

УДК: 004.65:577.214.625:575.1/2:581:602.6

TGP – база данных промоторов для трансгеноза растений

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

Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, г. Новосибирск, Новосибирская область, 630090, Россия

Аннотация. При создании трансгенных растений важную роль играет выбор промотора, обеспечивающего адекватный паттерн экспрессии трансгена. В настоящее время нет баз данных, содержащих информацию о промоторах для трансгеноза. В статье описывается база данных, предназначенная для систематизации информации о промоторах, активность которых была проверена в экспериментах с трансгенными растениями. Использование платформы и интерфейса Sequence Retrieval System (SRS) позволяет проводить быстрый поиск промоторов с определенными характеристиками и получать нуклеотидную последовательность выбранного промотора. База данных доступна по адресу http://www.mgs.bionet.nsc.ru/mgs/dbases/tgp_ru/home.html.

Ключевые слова: Базы данных, информационный ресурс, промотор, генетическая инженерия, трансгенные растения.

ВВЕДЕНИЕ

В настоящее время трансгенные организмы различной таксономической принадлежности активно используются в различных отраслях современной биотехнологии. Как системы для наработки фармацевтически значимых белков, растения более безопасны по сравнению с микробами и животными: у них нет патогенов человека, онкогенных последовательностей ДНК, а также эндотоксинов. Список фармацевтических белков, производимых различными растительными системами, включает антитела, вакцины, ростовые факторы, гормоны [1, 2]. Наряду с фармацевтической направленностью, генная инженерия используется для повышения урожайности и качества сельскохозяйственных культур.

Согласно современным представлениям, генно-инженерный эксперимент требует эффективного планирования и включает ряд последовательных этапов, одним из которых является дизайн генетической конструкции. Для обеспечения необходимого паттерна экспрессии перенесенного гена необходимо использовать адекватные регуляторные последовательности и сигналы экспрессии. Промотор является наиболее важным элементом генетической конструкции. В качестве примера удачных в биотехнологическом плане промоторов можно привести коммерческое применение трансгенных растений со стадиейспецифичным промотором SAG12 Арабидопсиса. При использовании в качестве трансгена изопентилтрансферазы *Agrobacterium tumefaciens*, специфическая активность промотора AtSAG12 на поздних стадиях развития у

* planta@bionet.nsc.ru

большого числа видов трансгенных растений приводит к продлению вегетации, повышению урожая и его сохранности [3–8]. К сожалению, в настоящее время генными инженерами активно используется ограниченный набор промоторов, что препятствует развитию экспериментов с трансгенными растениями и недостаточно для решения сложных биотехнологических задач [9–11].

Для успешного выполнения биотехнологических задач с помощью методов генной инженерии целесообразно использовать ряд имеющихся в открытом доступе баз данных (БД) и программных комплексов. В мире разработан ряд БД, содержащих информацию о нуклеотидных последовательностях промоторов и сайтов связывания транскрипционных факторов (ССТФ). Большинство из них накапливают информацию о ССТФ. Три взаимосвязанных БД AtTFDB, AtcisDB и AtRegNet информационного сервера AGRIS (Arabidopsis Gene Regulatory Information Server) содержат информацию о последовательностях промоторов Арабидопсиса, транскрипционных факторах и их целевых генах. Один из трех модулей базы, AtcisDB, содержит около 33,000 вышележащих районов аннотированных генов Арабидопсиса с описанием экспериментально проверенных и предсказанных ССТФ [12]. БД PlantProm содержит нуклеотидные последовательности промоторов растений с экспериментально проверенным стартом транскрипции [13]. БД PLACE позволяет находить регуляторные цис-элементы в последовательности ДНК растительных генов [14]. Навигатор PlantPAN предназначен для узнавания комбинаторных цис-элементов в генах растений [15]. БД Athena и Osiris обеспечивают визуализацию ССТФ и предсказанные структуры промоторов Арабидопсиса [16, 17]. Athena содержит нуклеотидные последовательности промоторов размером до 3 kb для предсказанных генов Арабидопсиса и нуклеотидные последовательности консенсусов для 105 охарактеризованных ранее ССТФ, импортированных из PLACE и AGRIS. БД Osiris содержит нуклеотидные последовательности промоторов, предсказанных ССТФ, аннотацию онтологии генов и данные анализа транскрипции на микрочипах для 24209 генов из генома риса. БД AthaMap представляет полногеномное картирование потенциальных ССТФ и содержит сайты для 115 различных транскрипционных факторов [18]. Имеющиеся информационные ресурсы преимущественно представляют данные о ССТФ и не содержат информацию об экспериментально проверенной промоторной активности фрагментов ДНК растений.

Следует отметить, что аннотированная в этих базах информация не оптимизирована для эффективного планирования генно-инженерного эксперимента. В принципе, присутствие ССТФ может быть использовано для выбора промотора при создании генетической конструкции. Однако существует много примеров ошибочности таких предсказаний. Например, промотор *GluC* обеспечивает высокий уровень транскрипции в эндосперме, хотя не содержит мотивов *GCN4*, *AACA*, *prolam* box, специфичных для экспрессии в этой ткани [10]. Ген *Pptha1* персика был обнаружен благодаря повышенной экспрессии в условиях холода, но промоторный район этого гена не обеспечивал аналогичный паттерн экспрессии в трансгенных растениях арабидопсиса [19]. С нашей точки зрения, наиболее подходящим источником данных для выбора промотора может быть информация о фрагментах ДНК с определенной транскрипционной активностью, полученная в экспериментах с трансгенными растениями [20–23]. В литературе содержится большое количество экспериментальных данных о нуклеотидных последовательностях ДНК, обладающих промоторной активностью. Эти данные были получены в экспериментах, направленных на изучение паттернов экспрессии генов растений, а также выявление соответствующих ССТФ. В таких случаях типовая схема экспериментального анализа структурно-функциональной организации промотора в качестве одного из этапов включает получение делеционных вариантов 5'-концевого участка ДНК, расположенного выше белок-кодирующей области изучаемого гена. Под управление этих участков помещают ген-репортер,

обеспечивающий визуализацию их транскрипционной активности. Эти генетические конструкции переносят в геном какого-либо вида растений и проводят анализ транскрипционной активности делеционных вариантов промоторного района. В результате в статьях приводится информация о фрагментах геномной ДНК, обладающих определенным паттерном транскрипционной активности, которые могут использоваться и в других экспериментах. Важно отметить, что эти данные не систематизированы и не существует информационных ресурсов, в которых они содержатся. Можно привести несколько примеров, выбранных нами из литературы.

Промотор гена пшеницы TaPT2 обеспечивает специфическую активность репортерного гена бета-глюкуронидазы (GUS) в корнях трансгенных арабидопсиса и пшеницы при фосфорном голодании [24]. Промоторы генов глиадинов пшеницы обеспечивают высокую активность GUS в эндосперме трансгенной пшеницы [25,26]. Промотор гена глобулина овса обеспечивает эндосперм-специфичную экспрессию GUS в семенах ячменя (до 10% растворимого белка [27]). Промотор TaPSG719 пшеницы проявляет специфическую активность исключительно в пыльце трансгенного табака [28]. Этот промотор может быть с успехом использован в сельском хозяйстве при создании форм с мужской стерильностью. В некоторых случаях показано, что такая тканеспецифическая экспрессия является более выгодной, чем конститутивная, которая обеспечивается промоторами 35S вируса мозаики цветной капусты или убиквитина кукурузы [29–31]. Развитие новых методов для повышения устойчивости растений к различным стрессам также требует использование промоторов со специфическим паттерном экспрессии. Например, активность промоторов TaAIDFa и Cog/Lea генов пшеницы, промотора DREB1 карликовой яблони возрастает после воздействия засухи, соли, низких температур и абсцизовой кислоты [32–35]. Эти промоторы могут быть с успехом использованы для повышения устойчивости сельскохозяйственных культур к неблагоприятным воздействиям окружающей среды.

Согласно проведенному нами анализу специализированные БД промоторов для планирования генно-инженерных экспериментов отсутствуют, имеющиеся аналоги приспособлены для решения задач фундаментального характера и не могут быть эффективно использованы для этой цели.

Таким образом, мы считаем обоснованным развитие специализированной БД промоторов для трансгенеза растений, основанной на аннотировании литературных данных.

КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ ДАННЫХ

С нашей точки зрения для выбора промотора в качестве элемента биотехнологической генетической конструкции необходимо и достаточно знать следующее: определенная нуклеотидная последовательность способна направлять транскрипцию репортерного гена в трансгенном организме по определенному паттерну и на определенном уровне. Если такая информация будет оформлена в виде БД, то специалист в области генной инженерии может выбирать потенциальные промоторы по следующим полям: (1) организм – донор промотора; (2) организм – реципиент промотора (в котором была оценена его экспрессия); (3) паттерн и уровень экспрессии промотора. Следует отметить, что объем доступной информации этого типа достаточно велик, поскольку характеристика транскрипционной активности промоторов часто используется в процессе изучения функций генов. Контроль экспрессии гена часто изучается с использованием репортерной конструкции, в которой промотор помещается перед репортерным геном (бета-глюкуронидазы *E. coli*, люциферазы, зеленого флюоресцирующего белка). Анализ активности репортерного белка в трансгенных растениях позволяет оценивать функциональные характеристики изучаемого промотора.

Таким образом, на основе аналитического обзора литературных данных нами были разработаны основные требования к БД промоторов для трансгенеза. БД должна содержать следующую необходимую информацию:

- а) название организма – донора нуклеотидной последовательности промотора;
- б) название гена – донора нуклеотидной последовательности промотора;
- в) уровень экспрессии гена репортера, направляемого данным промотором;
- г) название репортерного гена;
- д) паттерн экспрессии промотора (индуцибельность, ткане-, органо-, стадийспецифичность наработки белка-репортера);
- е) нуклеотидную последовательность промотора.

Эта информация была разбита на три тематических раздела.

ЛОГИЧЕСКАЯ СТРУКТУРА БД TGP

БДП включает в себя три составляющие:

- 1) таблицу промоторов TGP_PROMOTER, содержащую информацию о промоторах, полученную при анализе научных статей;
- 2) таблицу нуклеотидных последовательностей промоторов TGP_SEQUENCE, полученных из БД GenBank в соответствии с данными научных публикаций;
- 3) таблицу генов TGP_GENE, содержащую информацию о генах, полученную при анализе научных статей.

Структура БД TGP в графическом виде представлена рисунке 1.

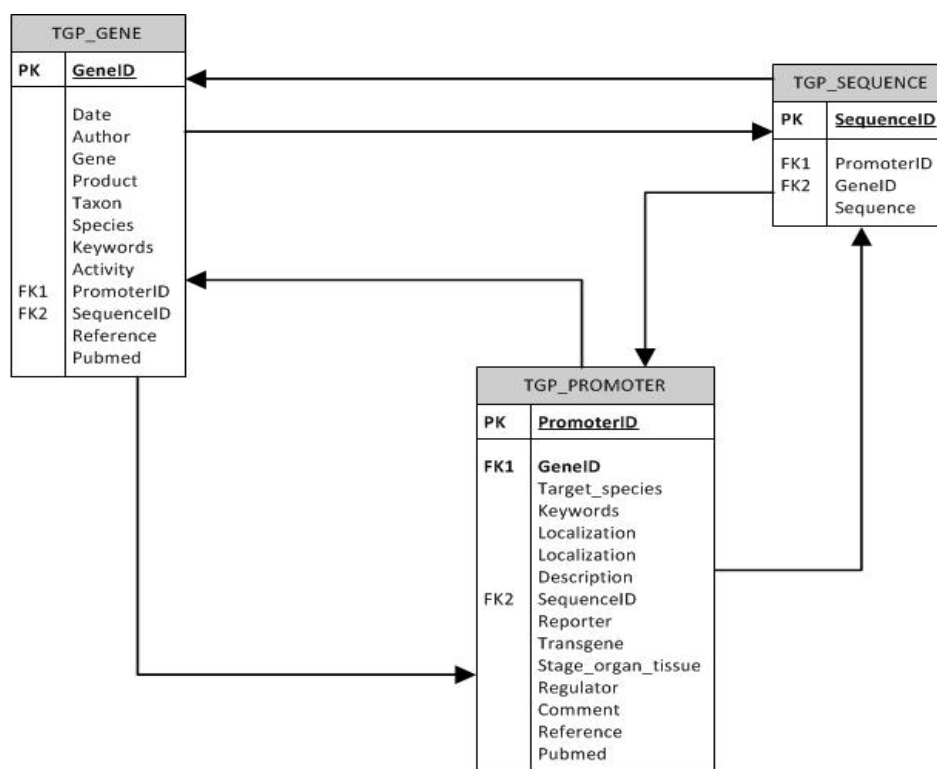


Рис. 1. Структура базы данных промоторов для трансгенеза.

БД TGP включает в себя три составляющие (таблицы) и 6 отношений (табл. 1).

Таблица 1. Описание связей между таблицами БД TGP

Таблица 1	Поле таблицы 1	Тип связи	Таблица 2	Поле таблицы 2
TGP_PROMOTER	SequenceID	1:1	TGP_SEQUENCE	SequenceID
TGP_PROMOTER	GeneID	N:1	TGP_GENE	GeneID
TGP_SEQUENCE	PromoterID	1:1	TGP_PROMOTER	PromoterID
TGP_SEQUENCE	GeneID	N:1	TGP_GENE	GeneID
TGP_GENE	PromoterID	1:N	TGP_PROMOTER	PromoterID
TGP_GENE	SequenceID	1:N	TGP_SEQUENCE	SequenceID

Блок PROMOTER (описание собственно промотора) включает 14 полей, описание которых приведено в таблице 2.

Таблица 2. Структура карточки блока PROMOTER (+ обозначает индексируемое SRS поле, по которому пользователь может осуществлять поиск информации)

Название поля	Содержание поля
PROMOTER_ID	Идентификатор карточки в блоке PROMOTER (+)
GENE_ID	Ссылка на идентификатор гена в блоке GENE (+)
TARGET SPECIES	Видовое название трансгенного организма (+)
KEYWORDS	Ключевые слова (+)
LOCALIZATION	Границы промотора относительно точки отсчета, ссылка на БД GenBank с указанием старта транскрипции или трансляции
DESCRIPTION	Данные о структуре промотора (комментарий) (+)
SEQUENCE_ID	Ссылка на идентификатор последовательности промотора в блоке SEQUENCE (+)
REPORTER	Название гена-репортера
TRANSGENE	Название и видовое происхождение трансгена (+)
STAGE_ORGAN_TISSUE	Название стадий развития, органов и тканей, в которых наблюдается активность промотора (+)
REGULATOR	Индуктор (репрессор), влияющий на активность промотора (+)
COMMENT	Характеристика активности промотора (комментарий) (+)
REFERENCE	Название статьи, из которой была взята информация о промоторе (+)
PUBMED	ссылка на БД публикаций Entrez-PUBMED (+)
END	Конец блока описания промотора

Блок SEQUENCE (описание нуклеотидной последовательности промотора) включает 5 полей (табл. 3).

Таблица 3. Структура карточки блока SEQUENCE (+ обозначает индексируемое SRS поле, по которому пользователь может осуществлять поиск информации)

Название поля	Содержание поля
SEQUENCE_ID	Идентификатор карточки в блоке SEQUENCE (+)
PROMOTER_ID	Перекрестная ссылка на идентификатор соответствующей карточки в блоке PROMOTER
GENE_ID	Перекрестная ссылка на идентификатор соответствующей карточки в блоке GENE
SEQUENCE	Нуклеотидная последовательность
END	Конец блока описания последовательности

Блок GENE (описание гена - донора промотора) включает 14 полей (табл. 4).

Таблица 4. Структура карточки блока GENE (+ обозначает индексируемое SRS поле, по которому пользователь может осуществлять поиск информации)

Название поля	Содержание поля
GENE_ID	Идентификатор карточки в блоке GENE (+)
DATE	Дата последнего редактирования
AUTHOR	Имена составителей карточки
GENE	Название и синонимы гена (+)
PRODUCT	Название и синонимы продукта гена (+)
TAXON	Таксономическая принадлежность (+)
SPECIES	Видовая принадлежность (+)
KEYWORDS	Ключевые слова (+)
ACTIVITY	Описание экспрессии гена (комментарий) (+)
PROMOTER_ID	Перекрестная ссылка на идентификатор соответствующей карточки в блоке PROMOTER (+)
SEQUENCE_ID	Перекрестная ссылка на идентификатор соответствующей карточки в блоке SEQUENCE (+)
REFERENCE	Ссылка на статью, из которой взяты данные (+)
PUBMED	Ссылка на БД публикаций Entrez-PUBMED (+)
END	Конец блока описания гена

Ниже приведены примеры аннотации заполнения карточек на основе литературных данных (табл. 5, 6, 7).

Таблица 5. Пример записи данных в блоке PROMOTER

<p>PROMOTER_ID Nt:PR2D_P2 GENE_ID Nt:PR2D TARGET SPECIES tobacco (Nicotiana tabacum) KEYWORDS stress response, tobacco mosaic virus-induced, salicylic acid-induced, tissue-specific, flower, developmentally-regulated, seedling, seed, cotyledon, sepal, placenta, ovary LOCALIZATION from -554 to +104; GenBank; X69794; TSS: 1707; from 1153 to 1810 DESCRIPTION The regulatory region upstream of the coding part of the reporter gene includes a promoter fragment (554 bp upstream of the transcription start site), the 5'UTR (63 bp), and 51 bp of the coding region of the Pr2d gene. SEQUENCE_ID Nt:PR2D_P2S REPORTER GUS STAGE_ORGAN_TISSUE leaf, flower, sepal, placenta, ovary REGULATOR tobacco mosaic virus, salicylic acid COMMENT A 5-fold increase in GUS activity (253 pmol 4-MU/min per mg protein) over background level was observed after flotation of leaf disk of six- to eight-week-old transgenic tobacco plants (N. tabacum cv. Xanthi) on water. Salicylic acid treatment increased GUS activity of the -554-promoter over 9-fold (2333 pmol 4-MU/min per mg protein). Tobacco mosaic virus infection increased the promoter activity over 18-fold. Deletion of sequences upstream of -554 did not qualitatively affect tissue-specific expression. GUS activity of the -554-construct appeared to be weaker than with longer 5' regions but was evident in the same</p>
--

tissues: sepal, placenta, ovary.

REFERENCE Hennig J., Dewey R.E., Cutt J.R., Klessig D.F. Pathogen, salicylic acid and developmental dependent expression of a beta-1,3-glucanase/GUS gene fusion in transgenic tobacco plants. Plant J., 1993, 4, 481-493.

PUBMED [8220491](#)

END

Таблица 6. Пример записи данных в блоке SEQUENCE

SEQUENCE_ID Nt:PR2D_P2S

PROMOTER_ID [Nt:PR2D_P2](#)

GENE_ID [Nt:PR2D](#)

SEQUENCE

-554

```

ttccaacg agggataaaa caaccttttc taataactcc catcaactaa tctaacaatt ttgtcatcta
tcactcacca attaatTTTg aaaagtgtac aaatttacc c gatgaactt ttgagtatag ttcagaatta
caagctagaa aaataatagt attaaattaa agtcacgaaa ttgctcaatc tcgcatgccc tatagtatag
gggcagcttc gacctccttc tccgaatatt cgaccatctt tttctcgtaa aatcaaacgt ctccatcact
gtactagatt tgaactttct acattttttt tccacttggt acataaaata cactctctat acttgggatg
acttggcttc actgggcttg acttggttat taactctact taatacactg tctatattcc atgacgtgct
tcaactgggta tgacttggtt aataaaatta cactattttt ttaactgctt cacagggcaa tggactttgg

```

+1

```

ctttttaagt cttttaggt cttcttctac ttgtctatat aagaagcagc ctaatgggtc ctgaaaCgca
caatttcagc tcaagtgtat ctaattattc tctcatTTTc attttagcta tggctttATG CATTAAAAAT
GGCTTCTTG CAGCTGCCCT TGTACTTGTT +104

```

END

Таблица 7. Пример записи данных в блоке GENE

GENE_ID Nt:PR2D

DATE 28.03.2012

AUTHOR Smirnova O.

GENE Pr2d

PRODUCT acidic beta-glucanase, acidic beta-(1->3)-glucanase, PR-2D, glucan endo-(1->3)-beta-glucosidase, (1->3)-beta-glucan endohydrolase

TAXON Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiids; Solanales; Solanaceae; Nicotiana.

SPECIES tobacco (Nicotiana tabacum)

KEYWORDS stress response, tobacco mosaic virus-induced, salicylic acid-induced, tissue-specific, flower, developmentally-regulated, seedling

ACTIVITY Expression was not found in healthy tissues. Infection of tobacco by tobacco mosaic virus and salicylate treatment strongly induce the acidic glucanase gene. Ethylene treatment and wounding have little effect on the acidic glucanase gene.

PROMOTER_ID [Nt:PR2D_P1](#) [Nt:PR2D_P2](#) [Nt:PR2D_P3](#)

SEQUENCE_ID [Nt:PR2D_P1S](#) [Nt:PR2D_P2S](#) [Nt:PR2D_P3S](#)

REFERENCE Hennig J., Dewey R.E., Cutt J.R., Klessig D.F. Pathogen, salicylic acid and developmental dependent expression of a beta-1,3-glucanase/GUS gene fusion in transgenic tobacco plants. Plant J., 1993, 4, 481-493.

PUBMED [8220491](#)

END

Следует отметить, что рабочий язык информационного ресурса английский, так как основная часть публикаций по биотехнологии растений идет на английском языке. При этом интерфейс TGP и соответствующий тьюториал созданы на русском языке.

ТЕХНОЛОГИИ РЕАЛИЗАЦИИ БД TGP

БД TGP управляется средствами СУБД SRS 6.1, которая развернута на сервере баз данных под управлением Red Hat Enterprise Linux 5.7. В качестве платформы для выполнения БД TGP использована система Sequence Retrieval System (SRS) [36], специально разработанная для формализованного описания биологических данных по заказу European Bioinformatics Institute. Большая часть БД EMBO расположена на этой платформе [37]. Средства SRS позволяют индексировать большинство полей в карточках блоков и эффективно осуществлять перекрестную связь полей в блоках БД, что необходимо для построения эффективных пользовательских запросов и свободной навигации между полями и карточками в различных блоках. 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).

```

$TGP_PROMOTER_DB=$Library:[TGP_PROMOTER group:$TGP_LIBS
  format:$TGP_PROMOTER_FORMAT maxNameLen:32
  files:{ $LibFile:tgp_promoter }
links:{
  $Link:[$TGP_PROMOTER_DB to:$TGP_GENE_DB
    fromField:$TGP_PROMF_GENE_ID toField:$TGP_GENEF_GENE_ID
  ]
  $Link:[$TGP_PROMOTER_DB to:$TGP_SEQUENCE_DB
    fromField:$TGP_PROMF_SEQUENCE_ID toField:$TGP_SEQF_SEQUENCE_ID
  ]
}
]

$TGP_PROMOTER_FORMAT=$LibFormat:[
  fileType:$TGP_PROMOTER_FILE
  syntax:$TGP_PROMOTER_SYNTAX
fields:{
  $Field:[$TGP_PROMF_PROMOTER_ID code:promoter_id index:id indexToken:promoter_id]
  $Field:[$TGP_PROMF_GENE_ID code:gene_id index:str indexToken:gene_id]
  $Field:[$TGP_PROMF_TSPEC code:tspec index:str indexToken:tspec]
  $Field:[$TGP_PROMF_KEYWORDS code:keywords index:str indexToken:keywords]
  $Field:[$TGP_PROMF_LOCALIZATION code:localization]
  $Field:[$TGP_PROMF_DESCRIPTION code:description]
  $Field:[$TGP_PROMF_SEQUENCE_ID code:sequence_id index:str indexToken:sequence_id]
  $Field:[$TGP_PROMF_REPORTER code:reporter index:str indexToken:reporter]
  $Field:[$TGP_PROMF_TRANSGENE code:transgene index:str indexToken:transgene]
  $Field:[$TGP_PROMF_STAGE code:stage index:str indexToken:stage]
  $Field:[$TGP_PROMF_REGULATOR code:regulator index:str indexToken:regulator]
  $Field:[$TGP_PROMF_COMMENT code:comment index:str indexToken:comment]
  $Field:[$TGP_PROMF_REFERENCE code:reference index:str indexToken:reference]
  $Field:[$TGP_PROMF_PUBMED code:pubmed index:str indexToken:pubmed]
}
]

$TGP_PROMF_PROMOTER_ID=$SrsField:[PromoterID short:pri group:$DF_ALL]
$TGP_PROMF_GENE_ID=$SrsField:[GeneID short:gni group:$DF_ALL]
$TGP_PROMF_TSPEC=$SrsField:['Target_species' short:tspec group:$DF_ALL]
$TGP_PROMF_KEYWORDS=$SrsField:[Keywords short:kw group:$DF_ALL]
$TGP_PROMF_LOCALIZATION=$SrsField:[Localization short:loc group:$DF_ALL]
$TGP_PROMF_DESCRIPTION=$SrsField:[Description short:desc group:$DF_ALL]
$TGP_PROMF_SEQUENCE_ID=$SrsField:[SequenceID short:seq group:$DF_ALL]
$TGP_PROMF_REPORTER=$SrsField:[Reporter short:rep group:$DF_ALL]
$TGP_PROMF_TRANSGENE=$SrsField:[Transgene short:transgene group:$DF_ALL]
$TGP_PROMF_STAGE=$SrsField:[Stage_organ_tissue short:stg group:$DF_ALL]
$TGP_PROMF_REGULATOR=$SrsField:[Regulator short:regulator group:$DF_ALL]
$TGP_PROMF_COMMENT=$SrsField:[Comment short:cmt group:$DF_ALL]
$TGP_PROMF_REFERENCE=$SrsField:[Reference short:ref group:$DF_ALL]
$TGP_PROMF_PUBMED=$SrsField:[Pubmed short:pmb group:$DF_ALL]

$TGP_PROMOTER_SYNTAX=$Syntax:[file:'SRSSITE:tgp_promoter.is']
$TGP_PROMOTER_FILE=$FileType:[typeName:dat maxline:200]

```

Рис. 2. Описание таблицы TGP_PROMOTER базы данных TGP на языке SRS.

Содержимое БД заполняется администратором БД по мере обновления информации БД. Средством заполнения является программа, написанная на языке Icarus системы SRS.

Пример описания таблицы TGP_PROMOTER БД TGP на языке SRS представлен на рисунке 2.

ОРГАНИЗАЦИЯ ВЕДЕНИЯ БД TGP

Системные программные средства, необходимые для работы оператора при заполнении БД TGP, должны быть представлены одной из операционных систем: Linux, Windows XP, Windows Vista, Windows 7. Необходимо наличие интернет-браузера Mozilla Firefox (версии не ниже 7.0) или Microsoft Internet Explorer (версии не ниже 8.0). В состав используемых аппаратных средств должны входить персональный компьютер, оснащенный процессором класса не ниже Intel Core 2/AMD Athlon 64 X2, объемом ОЗУ не менее 1 Гб с выходом в Интернет.

МОДУЛЬ ИНТЕРФЕЙСА БАЗЫ ДАННЫХ ПРОМОТОРОВ

Модуль интерфейса для БД TGP представляет собой программный компонент (модуль), обеспечивающий интерфейс пользователя с БД TGP. Доступ к БД TGP может быть осуществлен по адресу http://wwwmgs.bionet.nsc.ru/mgs/dbases/tgp_ru/home.html (рис. 3).



Рис. 3. Пользовательский интерфейс базы данных TGP, домашняя страница.

СОДЕРЖАНИЕ БД TGP

Содержание БД TGP отражено в специальном разделе. Текущий релиз базы содержит информацию о 289 промоторах, 289 нуклеотидных последовательностях и 158 генах. Представленные промоторы принадлежат 27 видам растений. Их активность описана более чем в 40 различных органах и тканях трансгенных растений. Список видов трансгенных растений, использованных для изучения активности промоторов,

включает 33 наименования. Описано действие более 40 видов регуляторов на активность промоторов.

Интерфейс пользователя базы позволяет выбирать потенциальные промоторы по нескольким полям, отмеченным знаком плюс в таблицах 2-4.

Чтобы ознакомиться с примерами типовых запросов в базе промоторов, в модуле интерфейса базы данных промоторов сделан выход в раздел тьюториала «Как искать».

Типичные примеры запросов для БД TGP:

- Найти промоторы, которые работают в конкретном виде растений (поиск в поле Target species);
- Найти промоторы, на которые влияет конкретный регулятор (поиск в поле Regulator);
- Найти промоторы, которые работают в конкретном виде растений и на которые влияет конкретный регулятор (поиск в полях Target species и Keywords и/или Regulator);
- Найти промоторы, выделенные из определенного вида растений (поиск в поле PromoterID);
- Найти промоторы, на которые влияют несколько различных регуляторов;
- Найти промоторы, которые активны в определенном органе или ткани (поиск в полях STAGE_ORGAN_TISSUE и/или COMMENT);
- Найти промоторы, которые активны в определенном органе или ткани (поиск в полях REGULATOR и STAGE_ORGAN_TISSUE или COMMENT).

Рассмотрим особенности работы интерфейса БД промоторов при выполнении этих запросов. Чтобы найти промоторы, которые работают в конкретном виде растений, необходимо перейти по ссылке Target_species на главной странице таблицы TGP_PROMOTER. Чтобы просмотреть список видов трансгенных растений, необходимо нажать кнопку «List Values» (рис. 4).

The screenshot shows the TGP_PROMOTER database interface. The main table lists various fields with their short names, types, and the number of entries and references. An arrow points to the 'Target species' field, which is expanded to show a search interface with a 'List Values' button.

Name	Short Name	Type	No of Keys	No of Entry References	Indexing Date	Status
PromoterID	pri	id	289			
GeneID	gni	index	158			
Target species	tspec	index	33			
Keywords	kw	index	145			
Localization	loc	show				
Description	dsc	show				
SequenceID	seq	index	289			
Reporter	rep	index	9			
Transgene	transgene	index	20			
Stage organ tissue	stg	index	89			
Regulator	regulator	index	66			
Comment	cmt	index	3399			
Reference	ref	index	169			
Pubmed	pmb	index	161			

The expanded 'Target species' field shows a search interface with a 'List Values' button and a search form: 'List Values' that match [] and occur in at least [] entries.

Рис. 4. Поиск организмов - мишеней для трансгенных промоторов, доступных в базе из таблицы TGP_PROMOTER.

Например, чтобы найти промоторы, активность которых была проверена в трансгенном ячмене, можно перейти по гиперссылке «barley (hordeum vulgare)».

Появляется новая страница с идентификаторами промоторов, активных в трансгенном ячмене (ID содержит информацию о соответствующем виде, например, Hv означает *Hordeum vulgare*, Ta, *Triticum aestivum*, Td, *Triticum turgidum subspecies durum*) (рис. 5).

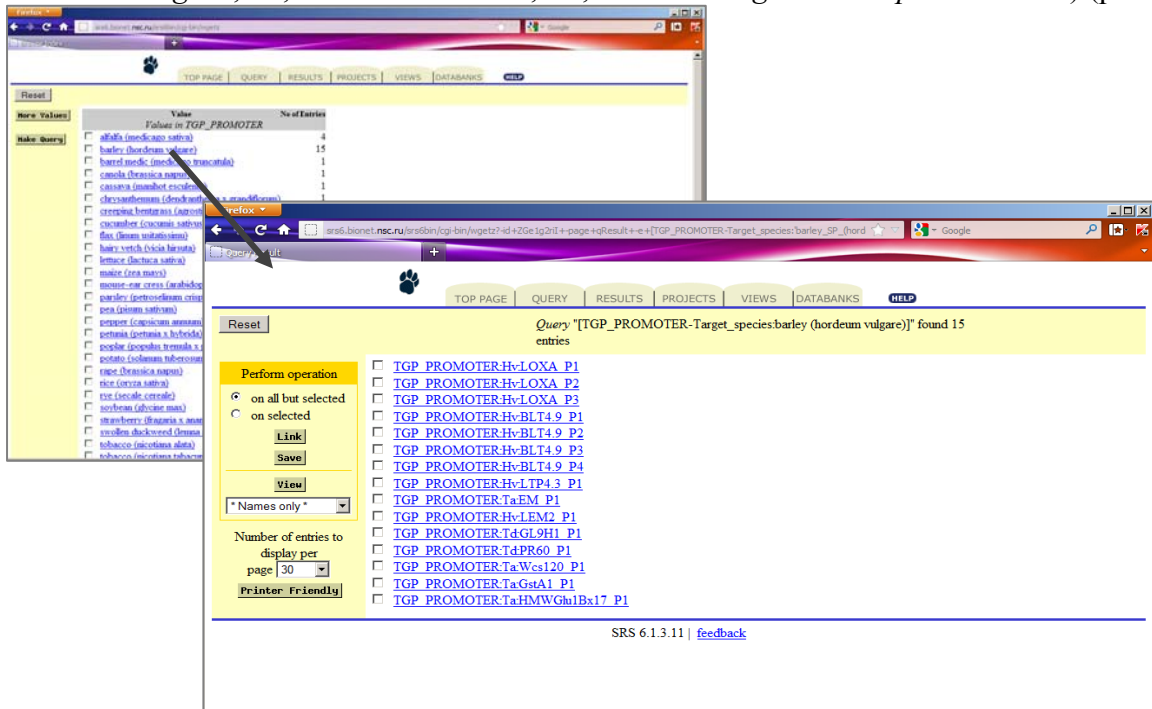


Рис. 5. Выбор списка промоторов, для которых генетические конструкции были использованы в трансгенном ячмене.

Чтобы просмотреть список регуляторов, необходимо выбрать поле «Regulator» на главной странице таблицы TGP_PROMOTER (рис. 4). На следующей странице необходимо нажать кнопку «List Values» (рис. 6).

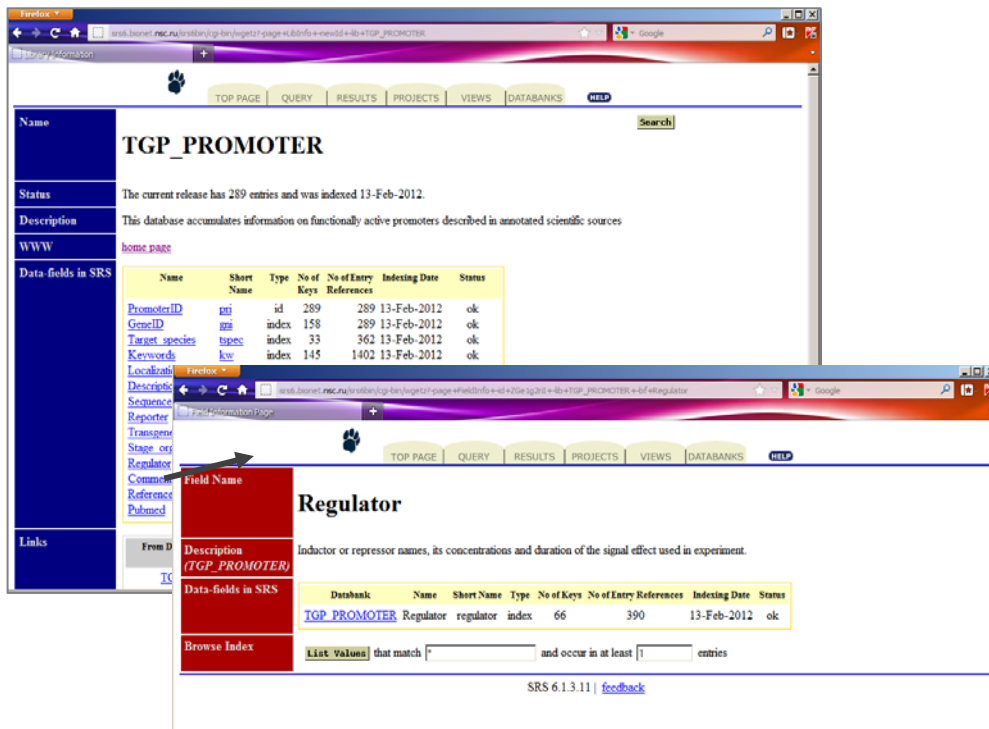


Рис. 6. Поиск регуляторов активности промоторов, доступных в базе из таблицы TGP_PROMOTER.

В результате будет получен список регуляторов активности промоторов. По этому списку можно найти, например, промоторы, чувствительные к низкой температуре, перейдя по ссылке «cold» для получения списка промоторов, регулируемых низкой температурой (рис. 7).

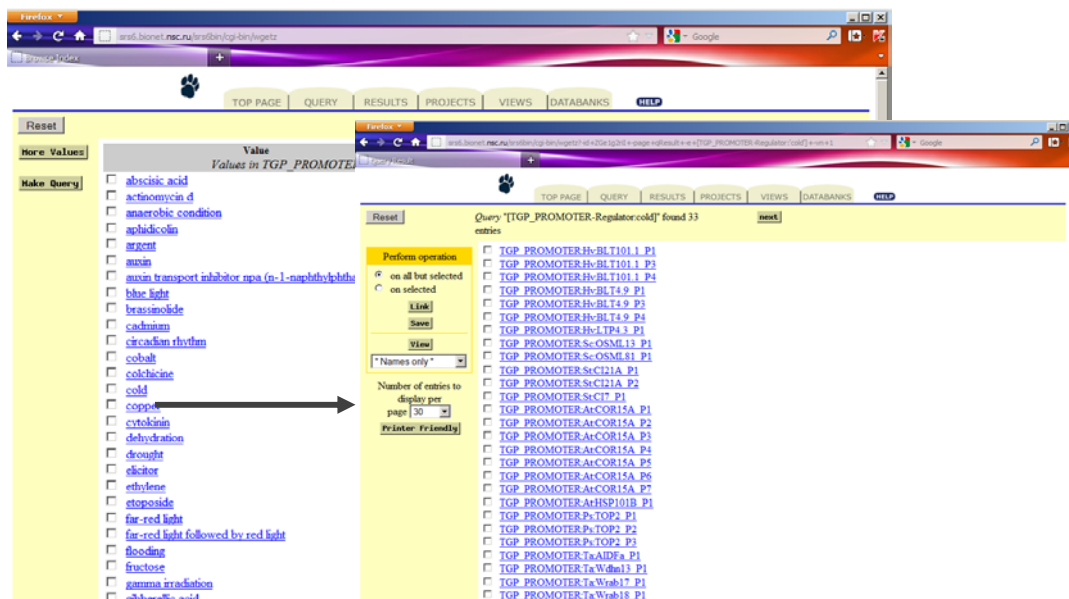


Рис. 7. Выбор списка промоторов, для которых активатором является холод.

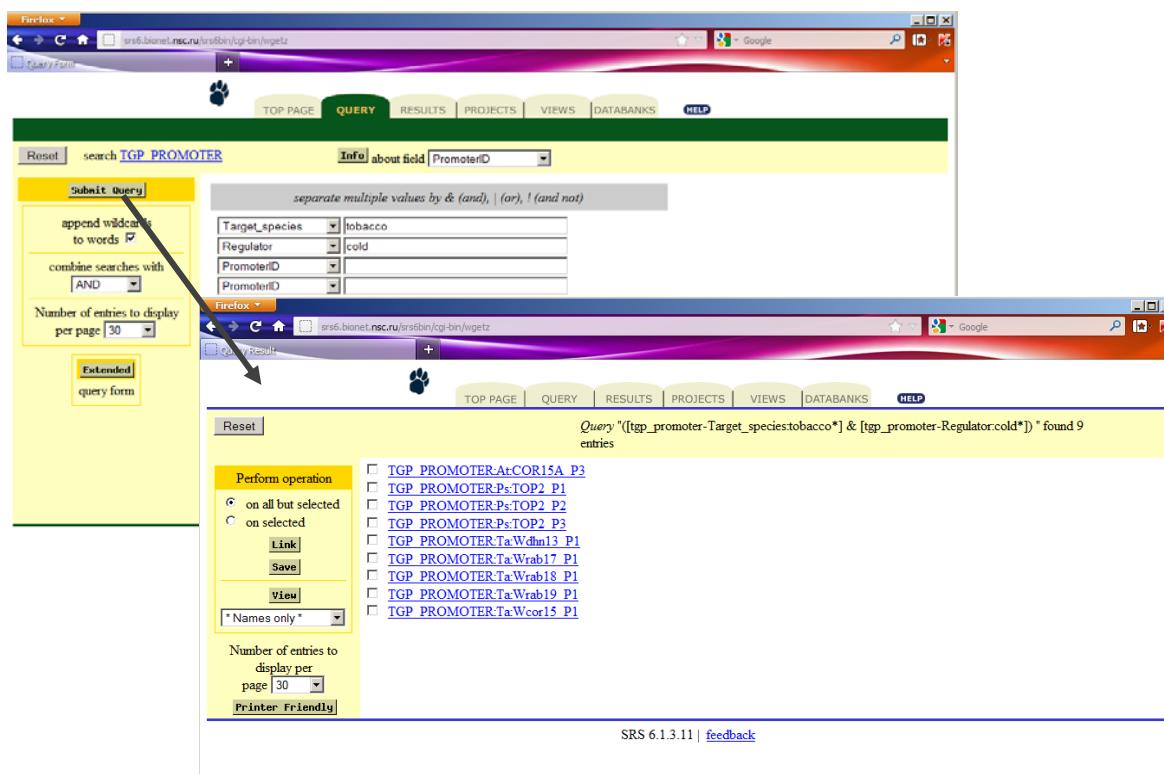


Рис. 8. Стандартная поисковая форма для таблицы TGP_PROMOTER.

Чтобы найти промоторы, которые активны в конкретном виде растений и активируются конкретным регулятором, необходимо перейти на страницу расширенного запроса (рис. 8), нажав кнопку «Search» на странице с типовым

описанием структуры полей TGP_PROMOTER. Поставленный запрос является «сложным» и для его выполнения необходимо объединить результаты поиска в нескольких полях одновременно с применением логических выражений.

Выбор комбинации запросов в этой форме осуществляется в меню «combine searches with». Выбор полей поиска для выполнения одновременного запроса осуществляется выпадающим меню в левой части поисковой формы. Для поиска мишеней, необходимо выбрать опцию «Target_species» и ввести название вида (например, «tobacco»). Затем в левом столбце выбирается поле Regulator, с указанием в правом столбце условия (например, cold). Запрос выполняется нажатием кнопки «Submit query». Результатом будет получен список записей базы данных, удовлетворяющих данному запросу (рис. 8).

Переход по ссылке из этого списка позволяет получить визуализацию информации о промоторе, а так же содержит ссылки на информацию о гене и его последовательности (рис. 9).



Рис. 9. Представление информации о промоторе, его последовательности и гене в БД TGP.

При работе с таблицей TGP_GENE можно решать следующие задачи:

- Найти гены, которые активируются или репрессируются определенным регулятором (поиск по полям KEYWORDS и ACTIVITY);
- Найти гены, которые экспрессируются в конкретных тканях или на определенных стадиях развития (поиск по полям KEYWORDS и ACTIVITY);
- Найти гены, кодирующие определенный белок (поиск по полю PRODUCT);
- Найти гены конкретного вида (поиск по полю SPECIES);
- Найти гены конкретного вида, чувствительные к определенному регулятору (поиск по полям SPECIES и KEYWORDS и/или ACTIVITY).

Например, чтобы найти в базе ген Delta1-pyrroline-5-carboxylate synthetase, необходимо с главной страницы базы перейти к работе с разделом TGP_GENE, затем нажать на странице этого раздела кнопку «Search». В форме запроса необходимо выбрать название поля «Gene» а в поле ввода текста либо название гена целиком, либо фрагмент названия, например, «*p5c*». После этого выполнить запрос, нажав кнопку «Submit Query». В результате система обработки запросов выдаст карточку гена TGP_GENE:At:P5CSA (рис. 10).

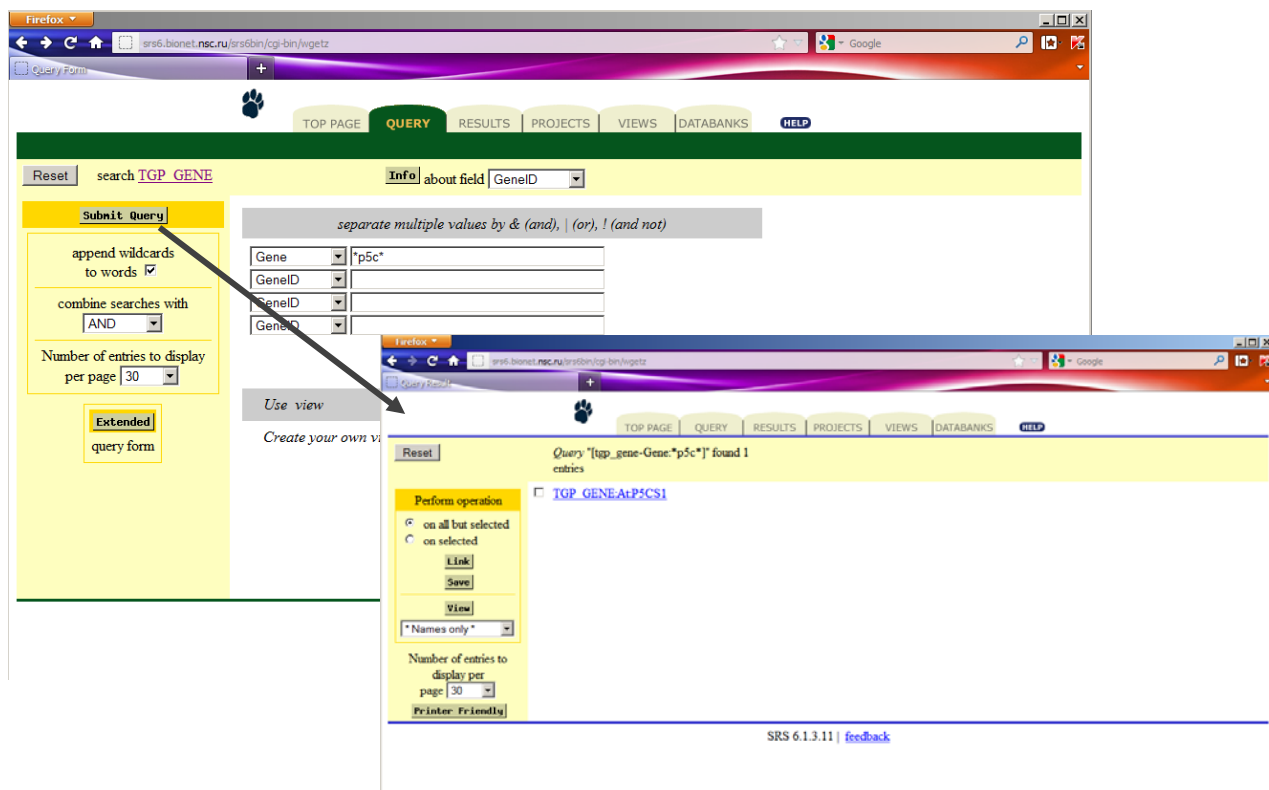


Рис. 10. Поиск гена в таблице TGP_GENE по фрагменту названия.

Таким образом, интерфейс БД TGP позволяет быстро находить информацию о промоторах - кандидатах генетических конструкций у растений, их генах и последовательностях.

ЗАКЛЮЧЕНИЕ

Разработанная БД представляет собой удобный инструмент для хранения, визуализации и поиска данных о промоторах. База TGP предназначена для накопления информации, необходимой для планирования генно-инженерных экспериментов с целью создания растительных организмов с качественно новыми или улучшенными свойствами. База содержит данные об экспрессии биотехнологически-значимых генов растений, активности их промоторов и делеционных мутантов этих промоторов, индукторах, влияющих на активность промоторов и нуклеотидных последовательностях описанных промоторов. Формат представления данных обеспечивает быстрый поиск ткане-, органо- и стадиспецифических растительных промоторов, чувствительных к различным индукторам, с целью обеспечения заданного пользователем паттерна транскрипции трансгена при решении конкретных генно-инженерных задач.

Следует также отметить, что TGP является одним из модулей информационного портала «Биотехнология растений», разрабатываемого в ИЦиГ СО РАН. Этот

информационный ресурс представляет собой платформу для решения различных задач в области генной инженерии и биотехнологии растений, содержащую модули разного типа. В частности, модуль TGP планируется использовать в комбинации с модулем TRANSIG, представляющим собой БД трансляционных энхансеров.

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). Описанная в работе БД TGP является модулем БДП информационного портала «Биотехнология растений», разработанного в рамках данного проекта.

СПИСОК ЛИТЕРАТУРЫ

1. Boehm R. Bioproduction of therapeutic proteins in the 21st century and the role of plants and plant cells as production platforms. *Ann. N Y Acad. Sci.* 2007. V. 1102. P. 121–134.
2. Xu J., Dolan M.C., Medrano G., Cramer C.L., Weathers P.J. Towards high-yield production of pharmaceutical proteins with plant cell suspension cultures. *Biotechnol. Adv.* 2011. V. 29. № 3. P. 278–299.
3. Chang H., Jones M.L., Banowitz G.M., Clark D.G. Overproduction of cytokinins in petunia flowers transformed with P(SAG12)-IPT delays corolla senescence and decreases sensitivity to ethylene. *Plant Physiol.* 2003. V. 132. № 4. P. 2174–2183.
4. Sykorova B., Kuresova G., Daskalova S., Trckova M., Hoyerova K., Raimanova I., Motyka V., Travnickova A., Elliott M.C., Kaminek M. Senescence-induced ectopic expression of the *A. tumefaciens ipt* gene in wheat delays leaf senescence, increases cytokinin content, nitrate influx, and nitrate reductase activity, but does not affect grain yield. *J. Exp. Bot.* 2008. V. 59. № 2. P. 377–387.
5. Xu Y., Tian J., Gianfagna T., Huang B. Effects of SAG12-ipt expression on cytokinin production, growth and senescence of creeping bentgrass (*Agrostis stolonifera* L.) under heat stress. *Plant Growth Regulation.* 2009. V. 57. P. 281–291.
6. Xu Y., Gianfagna T., Huang B. Proteomic changes associated with expression of a gene (*ipt*) controlling cytokinin synthesis for improving heat tolerance in a perennial grass species. *J. Exp. Bot.* 2010. V. 61. № 12. P. 3273–3289.
7. Zhang P., Wang W.Q., Zhang G.L., Kaminek M., Dobrev P., Xu J., Gruissem W. Senescence-inducible expression of isopentenyl transferase extends leaf life, increases drought stress resistance and alters cytokinin metabolism in cassava. *J. Integr. Plant Biol.* 2010. V. 52. № 7. P. 653–669.
8. Merewitz E.B., Gianfagna T., Huang B. Photosynthesis, water use, and root viability under water stress as affected by expression of SAG12-ipt controlling cytokinin synthesis in *Agrostis stolonifera*. *J. Exp. Bot.* 2011. V. 62. № 1. P. 383–395.
9. Furtado A., Henry R.J., Takaiwa F. Comparison of promoters in transgenic rice. *Plant Biotechnol. J.* 2008. V. 6. P. 679–693.
10. Qu le Q., Xing Y.P., Liu W.X., Xu X.P., Song Y.R. Expression pattern and activity of six glutelin gene promoters in transgenic rice. *J. Exp. Bot.* 2008. V. 59. P. 2417–2424.
11. Furtado A., Henry R.J., Pellegrineschi A. Analysis of promoters in transgenic barley and wheat. *Plant Biotechnol. J.* 2009. V. 7. P. 240–253.
12. Yilmaz A., Mejia-Guerra M.K., Kurz K., Liang X., Welch L., Grotewold E. AGRIS: the arabidopsis gene regulatory information server, an update. *Nucleic Acids Res.* 2011. V. 39. P. D1118–D1122. URL: <http://arabidopsis.med.ohio-state.edu/> (дата обращения: 10.07.2012).
13. Shahmuradov I.A., Gammerman A.J., Hancock J.M., Bramley P.M., Solovyev V.V. PlantProm: a database of plant promoter sequences. *Nucleic Acids Res.* 2003. V. 31. P. 114–117. URL:

- <http://linux1.softberry.com/berry.phtml?topic=plantprom&group=data&subgroup=plantprom> (дата обращения: 10.07.2012).
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. URL: <http://www.dna.affrc.go.jp/PLACE/signalscan.html> (дата обращения: 10.07.2012).
 15. Chang W.C., Lee T.Y., Huang H.D., Huang H.Y., Pan R.L. PlantPAN: plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. *BMC Genomics.* 2008. V. 9. P. 561. URL: <http://plantpan.mbc.nctu.edu.tw> (дата обращения: 10.07.2012).
 16. O'Connor T.R., Dyreson C., Wyrick J.J. Athena: a resource for rapid visualization and systematic analysis of Arabidopsis promoter sequences. *Bioinformatics.* 2005. V. 21. P. 4411–4413. URL: <http://www.bioinformatics2.wsu.edu/Athena> (дата обращения: 10.07.2012).
 17. Morris R.T., O'Connor T.R., Wyrick J.J. Osiris: an integrated promoter database for *Oryza sativa* L. *Bioinformatics.* 2008. V. 24. P. 2915–2917. URL: <http://www.bioinformatics2.wsu.edu/Osiris> (дата обращения: 10.07.2012).
 18. Bülow L., Brill Y., Hehl R. AthaMap-assisted transcription factor target gene identification in *Arabidopsis thaliana*. *Database (Oxford)*. 2010. baq034. URL: http://www.athamap.de/gene_ident.php (дата обращения: 10.07.2012).
 19. Tittarelli A., Santiago M., Morales A., Meisel L.A., Silva H. Isolation and functional characterization of cold-regulated promoters, by digitally identifying peach fruit cold-induced genes from a large EST dataset. *BMC Plant Biol.* 2009. V. 9. P. 121.
 20. Potenza C., Aleman L., Sengupta-Gopalan C. Targeting transgene expression in research, agricultural, and environmental applications: promoters used in plant transformation. *In Vitro Cell Dev. Biol. Plant.* 2004. V. 40. P. 1–22.
 21. Jones H.D., Sparks C.A. Promoter sequences for defining transgene expression. *Methods Mol. Biol.* 2009. V. 478. P. 171–184.
 22. Peremarti A., Twyman R.M., Gyzem-Galera S., Naqvi S., Farrq G., Sabalza M., Miralpeix B., Dashevskaya S., Yuan D., Ramessar K., Christou P., Zhu C., Bassie L., Capell T. Promoter diversity in multigene transformation. *Plant Mol. Biol.* 2010. V. 73. P. 363–378.
 23. Xiao Y.L., Redman J.C., Monaghan E.L., Zhuang J., Underwood B.A., Moskal W.A., Wang W., Wu H.C., Town C.D. High throughput generation of promoter reporter (GFP) transgenic lines of low expressing genes in *Arabidopsis* and analysis of their expression patterns. *Plant Methods.* 2010. V. 6. P. 18.
 24. Tittarelli A., Milla L., Vargas F., Morales A., Neupert C., Meisel L.A., Salvo-G H., Penaloza E., Munoz G., Corcuera L.J., Silva H. Isolation and comparative analysis of the wheat TaPT2 promoter: identification in silico of new putative regulatory motifs conserved between monocots and dicots. *J. Exp. Bot.* 2007. V. 58. P. 2573–2582.
 25. Van Herpen T.W., Riley M., Sparks C., Jones H.D., Gritsch C., Dekking E.H., Hamer R.J., Bosch D., Salentijn E.M., Smulders M.J., Shewry P.R., Gilissen L.J. Detailed analysis of the expression of an alpha-gliadin promoter and the deposition of alpha-gliadin protein during wheat grain development. *Ann. Bot.* 2008. V. 102. P. 331–342.
 26. Piston F., Marin S., Hernando A., Barro F. Analysis of the activity of a gamma-gliadin promoter in transgenic wheat and characterization of gliadin synthesis in wheat by MALDI-TOF during grain development. *Mol. Breed.* 2009. V. 23. P. 655–667.
 27. Vickers C.E., Xue G., Gresshoff P.M. A novel cis-acting element, ESP, contributes to high-level endosperm-specific expression in an oat globulin promoter. *Plant Mol. Biol.* 2006. V. 62. P. 195–214.
 28. Chen L., Tu Z., Hussain J., Cong L., Yan Y., Jin L., Yang G., He G. Isolation and heterologous transformation analysis of a pollen-specific promoter from wheat (*Triticum aestivum* L.). *Mol. Biol. Rep.* 2010. V. 37. P. 737–744.

29. Qu L.Q., Takaiwa F. Evaluation of tissue specificity and expression strength of rice seed component gene promoters in transgenic rice. *Plant Biotechnol. J.* 2004. V. 2. P. 113–125.
30. Tiwari S., Spielman M., Day R.C., Scott R.J. Proliferative phase endosperm promoters from *Arabidopsis thaliana*. *Plant Biotechnol. J.* 2006. V. 4. P. 393–407.
31. Eskelin K., Ritala A., Suntio T., Blumer S., Holkeri H., Wahlström E.H., Baez J., Mäkinen K., Maria N.A. Production of a recombinant full-length collagen type I alpha-1 and of a 45-kDa collagen type I alpha-1 fragment in barley seeds. *Plant Biotechnol. J.* 2009. V. 7. P. 657–672.
32. Xu Z.S., Ni Z.Y., Liu L., Nie L.N., Li L.C., Chen M., Ma Y.Z. Characterization of the TaAIDFa gene encoding a CRT/DRE-binding factor responsive to drought, high-salt, and cold stress in wheat. *Mol. Genet. Genomics.* 2008. V. 280. P. 497–508.
33. Egawa C., Kobayashi F., Ishibashi M., Nakamura T., Nakamura C., Takumi S. Differential regulation of transcript accumulation and alternative splicing of a DREB2 homolog under abiotic stress conditions in common wheat. *Genes Genet. Syst.* 2006. V. 81. P. 77–91.
34. Kobayashi F., Ishibashi M., Takumi S. Transcriptional activation of Cor/Lea genes and increase in abiotic stress tolerance through expression of a wheat DREB2 homolog in transgenic tobacco. *Transgenic Res.* 2008. V. 17. P. 755–767.
35. Yang W., Liu X.D., Chi X.J., Wu C.A., Li Y.Z., Song L.L., Liu X.M., Wang Y.F., Wang F.W., Zhang C., Liu Y., Zong J.M., Li H.Y. Dwarf apple MbDREB1 enhances plant tolerance to low temperature, drought, and salt stress via both ABA-dependent and ABA-independent pathways. *Planta.* 2011. V. 233. P. 219–229.
36. Zdobnov E.M., Lopez R., Apweiler R. and Etzold T. The EBI SRS server - recent developments. *Bioinformatics.* 2002. V. 18. P. 368–373.
37. *Databases at the EBI.* URL: <http://www.ebi.ac.uk/Databases/> (дата обращения: 10.07.2012).

Материал поступил в редакцию 16.07.2012, опубликован 30.07.2012.