Highlights in BioScience
ISSN: 2682-4043

Review Article **Open Access**

Plant Resistance Gene, SNP and Genome Annotation an Overview

Nor S. Zaimah

Sciences Research Institute, Selangor, Malaysia

Abstract

Plants have developed systems of effective and passive protection to safeguard themselves from pathogens. Active processes include kinds of immune responses that are adaptive and intrinsic. Adaptive immunity is focused on reaction type RNAi and works primarily against viruses. Biological immunity is more general and allows the plant to protect itself against a wide range of pathogens through bacterial and model resistance receptors (PPRs) and forms of resistance (R proteins). PPRs identify molecular models associated with microbes or pathogen that are preserved in a specific category of pathogens. With such big marker amounts, it has become feasible to scan the entire genome for interactions of individual markers with particular quantitatively hereditary traits called whole-genome studies (WGS), genome-wide association trials (GWAS), or association genetics at exceptionally elevated marker densities. Several SNP recognition methods have been used in a specified plant to identify significant amounts of SNPs. These include: identification of SNP based on EST feature information; identification of SNP from sequenced genomes; re-sequencing of amplicons; identification of SNP using sequencing techniques of the next generation and identification of SNP based on cluster tests. Pathologic recognizes parts of Genbank completely annotated genome and MetaCyc has been used as a database for the reference pathway. In contrast to sequence similarity information used in other systems, Pathologic uses Genbank annotation information and the EC assignment as evidence of the presence of each pathway in the genome of interest reference database. When the matching task is finished, pathologic will infer a number of reactions expected to take place in the destination genome and will determine which one of those pathways in the target genome are likely to exist.

Keywords: Resistance proteins, genome-wide association, SNP, biological pathways, genome annotation .

Citation: Zaimah N. S. (2019) Plant resistance gene, SNP and genome annotation an overview. Highlights in BioScience, Volume 2.Article ID 20192, dio:10.36462/ H.BioSci.20192

Received: June 19, 2019

Accepted: July 13, 2019

Published: August 12, 2019

Data Availability Statement: All relevant data are within the paper

Funding: The authors have no support or funding to report.

Introduction

Proteins of resistance, in turn, identify distinctive elements of avirulence (Avr) which are not conserved between many pathogens. Induced-signaling resistance protein contributes to the manufacturing of reactive oxygen species and the induction of a particular form of programmed cell death, called the hypersensitive response, which damages the impacted cells (1). Also called gene-to-gene resistance is the resistance of protein-mediated innate immunity, as each R gene reacts to a particular pathogenic Avr gene (1). As a result, a big variety of R genes per plant genome are anticipated to be prepared to confer resistance to a wide range of pathogens. R genes are also under selection diversification to maintain pace with the fast pathogen evolution. Although distinct R genes react to very unique pathogens, they share multiple regions (domains) that have been conserved. R proteins can be split into four subclasses on the basis of these domains.

SNP markers have acquired a great deal of concern in the science and breeding society (2) over the past two centuries. They happen in nearly infinite amounts as variations between individual nucleotides and each SNP is a possibly helpful marker in single copy DNA. In the study of human genome, the potential of SNP markers is obviously illustrated. Several million SNP markers have been recognized (3) and techniques have been created concurrently (mainly up to 1 million) to evaluate big amounts of SNP markers based on huge study attempts and the complete description of the human genome.

As more genome sequences became accessible, comparative assessment of various genomes became a very significant research method. Compared to various genomes of selection, biological pathway assessment is conducted using a range of computational techniques and databases (4,5). There are a range of enzymes and their substrates and products involved in a biological pathway. Pathways also communicate with each other. Thus, comparison pathway assessment is quite complex and can hardly be achieved without well-designed software systems for pathway assessment (4). We review the concept of plant resistance genes, plant SNP markers and plant Genome annotation.

Resistance genes

Most R proteins comprise a key nucleotide binding site (NBS) that functions as a molecular switch to regulate the protein's activity status, and a C-terminal, leucine-rich repeat region (LRR) needed to recognize Avr factor. Thus, the ranking of R proteins is focused on N-terminal domain variety (1). Drosophila Toll and human interleukin receptors categorized as TIR-NB-LRR proteins are homologous to NBS-LRR form R proteins with N-terminals. Non-TIR NBS-LRR proteins are known to as CC-NBS-LRR proteins since some non-TIR proteins in their N terminus form a coiled coil (CC) domain (6).

Furthermore, there are two categories of R proteins in their N terminus that contain an extracellular LRR. One of these groups, called kinases (RLKs) receptor, includes a kinase domain of cytoplasmic protein (7). This cytoplasmic protein kinase domain is also lacking in receptors such as proteins (RLPs). Since R genes from distinct plant species combine conserved domains, they could be used to monitor plant genomes for R genes and putative R genes (e.g., analogs of resistance genes, RGAs) and to generate molecular markers (7).

Meyers et al. (8) investigated a genome-wide assessment of 149 NBS-LRR processing genes in *Arabidopsis* and verified either 55 CC-NBS-LRR (CNL) or 94 TIR-NBS-LRR (TNL) proteins in two significant groups. Eight significant motifs varied in their deviation within and between CNL and TNL clusters, and in the model found in particular for plant R protein homologues. Introns may be more prevalent in cereals in the NBS region than in dicots. Only participants of the Arabidopsis Rpp8/Hrt gene class contain introns in the NBS domain in 20 named dicot NBS-LRR genes. Nevertheless, in their NBS region, Pib (9), Pi-ta (10) and Mla1 (11), three distinguished plant resistance genes have introns. Repeats rich in leucine (LRRs) consist of redundant incomplete amino acid sections that fold into solvent-exposed β-cell β-loop constructions and this domain is believed to be engaged in ligand binding and disease identification (12).

Alternating conservation motifs and hyper variance marked LRR areas. The variation is highest for codons (x) situated in the LRR consensus xxLxLxx around the two preserved aliphatic amino acids, and the number of LRR repeats ranges among family members. There were approximately 65 amino acids in TNL proteins between the NBS and LRR domains in the genome-wide assessment of *Arabidopsis* LRR domains (12).

In *Arabidopsis*, in CNL proteins, LRRs represent about half of the C-terminal region in the TNL proteins and almost the entire C-terminal region. The median domain of TNL LRR and CNL LRR comprised an average of 14 LRRs with ~10 separate MEME patterns spanning as many as 350 amino acids as possible. In the rice proteins, a total of 25 distinct LRR motifs have been recognized. In any one gene, the amount of LRR clusters varied from 3 to 40 (13) . There are more than 150 genes of NBS-LRR in the *Arabidopsis thaliana* DNA. A sum of 166 NBS-LRR sequences were mentioned in (13), along with 33 truncated sequences. These NBS-LRR genes happen in their chromosome layout as 51 singletons and 40 clusters. Meyers *et al*. (2003) identified more NBS-LRR mutations by using comprehensive manual re-annotation of the same species genomic structure. They identified 149 NBS-LRR genes as well as 58 truncated genes; spread as 40 singletons and 43 clusters were the 149 non-truncated genes.

In *Arabidopsis thaliana*, TIR-NBS-LRR genes family was detected out CC-NBS-LRR genes by approximately two to one, showing either a latest amplification of the former group or reduction of the latter gene group (13–15). NBS-LRR gene loci *Arabidopsis* are often not dispersed uniformly in the chromosomal genome. There are super groups on chromosomes 1 and 5 whereas chromosomes 2 and 3 were comparatively lacking in the genes of NBS-LRR (13,14). The clusters are believed to be engaged in both R-gene diversity generation and conservation.

Meyers et al. (14) noted that the NBS gene chromosomal allocation is substantially non-random in maize: chromosome 11 includes about one-quarter of the NBS genes. Five hundred and thirty-five NBS encoding genes have been recognized in rice, including 480 non-TIR NBS-LRR genes. TIR-NBS-LRR genes have been not detected in the rice genomic DNA. In 44 gene clusters, two hundred and sixty-three genes (51 percent) occurred. In the "clustered" allocation category, there are 40 pairs and 17 triplets, 394 genes. A total of 125 singletons of NBS have been distributed across all chromosomes. In the plant genome, the proportion of singletons to the complete amount of NBS chromosomes (24.1%) was comparable to that of *Arabidopsis* (26.8%).

Baumgarten et al. (16) proposed that most of the genomic dispersion of NBS-LRR genes was produced by duplication and translocation of whole chromosome sections (segmental duplication), rather than by small-scale ectopic duplication. Most of the dramatic variation happens within local chromosomal areas of the NBS-LRR gene copy number. Zhou et al. (17) revealed that 51% of NBS genes in rice originated in 44 gene clusters, where a cluster is an area with four and sometimes more genes within 200 kb or less. Many surveys of NBS-LRR sequences or analogs of resistance genes have shown that R genes or NBS-LRR sequences are also structured in big clusters in other plant species.

High-throughput genomic studies and plant genome sequences accessible in international databases provide unprecedented possibilities to recognize novel R-genes, investigate their role and method of diversification, find new genes for resistance, and eventually elucidate their interaction processes between pathogens and their crop hosts (18). In 2009, Sanseverino et al. (18) launched the Plant Disease Resistance Gene database (PRGdb), a comprehensive repository of R-genes across hundreds of plant species, with the intention of facilitating research on this agriculturally important gene family. A total of 16,844 gene records were included in PRGdb version 1.0. Of these 73 R-genes (e.g. the ' reference ' data collection) were recognized and manually curated, 6,308 were putative R-genes recovered from NCBI Genbank, and 10,463 were putative R-genes computationally anticipated from NCBI UniGene information.

Many plant genome-sequencing initiatives have developed quickly over the past few years. For example, potato (19), tomato (20), and melon (21) genomes have been finished, providing an chance to find extra R-genes (22). Sanseverino et al. (22) presented an overview of the PRGdb database of crop resistance genes. This database has been extended to include more than 6-fold valuable biological data from 233 plant species on a total of 104,459 R-genes. Of these papers, 112 are defined in the literature as manual-curated R-genes to confer resistance to 122 distinct pathogens. All other genes were predicted gene.

SNP identification based on EST sequence data

Large amounts of expressed sequence tags (ESTs) were produced for many plant species (23). The number of accessible ESTs in the NCBI EST database ranges from less than 10,000 to over one million ESTs in significant crop plants and 1.5 million ESTs for the Arabidopsis model plant. These ESTs were obtained in many cases within the framework of international efforts and were accumulated from a limited set of different lines and could therefore represent as a source for SNP identification. In some instances, ESTs were specifically produced for SNP detection from distinct lines as in *Arabidopsis thaliana* (24) and in other instances, using bioinformatics assessment techniques (25), ESTs from heterozygous extremely polymorphic samples were used to identify SNP.

SNP identification from sequenced genomes

Currently, many crop species genomic sequences have been released. Sequenced genomes can be used to identify big amounts of SNPs in several respects. SNPs can be mined immediately in the genomic sequence in the event of heterozygous species such as grape or poplar, since two genome models have actually been produced. Two distinct specimens were sequenced in the scenario of rice (26) and grape (27) so that SNPs can be mined by comparing the two genome models (28). For *Arabidopsis thaliana* , genome re-sequencing panels were built based on the Col-0 chromosome sequences and used among 20 genotypes to identify SNPs based on hybridization (29,30).

As a public database (https://www.ncbi.nlm.nih.gov/), the Single Nucleotide Polymorphism database (dbSNP) was created. This database contains 1,648,103,041 SNPs for many organisms, e.g., *Brassica napus* (901.5 thousand SNPs); *Arabidopsis thaliana* (1.1 million SNPs); *Cicer arietinum* (519.1 thousand SNPs); *Phoenix dactylifera* (3.5 million SNPs); *Glycine max* (16.9 million SNPs) and *Zea mays* (54.3 million).

Genome annotation

Karp et al. (31) established Pathway methods that use the pathological algorithm to assess the enzymatic responses catalyzed for each gene product in a query genome and then combine the identified response list from a reference database against all accessible processes. Thompson et al. (32) produced the KEGG Automatic Annotation Server (KAAS) towards the manually mapped KEGG/GENES registry database (33) providing functional gene annotation through BLAST comparisons, single best hit (SBH) and bidirectional best scores (BBH). As yield were produced KO tasks to genes and anticipated KEGG processes. Overbeek *et al*. (2005) launched the SEED platform that offers a web-based,' subsystems'-based, relative genome annotation method. Subsystems were a collection of functional features observed in any prevalent biological system, including cellular processes, phenotypes, or complicated multi-subunit constructions.

Haft et al. (35) established the TIGR Comprehensive Microbial Resources (CMR) to enable users from accessing all completed sequences of bacterial genomes. CMR offers two kinds of assets for annotation: main annotation from the genome sequencing centre and TIGR annotation produced through an interactive TIGR annotation method. The CMR Pathway Tool Kit comprises of three classifications of pathway assessment tools: ' Genome Properties ' offers information on the features of species extracted from genomic data and literature references ; ' Genome Properties Detailed Comparison ' offers direct step-by-step information for selected clients of a collection of genomes ; and ' KEGG Pathway Display ' shows KEGG's presence-based pathway measures.

Conesa et al. (36) developed the Blast2GO (B2G), a universal guide for GO annotation, visualization and stats that provides sophisticated functional evaluation for non-model organism genomics studies. B2G is intended to enable annotation of instant and high-throughput succession and incorporate annotation-based data mining features. B2G utilizes BLAST (37) to discover homologs for fasta-formatted sequences of inputs. The program excerpts GO conditions by referring to existing annotation associations for each hit acquired. Finally, an annotation rule gives GO conditions to the sequence of queries. It is possible to visualize annotation and functional analysis in a graph form that reconstructs the GO interactions and highlights the most appropriate regions.

In an embedded genome framework, Markowitz et al. (38) introduced the Integrated Microbial Genomes (JGI) for comparative microbial genome assessment. The data model comprising the IMG scheme includes main genomic sequence data, algorithmically forecast and ordered gene models, pre-computed sequence resemblance data, functional annotation, and pathway data. Microbial gene statistical analysis in IMG is conducted in a relative framework of various microbial genomes where a range of instruments can be used to assess genomes in terms of genome-specific statistics, genes and sequence conserved.

The KEGG Orthology-Based Annotation System (KOBAS) was implemented by Wu et al. (39) that can provide statistical significance examinations for anticipated pathways. KO terms have been chosen based either on sequence resemblance to KEGG/GENES entries or on KEGG/GENES cross-database links when a list of sequence classifications is accessible in the databases. Compared to the background model, regularly occurring or statistically significantly enriched query sequence pathways are recognized forward to KO assignment.

Pireddu et al. (40) created a Path-A (Pathway Analyst) software that uses a number of query protein samples from a genome and defines which sequences are probable to occur using multiple sequence evaluation methods (e.g. SVM, BLAST and HMM) in any of its endorsed model pathways. The approach of the model pathway allows the pathway prediction algorithm to predict instances of a pathway with pathway structure variations that have never been observed in the training pathway set.

At present, Path-A offers abstract models for 10 pathways, covering 125 cases of genome-specific processes. Choi and Kim (4) established a Comparative Pathway Workbench System (ComPath) that enables researchers to compare biological processes in multiple genomes using a spreadsheet-style online interface where different sequence-based analyzes can be conducted either to match enzymes (e.g. sequence clustering) or processes (e.g. pathway hole recognition) to search for a *de novo* enzyme forecast algorithm.

References

1. Van Ooijen G, van den Burg HA, Cornelissen BJC, Takken FLW. Structure and function of resistance proteins in solanaceous plants. Annu Rev Phytopathol. 2007;45:43–72.

2. Rafalski A. Applications of single nucleotide polymorphism in crop genetics. Curr Opin Plant Biol. 2002;5:94–100.

3. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007; 449(7164):851–61.

4. Choi K, Kim S. ComPath: comparative enzyme analysis and annotation in pathway/subsystem contexts. BMC Bioinformatics. 2008;9:145.

5. Mokhtar MM, Adawy SS, El-Assal SE-DS, Hussein EHA. Genic and Intergenic SSR Database Generation, SNPs Determination and Pathway Annotations, in Date Palm (Phoenix dactylifera L.). PLoS One. Public Library of Science; 2016;11(7):e0159268.

6. Martin GB, Bogdanove AJ, Sessa G. Understanding the functions of plant disease resistance proteins. Annu Rev Plant Biol. 2003;54:23–61.

7. Takken FLW, Albrecht M, Tameling WIL. Resistance proteins: molecular switches of plant defence. Curr Opin Plant Biol. 2006;9(4):383–90.

8. Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J. 1999;20(3):317–32.

9. Wang Z-X, Yano M, Yamanouchi U, Iwamoto M, Monna L, Hayasaka H, et al. The Pib gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. Plant J. 1999;19(1):55–64.

10. Bryan GT, Wu K-S, Farrall L, Jia Y, Hershey HP, McAdams SA, et al. A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene Pi-ta. Plant Cell. 2000;12 (11):2033–45.

11. Zhou F, Kurth J, Wei F, Elliott C, Valè G, Yahiaoui N, et al. Cell-autonomous expression of barley Mla1 confers race-specific resistance to the powdery mildew fungus via a Rar1-independent signaling pathway. Plant Cell. 2001;13(2):337–50.

12. Parker JE, Coleman MJ, Szabò V, Frost LN, Schmidt R, van der Biezen EA, et al. The Arabidopsis downy mildew resistance gene RPP5 shares similarity to the toll and interleukin-1 receptors with N and L6. Plant Cell. 1997;9(6):879–94.

13. Richly E, Kurth J, Leister D. Mode of amplification and reorganization of resistance genes during recent Arabidopsis thaliana evolution. Mol Biol Evol. 2002;19(1):76–84.

14. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR--encoding genes in Arabidopsis. Plant Cell. 2003;15(4):809–34.

15. Cannon SB, Zhu H, Baumgarten AM, Spangler R, May G, Cook DR, et al. Diversity, distribution, and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. J Mol Evol. 2002;54(4):548–62.

16. Baumgarten A, Cannon S, Spangler R, May G. Genome-level evolution of resistance genes in Arabidopsis thaliana. Genetics. 2003;165(1):309–19.

17. Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, et al. Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. Mol Genet Genomics. 2004; 271:402–15.

18. Sanseverino W, Roma G, De Simone M, Faino L, Melito S, Stupka E, et al. PRGdb: a bioinformatics platform for plant resistance gene analysis. Nucleic Acids Res. 2010; 38:D814–21.

19. Potato Genome Sequencing Consortium. Genome sequence and analysis of the tuber crop potato. Nature. 2011;475(7355):189–95.

20. Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. Nature. 2012;485(7400):635–41.

21. Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, González VM, et al. The genome of melon

(Cucumis melo L.). Proc Natl Acad Sci. 2012;109 (29):11872–7.

22. Sanseverino W, Hermoso A, D'Alessandro R, Vlasova A, Andolfo G, Frusciante L, et al. PRGdb 2.0: Towards a community-based database model for the analysis of R-genes in plants. Nucleic Acids Res. 2013;41(D1171) :1167–71.

23. Adawy SS, Mokhtar MM, Alsamman AM, Sakr MM. Development of annotated EST-SSR database in olive (Olea europaea). Int J Sci Res. 2015;4(9):1063–73.

24. Schmid KJ, Sörensen TR, Stracke R, Törjék O, Altmann T, Mitchell-Olds T, et al. Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in Arabidopsis thaliana. Genome Res. 2003;13(6):1250–7.

25. Pavy N, Parsons LS, Paule C, MacKay J, Bousquet J. Automated SNP detection from a large collection of white spruce expressed sequences: contributing factors and approaches for the categorization of SNPs. BMC Genomics. 2006;7(1):174.

26. Yu J, Wang J, Lin W, Li S, Li H, Zhou J, et al. The genomes of Oryza sativa: a history of duplications. PLoS Biol. 2005;3(2):e38.

27. Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, et al. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS One. 2007;2(12):e1326.

28. Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH. An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. Genome Res. 2004;14(9):1812–9.

29. Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, et al. Common sequence polymorphisms shaping genetic diversity in Arabidopsis thaliana. Science (80-). 2007;317(5836):338–42.

30. Zeller G, Clark RM, Schneeberger K, Bohlen A, Weigel D, Rätsch G. Detecting polymorphic regions in Arabidopsis thaliana with resequencing microarrays. Genome Res. 2008;18(6):918–29.

31. Karp PD, Paley S, Romero P. The pathway tools software. Bioinformatics. 2002;18:S225–32.

32. Thompson W, Rouchka EC, Lawrence CE. Gibbs Recursive Sampler: finding transcription factor binding sites. Nucleic Acids Res. 2003;31(13):3580–5.

33. Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. Nucleic Acids Res. 2002;30(1):42–6.

34. Overbeek R, Begley T, Butler RM, Choudhuri J V, Chuang H-Y, Cohoon M, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. 2005;33(17) :5691–702.

35. Haft DH, Selengut JD, Brinkac LM, Zafar N, White O. Genome Properties: a system for the investigation of prokaryotic genetic content for microbiology, genome annotation and comparative genomics. Bioinformatics. 2005;21(3):293–306.

36. Conesa A, Götz S, García-gómez JM, Terol J, Talón M, Robles M, et al. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics.2005;21(18):3674–6.

37. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215(3):403–10.

38. Markowitz VM, Korzeniewski F, Palaniappan K, Szeto E, Werner G, Padki A, et al. The integrated microbial genomes (IMG) system. Nucleic Acids Res. 2006; 34(suppl 1):D344–8.

39. Wu J, Mao X, Cai T, Luo J, Wei L. KOBAS server: a web-based platform for automated annotation and pathway identification. Nucleic Acids Res. 2006;34: W720–4.

40. Pireddu L, Szafron D, Lu P, Greiner R. The Path-A metabolic pathway prediction web server. Nucleic Acids Res. 2006;34:W714–9.