PRSRW: AN EXPERT SYSTEM FOR POSTULATING AND INFERRING RESISTANCE GENES TO WHEAT STRIPE RUST

Yu Yang [1], Lianzhi Wang [1,*], Shichang Xu [2], Qiang Bian [1], Fengle Wang [2]

[1] *College of Information and Electrical Engineering, China Agricultural University, Beijing, P. R. China, 100083*

[2] *Institute of Plant Protection, CAAS, Beijing 100193*

[*] *Corresponding author, Address: College of Information and Electrical Engineering, China Agricultural University, Beijing 100083 P. R. China, Email: ndjsj862@cau.edu.cn*

Abstract: Postulating and inferring resistance genes to wheat stripe rust are a complicated process and need abundant expertise. An expert system for postulating and inferring resistance genes to wheat stripe rust (PRSRW) was developed by China Agricultural University. The process of PRSRW was described on the basis of the user's requirement. The system structure and its main components were introduced, including database, inference process and user interface. Some issues regarding knowledge acquisition and representation of the expert system were concerned. A mount of experimental results showed this system was feasible and effective. At last, a conclusion was summarized.

Keyword: expert system; resistance genes; wheat stripe rust; knowledge acquisition and representation

1. INTRODUCTION

Wheat stripe rust caused by *Puccinia striiformis* f. sp. *tritici*, is the most destructive disease of wheat in China. In terms of area affected by stripe rust, China is the largest epidemic region in the world (Stubbs 1988). Stripe rust is the most destructive to autumn-sown wheat in northwest and southwest China when susceptible cultivars are grown and the weather is favorable for the disease (Wan et al. 2004).The epidemics of 1950, 1964, 1990 and 2002 in China caused yield losses up to 6.0, 3.0, 1.8, and 1.3 million tons (Li & Zeng 2002; Wan et al.2004). The wheat's resistance to stripe rust mainly depends on the resistance genes. The use of resistant cultivars is the most economical and environmentally sound method to reduce damage caused by stripe rust. Gene postulation and inference helps to undertake a quick identification of the probable stripe rust resistance genes.

Gene postulation and inference is a complex and time-consuming process in gene analysis activities. It contains three important steps: resistance spectrum comparison, pedigree analysis, and integral information analysis. Resistance spectrum comparison not only consumes time, but often operates errors. It is difficult to make the conclusion of the probable resistance genes which the wheat cultivar contains, due to the shortage of pedigree information and integral information of the wheat cultivar. Thus, the system of a postulating and inferring gene developing becomes an urgent issue.

Computer application, which is revolutionizing information technology, is developing so rapidly that it is creating a huge opportunity for developing large scale application system. Expert system is an intelligent system which can treat a certain problem of a special area as an expert do by emulating human thinking.

PRSRW is an Expert System for postulating and inferring resistance genes to wheat stripe rust. It is supported by NSFC (National Nature Science Foundation of China).This intelligent system can simulate human wheat expert and postulate resistance genes to wheat stripe rust with a user-friendly interface. PRSRW contains a large amount of wheat information, which is needed to postulating and inferring resistance genes to wheat strip rust.

In this paper, we described the PRSRW. This paper was organized as follows: In Section 2 we described the domain background; Section3 emphasized the knowledge acquisition and representation of PRSRW; Section 4 showed the system architecture and development; Section 5 described the implementation of PRSRW; then Section 6 discussed and drew some conclusions.

2. DOMAIN BACKGROUND

There are five steps involved in postulating and inferring resistance genes to wheat stripe rust as general practice:

(1) Experiment. The aim of this step is to obtain the resistance spectrum. The method is using wheat cultivars inoculated with selected different isolates of Puccinia striiformis， and obtain the resistance spectrum. The resistance spectrum was the elementary information for the further postulating.

(2) Comparison of resistance spectrum. This step is crucial in gene postulation which includes resistance spectrum comparison and analysis. Expert would compare the resistance spectrum between the detected wheat cultivars and the known ones, and then derive the probable genes in the detected wheat cultivars.

According to the results of step (1) and step (2), it is easier to estimate the possible resistance genes. However, this possibility should be validated by more expertise. It is necessary to carry out the further steps which involve analyzing the wheat's pedigree and integral information, such as time, space, etc.

(3) Pedigree analysis. A pedigree is a diagram of family relationships that uses symbols to represent genetic relationships. Using genetic principles, the information presented in a pedigree can be analyzed to determine whether a given gene is inherited or not and what the pattern of inheritance is. The experts would analyze genes that the ancestors contain. If the ancestors contain the gene which wheat cultivars had been examined out from wheat cultivars, it should be ascertained that the wheat cultivar contains, otherwise go to step 4 to check the Integral information.

(4) Integral information analysis. It is an assistant part of the pedigree analysis. Information in terms of time and space should be considered as the fundamental data in gene analysis. Integral information analysis refers to the process of detailed research, such as when the wheat comes into china and where the original wheat comes from.

(5) Decision making. Normally, the first two steps may identify the possible resistance genes of the wheat and the latter two steps shall further confirm this result. By precise experiment and postulating exactly, a conclusion can be reached.

Before developing an expert system of postulating and inferring resistance genes to wheat stripe rust, it is important to process a detailed analysis of domain problems. Methods and functions should be considered on how to simulate the intellective process of the domain expert and how to cope with the problems in computer system.

3. KNOWLEDGE ACQUISITION AND REPRESENTATION

3.1 Knowledge acquisition

Knowledge acquisition plays an important role in the expert system. Many knowledge acquisition tools had been developed for transferring expert knowledge into knowledge base. Knowledge acquisition is a difficult and time-consuming process which is commonly recognized as a bottleneck in the development of an expert system (Hayes-Roth et al., 1983).In this study, we used a multiple knowledge acquisition approach: human experts interviewing, machine learning and knowledge acquisition system.

(1) Human experts interviewing. Knowledge engineers get domain knowledge from human experts directly. Then transform the knowledge into the computerized representation forms. This is a time-consuming job because of the gap between the Knowledge engineers and Knowledge experts. It is also very difficult to elicit and integrate knowledge from multiple experts (Chu & Hwang, 2007). Knowledge engineers should cope with these problems carefully and make the knowledge base clearly.

(2) Machine learning. The machine learning approaches could learn the useful static knowledge of well-known objects by collecting many useful cases and instances with/without the involvement of domain experts (Lin et al, 2008). Neural network approach was used to train parts of the data information which were fuzzy and not clear. Then some useful results were acquired.

(3) Knowledge acquisition system. A knowledge acquisition system had been developed to help the human experts and knowledge engineers input the facts and the rules. The interface could collect data about the resistance spectrum of the wheat, pedigree and also its time and space information. This knowledge acquisition system can only be used by experts who had the authority. It also could facilitate experts to input, modify, delete and search facts and rules.

3.2 Knowledge representation

The main job of knowledge representation was translating the knowledge acquired from human expert into the system knowledge base. The translation of acquired knowledge into a system usable representation represents a roadblock to knowledge-based system (KBS) development (Walczak, 1998).Object-orientation provides several solutions to persistent knowledge acquisition and knowledge representation problems including transportability, knowledge reuse, and knowledge growth (Walczak, 1998).

The object-orientation method was adopted as the formation of the knowledge representation in PRSRW. The core of the Object-orientation method is class. Three main classes were defined in the expert system. The separation of knowledge from representation structure combined with class enabled the system to present knowledge models to experts and users in any flexible format. It was also convenient for engineer to develop the system.

4. SYSTEM ARCHITECTURE AND DEVELOPMENT

PRSRW was consisted of six main components: database, knowledge base, integral database, knowledge acquisition subsystem, inferring engine and user interface. The structure was showed in Fig.1.

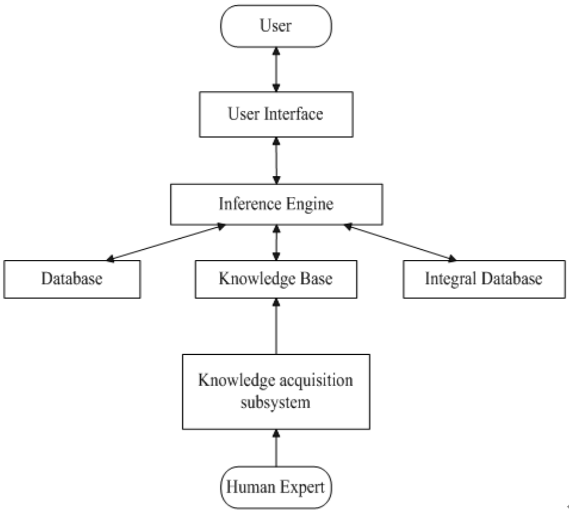

Fig.1: Structure of PRSRW

4.1 Development software and tools

The system used Visual Studio .NET as development tools, selected C# programming language as the development software and chose SQL language to index the database. Visual Studio .NET has the following characteristics: it is a Microsoft's integrated development environment for creating, documenting, running and debugging programs written in a variety of .NET programming language. Visual Studio.Net is a powerful and sophisticated tool for creating business-critical and mission-critical applications. The C# programming language is designed specifically for the .NET platform. It is a fully object-oriented, visual programming language. A programmer can create, run and test programs conveniently.

Database is the central and fundamental component of the expert system. It is responsible for storing all the data information used for the gene postulating and inferring in PRSRW. The computation results were then stored back in the database and displayed to user through user interface. This database was designed with SQL Server 2000.

The Database included a wheat information table, a gene information table, a pedigree database, an image table and a result table. These tables contained information of wheat's name, wheat's gene, wheat's pedigree, etc. The image database contained some pictures of the wheat.

4.2 User interface

User interface is the only communication between the system and its users. The target users of PRSRW were lab assistants, who were not good at computer. Therefore, the interface had been designed for facilitating easily access to input and update information. Users can also get the required information by selecting different conditions.

4.3 Inference process

Inference process simulates the intellective process of the domain expert. It is the core of the expert system. PRSRW users could query the system using an inference process that automatically matched the fact.

For example, through the experiment, users got the resistance spectrum, and then system would process the comparison with the resistance spectrums which were already known. A possible resistance gene of the wheat would be found. System would automatically query the knowledge base and found out its pedigree. If pedigree analysis found out the same gene from its ancestor, the system would confirm the possible resistance gene and showed the result by the interface. If pedigree analysis could not match the fact, a further step would involve analyzing the integral information of the wheat. Through the integral information analysis, there would be two results. One was confirming the possible resistance gene; the other was not containing the gene and the system would suggest the user to contact the human experts directly. Toward all the steps, the system would explain the inference process and indicated the final result on the user interface. After a process of the postulation and inference of the resistance gene, the system would save the results automatically. Fig.2 showed the inference process of the system.

5. TESTING AND IMPLEMENTATION

System testing is important before the actual use. It is to ensure whether the system would work accurately or not. System testing, such as debugging, rule checking and system maintenance would be carried out by system developers.

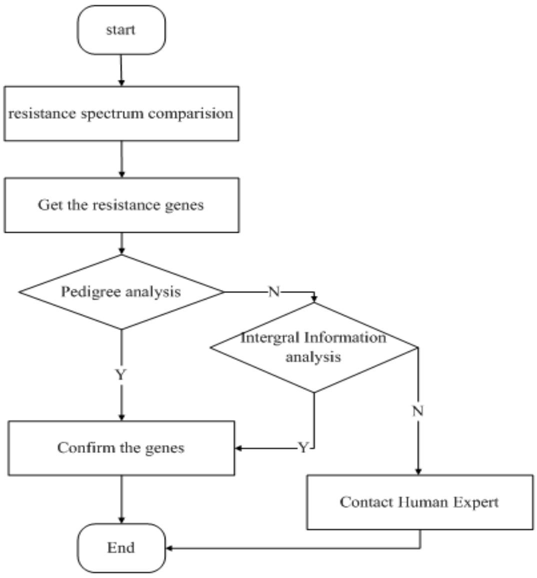

Fig.2: inference process of the system

The experiment data were collected from Chinese Academy of Agricultural Sciences, which is the research center of the wheat stripe rust. Some data were selected as the test data. After the system testing, user feedback was collected by conducting interviews. Totally, the system was an effective companion to lab assistant, human experts and other users.

Through some analysis of the system implementation tests and user feedbacks, the system showed the following characteristics:

The PRSRW expert system was regarded effectively in postulating and inferring resistance genes to wheat stripe rust, and it was saving users' time.

PRSRW was seemed as a useful tool for postulating and inferring resistance genes to wheat stripe rust. It could meet users' needs in different situation.

The friendly user interface was convenient and welcomed by users, and also the detailed comprehensive explanation of the result could help the users easily accept.

Some test results were showed in Fig.3

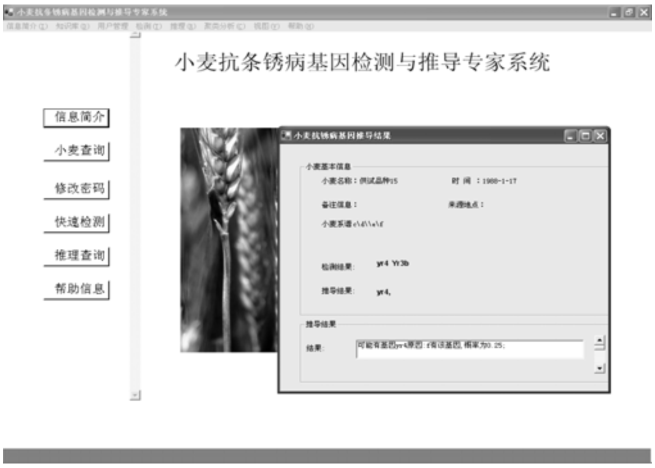

Fig.3: inferring result of the probable resistance gene

6. CONCLUSION AND DISCUSSION

In this research, an experts system called PRSRW was developed to postulate and infer resistance genes to wheat stripe rust. The System was able to facilitate experts to postulate and infer resistance genes quickly and accurately. The advantages were listed as following:

- PRSRW had a friendly user interface. It provided an easy way for the user accessing to the system, and also the expert could input and update the data information conveniently.
- Knowledge base and integral database contained about 400 data and 34 gene information for different type of wheat cultivars. PRSRW saved the results that facilitate users to search and review.
- PRSRW simulated the process of real human experts postulating and inferring the resistance gene by focusing on the resistance spectrum comparison and pedigree analysis.
- Knowledge base and database from human experts had been combined with the system successfully.
- PRSRW was designed as a multiple class user system. The low-level users only had parts of the authority to the system, while the high-level users had the whole authority.

Of course, no system is perfect and the PRSRW also had some disadvantages, such as:

- PRSRW was a stand-alone system. It only could be installed or delivered to users who need it. With the rapid development of the Internet and Intranet, a web-based system was convenient to the Internet users. Developing a web-based PRSRW would be our future work.

- Any expert system has limitations. Some problems might not be solved or achieve best results by PRSRW. A sub system should be developed that the users could communicate with the human experts directly.

In summary, postulation and inference of resistance genes to wheat stripe rust is a time-consuming job, and it often operates errors， the development of PRSRW could postulate and infer the resistance genes quickly and confirm the gene exactly. The research result demonstrated that the PRSRW was a valuable companion to dealing with the complexity problems.

ACKNOWLEDGEMENTS

This program was supported by NSFC (National Nature Science Foundation of China) under Grant NO.30471131. We would like to thank lots of domain experts of Chinese Academy of Agricultural Sciences for their support to the expert system of PRSRW.

REFERENCES

Chu, H.-C., Hwang, G.-J. A Delphi-based approach to developing expert systems with the cooperation of multiple experts Expert Systems with Applications.(2007)

Duan, Y., Fu, Z., & Li, D. Toward developing and using Web-based tele-diagnosis in aquaculture Expert Systems with Applications 25 (2003) 247–254

Fu, Z., Xu, F., Zhou, Y., & Zhang, X. Pig-vet: a web-based expert system for pig disease diagnosis. Expert Systems with Applications 29(2005), 93–103

Hayes-Roth, F., Waterman, D.A. & Lenat, D.B. (1983) Building expert systems, Addison Wesley, Reading, MA (1983).

Keles, A., Keles, A. ESTDD: Expert system for thyroid diseases diagnosis Expert Systems with Applications 34 (2008) 242–246

Kolmer, J.A. Postulation of leaf rust resistance genes in selected soft red winter wheat. Crop Sci. 2003,43:1266–1274

Li, D., Fu, Z., & Duan, Y. Fish-expert: a web-based expert system for fish disease diagnosis. Expert Systems with Applications, 23(2002), 311–320.

Li, Z., & Zeng, S. Wheat rusts in China. Beijing: China Agriculture Press.

Lin, H.-C., Tseng, S.-S., Teng, C.-W. Dynamic EMCUD for knowledge acquisition Expert Systems with Applications 34 (2008) 833–844