D814–D821 *Nucleic Acids Research, 2010, Vol. 38, Database issue* *Published online 11 November 2009*
doi:10.1093/nar/gkp978

PRGdb: a bioinformatics platform for plant resistance gene analysis

Walter Sanseverino[1], Guglielmo Roma[2], Marco De Simone[3], Luigi Faino[1], Sara Melito[1], Elia Stupka[4,5], Luigi Frusciante[1] and Maria Raffaella Ercolano[1,*]

[1]Department of Soil, Plant, Environmental and Animal Production Sciences, University of Naples 'Federico II', Via Università 100, 80055 Portici, Italy, [2]Center for Genomic Regulation, Dr Aiguader 88, 08003 Barcelona, Spain, [3]CBM S.c.r.l., Area Science Park, 34012 Trieste, Italy, [4]UCL Cancer Institute, Paul O'Gorman Building, University College London, Gower Street, London, WC1E 6BT and [5]The Blizard Institute, Barts and The London School of Medicine and Dentistry, 4 Newark Street, London, E1 2AT, UK

Received August 13, 2009; Revised October 2, 2009; Accepted October 15, 2009

ABSTRACT

PRGdb is a web accessible open-source (http://www.prgdb.org) database that represents the first bioinformatic resource providing a comprehensive overview of resistance genes (R-genes) in plants. PRGdb holds more than 16 000 known and putative R-genes belonging to 192 plant species challenged by 115 different pathogens and linked with useful biological information. The complete database includes a set of 73 manually curated reference R-genes, 6308 putative R-genes collected from NCBI and 10463 computationally predicted putative R-genes. Thanks to a user-friendly interface, data can be examined using different query tools. A home-made prediction pipeline called Disease Resistance Analysis and Gene Orthology (DRAGO), based on reference R-gene sequence data, was developed to search for plant resistance genes in public datasets such as Unigene and Genbank. New putative R-gene classes containing unknown domain combinations were discovered and characterized. The development of the PRG platform represents an important starting point to conduct various experimental tasks. The inferred cross-link between genomic and phenotypic information allows access to a large body of information to find answers to several biological questions. The database structure also permits easy integration with other data types and opens up prospects for future implementations.

INTRODUCTION

In their constant struggle for survival, plants have developed a wide range of defence mechanisms to protect themselves against the attack of pathogens. While some of these resistance strategies rely on simple physical or chemical barriers, more sophisticated biochemical mechanisms based on gene-for-gene interactions between plants and their infectious agents have been reported (1).

Plant disease resistance genes (R-genes) play a key role in recognizing proteins expressed by specific avirulence (Avr) genes of pathogens (2). R-genes originate from a phylogenetically ancient form of immunity that is common to plants and animals. However, the rapid evolution of plant immunity systems has led to enormous gene diversification (3,4). Although little is known about these agriculturally important genes, some fundamental genomic features have already been described. It has been recently shown that proteins encoded by resistance genes display modular domain structures and require several dynamic interactions between specific domains to perform their function. Some of these domains also seem necessary for proper interaction with Avr proteins and in the formation of signalling complexes that activate an innate immune response which arrests the proliferation of the invading pathogen (5).

R-genes can be functionally grouped in five distinct classes based on the presence of specific domains (6,7): the CNL class comprises resistance genes encoding proteins with at least a coiled-coil domain, a nucleotide binding site and a leucine-rich repeat (CC-NB-LRR); the TNL class includes those with a Toll-interleukin receptor-like domain, a nucleotide binding site and a leucine-rich repeat (TIR-NB-LRR); the RLP class, acronym for

*To whom correspondence should be addressed. Tel: +39 81 253 9431; Fax: +39 81 775 7935; Email: ercolano@unina.it

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

receptor-like protein, groups those with a receptor serine–threonine kinase-like domain, and an extracellular leucine-rich repeat (ser/thr-LRR); the RLK class contains those with a kinase domain, and an extracellular leucine-rich repeat (Kin-LRR); the 'Others' class includes all other genes which have been described as conferring resistance through different molecular mechanisms, e.g. *mlo* and *asc-1* (8,9).

Although many R-genes have been isolated to date, the exact reason why proteins exert their resistance function is still unknown. This is also due to the fact that single R-genes have evolved through a range of evolutionary mechanisms. The main models reported are positive, diversifying and balancing selection (10). Different mechanisms of mediation such as intra and interlocus sequence exchanges, insertion of transposon elements and base methylation changes have been shown to be involved in this process (11,12). Furthermore, resistance can be overcome through a co-evolution process between plant and pathogen, which is why advances in knowledge in this research field are required. This complex phenomenon requires an increase of research effort. New findings are expected for this genes family using bioinformatics supports. In fact, the peculiar features of R-genes, above described, make them ideal candidates to benefit of these tools. However, extrapolated specific data from automated database can present great difficulties. Sequence redundancy, annotation errors, irrelevant sequences contamination, can invalidate this task. Thus, a dedicated repository of the R-gene family can be useful to highlight gene diversification process, to discover new resistance capacity and to elucidate mechanisms of interaction between pathogens and their plant hosts.

In this study we present the plant resistance gene database (PRGdb), which is the first comprehensive bioinformatics resource dedicated to known and predicted plant disease resistance genes. This resource aims to provide scientists working in this field of research a comprehensive, up-to-date collection of manually curated R-genes extracted from the literature as well as an unprecedented set of more than 16 000 novel potential R-genes discovered among several plant species using an in-house developed bioinformatics pipeline. To share this resource with the scientific community, we designed and implemented a web interface that is freely accessible at http://www.prgdb.org. Since the PRG database can easily integrate external information, we do invite researchers interested in providing PRG data to contact us.

RESULTS

PRG data and tools

Semi-automated approach towards the creation of a comprehensive R-gene catalogue. To our knowledge, the PRG database represents the first collection of resistance genes publicly available to the scientific community. The complete dataset contains a total of 16 846 sequences obtained through a combination of manually curated and computational approaches, as shown in Figure 1.

First, we used a manual curation approach by searching the primary literature to identify a total of 73 R-genes isolated from 22 plant species interacting with 31 pathogens (Figure 1A). This represents the largest manually curated dataset published so far for plant disease resistance genes. Hence we refer to it from hereon as our 'reference' dataset (Table 1). A list of literature sources for each characterized gene is provided at home page by clicking 'see references'.

These genes have been mostly isolated from the Solanaceae family (33 genes) (7,13), although others have been studied in other plants, such as *Arabidopsis thaliana* (21 R-genes) (14), *Oryza sativa* (rice, four R-genes) (15,16), *Phaseulus vulgaris* (bean, one R-genes) (17), *Glicine max* (soybean, two R-genes) (18), *Zea mais* (mais, two R-genes) (19) and *Hordeum vulgare* (barley, three R-genes) (8,20,21), *Cucumis melo* (melon, two R-genes) (22), *Lactuca sativa* (lettuce, one R-genes) (23), *Beta vulgaris* (beet, one R-genes) (24) *Linum usatissimum* (linum, three R-genes) (25–27). Data related to these genes, such as nucleotide and protein sequences, genomic location, known genetic markers and relevant information about resistance to specific diseases and pathogens, were gathered from the literature and several publicly available resources such as NCBI nucleotide, NCBI taxonomy (28) and SOL network databases (29), and manually inserted into the PRG database through a web-based system. This dataset was used both to retrieve all putative R-gene sequences from NCBI database and to build up an R-gene prediction system.

In this way, a set of 6308 annotated R-genes from 161 plants was obtained automatically using an NCBI query (see Methods section) (Figure 1B). Information such as nucleotide and protein sequences, genomic locations and structural information were automatically retrieved and imported into the PRG database. Since these genes could have been annotated in NCBI as R-genes from other predictive tools, we will refer to them from here on as 'putative R-Genes collected from NCBI'.

Furthermore, we were able to computationally predict novel 'putative' R-genes from the UniGene dataset, using a home-made developed bioinformatic pipeline, Disease Resistance Analysis and Gene Orthology, (DRAGO, see 'Methods' section) (Figure 1C). A total of 604 981 non-redundant Unigene transcript sequences expressed in 33 different plants were translated into 488 250 potential protein sequences. Finally, a total of 10 463 sequences were identified as 'putative R-Genes predicted from NCBI UniGene' based on their sequence similarity and protein domain composition and imported into the PRG database.

These three distinct approaches yielded a total of 16 844 protein sequences annotated in our database as potential plant resistance genes. Of 194 plant species analyzed, 172 contained sequences related to resistance genes. A complete list of retrieved plants is available on the PRG web site under the 'plant search' section. In this section all putative resistance genes are divided by plant species to allow specific searches to be conducted.

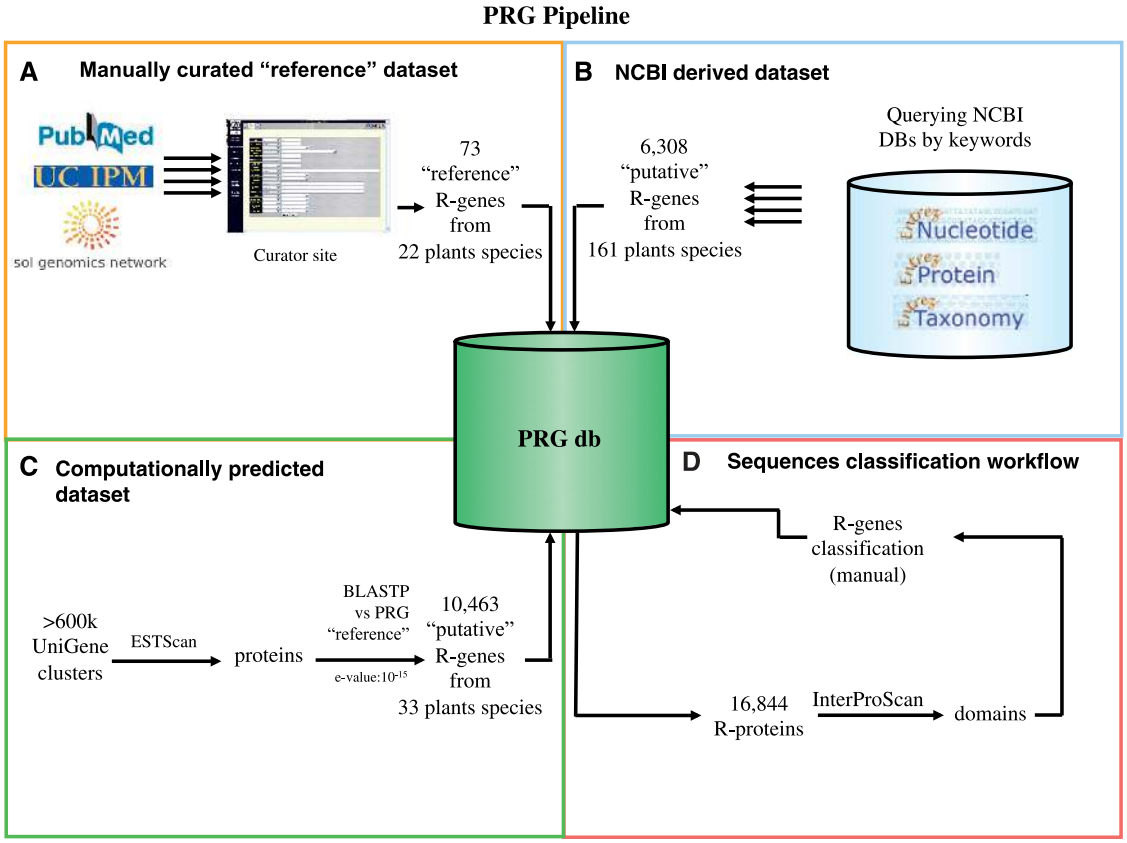

Figure 1. A schematic view of the PRG database showing the origin of dataset used and the sequences characterization. (**A**) The manually curated dataset that contains 73 literature cited R-genes from 22 different plants. (**B**) The NCBI dataset containing 6308 sequences related to reference R-genes retrieved by the NCBI database. (**C**) The computationally predicted dataset using the DRAGO pipeline containing 10 463 putative R-genes. (**D**) Workflow of conserved domain analysis and sequence classification.

PRG web interface

The PRG data is stored in a MySQL database and is freely accessible through a web interface at the address: http://www.prgdb.org. The PRG web site was designed to provide plant researchers with user-friendly tools to retrieve relevant information in our complete R-gene catalogue. Researchers interested only in the manually curated 'reference' dataset can search it by a combination of controlled key terms provided, such as reference R-gene name, Avr gene name, plant species, pathogen species and disease name.

The complete dataset of 16 844 R-genes comprising all the three different categories described in this article (such as 'reference' R-genes, putative R-Genes collected from NCBI, putative R-Genes predicted from NCBI UniGene) can be accessed through several entry points:

(i) Searching by single or combined query fields provided in the homepage, such as sequence category, one or more resistance domain types, plant species and pathogen species;
(ii) Searching by sequence comparison against a local database of R-gene sequences through the BLAST algorithm; both nucleotide and amino acid sequences are allowed;
(iii) Choosing a plant species by clicking on the image provided in the 'plant search' section;
(iv) Choosing a pathogen species by clicking on the image provided in the 'pathogen search' section.

Each of these queries generates a list of resistance genes that meet the search criteria. By clicking on a gene name, information regarding the gene of interest is visualized in a specific page including gene name, genome locations, known genetics markers, external links to several public resources and to Pubmed, transcript sequence, protein sequence, domains, as well as curated information related to the diseases and the plant–pathogen interactions. Moreover, a picture showing the gene structure is generated dynamically using BioPerl's Bio-Graphics module (Figure 2).

Mining PRG data

In order to further verify whether the sequences retrieved using the approaches described above were plausible candidates to exert the resistance function, we inspected them for the presence of specific R-protein signatures using InterProScan and the InterPro database. Based on these results, we proceeded to assign each sequence to one of the four already known R-gene classes. A schematic view of the single domains predicted and of four major

Table 1. Plant functional resistance genes identified to date in the plant kingdom with indication of donor species, related disease and pathogen

Gene Name	Donor Species	Disease	Pathogen
Asc1	*Solanum lycopersicum*	Alternaria stem canker	*Alternaria alternata*
At1	*Cucumis melo*	Cucurbit downy mildew	*Pseudoperonospora cubensis*
At2	*Cucumis melo*	Cucurbit downy mildew	*Pseudoperonospora cubensis*
Bs2	*Capsicum chacoense*	Bacterial spot	*Xanthomonas campestris pv. vesicatoria str. 85-10*
Bs3	*Capsicum annuum*	Bacterial spot	*Xanthomonas campestris pv. vesicatoria str. 85-10*
Bs3-E	*Capsicum annuum*	Bacterial spot	*Xanthomonas campestris pv. vesicatoria str. 85-10*
Bs4	*Solanum lycopersicum*	Bacterial spot	*Xanthomonas campestris*
Cf2	*Solanum pimpinellifolium*	Leaf mould	*Passalora fulva*
Cf4	*Solanum habrochaites*	Leaf mould	*Passalora fulva*
Cf4A	*Solanum habrochaites*	Leaf mould	*Passalora fulva*
Cf5	*Solanum lycopersicum var. cerasiforme*	Leaf mould	*Passalora fulva*
Cf9	*Solanum pimpinellifolium*	Leaf mould	*Passalora fulva*
Cf9B	*Solanum pimpinellifolium*	Leaf mould	*Passalora fulva*
Dm-3	*Lactica sativa*	Downy mildew	*Bremia lactucae*
EFR	*Arabidopsis thaliana*	Eliciting bacteria	*Bacteria with flagellum*
ER-Erecta	*Arabidopsis thaliana*	Bacterial wilt (Arabidopsis)	*Ralstonia solanacearum*
FLS2	*Arabidopsis thaliana*	Eliciting bacteria	*Bacteria with flagellum*
Gpa2	*Solanum tuberosum*	Yellow potato cyst nematode	*Globodera*
Gro1.4	*Solanum tuberosum*	Late blight potato	*Phytophthora infestans*
Hero	*Solanum lycopersicum*	Yellow potato cyst nematode	*Globodera*
Hm1	*Zea mays*	Leaf spot	*Bipolaris zeicola*
Hm2	*Zea mays*	Leaf spot	*Bipolaris zeicola*
HRT	*Arabidopsis thaliana*	Turnip crinkle virus	*Turnip crinkle virus*
Hs1	*Beta procumbens*	Beet cyst nematode	*Heterodera schachtii*
I2	*Solanum lycopersicum*	Fusarium wilt	*Fusarium oxysporum*
L6	*Linum usitatissimum*	Flax rust	*Melampsora lini*
LeEIX1	*Solanum lycopersicum*	Eliciting fungus	*Fungal ethylene-inducing xylanase*
LeEIX2	*Solanum lycopersicum*	Eliciting fungus	*Fungal ethylene-inducing xylanase*
M	*Linum usitatissimum*	Flax rust	*Melampsora lini*
Mi1.2	*Solanum lycopersicum*	Root-knot nematode	*Meloidogyne, Paratrichodorus minor*
MLA10	*Hordeum vulgare*	Powdery mildew (barley)	*Blumeria graminis*
Mlo	*Hordeum vulgare*	Powdery mildew (barley)	*Blumeria graminis*
N	*Nicotiana glutinosa*	Tobacco mosaic Virus	*Tobacco mosaic virus*
P2	*Linum usitatissimum*	Flax rust	*Melampsora lini*
PEPR1	*Arabidopsis thaliana*	Damping off	*Pythium*
PGIP	*Phaseolus vulgaris*	Eliciting fungus	*Fungus producing polygalacturonases*
Pi33	*Oryza sativa*	Rice blast disease	*Magnaporthe grisea*
Pi-ta	*Oryza sativa Japonica Group*	Rice blast disease	*Magnaporthe grisea*
Prf	*Solanum pimpinellifolium*	Bacterial speck	*Pseudomonas syringae*
Pto	*Solanum pimpinellifolium*	Bacterial speck	*Pseudomonas syringae*
R1	*Solanum demissum*	Late blight tomato	*Phytophthora infestans*
R3a	*Solanum tuberosum*	Late blight tomato	*Phytophthora infestans*
RCY1	*Arabidopsis thaliana*	Cucumber mosaic virus	*Cucumber mosaic virus*
RFO1	*Arabidopsis thaliana*	Fusarium wilt	*Fusarium oxysporum*
Rmd-c	*Glycine max*	Powdery mildew	*Microsphaera sparsa*
RPG1	*Hordeum vulgare*	Stem rust	*Puccinia Graminis*
Rpi-blb1	*Solanum bulbocastanum*	Late blight tomato	*Phytophthora infestans*
Rpi-blb2	*Solanum bulbocastanum*	Late blight tomato	*Phytophthora infestans*
RPM1	*Arabidopsis thaliana*	Bacterial blight	*Pseudomonas syringae*
RPP13nd	*Arabidopsis thaliana*	Downy mildew	*Hyaloperonospora parasitica*
RPP4	*Arabidopsis thaliana*	Downy mildew	*Peronospora parasitica*
RPP5	*Arabidopsis thaliana*	Downy mildew	*Hyaloperonospora parasitica*
RPP8	*Arabidopsis thaliana*	Downy mildew	*Hyaloperonospora parasitica*
Rps1-k-1	*Glycine max*	Phytophthora root	*Phytophthora sojae*
Rps1-k-2	*Glycine max*	Phytophthora root	*Phytophthora sojae*
Rps2	*Arabidopsis thaliana*	Bacterial blight	*Pseudomonas syringae*
Rps4	*Arabidopsis thaliana*	Bacterial blight	*Pseudomonas syringae*
RPS5	*Arabidopsis thaliana*	Bacterial blight	*Pseudomonas syringae*
RPW8.1	*Arabidopsis thaliana*	Powdery mildew	*Golovinomyces cichoracearum*
RPW8.2	*Arabidopsis thaliana*	Powdery mildew	*Golovinomyces cichoracearum*
RRS1	*Arabidopsis thaliana*	Bacterial wilt	*Ralstonia solanacearum*
RTM1	*Arabidopsis thaliana*	Synergistic disease syndromes	*Tobacco etch virus*
RTM2	*Arabidopsis thaliana*	Synergistic disease syndromes	*Tobacco etch virus*
Rx	*Solanum tuberosum*	Latent mosaic	*Potato virus X*
Rx2	*Solanum acaule*	Latent mosaic	*Potato virus X*
RY1	*Solanum tuberosum subsp andigena*	Potato virus Y	*Potato virus Y*
Sw5	*Solanum lycopersicum*	Tomato spotted wilt	*Tomato spotted wilt virus*
Tm2	*Solanum lycopersicum*	Tobacco mosaic virus	*Tobacco mosaic virus*
Tm2a	*Solanum lycopersicum*	Tobacco mosaic virus	*Tobacco mosaic virus*
Ve1	*Solanum lycopersicum*	Verticillium wilt potato	*Verticillium*
Ve2	*Solanum lycopersicum*	Verticillium wilt potato	*Verticillium*
Xa1	*Oryza sativa*	Bacterial blight	*Xanthomonas oryzae*
Xa21	*Oryza sativa Indica group*	Bacterial blight	*Xanthomonas oryzae*

Figure 2. A PRGdb web page reporting an R-gene description. The following information is displayed: gene name; CDS, RNA, protein sequences and domains position; Genbank ID; original resistant species (donor organism); related molecular markers; literature; disease description, related pathogen and corresponding avirulence gene. Words in green and red represent hypertext links.

classes identified is shown in Figure 3A and B. Of all the 16 885 sequences, the following were assigned to known classes: 1150 to CNL, 341 to TNL, 1930 to RLP and 2236 to RLK, while other proteins fall in new putative classes.

Mining the protein domain data highlighted the fact that quite a substantial number of genes do not fall within existing classes, as some of them present new domain combinations which had not yet been described

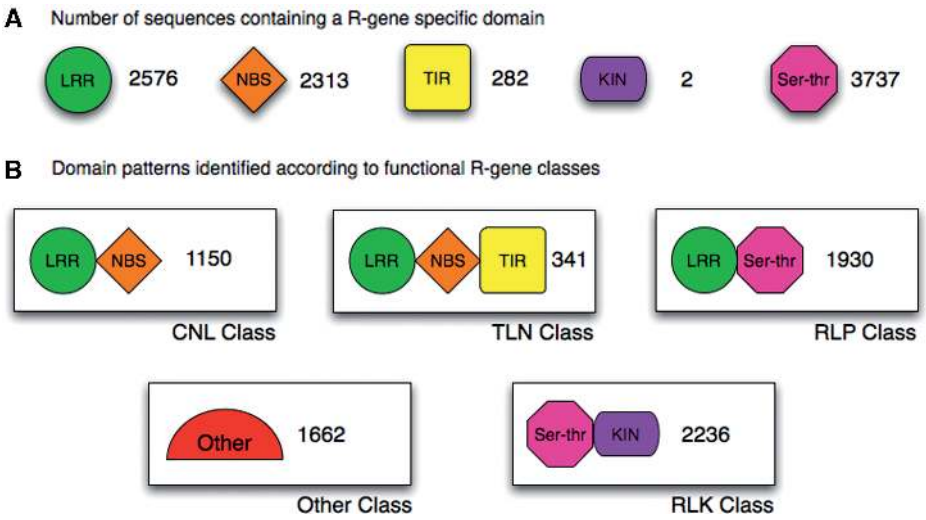

Figure 3. DRAGO predicted sequences divided by domains and identified by class. (**A**) Number of sequences containing an R-gene specific domain; LRR, leucine-rich repeat; NBS, nucleotide binding site; TIR, Toll interleukine receptor-like; KIN, kinase; Ser–Thr, serine–threonine. (**B**) Domain patterns identified according to functional R-gene classes.

in previous studies. A further class called ''other'' had to be included to represent sequences with specific roles in plant defence mechanisms: sequences in this class are not classifiable as they do not contain any specific R-protein domain. The PRG database allowed us to search new combinations of resistance gene domains, thus discovering new putative R-gene classes. Figure 4A shows a statistical Venn in which are showed all R-gene classes according with new and known conserved domain combination. Moreover, Figure 4B shows three examples of hitherto undescribed protein classes: the first class contains four Arabidopsis sequences (At.66955, F10C21.20, T1E4.9, WRKY19) with typical CNL class domains as well as a kinase domain. The second consists of 22 sequences with typical CNL class domains and a Ser–Thr domain. The third class contains two Poplar Unigene PHT16062 and the Arabidopsis RPP1 gene structured like a typical TNL class with the addition of a Ser–Thr domain.

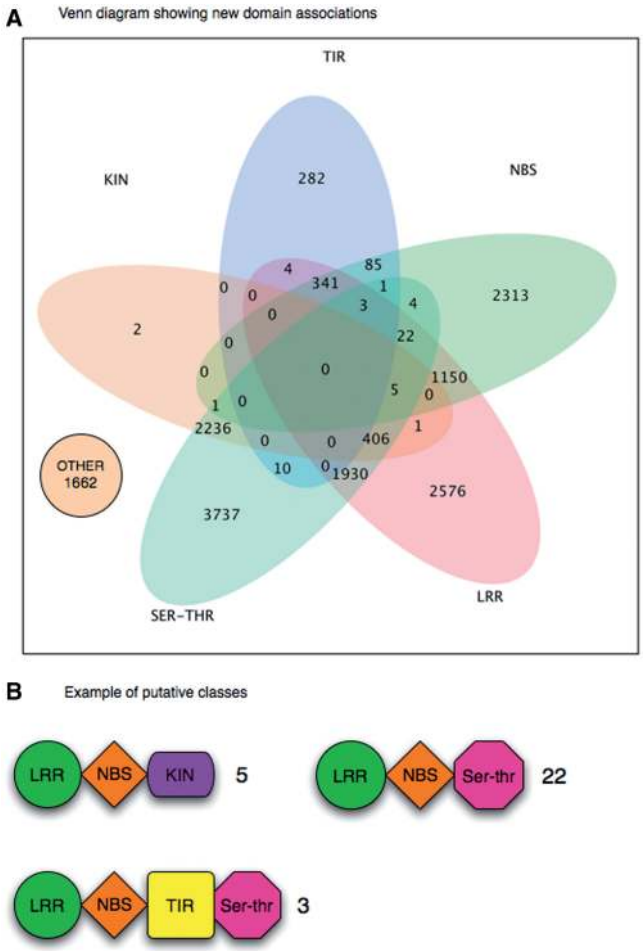

Figure 4. (**A**) A Venn diagram showing all possible combinations among domain classes produced by DRAGO pipeline. Each intersection represents a new or know domains association. Proteins numbers falling in each class are reported. (**B**) Examples of three unknown putative classes containing new domain combinations.

DATA SOURCES AND ANALYSIS PIPELINE

PRG site architecture and implementation

PRG data are stored within a relational database management system, MySQL (http://www.mysql.com). Our bioinformatics software is written in Perl and uses the Bioperl toolkit (30). The website was developed using the PHP language (http://www.php.net) and the Apache web server (http://www.apache.org). The annotation pipeline runs on a Linux cluster running the Gentoo Linux distribution (http://www.gentoo.org) and the PBS scheduling system (http://www.openpbs.org).

Automatic download of plant resistance genes

We developed a Perl script to automatically download known R-genes from NCBI using the following query: *plants AND ('disease resistance gene' OR 'disease resistance protein') NOT bacteria NOT virus*. The data

obtained were parsed and used to populate the PRG database.

Disease Resistance Analysis and Gene Orthology pipeline

Unigene sequences from 33 plant species were translated into potential protein sequences using the ESTScan program, version 3.0.2 (31), with default parameters and coupled with the *Arabidopsis thaliana* codon usage/log odds probability matrices. The resulting translations were subsequently checked for sequence homology with at least one resistance protein contained in the 'reference' dataset using the BLAST algorithm with a stringent *e*-value cut-off of 1×10^{-15}.

Domain analysis of selected sequences was performed using InterProScan version 3.0.2 (32), with standard options and last InterPro database release. Genes were divided into five already known classes according to their domains and gene structure. The resulting set of sequences was loaded into the PRG database.

The goodness of Disease Resistance Analysis and Gene Orthology (DRAGO) predictor was evaluated running the pipeline on the hand-curated dataset. The comparisons showed a perfect match between reference genes manual classification and DRAGO prediction.

DISCUSSION

Despite a large amount of experimental data produced in recent years (ESTs, whole genome sequences, gene expression data), progress in understanding the function of R-genes has been slow for several reasons: the lack of a reference set of sequences to be used as a model for R-gene studies; the genomic feature of R-genes that usually cluster in genomic regions with a high number of homologues and pseudo genes; the difficulties in performing plant-pathogen interaction studies (33).

The main aim of PRGdb is to provide tools to support research in this field. We have developed an exhaustive plant community database, providing data for extensive studies. As of July 2009 the database contained 16 844 annotated sequences, comprising 73 reference genes and several thousand related sequences. The data quality is very high and is guaranteed by combining a large-scale automated approach and manual annotation. In particular, our in-depth review of the literature was fundamental to update and organize the current R-gene panorama and create a robust basis to perform *in silico* analysis. Rapid scientific progress makes information updates difficult and R-gene reviews can lack a number of cloned R-genes (7,34). The development of a PRG platform represents an important starting point to conduct various experimental tasks. The inferred cross-link between genomic and phenotypic information allows the creation of a resource to perform multidisciplinary studies merging queries between disparate resources. Moreover, several questions can be addressed by comparative analysis of gene patterns in closely related organisms.

Our prediction pipeline called DRAGO was built to offer end-users a flexible user-friendly tool to explore known and novel disease resistance genes. We were able to assign to know classes ~40% of retrieved sequences. Large genomes annotation display that a high number of genes with coding domains characteristic of plant resistance proteins is not yet characterized (14,35). Our prediction tool allowed us to observe unknown combinations of resistance domains, thus discovering new putative R-gene classes.

Plant–pathogen interaction of R-genes works not only by single gene-for-gene interaction but also by activating proteins, disrupting or modifying the stable conformation of the R-gene receptor surface. The complex signal transduction system is often driven by different protein classes (36). For these reasons our pipeline fished all possible sequences involved in the disease resistance process according to this hypothesis.

In conclusion, a database and a public web interface regarding an important class of genes across hundreds of species was developed on the basis of a novel, specific prediction pipeline. Information about the gene structure, domains and organization of R-genes was obtained and made available through a user-friendly interface. Inference of gene function is a long arduous task, a process which we aim to simplify by starting from a strong knowledge base using the PRG platform. It is hoped the PRG database will provide a new perspective on the analysis of R-genes by tapping into a large, unbiased but curator driven, survey of these proteins.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Gianpiero Lago for system support and Dr Vincenza Maselli for useful suggestions for improving PRGdb; Mark Walters for editing the manuscript Contribution no. 202 from the DISSPAPA.

FUNDING

Ministry of Education, University and Research (GenoPOM Project); Ministry of Agricultural, Food and Forestry Policies (Agronanotech Project). Funding for open access charges: Department of Soil, Plant, Environment and Animal Production Sciences, University of Naples 'Federico II', via Universitá 100, 80055, Portici, Italy.

Conflict of interest statement. None declared.

REFERENCES

1. Flor,H.H. (1971) Current status of the gene-for-gene concept. *Annual Rev. Phytopathol.*, **9**, 275–296.
2. Ellis,J., Dodds,P. and Pryor,T. (2000) The generation of plant disease resistance gene specificities. *Trends Plant Sci.*, **5**, 373–379.
3. Chisholm,S.T., Coaker,G., Day,B. and Staskawicz,B.J. (2006) Host-microbe interactions: shaping the evolution of the plant immune response. *Cell*, **124**, 803–814.
4. Means,T.K., Golenbock,D.T. and Fenton,M.J. (2000) The biology of Toll-like receptors. *Cytokine Growth Factor Rev.*, **11**, 219–232.
5. Mackey,D., Holt,B.F., Wiig,A. and Dangl,J.L. (2002) RIN4 interacts with Pseudomonas syringae type III effector molecules and

is required for RPM1-mediated resistance in Arabidopsis. *Cell*, **108**, 743–754.
6. Bent,A.F. (1996) Plant disease resistance genes: function meets structure. *Plant Cell*, **8**, 1757–1771.
7. Van Ooijen,G., van den Burg,H.A., Cornelissen,B.J. and Takken,F.L. (2007) Structure and function of resistance proteins in solanaceous plants. *Annual Rev. Phytopathol.*, **45**, 43–72.
8. Buschges,R., Hollricher,K., Panstruga,R., Simons,G., Wolter,M., Frijters,A., van Daelen,R., van der Lee,T., Diergaarde,P., Groenendijk,J. *et al.* (1997) The barley Mlo gene: a novel control element of plant pathogen resistance. *Cell*, **88**, 695–705.
9. Brandwagt,B.F., Mesbah,L.A., Takken,F.L., Laurent,P.L., Kneppers,T.J., Hille,J. and Nijkamp,H.J. (2000) A longevity assurance gene homolog of tomato mediates resistance to Alternaria alternata f. sp. lycopersici toxins and fumonisin B1. *Proc. Natl Acad. Sci. USA*, **97**, 4961–4966.
10. Tiffin,P. and Moeller,D.A. (2006) Molecular evolution of plant immune system genes. *Trends Genet.*, **22**, 662–670.
11. Kuang,H., Woo,S.S., Meyers,B.C., Nevo,E. and Michelmore,R.W. (2004) Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell*, **16**, 2870–2894.
12. McDowell,S.A.S. (2006) Recent insights into R gene evolution. *Mol. Plant Pathol.*, **7**, 437–448.
13. Romer,P., Hahn,S., Jordan,T., Strauss,T., Bonas,U. and Lahaye,T. (2007) Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science*, **318**, 645–648.
14. Meyers,B.C., Kozik,A., Griego,A., Kuang,H. and Michelmore,R.W. (2003) Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. *Plant Cell*, **15**, 809–834.
15. Song,W.Y., Wang,G.L., Chen,L.L., Kim,H.S., Pi,L.Y., Holsten,T., Gardner,J., Wang,B., Zhai,W.X., Zhu,L.H. *et al.* (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science*, **270**, 1804–1806.
16. Liu,J., Liu,X., Dai,L. and Wang,G. (2007) Recent progress in elucidating the structure, function and evolution of disease resistance genes in plants. *J. Genet. Genomics*, **34**, 765–776.
17. Toubart,P., Desiderio,A., Salvi,G., Cervone,F., Daroda,L. and De Lorenzo,G. (1992) Cloning and characterization of the gene encoding the endopolygalacturonase-inhibiting protein (PGIP) of *Phaseolus vulgaris L. Plant J.*, **2**, 367–373.
18. Gao,H., Narayanan,N.N., Ellison,L. and Bhattacharyya,M.K. (2005) Two classes of highly similar coiled coil-nucleotide binding-leucine rich repeat genes isolated from the Rps1-k locus encode Phytophthora resistance in soybean. *Mol. Plant Microbe Interact.*, **18**, 1035–1045.
19. Zhang,L., Peek,A.S., Dunams,D. and Gaut,B.S. (2002) Population genetics of duplicated disease-defense genes, hm1 and hm2, in maize (Zea mays ssp. mays L.) and its wild ancestor (Zea mays ssp. parviglumis). *Genetics*, **162**, 851–860.
20. Halterman,D.A. and Wise,R.P. (2004) A single-amino acid substitution in the sixth leucine-rich repeat of barley MLA6 and MLA13 alleviates dependence on RAR1 for disease resistance signaling. *Plant J.*, **38**, 215–226.
21. Brueggeman,R., Rostoks,N., Kudrna,D., Kilian,A., Han,F., Chen,J., Druka,A., Steffenson,B. and Kleinhofs,A. (2002) The barley stem rust-resistance gene Rpg1 is a novel disease-resistance gene with homology to receptor kinases. *Proc. Natl Acad. Sci. USA*, **99**, 9328–9333.
22. Taler,D., Galperin,M., Benjamin.I, Cohen,Y. and Kenigsbuch,D. (2004) Plant eR genes that encode photorespiratory enzymes confer resistance against disease. *Plant Cell*, **16**, 172–184.
23. Shen,K.A., Chin,D.B., Arroyo-Garcia,R., Ochoa,O.E., Lavelle,D.O., Wroblewski,T., Meyers,B.C. and Michelmore,R.W. (2002) Dm3 is one member of a large constitutively expressed family of nucleotide binding site-leucine-rich repeat encoding genes. *Mol. Plant Microbe Interact.*, **15**, 251–261.
24. Cai,D., Kleine,M., Kifle,S., Harloff,H.J., Sandal,N.N., Marcker,K.A., Klein-Lankhorst,R.M., Salentijn,E.M., Lange,W., Stiekema,W.J. *et al.* (1997) Positional cloning of a gene for nematode resistance in sugar beet. *Science*, **275**, 832–834.
25. Lawrence,G.J., Finnegan,E.J., Ayliffe,M.A. and Ellis,J.G. (1995) The L6 gene for flax rust resistance is related to the Arabidopsis bacterial resistance gene RPS2 and the tobacco viral resistance gene N. *Plant Cell*, **7**, 1195–1206.
26. Anderson,P.A., Lawrence,G.J., Morrish,B.C., Ayliffe,M.A., Finnegan,E.J. and Ellis,J.G. (1997) Inactivation of the flax rust resistance gene M associated with loss of a repeated unit within the leucine-rich repeat coding region. *Plant Cell*, **9**, 641–651.
27. Dodds,P.N., Lawrence,G.J. and Ellis,J.G. (2001) Six amino acid changes confined to the leucine-rich repeat beta-strand/beta-turn motif determine the difference between the P and P2 rust resistance specificities in flax. *Plant Cell*, **13**, 163–178.
28. Sayers,E.W., Barrett,T., BensonD,A., Bryant,S.H., Canese,K., Chetvernin.V, Church,D.M., Di Cuccio,M., Edgar,R., Federhen,S. *et al.* (2009) Information resources at the National Center for Biotechnology Information. *Nucleic Acids Res.*, **37**, D5–D15.
29. Mueller,L.A., Solow,T.H., Taylor,N., Skwarecki,B., Buels,R., Binns,J., Lin,C., Wright,M.H., Ahrens,R., Wang,Y. *et al.* (2005) The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. *Plant Physiol.*, **138**, 1310–1317.
30. Stajich,J.E., Block,D., Boulez,K., Brenner,S.E., Chervitz,S.A., Dagdigian,C., Fuellen,G., Gilbert,J.G., Korf,I., Lapp,H. *et al.* (2002) The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.*, **12**, 1611–1618.
31. Iseli,C., Jongeneel,C.V. and Bucher,P. (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **1**, 138–148.
32. Hunter,S., Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Binns,D., Bork,P., Das,U., Daugherty,L., Duquenne,L. *et al.* (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res.*, **37**, D211–D215.
33. Friedman,A.R. and Baker,B.J. (2007) The evolution of resistance genes in multi-protein plant resistance systems. *Curr. Opin. Genet. Dev.*, **17**, 493–499.
34. Martin,G.B., Bogdanove,A.J. and Sessa,G. (2003) Understanding the functions of plant disease resistance proteins. *Annual Rev. Plant Biol.*, **54**, 23–61.
35. Goff,S.A., Ricke,D., Lan,T.H., Presting,G., Wang,R., Dunn,M., Glazebrook,J., Sessions,A., Oeller,P., Varma,H. *et al.* (2002) A draft sequence of the rice genome (Oryzasativa L. ssp. japonica). *Science*, **296**, 92–100.
36. Dangl,J.L. and Jones,J.D. (2001) Plant pathogens and integrated defence responses to infection. *Nature*, **411**, 826–833.