

База данных генов, повышающих устойчивость пшеницы и родственных ей злаков к патогенным грибам

О.Г. Смирнова , А.В. Кочетов

Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия
Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

Растения подвергаются воздействию огромного количества патогенных грибов. В последние годы идентифицировано большое количество генов, принимающих участие в формировании защитного ответа у пшеницы, и проведено секвенирование некоторых генов устойчивости. Секвенирование аллельных вариантов генов устойчивости у пшеницы и ее сородичей необходимо для понимания молекулярных механизмов формирования защитного ответа. В настоящей работе описывается информационный ресурс, предназначенный для сбора и хранения информации о генах, обеспечивающих устойчивость пшеницы и родственных ей злаков к заболеваниям, вызванным грибными патогенами. База PRG (Pathogenesis-Related Genes) содержит структурированную информацию о регуляции активности генов устойчивости. В базе представлены нуклеотидные последовательности генов устойчивости, а также данные об отдельных однонуклеотидных полиморфизмах генов, ассоциированных с различными уровнями устойчивости к заболеваниям. Собрана информация о продукте гена и его функциях. Суммированы данные о патогене и заболевании, хромосомной локализации гена устойчивости, даны ссылки на карточки в сопутствующих базах данных нуклеотидных последовательностей (GenBank) и белков (UniProt). Приводится важная информация об изменении экспрессии гена в ответ на действие патогенов, гормонов, а также на изменения условий окружающей среды. Информация вносится в базу в результате аннотирования научных публикаций. В настоящее время в базе данных содержится информация о 66 генах, представленных 75 аллельными вариантами. База PRG разработана на платформе SRS (Sequence Retrieval System). Средства SRS позволяют проводить множественный поиск по полям базы, что необходимо для эффективного построения запросов. Система SRS автоматически генерирует Web-интерфейс с поисковыми таблицами и визуализирует результаты поиска, которые могут быть сохранены в текстовом формате. Собранная информация и удобный интерфейс позволяют широкому кругу исследователей расширить представления о разнообразии генов, принимающих участие в формировании защитного ответа, способствуя тем самым эффективному созданию форм растений с повышенной устойчивостью к заболеваниям. База PRG доступна по адресу <http://srs6.bionet.nsc.ru/srs6bin/cgi-bin/wgetz?-page+top+-newld>.

Ключевые слова: база данных; информационный ресурс; гены устойчивости; патогенные грибы; зерновые культуры.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Смирнова О.Г., Кочетов А.В. База данных генов, повышающих устойчивость пшеницы и родственных ей злаков к патогенным грибам. Вавиловский журнал генетики и селекции. 2016;20(6):904-908. DOI 10.18699/VJ16.203

HOW TO CITE THIS ARTICLE:

Smirnova O.G., Kochetov A.V. A database on genes increasing resistance of wheat and relative species against pathogenic fungi. Vavilovskii Zhurnal Genetiki i Selektzii = Vavilov Journal of Genetics and Breeding. 2016;20(6):904-908. DOI 10.18699/VJ16.203

REVIEW

Received 07.10.2016 г.

Accepted for publication 29.11.2016 г.

© AUTHORS, 2016

 e-mail: planta@bionet.nsc.ru

A database on genes increasing resistance of wheat and relative species against pathogenic fungi

O.G. Smirnova , A.V. Kochetov

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia

Fungi belong to the major plant pathogens and investigation of plant resistance genes is a quite important task. During the last years many wheat resistance genes were identified. However, the sequencing of the *Triticum aestivum* L. genome is still going on and the nucleotide sequences of most resistance genes are not yet known. In addition, the study of allelic variants of resistance genes is important for better understanding of the molecular mechanisms of their action. In this paper we present an information resource for accumulation of data on sequenced genes of wheat and its relatives providing resistance against diseases caused by fungal pathogens. The database (Pathogenesis-Related Genes, PRG) contains information on gene chromosomal localizations and functional activities, nucleotide sequences and single nucleotide polymorphisms associated with their effects. PRG provides data on the proteins encoded, pathogens and diseases, as well as on the resistance gene expression patterns in response to pathogen inoculations, exposure to hormones and various external stimuli. It also has cross-references with related entries to the databases on nucleotide sequences (GenBank) and proteins (UniProt). Information entered into the database is a result of the annotation of scientific publications and manual curation. Currently PRG compiles data on 75 allelic variants of 66 resistance genes. The PRG database was developed on the basis of the SRS (Sequence Retrieval System) platform. This system allows the use of complex queries and visualization tools and automatically generates www-interface with the information in table or text formats. PRG may be useful for researchers studying plant biology or breeding new plant cultivars resistant to fungal diseases. It is available at the address: <http://srs6.bionet.nsc.ru/srs6bin/cgi-bin/wgetz?-page+top+-newld>.

Key words: databases; resistance gene; pathogenic fungus; crops.

Увеличение производства зерна мягкой пшеницы *Triticum aestivum* L. может осуществляться не только за счет роста урожайности, но и за счет снижения потерь, связанных с заболеваниями, вызываемыми грибными патогенами. Потери урожая при поражении листовой ржавчиной могут достигать 30–50 % (McIntosh et al., 1995). Создание устойчивых сортов является сложной задачей. Механизмы защиты растений связаны с функционированием генов устойчивости (R-генов – resistance genes) и их аналогов, обеспечивающих в первую очередь узнавание молекул патогена по присутствию в составе R-белков консервативных аминокислотных доменов. Среди изученных R-генов наиболее многочисленный класс представляют гены устойчивости, кодирующие белки, содержащие NBS (Nucleotide Binding Site) и LRR (Leucine-Rich Repeat) домены (Sanseverino et al., 2013). Тем не менее существенные компоненты защитной системы растений остаются еще не изученными. Более того, быстрая эволюционная изменчивость патогенов в некоторых случаях опережает усилия специалистов по защите растений и приводит к неэффективности расоспецифических генов устойчивости (Singh et al., 2005). Отдельный интерес представляют гены, формирующие неспецифическую устойчивость, поскольку она обеспечивает длительный период эффективности сразу к нескольким патогенам, что делает ее предпочтительной для селекционеров и исследователей. У пшеницы подробно охарактеризованы всего четыре гена *Lr34*, *Lr46*, *Lr67* и *Lr27*, определяющих замедленное развитие сразу нескольких заболеваний, два из которых, *Lr34* и *Lr67*, были секвенированы (Rosewarne et al., 2006; Mago et al., 2011; Spielmeyer et al., 2013).

К настоящему времени описано более 70 генов *Lr* (leaf rust), контролирурующих устойчивость пшеницы к листовой ржавчине и расположенных в разных хромосомах (McIntosh et al., 2013), и только четыре из них секвенировано. В связи с отсутствием данных о первичной структуре преобладающего большинства генов устойчивости к листовой ржавчине существует ряд проблем, связанных с маркированием этих генов. На сегодняшний день наиболее распространены маркеры, полученные исходя из информации о локализации генов на хромосомах. Эффективность таких маркеров снижается из-за возможных процессов рекомбинации между маркером и геном. Поэтому наиболее функциональными являются маркеры, разработанные на основании секвенирования соответствующих нуклеотидных последовательностей ДНК. Таким образом, для эффективной защиты пшеницы от патогенных грибов необходимо вести поиск новых генов устойчивости, используя данные по генам устойчивости у наиболее изученных в этом направлении видов растений и доступные сведения по структуре геномных нуклеотидных последовательностей пшеницы.

Компьютерные базы данных активно применяются для информационного обеспечения экспериментальных исследований, а также для поиска новых знаний с помощью методов компьютерной системной биологии. Разработано большое число целевых баз данных, без использования которых планирование экспериментов и интерпретация полученных результатов затруднительны. В области изучения генов устойчивости растений создана база Plant

Resistance Genes (<http://prgdb.org>), в которой приведены нуклеотидные последовательности 112 известных и более чем 100 000 предполагаемых R-генов, присутствующих у 233 видов растений (Sanseverino et al., 2013). Также следует отметить базу данных PathoPlant (Bülow et al., 2007), где аннотирована информация о генах *Arabidopsis thaliana*, экспрессия которых усиливается в ответ на инокуляцию различными патогенами.

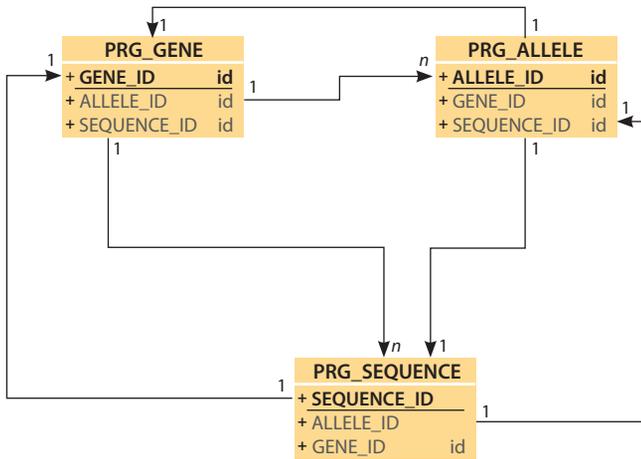
Нами создана целевая база данных генов (и аллелей генов) пшеницы, контролирующих устойчивость к фитопатогенным грибам. Было решено сконцентрироваться на аннотации генов устойчивости пшеницы и родственных ей злаков, поскольку пшеница является важной зерновой культурой и многие гены устойчивости были интрогрессированы в мягкую пшеницу из дикорастущих видов. В базу также внесена информация о генах, экспрессия которых изменяется после действия патогена. Продукты этих генов принимают участие в передаче сигнала и формировании защитного ответа, возникающего в клетке после контакта с патогеном. Концентрация такой информации в базе данных позволяет использовать ее в качестве базы знаний для анализа механизмов формирования защитного ответа пшеницы к действию патогенных грибов. В базу вносятся информация, полученная при анализе научных публикаций. Структура базы данных PRG, созданной на платформе SRS (Sequence Retrieval System), позволяет пользователю составлять сложные запросы и получать комплексную информацию, необходимую для разработки собственных маркеров при картировании аллелей генов устойчивости разной специфичности. Такая информация актуальна для планирования селекционно-генетических экспериментов при создании устойчивых сортов пшеницы в связи с появлением новых рас патогенных грибов и необходимостью постоянной работы по обновлению существующих сортов сельскохозяйственных растений.

Логическая структура и формат базы данных PRG

База данных PRG построена на платформе SRS и включает в себя три составляющие: таблицу генов PRG_GENE, таблицу аллелей PRG_ALLELE и таблицу последовательностей PRG_SEQUENCE, взаимосвязи между которыми устанавливаются согласно схеме (см. рисунок).

Таблица генов PRG_GENE включает информацию о названии организма, гена, локализации гена на хромосоме, патогене, заболевании и состоит из 14 полей. Таблица аллелей PRG_ALLELE содержит информацию о чувствительности к патогену, связанной с аллельным состоянием гена, и состоит из 12 полей. Таблица последовательностей PRG_SEQUENCE содержит информацию о последовательностях аллельных вариантов генов, описанных в таблице PRG_ALLELE, и состоит из 7 полей. Подробная информация о составе полей базы и их содержании приведена в табл. 1–3. Примеры заполнения полей представлены в табл. 4 и 5.

Все поля в таблицах PRG_GENE и PRG_ALLELE, за исключением поля END, являются индексируемыми, т. е. пользователь может осуществлять поиск информации по этим полям.



Schematic representation of the PRG database. Cross-indexed fields of the PRG_GENE, PRG_ALLELE, and PRG_SEQUENCE tables are shown.

Table 1. Fields of the PRG_GENE table

Field name	Description
GENE_ID	Gene identifier
ALLELE_ID	Cross-reference to ALLELE_ID in the PRG_ALLELE table
SEQUENCE_ID	Cross-reference to SEQUENCE_ID in the PRG_SEQUENCE table
DISEASE	Disease
PATHOGEN	Causative agent of the disease
HOST_SPECIES	The common and Latin names of the species
GENE	Gene names
MAPPING	Gene mapping on the chromosome
PRODUCT	Gene product
COMMENT	Comment
KEYWORDS	Keywords
REFERENCE	Bibliographic reference
PUBMED	Cross-reference to PUBMED
END	End of the entry

Технологии реализации и организация ведения базы PRG

База PRG управляется средствами системы SRS 6.1.3, которая развернута на сервере баз данных под управлением CentOS Linux 6.6. Система SRS специально разработана для формализованного описания биологических данных (Zdobnov et al., 2002). Средства SRS позволяют проводить множественный поиск по полям базы, что необходимо для эффективного построения запросов пользователями. Система SRS автоматически генерирует Web-интерфейс с поисковыми таблицами и визуализирует результаты поиска, которые могут быть сохранены в текстовом формате. Средства SRS позволяют индексировать большинство информационных полей и эффективно осуществлять

Table 2. Fields of the PRG_ALLELE table

Field name	Description
ALLELE_ID	Allele identifier
SEQUENCE_ID	Cross-reference to SEQUENCE_ID in the PRG_SEQUENCE table
GENE_ID	Cross-reference to GENE_ID in the PRG_GENE table
CULTIVAR	Variety or mutant strain
RESISTANCE	Disease resistance
GENBANK	Cross-reference to GenBank
UNIPROT	Cross-reference to UniProt
COMMENT	Comment
KEYWORDS	Keywords
REFERENCE	Bibliographic reference
PUBMED	Cross-reference to PUBMED
END	End of the entry

Table 3. Fields in the PRG_SEQUENCE section

Field name	Description
SEQUENCE_ID	Sequence identifier
ALLELE_ID	Cross-reference to ALLELE_ID in the PRG_ALLELE table
GENE_ID	Cross-reference to GENE_ID in the PRG_GENE table
LOCALIZATION	Gene positions according to the GenBank notation
COMMENT	Comment
SEQUENCE	Sequence
END	End of the entry

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 базе данных (формы запроса, визуализация данных, гиперссылки на документы в базе и интернет-ресурсы, настройки способа визуализации). Эксперт-аннотатор заполняет содержимое полей в записи базы данных на основании анализа научной литературы. Данные представляются для каждой таблицы отдельно в виде текстового файла, как показано в табл. 1–3. Файлы с информацией передаются администратору системы, который вводит их согласно инструкции системы SRS. Структура базы данных позволяет расширять список аннотируемых генов по мере появления новых литературных данных.

Пользовательский интерфейс

Модуль интерфейса является программным компонентом и служит для взаимодействия пользователя с базой данных. Модуль интерфейса для базы данных PRG обеспе-

Table 4. An exemplary entry in the PRG_GENE table

Field name	Description
GENE_ID	TaLr34
ALLELE_ID	TaLr34-CS TaLr34-Renan
SEQUENCE_ID	TaLr34-CS_S TaLr34-Renan_S
DISEASE	Leaf rust, stripe rust, powdery mildew, stem rust
PATHOGEN	<i>Puccinia triticina</i> , <i>Puccinia striiformis</i> (tritici), <i>Blumeria graminis</i> (tritici), <i>Puccinia graminis</i>
HOST_SPECIES	<i>Triticum aestivum</i> (bread wheat)
GENE	Lr34, Lr34 allele D, Lr34-D, Yr18, Sr57, Pm38, Sb1, Ltn1
MAPPING	Chromosome 7DS
PRODUCT	LR34 protein, ABC transporter, ATF-binding cassette transporter, adenosine triphosphate-binding cassette transporter
COMMENT	Alleles of <i>Lr34</i> conferring resistance or susceptibility differ by three genetic polymorphisms. The <i>Lr34</i> gene, which functions in the adult plant, stimulates senescence-like processes in the flag leaf tips and edges. The <i>Lr34</i> gene, then called <i>Yr18</i> , also confers moderate resistance to stripe or yellow rust (<i>Puccinia striiformis</i>) and powdery mildew (<i>Blumeria graminis</i>), where it is called <i>Pm38</i> . The locus is further associated with tolerance to stem rust (<i>Puccinia graminis</i>) and barley yellow dwarf virus, Bdv1. Flag leaves of many wheat cultivars containing <i>Lr34</i> in certain environments develop a necrotic leaf tip. This morphological marker is referred to as leaf tip necrosis
KEYWORDS	Multiple resistance, leaf rust, stripe rust, powdery mildew, chromosome 7DS, Lr34
REFERENCE	Krattinger SG, Lagudah ES, Spielmeier W, Singh RP, Huerta-Espino J, McFadden H, Bossolini E, Selter LL, Keller B. A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. <i>Science</i> . 2009, 323(5919):1360-3
PUBMED	19229000
END	

Table 5. An exemplary entry in the PRG_ALLELE table

Field name	Description
ALLELE_ID	TaLr34-CS
SEQUENCE_ID	TaLr34-CS_S
GENE_ID	TaLr34
CULTIVAR	Chinese Spring
RESISTANCE	Resistant (leaf rust)
GENBANK	FJ436983
UNIPROT	COJSA8
COMMENT	Allelic variation in <i>Lr34</i> between the resistant form in cv. Chinese Spring and the susceptible form in cv. Renan involves three polymorphisms. The first one is an A/T single-nucleotide polymorphism (SNP) in intron 4, the second one is a 3-bp ins/del in exon 11, and the third one is a C/T SNP in exon 12, where an amino acid is altered. No difference within 2 kb of the putative <i>Lr34</i> promoter region was detected between the two alleles
KEYWORDS	Chinese Spring, resistant, Lr34
REFERENCE	Krattinger SG, Lagudah ES, Spielmeier W, Singh RP, Huerta-Espino J, McFadden H, Bossolini E, Selter LL, Keller B. A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. <i>Science</i> . 2009, 323(5919):1360-3
PUBMED	19229000
END	

чивает поиск и визуализацию информации о генах устойчивости растений. Разработаны словари для поисковых полей базы данных по генам устойчивости к патогенным грибам. Пользователю предоставляется возможность просматривать списки видов растений, генов и их продуктов, сортов и мутантных линий, устойчивых и чувствительных

генотипов, заболеваний и вызывающих их патогенных организмов, а также ключевых слов, которые содержатся в базе. Кроме этого, пользователь имеет возможность осуществлять поиск по большому числу поисковых полей таблиц PRG_GENE и PRG_ALLELE. Поиск проводится отдельно для таблиц PRG_GENE и PRG_ALLELE.

Поиск в таблице PRG_ALLELE позволяет найти сорта, устойчивые или чувствительные к определенным заболеваниям (поиск по полю RESISTANCE). Затем пользователь имеет возможность получить нуклеотидную последовательность аллельного варианта гена, обеспечивающего устойчивость, кликнув на идентификатор последовательности в поле SEQUENCE_ID таблицы PRG_ALLELE.

Поиск в таблице PRG_GENE позволяет находить гены по следующим критериям:

- названию (поиск по полю GENE);
- расположению в определенной хромосоме (поиск по полям MAPPING и KEYWORDS);
- принадлежности определенному организму (поиск по полю HOST_SPECIES);
- участию в формировании защитного ответа при действии определенного патогена (поиск по полю PATHOGEN);
- участию в формировании защитного ответа при определенном заболевании (поиск по полю DISEASE).

Если найденный сигнал удовлетворяет требованиям, пользователь может получить дополнительную информацию об уровне экспрессии гена из поля COMMENT и нуклеотидную последовательность из поля SEQUENCE_ID таблицы PRG_GENE. В таблице PRG_GENE существует возможность проводить комплексный поиск по нескольким полям. Для этого необходимо перейти на страницу сложного запроса, нажав кнопку «Search» на странице с типовым описанием структуры полей таблицы PRG_GENE. На этой странице меню «Combine searches with» задает логическую операцию для выполнения совместных запросов по нескольким полям. Например, для поиска генов, устойчивых к действию *Puccinia triticina* и локализованных в пятой хромосоме пшеницы, необходимо установить значение этого меню «AND», в левом столбце формы запроса выбрать поле «PATHOGEN», напротив в правом столбце ввести его название – *Puccinia triticina*. В следующей строке формы в левом столбце выбрать «MAPPING», а в правом ввести часть названия хромосомной локализации со звездочкой, чтобы сделать расширенный поиск (*5), и выполнить запрос. В результате выполнения запроса система выдает список генов, расположенных в 5 хромосоме, которые принимают участие в защитной реакции на действие патогена *P. triticina*, вызывающего листовую ржавчину, например PRG_GENE:TaLr1.

Заключение

Разработана база данных PRG, предназначенная для накопления информации, необходимой для развития представлений о механизмах формирования защитного ответа у растений и поиска новых генов устойчивости. База данных PRG содержит структурированную информацию

о генах устойчивости и защитного ответа. Текущая версия базы данных генов устойчивости к фитопатогенам состоит из трех таблиц (генов, аллелей, нуклеотидных последовательностей) и содержит 216 уникальных записей. Формат представления данных обеспечивает быстрый поиск последовательностей генов, принимающих участие в формировании защитного ответа на действие патогенных грибов у пшеницы и родственных ей злаков, что необходимо для создания дополнительных маркеров, обеспечивающих поиск новых генов устойчивости у растений.

Acknowledgments

This work was supported by the Ministry of Education and Science of the Russian Federation, grant agreement 14.604.21.0107 of August 7, 2014, unique identifier RFMEFI60414X0107.

Conflict of interest

The authors declare no conflict of interest.

References

- Bülow L., Schindler M., Hehl R. PathoPlant[®]: a platform for microarray expression data to analyze co-regulated genes involved in plant defense responses. *Nucl. Acids Res.* 2007;35(1):D841-45. DOI 10.1093/nar/gkl835.
- Mago R., Tabe L., McIntosh R.A., Pretorius Z., Kota R., Paux E., Wicker T., Breen J., Lagudah E.S., Ellis J.G., Spielmeier W. A multiple resistance locus on chromosome arm 3BS in wheat confers resistance to stem rust (*Sr2*), leaf rust (*Lr27*) and powdery mildew. *Theor. Appl. Genetics.* 2011;123(4):615-623. DOI 10.1007/s00122-011-1611-y.
- McIntosh R.A., Wellings C.R., Park R.F. Wheat Rusts: An Atlas of Resistance Genes. Australia, Victoria: CSIRO Publ., 1995.
- McIntosh R.A., Yamazaki Y., Dubcovsky J., Rogers J., Morris C., Somers D.J., Appels R., Devos K.M. Catalogue of gene symbols for wheat. 2013. <http://www.shigen.nig.ac.jp/wheat/komugi/genes/symbolClassList.jsp> (Accessed 19 Sept. 2016).
- Rosewarne G.M., Singh R.P., Huerta-Espino J., William H.M., Bouchet S., Cloutier S., McFadden H., Lagudah E.S. Leaf tip necrosis, molecular markers and β 1-proteasome subunits associated with the slow rusting resistance genes *Lr46/Yr29*. *Theor. Appl. Genetics.* 2006;112(3):500-508. DOI 10.1007/s00122-005-0153-6.
- Sanseverino W., Hermoso A., D'Alessandro R., Vlasova A., Andolfo G., Frusciant L., Lowy E., Roma G., Ercolano M.R. PRGdb 2.0: towards a community-based database model for the analysis of R-genes in plants. *Nucl. Acids Res.* 2013;41(D1):D1167-1171. DOI 10.1093/nar/gks1183.
- Singh R.P., Huerta-Espino J., William H.M. Genetics and breeding for durable resistance to leaf and stripe rusts in wheat. *Turk. J. Agric. For.* 2005;29:121-127.
- Spielmeier W., Mago R., Wellings C., Ayliffe M. *Lr67* and *Lr34* rust resistance genes have much in common—they confer broad spectrum resistance to multiple pathogens in wheat. *BMC Plant Biology.* 2013;13:96. DOI 10.1186/1471-2229-13-96.
- Zdobnov E.M., Lopez R., Apweiler R., Eizold T. The EBI SRS server – recent developments. *Bioinformatics.* 2002;18(2):368-373.