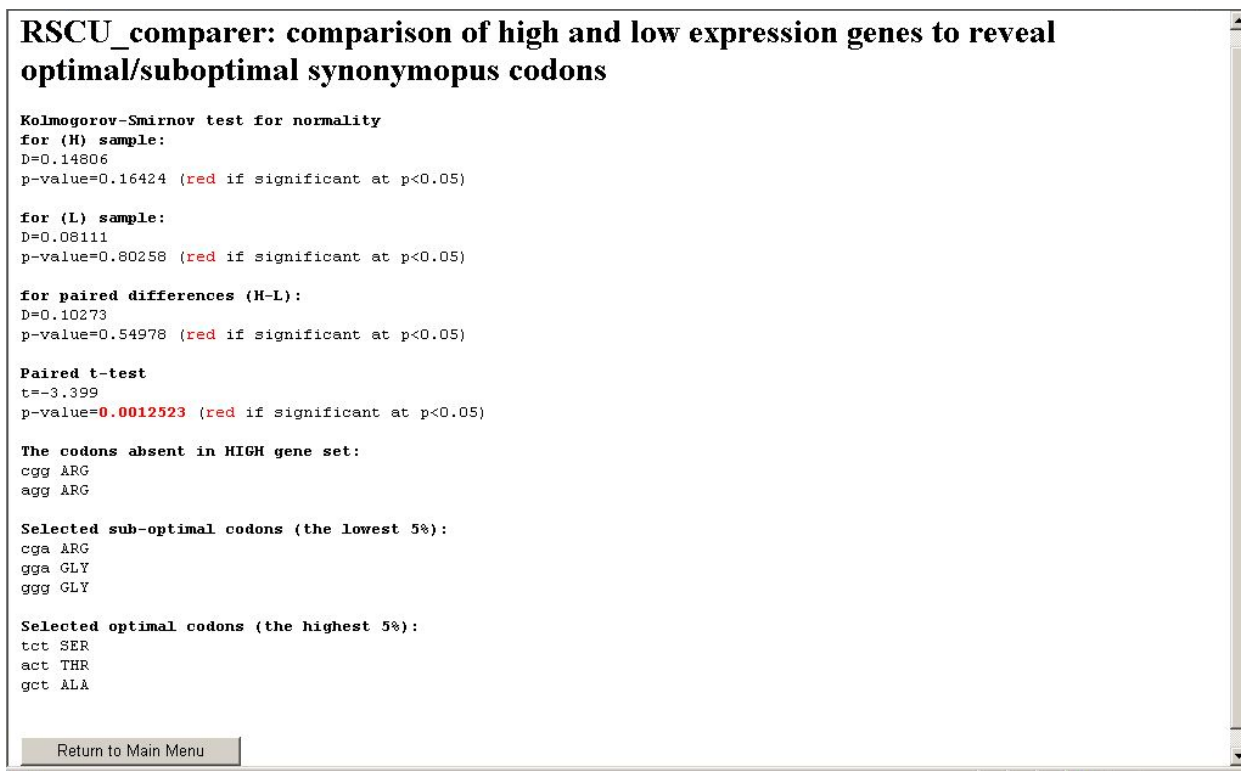


Рисунок 40.

Эта информация для дальнейшего использования должна сохраняться пользователем в отдельном файле; она может в дальнейшем использоваться в качестве независимого источника данных для работы программы «SynCodeOpt».

2.2.3. Программные компоненты для оптимизации структуры транскена с целью максимизации его экспрессии.

2.2.3.1. Оптимизации кодонного состава транскена

Программа **SynCodeOpt** предназначена для внесения модификаций в кодонный состав белок-кодирующей последовательности мРНК транскена (биологическое обоснование необходимости такой оптимизации приведено в описании программы RSCU_comparer). Программа предоставляет пользователю экспертную информацию о частотах синонимических кодонов в генах определенной группы растений, для которых такие данные были получены нами ранее. В случае, если такой информации о растении-реципиенте в программе нет, ее можно получить с помощью программной компоненты RSCU_comparer.

В п/с «Трансгенез» возможна оценка эффективности сигналов инициации и терминции трансляции в определенных генетически-модифицированных организмах (выбор осуществляется из списка *Zea mays*, *Solanum tuberosum*, *Lycopersicon esculentum*, *Nicotiana tabacum*, *Spinacia oleracea*, *Hordeum vulgare*, *Triticum aestivum*, *Daucus carota*, *Sorghum bicolor*, *Medicago sativa*, *Pisum sativum*, *Oryza sativa*, *Arabidopsis thaliana*). В случае, если организма-реципиента в этом списке нет, есть возможность ввода дополнительной информации из файла, содержащего данные о средних частотах кодонов в генах данного организма (выбирается опция Use CUTG file и производится загрузка файла с компьютера пользователя). В этом случае пользователь должен подготовить файл самостоятельно.

2.2.3.1.1. Описание работы программы SynCodeOpt

Пользователь должен загрузить html-страницу с интерфейсом п/с «Трансгенез», нажать кнопку «**Эукариоты**» и загрузить нуклеотидную последовательность таким образом, как это описано в пункте «Операции для ввода данных в подсистему «Трансгенез»».

На странице выбора программ выбирается «**SynCodeOpt**»: оптимизация кодонного состава белок-кодирующей части мРНК (**SynCodeOpt**: optimization of mRNA codon content).

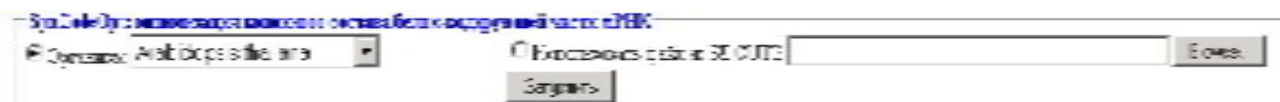


Рисунок 41.

Кроме этого необходимо выбрать организм-реципиент трансгена (выбор осуществляется из списка *Zea mays*, *Solanum tuberosum*, *Lycopersicon esculentum*, *Nicotiana tabacum*, *Spinacia oleracea*, *Hordeum vulgare*, *Triticum aestivum*, *Daucus carota*, *Sorghum bicolor*, *Medicago sativa*, *Pisum sativum*, *Oryza sativa*, *Arabidopsis thaliana*). В случае, если организма-реципиента в этом списке нет, есть возможность ввода дополнительной информации из файла, содержащего данные о средних частотах кодонов в генах данного организма (выбирается опция **Использовать файл из БД CUTG** и производится загрузка файла с компьютера пользователя).

UUU 12.5 (1560)	UCU 11.3 (1409)	UAU 9.3 (1160)	UGU 6.4 (798)
UUC 27.8 (3461)	UCC 17.2 (2143)	UAC 19.5 (2424)	UGC 13.6 (1689)
UUA 5.0 (621)	UCA 10.1 (1255)	UAA 0.7 (86)	UGA 1.2 (146)
UUG 12.9 (1604)	UCG 8.8 (1097)	UAG 0.6 (80)	UGG 14.2 (1766)
CUU 15.6 (1941)	CCU 12.0 (1496)	CAU 10.0 (1240)	CGU 6.5 (806)
CUC 28.4 (3536)	CCC 13.5 (1684)	CAC 13.2 (1646)	CGC 11.3 (1405)
CUA 7.6 (949)	CCA 13.7 (1706)	CAA 13.6 (1697)	CGA 3.5 (439)
CUG 23.1 (2874)	CCG 15.1 (1875)	CAG 24.0 (2985)	CGG 9.7 (1206)
AUU 14.5 (1803)	ACU 9.2 (1145)	AAU 13.8 (1714)	AGU 7.6 (946)
AUC 25.0 (3107)	ACC 18.8 (2344)	AAC 26.9 (3345)	AGC 17.2 (2145)
AUA 8.8 (1091)	ACA 9.7 (1206)	AAA 13.7 (1699)	AGA 7.7 (960)
AUG 25.0 (3116)	ACG 11.3 (1407)	AAG 38.1 (4748)	AGG 12.7 (1580)
GUU 14.7 (1825)	GCU 18.7 (2323)	GAU 22.3 (2773)	GGU 15.9 (1974)
GUC 22.6 (2810)	GCC 32.8 (4088)	GAC 32.7 (4070)	GGC 32.4 (4028)
GUA 6.0 (751)	GCA 15.1 (1877)	GAA 17.7 (2209)	GGA 13.5 (1679)
GUG 26.8 (3341)	GCG 21.6 (2687)	GAG 37.9 (4715)	GGG 17.4 (2168)

Рисунок 42. Образец файла с значениями индекса CUTG (источник - <http://www.kazusa.or.jp/codon/>).

После выбора организма-реципиента трансгена из списка или загрузки файла в формате CUTG нажимается кнопка «Запустить». Открывается окно с экспертной информацией о кодонном составе в генах данного организма-реципиента трансгена: среднее значение RSCU для каждого синонимического кодона (Average RSCU in sequence under analysis); средняя частота кодона в нуклеотидной последовательности трансгена (Average codon frequency in sequence under analysis); средние частоты синонимических кодонов из БД CUTG (Average species-specific codon freq (CUTG)); а также доступны средние значения RSCU для кодонов в выборках высокого и низкого уровней экспрессии (Average RSCU in a sample of high expression genes; Average RSCU in a sample of low expression genes, соответственно); разница в значениях RSCU между генами высокого и низкого уровней экспрессии (RSCU(H-L)).

SynCodeOpt: Оптимизация кодонного состава белок-кодирующей части мРНК

Выберите неоптимизированные синонимические кодоны для их замещения.

Таблица содержит информацию о (1) Содержании синонимических кодонов в белок-кодирующей части трансгена (индекс RSCU), (2) Информацию о средних частотах кодонов в генах выбранного организма-реципиента (взяты из БД CUTG), Средние значения индекса RSCU, характерные для генов высокого и низкого уровней экспрессии выбранного организма (если они есть в БД программы).

Кнопка замещает все синонимические кодоны в группе на наиболее часто встречающийся вариант в БД CUTG.

Организм: Arabidopsis thaliana							
Синонимические кодоны	Средние значения индекса RSCU в анализируемой мРНК	Средние частоты кодонов в анализируемой мРНК	Средние видоспецифические значения индекса RSCU	Средние значения индекса RSCU в выборке генов высокого уровня экспрессии	Средние значения индекса RSCU в выборке генов низкого уровня экспрессии	RSCU(H-L)	Выбор кодонов для замещения
Ala							
GCA	1.255	9.1	17.5	0.786	1.051	-0.265	заменить на наилучший
GCC	0.784	5.7	10.4	0.802	0.636	0.166	заменить на наилучший
GCG	0.392	2.8	9.0	0.372	0.577	-0.205	заменить на наилучший
GCT	1.569	11.4	28.4	2.040	1.736	0.304	заменить на наилучший
Arg							
AGA	1.920	4.6	18.9	1.726	2.296	-0.570	заменить на наилучший
AGG	0.480	1.1	10.9	1.689	1.270	0.419	заменить на наилучший
CGA	0.480	1.1	6.3	0.300	0.654	-0.354	заменить на наилучший
CGC	1.440	3.4	3.8	0.399	0.354	0.045	заменить на наилучший
CGG	0.720	1.7	4.9	0.213	0.513	-0.300	заменить на наилучший
CGT	0.960	2.3	9.0	1.673	0.913	0.760	заменить на наилучший
Asn							
AAC	0.600	3.4	20.9	1.372	1.020	0.352	заменить на наилучший
AAT	1.400	8.0	22.3	0.628	0.980	-0.352	заменить на наилучший
Asp							
GAC	0.200	1.1	17.3	0.857	0.619	0.238	заменить на наилучший
GAT	1.800	10.3	36.7	1.143	1.381	-0.238	заменить на наилучший
Cys							
TGC	1.200	1.7	7.1	0.948	0.742	0.206	заменить на наилучший
TGT	0.800	1.1	10.5	1.052	1.258	-0.206	заменить на наилучший

Рисунок 43.

На этой странице пользователь имеет возможность выбора паттерна замещения в группах синонимических кодонов – например, кодон GCA может быть заменен на любой из группы {GCA, GCG, GCC, GCU} – для этого используются меню выбора в каждой их синонимических групп. По умолчанию синонимические замены не проводятся. Преимущество должны иметь кодоны с более высокими значениями индекса RSCU в группе высокоэкспрессирующихся генов и более высокой средней частотой в синонимической группе (CUTG). Для проведения оптимизации пользователь выбирает кодоны (рекомендуется заменять наиболее неоптимальные кодоны, которые легко отобрать на основе вышеописанной информации) и нажимает на кнопку справа от синонимической группы, что выводит ниспадающее меню с выбором кодона для замены.

Ala							best CUTG
GCA	0.333	2.1	17.5	0.786	1.051	-0.265	GCA ▼
GCC	0.500	3.1	10.4	0.802	0.636	0.166	GCA
GCG	1.667	10.4	9.0	0.372	0.577	-0.205	GCC
GCT	1.500	9.4	28.4	2.040	1.736	0.304	GCG
Arg							GCT
							best CUTG

Рисунок 44.

Для проведения глубокой оптимизации нажимается кнопка «заменить на наилучший» («(best CUTG)»), что приводит к замене всех кодонов в синонимической группе на наиболее часто встречающийся. После выбора паттерна замещения кодонов нажимается кнопка «Оптимизировать» (Optimize).

Thr							best CUTG
ACA	0.889	10.4	15.7	0.890	1.181	-0.291	ACA ▼
ACC	0.800	9.4	10.3	1.189	0.797	0.392	ACC ▼
ACG	1.156	13.6	7.7	0.382	0.703	-0.321	ACG ▼
ACT	1.156	13.6	17.6	1.539	1.319	0.220	ACT ▼
Tyr							best CUTG
TAC	0.696	8.4	13.8	1.393	0.967	0.426	TAC ▼
TAT	1.304	15.7	14.7	0.607	1.033	-0.426	TAT ▼
Val							best CUTG
GTA	1.020	13.6	9.9	0.259	0.579	-0.320	GTA ▼
GTC	1.098	14.6	12.8	0.919	0.726	0.193	GTC ▼
GTG	0.549	7.3	17.4	1.084	1.048	0.036	GTG ▼
GTT	1.333	17.7	27.3	1.739	1.646	0.093	GTT ▼
Optimize							
Return to Main Menu							

Рисунок 45.

Генерируется html-страница, на которой расположены два варианта нуклеотидных последовательностей трансгена (нативная и оптимизированная), выровненных друг относительно друга. В оптимизированном варианте нуклеотидной последовательности красным цветом выделены оптимизированные кодоны. Далее пользователь имеет возможность либо сохранить модифицированный вариант нуклеотидной последовательности (нажатие «Save & Main Menu», дальнейший анализ будет осуществляться с уже модифицированной нуклеотидной последовательностью трансгена), либо отказаться от модификаций (нажатие «Drop & Main Menu») дальнейший анализ будет осуществляться с использованием нативного варианта трансгена).

```
SynCodeOpt: optimization of mRNA codon content

5'UTR:61 CDS:657 3'UTR:242 TOTAL:960

CAATCCACTA ACGATCCCTA ACCGAAAACA GAGTAGTCAA GAAACAGAGT ATTTTTCGA ; 60
CAATCCACTA ACGATCCCTA ACCGAAAACA GAGTAGTCAA GAAACAGAGT ATTTTTCGA ; optimized

C ; 61
C ; optimized

Met Asp Pro Phe Leu Ile Gln Ser Pro Phe Ser Gly Phe Ser Pro Glu Tyr Ser Ile Gly ;
ATG GAT CCA TTT TTA ATT CAG TCC CCA TTC TCC GGC TTC TCA CCG GAA TAT TCT ATC GGA ; 121
ATG GAT CCT TTT CTT ATT CAG TCC CCT TTC TCC GGC TTC TCA CCT GAA TAT TCT ATC GGA ; optimized

Ser Ser Pro Asp Ser Phe Ser Ser Ser Ser Ser Asn Asn Tyr Ser Leu Pro Phe Asn Glu ;
TCT TCT CCA GAT TCT TTC TCA TCC TCT TCT TCT AAC AAT TAC TCT CTT CCC TTC AAC GAG ; 181
TCT TCT CCT GAT TCT TTC TCA TCC TCT TCT TCT AAC AAT TAC TCT CTT CCT TTC AAC GAG ; optimized

Asn Asp Ser Glu Glu Met Phe Leu Tyr Gly Leu Ile Glu Gln Ser Thr Gln Gln Thr Tyr ;
AAC GAC TCA GAG GAA ATG TTT CTC TAC GGT CTA ATC GAG CAG TCC ACG CAA CAA ACC TAT ; 241
AAC GAC TCA GAG GAA ATG TTT CTT TAC GGT CTT ATC GAG CAG TCC ACT CAA CAA ACT TAT ; optimized

Ile Asp Ser Asp Ser Gln Asp Leu Pro Ile Lys Ser Val Ser Ser Arg Lys Ser Glu Lys ;
ATT GAC TCG GAT AGT CAA GAC CTT CCG ATC AAA TCC GTA AGC TCA AGA AAG TCA GAG AAG ; 301
ATT GAC TCG GAT AGT CAA GAC CTT CTT ATC AAA TCC GTA AGC TCA AGA AAG TCA GAG AAG ; optimized

Ser Tyr Arg Gly Val Arg Arg Arg Pro Trp Gly Lys Phe Ala Ala Glu Ile Arg Asp Ser ;
TCT TAC AGA GGC GTA AGA CGA CGG CCA TGG GGG AAA TTC GCG GCG GAG ATA AGA GAT TCG ; 361
TCT TAC AGA GGC GTA AGA CGA CGG CTT TGG GGG AAA TTC GCG GCG GAG ATA AGA GAT TCG ; optimized

Thr Arg Asn Gly Ile Arg Val Trp Leu Gly Thr Phe Glu Ser Ala Glu Glu Ala Ala Leu ;
ACT AGA AAC GGT ATT AGG GTT TGG CTC GGG ACG TTC GAA AGC GCG GAA GAG GCG GCT TTA ; 421
ACT AGA AAC GGT ATT AGG GTT TGG CTT GGG ACT TTC GAA AGC GCG GAA GAG GCG GCT CTT ; optimized

Ala Tyr Asp Gln Ala Ala Phe Ser Met Arg Gly Ser Ser Ala Ile Leu Asn Phe Ser Ala ;
GCC TAC GAT CAA GCT GCT TTC TCG ATG AGA GGG TCC TCG GCG ATT CTC AAT TTT TCG GCG ; 481
GCC TAC GAT CAA GCT GCT TTC TCG ATG AGA GGG TCC TCG GCG ATT CTT AAT TTT TCG GCG ; optimized
```

Рисунок 46.

После выполнения запланированных процедур пользователь имеет возможность визуализировать различия между исходным вариантом нуклеотидной последовательности трансгена и конечным вариантом модифицированной последовательности (нажатие «Показать различия» – «Show Differences»), а также пользователь может сохранить исходный и модифицированный варианты нуклеотидной последовательности трансгена в файле с помощью кнопки Сохранить – «Save all».

2.2.3.2. Элиминация потенциальных сайтов сплайсинга и полиаденилирования

Поскольку контекстные характеристики трансгена характерны для организма-донора и могут отличаться от типичных характеристик генов организма-реципиента, то некоторые комбинации нуклеотидов могут восприниматься системой экспрессии клеток организма-хозяина как сигналы экспрессии. Поэтому одна из задач, решение которой часто необходимо для эффективного трансгенеза, заключается в элиминации таких ложных сигналов сплайсинга (Haseloff et al., 1997; Rouwendal et al., 1997) и полиаденилирования (Diehn et al., 1998; Grec et al., 2000). В настоящее время существует несколько компьютерных программ, сделанных на основе информации о геноме арабидопсиса и позволяющих предсказывать сайты сплайсинга (Tolstrup et al., 1997; Brendel et al., 1998), однако эффективность их применения для других видов растений может быть ниже. К сожалению, методы для распознавания потенциальных сайтов полиаденилирования для генов растений пока не разработаны. Тем не менее, в некоторых случаях такая модификация структуры трансгена может быть необходимым условием достижения эффективной экспрессии чужеродных генов в растениях.

Для предсказания потенциальных сайтов сплайсинга и полиаденилирования нами были разработаны программные элементы FalseSpliceSitePred и FalsePolyASitePred, а также разработана схема их элиминации, включающая внесение точечных мутаций в критические позиции этих сайтов. Схема оптимизации основана на том, что при внесении точечных

мутаций в белок-кодирующую часть мРНК, аминокислотная последовательность трансгенного белка сохраняется интактной. Более того, если растение-реципиент содержится в базе данных программы, точечные мутации вносятся таким образом, чтобы при новом синонимическом кодоне был наиболее оптимальным для данного вида.

2.2.3.2.1. Описание работы FalseSpliceSitePred

Программа предназначена для выявления потенциальных сигналов экспрессии (сайтов сплайсинга и поли (А)-сигналов), расположенных в составе мРНК трансгена.

Пользователь должен ввести нуклеотидные последовательности 5'-НТП, CDS и 3'-НТП мРНК трансгена суммарной длиной не менее 60 нт, то есть длина 5'-НТП + длина CDS + длина 3'-НТП $\geq$ 60 нт.

Пользователь должен проделать следующие действия:

загрузить html-страницу с интерфейсом п/с «Трансгенез»;

нажать кнопку **Эукариоты** и загрузить нуклеотидную последовательность таким образом, как это описано в пункте «Операции для ввода данных в подсистему «Трансгенез»». выбирать программу **FalseSpliceSitePred**: предсказание и элиминация потенциальных сайтов сплайсинга;

нажимается кнопка **Запустить**.



Рисунок 47.

Открывается окно программы, в котором предоставляется возможность выбора скоров распознавания сайтов интрон-экзон и экзон-интрон. По умолчанию выставлены значения = 0.6, давшие оптимальное соотношение величин ошибок первого и второго рода при тестировании программы на обучающей и контрольной выборках.

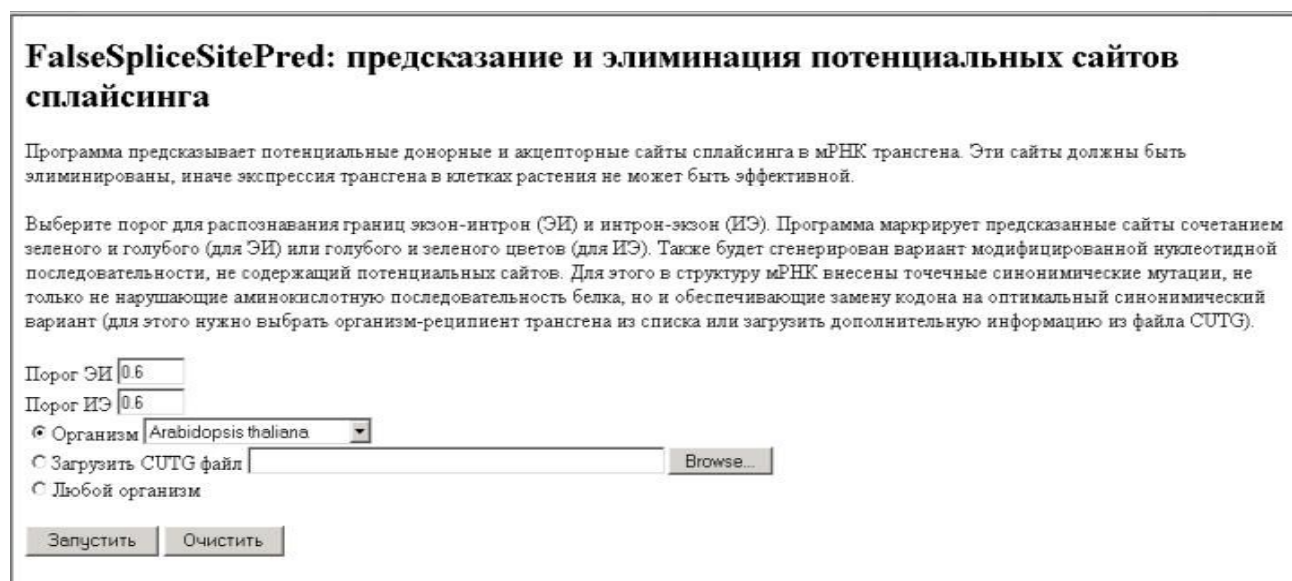
The image shows the main interface of the FalseSpliceSitePred program. It has a title bar "FalseSpliceSitePred: предсказание и элиминация потенциальных сайтов сплайсинга". Below the title bar is a paragraph explaining the program's purpose: "Программа предсказывает потенциальные донорные и акцепторные сайты сплайсинга в мРНК трансгена. Эти сайты должны быть элиминированы, иначе экспрессия трансгена в клетках растения не может быть эффективной." Below this is a paragraph of instructions: "Выберите порог для распознавания границ экзон-интрон (ЭИ) и интрон-экзон (ИЭ). Программа маркирует предсказанные сайты сочетанием зеленого и голубого (для ЭИ) или голубого и зеленого цветов (для ИЭ). Также будет сгенерирован вариант модифицированной нуклеотидной последовательности, не содержащий потенциальных сайтов. Для этого в структуру мРНК внесены точечные синонимические мутации, не только не нарушающие аминокислотную последовательность белка, но и обеспечивающие замену кодона на оптимальный синонимический вариант (для этого нужно выбрать организм-реципиент трансгена из списка или загрузить дополнительную информацию из файла CUTG)." Below the instructions are input fields for "Порог ЭИ" (set to 0.6) and "Порог ИЭ" (set to 0.6). There is a radio button for "Организм" with a dropdown menu showing "Arabidopsis thaliana". There is a radio button for "Загрузить CUTG файл" with a text input field and a "Browse..." button. There is a radio button for "Любой организм". At the bottom are two buttons: "Запустить" and "Очистить".

Рисунок 48.

Пользователь имеет возможность увеличивать чувствительность программы (понижая скор) или уменьшать вероятность ошибочного распознавания сайта (увеличивая скор). Рекомендуется работать в диапазоне значений сора от 0.6 до 0.65. Кроме этого есть возможность выбора паттерна видоспецифической оптимизации нуклеотидной последовательности. Например, выбор организма-реципиента из списка, или использование информации о частотах синонимических кодонов из файла («Загрузить CUTG файл»), или использование опции «Любой организм». Эти параметры программы имеют отношение к выполнению другого функционального требования (компьютерный дизайн структуры модифицированного варианта трансгена с целью достижения оптимальных параметров его экспрессии в конкретном молекулярно-генетическом окружении клетки организма-реципиента), и их тестирование будет описано в соответствующем разделе. После выбора скоров распознавания нажимается кнопка «Запустить». Генерируется html-страница, на которой расположены два варианта нуклеотидных последовательностей трансгена (нативная и оптимизированная), выровненные друг относительно друга. В нативном варианте нуклеотидной последовательности цветом выделены потенциальные сайты сплайсинга (границы интрон-экзон и экзон-интрон различаются чередованием синего и зеленого цветов).

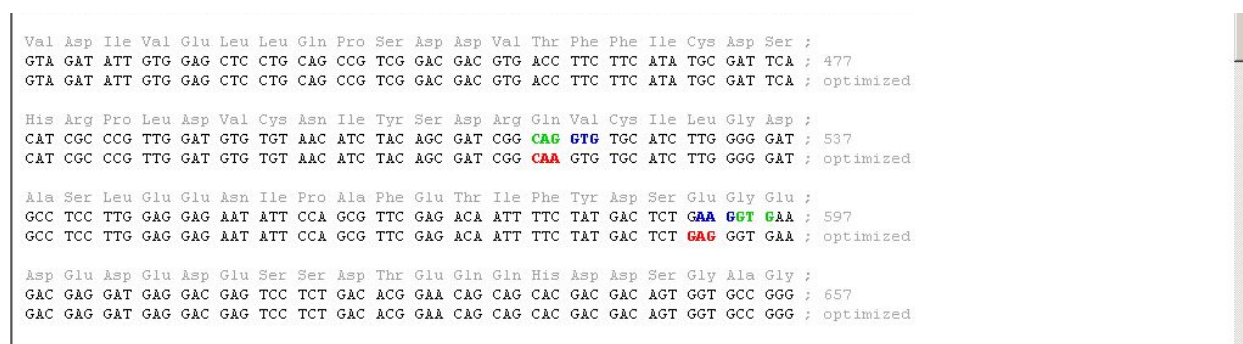


Рисунок 49.

Далее пользователь имеет возможность либо сохранить модифицированный вариант нуклеотидной последовательности (нажатие **Save & Main Menu**, дальнейший анализ будет осуществляться с этой уже модифицированной нуклеотидной последовательностью трансгена), либо отказаться от модификаций (нажатие «**Drop & Main Menu**», дальнейший анализ будет осуществляться с использованием нативного варианта трансгена).

2.2.3.2.2. Описание работы FalsePolyASitePred

Программа предназначена для выявления локализации потенциальных сайтов сплайсинга и полиаденилирования

На странице выбора программ выбирается программа «**FalsePolyASitePred**»: предсказание и элиминация потенциальных сайтов полиаденилирования. Рядом с названием программы приведено окно выбора сора распознавания (по умолчанию приведено оптимизированное значение = 0.612, показавшее наилучшие результаты при тестировании программы на контрольной выборке).



Рисунок 50.

Пользователь имеет возможность увеличивать чувствительность программы (понижая скор) или уменьшать вероятность ошибочного распознавания сайта (увеличивая скор). Рекомендуется работать в границах скор между 0.60 и 0.65. После выбора скор распознавания нажимается кнопка «Запустить». Генерируется html-страница, на которой расположены два варианта нуклеотидных последовательностей трансгена (нативная и оптимизированная), выровненных друг относительно друга. В нативном варианте нуклеотидной последовательности красным цветом выделены потенциальные сайты.

```

Val Val Ala Leu Lys Arg Lys His Ser Met Arg Arg Arg Met Thr Asn Lys Lys Thr Lys ;
GTT GTG GCG TTG AAG AGG AAA CAC TCG ATG AGA CGG AGA ATG ACC AAT AAG AAG ACG AAA ; 601
GTT GTG GCG TTG AAG AGG AAA CAC TCG ATG AGA CGG AGA ATG ACC AAT AAG AAG ACG AAA ; optimized

Asp Ser Asp Phe Asp His Arg Ser Val Lys Leu Asp Asn Val Val Val Phe Glu Asp Leu ;
GAT AGT GAC TTT GAT CAC CGC TCC GTG AAG TTA GAT AAT GTA GTT GTC TTT GAG GAT TTG ; 661
GAT AGT GAC TTT GAT CAC CGC TCC GTG AAG TTA GAT AAT GTA GTT GTC TTT GAG GAT TTG ; optimized

Gly Glu Gln Tyr Leu Glu Glu Leu Leu Gly Ser Ser Glu Asn Ser Gly Thr Trp End ;
GGA GAA CAG TAC CTT GAG GAG CTT TTG GGG TCT TCT GAA AAT AGT GGG ACT TGG TGA ; 718
GGA GAA CAG TAC CTT GAG GAG CTT TTG GGG TCT TCT GAA AAT AGT GGG ACT TGG TGA ; optimized

AAGATTAGGA TTGTATTAG GGACCTTAAG TTGAAAGTGG TTGATTAATT TTAACCCCTAA ; 778
AAGATTAGGA TTGTATTAG GGACCTTAAG TTGAAAGTGG TTGATGAATT TTAAGCCTAA ; optimized

TATGTTTTTT GTTGCTTAA ATATTTGATT CTATTGAGAA ACATCGAAAA CAGTTTGTAT ; 838
TATGTTTTTT GTTGCTTAA ATATTTGATT CTATTGAGAA ACATCGAAAA CAGTTTGTAT ; optimized

GTATTTTGTG GATACTTGAC TTTATCAATT TCAAGTTAGT GTAAAAGTCT TTGTCTCTTA ; 898
GTATTTTGTG GATACTTGAC TTTATCAATT TCAAGTTAGT GTAAAAGTCT TTGTCTCTTA ; optimized

CTATTCAACT AGTAAGAAAA CTCTCATTTT TCTTATTTAT CGTAATGAGG ATTGTCTTTA ; 958
CTATTCAACT AGTAAGAAAA CTCTCATTTT TCTTATTTAT CGTAATGAGG ATTGTCTTTA ; optimized

AG ; 960
AG ; optimized

```

Save & Main Menu Drop & Main Menu

Рисунок 51.

Далее пользователь имеет возможность либо сохранить модифицированный вариант нуклеотидной последовательности (нажатие «**Save & Main Menu**», дальнейший анализ будет осуществляться с этой уже модифицированной нуклеотидной последовательностью трансгена), либо отказаться от модификаций (нажатие «**Drop & Main Menu**», дальнейший анализ будет осуществляться с использованием нативного варианта трансгена). В данном случае возможен выбор любого из этих вариантов

Информация о локализации потенциальных сайтов сплайсинга и полиаденилирования выдается в виде изображения нуклеотидной последовательности мРНК трансгена с выделенными цветом позициями сайтов. Сохранение этой информации средствами п/с «Трансгенез» не предусмотрено, однако пользователь легко сохранить эту информацию самостоятельно – например, в виде html-файла с помощью стандартных функций интернет-браузера.

3. Полезные ссылки.

1. Кочетов А. В., Шумный В. К. Влияние структуры мРНК на процесс инициации трансляции в клетках растений // Успехи совр. биол. 1998. V. 118. P. 754-770.
2. Ananko E.A., Podkolodny N.L., Stepanenko I.L., Podkolodnaya O.A., Rasskazov D.A., Miginsky D.S., Likhoshvai V.A., Ratushny A.V., Podkolodnaya N.N., Kolchanov N.A. GeneNet in 2005. // *Nucleic Acids Res.* 2005. V. 33. P. 425-427.

3. Biemelt S., Sonnewald U., Galmbacher P., Willmitzer L. et al. Production of human papillomavirus type 16 virus-like particles in transgenic plants // *J. Virol.* 2003. V. 77. P. 9211-9220.
4. Brendel V., Kleffe J. Prediction of locally optimal splice sites in plant pre-mRNA with applications to gene identification in *Arabidopsis thaliana* genomic DNA // *Nucleic Acids Res.* 1998. V. 26. P. 4748-4757.
5. Chiapello H., Lisacek F., Caboche M., Henaut A. Codon usage and gene function are related in sequences of *Arabidopsis thaliana* // *Gene.* 1998. V. 209. GC1-GC38.
6. DellaPenna D. Plant metabolic engineering // *Plant Physiol.* 2001. V. 125. P. 160-163.
7. Diehn S. H., Chiu W-L., De Rocher E. J., Green P. J. Premature polyadenylation at multiple sites within a *Bacillus thuringiensis* toxin gene-coding region // *Plant Physiol.* 1998. V. 117. P. 1433-1443.
8. Duret L., Mouchiroud D. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis* // *Proc. Natl. Acad. Sci. U S A.* 1999. V. 96. P. 4482-4487.
9. Gallie D. R. Controlling gene expression in transgenics // *Curr. Opin. Plant Biol.* 1998. V. 1. P. 166-172.
10. Grec S., Wang Y., Le Guen L. et al. Cryptic polyadenylation sites within the coding sequence of three yeast genes expressed in tobacco // *Gene.* 2000. V. 242. P. 87-95.
11. Guttierrez R. A., Macuntosh G. C., Green P. J. Current perspectives on mRNA stability in plants: multiple levels and mechanisms of control // *Trends Plant Sci.* 1999. V. 4. P. 429-438.
12. Jacobs G. H., Rackham O., Stockwell P. A., et al. Transterm: a database of mRNAs and translational control elements // *Nucleic Acids Res.* 2002. V. 30. P. 310-311.
13. Haseloff J., Siemering K. S., Prasher D. C., Hodge S. Removal of cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly // *Proc. Natl. Acad. Sci. USA.* 1997. V. 94. P. 2122-2127.
14. Higo K., Ugawa Y., Iwamoto M., Korenaga T. Plant cis-acting regulatory DNA elements (PLACE) database:1999 // *Nucleic Acids Res.* 1999. V. 27. P. 297-300.
15. Kanehisa M., Goto S., Kawashima S., Nakaya A. The KEGG databases at GenomeNet // *Nucleic Acids Res.* 2002. V. 30. P. 42-46.
16. Khanna H. K., Raina S. K. Elite *Indica* transgenic rice plants expressing modified Cry1Ac endotoxin of *Bacillus thuringiensis* show enhanced resistance to yellow stem borer (*Scirpophaga incertulas*) // *Transgenic Res.* 2002. V. 11. P. 411-423.
17. Kochetov A. V., Grigorovich D., Kolchanov N. A., Sarai A. Database on mRNA located eukaryotic expression signals influencing translation efficiency and specificity // *Genome Informatics. Ser. 12 / Edt: Matsuda et al. Universal Academy Press: Tokyo,* 2001. P. 492-493.
18. Kochetov A. V., Ischenko I. V., Vorobiev D. G., Kel A. E. et al. Eukaryotic mRNAs encoding abundant and scarce proteins are statistically dissimilar in many structural features // *FEBS Lett.* 1998. V. 440. P. 351-355.
19. Kochetov A. V., Ponomarenko M. P., Kisselev L. L., Kolchanov N. A. Prediction of eukaryotic mRNA translational properties // *Bioinformatics.* 1999. V. 15. P. 704-712.
20. Kolchanov N. A., Ignatieva E. V., Ananko E. A. et al. Transcription Regulatory Regions Database (TRRD): its status in 2002. *Nucleic Acids Res.* 2002a. V. 30. P. 312-317.
21. Komar A. A., Lesnik T., Reiss C. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation // *FEBS Lett.* 1999. V. 462. P. 387-391.
22. Kozak M. Pushing the limits of the scanning mechanism for initiation of translation // *Gene.* 2002. V. 299. P. 1-34.
23. Larrick J. W., Thomas D. W. Producing proteins in transgenic plants and animals // *Curr. Opin. Biotechnol.* 2001. V. 12. P. 411-418.

24. Le Hir H., Nott A., Moore M. J. How introns influence and enhance eukaryotic gene expression // Trends Biochem. Sci. 2003. V. 28 P. 215-220.
25. Lescot M., Dehais P., Thijs G. et al.. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences // Nucleic Acids Res. 2002. V. 30. P. 325-327.
26. Loktev, K.A., Tkachev Yu.A., Ananko E.A., Podkolodny, N.L. (2002). A system for visual modelling of gene networks' structural and functional organization. Proc. Third International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2002). ICG, Novosibirsk, 2002, V. 2 . P. 132-135.
27. Lukaszewicz M., Feuermann M., Jerouville B., Stas A., Boutry M. *In vivo* evaluation of the context sequence of the translation initiation codon in plants // Plant Sci. 2000. V. 154. P. 89-98.
28. McCarthy J. E. G. Posttranscriptional control of gene expression in yeast // Microbiol Mol. Biol. Rev. 1998. V. 62. P. 1492-1553.
29. Mueller L.A., Zhang P., Rhee S.Y. AraCyc. A Biochemical Pathway Database for Arabidopsis // Plant Physiol. 2003. V. 132. P. 453-460.
30. Pesole G., Liuni S., Grillo G. et al. UTRdb and UTRsite: specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs. Update 2002 // Nucleic Acids Res. 2002. V. 30. P. 335-340.
31. Pestova T. V., Kolupaeva V. G., Lomakin I. B. et al. Molecular mechanisms of translation initiation in eukaryotes // Proc. Natl. Acad. Sci. U S A. 2001. V. 98. P. 7029-7036.
32. Rouwendal G. J. A., Mendes O., Wolbert E. J. H., de Boer A. D. Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage // Plant Mol. Biol. 1997. V. 33. P. 989-999.
33. Sawant S. V., Kiran K., Singh P. K., Tuli R. Sequence architecture downstream of the initiator codon enhances gene expression and protein stability in plants // Plant Physiol. 2001. V. 126. P. 1630-1636.
34. Schacherer F., Choi C., Gotze U. et al. The TRANSPATH signal transduction database: a knowledge base on signal transduction networks // Bioinformatics. 2001. V. 17. P. 1053-1057.
35. Tolstrup N., Rouze P., Brunak S. A branch point consensus from Arabidopsis found by non-circular analysis allows for better prediction of acceptor sites // Nucleic Acids Res. 1997. V. 25. P. 3159-3163.
36. Wingender E., Chen X., Fricke E. et al. The TRANSFAC system on gene expression regulation // Nucleic Acids Res. 2001. V. 29. P. 281-283.
37. van Hoof A., Green P. J. Rare codons are not sufficient to destabilize a reporter gene transcript in tobacco // Plant Mol. Biol. 1997. V. 35. P. 383-387.