

Sàng lọc và phân tích đặc tính của các gene có phản ứng khác biệt  
với lây nhiễm nhân tạo bệnh bạc lá do vi khuẩn  
*Xanthomonas oryzae* pv. *oryzae* ở cây lúa (*Oryza sativa*)

Đông Huy Giới<sup>1</sup>, Phạm Ngọc Minh<sup>1,2</sup>, Nguyễn Thị Minh Nguyệt<sup>3</sup>,  
Thân Thị Hoa<sup>4</sup>, Chu Đức Hà<sup>2\*</sup>, Bùi Thị Thu Hương<sup>1</sup>

<sup>1</sup>Học viện Nông nghiệp Việt Nam

<sup>2</sup>Trường Đại học Công nghệ

<sup>3</sup>Viện Di truyền Nông nghiệp

<sup>4</sup>Trường Đại học Nông lâm Bắc Giang

Screening and characterization of differentially expressed genes  
under the inoculation of *Xanthomonas oryzae* pv. *oryzae*  
causing bacterial leaf blight in rice (*Oryza sativa*)

Dong Huy Gioi<sup>1</sup>, Pham Ngoc Minh<sup>1,2</sup>, Nguyen Thi Minh Nguyet<sup>3</sup>,  
Than Thi Hoa<sup>4</sup>, Chu Duc Ha<sup>2\*</sup>, Bui Thi Thu Huong<sup>1</sup>

<sup>1</sup>Vietnam National University of Agriculture

<sup>2</sup>University of Engineering and Technology

<sup>3</sup>Agricultural Genetics Institute

<sup>4</sup>Bac Giang Agriculture and Forestry University

\*Corresponding author: cd.ha@vnu.edu.vn

<https://doi.org/10.55250/jo.vnu.12.5.2023.039-050>

**TÓM TẮT**

Bệnh bạc lá là một trong những bệnh hại chính trong sản xuất lúa (*Oryza sativa*) ở Việt Nam. Công nghệ chọn tạo giống hiện đại giúp phát triển các dòng lúa kháng bạc lá, tuy nhiên, cơ chế phản ứng của cây lúa với bạc lá ở cấp độ phân tử vẫn chưa được khám phá toàn diện. Nghiên cứu này được thực hiện nhằm khai thác các microarray bằng phân tích tin sinh học dữ liệu lớn để sàng lọc và đánh giá đặc tính của gene có biểu hiện khác biệt (DEG) với lây nhiễm bạc lá. Cụ thể, tổng số 1630 DEG ( $|\text{fold-change}| \geq 2$ ), bao gồm 895 gene được tăng cường biểu hiện và 735 gene bị kìm hãm biểu hiện được sàng lọc từ dữ liệu microarray GSE57670 và GSE108504 liên quan đến lây nhiễm bệnh bạc lá ở cây lúa. Trong đó, nghiên cứu đã lựa chọn 51 DEG có đáp ứng mạnh ( $|\text{fold-change}| \geq 10$ ) để phân tích biểu hiện gene, chú giải chức năng, đặc tính lý hóa và vị trí cư trú nội bào. Kết quả cho thấy, 51 DEG có biểu hiện đa dạng tại 10 vị trí chính trong điều kiện thường. Các DEG này mã hóa chủ yếu cho protein điều hòa, điển hình như enzyme, nhân tố phiên mã và protein chức năng. Phân tích cho thấy các protein có tính chất lý hóa rất đa dạng, trong khi các protein này được dự đoán là cư trú tại các bào quan chính trong tế bào. Kết quả của nghiên cứu đã cung cấp những dẫn liệu quan trọng cho phân tích chức năng gene làm cơ sở để cải thiện tính kháng bệnh bạc lá ở cây lúa.

**ABSTRACT**

Bacterial leaf blight has been regarded as one of the major diseases that occurred in rice (*Oryza sativa*) cultivation in Vietnam. Recent achievements in the development of bacterial blight-resistant rice lines have been recorded, however, the mechanism of bacterial blight responsiveness at the molecular level has been not clearly described. This current study was performed to explore the previous microarray datasets by using various big-data bioinformatics tools and characterize putative differentially expressed genes (DEG). Particularly, a total of 1630 DEGs ( $|\text{fold-change}| \geq 2$ ), including 895 up-regulated genes and 735 down-regulated genes has been obtained from two microarray datasets related to the inoculation of bacterial blight in rice plants, namely GSE57670 and GSE108504. Among them, we selected 51 putative strongly-responsive DEGs ( $|\text{fold-change}| \geq 10$ ) for further expression analysis, gene annotation, feature analysis and subcellular localization. Our study revealed that 51 DEGs exhibited a variable expression level in 10 major tissues in the normal condition. These DEGs

**Thông tin chung:**

Ngày nhận bài: 04/08/2023

Ngày phản biện: 08/09/2023

Ngày quyết định đăng: 27/09/2023

**Từ khóa:**

biểu hiện gene, bệnh bạc lá, chú giải gene, đặc tính, gene có biểu hiện khác biệt, lúa gạo.

**Keywords:**

bacterial blight, characteristic, differentially expressed gene, expression level, gene annotation, rice.

mostly encoded regulatory proteins, especially enzymes and transcription factors, and functional proteins. Next, the physico-chemical properties of these proteins were greatly variable, while they were mostly localized in major organelles. Taken together, our study could provide important evidence for further functional characterizations toward the improvement of bacterial blight resistance in rice plants.

## 1. ĐẶT VẤN ĐỀ

Sản xuất lúa gạo (*Oryza sativa*) là một trong những nhiệm vụ chính trị quan trọng của ngành trồng trọt, đóng vai trò xương sống cho nền kinh tế và đảm bảo tình hình an ninh lương thực tại Việt Nam [1]. Là loại cây trồng chính trong hệ thống sản xuất nông nghiệp, lúa gạo cung cấp nguồn thực phẩm cơ bản cho dân số và tạo ra nguồn thu nhập chính cho hơn 70% hộ gia đình nông dân. Việt Nam là một trong những quốc gia sản xuất và xuất khẩu gạo hàng đầu thế giới, góp phần quan trọng vào tổng sản phẩm nội địa và cải thiện cán cân thương mại quốc gia [1]. Tuy nhiên, phát triển lúa gạo tại các địa phương hiện nay đang gặp nhiều khó khăn gây ra bởi tình trạng biến đổi khí hậu [2]. Bên cạnh các điều kiện bất thuận [2], sâu bệnh hại cũng là một trong những bất lợi sinh học chính gây ảnh hưởng rõ rệt đến sinh trưởng, phát triển, năng suất và chất lượng của cây lúa [1, 3]. Trong đó, bệnh bạc lá lúa, gây ra bởi vi khuẩn *Xanthomonas oryzae* pv. *oryzae* [4-6], là một trong những tác nhân gây sụt giảm năng suất nặng nề do bệnh có khả năng phát tán nhanh và gây tổn thương nghiêm trọng đến lá [6]. Do vậy, trong chiến lược phát triển lúa gạo hiện nay, tăng cường tính kháng bệnh bạc lá sẽ góp phần quan trọng trong việc đảm bảo an ninh lương thực và tạo ra sự đa dạng trong sản xuất lúa gạo của Việt Nam [1].

Đến nay, nhiều nghiên cứu đã được tiến hành nhằm tích hợp các gene kháng bạc lá phân lập từ nhiều nguồn giống cho gene [5-8]. Khoảng 46 gene kháng bạc lá đã được công bố, trong khi số lượng vi khuẩn *X. oryzae* được ghi nhận là hơn 100 chủng [6], điều này đặt ra nhiệm vụ tìm kiếm các gene có phổ kháng rộng. Việc tiếp cận bằng công cụ tin sinh học dữ liệu lớn đã cho phép sàng lọc và đưa ra danh sách các gene có phản ứng khác biệt (differentially expressed

gene, DEG) liên quan đến xử lý bất lợi ở cây trồng [9, 10]. Bằng việc kết hợp nhiều dữ liệu microarray trên National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) [11], tập hợp gồm 10 DEG đã được xác định có đáp ứng mạnh với điều kiện hạn hán và nhiệt độ cao, đồng thời cảm ứng với con đường tín hiệu auxin trên đối tượng cây cà chua (*Solanum lycopersicum*) [10]. Trong khi đó, kết hợp các dữ liệu microarray liên quan đến xử lý hạn và mặn ở mẫu rễ của cây đậu gà (*Cicer arietinum*) trên NCBI GEO [11] đã sàng lọc được tổng số 41 DEG [10]. Các DEG tìm được đều mã hóa cho các nhóm nhân tố phiên mã (transcription factor, TF), protein vận chuyển và enzyme chuyển hóa tham gia vào quá trình sinh lý quan trọng trong tế bào [9, 10]. Những kết quả này đã tạo phương pháp luận quan trọng cho việc sàng lọc và phân tích các DEG liên quan đến cơ chế phản ứng với bệnh bạc lá ở cây lúa gạo. Ví dụ, 115 DEG đã được xác định có mức độ phiên mã đáp ứng với lây nhiễm bệnh bạc lá trên giống lúa Y73 [25].

Mục đích của nghiên cứu nhằm phân tích toàn bộ các DEG với điều kiện lây nhiễm nhân tạo bệnh bạc lá ở cây lúa. Theo đó, các dữ liệu microarray liên quan đến lây nhiễm bệnh bạc lá trên đối tượng cây lúa đã được tái phân tích nhằm tìm ra toàn bộ các DEG đáp ứng với điều kiện lây nhiễm bệnh. Từ đó, một danh sách DEG có đáp ứng mạnh với lây nhiễm bệnh bạc lá được chọn lọc để chú giải, phân tích đặc tính và xác định biểu hiện tại các cơ quan chính trên cây.

## 2. PHƯƠNG PHÁP NGHIÊN CỨU

### 2.1. Vật liệu nghiên cứu

Dữ liệu genome và proteome của *O. sativa* được khai thác trên cơ sở dữ liệu Rice Genome Annotation Project [12, 13], NCBI và Phytozome [14].

Dữ liệu microarray trong điều kiện lây nhiễm nhân tạo bệnh bạc lá, bao gồm GSE57670 và GSE108504 [15] được thu thập trên cơ sở dữ liệu NCBI GEO [11].

## 2.2. Phương pháp nghiên cứu

Phương pháp phân tích dữ liệu microarray: Dữ liệu GSE57670 và GSE108504 [15] được truy cập để tái phân tích các DEG dựa theo mô tả trong nghiên cứu trước đây [9, 10]. Cụ thể, DEG được quy định là gene có sự thay đổi mức độ phiên mã khác biệt, với giá trị fold-change  $\geq 2$  (tăng cường biểu hiện) hoặc  $\leq -2$  (kìm hãm biểu hiện) [9, 10]. Theo đó, các dữ liệu microarray được phân tích và biểu diễn bằng ngôn ngữ R theo mô tả trong nghiên cứu trước đây với chỉnh sửa nhỏ. Công cụ VennPainter [16] được sử dụng để kết hợp các DEG tìm thấy trong hai dữ liệu microarray [15]. Bên cạnh đó, các DEG tiếp tục được sử dụng để đánh giá biểu hiện tại các mô cơ quan chính trong cây lúa gạo trong điều kiện thường thông qua NCBI Sequence Read Archive (SRA) [17]. Cụ thể, mã định danh của DEG được truy vấn trên Rice Genome Annotation Project [12, 13] để phân tích biểu hiện tại 10 mẫu mô thu thập trên cây lúa, bao gồm lá 20 ngày tuổi, cụm hoa trước khi xuất hiện, cụm hoa sau khi xuất hiện, bao phấn, nhụy hoa, thân, hạt 5 ngày sau thụ phấn, hạt 10 ngày sau thụ phấn, phôi 25 ngày sau thụ phấn và nội nhũ 25 ngày sau thụ phấn. Gene được quy ước là “có biểu hiện” khi giá trị Fragments per kilobase million (FPKM)  $> 10$ , “tăng cường biểu hiện” khi FPKM  $> 30$ , “biểu hiện mạnh” khi FPKM  $> 50$ , “biểu hiện đặc thù” khi FPKM  $> 100$  và “dưới ngưỡng phát hiện” khi FPKM  $< 10$  [17].

Phương pháp chú giải chức năng gene: Các DEG được chú giải dựa vào hệ quy chiếu của cây lúa gạo trên Rice Genome Annotation Project [12, 13], NCBI và Phytozome [14]. Theo đó, mã định danh, bao gồm mã định danh gene, mã định danh protein và mã định danh locus của các DEG được thu thập kèm theo trình tự protein đầy đủ, trình tự exon mã hóa và trình tự vùng gene. Vùng domain quy định chức năng

của protein được phân tích trên Pfam [18].

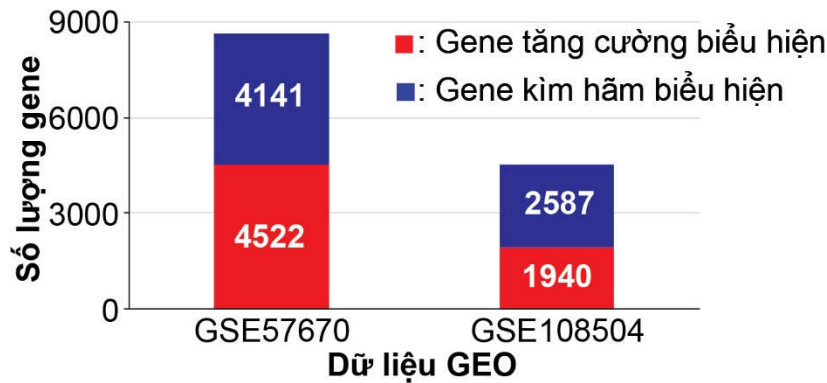
Phương pháp phân tích đặc tính của protein: Đặc điểm cơ bản của protein được phân tích trên công cụ ExPasy ProtParam [19] dựa vào mô tả trước đây [9, 10, 20-23]. Cụ thể, trình tự protein đầy đủ được sử dụng để phân tích các tính chất cơ bản, bao gồm kích thước protein, trọng lượng phân tử (molecular weight, mW) và điểm đẳng điện lý thuyết (isoelectric point, pI) [19]. Theo đó, tính acid và base của protein được thể hiện qua giá trị pI, tương ứng với  $pI < 7$  và  $> 7$  [19].

Phương pháp xác định vị trí cư trú nội bào của protein: Trình tự protein đầy đủ được truy vấn trên Yloc [24] để tìm kiếm đoạn trình tự peptide tín hiệu đặc trưng cho các bào quan trong tế bào dựa trên mô tả trong nghiên cứu trước đây [9, 10, 20-23]. Theo đó, tổng số 10 bào quan, bao gồm nhân tế bào, tế bào chất, ty thể, màng sinh chất, khoảng gian bào, mạng lưới nội chất, thể Golgi, peroxisome, không bào và lục lạp được sử dụng để xây dựng cho mô hình phân bố của phân tử protein truy vấn trong cấu trúc tế bào thực vật [24].

## 3. KẾT QUẢ VÀ THẢO LUẬN

### 3.1. Sàng lọc các DEG có phản ứng khác biệt với điều kiện lây nhiễm bệnh bạc lá từ dữ liệu microarray

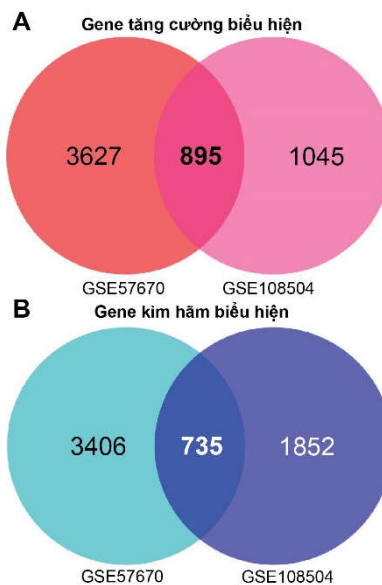
Trong nghiên cứu này, hai dữ liệu lây nhiễm bệnh bạc lá trên cây lúa, bao gồm GSE57670 và GSE108504 [15] đã được khai thác để tái phân tích và sàng lọc nhằm tìm ra một danh sách các DEG đáp ứng với điều kiện xử lý bệnh. Kết quả tái phân tích đã cho thấy số lượng lớn DEG liên quan đến cơ chế đáp ứng của cây lúa khi nhiễm bệnh bạc lá (Hình 1). Cụ thể, trong dữ liệu GSE57670, tổng số 8.663 DEG đã được xác định, bao gồm 4.522 gene được tăng cường biểu hiện và 4.141 gene bị kìm hãm biểu hiện ở mô lá khi lây nhiễm bệnh bạc lá. Trong khi đó, 4.527 DEG, bao gồm 1.940 gene tăng cường biểu hiện và 2.587 gene bị giảm biểu hiện ở mô lá nhiễm bệnh bạc lá ở dữ liệu GSE108504. Các DEG này được kết hợp bằng công cụ VennPainter [16] nhằm tìm ra một danh sách các gene có phản ứng với việc lây nhiễm bệnh bạc lá ở cây lúa gạo.



Hình 1. Số lượng gene có biểu hiện khác biệt trong dữ liệu microarray liên quan đến lây nhiễm bệnh bạc lá ở cây lúa gạo

Kết quả kết hợp các dữ liệu microarray đã xác định được tổng số 1.630 DEG có đáp ứng, bao gồm 895 gene được tăng cường biểu hiện ( $\text{fold-change} \geq 2$  trong cả hai dữ liệu microarray) và 735 gene bị giảm biểu hiện ( $\text{fold-change} \leq -2$  trong cả hai dữ liệu microarray) (Hình 2). Theo đó, nghiên cứu đã

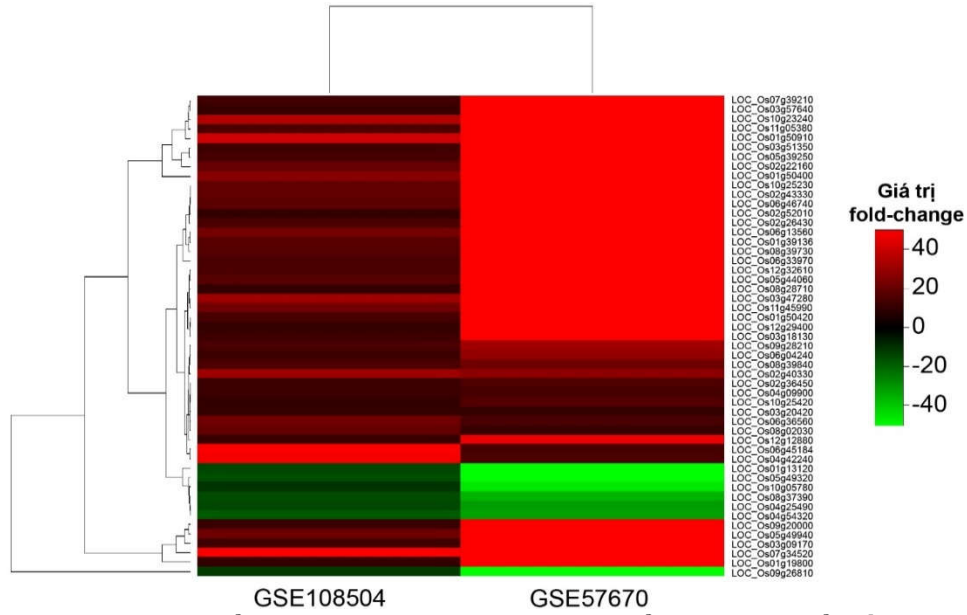
lựa chọn những DEG có mức độ biểu hiện thay đổi rất mạnh ( $|\text{fold-change}| \geq 10$ ) để tiếp tục định danh và phân tích đặc tính (Hình 3). Kết quả thu được 51 gene có mức độ biểu hiện rất mạnh ở mẫu lá trong cả hai dữ liệu microarray (Hình 3). Các DEG này tiếp tục sử dụng để tiến hành phân tích *in silico* tiếp theo.



Hình 2. Biểu đồ Venn sàng lọc các gene có biểu hiện khác biệt đáp ứng với lây nhiễm bệnh bạc lá trên hai dữ liệu microarray

Trong nghiên cứu trước đây, việc khai thác dữ liệu microarray để phân tích các DEG nhằm giải thích cơ chế đáp ứng của cây lúa gạo với bệnh bạc lá ở cấp độ phân tử đã được triển khai. Ví dụ, lây nhiễm đồng thời chủng vi khuẩn bạc lá X11-5A và gây nóng nhân tạo trên giống lúa IRBB61 đã xác định được tổng số 8.499 DEG có phản ứng với các điều kiện xử lý [5]. Lây nhiễm chủng vi khuẩn bạc lá T7174 trên dòng

lúa chuyển gene biểu hiện quá mức *WRKY45* cho thấy khoảng 1.664 gene tăng cường biểu hiện so với giống đối chứng Nipponbare [7]. Trước đó, phân tích RNA-Seq trên giống lúa kháng bạc lá Y73 lây nhiễm với các chủng vi khuẩn bạc lá PXO341 đã cho thấy 115 DEG có đáp ứng ở thời điểm 24 giờ sau lây nhiễm, bao gồm 103 gene được tăng cường biểu hiện và 12 gene bị giảm biểu hiện [8].

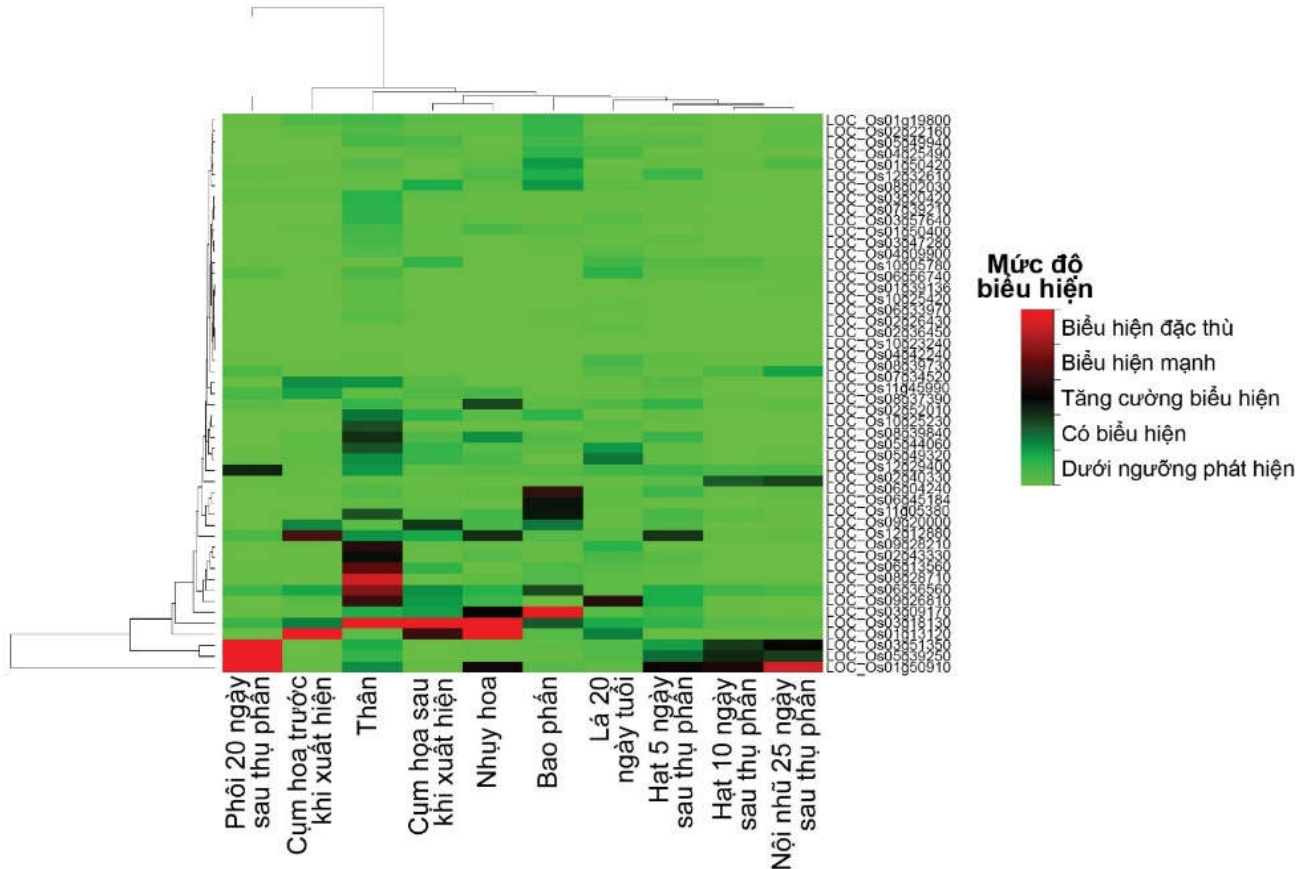


Hình 3. Mức độ biểu hiện của 51 gene có mức độ biểu hiện thay đổi rất mạnh thu thập từ hai dữ liệu microarray liên quan đến lây nhiễm bệnh bạc lá

### 3.2. Đánh giá mức độ biểu hiện của các DEG phản ứng mạnh với bệnh bạc lá trong điều kiện thường

Trong nghiên cứu này, mức độ biểu hiện của 51 DEG phản ứng mạnh với bệnh bạc lá được

đánh giá trong điều kiện thường tại 10 vị trí trên cây lúa gạo. Dựa trên công cụ NCBI SRA [17], các DEG có mức độ biểu hiện khác nhau ở các vị trí cơ quan khác nhau trong cây ở điều kiện thường (Hình 4).



Hình 4. Mức độ biểu hiện của 51 gene có mức độ biểu hiện thay đổi rất mạnh thu thập từ hai dữ liệu microarray liên quan đến lây nhiễm bệnh bạc lá

Kết quả cho thấy hai DEG, bao gồm LOC\_Os05g39250 và LOC\_Os03g51350 có biểu hiện đặc thù tại phôi 20 ngày sau thụ phấn, trong khi LOC\_Os01g50910 có biểu hiện đặc thù tại phôi 20 ngày sau thụ phấn và nội nhũ 25 ngày sau thụ phấn. Trong khi đó, LOC\_Os01g13120 có biểu hiện đặc thù tại cụm hoa trước khi xuất hiện và nhụy hoa. Đáng chú ý, biểu hiện của LOC\_Os03g18130 được ghi nhận đặc trưng tại thân, nhụy hoa và cụm hoa sau khi xuất hiện. Một DEG khác, LOC\_Os03g09170 được cho là biểu hiện đặc thù tại bao phấn, trong khi 6 DEG, bao gồm LOC\_Os09g26810, LOC\_Os06g36560, LOC\_Os08g28710, LOC\_Os06g13560, LOC\_Os02g43330 và LOC\_Os09g28210 có biểu hiện mạnh và đặc thù tại thân (Hình 4).

### 3.3. Chú giải chức năng của DEG phản ứng mạnh với điều kiện lây nhiễm bệnh bạc lá

Để làm rõ về cơ chế đáp ứng với bệnh bạc lá ở cấp độ phân tử ở cây lúa, nghiên cứu đã tiến hành chú giải chức năng của 51 DEG có phản ứng mạnh với điều kiện lây nhiễm bệnh. Kết quả đối chiếu trên các cơ sở dữ liệu [12-14] cho thấy, hầu hết các DEG đều mã hóa cho những nhóm protein điều hòa (regulatory protein, RP) và protein chức năng (functional protein, FP) quan trọng (Bảng 1). Nghiên cứu đã chú giải chức năng của 36 DEG, bao gồm 12 DEG mã hóa cho các FP và 24 DEG mã hóa cho các RP. Bên cạnh đó, 15 DEG chưa được chú giải chức năng trên Pfam [18].

Cụ thể, 24 (trên tổng số 51) DEG mã hóa cho các RP, trong đó 14 DEG mã hóa cho enzyme, 7 DEG mã hóa cho TF và 3 DEG được cho là mã hóa cho RP khác (Bảng 1). Theo đó, LOC\_Os01g50400 và LOC\_Os01g50420 được ghi nhận là mã hóa cho mitogen-activated protein kinase kinase kinase 15, hay còn gọi là apoptosis signal-regulating kinase 3, tham gia vào quá trình phosphoryl hóa các nhân tố phiên mã nhằm điều hòa sự biểu hiện gene, điều hòa quá trình chết theo chương trình của tế bào, đồng thời liên quan đến con đường tín hiệu mitogen-activated protein kinase. Trước đó, một thành viên thuộc nhóm mitogen-activated protein kinase kinase kinase đã được chứng minh là điều hòa âm tính trong con đường mitogen-activated protein kinase, từ đó thể hiện

phổ kháng rộng với bệnh bạc lá [25]. Trong khi đó, LOC\_Os04g09900 mã hóa cho ent-kaurene synthase, một enzyme tham gia vào quá trình sinh tổng hợp hormone gibberellin. Gibberellin được cho là gây giảm tính kháng ở cây trồng với vi khuẩn [3, 26], như *X. oryzae* gây bệnh trên cây sắn (*Manihot esculenta*) [27] và trên cây lúa [28]. Tiếp theo, LOC\_Os08g39730 và LOC\_Os11g05380 mã hóa cho cytochrome P450, là enzyme xúc tác chuỗi chuyển hóa thứ cấp nội bào, được chứng minh có liên quan đến tính kháng bệnh ở thực vật thông qua điều hòa sinh tổng hợp oxylipin và hormone jasmonic acid [29]. Trước đây, một thành viên của cytochrome P450, đặt tên là CYP71P1 được chứng minh là tăng cường tính kháng bệnh bạc lá ở cây lúa thông qua xúc tác quá trình chuyển hóa tryptamine thành serotonin [29]. LOC\_Os10g25420 mã hóa cho GDSL-like lipase/acylhydrolase, một enzyme thủy phân lipid và điều hòa cân bằng nội môi lipid [30]. Vai trò của nhóm enzyme này liên quan đến tính kháng bệnh bạc lá ở cây lúa đã được chứng minh trong nghiên cứu trước đây của Gao và cộng sự (2017) [30]. Cụ thể, hai thành viên của nhóm enzyme này, bao gồm OsGLIP1 và OsGLIP2 bị kìm hãm ở mẫu lá lúa khi nhiễm bệnh bạc lá và xử lý với hormone salicylic acid [30]. Biểu hiện quá mức hai gene này làm tăng cường tích lũy monogalactosyldiacylglycerol và digalactosyldiacylglycerol, từ đó làm tăng tính nhạy cảm của dòng cây chuyển gene với bệnh bạc lá [30]. Một số DEG khác, như LOC\_Os06g36560 mã hóa cho inositol oxygenase (sinh tổng hợp myo-inositol), LOC\_Os06g13560 mã hóa cho S-adenosylmethionine dependent carboxyl methyltransferase (chuyển gốc methyl từ S-adenosylmethionine sang chất nhận), LOC\_Os04g25490 mã hóa cho cytokinin-O-glucosyltransferase 2 (xúc tác cho phản ứng chuyển hóa glucose trong sinh tổng hợp hormone cytokinin), LOC\_Os07g34520 mã hóa cho isocitrate lyase (tham gia vào chuỗi chuyển hóa citric acid), LOC\_Os08g02030 mã hóa cho transferase (xúc tác chuyển hóa nhóm chức), LOC\_Os08g39840 mã hóa cho lipoxygenase (xúc tác phản ứng oxi hóa các acid béo chưa no), LOC\_Os03g18130 mã hóa cho asparagine

synthetase (sinh tổng hợp asparagine) và LOC\_Os03g20420 mã hóa cho alpha-N-arabinofuranosidase A (xúc tác phản ứng thủy

phân các gốc alpha-L-arabinofuranoside ở alpha-L-arabinoside) cũng đã được xác định có đáp ứng mạnh với điều kiện lây nhiễm bệnh bạc lá.

**Bảng 1. Thông tin chú giải và đặc tính của 51 gene có mức độ biểu hiện thay đổi rất mạnh thu thập từ hai dữ liệu microarray liên quan đến lây nhiễm bệnh bạc lá**

TT	Mã định danh locus	Chú giải chức năng	Vai trò	L	mW	pI
1	LOC_Os01g50400	mitogen-activated protein kinase kinase kinase 15	Enzyme	419	44,24	4,35
2	LOC_Os01g50420	mitogen-activated protein kinase kinase kinase 15	Enzyme	542	58,59	4,58
3	LOC_Os04g09900	ent-kaurene synthase	Enzyme	768	87,22	5,40
4	LOC_Os06g36560	inositol oxygenase	Enzyme	309	35,80	4,75
5	LOC_Os06g13560	S-adenosylmethionine dependent carboxyl methyltransferase	Enzyme	375	42,34	6,51
6	LOC_Os04g25490	cytokinin-O-glucosyltransferase 2	Enzyme	476	51,33	4,88
7	LOC_Os07g34520	isocitrate lyase	Enzyme	573	63,05	7,23
8	LOC_Os08g02030	transferase	Enzyme	469	49,56	6,12
9	LOC_Os08g39730	cytochrome P450	Enzyme	507	55,74	6,87
10	LOC_Os10g25420	GDSL-like lipase/acylhydrolase	Enzyme	339	36,18	9,20
11	LOC_Os11g05380	cytochrome P450	Enzyme	521	56,58	8,41
12	LOC_Os08g39840	lipoxygenase	Enzyme	925	102,82	6,26
13	LOC_Os03g18130	asparagine synthetase	Enzyme	605	67,32	6,51
14	LOC_Os03g20420	alpha-N-arabinofuranosidase A	Enzyme	676	74,01	4,69
15	LOC_Os01g19800	zinc finger, C3HC4 type	TF	221	24,67	7,11
16	LOC_Os03g09170	ethylene-responsive transcription factor	TF	298	31,79	6,51
17	LOC_Os02g43330	homeobox associated leucine zipper	TF	262	28,54	4,61
18	LOC_Os09g28210	bHelix-loop-helix transcription factor	TF	237	25,35	6,81
19	LOC_Os10g25230	ZIM domain containing protein (JA)	TF	167	16,95	9,74
20	LOC_Os02g22160	DNA binding protein	TF	440	48,13	9,24
21	LOC_Os02g26430	WRKY42	TF	254	27,14	9,93
22	LOC_Os06g46740	early nodulin 20 precursor	FP	280	27,35	4,78
23	LOC_Os02g52010	phosphate-induced protein 1 conserved region domain containing protein	FP	329	34,18	8,91
24	LOC_Os05g39250	phosphatidylethanolamine-binding protein	FP	168	18,24	6,08
25	LOC_Os02g36450	transporter family protein	FP	519	55,06	8,72
26	LOC_Os01g50910	late embryogenesis abundant protein	FP	216	22,13	9,67
27	LOC_Os11g45990	von Willebrand factor type A domain containing protein	FP	634	67,71	6,44
28	LOC_Os03g57640	gibberellin receptor	FP	361	38,46	4,79
29	LOC_Os08g28710	receptor protein kinase CRINKLY4 precursor	FP	488	53,65	6,46
30	LOC_Os12g29400	GRAM domain containing protein	FP	299	31,56	6,80
31	LOC_Os01g13120	aquaporin protein	FP	252	25,05	6,79
32	LOC_Os09g26810	chlorophyll A-B binding protein	FP	265	28,90	6,24
33	LOC_Os10g05780	POT domain containing peptide transporter	FP	554	58,38	7,49
34	LOC_Os03g47280	VQ domain containing protein	RP	144	14,08	9,50
35	LOC_Os06g33970	VQ domain containing protein	RP	145	14,71	8,45
36	LOC_Os08g37390	Cyclin	RP	384	41,66	4,67
37	LOC_Os04g54320	-	-	164	18,05	5,37
38	LOC_Os07g39210	-	-	165	18,74	4,28
39	LOC_Os04g42240	-	-	102	11,26	12,04
40	LOC_Os03g51350	-	-	174	18,05	6,98
41	LOC_Os10g23240	-	-	70	7,79	10,11

TT	Mã định danh locus	Chú giải chức năng	Vai trò	L	mW	pI
42	LOC_Os05g44060	-	-	181	18,00	6,80
43	LOC_Os05g49940	-	-	141	14,02	11,00
44	LOC_Os06g04240	-	-	91	10,18	7,37
45	LOC_Os12g32610	-	-	209	20,89	12,15
46	LOC_Os12g12880	-	-	216	23,04	5,58
47	LOC_Os02g40330	retrotransposon protein	-	115	12,78	8,21
48	LOC_Os06g45184	retrotransposon protein	-	165	18,24	10,94
49	LOC_Os01g39136	retrotransposon protein	-	111	11,85	11,25
50	LOC_Os05g49320	ribosomal protein L7/L12 C-terminal domain containing protein	-	180	18,21	6,28
51	LOC_Os09g20000	heavy metal-associated domain containing protein	-	161	17,65	10,09

Ghi chú: - : Chưa xác định; TF: Nhân tố phiên mã; RP: Protein điều hòa; FP: Protein chức năng; L: Kích thước protein; mW: Trọng lượng phân tử; pI: Điểm đẳng điện.

Báo cáo trước đây đã chứng minh rằng, TF cũng là các nhóm RP quan trọng điều hòa sự biểu hiện gene, tham gia vào cơ chế đáp ứng với tác nhân gây bệnh ở thực vật nói riêng và cây lúa nói chung [31]. Trong nghiên cứu này, các DEG mã hóa một số nhóm TF đã được xác định có biểu hiện thay đổi mạnh mẽ trong điều kiện lây nhiễm bệnh bạc lá (Bảng 1). Ví dụ, LOC\_Os01g19800 mã hóa cho zinc finger (C3HC4 type), là một trong những nhóm TF điển hình ở thực vật, được chứng minh liên quan đến đáp ứng bất lợi thông qua các con đường tín hiệu hormone [32]. Điển hình, trong số 67 thành viên của nhóm zinc finger ở cây lúa, hoạt hóa sự biểu hiện của gene *C3H12* giúp tăng cường tính kháng chủng vi khuẩn bạc lá PXO61 trên giống lúa Minghui 63 thông qua con đường tín hiệu phụ thuộc vào jasmonic acid [32]. Tiếp theo, LOC\_Os03g09170 mã hóa cho ethylene-responsive transcription factor (Bảng 1), là một trong những nhóm TF đặc trưng ở thực vật, đóng vai trò rất quan trọng trong đáp ứng điều kiện bất thuận và tác nhân gây bệnh [33]. Nhóm TF này có thể cảm ứng với nhiều con đường tín hiệu hormone, điển hình như jasmonic acid, ethylene, salicylic acid và abscisic acid, trong khi các thành viên trong nhóm này đã được sử dụng để tạo ra một số dòng lúa chuyển gene có khả năng kháng bạc lá và đạo ôn [33]. Kết quả cũng ghi nhận một số DEG mã hóa cho các nhóm TF quan trọng khác, như LOC\_Os02g43330 mã hóa cho homeobox associated leucine zipper, LOC\_Os09g28210

mã hóa bHelix-loop-helix transcription factor, LOC\_Os10g25230 mã hóa cho ZIM domain containing protein, LOC\_Os02g26430 mã hóa cho WRKY42 (Bảng 1). Đây đều là những TF quan trọng, đều cảm ứng với con đường tín hiệu jasmonic acid [31]. Bên cạnh đó, LOC\_Os02g22160 mã hóa cho DNA binding protein, chưa được chú giải chức năng một cách rõ ràng.

### 3.4. Phân tích đặc tính của protein mã hóa bởi DEG phản ứng mạnh với điều kiện lây nhiễm bệnh bạc lá

Phân tích đặc tính cơ bản của protein có thể gợi mở về chức năng tiềm năng của gene trong sinh trưởng và phát triển của cây trồng [9, 10, 20-23]. Theo đó, ba đặc tính lý hóa, bao gồm kích thước protein, trọng lượng phân tử và giá trị điểm đẳng điện của protein mã hóa bởi các DEG được phân tích trên công cụ ExPasy ProtParam [19]. Kết quả phân tích được thể hiện ở Bảng 1.

Cụ thể, các protein mã hóa bởi DEG có kích thước dao động từ 70 (protein chưa rõ chức năng mã hóa bởi LOC\_Os10g23240) đến 925 gốc amino acid (lipoxygenase mã hóa bởi LOC\_Os08g39840), với giá trị trung bình đạt khoảng 326 gốc amino acid. Tương đồng với kích thước, trọng lượng của các protein dao động từ 7,79 đến 102,82 kDa, với giá trị trung bình đạt khoảng 35,24 kDa. Tiếp theo, 29 (trên tổng số 51) protein mã hóa bởi DEG có tính acid, với giá trị điểm đẳng điện < 7, dao động từ 4,28 (protein chưa rõ chức năng mã hóa bởi



LOC\_Os07g39210) đến 6,98 (protein chưa rõ chức năng mã hóa bởi LOC\_Os03g51350). Hai mươi một (trên tổng số 51) protein mã hóa bởi DEG còn lại thể hiện tính base, với giá trị điểm

đẳng điện > 7, đạt từ 7,23 (isocitrate lyase mã hóa bởi LOC\_Os07g34520) đến 12,15 (protein chưa rõ chức năng mã hóa bởi LOC\_Os12g32610).

**Bảng 2. Vị trí cư trú nội bào của 51 protein mã hóa bởi DEG có mức độ biểu hiện thay đổi rất mạnh thu thập từ hai dữ liệu microarray liên quan đến lây nhiễm bệnh bạc lá**

TT	Protein	Vị trí phân bố nội bào	Xác suất của phép dự đoán (%)	Độ tin cậy của phép dự đoán
1	LOC_Os01g50400	Ty thể	52,22	Trung bình (0,42)
2	LOC_Os01g50420	Ty thể	79,42	Cao (0,92)
3	LOC_Os04g09900	Tế bào chất	99,51	Trung bình (0,66)
4	LOC_Os06g36560	Tế bào chất	99,77	Rất cao (0,95)
5	LOC_Os06g13560	Tế bào chất	94,58	Cao (0,95)
6	LOC_Os04g25490	Peroxisome	49,65	Thấp (0,05)
7	LOC_Os07g34520	Peroxisome	91,78	Rất cao (0,98)
8	LOC_Os08g02030	Tế bào chất	36,32	Thấp (0,07)
9	LOC_Os08g39730	Màng lưới nội chất	97,48	Trung bình (0,49)
10	LOC_Os10g25420	Thể Golgi	65,90	Trung bình (0,64)
11	LOC_Os11g05380	Thể Golgi	54,48	Thấp (0,04)
12	LOC_Os08g39840	Tế bào chất	72,00	Cao (0,89)
13	LOC_Os03g18130	Tế bào chất	99,91	Cao (0,82)
14	LOC_Os03g20420	Thể Golgi	81,10	Cao (0,83)
15	LOC_Os01g19800	Tế bào chất	78,23	Thấp (0,29)
16	LOC_Os03g09170	Nhân tế bào	99,99	Cao (0,81)
17	LOC_Os02g43330	Nhân tế bào	100,00	Rất cao (1,00)
18	LOC_Os09g28210	Nhân tế bào	100,00	Rất cao (0,98)
19	LOC_Os10g25230	Tế bào chất	68,15	Trung bình (0,65)
20	LOC_Os02g22160	Thể Golgi	60,61	Trung bình (0,40)
21	LOC_Os02g26430	Nhân tế bào	99,98	Rất cao (1,00)
22	LOC_Os06g46740	Gian bào	56,10	Thấp (0,29)
23	LOC_Os02g52010	Không bào	46,18	Thấp (0,10)
24	LOC_Os05g39250	Tế bào chất	81,43	Cao (0,92)
25	LOC_Os02g36450	Màng sinh chất	99,46	Trung bình (0,34)
26	LOC_Os01g50910	Nhân tế bào	80,38	Rất cao (1,00)
27	LOC_Os11g45990	Không bào	53,45	Cao (0,93)
28	LOC_Os03g57640	Peroxisome	70,18	Trung bình (0,36)
29	LOC_Os08g28710	Tế bào chất	63,84	Thấp (0,01)
30	LOC_Os12g29400	Nhân tế bào	79,99	Cao (0,87)
31	LOC_Os01g13120	Màng sinh chất	92,67	Thấp (0,13)
32	LOC_Os09g26810	Lục lạp	100,00	Rất cao (1,00)
33	LOC_Os10g05780	Peroxisome	55,36	Thấp (0,02)
34	LOC_Os03g47280	Nhân tế bào	60,78	Thấp (0,03)
35	LOC_Os06g33970	Nhân tế bào	67,01	Trung bình (0,46)
36	LOC_Os08g37390	Nhân tế bào	99,88	Rất cao (0,99)
37	LOC_Os04g54320	Ty thể	95,04	Cao (0,90)
38	LOC_Os07g39210	Nhân tế bào	70,78	Rất cao (0,99)
39	LOC_Os04g42240	Nhân tế bào	48,61	Trung bình (0,69)
40	LOC_Os03g51350	Nhân tế bào	41,59	Thấp (0,22)
41	LOC_Os10g23240	Gian bào	96,52	Thấp (0,02)
42	LOC_Os05g44060	Lục lạp	62,77	Thấp (0,02)
43	LOC_Os05g49940	Lục lạp	99,77	Thấp (0,19)

TT	Protein	Vị trí phân bố nội bào	Xác suất của phép dự đoán (%)	Độ tin cậy của phép dự đoán
44	LOC_Os06g04240	Tế bào chất	66,16	Trung bình (0,33)
45	LOC_Os12g32610	Ty thể	98,30	Thấp (0,13)
46	LOC_Os12g12880	Nhân tế bào	99,46	Rất cao (0,98)
47	LOC_Os02g40330	Nhân tế bào	92,93	Cao (0,92)
48	LOC_Os06g45184	Nhân tế bào	88,67	Rất cao (0,97)
49	LOC_Os01g39136	Ty thể	69,30	Thấp (0,23)
50	LOC_Os05g49320	Lục lạp	99,27	Rất cao (1,00)
51	LOC_Os09g20000	Tế bào chất	99,49	Thấp (0,04)

Để xác định vị trí cư trú nội bào của protein mã hóa bởi DEG, công cụ Yloc [24] được sử dụng để phân tích trình tự protein đầy đủ nhằm rà soát các đoạn trình tự peptide tín hiệu đặc trưng. Kết quả mô tả ở Bảng 2 cho thấy, protein mã hóa bởi DEG có thể cư trú ở rải rác tại các bào quan quan trọng trong tế bào. Cụ thể, protein mã hóa bởi 4 DEG, bao gồm LOC\_Os05g44060, LOC\_Os05g49940, LOC\_Os05g49320 và LOC\_Os09g26810 cư trú tại lục lạp, với xác suất của phép dự đoán đạt từ 62,77 đến 100%. Tiếp theo, 12 protein có thể phân bố trong tế bào chất, trong đó, độ tin cậy của phép dự đoán vị trí cư trú nội bào của LOC\_Os05g39250, LOC\_Os06g13560, LOC\_Os06g36560 và LOC\_Os08g39840 được ghi nhận là mạnh - rất mạnh. Một protein khác, LOC\_Os08g39730 có thể cư trú tại mạng lưới nội chất, trong khi hai protein, LOC\_Os06g46740 và LOC\_Os10g23240 nằm ở khoảng gian bào. Nghiên cứu cũng đã chỉ ra 4 protein, bao gồm LOC\_Os02g22160, LOC\_Os03g20420, LOC\_Os10g25420 và LOC\_Os11g05380 nằm ở thể Golgi, và 5 protein, bao gồm LOC\_Os01g39136, LOC\_Os01g50400, LOC\_Os01g50420, LOC\_Os12g32610 và LOC\_Os04g54320 cư trú tại ty thể. Bên cạnh đó, 4 protein, như LOC\_Os03g57640, LOC\_Os07g34520, LOC\_Os04g25490 và LOC\_Os10g05780 cư trú tại peroxisome, trong khi 2 và 2 protein, LOC\_Os02g36450 và LOC\_Os01g13120, và LOC\_Os02g52010 và LOC\_Os11g45990 được ghi nhận phân bố tại màng sinh chất và không bào. 15 protein còn lại được báo cáo nằm trong nhân tế bào.

#### 4. KẾT LUẬN

Nghiên cứu đã xác định được 895 gene được tăng cường biểu hiện và 735 gene bị kìm hãm

biểu hiện từ 2 dữ liệu microarray liên quan đến lây nhiễm bệnh bạc lá ở cây lúa. Trong đó, 51 DEG đã được xác định là có mức biểu hiện thay đổi rất mạnh với giá trị  $|\text{fold-change}| \geq 10$ .

Khai thác dữ liệu biểu hiện cho thấy các DEG có biểu hiện khác nhau tại 10 bộ phận trên cây lúa trong điều kiện thường. Một số DEG có biểu hiện đặc thù tại các cơ quan sinh sản sinh thực, bao gồm phôi 20 ngày sau thụ phấn, nội nhũ 25 ngày sau thụ phấn, cụm hoa trước khi xuất hiện, cụm hoa sau khi xuất hiện, bao phấn và nhụy hoa, trong khi có 6 DEG có xu hướng biểu hiện đặc thù tại thân.

Chú giải chức năng gene cho thấy 51 DEG mã hóa cho 14 enzyme, 7 TF, 3 RP và 12 FP khác và 14 protein chưa rõ chức năng. Nhiều nhóm enzyme đóng vai trò điều hòa âm tính trong đáp ứng với tác nhân gây bệnh ở cây lúa gạo, trong khi các nhóm TF được ghi nhận cảm ứng với con đường tín hiệu jasmonic acid, ethylene, salicylic acid và abscisic acid.

Phân tích đặc tính cơ bản cho thấy các protein mã hóa bởi DEG thể hiện tính đa dạng về kích thước, trọng lượng phân tử và giá trị điểm đẳng điện. Kích thước và trọng lượng phân tử của protein đạt từ 70 đến 925 gốc amino acid, tương đương 7,79 đến 102,82 kDa, trong khi giá trị điểm đẳng điện nằm trong khoảng từ acid (4,28) đến base (12,15). Nghiên cứu đã xác định được phần lớn protein phân bố trong tế bào chất, nhân tế bào và một số bào quan chính, như thể Golgi, ty thể, lục lạp và peroxisome, trong khi các protein còn lại nằm rải rác tại màng sinh chất, không bào, mạng lưới nội chất, khoảng gian bào.

#### Lời cảm ơn

Công trình này được hỗ trợ bởi Học viện Nông nghiệp Việt Nam thông qua Đề tài khoa học và công nghệ có mã số T2023-12-09TD.

**TÀI LIỆU THAM KHẢO**

- [1]. Tran Dang Khanh, Vu Xuan Duong, Phi Cong Nguyen, Tran Dang Xuan, Nguyen Thanh Trung, Khuat Huu Trung, Dong Huy Gioi, Nguyen Huy Hoang, Hoang Dung Tran, Do Minh Trung & Bui Thi Thu Huong (2021). Rice breeding in Vietnam: Retrospects, challenges and prospects. *Agriculture*. 11(5): 397.
- [2]. Kang H., Sridhar V., Mainuddin M. & Le Duc Trung (2021). Future rice farming threatened by drought in the Lower Mekong Basin. *Scientific Reports*. 11(1): 9383.
- [3]. Verma V., Ravindran P. & Kumar P. P. (2016). Plant hormone-mediated regulation of stress responses. *BMC Plant Biology*. 16: 86.
- [4]. Zhang H. & Wang S. (2013). Rice versus *Xanthomonas oryzae* pv. *oryzae*: a unique pathosystem. *Current Opinion in Plant Biology*. 16(2): 188-95.
- [5]. Cohen S. P., Liu H., Argueso C. T., Pereira A., Vera Cruz C., Verdier V. & Leach J. E. (2017). RNA-Seq analysis reveals insight into enhanced rice Xa7-mediated bacterial blight resistance at high temperature. *PLoS One*. 12(11): e0187625.
- [6]. Abdul Fiyaz R., Shivani D., Chaithanya K., Mounika K., Chiranjeevi M., Laha G. S., Viraktamath B. C., Subba Rao L. V. & Sundaram R. M. (2022). Genetic improvement of rice for bacterial blight resistance: Present status and future prospects. *Rice Science*. 29(2): 118-132.
- [7]. Shimono M., Koga H., Akagi A., Hayashi N., Goto S., Sawada M., Kurihara T., Matsushita A., Sugano S., Jiang C. J., Kaku H., Inoue H. & Takatsuji H. (2012). Rice WRKY45 plays important roles in fungal and bacterial disease resistance. *Molecular Plant Pathology*. 13(1): 83-94.
- [8]. Xu-Ming W., Jie Z., Yong Y., Fei-Bo Y., Juan C., Chu-Lang Y., Fang W., Ye C., Cheng-Qi Y. & Jian-Ping C. (2013). Transcriptome analysis of a progeny of somatic hybrids of cultivated rice (*Oryza sativa* L.) and wild rice (*Oryza meyeriana* L.) with high resistance to bacterial blight. *Journal of Phytopathology*. 161(5): 324-334.
- [9]. Tống Văn Hải, Nguyễn Quốc Trung, Cao Việt Bách, Lê Thị Ngọc Quỳnh, La Việt Hồng, Hà Thị Quyên, Phạm Minh Triển & Chu Đức Hà (2021). Phân tích đặc điểm của các gene cảm ứng tín hiệu auxin đáp ứng mạnh với điều kiện hạn và nóng ở cà chua (*Solanum lycopersicum*). *Kỷ yếu Hội nghị công nghệ sinh học toàn quốc 2021*. 570-574.
- [10]. Nguyễn Quốc Trung, Tống Văn Hải, Trịnh Thị Lam Hồng, La Việt Hồng, Phan Thị Thu Hiền, Trần Văn Tiến & Chu Đức Hà (2022). Phân tích nhóm gen chính đáp ứng với stress hạn và mặn ở cây đậu gà (*Cicer arietinum*) bằng phân tích dữ liệu giải mã hệ phiên mã. *Tạp chí Khoa học và Công nghệ - Đại học Thái Nguyên*. 227(05): 163-170
- [11]. 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 Res*. 41(Database issue): D991-D995.
- [12]. Tanaka T., Antonio B. A., Kikuchi S., Matsumoto T., Nagamura Y., Numa H., Sakai H., Sato Y., Souvorov A., Smith-White B., Tatusova T., An S., An G., Ota O., Fuks G., Fuks G., Lu C., Meyers B. C., Chaparro C., Piegu B., Panaud O. & Echeverria M. (2008). The Rice Annotation Project Database (RAP-DB): 2008 update. *Nucleic Acids Research*. 36(Database issue): D1028-D1033.
- [13]. Li J. Y., Wang J. & Zeigler R. S. (2014). The 3,000 rice genomes project: new opportunities and challenges for future rice research. *Gigascience*. 3: 8.
- [14]. 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(Database issue): D1178-D1186.
- [15]. Tran Tuan Tu, Perez-Quintero A. L., I. Wonni, S. C. D. Carpenter, Y. Yu, L. Wang, J. E. Leach, V. Verdier, S. Cunnac, A. J. Bogdanove, R. Koebnik, M. Hutin & B. Szurek (2018). Functional analysis of African *Xanthomonas oryzae* pv. *oryzae* TALomes reveals a new susceptibility gene in bacterial leaf blight of rice. *PLoS Pathog*. 14(6): e1007092.
- [16]. 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*. 11(4): e0154315.
- [17]. Leinonen R., Sugawara H. & Shumway M. (2011). The sequence read archive. *Nucleic Acids Res*. 39(Database issue): D19-D21.
- [18]. 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(D1): D427-D432.
- [19]. 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*. 31(13): 3784-3788.
- [20]. La Viet Hong, Chu Duc Ha, Tran Duy Cuong, Nguyen Huu Kien, Le Thi Ngoc Quynh, Hoang Minh Chinh, Cao Phi Bang, Pham Cong Tuyen Anh, 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.

- [21]. Le Thi Man, Tran Thi Thanh Huyen, Vu Xuan Quyen, Chu Duc Ha, Pham Chau Thuy, Le Thi Ha, Le Thi Ngoc Quynh, La Viet Hong & Cao Phi Bang (2022). Genome-wide identification and analysis of genes encoding putative Heat shock protein 70 in papaya (*Carica papaya*). Pakistan Journal of Biological Sciences. 25(6): 468-475.
- [22]. La Viet Hong (2022). Genome-wide identification and analysis of Heat shock Protein 70 family in *Theobroma cacao*. Pakistan Journal of Biological Sciences. 25(7): 608-618.
- [23]. Tran Duy Cuong, Chu Duc Ha, Nguyen Huu Kien, Yasuko Watanabe, La Viet Hong, Tran Dang Khanh & Tran Phan Lam-Son (2018). Genome-wide identification of the TCP transcription factor family in chickpea (*Cicer arietinum* L.) and their transcriptional responses to dehydration and exogenous abscisic acid treatments. Journal of Plant Growth Regulation. 37(4): 1286-1299.
- [24]. Briesemeister S., Rahnenfuhrer J. & Kohlbacher O. (2010). YLoc-an interpretable web server for predicting subcellular localization. Nucleic Acids Research. 38(Web Server issue): W497-502.
- [25]. Chen J., Wang L., Yang Z., Liu H., Chu C., Zhang Z., Zhang Q., Li X., Xiao J., Wang S. & Yuan M. (2021). The rice Raf-like MAPKKK OsILA1 confers broad-spectrum resistance to bacterial blight by suppressing the OsMAPKK4-OsMAPK6 cascade. Journal of Integrative Plant Biology. 63(10): 1815-1842.
- [26]. Ricardo C., Conchi S., Nieves V. & Jesús M. (2022). Interactions of gibberellins with phytohormones and their role in stress responses. Horticulturae. 8(3): 241.
- [27]. Li X., Liu W., Li B., Liu G., Wei Y., He C. & Shi H. (2018). Identification and functional analysis of cassava DELLA proteins in plant disease resistance against cassava bacterial blight. Plant Physiology Biochemistry. 124: 70-76.
- [28]. Dong-Lei Y., Qun L., Yi-Wen D., Yong-Gen L., Mu-Yang W., Guo-Xing Z., Ying-Ying Z. & Zu-Hua H. (2008). Altered disease development in the eui mutants and Eui overexpressors indicates that gibberellins negatively regulate rice basal disease resistance. Molecular Plant. 1(3): 528-537.
- [29]. Wang A., Ma L., Shu X., Jiang Y., Liang J. & Zheng A. (2022). Rice (*Oryza sativa* L.) cytochrome P450 protein 716A subfamily CYP716A16 regulates disease resistance. BMC Genomics. 23(1): 343.
- [30]. Gao M., Yin X., Yang W., Lam S. M., Tong X., Liu J., Wang X., Li Q., Shui G. & He Z. (2017). GDSL lipases modulate immunity through lipid homeostasis in rice. PLoS Pathogen. 13(11): e1006724.
- [31]. Helliwell E. E. & Yang Y. (2013). Molecular strategies to improve rice disease resistance. Methods in Molecular Biology. 956: 285-309.
- [32]. Deng H., Liu H., Li X., Xiao J. & Wang S. (2012). A CCCH-type zinc finger nucleic acid-binding protein quantitatively confers resistance against rice bacterial blight disease. Plant Physiology. 158(2): 876-89.
- [33]. Zhou Y., Xu S., Jiang N., Zhao X., Bai Z., Liu J., Yao W., Tang Q., Xiao G., Wang K., Hu X., Tan J. & Yang Y. (2022). Engineering of rice varieties with enhanced resistances to both blast and bacterial blight diseases via CRISPR/Cas9. Plant Biotechnol Journal. 20(5): 876-885.