Screening and characterization of highly responsive genes related to the blast disease caused by *Pyricularia oryzae* in rice (*Oryza sativa*)

Dong Huy Gioi[1], Chu Duc Ha[2], Nguyen Quoc Trung[1], Bui Van Thang[3], Tran Van Tien[4], Bui Thi Thu Huong[1*]

[1]*Vietnam National University of Agriculture*
[2]*University of Engineering and Technology*
[3]*Vietnam National University of Forestry*
[4]*National Academy of Public Administration*

Sàng lọc và phân tích đặc tính của các gene có biểu hiện khác biệt với bệnh đạo ôn gây ra bởi nấm *Pyricularia oryzae* ở cây lúa (*Oryza sativa*) bằng công cụ tin sinh học dữ liệu lớn

Đồng Huy Giới[1], Chu Đức Hà[2], Nguyễn Quốc Trung[1], Bùi Văn Thắng[3], Trần Văn Tiến[4], Bùi Thị Thu Hương[1*]

[1]*Học viện Nông nghiệp Việt Nam*
[2]*Trường Đại học Công nghệ*
[3]*Trường Đại học Lâm nghiệp*
[4]*Học viện Hành chính Quốc gia*
[*]*Corresponding author: btthuonghp@gmail.com*

https://doi.org/10.55250/jo.vnuf.8.2.2023.019-027

Article info:
Received: 19/09/2023
Revised: 20/10/2023
Accepted: 06/11/2023

Keywords:
Bioinformatics, blast disease, characteristic, differentially expressed gene, Pyricularia oryzae, rice.

ABSTRACT

Blast, caused by Pyricularia oryzae, has been regarded as one of the most critical diseases in rice plants (Oryza sativa). Seeking a list of candidate genes is a major step for resistance breeding programs. In this study, at least 24 RNA-Seq datasets related to blast disease infection in rice plants have been obtained from databases. Among them, four microarray datasets obtained from blast-infected leaf samples in Indica rice varieties, including GSE122258, GSE126961, GSE78266 and GSE39635 were selected for further in silico analysis. Particularly, a total of 216 differentially expressed genes in blast-treated leaves were obtained by exploring four microarray atlas. Among them, at least 32 blast-responsive genes that exhibited similar expression patterns were selected for further bioinformatics analysis. Particularly, these differentially expressed genes encode important transcription factors and enzyme families. By using Expasy and Yloc tools, the physic-chemical properties of these proteins, including lengths, molecular weights, iso-electric points and grand average of hydropathy were greatly variable and the majority of these blast-responsive gene-encoding proteins was localized in the cytoplasm and nucleus. To sum up, this current study has provided fundamental knowledge about potential candidate differentially expressed genes for further functional characterization of genes aiming to control blast resistance in rice.

Từ khóa:
Bệnh đạo ôn, đặc tính, gene có biểu hiện khác biệt, lúa gạo, Pyricularia oryzae, tin sinh học.

TÓM TẮT

Bệnh đạo ôn, gây ra bởi nấm Pyricularia oryzae được xem là một trong những tác nhân gây bệnh với tổn thất nghiêm trọng nhất ở lúa gạo (Oryza sativa). Việc tìm kiếm các gene lúa ứng viên là bước cần thiết cho công tác chọn tạo giống kháng bệnh đạo ôn. Trong nghiên cứu này, ít nhất 24 dữ liệu RNA-Seq liên quan đến lây nhiễm bệnh đạo ôn trên cây lúa đã được thu thập từ các cơ sở dữ liệu. Trong số đó, bốn dữ liệu microarray từ mẫu lá lây nhiễm bệnh đạo ôn ở các giống lúa Indica, bao gồm GSE122258, GSE126961, GSE78266 và GSE39635 đã được lựa chọn cho các phân tích in silico tiếp theo. Cụ thể, tổng số 216 gene có biểu hiện khác biệt đã được sàng lọc từ các dữ liệu microarray liên quan đến lây nhiễm bệnh đạo ôn trên cây lúa. Trong

đó, tổng số 32 gene đáp ứng có xu hướng biểu hiện tương tự nhau đã được lựa chọn để tiến hành các phân tích tin sinh học. Các gene có biểu hiện khác biệt mã hóa cho nhóm nhân tố phiên mã và enzyme quan trọng. Bằng công cụ Expasy và Yloc, các tính chất lý hóa, bao gồm kích thước, trọng lượng phân tử, điểm đẳng điện và độ ưa nước trung bình của protein đa dạng và đa số các phân tử này phân bố tại tế bào chất và nhân tế bào. Kết quả của nghiên cứu này đã cung cấp thông tin về gene có biểu hiện khác biệt ứng viên cho phân tích chức năng gene nhằm kiểm soát tính kháng bệnh đạo ôn ở lúa.

1. INTRODUCTION

Rice (*Oryza sativa*), a fundamental cereal crop with a history dating back over 10,000 years to the Yangtze River basin in China and the Ganges River delta in India [1], is a crucial dietary staple for over half the world's population, valued for its adaptability to diverse environments and high caloric yield [2]. Beyond its nutritional significance, rice has deep cultural and economic importance, especially in Asian societies [3], symbolizing prosperity and fertility and forming the primary income source for numerous smallholder farmers [3, 4]. However, rice production faces significant challenges, notably the blast disease caused by *Pyricularia oryzae* [5], which leads to major yield losses and poses a threat to global food security [6]. In Vietnam, the total area impacted by blast disease in 2012 was approximately 366,412 hectares, of which 11,400 hectares were affected by severe infections and at least eight hectares were completely lost; milling output was reduced by 10 - 25% [7]. This critical disease, characterized by lesions on various plant parts, necessitates costly and environmentally impactful management strategies, highlighting the need for sustainable practices [5]. Research efforts are thus directed towards developing resistant rice varieties, employing improved agricultural techniques, and exploring eco-friendly disease control methods, underscoring the criticality of a multifaceted approach to ensure the continuous, sustainable cultivation of this essential global food source [4].

Concurrently, there is an intensified focus on elucidating the pathogen's biological characteristics, particularly its genetic variability and infection modalities. Refinements in agronomic practices, including crop rotation, strategic timing of planting, meticulous water management, and judicious fertilization, are being pursued to mitigate disease impact. Current research endeavors in the realm of rice blast disease management, attributable to *P. oryzae*, adopt a holistic strategy. This encompasses the development of genetically resistant rice cultivars, leveraging both conventional breeding methodologies [8] and cutting-edge genetic modification techniques such as CRISPR/Cas9 [9]. To date, the significance of regulatory proteins and functional proteins in the response mechanism to blast disease in plants has been understood [10]. The majority of the responsive genes were found to encode transcription factors, metabolic enzymes, and transport proteins via the phytohormone signaling pathway [11]. The recent advancement of RNA-Seq technology has enabled the investigation of gene transcription levels under experimental conditions. For example, analysis of gene expression levels in the mutant rice line MIR7695-Ac showed 281 differently expressed genes (153 and 128 induced and reduced genes, respectively) under blast fungal infection circumstances [12]. Similarly, other studies have used microarray data to investigate gene expression in the rice genome in response to artificial blast infection [13-15]. Although blast resistance genes have been identified in the rice genome [6], recent microarray data [12-15] combined with the available rice assembly [16] can enable to screening and listing of differently expressed genes associated with blast disease, thereby providing candidate genes for functional analysis.

There has been little genetic research on blast resistance genes in Vietnamese rice varieties. Four resistance genes, *Pi1*, *Pik-h*, *Pita*, and an unknown gene, were discovered in the genetic background of the Vietnamese landrace ‘Tẻ

tép', which was previously known to be resistant to blast disease [17]. It discovered the resistance gene, namely *Pi-VT7*, in the Vietnamese landrace 'Chiêm bạc' on chromosome 12 in the same location as *Pita* and *Pi12*(t) [18]. After screening 500 accessions from the Mekong River Delta in South Vietnam, at least 23 highly resistant cultivars and 80 moderately resistant cultivars have been discovered [19]. However, the survey of blast-responsive genes at the genome scale has been not considered. In this present study, a comprehensive search was performed against the datasets related to blast disease infection in rice plants. A potential list of highly responsive genes related to blast disease infection in rice plants was provided. The annotation of each gene was then investigated by using the rice assembly. Finally, the features and predicted subcellular localization of these putative genes were fully-characterized.

2. RESEARCH METHODOLOGY

2.1. Materials

Recent rice assembly, including genome and proteome [16] has been explored in the NCBI (NCBI accession: PRJNA1040485) and Phytozome databases (Phytozome genome identifier: 323) [20]. All microarray datasets related to the blast disease infection in rice plants were sought against the NCBI GEO portal [21].

2.2. Methods

- Screening of microarray datasets: To seek all datasets related to the blast disease infection, the NCBI GEO portal [21] was used as previously described [22] with minor changes. Particularly, two keywords, like "blast disease" and "*Oryza sativa*" were used to collect all putative datasets related to the blast disease infection in rice plants. All necessary information, including samples, cultivars and platforms was collected for further analysis.

- Re-analysis of microarray datasets: All datasets were analyzed by using R script [23] as previously described [22]. In this study, an up-regulated gene was defined as fold-change ≥ 2.0, whereas a down-regulated gene was considered as fold-change ≤ -2.0. Among them, the fold-change was calculated as the ratio between the Fragments Per Kilobase of transcript per Million mapped reads in treated samples and those in controls [22]. The overlapped genes were sorted by using the VennPainter tool [24].

- Annotation of differently expressed genes: All potential genes were retrieved against the genome and proteome of rice [16] available in NCBI and Phytozome databases [20] to gain the full-length protein sequence and coding DNA sequence. Gene function was predicted by using the Blast2GO software [25] as previously reported [22]. The conserved domain of each protein was validated by using the Pfam database [26].

- Analysis of protein features: The common characteristics of each protein were identified as previously described [22, 27]. Briefly, the full-length protein sequence was analyzed via the Expasy Protparam tool [28]. Four features, including molecular mass (kilo Dalton, kDa), protein length (amino acid residues), iso-electric point and grand average of hydropathy were estimated. Among them, iso-electric point < 7, = 7 and > 7 exhibited acidic, neutral and base, while the grand average of hydropathy < 0 and > 0 exhibited hydrophilicity and hydrophobicity [28].

- Prediction of subcellular localization of proteins: All full-length protein sequences were searched against the Yloc web-based tool [29] to seek organelle-specific signal peptides as previously described [27]. Based on the signal peptides, major organelles, including the nucleus, cytoplasm, mitochondrion, plasma membrane, extracellular space, endoplasmic reticulum, peroxisome, Golgi apparatus, vacuole and chloroplast were targeted for the plant model [29].

3. RESULTS AND DISCUSSION

3.1. Exploration of all microarray datasets related to the blast disease infection in rice

In order to get insight into the molecular mechanism of how rice plants respond to artificial blast disease, all microarray datasets were screened in the NCBI GEO tool [21]. As a result, a total of 24 RNA-Seq datasets related to the blast disease infection in rice plants have been fully reported (Figure 1).

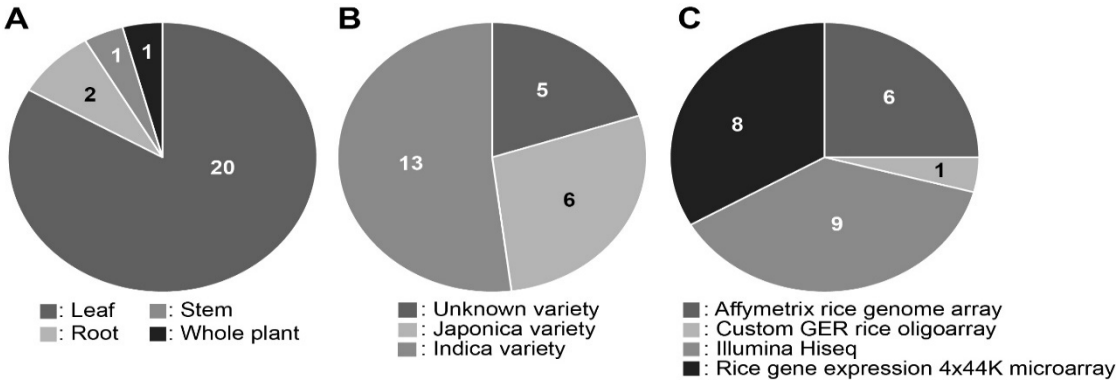

Figure 1. Summary of 24 microarray datasets related to the blast disease infection in rice

Particularly, 20 (out of 24) datasets were recorded in infected leaf samples, whereas two (out of 24) datasets were analyzed in roots from infected rice plants (Figure 1A). Next, a number of datasets (13 out of 24) was collected from Indica rice varieties, whereas six (out of 24) datasets were obtained from Japonica rice varieties (Figure 1B). The remaining datasets (five out of 24) were recorded in unknown rice varieties. Finally, four platforms, including Affymetrix rice genome array (six out of 24), custom GER rice oligoarray (one out of 24), Illumina Hiseq (nine out of 24) and rice gene expression 4×44K microarray (eight out of 24) were used to perform these RNA-Seq datasets (Figure 1C). In this study, the criteria of selection was based on the platform, sample and rice variety. As expected, four microarray datasets obtained in infected leaf samples from Indica rice varieties, including GSE122258 [12], GSE126961 [13], GSE78266 [14] and GSE39635 [15] were selected for further *in silico* analyses.

Previously, in order to investigate the molecular mechanism of how rice plants respond to cytokinin, at least five distinct RNA-Seq datasets of cytokinin-treated roots were selected for re-analysis [30]. Kong et al. (2019) reported a core collection of salt-responsive genes in various rice genotypes during the seedling period by re-analyzing 96 RNA-Seq datasets [31]. A similar approach has been explored to identify a set of genes involved in iron homeostasis in rice [32]. These principles of collecting data were slightly similar to this present study.

3.2. Establishment of a core set of differentially expressed genes under the blast disease infection in rice

To provide a core set of differentially expressed genes in blast disease-infected leaf samples, four RNA-Seq datasets in previous studies [12-15] were re-analyzed and overlapped all results. By using R script [23], a total of 216 highly responsive genes (|fold-change| ≥ 2) was found in four datasets (Figure 2).

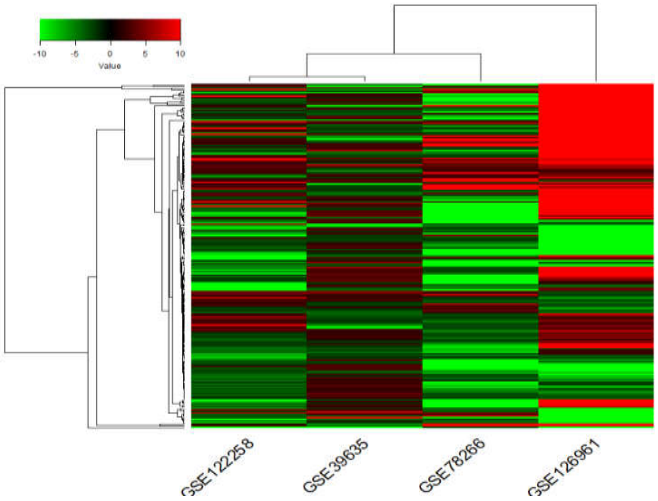

Figure 2. Expression profiles of a core set of differentially expressed genes in blast disease-infected leaf samples

In this study, 32 differentially expressed genes exhibited similar expression patterns in four datasets (Table 1). Among them, eight (out of 32) genes were up-regulated in all treatments, whereas 24 (out of 32) genes were down-regulated in four datasets. Particularly, *Os06g0278000* was noted to exhibit the highest level of expression in blast-infected leaf samples, by approximately 101,83-fold), whereas *Os03g0423300* was the most reduced gene (approximately -1203,82-fold). All detailed expression levels of 32 blast-responsive genes were well-described in Table 1.

Table 1. Expression profiles of a core set of 32 differentially expressed genes in four microarray datasets

No.	GeneID	GSE39635	GSE78266	GSE126961	GSE122258
1	*Os03g0172100*	2,22	**22,34**	18,84	2,18
2	*Os04g0606200*	2,14	**17,60**	3,29	2,17
3	*Os06g0174700*	4,42	7,89	**9,01**	7,58
4	*Os06g0278000*	2,59	6,49	**101,83**	3,22
5	*Os06g0714800*	**9,75**	3,04	6,44	3,52
6	*Os09g0498500*	3,74	3,91	**6,50**	3,06
7	*Os10g0364900*	2,03	6,75	**12,71**	3,37
8	*Os11g0569600*	3,49	3,24	**101,43**	8,68
9	*Os01g0293000*	-4,05	-2,61	-40,34	**-56,92**
10	*Os01g0370900*	-2,53	-5,24	-2,90	**-6,82**
11	*Os02g0214900*	-2,16	-2,87	-2,95	**-7,64**
12	*Os02g0258300*	-2,09	-3,91	**-113,66**	-8,18
13	*Os02g0320800*	-2,07	-2,51	-3,35	**-7,68**
14	*Os03g0266900*	-2,83	**-16,32**	-3,65	-3,19
15	*Os03g0416500*	-2,99	-2,03	**-20,70**	-3,37
16	*Os03g0423300*	-9,88	-2,63	**-1203,82**	-55,43
17	*Os03g0752800*	-2,29	-2,40	**-260,10**	-6,73
18	*Os04g0121100*	-3,46	**-7,51**	-7,70	-2,94
19	*Os04g0690500*	-2,15	-2,36	**-6,91**	-4,70
20	*Os05g0406800*	-2,11	-4,06	**-23,33**	-2,65
21	*Os05g0477600*	-4,15	**-6,94**	-2,56	-4,03
22	*Os06g0212900*	-2,16	-10,93	**-21,05**	-8,66
23	*Os07g0122000*	**-79,64**	-5,43	-11,82	-5,91
24	*Os07g0440100*	-3,01	**-33,85**	-10,15	-4,99
25	*Os07g0663600*	-3,49	**-25,63**	-7,41	-2,80
26	*Os08g0452500*	-2,55	-3,87	**-16,77**	-3,99
27	*Os09g0437100*	-2,03	-11,86	**-24,97**	-8,47
28	*Os09g0502100*	-2,22	-2,39	**-55,30**	-5,05
29	*Os10g0540800*	-4,21	-2,73	**-136,03**	-10,81
30	*Os10g0556100*	-2,26	**-16,99**	-9,08	-2,98
31	*Os11g0106900*	-3,40	-14,44	**-26,80**	-2,53
32	*Os12g0143800*	-4,83	-3,92	**-9,74**	-2,54

Previously, a core set of responsive genes was also recorded in rice plants under adverse environmental conditions. For example, at least 291 differentially expressed genes, including 205 induced and 86 reduced genes have been found in cytokinin-treated roots in rice plants [30]. Meanwhile, in order to investigate the salt stress response in rice plants at the molecular scale, a total of 5559 key genes were found to be highly altered under salt stress, and 3210 differentially expressed genes were found during the recovery process [31].

3.3. Annotation of a core set of differentially expressed genes under the blast disease infection in rice

In this study, 32 differentially expressed genes were continued to analyze their potential function via Blast2GO [25] and Pfam tools [26]. As expected, a majority of blast-responsive genes (22 out of 32) was found their putative function, of which 9 and 13 genes encoded regulatory and functional proteins, respectively.

In contrast, 10 (out of 32) differentially expressed genes, including *Os04g0606200, Os06g0174700, Os06g0714800, Os10g0364900, Os04g0121100, Os04g0690500, Os07g0122000, Os07g0440100, Os09g0502100* and *Os10g0540800* were not fully-annotated (Table 2).

Table 2. Gene function of a core set of 32 differentially expressed genes in four microarray datasets

No.	GeneID	Gene function	Categorization
1	*Os03g0172100*	Leucine zipper protein	Regulatory protein
2	*Os04g0606200*	-	-
3	*Os06g0174700*	-	-
4	*Os06g0278000*	Glucoamylase	Functional protein
5	*Os06g0714800*	-	-
6	*Os09g0498500*	FAD dependent oxidoreductase	Functional protein
7	*Os10g0364900*	-	-
8	*Os11g0569600*	Receptor kinase-like protein	Functional protein
9	*Os01g0293000*	S-adenosylmethionine synthetase 1	Functional protein
10	*Os01g0370900*	Glutathione transferase	Functional protein
11	*Os02g0214900*	Class-I type histone deacetylase	Functional protein
12	*Os02g0258300*	Zinc finger	Regulatory protein
13	*Os02g0320800*	Iron/ascorbate-dependent oxidoreductase	Functional protein
14	*Os03g0266900*	Heat shock protein	Regulatory protein
15	*Os03g0416500*	6-phosphogluconolactonase	Functional protein
16	*Os03g0423300*	Stearoyl-acyl carrier protein desaturse	Functional protein
17	*Os03g0752800*	MADS-box transcription factor	Regulatory protein
18	*Os04g0121100*	-	-
19	*Os04g0690500*	-	-
20	*Os05g0406800*	Leucine-rich repeat	Regulatory protein
21	*Os05g0477600*	Alpha-expansin	Functional protein
22	*Os06g0212900*	Heat shock protein	Regulatory protein
23	*Os07g0122000*	-	-
24	*Os07g0440100*	-	-
25	*Os07g0663600*	Short-chain dehydrogenase/reductase	Functional protein
26	*Os08g0452500*	Auxin responsive SAUR protein	Regulatory protein
27	*Os09g0437100*	Auxin responsive SAUR protein	Regulatory protein
28	*Os09g0502100*	-	-
29	*Os10g0540800*	-	-
30	*Os10g0556100*	Beta-expansin	Functional protein
31	*Os11g0106900*	Lateral organ boundaries	Functional protein
32	*Os12g0143800*	Disrupted meiotic cDNA 1 protein	Regulatory protein

Note: -: Unknown function.

This analysis indicated that 22 (out of 32) blast-responsive genes were annotated to function as transcription factors and specific enzymes related to the reduction of reactive oxygen species. Particularly, seven (out of 22) annotated genes, including *Os03g0172100* (encoded Leucine zipper protein), *Os02g0258300* (encoded zinc finger), *Os03g0266900* and *Os06g0212900* (encoded Heat shock protein), *Os03g0752800* (encoded MADS-box), *Os08g0452500* and *Os09g0437100* (encoded Auxin responsive SAUR protein) were recorded to encode plant-specific transcription factor families. Meanwhile, this current study revealed several blast-responsive genes related to enzymes involved in cellular metabolism. For example, *Os06g0278000* encodes glucoamylase, which was demonstrated to degrade protein molecules [33]. Previously, *Os06g0278000* was also reported to be highly expressed under the submergence stress condition [33].

Recently, the function of the core set of responsive genes related to adverse environmental conditions has been investigated. For example, the core set of cytokinin-

responsive genes was recorded to include cytokinin oxidases/dehydrogenases and the type-A response regulators [30]. To understand the responsiveness of rice plants against salt stress conditions, a gene ontology has been performed in a collection of differentially expressed genes [31]. More specifically, the key salt-responsive genes encoded the mitogen-activated protein kinase, Ca^{2+} signal transduction pathway, transcription factors and other important functional proteins [31].

3.4. Characterization of a core set of differentially expressed genes under the blast disease infection in rice

To understand the characteristics of a core set of differentially expressed genes under the blast disease infection in rice, the physic-chemical features and sub-cellular localization of each protein molecule were then analyzed by using various bioinformatics tools [28, 29]. All features and sub-cellular localization of 32 proteins were provided in Table 3.

Table 3. Protein feature and sub-cellular localization of a core set of 32 differentially expressed genes in four microarray datasets

No.	Protein name	Features				Sub-cellular localization
		L	mW	pI	GRAVY	
1	Os03g0172100	82	8.23	4.93	0.31	C
2	Os04g0606200	71	7.60	9.36	0.13	C
3	Os06g0174700	407	42.60	9.20	-0.55	N
4	Os06g0278000	333	36.64	5.17	-0.24	V
5	Os06g0714800	136	14.63	6.83	-0.41	C
6	Os09g0498500	416	43.98	5.86	-0.11	P
7	Os10g0364900	80	9.23	6.82	-1.21	N
8	Os11g0569600	1102	118.08	6.19	0.07	V
9	Os01g0293000	396	43.31	5.22	-0.27	C
10	Os01g0370900	248	28.15	9.78	-0.08	C
11	Os02g0214900	510	56.50	5.54	-0.47	N
12	Os02g0258300	842	93.05	7.29	-0.77	N
13	Os02g0320800	302	33.89	5.88	-0.45	C
14	Os03g0266900	154	17.37	6.18	-0.71	C
15	Os03g0416500	296	33.27	4.96	-0.35	C
16	Os03g0423300	418	45.37	7.77	-0.21	Chl
17	Os03g0752800	246	28.42	9.12	-0.77	N
18	Os04g0121100	752	81.25	6.50	-0.04	V
19	Os04g0690500	227	24.69	10.35	-0.54	C
20	Os05g0406800	394	42.03	5.24	0.23	PM
21	Os05g0477600	246	25.88	8.13	-0.12	ExS
22	Os06g0212900	470	50.53	6.64	0.03	ER
23	Os07g0122000	414	47.02	5.81	-0.40	C
24	Os07g0440100	422	32.95	10.49	0.07	PM
25	Os07g0663600	302	31.37	6.71	0.14	Chl
26	Os08g0452500	133	14.17	6.19	-0.18	M
27	Os09g0437100	165	16.64	8.78	-0.20	Chl
28	Os09g0502100	401	43.70	4.68	-0.90	N
29	Os10g0540800	853	95.56	6.39	-0.11	C
30	Os10g0556100	286	31.36	5.96	-0.32	ExS
31	Os11g0106900	156	17.40	6.42	-0.25	N
32	Os12g0143800	344	37.54	6.04	-0.10	N

Note: L: Length (amino acid residues), mW: Molecular weight (kDa), pI: Iso-electric point, GRAVY: Grand average of hydropathy, C: Cytoplasm, N: Nucleus, M: Mitochondrion, ExS: Extracellular space, PM: Plasma membrane, V: Vacuole, Chl: Chloroplast, ER: Endoplasmic reticulum, P: Peroxisome.

It has been found that the sizes of proteins were varied from 71 (Os04g0606200) to 1102 amino acid residues (Os11g0569600). The molecular masses of 32 proteins ranged from 7,6 to 118,08 kDa. A majority of proteins (22 out of 32) exhibited iso-electric points < 7.0, ranging from 4.68 (Os09g0502100) to 6.83 (Os06g0714800). The remaining proteins, (10 out of 32), including Os02g0258300, Os03g0423300, Os05g0477600, Os09g0437100, Os03g0752800,

Os06g0174700, Os04g0606200, Os01g0370900, Os04g0690500 and Os07g0440100 had iso-electric points > 7.0. Next, a large number of proteins (25 out of 32) had the grand average of hydropathy < 0, ranging from -1.21 (Os10g0364900) to -0.04 (Os04g0121100), whereas seven (out of 32) proteins, including Os06g0212900, Os11g0569600, Os07g0440100, Os04g0606200, Os07g0663600, Os05g0406800 and Os03g0172100 exhibited the grand average of hydropathy > 0.

It has been believed that the sub-cellular localization of proteins could reveal the potential function of these molecules [34]. Thus, the sub-cellular localization of 32 proteins was predicted and well-described in Table 3. Particularly, 11 (out of 32) proteins were localized in the cytoplasm, while eight (out of 32) proteins were distributed in the nucleus. Additionally, three proteins, including Os06g0278000, Os04g0121100 and Os11g0569600 were predicted in vacuole, while three proteins, such as Os03g0423300, Os09g0437100 and Os07g0663600 were found in chloroplast. Two proteins, including Os07g0440100 and Os05g0406800 were noted to be localized in the plasma membrane, while Os10g0556100 and Os05g0477600 were suggested to be in extracellular space. Finally, Os09g0498500 was localized in the peroxisome, while Os08g0452500 and Os06g0212900 were distributed in the mitochondrion and endoplasmic reticulum, respectively. The prediction of sub-cellular localization of protein molecules could be important evidence for further analysis of interaction with *Avr* genes from pathogens [35]. For example, *Pi54*, a blast-resistant gene, encoded a cytoplasm-specific protein [35]. Overexpression of *GFP::Pi54* confers blast disease resistance in rice [35].

4. CONCLUSION

In this study, a meta-analysis of transcriptomic studies of blast disease-treated rice leaf samples was performed to define a core set of blast-responsive genes. Particularly, a total of 24 RNA-Seq datasets related to blast disease infection in rice plants was intensively surveyed to select four suitable microarray datasets for re-analysis. A core set of 32 blast-responsive genes, which mostly include genes encoding transcription factors and specific enzymes related to the reduction of reactive oxygen species was proposed. Next, the structural analysis indicated that these 32 proteins were variable in sizes, masses, iso-electric points, grand average of hydropathy and sub-cellular localization. To sum up, further studies may extend this data to demonstrate the potential functions of unannotated blast-responsive genes.

REFERENCES

[1]. Huang X., Kurata N., Wei X., Wang Z., Wang A., Zhao Q., Zhao Y., Liu L., Lu H., Li W., Guo Y., Lu Y., Zhou C., Fan D., Weng Q., Zhu C., Huang T., Zhang L., Wang Y., Feng L., Furuumi H., Kubo T., Miyabayashi T., Yuan X., Xu Q., Dong G., Zhan Q., Li C., Fujiyama A. & Toyoda A. (2012). A map of rice genome variation reveals the origin of cultivated rice. Nature. 490: 497-501.

[2]. Wenefrida I., Utomo H. S. & Linscombe S. D. (2013). Mutational breeding and genetic engineering in the development of high grain protein content. Journal of Agricultural and Food Chemistry. (48): 11702-11710.

[3]. Shi J., An G., Weber A. P. M. & Zhang D. (2023). Prospects for rice in 2050. Plant, Cell & Environment. (4): 1037-1045.

[4]. Tran Dang Khanh, Vu Xuan Duong, Phi Cong Nguyen, Tran Dang Xuan, Nguyen Thanh Trung, Khuat Huu Trung, Dong Huy Gioi, Nguyen Huy Hoang, Tran Hoang Dung, Do Minh Trung & Bui Thi Thu Huong. (2021). Rice breeding in Vietnam: Retrospects, challenges and prospects. Agriculture. (5): 397.

[5]. Nguyen Thi Minh Nguyet, Hoang Hoa Long, Nguyen Ba Ngoc, Nguyen Thi Nhai, Nguyen Thi Thanh Thuy, Hayashi N. & Fukuta Y. (2019). Diversity and distribution of rice blast (*Pyricularia oryzae* Cavara) races in Vietnam. Plant Diseases. (2): 381-387.

[6]. Ashkani S., Yusop M. R., Shabanimofrad M., Azady A., Ghasemzadeh A., Azizi P. & Latif M. A. (2015). Allele mining strategies: Principles and utilisation for blast resistance genes in rice (*Oryza sativa* L.). Current Issues in Molecular Biology. (17): 57-73.

[7]. Statistical Yearbook of Vietnam. (2013). Socialist Republic of Vietnam, General Statistics Office, Statistical Publishing House, Hanoi. 935.

[8]. Miah G., Rafii M. Y., Ismail M. R., Puteh A. B., Rahim H. A., Asfaliza R. & Latif M. A. (2013). Blast resistance in rice: a review of conventional breeding to molecular approaches. Molecular Biology Reports. (3): 2369-2388.

[9]. Wang F., Wang C., Liu P., Lei C., Hao W., Gao Y., Liu Y. G. & Zhao K. (2016). Enhanced rice blast resistance by CRISPR/Cas9-targeted mutagenesis of the ERF transcription factor gene *OsERF922*. PLoS One. (4): e0154027.

[10]. Srivastava D., Shamim K., Mahesh M., Anurag P., Pramila K., Deepak Y., Prashant S., Mohammed H. & Singh K. N. (2017). Current status of conventional and molecular interventions for blast resistance in rice. Rice Science. (6): 299-321.

[11]. Helliwell E. E. & Yang Y. (2013). Molecular strategies to improve rice disease resistance. Methods in Molecular Biology. (956): 285-309.

[12]. Sánchez-Sanuy F., Peris-Peris C., Tomiyama S., Okada K., Hsing Y. I., San Segundo B. & Campo S. (2019). Osa-miR7695 enhances transcriptional priming in defense responses against the rice blast fungus. BMC Plant Biology. (1): 563.

[13]. Norvienyeku J., Lin L., Waheed A., Chen X., Bao J., Aliyu S. R., Lin L., Shabbir A., Batool W., Zhong Z., Zhou J., Lu G. & Wang Z. (2021). Bayogenin 3-O-cellobioside confers non-cultivar-specific defence against the rice blast fungus Pyricularia oryzae. Plant Biotechnology Journal. (3): 589-601.

[14]. Kadotani N., Akagi A., Takatsuji H., Miwa T. & Igarashi D. (2016). Exogenous proteinogenic amino acids induce systemic resistance in rice. BMC Plant Biology. 16: 60.

[15]. Chujo T., Miyamoto K., Shimogawa T., Shimizu T., Otake Y., Yokotani N., Nishizawa Y., Shibuya N., Nojiri H., Yamane H., Minami E. & Okada K. (2013). OsWRKY28, a PAMP-responsive transrepressor, negatively regulates innate immune responses in rice against rice blast fungus. Plant Molecular Biology. (1-2): 23-37.

[16]. Rice genomes. (2014). The 3,000 rice genomes project. Gigascience. (1): 7.

[17]. Wang L., Zhao L., Zhang X., Zhang Q., Jia Y., Wang G., Li S., Tian D., Li W. H. & Yang S. (2019). Large-scale identification and functional analysis of NLR genes in blast resistance in the Tetep rice genome sequence. Proceedings of the National Academy of Sciences of the United States of America. (37): 18479-18487.

[18]. Nguyen Thi Thuan, Bigirimana J., Roumen E., Hofte M. & Straeten D. (2003). Genetic analysis of resistance to blast in the Vietnamese rice cultivar 'Chiembac'. Communications in Agricultural and Applied Biological Sciences. (68): 375-380.

[19]. Luu Van Quynh & Bui Ba Bong. (1999). Study on durable resistance of rice varieties to blast disease in the Mekong Delta of Vietnam. Omonrice. (7): 9-14.

[20]. Goodstein D. M., Shu S., Howson R., Neupane R., Hayes R. D., Fazo J., Mitros T., Dirks W., Hellsten U., Putnam N. & Rokhsar D. S. (2012). Phytozome: A comparative platform for green plant genomics. Nucleic Acids Research. (40): D1178-D1186.

[21]. Barrett T., Wilhite S. E., Ledoux P., Evangelista C., Kim I. F., Tomashevsky M., Marshall K. A., Phillippy K. H., Sherman P. M., Holko M., Yefanov A., Lee H., Zhang N., Robertson C. L., Serova N., Davis S. & Soboleva A. (2013). NCBI GEO: Archive for functional genomics data sets - Update. Nucleic Acids Research. (41): D991-D995.

[22]. Tong Van Hai, Nguyen Quoc Trung, Cao Viet Bach, Le Thi Ngoc Quynh, La Viet Hong, Ha Thi Quyen, Pham Minh Trien & Chu Duc Ha. (2021). Investigation on the structure of auxin-regulated genes induced in drought and heat stresses in tomato (*Solanum lycopersicum*). Proceedings of 2021 Vietnam National Conference of Biotechnology. 570-574.

[23]. Liao Y., Smyth G. K. & Shi W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Research. (8): e47.

[24]. Lin G., Chai J., Yuan S., Mai C., Cai L., Murphy R. W., Zhou W. & Luo J. (2016). VennPainter: A tool for the comparison and identification of candidate genes based on Venn diagrams. PLoS One. (4): e0154315.

[25]. Gotz S., Garcia-Gomez J. M., Terol J., Williams T. D., Nagaraj S. H., Nueda M. J., Robles M., Talon M., Dopazo J. & Conesa A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research. (10): 3420-3435.

[26]. El-Gebali S., Mistry J., Bateman A., Eddy S. R., Luciani A., Potter S. C., Qureshi M., Richardson L. J., Salazar G. A., Smart A., Sonnhammer E. L. L., Hirsh L., Paladin L., Piovesan D., Tosatto S. C. E. & Finn R. D. (2019). The Pfam protein families database in 2019. Nucleic Acids Research. (47): D427-D432.

[27]. La Viet Hong, Chu Duc Ha, Tran Duy Cuong, Nguyen Huu Kien, Le Thi Ngoc Quynh, Hoang Minh Chinh, Cao Phi Bang, Pham Cong Anh Tuyen, Nguyen Duc Bach, Nguyen Quoc Trung, Nguyen Van Loc, Ha Van Chien, Le Thi Hien, Le Huy Ham, Le Tien Dung & Tran Phan Lam Son. (2022). Insights into the gene and protein structures of the CaSWEET family members in chickpea (*Cicer arietinum*), and their gene expression patterns in different organs under various stress and abscisic acid treatments. Gene. (819): 146210.

[28]. Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R. D. & Bairoch A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Research. (13): 3784-3788.

[29]. Briesemeister S., Rahnenfuhrer J. & Kohlbacher O. (2010). YLoc--an interpretable web server for predicting subcellular localization. Nucleic Acids Research. (38): W497-W502.

[30]. Polko J. K., Potter K. C., Burr C. A., Schaller G. E. & Kieber J. J. (2021). Meta-analysis of transcriptomic studies of cytokinin-treated rice roots defines a core set of cytokinin response genes. The Plant Journal. (5): 1387-1402.

[31]. Kong W., Zhong H., Gong Z., Fang X., Sun T., Deng X. & Li Y. (2019). Meta-analysis of salt stress transcriptome responses in different rice genotypes at the seedling stage. Plants . (3): 64.

[32]. Shekhawat P., Kanwar S., Shaswati Y., Banita S., Prafull S. P. & Ram H. (2023). Meta-analysis of transcriptomics studies identifies novel attributes and set of genes involved in iron homeostasis in rice. Functional & Integrative Genomics. (4): 336.

[33]. Wu Y. S. & Yang C. Y. (2020). Comprehensive transcriptomic analysis of auxin responses in submerged rice coleoptile growth. International Journal of Molecular Sciences. (4): 1292.

[34]. Kiraga J., Mackiewicz P., Mackiewicz D., Kowalczuk M., Biecek P., Polak N., Smolarczyk K., Dudek M. R. & Cebrat S. (2007). The relationships between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. BMC Genomics. (8): 163.

[35]. Singh J., Gupta S. K., Devanna B. N., Singh S., Upadhyay A. & Sharma T. R. (2020). Blast resistance gene *Pi54* over-expressed in rice to understand its cellular and sub-cellular localization and response to different pathogens. Scientific Reports. (1): 5243.