

УДК 004.65

ИНФОРМАЦИОННАЯ ПОДДЕРЖКА ЭКСПЕРИМЕНТОВ ПО ТРАНСГЕНЕЗУ РАСТЕНИЙ В БАЗЕ ДАННЫХ ТРАНСЛЯЦИОННЫХ ЭНХАНСЕРОВ

© 2012 г. О.Г. Смирнова, Д.А. Рассказов, А.В. Кочетов

Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия, e-mail: ak@bionet.nsc.ru

Поступила в редакцию 15 июля 2012 г. Принята к публикации 31 августа 2012 г.

Разработана база данных для подбора трансляционных энхансеров, обеспечивающих дополнительный контроль экспрессии чужеродного гена в растениях на уровне трансляции мРНК. База данных содержит структурированную информацию о локализованных в мРНК трансляционных энхансерах, которые контролируют экспрессию генов на посттранскрипционном уровне. Эта информация полезна для планирования генно-инженерных экспериментов. Использование платформы и интерфейса Sequence Retrieval System позволяет проводить быстрый поиск трансляционных энхансеров с определенными характеристиками и получать нуклеотидные последовательности выбранных энхансеров. База данных доступна по адресу <http://www.mgs.bionet.nsc.ru/mgs/dbases/trsig/>.

Ключевые слова: база данных, информационный ресурс, трансляционный энхансер, генетическая инженерия, трансгенные растения.

ВВЕДЕНИЕ

Генетическая конструкция, которая используется для переноса чужеродного генетического материала при трансгенезе, помимо целевого гена, содержит дополнительные регуляторные последовательности, обеспечивающие транскрипцию и трансляцию этого гена. Для обеспечения необходимого паттерна экспрессии перенесенного гена необходимо использовать адекватные регуляторные последовательности и сигналы экспрессии. Дизайн генетической конструкции включает ряд последовательных этапов и требует эффективного планирования. Помимо сигналов транскрипционного контроля, большое значение для эффективного проведения научно-исследовательских работ в этой области может иметь оптимизация экспрессии трансгена на посттранскрипционном уровне. В составе эукариотических мРНК часто содержатся сигналы, контролирующие эффективность трансляции или цитоплазматическую стабильность матрицы. Применение таких сигналов в дизайне трансгена может существенно

увеличить эффективность трансгенеза. Обычно генетические конструкции не содержат сайтов сплайсинга, и в них используются аутентичные сигналы полиаденилирования, взятые из генов организма-реципиента, поэтому с этой фазой экспрессии трансгена возникает меньше проблем. Таким образом, практически важным становится адекватное планирование эффективности трансляции мРНК.

Одним из способов управления экспрессией гена являются трансляционные сигналы экспрессии, обычно расположенные в составе 5'- или 3'-нетранслируемых районов мРНК (Liu *et al.*, 2009). Некоторые из таких сигналов могут определять общую трансляционную активность мРНК: например, если в составе генетической конструкции используется 5'-НТП мРНК вируса табачной мозаики (размером 68 нуклеотидов), то это в большинстве случаев увеличивает уровень синтеза белкового продукта в несколько раз (Gallie, 2002). Этот энхансер активно используется в биотехнологии растений. Кроме неспецифических усилителей трансляционной активности мРНК, известны специфические

сигналы (например, IRE, регулирующий трансляцию мРНК гена ферритина в зависимости от присутствия железа; с этим сигналом связывается специфический белок IRP, конформация и активность которого зависят от присутствия железа, что функционально эквивалентно ситуации транскрипционного контроля и сайтам связывания транскрипционных факторов в промоторах). Однако, в отличие от сайтов связывания транскрипционных факторов, сигналы трансляционного контроля экспрессии мало изучены, что связано со специфическими особенностями одноцепочечных молекул РНК, способных в цитоплазме существовать в виде различных конформеров. Большинство сигналов такого типа представляют собой комбинацию контекстных и структурных элементов, что делает их предсказание очень сложным.

Информационные ресурсы в этой области представлены базами данных (БД) UTRsite (Grillo *et al.*, 2010) и TransTerm (Jacobs *et al.*, 2009). В этих БД аннотировано крайне малое количество (всего несколько десятков) трансляционных сигналов разной таксономической принадлежности (по некоторым оценкам сигналы этого типа могут присутствовать у 10–15 % мРНК генов эукариот). Причина такой недостаточной аннотации заключается в том, что в этих БД рассматриваются только те трансляционные сигналы, у которых точно известна тонкая структура (контекстные и конформационные элементы сигнала). При этом для планирования биотехнологических опытов, как правило, достаточно знать, как изменяется паттерн трансляционной активности мРНК трансгена в присутствии тех или иных участков мРНК с известной трансляционной активностью. Таким образом, представленная в существующих базах данных по трансляционным сигналам информация, с одной стороны, недостаточна, а с другой – чрезмерно детализирована для целей планирования трансгенетических экспериментов. В то же время следует отметить, что экспериментальной информации такого рода в литературе достаточно много, что говорит о возможности создания специализированного информационного ресурса.

В настоящей работе описывается информационный ресурс БДТЭ (База данных трансляционных энхансеров), который разработан нами для решения задачи сбора, хранения и

систематизации информации о трансляционных энхансерах. Представленная информация и удобный интерфейс позволяют осуществлять выбор такого важного элемента генетической конструкции, как энхансер трансляции, способствуя тем самым эффективному созданию форм растений с улучшенными и качественно новыми свойствами.

Формат и логическая структура БД трансляционных энхансеров

Согласно проведенному нами анализу литературных данных, специализированные БД трансляционных энхансеров для планирования генно-инженерных экспериментов отсутствуют. Имеющиеся аналоги приспособлены для решения более широкого круга задач (в основном фундаментального характера) и не могут быть эффективно использованы для этой цели. С нашей точки зрения, основным недостатком существующих информационных ресурсов является тот факт, что процедура аннотации энхансеров трансляции на основе анализа литературных данных в них высокоизбирательна, т. е. необходимы детальные знания об их тонкой организации.

Однако для решения биотехнологических задач эта информация не нужна. Для выбора трансляционного энхансера в качестве элемента биотехнологической генетической конструкции необходимо и достаточно знать, какой уровень и паттерн трансляции мРНК репортерного гена обеспечивает определенная нуклеотидная последовательность (прототип энхансера) в генетически модифицированном организме.

База данных трансляционных энхансеров позволяет специалисту в области генной инженерии выбирать потенциальные энхансеры по следующим полям: 1) организм – донор энхансера; 2) организм – реципиент энхансера (в котором была оценена его экспрессия); 3) паттерн трансляционной активности (ткане-, органо-, стадийспецифичность наработки белка-репортера); 4) уровень трансляционной активности мРНК гена репортера, содержащей данный энхансер.

БДТЭ включает в себя две составляющие:

1) таблицу объектов (TRANSIG_OBJ), содержащую информацию о прототипе исследуемого объекта (обычно – природном варианте

трансляционного энхансера) и о его свойствах, полученную при анализе научных статей. Эта информация помогает пользователю выбрать тип регуляторного сигнала с нужными свойствами;

2) таблицу энхансеров (TRANSIG_ENH), содержащую информацию об энхансерах, полученную при анализе научных статей. Под энхансером понимаются нуклеотидные последовательности, обычно представляющие собой модифицированный вариант объекта (прототипа, природного варианта), содержащий делеции или инсерции, активность которого была описана в аннотированной статье. Характерной особенностью энхансера является привязка к конкретным экспериментальным данным (нуклеотидная последовательность, вид растений, в котором была проведена проверка свойств этой нуклеотидной последовательности, репортерный ген и т. п.). Каждому прототипу (объекту) может соответствовать несколько вариантов энхансеров (экспериментальных последовательностей, как минимум, один вариант).

Структура БДТЭ в графическом виде представлена на рис. 1.

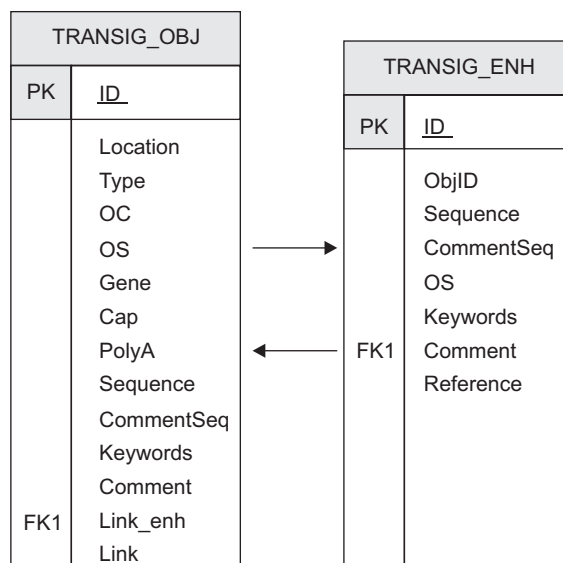


Рис. 1. Структура базы данных ТЭ.

Описание таблицы TRANSIG_OBJ включает 15 полей (табл. 1). Описание таблицы TRANSIG_ENH включает 9 полей (табл. 2). В табл. 3 и 4 приведены примеры записей в информационных таблицах TRANSIG_OBJ

Таблица 1

Структура записи в таблице TRANSIG_OBJ (+ обозначает индексируемое SRS поле, по которому пользователь может осуществлять поиск информации)

ID	Идентификатор записи TRANSIG_OBJ (+)
LOCATION	Расположение энхансера (5'UTR, 3'UTR, CDS) (+)
TYPE	Тип энхансера (stress-specific и др.) (+)
OC	Таксономическая классификация (+)
OS	Название вида (+)
GENE	Название гена (+)
CAP	Наличие кепы на 5'-конце мРНК (+)
POLYA	Наличие поли(А)-участка на 3'-конце мРНК (+)
SQ	Собственно нуклеотидная последовательность энхансера
COMMENTSEQ	Комментарий о происхождении нуклеотидной последовательности и ее расположении в составе генетических конструкций (+)
KEYWORD	Ключевые слова (+)
COMMENT	Развернутый комментарий о специфичности и активности энхансера, эффективности его использования в различных видах организмов реципиентов (+)
LINK_ENH	Ссылка на идентификатор записи таблицы TRANSIG_ENH (+)
LINK	Ссылка на банк данных нуклеотидных последовательностей (+)

Таблица 2

Структура записи в таблице TRANSIG_ENH (+ обозначает индексируемое SRS поле, по которому пользователь может осуществлять поиск информации)

ID	Идентификатор записи TRANSIG_ENH (+)
OBJID	Ссылка на идентификатор записи таблицы TRANSIG_OBJ (+)
SEQUENCE	Нуклеотидная последовательность функционального района (5'-НТП, 3'-НТП), содержащего энхансер
COMMENTSEQ	Комментарий к структуре экспериментальной конструкции (+)
ORGANISM	Видовое название организма, на котором проводили эксперименты (+)
KEYWORD	Ключевые слова (+)
COMMENT	Развернутый комментарий о специфичности и активности энхансера, эффективности его использования в различных видах организмов реципиентов (+)
REFERENCE	Название статьи и ссылка на БД PubMed (+)

Таблица 3

Пример заполнения записи в таблице TRANSIG_OBJ

ID	ADHZM5
LOCATION	5'UTR
TYPE	Stress-specific enhancer
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea
OS	Zea mays
GENE	ADH1, alcoholdehydrogenase I
CAP	Capped
POLYA	Polyadenylated
SQ	ATTTTCTCGCTCCTCACAGGCTCATCTCGTTTGGATCGATTG GTTTCGTAAGTGGTGAAGGACTGAGGGTCTCGGAGTGGATCG ATTTGGGATTCTGTTTCAAGATTGCGGAGGGGGGCA
COMMENTSEQ	5'UTR of ADH1 gene mRNA
KEYWORD	Enhancer, hypoxia, anoxia, anaerobiosis, stress
COMMENT	It was found that translation of alcoholdehydrogenase mRNA was efficient under oxygen deprivation conditions whereas translation of many other mRNAs was stopped. No changes in mRNA stability were detected so the effect observed could result from the changes in stress-specific translation rate. Deletions of ADH 5'UTR decreased stress-specific translatability; the influence of possible changes in secondary structure was not tested or discussed...
LINK_ENH	ADHZM5a
LINK	EMBL_AC X00580

Таблица 4

Пример заполнения записи в таблице TRANSIG_ENH

ID	ADHZM5a
OBJID	ADHZM5
SEQUENCE	ATAGGGAGACCGAATTCGAGCTCATTTTCTCGCTCCTCACAGGCTCATCT CGTTTGGATCGATTGGTTTCGTAAGTGGTGAAGGACTGAGGGTCTCGGAG TGGATCGATTGGGATTCTGTTCGAAGATTTGCGGAGGGGGGCA
COMMENTSEQ	Design of mRNA 5'UTR of GUS reporter gene: first 23 nt were taken from vector sequence followed by 108-nt long 5'UTR of ADH1. In this construct CDS consisted from 18 codons of ADH1 CDS fused to 8 codons derived from vector polylinker sequence and GUS CDS downstream (see LONG). 3'UTR was represented by ADH
ORGANISM	Zea mays
KEYWORD	Hypoxia, stress, 5'UTR, enhancer
COMMENT	Translational efficiencies of reporter mRNAs containing UTR sequences of maize alcoholdehydrogenase gene mRNA were tested in maize protoplasts under normal or oxygen deprivation conditions. No changes in mRNA stability were detected so the effect observed resulted from the changes in stress-specific translation rate. Interestingly, the presence of ADH 5'UTR did not affect translation under aerobic conditions. Generally, deletions of ADH-derived fragments decreased stress-specific translatability: either deletion of first 18 ADH-derived codons or fragments of 5'UTR or 3'UTR. Note, that 5'portion of 5'UTR was presented in all constructions. The influence of possible changes in secondary structure was not tested or discussed. As was found ADH1 3'UTR mRNA increase hypoxia-specific translation 3.5-fold but decrease aerobic translation 3-fold.
REFERENCE	Bailey-Serres J., Dawe R.K. Both 5' and 3' sequences of maize adh1 mRNA are required for enhanced translation under low-oxygen conditions. Plant Physiol. 1996. 112. 685-695 PMID:8883381

и TRANSIG_ENH. Структура базы данных позволяет расширять список аннотируемых энхансеров по мере появления новых литературных данных.

Технологии реализации БДТЭ

БДТЭ управляется средствами информационно-поисковой системы Sequence Retrieval System (SRS) 6.1, которая развернута на сервере баз данных под управлением Red Hat Enterprise Linux 5.7. Система SRS (Zdobnov *et al.*, 2002) специально разработана для формализованного описания биологических данных по заказу European Bioinformatics Institute. Средства SRS позволяют индексировать большинство информационных полей и эффективно осуществлять перекрестную связь полей в таблицах баз данных, что необходимо для построения эффективных пользовательских запросов и

свободной навигации между полями и записями в различных таблицах. Эта система также автоматически генерирует Web-интерфейс для обеспечения поиска и визуализации информации в БД (формы запроса, визуализация данных, гиперссылки на документы в базе и Интернет-ресурсы, настройки способа визуализации).

Пользовательский интерфейс

Модуль интерфейса для БДТЭ представляет собой программный компонент, обеспечивающий интерфейс пользователя с БД. Доступ к БДТЭ может быть осуществлен по адресу <http://www.mgs.bionet.nsc.ru/mgs/dbases/trsig/> (рис. 2).

Пользователю предоставляется возможность просматривать списки генотипов, ключевых слов, типов энхансеров, которые содержатся в базе. Кроме этого, пользователь имеет возможность осуществлять поиск по большому числу

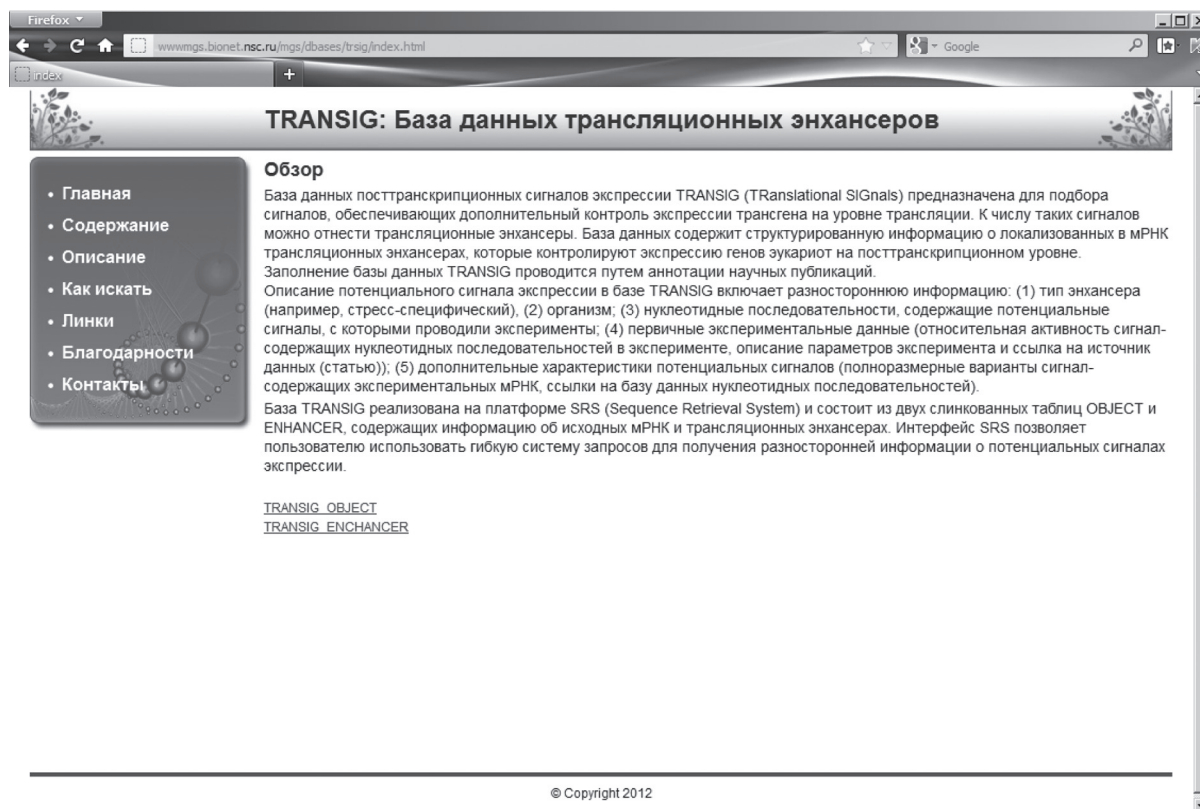


Рис. 2. Скриншот титульной страницы интерфейса БДТЭ.

поисковых полей таблиц TRANSIG_OBJ и TRANSIG_ENH.

Поиск проводится отдельно для таблиц TRANSIG_OBJ и TRANSIG_ENH. Поиск в таблице TRANSIG_OBJ позволяет решать следующие задачи:

- найти сигналы, локализованные в 5'UTR, 3'UTR или internal fragment (поиск по полю LOCATION);
- найти тканеспецифические сигналы (поиск по полям TYPE, KEYWORDS и COMMENT);
- найти энхансеры, принадлежащие определенному организму (поиск по полю OS);
- найти энхансеры, расположенные в мРНК определенного гена (поиск по полю GENE);
- найти энхансеры, выделенные из мРНК с кепом на 5'-конце (поиск по полю CAP);
- найти энхансеры, выделенные из мРНК с поли(A)-участком на 3'-конце (поиск по полю POLYA).

Если найденный сигнал удовлетворяет требованиям пользователя, нуклеотидная последовательность из поля «Sequence» таблицы TRANSIG_OBJ может в дальнейшем использо-

ваться как специфический сигнал экспрессии при дизайне регуляторных элементов трансгена.

Чтобы найти сигнал по названию гена, в мРНК которого он был обнаружен, в поисковой таблице необходимо выбрать название поля «Gene» и ввести часть названия гена со звездочкой (чтобы сделать расширенный поиск). Например, ген «alcoholdehydrogenase I» может быть обозначен как ADH1. В результате выполнения запроса система выдает список генов TRANSIG_OBJ: ADH1 и TRANSIG_OBJ:ADH1.

Чтобы найти сигналы, которые влияют на уровень мРНК в конкретном виде растений и зависят от конкретного регулятора, необходимо провести комплексный поиск. Для этого необходимо перейти на страницу сложного запроса, нажав кнопку «Search» на странице с типовым описанием структуры полей таблицы TRANSIG_ENH.

На этой странице меню «combine searches with» задает логическую операцию для выполнения совместных запросов нескольких полей. Например, для поиска энхансеров табака, которые реагируют на освещенность, необходимо

установить значение этого меню «AND», в левом столбце формы запроса выбрать поле «OS», а в правом столбце ввести название вида («tobacco»). В следующей строке формы в левом столбце выбрать «KEYWORDS», а в правом ввести условие («light») и выполнить запрос. Будет получен список сигналов, действие которых в растениях трансгенного табака зависело от освещения: TRANSIG_ENH:PSILRE5a.

ЗАКЛЮЧЕНИЕ

Разработана БДТЭ, предназначенная для накопления информации, необходимой для планирования генно-инженерных экспериментов с целью создания организмов с качественно новыми или улучшенными свойствами. БДТЭ содержит структурированную информацию о локализованных в мРНК трансляционных энхансерах, контролирующих экспрессию генов растений на посттранскрипционном уровне. Текущая версия БДТЭ содержит 58 аннотированных объектов (природных прототипов энхансеров) и 68 экспериментально исследованных энхансеров.

Формат представления данных обеспечивает быстрый поиск трансляционных энхансеров, чувствительных к различным индукторам, с целью обеспечения дополнительного к действию промотора паттерна транскрипции трансгена, необходимого для решения конкретных генно-инженерных задач.

БДТЭ является одним из модулей информационного портала «Биотехнология растений», разрабатываемого в ИЦиГ СО РАН (Кочетов и др., 2012). Этот информационный ресурс представляет собой платформу для решения различных задач в области генной инженерии и биотех-

нологии растений, содержащую модули разного типа. В частности, модуль БДТЭ планируется использовать в комбинации с модулем БДП, представляющим собой базу данных промоторов для трансгенеза (Смирнова и др., 2012).

Работа поддержана грантом Министерства образования и науки РФ в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 гг.» (07.514.11.4052).

ЛИТЕРАТУРА

- Кочетов А.В., Смирнова О., Ибрагимова С. и др. Информационный портал «Биотехнология растений» – Интернет-ресурс для поддержки экспериментов в области генной инженерии растений, генетики и селекции пшеницы // Вавилов. журн. генет. и селекции. 2012. Т. 16. № 4/1. С. 838–848.
- Смирнова О.Г., Рассказов Д.А., Афонников Д.А., Кочетов А.В. TGP – база данных промоторов для трансгенеза растений // Матем. биология и биоинформатика. 2012. Т. 7. № 2. С. 444–460.
- Gallie D.R. The 5'-leader of tobacco mosaic virus promotes translation through enhanced recruitment of eIF4F // Nucl. Acids Res. 2002. V. 30. No. 15. P. 3401–3411.
- Grillo G., Turi A., Licciulli F. *et al.* UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs // Nucl. Acids Res. 2010. V. 38. Database issue. P. D75–D80.
- Jacobs G.H., Chen A., Stevens S.G. *et al.* Transterm: a database to aid the analysis of regulatory sequences in mRNAs // Nucl. Acids Res. 2009. V. 37. Database issue. P. D72–D76.
- Liu Y., Wimmer E., Paul A.V. Cis-acting RNA elements in human and animal plus-strand RNA viruses // Biochim. Biophys. Acta. 2009. V. 1789. No. 9/10. P. 495–517.
- Zdobnov E.M., Lopez R., Apweiler R., Etzold T. The EBI SRS server – recent developments // Bioinformatics. 2002. V. 18. P. 368–373.

A DATABASE ON TRANSLATIONAL ENHANCERS TO SUPPORT EXPERIMENTS WITH TRANSGENIC PLANTS

O.G. Smirnova, D.A. Rasskazov, A.V. Kochetov

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: ak@bionet.nsc.ru

Summary

A database on translational enhancers providing additional control of foreign gene expression at the mRNA translation level has been developed. It contains structured information on the presence of enhancers located within mRNAs, which control gene expression at the posttranscriptional stage. These data can be used to design genetic constructs for plant transgenesis. The database is based on the platform of the Sequence Retrieval System (SRS), allowing users to make a rapid search for enhancers with defined properties and retrieve corresponding nucleotide sequences. The database is available at <http://wwwmgs.bionet.nsc.ru/mgs/dbases/trsig/>.

Key words: database, information resource, translational enhancer, genetic engineering, transgenic plants.