BỘ GIÁO DỤC
VÀ ĐÀO TẠO

VIỆN HÀN LÂM
KHOA HỌC VÀ CÔNG NGHỆ VN

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

Nguyễn Thị Thủy Tiên

NGHIÊN CỨU PHÂN TÍCH PHÁT SINH LOÀI CỦA MỘT SỐ LOÀI VI KHUẨN THUỘC CHI *Bacillus* BẰNG KỸ THUẬT MULTILOCUS SEQUENCING ANALYSIS (MLSA)

Chuyên ngành: Sinh học thực nghiệm
Mã số: 8420114

**LUẬN VĂN THẠC SĨ NGÀNH
SINH HỌC THỰC NGHIỆM**

NGƯỜI HƯỚNG DẪN KHOA HỌC:
Ts. Lê Thị Huỳnh Trâm

Thành phố Hồ Chí Minh - Năm 2023

LỜI CAM ĐOAN

Tôi xin cam đoan đề tài nghiên cứu trong luận văn này là công trình nghiên cứu của tôi dựa trên những tài liệu, số liệu do chính tôi tự tìm hiểu và nghiên cứu. Chính vì vậy, các kết quả nghiên cứu đảm bảo trung thực và khách quan nhất. Đồng thời, kết quả này chưa từng xuất hiện trong bất cứ một nghiên cứu nào. Các số liệu, kết quả nêu trong luận văn là trung thực nếu sai tôi hoàn chịu trách nhiệm.

Nguyễn Thị Thủy Tiên

LỜI CẢM ƠN

Tôi xin gửi lời cảm ơn đến Tiến sĩ Lê Thị Huỳnh Trâm, người đã hướng dẫn tận tình để tôi có thể hoàn thành nghiên cứu này.

Tôi xin gửi lời cảm ơn đến những người đã giúp đỡ tôi trong quá trình thực hiện nghiên cứu này gồm có Thạc sĩ Đạo Nữ Diệu Hồng, Cử Nhân Trang Hoàng Long.

Tôi cũng xin gửi lời cảm ơn đến tập thể phòng Công nghệ Vi sinh đã hỗ trợ tôi rất nhiều trong quá trình thực hiện Luận văn. Cám ơn ban lãnh đạo Trung tâm Công nghệ Sinh học, đã tạo điều kiện cho tôi vừa công tác vừa có thể theo học chương trình Thạc sĩ.

Và cuối cùng tôi xin gửi lời cảm ơn chân thành đến phòng Đào tạo, các phòng chức năng của Học viện Khoa học và Công nghệ để luận văn được hoàn thành.

MỤC LỤC

MỞ ĐẦU **1**

NỘI DUNG **4**

Chương 1. TỔNG QUAN NGHIÊN CỨU **4**

1. Phương pháp phát sinh loài 4

1.1 Phân tích phát sinh loài dựa vào trình tự phân tử 4

1.2 Xây dựng cây phát sinh loài 5

1.2.1 Cây khoảng cách 6

1.2.2 Cây Likelihood 6

1.2.3 Cây Parsimony 7

2. Chi *Bacillus* 7

3. Multilocus Sequencing Analysing (MLSA) 9

2.1 So sánh với các phương pháp phân loại khác 11

2.2 Tình hình nghiên cứu 12

4. Housekeeping gene – Gen giữ nhà 12

2.1 Gen *16S rRNA* 13

2.2 Gen *rpoD* 14

2.3 Gen *glpF* 15

2.1 Gen *pta* 15

2.2 Gen *pycA* 16

2.3 Gen *purH* 16

Chương 2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU **18**

2.1. Đối tượng nghiên cứu 18

2.2. Phương pháp nghiên cứu 20

2.2.1 Nuôi cấy 20

2.2.2 Tách chiết DNA tổng số của các chủng vi khuẩn: 20

2.2.3 Khảo sát nhiệt độ bắt cặp của các cặp mồi *rpoD, glpF, ptA, pycA và purH* 21

2.2.4 Giải trình tự 6 vùng gen *16S, rpoD, glpF, ptA, pycA và purH* 22

2.2.5 Phân tích trình tự 23

2.2.6 Phân tích phát sinh loài 23

Chương 3. KẾT QUẢ VÀ THẢO LUẬN **25**

3.1. Nuôi cấy 25

3.3. Giải trình tự và phân tích trình tự 6 vùng gen *16S rRNA, rpoD, glpF, pta, pycA* và *purH* 30

3.4 Phân tích phát sinh loài 33

3.4.1 Phân tích phát sinh loài từ trình tự các vùng gen *16S rRNA, rpoD, glpF, ptA, pycA và purH* 33

3.4.2 Khảo sát số lượng gen 42

KẾT LUẬN VÀ KIẾN NGHỊ **56**

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ CÁI VIẾT TẮT

Ký hiệu chữ viết tắt	Chữ viết đầy đủ
HGT	Horizontal gene transfer
MLSA	Multilocus Sequencing Analysis
MLST	Multilocus Sequencing Typing
PCR	Polymerase Chain Reaction
NJ	Neighbour-joining
ML	Maximum Likelihood
MP	Maximum Parsimony
MUSCLE	MUltiple Sequence Comparison by Log Expectation
IPS	Information Parsimony Site
HBCM	HCMBiotech Collection of Microorganisms
TSA	Tryptone Soya Agar
TSB	Tryptone Soya Broth

DANH MỤC CÁC BẢNG

Bảng 2. 1: Ký hiệu các chủng được sử dụng trong nghiên cứu này 18

Bảng 2. 2: Số GenBank accession của các chủng tham chiếu 18

Bảng 2. 3: Thông tin các cặp mồi được sử dụng 19

Bảng 2. 4: Gradient nhiệt độ bắt cặp 21

Bảng 3. 1: Kết quả hình thái đại thể, vi thể của các chủng mục tiêu 25

Bảng 3. 2: Nhiệt độ bắt cặp của các cặp mồi dùng để chạy PCR 28

Bảng 3. 3: Các đặc điểm của các gen và các gen ghép nối 32

Bảng 3. 4 Giá trị khoảng cách trình tự (%) giữa các loài khi thay đổi số lượng gen phân tích MLSA. 54

DANH MỤC CÁC HÌNH VẼ, ĐỒ THỊ

Hình 1. 1: Tế bào vi khuẩn *Bacillus* sp dưới kính hiển vi quang học 7

Hình 1. 2 Phân loại khoa học của chi *Bacillus*.....8

Hình 1. 3: Phương pháp Phân tích trình tự đa gen MLSA.....10

Hình 1. 4 Các tính chất của gen giữ nhà 13

Hình 1. 5 Sơ đồ phức hợp ribosome và gen *16S rRNA*. 14

Hình 1. 6 Sơ đồ của con đường acetate kinase (AckA)-phosphotransacetylase (*Pta*) chuyển hóa xen kẽ acetyl-CoA và acetate. Công thức phân tử của acetyl-P.....16

Hình 1. 7 Gen *pycA* và *pycB* mã hóa hai tiểu đơn vị của pyruvate carboxylase.....16

Hình 1. 8 Con đường Chuyển hóa formyl và phản ứng IMP cyclohydrolase xúc tác bởi ATIC.....17

Hình 1. 9 Cấu trúc gen *purHJ* mã hóa enzyme ATIC 17

Hình 2. 1 Quy trình nghiên cứu.....20

Hình 3. 1: Kết quả điện di sản phẩm PCR xác định nhiệt độ bắt cặp tối ưu của các cặp mồi cho phản ứng PCR khuếch đại các gen *purH, pycA, rpoD, pta, glpF*. 27

Hình 3. 2: Kết quả điện di sản phẩm PCR khuếch đại gen *16S rRNA* của 8 chủng *Bacillus* spp.....29

Hình 3. 3: Kết quả điện di sản phẩm PCR khuếch đại các gen *rpoD, glpF, pta, pycA* và *purH* của 8 chủng *Bacillus* spp.. 29

Hình 3. 4: Kết quả giải trình tự của 8 chủng mục tiêu với 6 gen *16S rRNA, rpoD, glpF, pta, pycA* và *purH*.....30

Hình 3. 5 Trình tự được căn chỉnh, sắp xếp thẳng hàng bằng thuật toán MUSCLE của phần mềm MEGA11 31

Hình 3. 6: Cây phát sinh loài từ trình tự gen *16S rRNA* 35

Hình 3. 7: cây phát sinh loài từ trình tự gen *glpF*.....37

Hình 3. 8: Cây phát sinh loài từ trình tự gen *rpoD*.....38

Hình 3. 9: Cây phát sinh loài từ trình tự gen *pta*. 39

Hình 3. 10: Cây phát sinh loài từ trình tự gen *purH*.....40

Hình 3. 11: cây phát sinh loài từ trình tự gen *pycA*. 41

Hình 3. 12: Cây phát sinh loài từ trình tự 2 gen *16S rRNA-rpoD*. 43

Hình 3. 13: Cây phát sinh loài từ trình tự 3 gen *16S rRNA-rpoD-pta*. 45

Hình 3. 14: Cây phát sinh loài từ trình tự 4 gen *16S rRNA-rpoD-pta-purH*.. 47
Hình 3. 15: Cây phát sinh loài từ trình tự 5 gen *16S rRNA-rpoD-pta-purH-glpF*.. 49
Hình 3. 16: cây phát sinh loài từ trình tự 6 gen *16S rRNA-rpoD-pta-purH-glpF-pycA*.. 51

MỞ ĐẦU

- **Lý do chọn đề tài:**

Chi *Bacillus* là một nhóm các vi khuẩn hiếu khí hình que, Gram dương, có khả năng tạo bào tử và có khả năng lên men. Tính đến năm 2020, chi *Bacillus* bao gồm 396 loài có tên được công bố hợp lệ (*http://www.bacterio.net/bacillus.html*). Các loài thuộc chi *Bacillus* được sử dụng rộng rãi trong y tế, dược phẩm, nông nghiệp và công nghiệp, có khả năng sản xuất một loạt các enzyme (protease, amylase và cellulase), kháng sinh (bacitracin, polymyxin), và các chất chuyển hóa khác. Trong đó, có nhiều loài đang được sử dụng rộng rãi trong sản xuất và đời sống như *B. subtilis, B. amyloliquefaciens, B. licheniformis*... Có 3 phương pháp chính để định danh vi sinh vật là dựa vào đặc điểm hình thái, sinh hóa và đặc biệt là phương pháp sinh học phân tử được xem là phương pháp chính xác nhất. Phương pháp sử dụng phổ biến nhất là định danh dựa trên trình tự gen *16S rRNA*, tuy nhiên phương pháp này không cung cấp đủ độ phân biệt giữa các loài có quan gần gũi, cho ra kết quả khác nhau và có sự nhầm lẫn giữa các loài như *B. subtilis* và *B. amyloliquefaciens* hay *B. safensis, B. amyloliquefaciens, B. cereus* và *B. lichenformis*. Ngược lại, phương pháp lai DNA-DNA và DNA-fingerprinting có thể được sử dụng để phân biệt giữa các loài có quan hệ họ hàng gần, nhưng cả hai phương pháp này đều tốn nhiều công sức, không mang lại dữ liệu tích lũy và khó áp dụng cho số lượng lớn các chủng.

Với các vấn đề nêu trên, cần một phương pháp chính xác và hiệu quả hơn để phân biệt các loài trong chi *Bacillus* và MLSA là phương pháp phù hợp để giải quyết các vấn đề trên, cung cấp dữ liệu chất lượng cao, có thể tích lũy để thiết lập các mối quan hệ trong và giữa các loài có quan hệ họ hàng gần. Mặc dù phương pháp MLSA đã được sử dụng rộng rãi để phân tích đa dạng di truyền của các loài có mối quan hệ gần gũi, tuy nhiên việc chọn số lượng gen sử dụng cho phương pháp MLSA vẫn chưa có các khuyến nghị chung, trái ngược với các khuyến nghị được cung cấp cho nghiên cứu phát sinh loài dựa trên gen *16S rRNA.* Do đó việc khảo sát số lượng gen ghép nối để xây dựng các cây phát sinh loài đã được phân tích để so sánh mức độ phân biệt các loài khi thay đổi số lượng gen sử dụng trong phương pháp MLSA nhằm đem lại cái nhìn tổng quan hơn về việc lựa chọn số lượng gen để đạt được sự cân bằng giữa thời gian, chi phí và mức độ phân biệt chấp nhận được.

- **Mục đích nghiên cứu:**

Xây dựng phương pháp định danh các loài thuộc chi *Bacillus* bằng kỹ thuật Multilocus Sequencing Analysis (MLSA) nhằm phân loại đến loài các vi khuẩn thuộc chi *Bacillus* thuộc bộ sưu tập giống vi sinh vật HBCM

- **Nội dung nghiên cứu:**
  - Giải trình tự và phân tích trình tự các gen *16S rRNA, rpoD, glpF, ptA, pycA và purH* của các chủng *Bacillus* spp.
  - Phân tích phát sinh loài các chủng *Bacillus* spp. từ trình tự các vùng gen *16S rRNA, rpoD, glpF, ptA, pycA và purH*
  - Phân tích phát sinh loài bằng phương pháp MLSA dựa trên các trình tự ghép nối từ các gen *16S rRNA, rpoD, glpF, ptA, pycA và purH.*
- **Cơ sở khoa học và tính thực tiễn của đề tài:**

Phương pháp MLSA lần đầu tiên được áp dụng bởi Brady và cộng sự [1] để phân tích phát sinh loài của chi *Pantoea.* Sau đó, phương pháp MLSA đã được áp dụng để hỗ trợ đề xuất cho một số loài *Pantoea* mới và các loài *Pantoea* khác đã được phân loại lại *Pantoea citrea*, *Pantoea punchtata* và *Pantoea terrea* lần lượt được chuyển sang các chi *Tatumella* thành *Tatumella citrea*, *Tatumella punctata* và *Tatumella terrea*[2]. Việc áp dụng phương pháp MLSA đối với các loài thuộc chi *Enterobacter* dẫn đến việc phân loại lại 11 loài *Enterobacter* thành ba chi mới được đề xuất, *Lelliottia*, *Pluralibacter* và *Kosakonia*[3].

Phương pháp này đã được sử dụng rộng rãi trong việc phân loại và xác định đa dạng vi khuẩn, để xác định mức độ trao đổi gen trong và giữa các loài và tương đối quan trọng để thiết lập việc tái tổ hợp trong di truyền học quần thể. Phương pháp MLSA cho thấy tiềm năng trong việc đánh giá mối quan hệ của các đơn vị phân loại vi khuẩn bằng cách sử dụng các mô hình biến đổi di truyền [4]. Như Hossein và cộng sự năm 2022 đã phân lập các loài *Mycobacterium*, vi khuẩn gây ra bệnh phổi lâm sàn bằng phương pháp MLSA để nối sáu gen *rpoB, ssrA, tuf, atpE, ku* và *dnaK* [5]. Phương pháp MLSA còn được sử dụng để phân biệt giữa các chủng thuộc nhóm *Bacillus cereus* [6], [7], [8]. Các nghiên cứu trước đây đã ghi nhận thành công phương pháp MLSA phân biệt các loài thuộc nhóm *B. pumilus* có liên kết chặt chẽ trong môi trường biển bằng cách sử dụng bảy gen giữ nhà (*gyrB, rpoB, pyrA, pyrE, aroE, mutL* và *trpB*) [9].

Trong một bài đánh giá được viết vào năm 2009, Schleifer đã tiếp thu gợi ý của Gevers và cộng sự năm 2005 và cũng đề cập đến MLSA như một phương pháp có tiềm năng thay thế phương pháp lai DNA-DNA (DDH) để phân định loài. Tác giả tuyên bố rằng cây phát sinh loài dựa trên các trình tự (bằng phương pháp MLSA) có thể được sử dụng để làm sáng tỏ các cụm phân nhánh sâu. Quan điểm của ông là phương pháp MLSA nên được áp dụng như một phương pháp bổ sung để xác định

kiểu gen nhóm trong một chi hoặc loài [10]. Năm 2010, Tindall và cộng sự cũng nhấn mạnh tiềm năng của phương pháp MLSA dựa trên các gen mã hóa protein, thường là gen giữ nhà, để bổ sung cho các phân tích DDH và *16S rRNA* để phân tích phân loại ở cấp độ loài [11]. Phương pháp MLSA được đề xuất sử dụng kết hợp một số gen giữ nhà để đánh giá lại định nghĩa loài trong vi khuẩn học, phương pháp MLSA đã làm mới về hệ thống sinh vật nhân sơ và phát sinh loài. Các mối quan hệ dựa trên trình tự gen *16S rRNA* cung cấp một khuôn khổ vô giá cho các nghiên cứu phân tích MLSA nhưng không cung cấp đủ độ phân biệt giữa các loài có quan hệ họ hàng gần [12]. Ngược lại, phương pháp lai DNA-DNA (DDH) và DNA-fingerprinting có thể được sử dụng để phân biệt giữa các loài có quan hệ họ hàng gần, nhưng cả hai phương pháp này đều tốn nhiều công sức, không mang lại dữ liệu tích lũy và khó áp dụng cho số lượng lớn các chủng. Phương pháp MLSA cung cấp một giải pháp thay thế cho DDH bằng cách cung cấp dữ liệu chất lượng cao, có thể tái tạo để thiết lập các mối quan hệ trong và giữa các loài [4].

- **Những đóng góp của luận văn**

Phương pháp định danh các loài thuộc chi *Bacillus* bằng phương pháp phân tích phát sinh loài dựa trên trình tự ghép nối của 6 đoạn gen *16S rRNA, rpoD, glpF, ptA, pycA và purH* và số lượng gen ghép nối cần thiết cho các mức độ phân loại khác nhau.

NỘI DUNG

Chương 1. TỔNG QUAN NGHIÊN CỨU

1. Phương pháp phát sinh loài

Phân tích phát sinh loài là thiết lập mối quan hệ tiến hóa giữa các gen hoặc đặc điểm của sinh vật. Và mặc dù sự sống có một nguồn gốc duy nhất và chỉ có một lịch sử tiến hóa thực sự, nhiệm vụ tái tạo lại lịch sử này có thể từ dễ dàng đến khó khăn tùy thuộc vào các đơn vị phân loại cụ thể hoặc các sinh vật được phân tích. Những khó khăn phát sinh với việc tái tạo phát sinh loài do thực tế là hầu như tất cả các suy luận về sự tiến hóa và các mối quan hệ đều dựa trên nghiên cứu về các sinh vật hiện tại. Hiếm khi có thể lấp đầy những thiếu sót để xác minh thông tin về sự sống thực sự diễn ra như thế nào. Những vấn đề này được phóng đại ở vi khuẩn vì chúng có ít đặc điểm có thể nhìn thấy và hầu như không có hồ sơ hóa thạch. Hơn nữa, vi khuẩn thì cổ xưa và đã đa dạng hóa qua hàng tỷ năm và là đối tượng của các quá trình chuyển gen ngang (Horizontal gene transfer - HGT), có khả năng tạo ra bất kỳ đặc điểm nào trong bất kỳ dòng dõi nào và che khuất các mô hình tổ tiên truyền thống của Darwin, do đó làm cho lịch sử của chúng thậm chí còn khó xác định hơn [13].

Trong hàng trăm năm đầu tiên của ngành vi sinh học, vi khuẩn được phân loại, phân biệt và nhóm lại chủ yếu theo môi trường sống, đặc điểm sinh trưởng, thuộc tính sinh hóa và tiềm năng độc lực của chúng. Nhưng trong nửa sau của thế kỷ 20, đã có một sự thay đổi rõ rệt đối với việc sử dụng thông tin di truyền phân tử, thu được thông qua phân tích trình tự protein và axit nucleic. Điều này cung cấp ba lợi thế ngay lập tức.

- Đầu tiên, các phân tích so sánh trở nên tập trung vào các ký tự được phân phối phổ biến, giúp loại bỏ các dòng được phân loại cùng nhau do thiếu các đặc điểm xác định cụ thể của chúng.
- Thứ hai, kiến thức chuyên sâu về các cơ chế vật lý mà theo đó các đại phân tử thông tin có thể thay đổi đã dẫn đến sự phát triển của các mô hình mạnh mẽ về tiến hóa trình tự.
- Thứ ba, việc sử dụng các trình tự phân tử đã làm tăng đáng kể số lượng các ký tự rời rạc, theo đó các vi sinh vật có thể được so sánh với mỗi nucleotide hoặc axit amin có thể đóng vai trò là một đặc điểm thông tin [13].

1.1 Phân tích phát sinh loài dựa vào trình tự phân tử

Việc sử dụng dữ liệu trình tự phân tử để suy ra các mối quan hệ là trọng tâm về quá trình tiến hóa. Những thay đổi ở cấp độ phân tử đã được sử dụng để thiết lập phát sinh loài của nhiều loại sinh vật. Đối với vi khuẩn, đặc biệt, nơi các dấu hiệu kiểu hình cổ điển thường không hiệu quả hoặc không thực tế để nhóm các phân lập vi khuẩn, các phương pháp phát sinh loài phân tử đã mang tính cách mạng [13].

- Sắp xếp trình tự

Điều kiện tiên quyết để tạo ra một kiểu phát sinh loài của sinh vật dựa trên trình tự phân tử là sự sắp xếp các trình tự trực giao (di truyền theo chiều dọc) gần đúng nhất với quá trình tiến hóa phân tử thực sự của chúng. Sự sắp xếp có thể được tạo bằng nhiều thuật toán, hầu hết trong số đó sử dụng một số dạng ma trận thay thế để định vị các nucleotide tương đồng hoặc axit amin để tối đa hóa số lượng vị trí giống hệt hoặc tương tự tại một vị trí nhất định. Clustal, hiện là phần mềm căn chỉnh phổ biến nhất, sử dụng ma trận IUB và ClustalW1.6 để sắp xếp DNA và ma trận PAM, BLOSUM và Gonnet để sắp xếp axit amin [13].

Khi một tập dữ liệu trình tự được đưa vào chương trình căn chỉnh, chẳng hạn như Clustal, thuật toán bắt đầu bằng cách tạo sự căn chỉnh theo cặp của tất cả các trình tự một cách độc lập. Sau đó, thuật toán sẽ tính toán tất cả khoảng cách giữa các cặp trình tự, tạo ra một ma trận khoảng cách mà sau đó được chuyển thành biểu đồ cây (dendrogram). Như tên gọi của nó, cây này phục vụ như một hướng dẫn để bắt đầu căn chỉnh nhiều trình tự, với các trình tự giống nhau nhất được thêm vào trước, tiếp theo là bổ sung dần dần các trình tự khác nhau hơn. Loại phương pháp căn chỉnh trình tự lũy tiến theo cặp đơn giản và tương đối hiệu quả này đã được sử dụng rộng rãi trong nhiều nghiên cứu phát sinh gen, nhưng kém hiệu quả hơn đối với các bộ dữ liệu cụ thể. Điều này đã thúc đẩy sự phát triển của các thuật toán căn chỉnh cải tiến hơn, chẳng hạn như MUSCLE, T-Coffee, POA, DIALIGN, SAGA và MAFFT, hiệu quả và chính xác trong việc khôi phục các căn chỉnh tối ưu từ các bộ dữ liệu khác nhau, được thiết lập bằng cách so sánh với BAliBASE (cơ sở dữ liệu sắp xếp tham chiếu được tinh chỉnh thủ công). Tuy nhiên, không phải tất cả các thuật toán này đều tương đương về thuật toán, chẳng hạn như MUSCLE, T-Coffee và MAFFT, thực hiện một loạt các phép toán lũy tiến, căn chỉnh và tối ưu hóa trong mỗi lần lặp lại để đảm bảo độ chính xác căn chỉnh cao hơn đáng kể. Sự liên kết tiến bộ, tinh tế được cung cấp bởi các phương pháp này là điều cần thiết để tái tạo chính xác các mối quan hệ phát sinh gen.

Bất kể chương trình hoặc thuật toán được sử dụng để tạo liên kết nhiều trình tự, mỗi liên kết phải được đánh giá trước khi sử dụng để tái tạo phát sinh gen. Biến thể trình tự trên một căn chỉnh có thể không đồng nhất, dẫn đến các vùng được bảo tồn nằm xen kẽ giữa các vùng biến đổi cao. Do sự liên kết của các vùng có thể siêu biến thường là vấn đề nên các phần này được che giấu hoặc loại bỏ tốt nhất. Điều này đặc biệt quan trọng bởi vì ngay cả những thay đổi nhỏ trong sự liên kết cũng sẽ bị nhiễu thông tin có khả năng che khuất sự phát sinh loài thực sự [13].

1.2 Xây dựng cây phát sinh loài

Khi một sự liên kết đáng tin cậy được tạo ra, nhiều phương pháp xây dựng cây có thể được sử dụng để chuyển đổi dữ liệu liên kết thành một cây phát sinh loài. Các

phương pháp xây dựng cây này được phân loại rộng rãi thành các phương pháp tiếp cận dựa trên khoảng cách, phân tích và xác suất [13].

1.2.1 Cây khoảng cách

Các cách tiếp cận dựa trên khoảng cách phân loại đại diện cụm dựa trên số lượng thay thế nucleotide hoặc axit amin giữa các trình tự. Một trong những phương pháp dựa trên khoảng cách đầu tiên và đơn giản nhất, được phát triển để tạo ra các bản sao kiểu hình, là UPGMA (Phương pháp nhóm cặp không trọng số với trung bình số học). UPGMA sử dụng phương pháp phân cụm liên tiếp, theo đó một ma trận của tất cả các điểm tương đồng của trình tự theo cặp được tạo ra và hai đơn vị phân loại có khoảng cách nhỏ nhất được phân cụm trước. Ma trận khoảng cách được tính toán lại, coi các đơn vị phân loại đã được nhóm thành một đơn vị phân loại và đơn vị phân loại có khoảng cách nhỏ nhất tiếp theo được thêm vào cây. Điều này được lặp lại cho đến khi tất cả các đơn vị phân loại được thêm vào và một bộ phát sinh loài cuối cùng được hình thành. Do cách giải thích đơn giản này về ma trận khoảng cách và giả định về tốc độ tiến hóa không đổi (đồng hồ phân tử) cho tất cả các trình tự, UPGMA có xu hướng xây dựng các cây phát sinh loài không được hỗ trợ (unsupported phylograms).

Phương pháp Neighbor-Joining (NJ)[14], có cách tiếp cận tương tự để phân nhóm theo phân loại như UPGMA, trước tiên phân cụm các chuỗi (láng giềng) gần nhất và sau đó tính toán lại khoảng cách giữa các cặp lân cận. Do đó, NJ không gặp phải những hạn chế giống như UPGMA và đã trở thành phương pháp dựa trên khoảng cách được ưa thích để tái tạo phát sinh gen[13].

1.2.2 Cây Likelihood

Việc áp dụng các phương pháp likelihood để tái cấu trúc phát sinh loài đã trở nên ngày càng phổ biến, phần lớn là do độ chính xác và tính nhất quán cao hơn một chút của chúng trong việc khôi phục một phát sinh loài chính xác, và do sự gia tăng đáng kể về khả năng và tốc độ tính toán. Maximum Likelihood (ML) [15] và phương pháp tiếp cận Bayesian [16] đưa ra hai cách tiếp cận xác suất riêng biệt nhưng có liên quan để xác định kiểu phát sinh loài tốt nhất. Các phương pháp ML cố gắng xác định cấu trúc liên kết cây có xác suất cao nhất với dữ liệu chuỗi được cung cấp. ML cần nhiều tính toán và không phù hợp với các tập dữ liệu rất lớn [13].

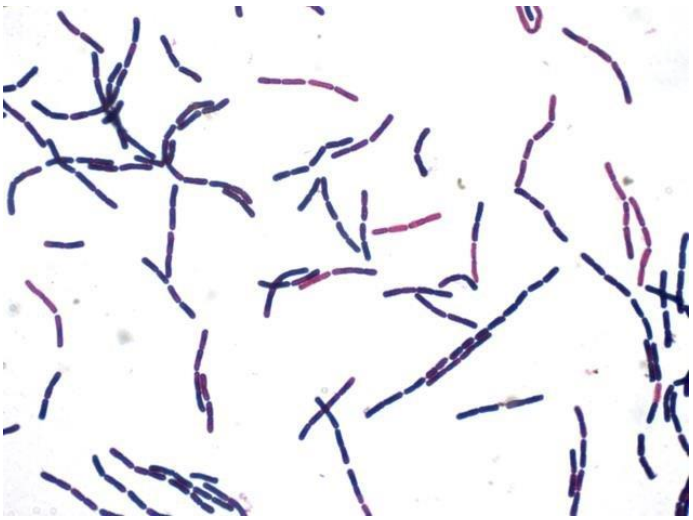
Phương pháp Bayesian để tái cấu trúc phát sinh loài được xem là một giải pháp thay thế lý tưởng cho phương pháp ML. Không giống như cách tiếp cận dựa trên khoảng cách, cách tiếp cận Bayes không tạo ra một cây duy nhất, nó lấy mẫu một loạt các cấu trúc liên kết cây có thể xảy ra dựa trên bộ dữ liệu và các phân bố xác suất trước (Prior probability). Bởi vì các xác suất sau đã được tính toán trong quá trình lấy mẫu, chúng được sử dụng để xác định giá trị độ tin cậy tại mỗi nút trên cây, không cần hỗ trợ thống kê thêm các giá trị như bootstrap hoặc jackknife [13].

1.2.3 Cây Parsimony

Các phương pháp phân tích để xây dựng cây phát sinh loài được đặt tiền đề để ủng hộ con đường tiến hóa với số lượng thay đổi ít nhất. Không giống như các phương pháp dựa trên khoảng cách, tính phân tích dựa trên suy luận dựa trên ký tự của các kiểu phát sinh loài, sử dụng dữ liệu trình tự ở trạng thái ban đầu thay vì chuyển đổi nó thành khoảng cách. Một trong những phương pháp phân tích phổ biến hơn là Maximum Parsimony (MP)[17], một phương pháp tính toán các bước tiến hóa cho tất cả các cấu trúc liên kết có khả năng và trình bày cây có các bước tiến hóa ít nhất. Thông thường, có nhiều cây chi tiêu ngang nhau, một số có cấu trúc liên kết khác hẳn so với những cây khác. Giống như cách tiếp cận Bayesian, một kiểu phát sinh loài thống nhất được xây dựng trên tập hợp cây này, do đó kết hợp tất cả các cấu trúc liên kết thành một cấu trúc liên kết cây gần đúng duy nhất [13].

2. Chi *Bacillus*

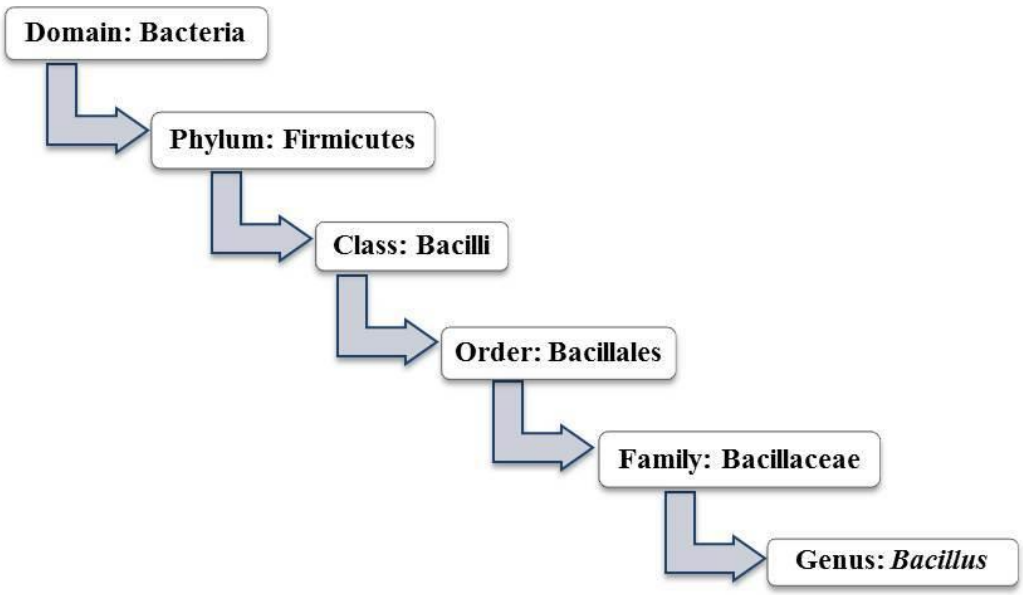
Các loài vi khuẩn thuộc chi *Bacillus* là vi khuẩn gram dương hình que (hình 1.1), hình thành nội bào tử hoặc kỵ khí tùy ý; ở một số loài có thể chuyển sang Gram âm theo thời gian nuôi cấy. Nhiều loài thuộc chi thể hiện nhiều khả năng sinh lý cho phép chúng sống trong mọi môi trường tự nhiên [18].



Hình 1. 1: Tế bào vi khuẩn *Bacillus* sp dưới kính hiển vi quang học [19]

Phân loại khoa học của chi *Bacillus* được thể hiện trong hình 1.2. Chi vi khuẩn *Bacillus* có lịch sử lâu dài và phong phú trong biên niên sử về vi sinh vật học. Việc sử dụng *Bacillus* để thúc đẩy tăng trưởng thực vật là một lĩnh vực có tiềm năng lớn cho nông nghiệp. Nhờ khả năng tạo nội bào tử thông qua một quá trình phát triển nguyên thủy nhưng phức tạp, chi *Bacillus* đã được nghiên cứu chuyên sâu trong lĩnh vực học thuật. Ở Châu Âu (trừ Vương quốc Anh), chế phẩm sinh học *Bacillus* được sử dụng để phòng ngừa các bệnh về đường tiêu hóa. Tầm quan trọng của *Bacillus* trong lĩnh vực y học đã được khẳng định vào cuối những năm 1800 khi Louis Pasteur

và A. Koch xác định *Bacillus anthracis* là tác nhân gây bệnh than. Thực vật là nguồn tài nguyên phong phú của các loài *Bacillus* mới, một số loài nội sinh và những loài khác có liên quan đến vùng rễ [20]. Nhiều đặc tính sinh lý và các chất chuyển hóa chuyên biệt của các loài *Bacillus* đã được sử dụng trong ngành dược phẩm, nông nghiệp và thực phẩm. Mặt khác, sức chống chịu của các bào tử đối với khử trùng và tiệt trùng rất mạnh khiến chúng trở thành chất gây ô nhiễm trong thực phẩm, vật tư y tế, quy trình phẫu thuật, v.v…[18]



Hình 1. 2 Phân loại khoa học của chi *Bacillus* [21]

Tính đến năm 2020, chi *Bacillus* bao gồm 396 loài có tên được công bố hợp lệ (*http://www.bacterio.net/bacillus.html*). Có 3 phương pháp chính để định danh vi sinh vật là dựa vào đặc điểm hình thái, sinh hóa và đặc biệt là phương pháp sinh học phân tử được xem là phương pháp chính xác nhất.

Các phương pháp sinh học phân tử thường được sử dụng trong phân loại chi *Bacillus* là:

- Giải trình tự DNA: dựa vào trình tự vùng gen như *16S rRNA, 23S rRNA* để định danh chi *Bacillus*, các vùng gen như *groEL, gyrB, recN, rpoB, spoIIA* để định danh tới loài và gen hag để định danh tới dưới loài; hoặc dựa vào các vùng gen được ghép nối tạo thành trình tự đa gen - phương pháp MLSA có thể phân loại đến dưới loài.
- DNA fingerprinting: gồm các kỹ thuật nghiên cứu tính đa hình chiều dài của các phân đoạn DNA dựa trên điểm cắt các enzyme giới hạn (RFLP), điện di trường xung đẩy (Pulsed Field Gel Electrophoresis - PFGE), Kỹ thuật dấu vân tay di truyền Rep-PCR (Repetitive DNA PCR fingerprinting), kỹ thuật khuếch

đại ngẫu nhiên DNA đa đình và kỹ thuật nhận dạng dấu vân tay dựa vào Oligonucleotide cho microarray (Oligonucleotide microarray fingerprinting) được sử dụng để phân loại đến dưới loài.

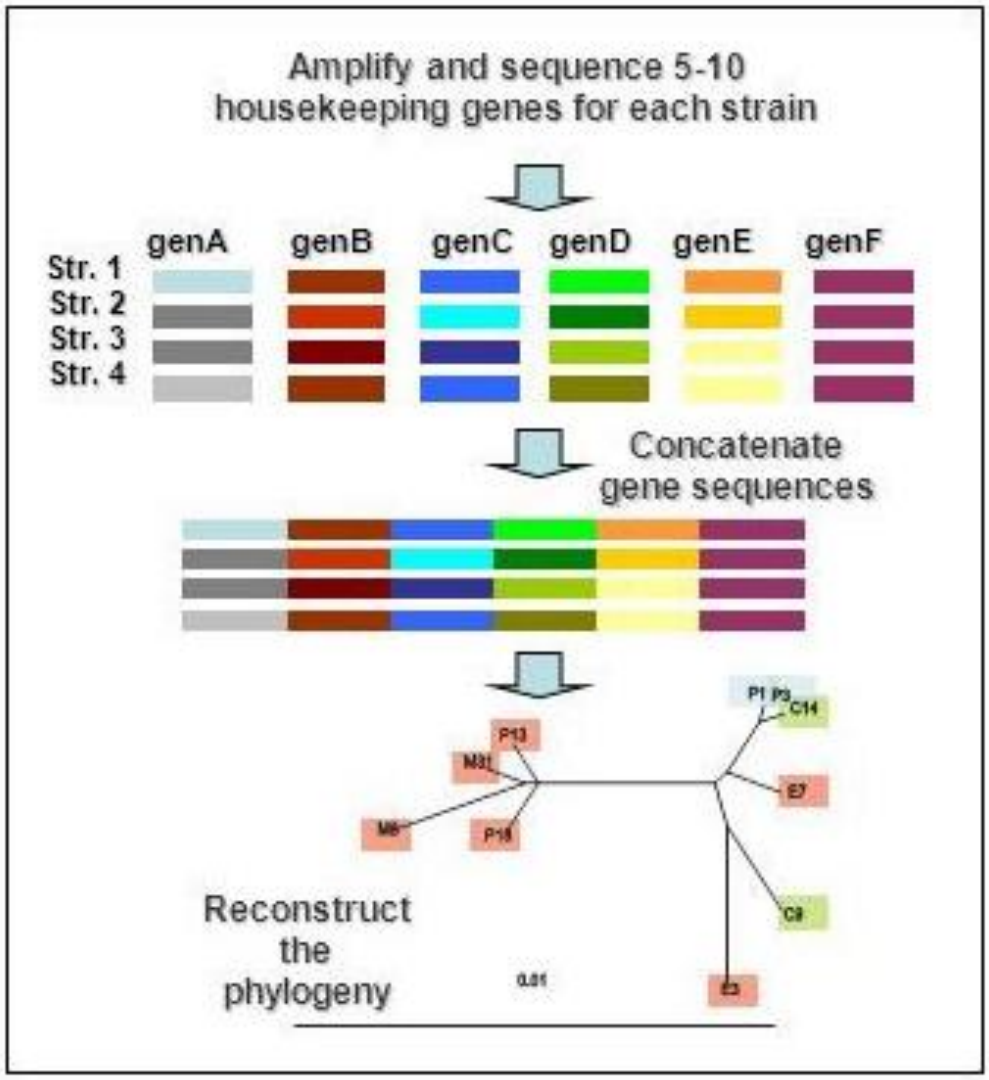
- Một số phương pháp khác cũng được sử dụng để định danh tới dưới loài trong chi *Bacillus* như phân tích các acid béo (Fatty acid profiling), quang phổ khối MALDI-TOF (MALDI-TOF mass spectroscopy of spore proteins), Điện di enzyme đa locus (Multilocus Enzyme Electrophoresis - MLEE). [20]

Mặc dù có nhiều phương pháp mang lại kết quả phân loại đến dưới loài trong chi *Bacillus* nhưng phương pháp dựa vào trình tự DNA là được sử dụng phổ biến nhất vì có nhiều ưu điểm như về chi phí và đòi hỏi kỹ thuật, thiết bị không quá phức tạp. Trong đó, trình tự gen *16S rRNA* được xem như là dữ liệu nền tảng trong việc phân loại vi khuẩn [22], như đã nêu ở trên các loài thuộc chi *Bacillus* có mối quan hệ gần gũi nên nếu chỉ dựa vào trình tự *16S rRNA* thì chưa đủ để phân biệt các loài trong chi. Từ đó có thể thấy phương pháp định danh dựa vào trình tự DNA của các gen ghép nối (MLSA) thích hợp cho việc phân biệt đến loài trong chi *Bacillus.*

3. Multilocus Sequencing Analysing (MLSA)

Multilocus sequence analysis/typing (MLSA/MLST) là một cách tiếp cận dựa trên trình tự nucleotide để mô tả đặc tính rõ ràng của sinh vật nhân sơ thông qua Internet, mô tả trực tiếp các biến thể trình tự DNA trong một bộ gen giữ nhà và đánh giá mối quan hệ giữa các chủng dựa trên cấu hình hoặc trình tự allelic độc đáo của chúng [23]. Phương pháp này đã được sử dụng rộng rãi trong việc phân loại và xác định đa dạng vi khuẩn, để xác định mức độ trao đổi gen trong và giữa các loài và tương đối quan trọng để thiết lập việc tái tổ hợp trong di truyền học quần thể. MLSA đang cung cấp cơ hội mới trong việc đánh giá mối quan hệ của các đơn vị phân loại vi khuẩn bằng cách sử dụng các mô hình biến đổi di truyền [4].

Phương pháp MLSA được phát triển từ việc áp dụng phương pháp Multilocus Sequencing Typing (MLST) để tái tạo lại các mối quan hệ tiến hóa giữa các sinh vật nhân sơ [24]. Phương pháp MLST so sánh dựa trên trình tự sắp xếp các đoạn 450–500 bp của 5-7 gen giữ nhà cung cấp thông tin về sự lan truyền của sự phân tán nucleotide trên các nhiễm sắc thể của quần thể được lấy mẫu. Các trình tự khác nhau dù chỉ một nucleotide cho mỗi gen được chỉ định là các alen khác nhau, do đó làm cho MLSA rất thích hợp để phát hiện những thay đổi di truyền trong và giữa các loài.



Hình 1. 3: Phương pháp Phân tích trình tự đa gen MLSA [25]

Phân tích phát sinh loài bằng phương pháp MLSA bao gồm các bước cơ bản sau:

- Bước 1: Lựa chọn các chủng và gen giữ nhà
- Bước 2: Tạo trình tự (khuếch đại phản ứng chuỗi polymerase (PCR) và giải trình tự DNA).
- Bước 3: Phân tích trình tự để xác định các vị trí tương đồng trong mỗi gen
- Bước 4: Sử dụng trình tự ghép nối.

Gen giữ nhà được chọn để phân tích MLSA phải là các trình tự duy nhất, orthologous (là các gen có liên quan đến sự hình thành loài) và phổ biến trong số tất cả các chủng được lấy mẫu. Chúng cũng cần được bảo tồn cao, không có sự mất cân bằng liên kết trên nhiễm sắc thể nhưng phải chứa đủ các vị trí nucleotide khác nhau để thiết lập chính xác mối quan hệ giữa các chủng liên quan chặt chẽ. Để đạt được sự cân bằng giữa sức mạnh nhận dạng chấp nhận được, thời gian và chi phí cho việc

phân tích chủng, khoảng 5-7 gen giữ nhà thường được sử dụng. Tuy nhiên, không có gì lạ khi sử dụng tới 10 gen giữ nhà, như được minh họa trong trường hợp chi *Nocardia* nơi 14 gen mã hóa protein đã được kiểm tra bởi Tamura và cộng sự năm 2012 [26]. Do đó, cả số lượng và loại gen giữ nhà được phân tích MLSA có thể khác nhau giữa các chi. PCR thường được sử dụng để tạo ra các đoạn trình tự. Hồ sơ bảo tồn của các gen mã hóa protein cung cấp các vùng được bảo tồn cao có thể được sử dụng để thiết kế các mồi khuếch đại và giải trình tự có tính đặc hiệu rộng đối với đa dạng phát sinh loài [4]. Bước phân tích dữ liệu, căn chỉnh trình tự kết hợp với cả việc kiểm tra chất lượng trước và sau của dữ liệu thô cho mỗi vị trí là điều kiện tiên quyết đối với phương pháp MLSA. Phân tích thống kê các trình tự của các đoạn gen được sử dụng trong các nghiên cứu MLSA, thống kê cho từng vị trí, chẳng hạn như số lượng và tỷ lệ các vị trí đa hình, hàm lượng G + C trung bình và chỉ số dN/dS (các tỷ lệ từ sự thay thế không đồng nghĩa đến hiện tượng đa hình), có thể được tóm tắt bằng cách sử dụng START2 (http://pubmlst.org/software/analysis/start2/), MEGA (http://www.megasoftware.net/) và DnaSP (http://www.ub.edu/dnasp/). Xây dựng cây phát sinh loài, sự tích lũy những thay đổi nucleotide trong gen là một quá trình tương đối chậm; do đó, các trình tự ghép nối của các dòng vi khuẩn phân lập đủ ổn định theo thời gian lý tưởng cho việc suy ra các mối quan hệ phát sinh loài. Để đảm bảo tính ổn định và độ tin cậy của các mối quan hệ phát sinh loài giữa các chủng dựa trên cách tiếp cận MLSA, cây phát sinh loài thường được xây dựng bằng cách sử dụng cả hai phương pháp dựa trên khoảng cách và đặc điểm, neighbour-joining (NJ) [14] là một trong những thuật toán được sử dụng thường xuyên nhất trong xây dựng cây phát sinh loài MLSA, vì nó có thể nhanh chóng được sử dụng để phản ánh khoảng cách theo từng cặp, đặc biệt là đối với các bộ dữ liệu có liên quan chặt chẽ. Cần lưu ý rằng các ước tính phát sinh loài có thể bị ảnh hưởng bởi độ dài nhánh, kích thước tập dữ liệu (cả số đơn vị phân loại và vị trí), tính không đồng nhất của trình tự, độ sâu tiến hóa, độ phức tạp của tập dữ liệu và khung phân tích. Ngay cả khi các cây phát sinh loài tối ưu được tạo ra thành công, chúng không phải lúc nào cũng cung cấp các mối quan hệ có ý nghĩa từ quan điểm sinh học. Điều này cho thấy chúng ta nên chú ý đến việc ước lượng phát sinh loài cũng như các kỹ thuật thuật toán để có được các giải pháp tối ưu [4].

2.1 So sánh với các phương pháp phân loại khác

Các mối quan hệ dựa trên dữ liệu trình tự gen *16S rRNA* cung cấp một khuôn khổ vô giá cho các nghiên cứu MLSA nhưng không đưa ra đủ giải pháp để phân biệt giữa các loài có liên quan chặt chẽ. Ngược lại, các phương pháp DNA-fingerprinting và phương pháp lai DNA-DNA (DDH) có thể được sử dụng để phân biệt giữa các loài

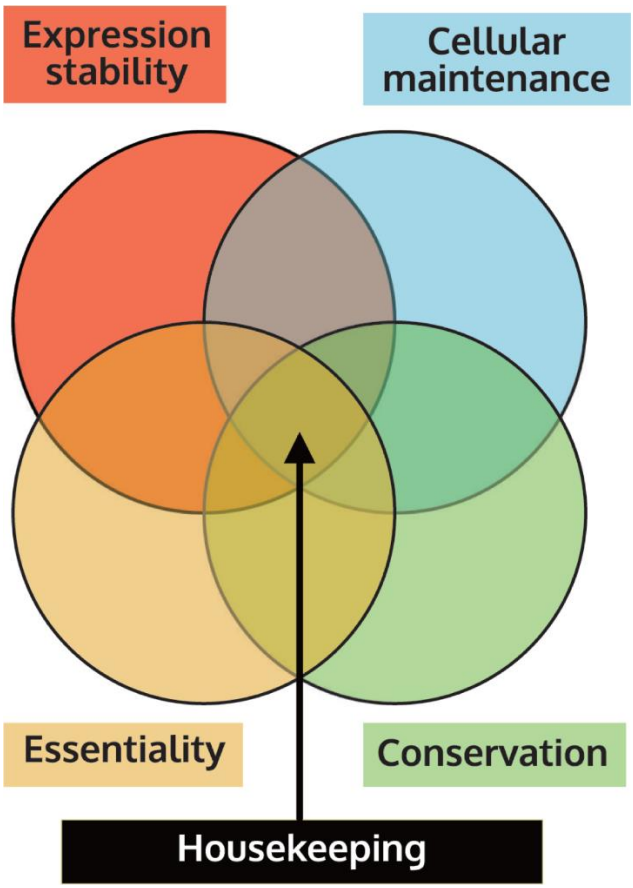
có quan hệ họ hàng gần, nhưng cả hai phương pháp này đều tốn nhiều công sức, không mang lại dữ liệu tích lũy và khó áp dụng cho số lượng lớn các chủng. MLSA cung cấp một giải pháp thay thế hấp dẫn cho DDH bằng cách cung cấp dữ liệu chất lượng cao, có thể tái tạo để thiết lập mối quan hệ trong và giữa các loài. Phương pháp MLSA cũng khác với DNA barcoding – mã vạch DNA, một phương pháp phân loại sử dụng các dấu hiệu di truyền ngắn để nhận ra các loài đã biết hoặc chưa biết và dựa trên DNA ty thể hoặc một số bộ phận của cistron DNA ribosome có tốc độ đột biến tương đối nhanh [4].

2.2 Tình hình nghiên cứu

Phương pháp MLSA đã được phát triển rộng rãi và được sử dụng để phân biệt giữa các chủng thuộc nhóm *Bacillus cereus* bởi Ko và cộng sự năm 2004[6], Priest và cộng sự năm 2004 [7], Soufiane và cộng sự năm 2013[8] . Các nghiên cứu trước đây đã ghi nhận thành công MLSA để phân biệt các nhóm thuộc nhóm *B. pumilus* có liên kết chặt chẽ trong môi trường biển bằng cách sử dụng bảy gen giữ nhà (*gyrB, rpoB, pyrA, pyrE, aroE, mutL* và *trpB*) [9]. Năm 2019, Lê Xuân Thế và công sự cũng đã sử dụng phương pháp MLSA để đánh giá đa dạng di truyền của *Bacillus* spp. từ các ao nuôi tôm công nghiệp ở Việt Nam bằng cách sử dụng bảy gen giữ nhà (*glpF, ilvD, ptA, purH, pycA, rpoD,* và *tpiA*) [27].

4. Housekeeping gene – Gen giữ nhà

Gen giữ nhà thường được định nghĩa là các gen có biểu hiện ổn định trong tất cả các tế bào và là điều kiện thiết yếu duy trì sự sống tế bào và được bảo tồn [28]. Khái niệm về gen giữ nhà đã hỗ trợ lý thuyết và ứng dụng sinh học bao gồm cả nghiên cứu về sự tiến hóa. Gen giữ nhà có thể được định nghĩa là tập hợp tối thiểu các gen cần thiết để duy trì sự sống [29]. Ở cấp độ thực tế, chúng có thể được định nghĩa là các gen biểu hiện ổn định trong tất cả các tế bào của một sinh vật bất kể loại mô, giai đoạn phát triển, trạng thái chu kỳ tế bào hoặc tín hiệu bên ngoài hoặc là dấu hiệu của trạng thái sinh học khỏe mạnh của sinh vật [30]. Ở cấp độ tiến hóa, chúng có thể cho phép chúng ta xác định loài và các đặc điểm bộ gen cụ thể của phân loại cao hơn và chức năng gen có thể thúc đẩy bảo tồn hoặc thay đổi. Do đó, kiến thức về gen giữ nhà có thể đóng góp đáng kể cho các nghiên cứu khám phá, cơ bản và mở rộng [28]. Bốn tính chất rất khác nhau của gen giữ nhà: sự ổn định biểu hiện (biểu hiện tương tự giữa các loại và điều kiện tế bào), chức năng (ví dụ: thuộc về xu hướng chính của tế bào), tính thiết yếu (mất chức năng này gây chết tế bào) và bảo tồn (trong trường hợp này, được biểu hiện ổn định và cần thiết trên các đơn vị phân loại) (Hình 1.3).



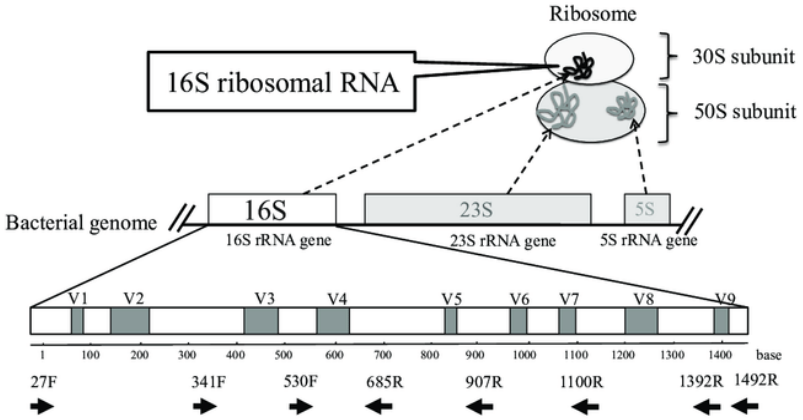
Hình 1. 4 Các tính chất của gen giữ nhà [28]

2.1 Gen *16S rRNA*

Cấu trúc ribosome và gen *16S rRNA* được thể hiện trong Hình. 1.3. Ribosome là một phức hợp protein và tiểu đơn vị RNA được tìm thấy trong tất cả các tế bào sống, đóng vai trò quan trọng trong tổng hợp protein sinh học (dịch mã). Ribosome bao gồm hai thành phần chính: tiểu đơn vị ribosome nhỏ (tiểu đơn vị ribosome 30S trong tế bào nhân sơ) và tiểu đơn vị lớn (tiểu đơn vị ribosome 50S trong tế bào nhân sơ). Mỗi tiểu đơn vị chứa một hoặc nhiều phân tử RNA ribosome (rRNA) và nhiều loại protein ribosome. Gen *16S rRNA* mã hóa một phân tử RNA ribosome của tiểu đơn vị ribosome 30S có trong tất cả các tế bào nhân sơ, bao gồm vi khuẩn và vi khuẩn cổ. 23S rRNA và 5S rRNA là các tiểu đơn vị rRNA chứa trong tiểu đơn vị ribosome 50S. Các gen mã hóa cho các thành phần của ribosome hầu hết được bảo tồn, có nghĩa là cấu trúc của chúng đã thay đổi rất ít theo thời gian do chức năng quan trọng của chúng, dịch mã mRNA thành protein. Việc phân loại ba miền (Eukarya, Vi khuẩn và Archaea) được đề xuất theo cây phát sinh loài dựa trên trình tự gen 16S *rRNA* [31].

Gen *16S rRNA* là một công cụ thường được sử dụng để xác định vi khuẩn vì nhiều lý do. Đầu tiên, gen này tương đối ngắn (khoảng 1.500 bp). Thứ hai, có mười

vùng trong trình tự gen *16S rRNA* phổ biến ở hầu hết các vi khuẩn (vùng được bảo tồn) và được tách thành chín vùng khác nhau (vùng siêu biến đổi) (Hình 1.4). Do đó, một số mồi phổ quát được thiết lập ở các khu vực được bảo tồn [32]. Thứ ba, các trình tự gen được đăng ký trong cơ sở dữ liệu công cộng đang tăng đáng kể, bởi vì trình tự gen là thông tin quan trọng để xác định và phân loại trong các nghiên cứu phân loại vi khuẩn [33].



Hình 1. 5 Sơ đồ phức hợp ribosome và gen *16S rRNA*.

Các mũi tên in đậm thể hiện vị trí gần đúng của các đoạn mồi phổ biến trên trình tự gen *16S rRNA* của *Escherichia coli*. □các vùng được bảo tồn, ■: các vùng siêu biến (V1-V9). [33]

2.2 Gen *rpoD*

RpoD (còn được gọi là sigma 70) là yếu tố sigma chính và được nghiên cứu kỹ lưỡng chịu trách nhiệm phiên mã các gen giữ nhà ở hầu hết các vi khuẩn [34]. Trong quá trình điều hòa biểu hiện gen của vi khuẩn, quá trình bắt đầu phiên mã, được trung gian bởi holoenzyme RNA polymerase (RNAP) phụ thuộc DNA, đóng vai trò quan trọng là bước đầu tiên trong quy trình điều hòa. Holoenzyme RNAP bao gồm một enzyme lõi (thành phần tiểu đơn vị α2ββ'ω) có hoạt tính xúc tác cho quá trình trùng hợp RNA và một tiểu đơn vị bổ sung được gọi là yếu tố sigma liên quan đến nhận dạng trình khởi động và làm tan chảy DNA [34]. Hầu hết các vi khuẩn có nhiều yếu tố sigma nhận ra các bộ khởi động khác nhau như là bộ điều chỉnh chính của phản ứng căng thẳng đối với những thay đổi môi trường và biểu hiện gen cơ bản. Các protein họ sigma 70, bao gồm *RpoD*, chứa bốn vùng được bảo tồn được chỉ định từ 1 đến 4 [35], [36]. Nói chung, hai miền liên kết DNA cư trú ở vùng 2 và 4 nhận ra các trình tự hexamer khởi đầu được bảo tồn (các yếu tố khởi đầu) xung quanh các vị trí ngược dòng tương ứng khoảng 10 và 35 nucleotide của các vị trí bắt đầu phiên mã (TSS) [34]. Các nghiên cứu trước đây đã cho thấy rằng ở nhiều vi khuẩn, *RpoD* tương

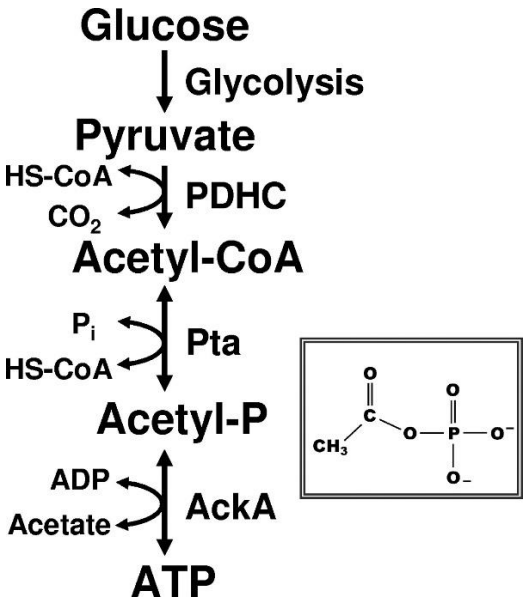
tác với hai yếu tố khởi đầu nguyên mẫu (–35 5′-TTGACA-3′ và –10 5′-TATAAT-3′ được phân tách bằng một miếng đệm khoảng 17 bp)[36].

2.3 Gen *glpF*

Gen Glycerol uptake facilitator (*glpF*) mã hóa cho protein Glycerol uptake facilitator được mô tả là trung gian khuếch tán glycerol qua màng tế bào chất thông qua cơ chế kiểu lỗ. Có khả năng thấm cao với glycerol, nhưng ít thấm nước hơn. Không vận chuyển ion. Nó cũng có thể có tính thấm hạn chế đối với nhiều chất nền khác bao gồm xylitol, erythritol, D-arabitol, L-arabitol, ribitol, D-galactitol, D-mannitol, D-sorbitol, urê, glycine, D/L-glyceraldehyde và hóa trị ba dạng vô cơ của asen và antimon. Tính chất đáng chú ý của tính dẫn nước hiệu quả kết hợp với việc loại trừ nghiêm ngặt tất cả các ion bao gồm cả proton được trung gian bởi hai asparagine được bảo tồn, buộc một phân tử nước trung tâm đóng vai trò là chất cho liên kết hydro với các phân tử nước lân cận của nó. Được hỗ trợ bởi điện thế tĩnh điện được tạo ra bởi hai vòng lặp kéo dài nửa màng, điều này quyết định hướng ngược lại của các phân tử nước trong hai nửa kênh, và do đó ngăn chặn sự hình thành 'dây proton', đồng thời cho phép nước khuếch tán nhanh chóng [37].

2.1 Gen *pta*

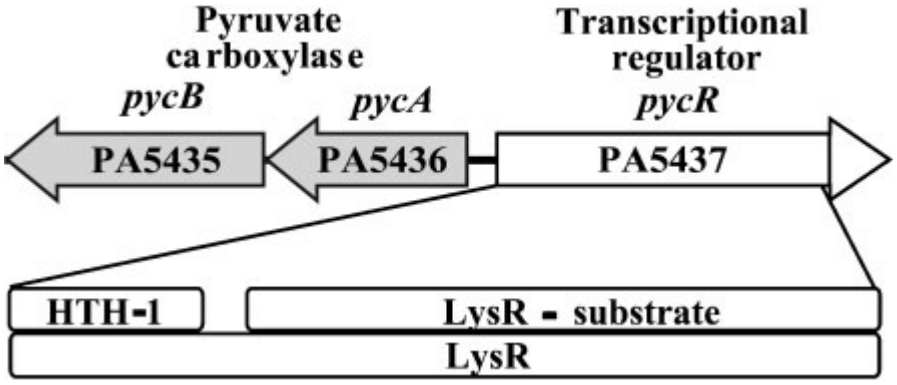
Gen *pta* mã hóa protein Phosphate acetyltransferase, là enzyme tham gia vào quá trình chuyển hóa acetate. Xúc tác cho quá trình chuyển hóa thuận nghịch giữa acetyl-CoA và acetyl phosphate. Hướng của phản ứng tổng thể thay đổi tùy thuộc vào điều kiện tăng trưởng. Trên môi trường tối thiểu acetyl-CoA được tạo ra. Trong môi trường giàu acetyl-CoA được chuyển thành acetate và cho phép tế bào loại bỏ lượng acetyl hóa dư thừa để đổi lấy năng lượng dưới dạng ATP. Con đường chính để sản xuất acetate trong giai đoạn lũy thừa. Hoạt tính xúc tác: acetyl-CoA + phosphat = acetyl phosphat + CoA. Phản ứng này tiến hành theo hướng thuận và ngược [38].



Hình 1. 6 Sơ đồ của con đường acetate kinase (AckA)-phosphotransacetylase (*Pta*) chuyển hóa xen kẽ acetyl-CoA và acetate. Công thức phân tử của acetyl-P [39].

2.2 Gen *pycA*

Gen *pycA* và *pycB* mã hóa hai tiểu đơn vị của pyruvate carboxylase, là enzyme xúc tác phản ứng 2 bước, bao gồm quá trình carboxyl hóa của biotin phụ thuộc ATP, liên kết cộng hóa trị trong bước đầu tiên và chuyển nhóm carboxyl thành pyruvate trong bước thứ hai, dẫn đến sản xuất oxaloacetate. Hoàn thành chức năng hình thành các chất trung gian của quá trình trao đổi chất ở *B. subtilis* vì nó cần thiết cho sự tăng trưởng trên glucose, nhưng không cần thiết cho sự hình thành bào tử. Hoạt động xúc tác: ATP + hydrocacbonat + pyruvate = ADP + H+ + oxaloacetate + phosphate.

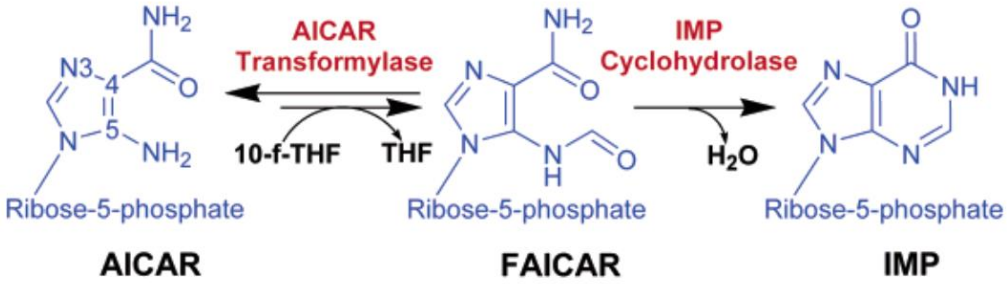


Hình 1. 7 Gen *pycA* và *pycB* mã hóa hai tiểu đơn vị của pyruvate carboxylase [40]

2.3 Gen *purH*

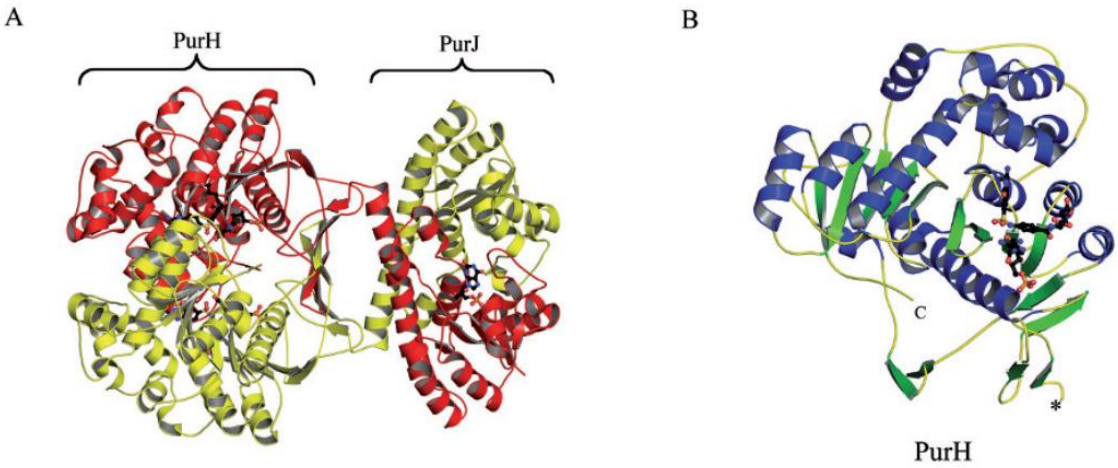
Gen *purH* mã hóa enzyme AICAR transformylase, một trong mười enzyme tạo ra inosine 5'-monophosphate trong con đường sinh tổng hợp purine [41]. Ở hầu hết các sinh vật, inosine 5'-monophosphate (IMP) được hình thành từ các tiền chất

phân tử nhỏ thông qua con đường sinh tổng hợp purine. Bước xúc tác áp chót của con đường tổng hợp purine de novo là chuyển đổi aminoimidazole-4-carboxamide ribonucleotide (AICAR) thành 5-formyl-AICAR đòi hỏi đồng yếu tố N10-formyl-tetrahydrofolate cung cấp formyl (Hình 1.7)[42].



Hình 1. 8 Con đường Chuyển hóa formyl và phản ứng IMP cyclohydrolase xúc tác bởi ATIC [42]

Phản ứng này được xúc tác bởi enzyme AICAR transformylase của enzyme lưỡng chức năng AICAR transformylase /inosine monophosphate cyclohydrolase (ATIC) được mã hóa bởi gen *purHJ* (hình 1.8).



Hình 1. 9 Cấu trúc gen *purHJ* mã hóa enzyme ATIC [41]

Chương 2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Đối tượng nghiên cứu

- Chủng giống vi sinh vật:

Các chủng vi khuẩn *Bacillus* spp sử dụng trong nghiên cứu thuộc bộ sưu tập giống vi sinh vật HBCM được phát triển bởi Trung tâm Công Nghệ Sinh Học Thành Phố Hồ Chí Minh.

Bảng 2. 1: Ký hiệu các chủng được sử dụng trong nghiên cứu này

STT	Mã số
1	BC-B0020
2	BC-B0027
3	BC-B0028
4	BC-B0029
5	BC-B0037
6	BC-B0039
7	BC-B0112
8	BC-B0116

- Các chủng tham chiếu được lấy từ ngân hàng dữ liệu Genbank NCBI, số GenBank accession của thể hiện trong bảng 2.2

Bảng 2. 2 Số GenBank accession của các chủng tham chiếu

STT	Loài	Chủng	Số GenBank accession	Ghi Chú
1	*B. amyloliquefaciens*	MBE1283	CP013727.1	
2		IT-45	CP004065.1	
3		S499	CP014700.1	
4		Y14	CP017953.1	
5		DH8030	CP041770.1	
6	*B. subtilis*	TL03	CP021169.1	
7		SRCM102754	CP028202.1	
8		MBI 600	CP033205.1	
9		ATCC 21228	CP020023.1	
10		ATCC 11774	CP026010.1	
11		SRCM103637	CP035414.1	
12		168	AL009126.3	
13		GOT9	CP047325.1	

14		CB0015	CP048852.1	
15	*B. velezensis*	JS25R	CP009679.1	
16		SRCM103788	CP035399.1	
17		UCMB5044	CP041144.1	
18	*B. subtilis subsp. spizizenii*	ATCC 6633	CP034943.1	
19		TU-B-10	CP002905.1	
20		W23	CP002183.1	
21	*B. licheniformis*	DSM 13	AE017333.1	
22		ATCC 9789	CP023729.1	
23		SRCM100027	CP021677.1	
24		SCDB 14	CP014842.1	
25	*B. atrophaeus*	BA59	CP024051.1	
26		SRCM101359	CP021500.1	
27	*B. pumilus*	UAMX	CP058951.1	
28	*B. subtilis subsp. inaquosorum*	KCTC 13429	CP029465.1	
29	*B. sonorensis*	SRCM101395	CP021920.1	
30	*B. vallismortis*	Bac111	CP033052.1	
31		DSM 11031	CP026362.1	
32	*B. mojavensis*	UCMB5075	CP051464.1	
33	*B. halotolerans*	ZB201702	CP029364.1	
34		F41-3	CP041357.1	
35	*Listeria monocytogenes*	ATCC 19117	NC_018584.1	Outgroup

- Mồi:

Vùng gen *16S rRNA*, *glpF*, *pta*, *purH*, *pycA* và *rpoD* đã được chọn để phân tích phát sinh loài (www.pubmlst.org/bsubtilis). Các đoạn mồi để chạy phản ứng PCR 5 gen được tham khảo từ nghiên cứu của Lê Xuân Thế và cộng sự [27], trình tự mồi được trình bày trong Bảng 2.

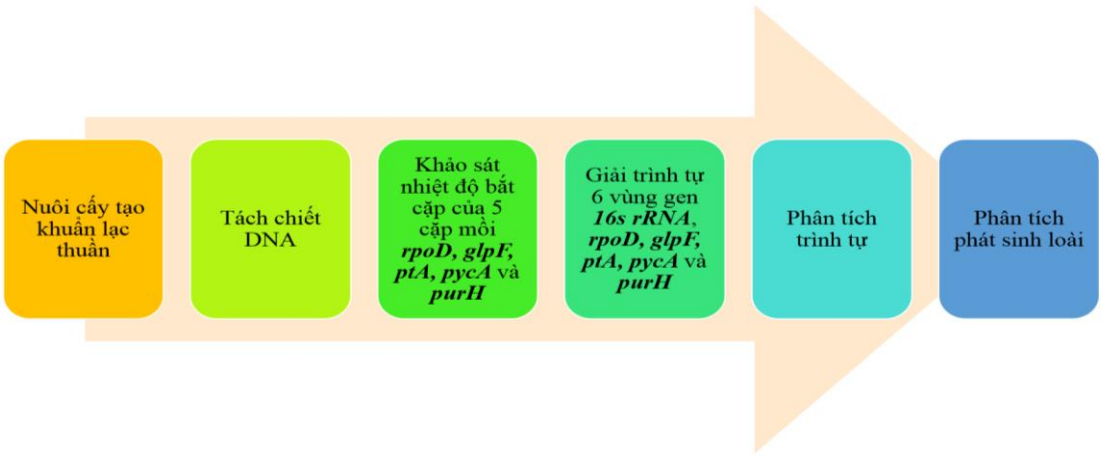
Bảng 2. 3: Thông tin các cặp mồi được sử dụng [27]

Gen mục tiêu	**Tên mồi**	**Trình tự (5’ – 3’)**	**Tm**	**Kích thước (bp)**
16S rRNA	*27F*	AGA GTT TGA TCM TGG CTC AG	65°C	1500
	1492R	TAC GGY TAC CTT GTT ACG ACTT		
rpoD	*rpoD*-F	GCY GAA GAA GAA TTT GAC CTB AA	54 °C	854
	rpoD-R	CGT TTR CTT CTG CTH GGA TGT CT		
glpF	*glpF*-F	WTG ACA GCA TTT TGG GG	54 °C	690

	glpF-R	GTA AAA TAC RCC GCC GA		
ptA	*ptA*-F	ATA CAT AYG AAG GVA TGG AAG	54 °C	610
	ptA-R	TAG CCG ATR TTY CCY GCT		
pycA	*pycA*-F	AAA TCA GAR GCG AAA GC	54 °C	545
	pycA-R	CCT GAG CGG TAA GCC AT		
purH	*purH*-F	AYA TTC ACG GMG GNC TBC T	54 °C	568
	purH-R	TGY TCK BTC GGY TCY CTT TT		

2.2. Phương pháp nghiên cứu

Quy trình nghiên cứu tóm tắt trong sơ đồ sau:



Hình 2. 1 Quy trình nghiên cứu

2.2.1 Nuôi cấy

Các chủng này được tăng sinh trong môi trường Tryptone Soya Broth (TSB) lắc 180 vòng/phút, 30 ^{o}C trong 24 giờ, cấy ria trên đĩa thạch Tryptone Soya Agar (TSA) để tạo các khuẩn lạc riêng lẻ, ủ ở 30^{o}C trong 24 giờ.

2.2.2 Tách chiết DNA tổng số của các chủng vi khuẩn:

- Thu thập sinh khối từ khuẩn lạc thuần trên đĩa thạch TSA.
- Thêm 500 µl STES Buffer, phá vỡ tế bào bởi các hạt bi thủy tinh. Thêm 500 µl phenol - chloroform - isoamyl alcohol (25:24:1).
- Vortex mạnh. Ly tâm 5 phút - tốc độ tối đa (15.000 vòng/phút). Chuẩn bị ống mới 1,5 ml, thêm 30 µl 3M CH3COONa pH 5,2 và 300 µl iso-propanol, lắc đều.
- Chuyển dịch nổi từ ống ly tâm sang ống mới, trộn đều.

- Ly tâm 5 phút - tốc độ tối đa (15.000 vòng/phút), loại bỏ phần nổi phía trên.
- Rửa bằng Ethanol 70%. Bỏ Ethanol. Sấy ở 65^{o}C - 15 phút.
- Thêm 50 µl TE với Rnase (0,01 mg / ml). Ủ ở 65^{o}C - 30 min để DNA tan hoàn toàn.
- Kiểm tra nồng độ và chất lượng DNA dựa trên độ hấp thụ ở 260/ 280 nm và 260/230 nm bằng cách sử dụng máy Nano-Drop.

2.2.3 Khảo sát nhiệt độ bắt cặp của các cặp mồi *rpoD, glpF, ptA, pycA và purH*

- Thành phần phản ứng PCR (thể tích 50 µl)

PCR Master Mix (2X)	25 µL
Mồi xuôi (10µM)	1 µl
Mồi ngược (10µM)	1 µl
DNA tổng số	30 - 100 ng
ddH_2O	vừa đủ 50 µl

- Chu trình nhiệt phản ứng PCR:

95^{o}C	5 phút	
95^{o}C	30 giây	35 chu kỳ
Nhiệt độ khảo sát	30 giây	
72^{o}C	1 phút	
72^{o}C	5 phút	

- Nhiệt độ bắt cặp được khảo sát trong khoảng nhiệt độ như sau:

Bảng 2. 4 Gradient nhiệt độ bắt cặp

Cặp mồi	Nhiệt độ							
rpoD *ptA* *glpF*	48,1	48,5	50,2	52,6	55,6	57,9	59,4	60,1

purH *pycA*	40,1	40,7	42,2	44,5	47,6	49,9	51,4	52

- Sản phẩm PCR được kiểm tra bằng điện di trên gel agarose 2%.
- Nhiệt độ tối ưu được sử dụng để chạy phản ứng PCR cho bước giải trình tự.

2.2.4 Giải trình tự 6 vùng gen *16S, rpoD, glpF, ptA, pycA và purH*

Tinh sạch sản phẩm PCR bằng kit GeneJET PCR Purification Kit (Thermo Scientific), các bước như sau:

- Bước 1: Thêm thể tích dung dịch Binding Buffer 1:1 vào sản phẩm PCR (ví dụ: cứ 100 µL sản phẩm PCR, thêm 100 µL dung dịch Binding Buffer). Trộn kỹ. Kiểm tra màu sắc của dung dịch. Màu vàng cho thấy độ pH tối ưu để liên kết DNA. Nếu màu của dung dịch là cam hoặc tím, thêm 10 µL natri axetat 3 M, dung dịch pH 5,2 và trộn. Màu của hỗn hợp trở thành màu vàng.
- Bước 2: nếu đoạn DNA ≤500 bp, thêm thể tích isopropanol 100% 1:2 (ví dụ: nên thêm 100 µL isopropanol vào 100 µL hỗn hợp PCR kết hợp với 100 µL dung dịch Binding Buffer). Trộn kỹ.
- Chuyển tối đa 800 µL dung dịch từ bước 1 (hoặc bước 2 tùy chọn) sang cột lọc GeneJET. Ly tâm trong 30-60 giây. Loại bỏ dung dịch.
- Thêm 700 µL Dung dịch Wash Buffer vào cột lọc GeneJET. Ly tâm trong 30-60 giây. Loại bỏ dung dịch và đặt cột lọc trở lại ống thu. Lly tâm cột lọc GeneJET trống thêm 1 phút để loại bỏ hoàn toàn dung dịch Wash Buffer còn sót lại.
- Chuyển cột lọc GeneJET sang tube sạch 1,5 mL. Thêm 50 µL dung dịch Elution Buffer vào giữa màng lọc của cột GeneJET và ly tâm trong 1 phút.
- Loại bỏ cột GeneJET và lưu trữ DNA tinh khiết ở -20°C.
- Kiểm tra nồng độ và chất lượng DNA dựa trên độ hấp thụ ở 260/280 nm và 260/230 nm bằng cách sử dụng máy Nano-Drop. Nồng độ DNA được sử dụng để chạy phản ứng Chain-termination PCR từ 10-100ng/ µL

Chạy phản ứng Chain-termination PCR (thể tích 20 µl)

- Thành phần:

BigDye™ Terminator 3.1 Ready Reaction Mix	8 µl
Mồi (3.2 µM)	2 µl
Mạch khuôn (sản phẩm PCR tinh sạch)	10 - 100 ng
ddH_2O	vừa đủ 20 µl

- Chu trình nhiệt:

96°C	1 phút	
96°C	10 giây	25 chu kỳ
50°C	5 giây	
60°C	4 phút	

Sản phẩm Chain-termination PCR được tinh sạch bằng Sephadex G-50, ủ ở nhiệt độ 95°C 2 phút, và được giải trình tự.

2.2.5 Phân tích trình tự

- Các đầu trình tự chất lượng thấp được loại bỏ bằng phần mềm ATGC (GENETYX CORPORATION). Chỉ sử dụng các trình tự nucleotide chất lượng cao, có nghĩa là chúng phải dựa trên các peak không gạch chân. Các biểu đồ điện tử thu được bằng giải trình tự Sanger phải luôn được kiểm tra theo cách thủ công, đầu 5'và đầu 3' "không rõ ràng" phải được cắt bỏ trước khi phân tích, các trình tự mồi trong quá trình khuếch đại PCR nên được loại bỏ khỏi phân tích vì các mồi có thể gây ra sai lệch trình tự [43]
- Trình tự các đoạn gen của các chủng tham chiếu được tìm kiếm bằng cách BLAST trình tự thu được trên cơ sở dữ liệu GenBank (http://www.ncbi.nlm.nih.gov)
- Sắp xếp thẳng hàng các trình tự nucleotide của các chủng mục tiêu và các chủng tham chiếu từ NCBI (43 chủng), bằng phương pháp MUltiple Sequence Comparison by Log Expectation – MUSCLE [44] của phần mềm MEGA 11, Với những vùng không có khả năng sắp xếp thẳng hàng, loại bỏ trước khi đưa vào phân tích
- Phân tích thống kê các đặc điểm của gen: đa dạng nucleotide, hàm lượng G+C, Information Parsimony Site của tất cả trình tự đã được tính toán bằng chương trình DnaSP, phiên bản 6 (//www.ub.es/dnasp)[45].

2.2.6 Phân tích phát sinh loài

- Các cây phát sinh loài từ trình tự các trình tự gen riêng lẻ được xây dựng bằng phương pháp phân cụm Neighbor-Joining [14]. Cây tối ưu được hiển thị. Tỷ lệ cây sao chép trong đó các đơn vị phân loại liên quan được nhóm lại với nhau trong thử nghiệm bootstrap (1000 lần lặp lại) được hiển thị bên cạnh các nhánh [46]. Khoảng cách tiến hóa được tính toán bằng phương pháp P-distance [47] và được tính theo đơn vị số lượng khác biệt cơ sở trên mỗi vị trí. Phân tích này liên quan đến 43 trình tự nucleotide. Tất cả các vị trí không rõ ràng đã bị xóa cho từng cặp trình tự (tùy chọn xóa theo cặp). Có tổng cộng 4580 vị trí trong bộ dữ liệu cuối cùng. Các phân tích tiến hóa được tiến hành trong MEGA 11 [48].
- Các cây phát sinh loài với số lượng các gen ghép nối tăng dần cũng được phân tích để so sánh mức độ phân biệt các loài khi thay đổi số lượng gen sử dụng trong phương pháp MLSA, tiêu chí để chọn lựa thứ tự gen ghép nối trong nghiên cứu này dựa vào giá trị khoảng cách trung bình của tập dữ liệu được phân tích bằng phương pháp p-distance. Cây phát sinh loài dựa vào trình tự ghép nối được xây dựng từ trình nối nhiều tệp chứa dữ liệu trình tự thành một liên kết trình tự duy nhất của phần mềm MEGA 11. Công cụ này được sử dụng như sau:
    - Tất cả các tệp đã được căn chỉnh, được nối phải được đặt cùng nhau vào một thư mục. Không được có tệp nào khác trong thư mục này và tất cả các tệp này phải là tệp có định dạng FASTA hoặc tệp có định dạng MEGA. Tất cả dữ liệu cũng phải cùng loại (không thể trộn dữ liệu DNA và axit amin).
    - Từ trang chính của MEGA, nhấp vào Data->Concatenate Sequence Alignments. Chọn thư mục chứa các tệp trình tự đã căn chỉnh.
    - MEGA xử lý các tệp đầu vào theo thứ tự bảng chữ cái, nối các chuỗi có cùng tên và thêm một chuỗi mới khi gặp tên mới.
    - Sau khi ghép xong, xây dựng cây phát sinh loài bằng phương pháp phân cụm Neighbor-Joining đã được nêu ở trên.
- Với mỗi cây phát sinh loài, một ma trận khoảng cách được phân tích để so sánh mức độ sai khác trình tự giữa các loài thuộc chi *Bacillus* trong phạm vi nghiên cứu. Ước tính về sự khác biệt giữa các chuỗi trình tự. Số lượng khác biệt cơ sở trên mỗi vị trí giữa các chuỗi trình tự được hiển thị, kết quả thu được bằng quy trình bootstrap (1000 lần lặp lại). Phân tích này liên quan đến 43 trình tự nucleotide. Tất cả các vị trí không rõ ràng bị xóa cho từng cặp trình tự (tùy chọn xóa theo cặp). Có tổng cộng 2615 vị trí trong bộ dữ liệu cuối cùng [48].

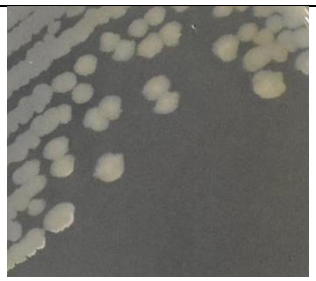
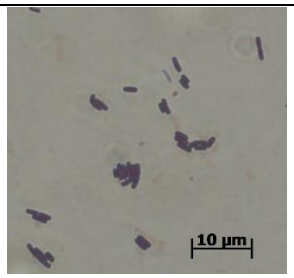
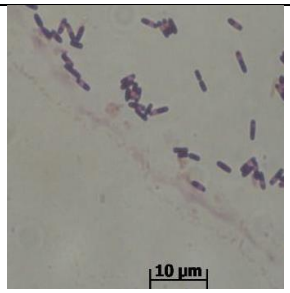
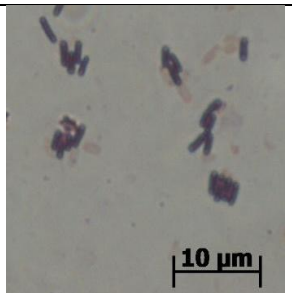
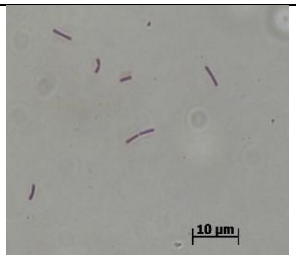
Chương 3. KẾT QUẢ VÀ THẢO LUẬN

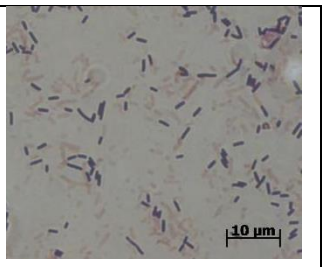
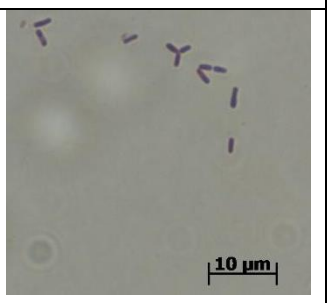
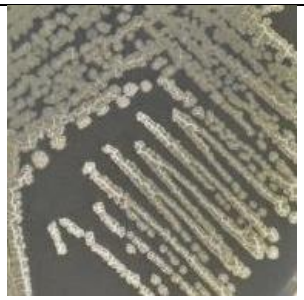
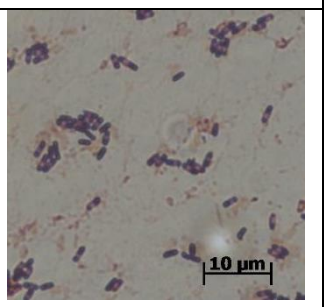
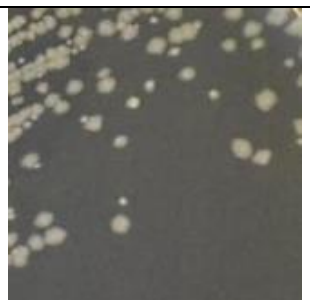
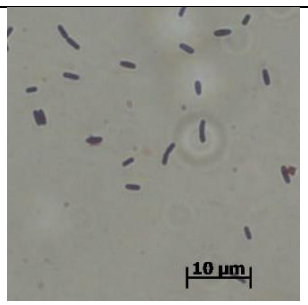
3.1. Nuôi cấy

Các chủng này được tăng sinh trong môi trường TSB lắc 180 vòng/phút, 30^{o}C trong 24 giờ, cấy ria trên đĩa thạch TSA, ủ ở 30^{o}C trong 24 giờ để tạo các khuẩn lạc thuần riêng lẻ, thu sinh khối tách chiết DNA tổng số.

Đặc điểm hình thái các chủng đều có dạng hình tròn hoặc không đều, rìa nguyên hoặc răng cưa, có màu trắng đục. Tế bào có dạng hình que ngắn hoặc dài, Gram dương. Sinh khối thu từ các khuẩn lạc thuần được tách chiết DNA.

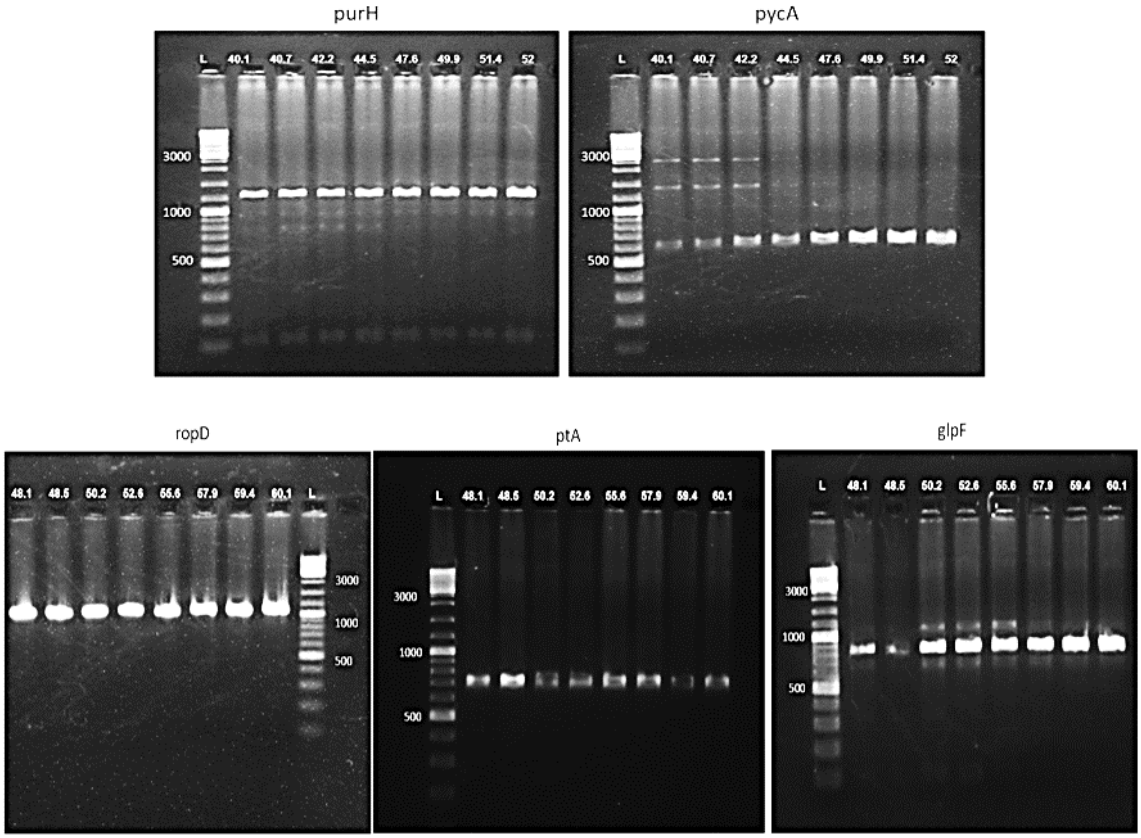
Bảng 3. 1: Kết quả hình thái khuẩn lạc và tế bào của các chủng *Bacillus* spp.

STT	Mã chủng	Hình ảnh khuẩn lạc	Hình ảnh tế bào
1	BC-B0020		
2	BC-B0027		
3	BC-B0028		
4	BC-B0029		

5	BC-B0037		
6	BC-B0039		
7	BC-B0112		
8	BC-B0116		

3.2. Khảo sát nhiệt độ bắt cặp của các cặp mồi

Dù đã chọn được hệ thống mồi thích hợp, các gen tương ứng vẫn có thể không được khuếch đại từ tất cả các chủng được khảo sát [49]. Trong trường hợp này, việc khảo sát nhiệt độ bắt cặp cho phản ứng PCR là cần thiết để khuếch đại các gen mục tiêu. Do đó, để tối ưu hóa phản ứng PCR khuếch đại các gen, nhiệt độ bắt cặp của 5 cặp mồi đã được khảo sát để lựa chọn nhiệt độ tối ưu cho phản ứng PCRvới từng cặp mồi khác nhau.



Hình 3. 1: Kết quả điện di sản phẩm PCR xác định nhiệt độ bắt cặp tối ưu của các cặp mồi cho phản ứng PCR khuếch đại các gen *purH, pycA, rpoD, pta, glpF*. Phản ứng PCR được thực hiện với mạch khuôn là gDNA của BC-B0029, sản phẩm PCR được kiểm ra bằng cách điện di bằng gel agarose 2%.

Kết quả khảo sát nhiệt độ bắt cặp của các cặp mồi được thể hiện ở hình 3.1, đối với gen *purH*, tại nhiệt độ bắt cặp là 40,1ºC, thì sản phẩm PCR khuếch đại gen *purH* cho kết quả điện di là một vạch duy nhất, còn đối với các nhiệt còn lại cho kết quả nhiều hơn một vạch chứng tỏ sản phẩm PCR tại các nhiệt độ này không đặc hiệu. Các nhiệt độ bắt cặp cho sản phẩm PCR không đặc hiệu không thể sử dụng để giải trình tự, vì có thể ra các đoạn trình tự không mong làm giảm chất lượng trình tự ở bước phân tích. Như vậy nhiệt độ bắt cặp tối ưu của cặp mồi cho phản ứng PCR khuếch đại gen *purH* là 40,1. Đối với gen *pycA* thì tại các nhiệt độ bắt cặp là 47,6 ºC, 49,9 ºC, 51,4 ºC và 52 ºC đều cho sản phẩm PCR là một vạch duy nhất và rõ nét. Như vậy có thể chọn nhiệt độ bất kỳ trong dải nhiệt độ từ 47,6 ºC đến 52 ºC cho phản ứng PCR khuếch đại gen *pycA.* Tương tự với gen *rpoD*, trong khoảng nhiệt độ khảo sát 48,1 ºC, 48,5 ºC, 50,2 ºC, 52,6 ºC, 55,6 ºC, 57,9 ºC, 59,4 ºC và 60,1 ºC, tất cả các mốc đều cho kết quả 1 vạch duy nhất rõ nét, do đó có thể kết luận các mốc nhiệt độ này đều thích hợp để chạy phản ứng PCR. Còn với gen *pta*, kết quả ở tất cả các nhiệt độ đều cho ra 1 vạch tuy nhiên chỉ có ở nhiệt độ 48,5ºC là vạch điện di sáng nhất, chứng

tỏ tại nhiệt độ này nồng độ DNA được khuếch đại là cao nhất, cho nên đây là nhiệt độ được chọn để chạy phản ứng PCR. Cuối cùng là gen *glpF*, kết quả điện di cho thấy ở các nhiệt độ 57,9 ^{o}C, 59,4 ^{o}C và 60,1 ^{o}C đều có thể sử dụng chạy phản ứng PCR với một vạch duy nhất và nồng độ DNA cao tương tự nhau (Hình 3.1).

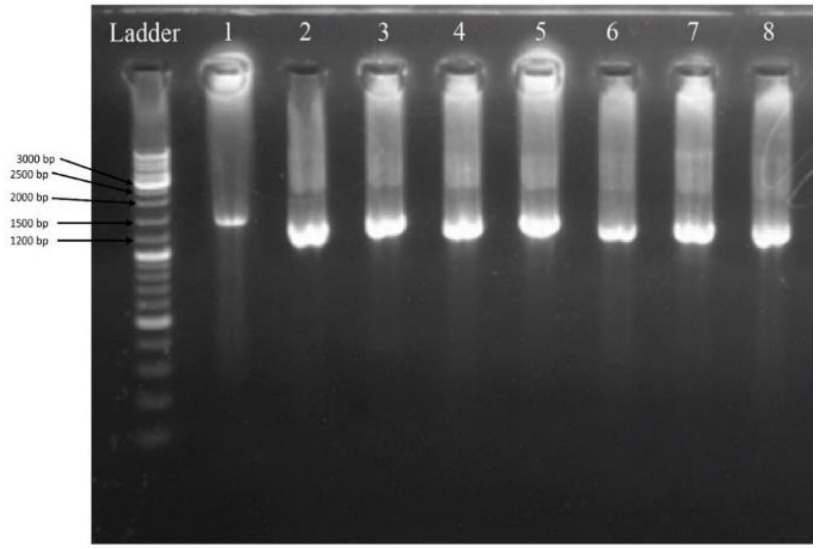
Nhiệt độ bắt cặp dùng để chạy phản ứng PCR cho bước giải trình tự được thể hiện trong bảng 3.2 như sau:

Bảng 3. 2: Nhiệt độ bắt cặp của các cặp mồi dùng để chạy PCR

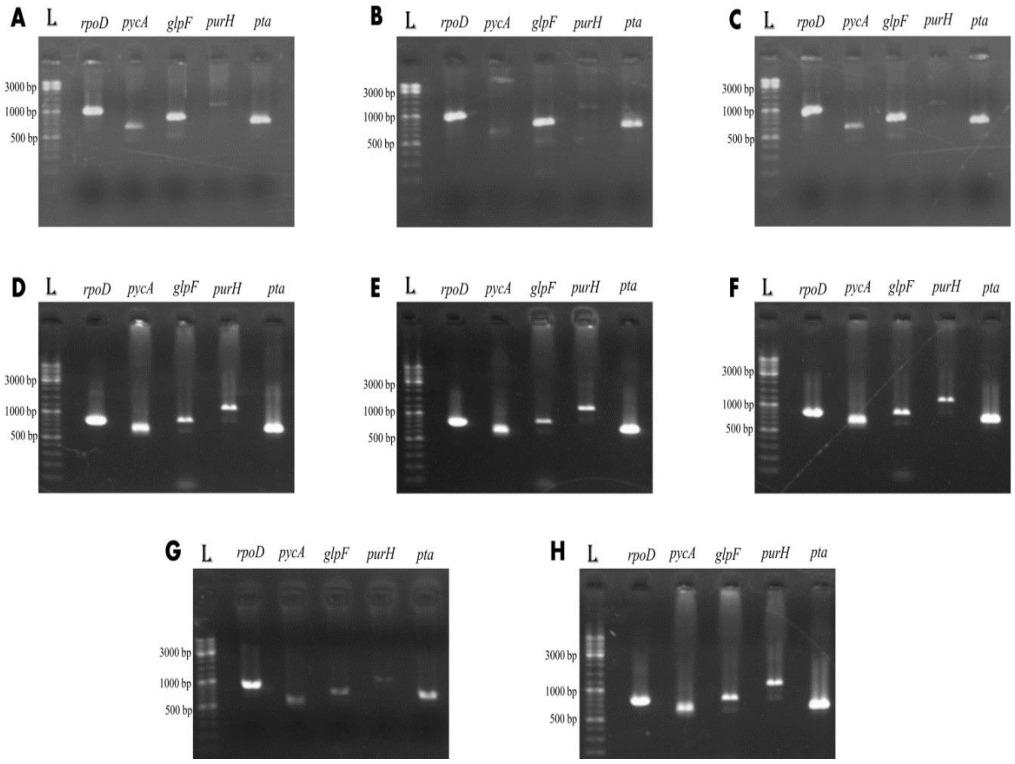
STT	**Mồi**	**Nhiệt độ bắt cặp**
1	*purH*	40,1^{o}C
2	*pycA*	52^{o}C
3	*rpoD*	57,9^{o}C
4	*pta*	48,5^{o}C
5	*glpF*	57,9^{o}C
6	*16S rRNA*	55 ^{o}C

Sau khi chọn được nhiệt độ bắt cặp tối ưu cho các phản ứng PCR, các gen *16S rRNA*, *rpoD*, *glpF*, *pta*, *pycA* và *purH* đã được khuếch đại bằng phản ứng PCR sử dụng DNA tổng số của các chủng BC-B0020, BC-B0027, BC-B0028, BC-B0029, BC-B0037, BC-B0039, BC-B0112, BC-B0116 làm mạch khuôn, trình tự các cặp mồi được sử dụng trong phản ứng PCR được liệt kê trong bảng 2.3. Kiểm tra sự hiện diện của DNA sau khi khuếch đại bằng chạy điện di trên gel agarose 2%, kết quả điện di được thể hiện ở hình 3.2. Tất cả gen *16S rRNA* của tám chủng phân tích được hiển thị dưới dạng điện di trên gel. Kích thước từ 1500-1600 kb tương ứng với kích thước mong muốn của gen *16S rRNA* ở các loài *Bacillus.*

Kết quả điện di sản phẩm PCR các gen *rpoD*, *glpF*, *pta*, *pycA* và *purH* của 8 chủng *Bacillus* spp được thể hiện ở hình 3.3 cho thấy tất cả các chủng mục tiêu đều được khuếch đại thành công bằng phản ứng PCR. 40 mẫu được tinh sạch sản phẩm PCR bằng kit GeneJET PCR Purification Kit (Thermo Scientific), sau đó chạy phản ứng Chain-termination PCR, sản phẩm được tinh sạch bằng Sephadex G-50 rồi tiến hành giải trình tự.



Hình 3. 2: Kết quả điện di sản phẩm PCR khuếch đại gen *16S rRNA* của 8 chủng *Bacillus* spp. gel agarose 2%. Từ 1 – 8 lần lượt là BC-B0020, BC-B0027, BC-B0028, BC-B0029, BC-B0037, BC-B0039, BC-B0112, BC-B0116



Hình 3. 3: Kết quả điện di sản phẩm PCR khuếch đại các gen *rpoD*, *glpF*, *pta*, *pycA* và *purH* của 8 chủng *Bacillus* spp. trên gel agarose 2%.

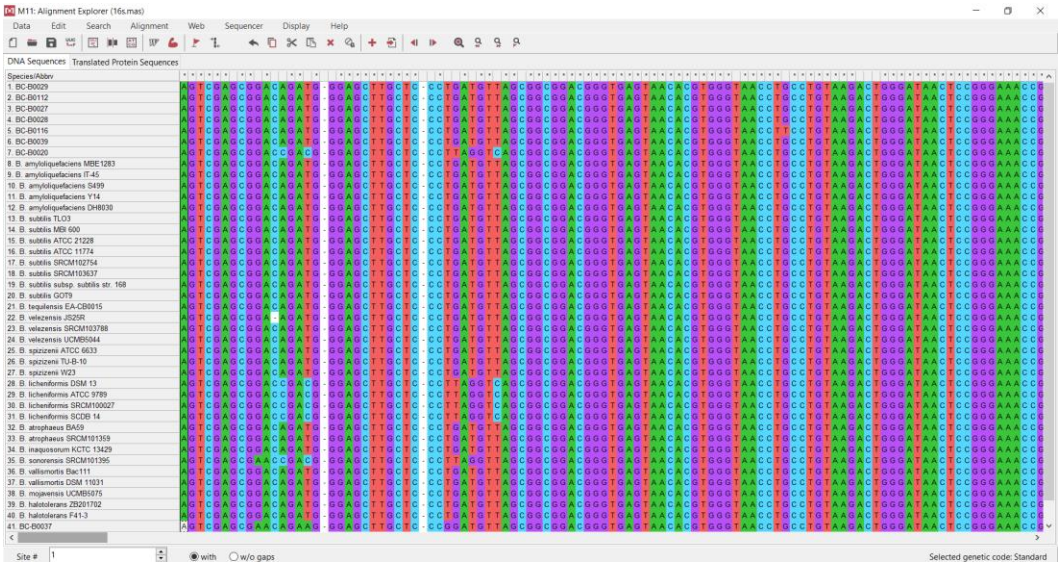
Chú thích: A, B, C, D, E, F, G, H lần lượt là BC-B0020, BC-B0027, BC-B0028, BC-B0029, BC-B0037, BC-B0039, BC-B0112, BC-B0116

3.3. Giải trình tự và phân tích trình tự 6 vùng gen *16S rRNA, rpoD, glpF, pta, pycA* và *purH*

Sau khi giải trình tự, tổng cộng thu được 48 trình tự từ 8 chủng *Bacillus spp*. ứng với 6 gen *16S rRNA, rpoD, glpF, pta, pycA* và *purH* được thể hiện tóm tắt trong hình 3.4. Kết quả trình tự cho thấy các peak rõ ràng, không chồng chéo lên nhau.

Các trình tự được xử lý bằng cách các đầu trình tự chất lượng thấp được loại bỏ bằng phần mềm ATGC (GENETYX CORPORATION), chỉ sử dụng các trình tự nucleotide chất lượng cao, có nghĩa là chúng phải dựa trên các peak rõ ràng, không gạch chân. Với mỗi gen, trình tự tham chiếu của 35 chủng đã được thu thập từ cơ sở dữ liệu GenBank (http://www.ncbi.nlm.nih.gov), kết quả thu được 6 tập dữ liệu cho 6 gen, mỗi tập dữ liệu bao gồm 8 chủng mục tiêu và 35 chủng tham chiếu (đã bao gồm 1 chủng outgroup).

Hình 3. 4: Kết quả giải trình tự của 8 chủng mục tiêu với 6 gen *16S rRNA, rpoD, glpF, pta, pycA* và *purH*



Hình 3. 5 Trình tự được căn chỉnh, sắp xếp thẳng hàng bằng thuật toán MUSCLE của phần mềm MEGA11

Với mỗi tập dữ liệu của mỗi gen được sắp xếp thẳng hàng các trình tự (Multiple sequence alignment) bằng thuật toán MUSCLE của phần mềm MEGA11.

Một số đặc điểm của gen được phân tích bao gồm chiều dài (số lượng nucleotide), Information Parsimony Site, hàm lượng G+C và khoảng cách trung bình giữa các trình tự để thảo luận và so sánh thêm (Bảng 3.3). Kết quả phân tích giá trị khoảng cách P- distance trung bình giữa các trình tự của gen *16S rRNA* trong bảng 3.3 là 0,010%. Tương tự, khoảng cách trung bình giữa các trình tự của các gen lần lượt là 0,129 % (*rpoD*), 0,132% (*pta*), 0,152% (*purH*), 0,188% (*pycA*), và 0,184% (*glpF*). Giá trị P-distance trung bình giữa các housekeeping gen được sử dụng trong phương pháp MLSA cho thấy sự khác biệt đáng kể so với các trình tự *16S rRNA*. Nhiều trình tự *16S rRNA* giữa các loài có 100% độ tương đồng, điều này khiến các nhà nghiên cứu khó phân biệt các chủng mới hoặc thậm chí các loài mới nếu không tiến hành nghiên cứu sâu hơn. Các Information Parsimony Site (IPS) (vị trí chứa ít nhất hai loại nucleotide hoặc axit amin và ít nhất hai trong số chúng xảy ra với tần số tối thiểu là hai) ở gen *16S rRNA* là 63, có nghĩa là ở gen *16S rRNA* có 63 vị trí có sự thay đổi nuleotide với tần suất từ 2 trở lên giữa 43 trình tự đã được căn chỉnh trên tổng số 1475 nucleotide, tương ứng với tỷ lệ là 4,27%. Tương tự, IPS ở các gen *rpoD*, *glpF*, *pta*, *pycA* và *purH* lần lượt chiếm 31,1%, 33,8%, 38,7%, 42,7%, và 45,2%, điều này cho thấy IPS của *16S rRNA* là thấp nhất trong 6 gen. Như vậy mức độ bảo tồn của gen *16S rRNA* là cao nhất trong tổng số 6 gen phân tích, tuy nhiên do mức độ bảo tồn cao, ít sai khác trong trình tự giữa các loài gần gũi nên các nghiên cứu dựa trên gen *16S rRNA* là không đủ để phân biệt các loài và chủng có liên quan chặt chẽ.

So sánh các giá trị IPS của các gen *rpoD, glpF, pta, pycA* và *purH* có thể thấy mức độ bảo tồn của các gen này thấp hơn gen *16S rRNA* do có nhiều vị trí thay đổi nucleotide giữa các trình tự hơn, chứng tỏ giữa các loài gần gũi thì ở các gen *rpoD, glpF, pta, pycA* và *purH* cũng cho thấy được nhiều sự khác biệt trình tự, có thể sử dụng các gen này với mục đích phân biệt các loài có mối quan hệ gần gũi. Tuy nhiên nếu chỉ sử dụng riêng lẻ các gen *rpoD, glpF, pta, pycA* hoặc *purH* để phân tích phát sinh loài có thể gây mâu thuẫn loài (kết quả cây phát sinh loài của từng gen hình 3.6 đến 3.11). Từ đó việc kết hợp giữa gen *16S rRNA* với các gen *rpoD, glpF, pta, pycA* và *purH* đã được tiến hành với mục tiêu tăng cả mức độ bảo tồn gen và mức độ phân biệt các loài gần gũi. Và việc ghép nối các gen với nhau dựa vào chỉ số IPS tăng dần với mục đích ưu tiên sự bảo tồn gen, thứ tự các gen ghép nối là: *16S rRNA-rpoD, 16S rRNA-rpoD-pta, 16S rRNA-rpoD-pta-purH, 16S rRNA-rpoD-pta-purH-glpF, 16S rRNA-rpoD-pta-purH-glpF-pycA* và các giá trị IPS của các trình tự ghép nối với số lượng gen tăng dần lần lượt là 12,2%, 18,3%, 24,1%, 26,0% và 27,3%. Kết quả cho thấy mức độ bảo tồn từ 1 gen lên 2 gen đã tăng lên đáng kể tương ứng với giá trị IPS giảm, như với gen *rpoD* từ 31,1% còn 12,2% khi ghép nối với gen *16S rRNA*. Và khi số lượng gen tăng lên thì giá trị IPS cũng tăng dần đồng nghĩa với mức độ bảo tồn lại càng giảm dẫn tới giảm độ tin cậy. Do đó không phải số lượng gen càng nhiều thì phân biệt loài càng tốt, tuy nhiên chưa có nghiên cứu nào đưa ra khuyến nghị rằng với giá trị IPS trong khoảng bao nhiêu thì có kết quả phân biệt tốt nhất, có lẽ do nó còn tùy thuộc vào các chủng, loài mục tiêu và phạm vi nghiên cứu nên chưa có khuyến nghị chung nào. Cần kết hợp thêm các dữ liệu từ cây phát sinh loài để đưa ra kết luận cho vấn đề này.

Hàm lượng GC của gen *16S rRNA* cao hơn so với các phân tích MLSA (Bảng 3.3), có nghĩa là trình tự *16S rRNA* có nhiều khả năng được bảo tồn hơn so với trình tự phân tích MLSA kết quả này phù hợp với nhận định trên về độ bảo tồn của gen *16S rRNA*.

Bảng 3. 3: Các đặc điểm của các gen và các gen ghép nối

Gen	Chiều dài (gần đúng)	Information Parsimony Site		Hàm lượng G+C (%)	P-distane (%)
		Số lượng	Tỉ lệ (%)		
16S rRNA	1475	63	4,27	54,9	0,010
rpoD	1100	351	31,1	46,3	0,129
pta	960	329	33,8	46,4	0,132

Gen	Chiều dài (gần đúng)	Information Parsimony Site		Hàm lượng G+C (%)	P-distane (%)
purH	1200	484	38,7	47,9	0,152
pycA	720	309	42,7	46	0,188
glpF	820	375	45,2	46,9	0,184
16s-rpoD	2500	305	12,2	49,9	0,063
16s-rpoD-pta	3479	635	18,3	48,9	0,083
16s-rpoD-pta-purH	5018	1209	24,1	48,6	0,104
16s-rpoD-pta-purH-glpF	5839	1518	26,0	48,4	0,111
16s-rpoD-pta-purH-glpF-pycA	6293	1718	27,3	48,1	0,120

3.4 Phân tích phát sinh loài

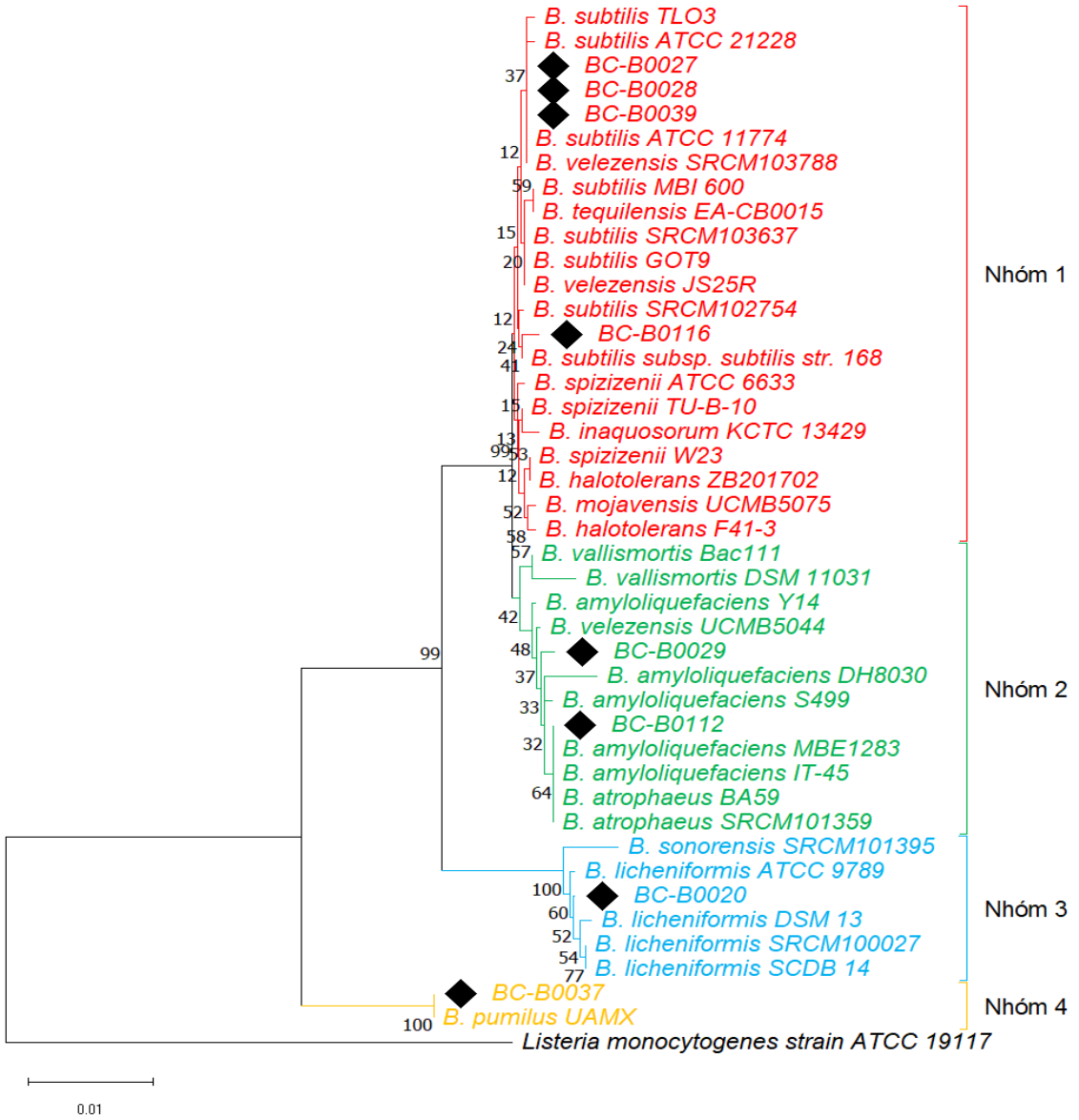
3.4.1 Phân tích phát sinh loài từ trình tự các vùng gen *16S rRNA, rpoD, glpF, ptA, pycA và purH*

Cây phát sinh loài được xây dựng để phân tích phát sinh chủng loại của 8 chủng mục tiêu so với 35 chủng tham chiếu. Dựa vào kết quả của cây phát sinh loài từ trình tự gen *16S rRNA* Hình 3.6, có thể chia thành bốn nhóm như sau:

- Nhóm 1 gồm *B. subtilis, B. subtilis subsp subtilis, B. subtilis subsp spizizenii* và *B. subtilis subsp inaquosorum, B. tequilensis, B. mojavensis, B. halotolerans* và *B. velezensis*
- Nhóm 2 gồm *B. amyloliquefaciens, B. vallismortis* và *B. atrophaeus, B. velezensis*
- Nhóm 3 gồm *B. licheniformis* và *B. sonorensis*
- Nhóm 4 gồm *B. pumilus*

B. velezensis được xuất hiện trong hai nhóm, nhóm 1 và nhóm 2. Các nghiên cứu dựa trên gen *16S rRNA* thường cung cấp độ phân loại thấp trong việc phân biệt giữa các loài có liên hệ chặt chẽ. Nhiều bài báo trước đây đã chỉ ra rằng một số loài chia sẻ trên 97% hoặc thậm chí 100% trình tự gen *16S rRNA* [50][51][52]. Như vậy, trình tự gen *16S rRNA* không thể sử dụng để phân biệt các loài có mối quan hệ gần gũi.. Hình 3.6 chỉ ra rằng, ở Nhóm 1, *B. subtilis* MBI 600 có mối quan hệ gần gũi với *B. tequilensis* EA-CB0015 hơn các chủng *B. subtilis* khác. Điều này là không phù hợp

vì các chủng thuộc loài *B. subtilis* phải có mối quan hệ gần gũi hơn với loài *B. tequilensis* (các chủng *Bacillus* spp. trên GeneBank được sử dụng trong nghiên cứu này đã được giải trình tự bộ gen nên loại trừ trường hợp các chủng này được định danh không chính xác). Ngoài ra, sự tương đồng về trình tự *16S rRNA* là giữa *B. subtilis* và *B. amyloliquefaciens, B. velezensis* và *B. amyloliquefaciens* cao hơn 99% (cụ thể lần lượt là 99,52%, 99,64%), điều này cho thấy rằng đây là các loài rất khó phân biệt. Hơn nữa, giá trị bootstrap thấp ở nhiều nhánh (đa số <70%), cây phát sinh loài không được coi là đáng tin cậy. Trong bảng ma trận khoảng cách giữa các loài được thể hiện bằng phần trăm sự khác biệt nucleotide của trình tự gen *16S rRNA* (bảng 3.4) ví dụ như giữa loài *B. subtilis* và *B. amyloliquefaciens* là 0,48%, *B. velezensis* và *B. amyloliquefaciens* là 0,36%, sự chồng chéo giữa các loài là rất lớn hầu như không phân biệt được. Kết hợp các dữ liệu trên, có thể nói rằng các nghiên cứu dựa trên gen *16S rRNA* đã được công nhận rộng rãi để xác định mối quan hệ phát sinh loài của vi sinh vật trong một thời gian rất dài, nó có ưu điểm là phân biệt tốt các loài vi sinh vật có quan hệ xa, nhưng lại không tốt cho các nghiên cứu sâu hơn nhằm phân loại chính xác các loài có mối quan hệ gần gũi, ví dụ như các thành viên của chi *Bacillus* trong nghiên cứu này. Cây phát sinh loài của gen *rpoD, glpF, ptA, pycA* và *purH* cũng được xây dựng để xem liệu một gen trong các gen này có thể phân biệt tốt hơn gen *16S rRNA* hay không.



Hình 3. 6: Cây phát sinh loài từ trình tự gen *16S rRNA*, sử dụng phương pháp Neighbor-Joining.

Tỷ lệ phần trăm bootstrap (1000 lần lặp lại) được hiển thị bên cạnh các nhánh Loài outgroup được sử dụng là *Listeria monocytogenes strain ATCC*. Scale bar = 0,01. Tám chủng mục tiêu được đánh dấu ♦.

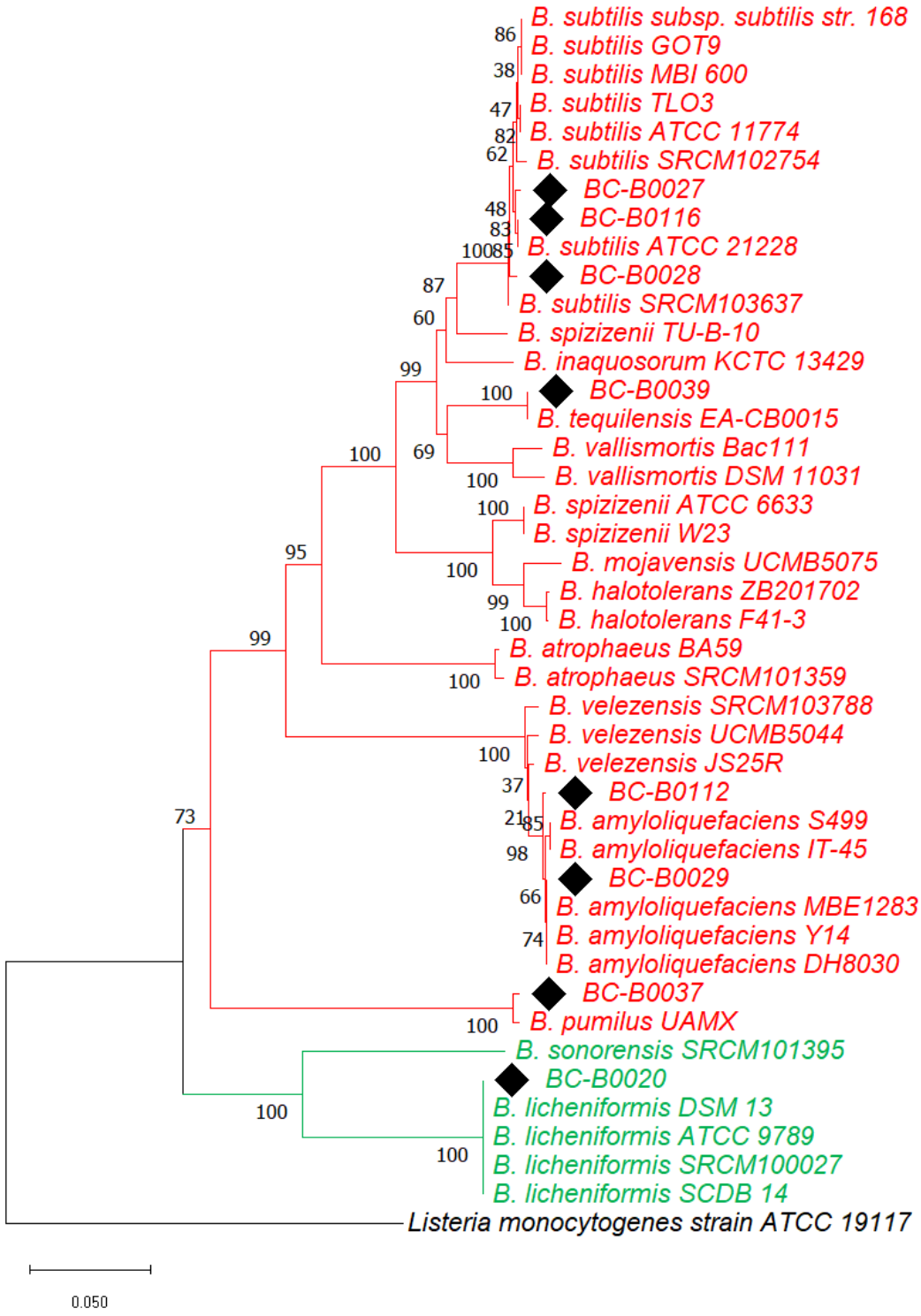
Kết quả cho thấy, cấu trúc cây phát sinh loài từ trình tự riêng lẻ của gen *glpF* và *rpoD* (hình 3.7 và 3.8) tương tự nhau, có thể phân biệt rõ ràng *B. licheniformis* và *B. sonorensis* (bootstrap đạt 100%), các nhánh dài hơn và tách biệt rõ ràng hơn so với cây phát sinh loài từ trình tự *16S rRNA*, tuy nhiên có nhiều giá trị bootstrap bé hơn 95% được cho là không đáng tin cậy. Bên cạnh đó, 2 cây phát sinh loài này cũng cho

thấy loài *B. pumilus* có quan hệ gần gũi với các loài còn lại hơn *B. licheniformis* và *B. sonorensis.*

Cây phát sinh loài được dựng từ trình tự gen *pta* (hình 3.9) và *16S rRNA* có cấu trúc tương tự nhau, ở gen *pta* các nhánh đã có sự phân tách rõ ràng hơn và có giá trị bootstrap cao hơn, tuy nhiên với giá trị bootstrap đáng tin cậy (>95%) chỉ có thể phân biệt ở loài *B. pumilis* (100%), còn lại giá trị bootstrap còn thấp tương tự mức độ phân loại của cây *16S rRNA*.

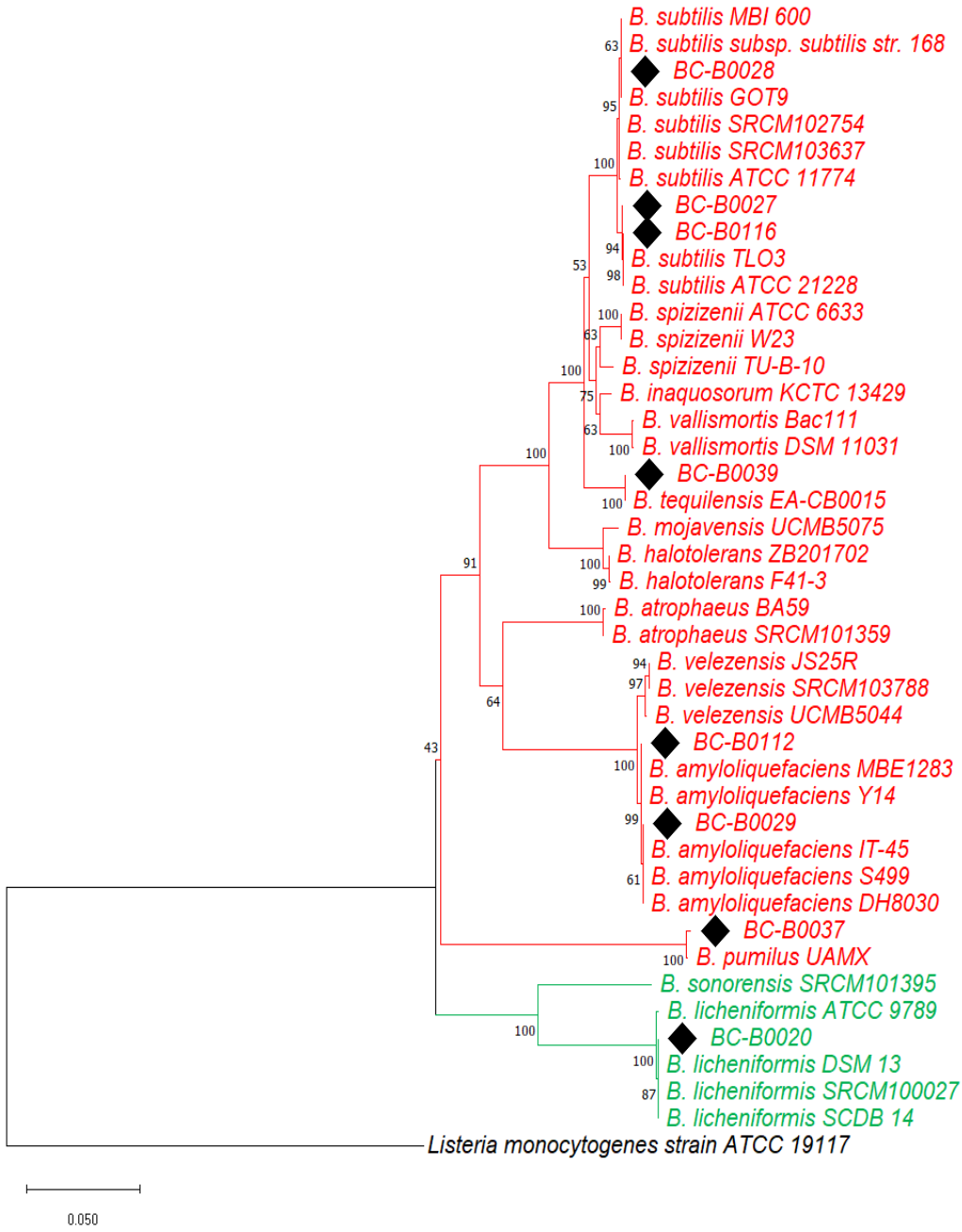
Tương tự với cây phát sinh loài từ trình tự gen *purH* (hình 3.10), các nhánh đã có sự phân tách rõ ràng hơn và có giá trị bootstrap cao hơn so với trình tự *16S rRNA*, có thể phân biệt được loài *B. pumilis*, *B. licheniformis* và *B. sonorensis* (100%). Cây phát sinh loài từ trình tự gen *pycA* (hình 3.11), các nhánh đã tách biệt rõ ràng hơn, tuy nhiên giá trị bootstrap vẫn rất thấp (58%).

Đặc điểm chung của cây phát sinh loài của các gen *rpoD, pta, purH, pycA và glpF* là có giá trị bootstrap cao ở các nhánh phụ và các nhánh chính thì giá trị bootstrap thấp, ngược lại với cây phát sinh loài của gen *16S rRNA*, các nhánh chính có giá trị bootstrap cao >99%, còn các nhánh phụ thì giá trị bootstrap rất thấp như nhánh của nhóm 1 giá trị bootstrap chỉ đạt 12%. Điều này phù hợp với lập luận ở trên rằng gen *16S rRNA* thì không thích hợp để phân tích phát sinh loài của các chủng có quan hệ gần gũi mà sử dụng các gen *rpoD, pta, purH, pycA và glpF* sẽ thích hợp hơn. Tuy nhiên, nếu chỉ sử dụng các gen *rpoD, pta, purH, pycA và glpF* để phân tích phát sinh loài thì độ tin cậy của cây phát sinh loài không cao, do các nhánh chính có giá trị bootstrap thấp. Kết quả này có thể ảnh hưởng bởi chỉ số IPS của các gen *rpoD, pta, purH, pycA và glpF* cụ thể lần lượt là 31,1%, 33,8%, 38,7%, 42,7%, và 45,2%, thể hiện mức độ bảo tồn của các gen này thấp hơn so với gen *16S rRNA*. Do đó việc kết hợp gen *16S rRNA* với các gen còn lại vừa đảm bảo mức độ bảo tồn gen vừa có thể tăng mức độ phân biệt các loài gần gũi, tăng độ tin cậy cho các cây phát sinh loài từ trình tự ghép nối.



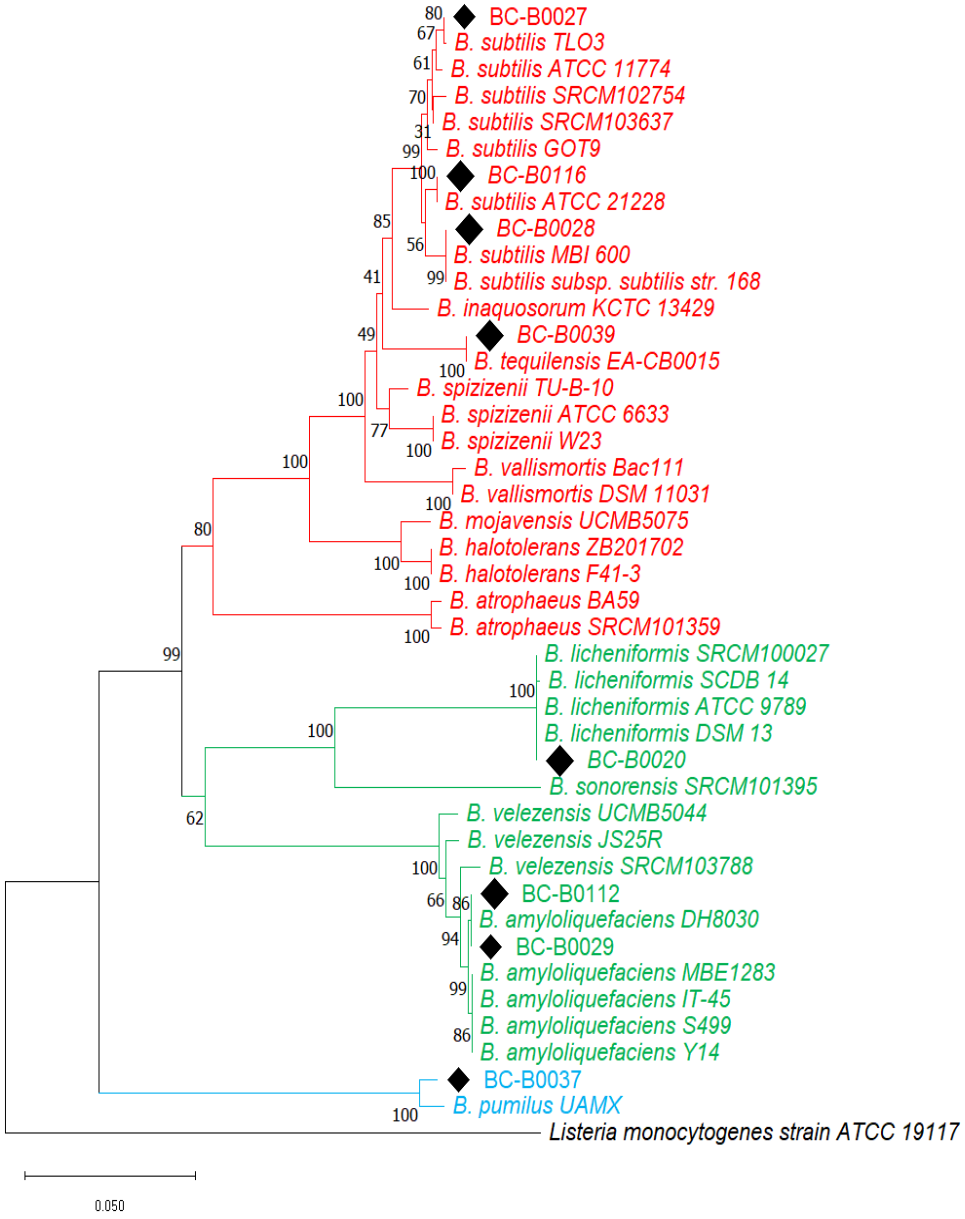
Hình 3. 7: cây phát sinh loài từ trình tự gen *glpF*, sử dụng phương pháp Neighbor-Joining.

Tỷ lệ phần trăm bootstrap (1000 lần lặp lại) được hiển thị bên cạnh các nhánh Loài outgroup được sử dụng là *Listeria monocytogenes strain ATCC*. Scale bar = 0,050. Tám chủng mục tiêu được đánh dấu ♦.



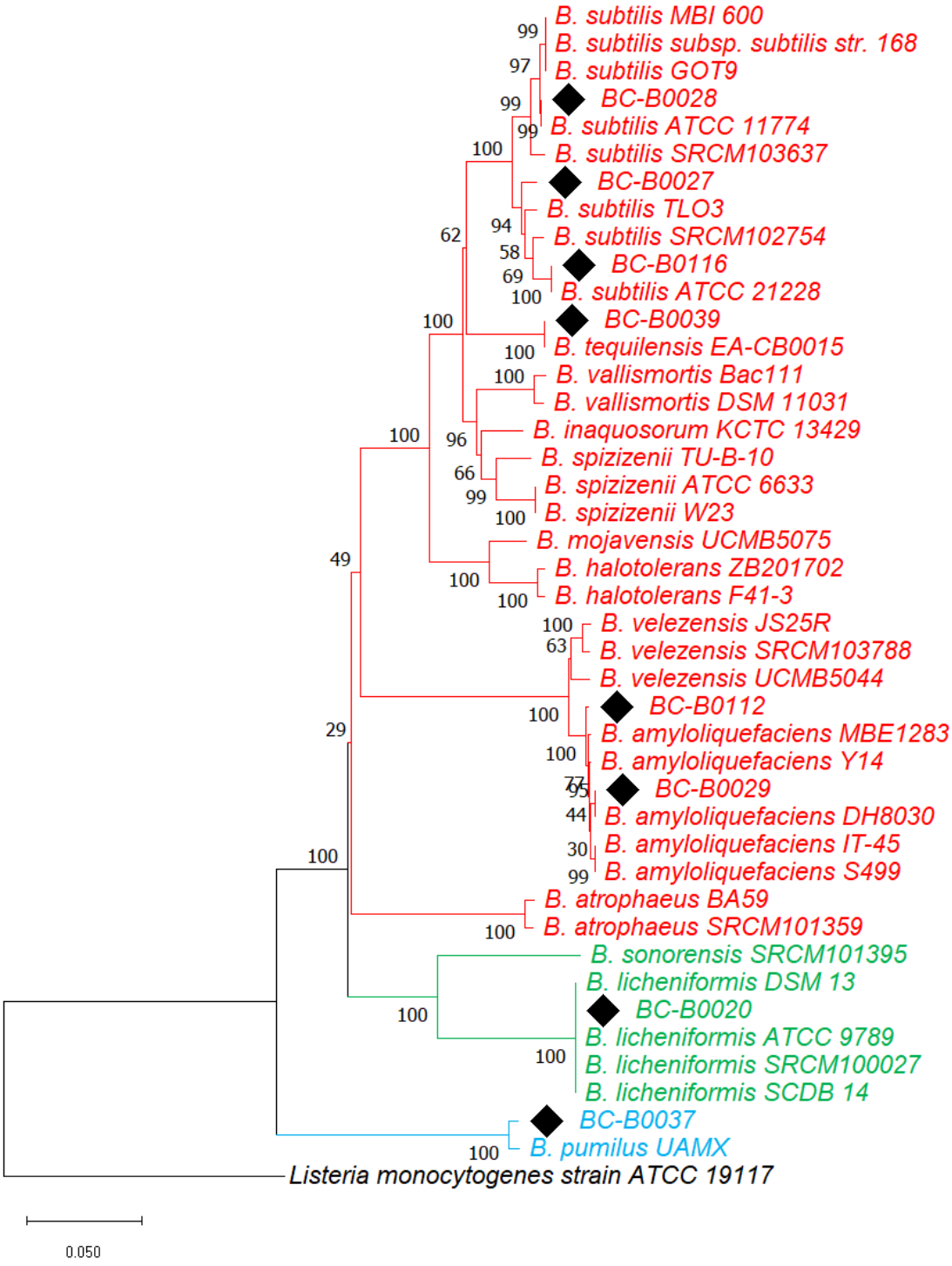
Hình 3. 8: Cây phát sinh loài từ trình tự gen *rpoD*, sử dụng phương pháp Neighbor-Joining.

Tỷ lệ phần trăm bootstrap (1000 lần lặp lại) được hiển thị bên cạnh các nhánh Loài outgroup được sử dụng là *Listeria monocytogenes strain ATCC*. Scale bar = 0,050. Tám chủng mục tiêu được đánh dấu ♦.



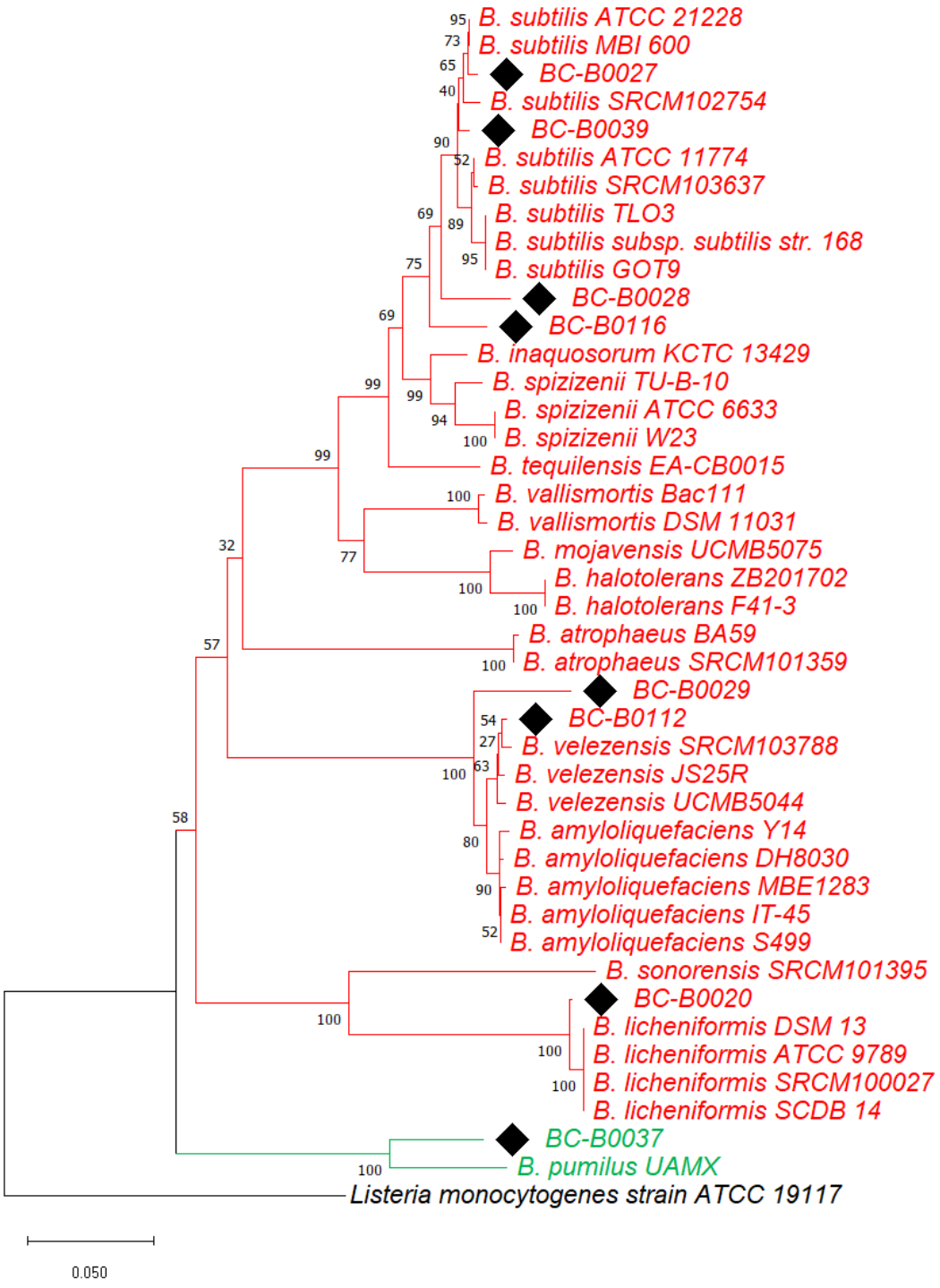
Hình 3. 9: Cây phát sinh loài từ trình tự gen *pta*, sử dụng phương pháp Neighbor-Joining.

Tỷ lệ phần trăm bootstrap (1000 lần lặp lại) được hiển thị bên cạnh các nhánh Loài outgroup được sử dụng là *Listeria monocytogenes strain ATCC*. Scale bar = 0,050. Tám chủng mục tiêu được đánh dấu ♦.



Hình 3. 10: Cây phát sinh loài từ trình tự gen *purH*, sử dụng phương pháp Neighbor-Joining.

Tỷ lệ phần trăm bootstrap (1000 lần lặp lại) được hiển thị bên cạnh các nhánh Loài outgroup được sử dụng là *Listeria monocytogenes strain ATCC*. Scale bar = 0,050. Tám chủng mục tiêu được đánh dấu ♦.



Hình 3. 11: cây phát sinh loài từ trình tự gen *pycA*, sử dụng phương pháp Neighbor-Joining.

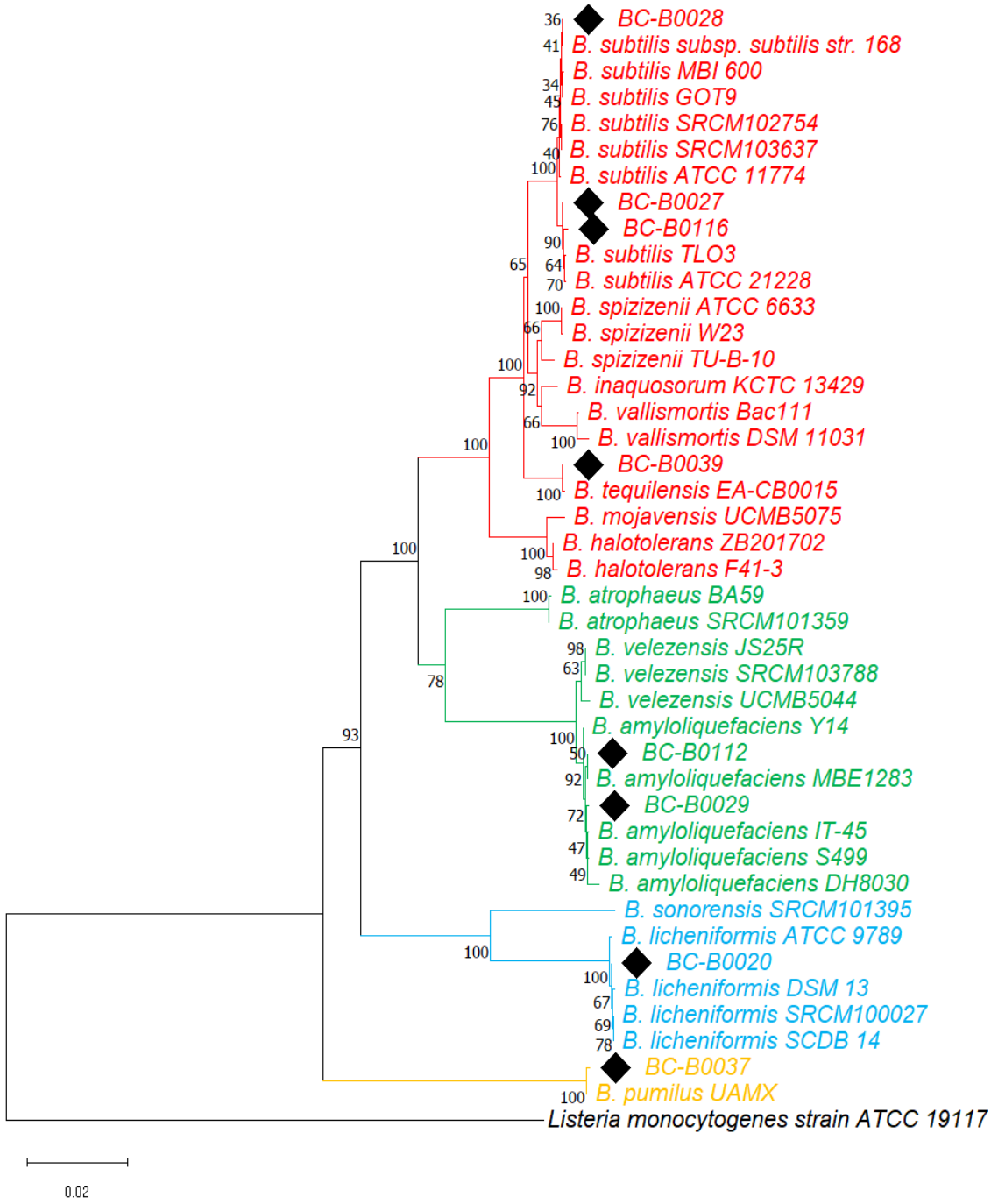
Tỷ lệ phần trăm bootstrap (1000 lần lặp lại) được hiển thị bên cạnh các nhánh Loài outgroup được sử dụng là *Listeria monocytogenes strain ATCC*. Scale bar = 0,050. Tám chủng mục tiêu được đánh dấu ♦.

3.4.2 Khảo sát số lượng gen

Số lượng gen ghép nối cần được xem xét, dựa vào các nghiên cứu khác nhau, ít nhất 4-5 gen có thể được sử dụng, thậm chí nhiều gen hơn nên được xem xét [53]. Tuy nhiên, điều này không phổ biến đối với các nghiên cứu phân tích MLSA. Một số nghiên cứu cho rằng, ít hơn bốn gen được coi là đủ (mặc dù với số lượng gen này có thể cho kết quả đáng nghi ngờ), Ví dụ, năm 2006, Konstantinidis và cộng sự cho thấy rằng sự phân tích phát sinh loài của *Escherichia coli* bằng phương pháp MLSA dựa trên việc lựa chọn ba gen thì phù hợp nhất, kết quả đã được xác nhận với độ phân biệt thu được từ ba gen tương đương kết quả từ dữ liệu của toàn bộ genome[54]. Điều này xảy ra là do các khuyến nghị chung cho nghiên cứu MLSA vẫn chưa được áp dụng, trái ngược với các khuyến nghị được cung cấp cho nghiên cứu phát sinh loài dựa trên gen *16S rRNA*. Như năm 2011, De Vos đã chỉ ra, số lượng gen được chọn thường dựa trên sự cân bằng giữa các mục tiêu như tính khả thi trong thực tế và sức mạnh phân biệt cần thiết [55]. Do đó việc khảo sát số lượng gen ghép nối để xây dựng các cây phát sinh loài đã được phân tích để so sánh mức độ phân biệt các loài khi thay đổi số lượng gen sử dụng trong phương pháp MLSA nhằm đem lại cái nhìn tổng quan hơn về việc lựa chọn số lượng gen.

Khi so sánh cây phát sinh loài dựng từ trình tự ghép nối của gen *16S rRNA-rpoD* (Hình 3.12) với *16S rRNA* (hình 3.6) cho thấy khác biệt rõ ràng nhất giữ 2 cây phát sinh loài này là ở cây phát sinh loài từ gen *16S rRNA* cho thấy 4 chủng mục tiêu BC-B0027, BC-B0028, BC-B0039 và BC-B0116 đều thuộc loài *B. subtilis*, còn ở cây phát sinh loài của gen *16S rRNA-rpoD* thì chỉ có 3 chủng BC-B0027, BC-B0028 và BC-B0116 thuộc loài *B. subtilis* và chủng BC-B0039 là *B. tequilensis* với giá trị bootstrap đạt (100%). *B. velezensis* cũng không còn bị lẫn với các loài *B. subtilis*, tuy nhiên giá trị bootstrap vẫn còn tương đối thấp như nhánh các loài *B. atrophaeus, B. amyloliquefaciens* và *B. velezensis chỉ đạt 78%*. Với giá trị bootstrap 100% có thể khẳng định chủng BC-0020 là *B. licheniformis*. Nhìn chung các giá trị bootstrap cao hơn so với cây phát sinh loài từ trình tự *16S rRNA*, các nhánh cũng có sự phân tách rõ ràng hơn nhưng vẫn chưa phân biệt được các loài khác, và giá trị bootstrap của nhánh chính còn thấp 93%. Như đã đề cập trước đó, 2 loài *B. amyloliquefaciens* và *B. subtilis* có mức độ tương đồng >99%, là 2 loài không thể phân biệt chỉ với trình tự *16S rRNA* nên khi ghép nối thêm trình tự *rpoD* vào thì giá trị khoảng cách của 2 đã tăng từ 0.48% lên 6.24% (bảng 3.4) cho thấy mức độ tương đồng trình tự của 2 loài này đã giảm nhiều. Kết quả từ cây phát sinh loài này chỉ ra rằng có thể phân biệt 2 loài *B. amyloliquefaciens* và *B. subtilis* này như sau: sau khi giải trình tự gen *16S*

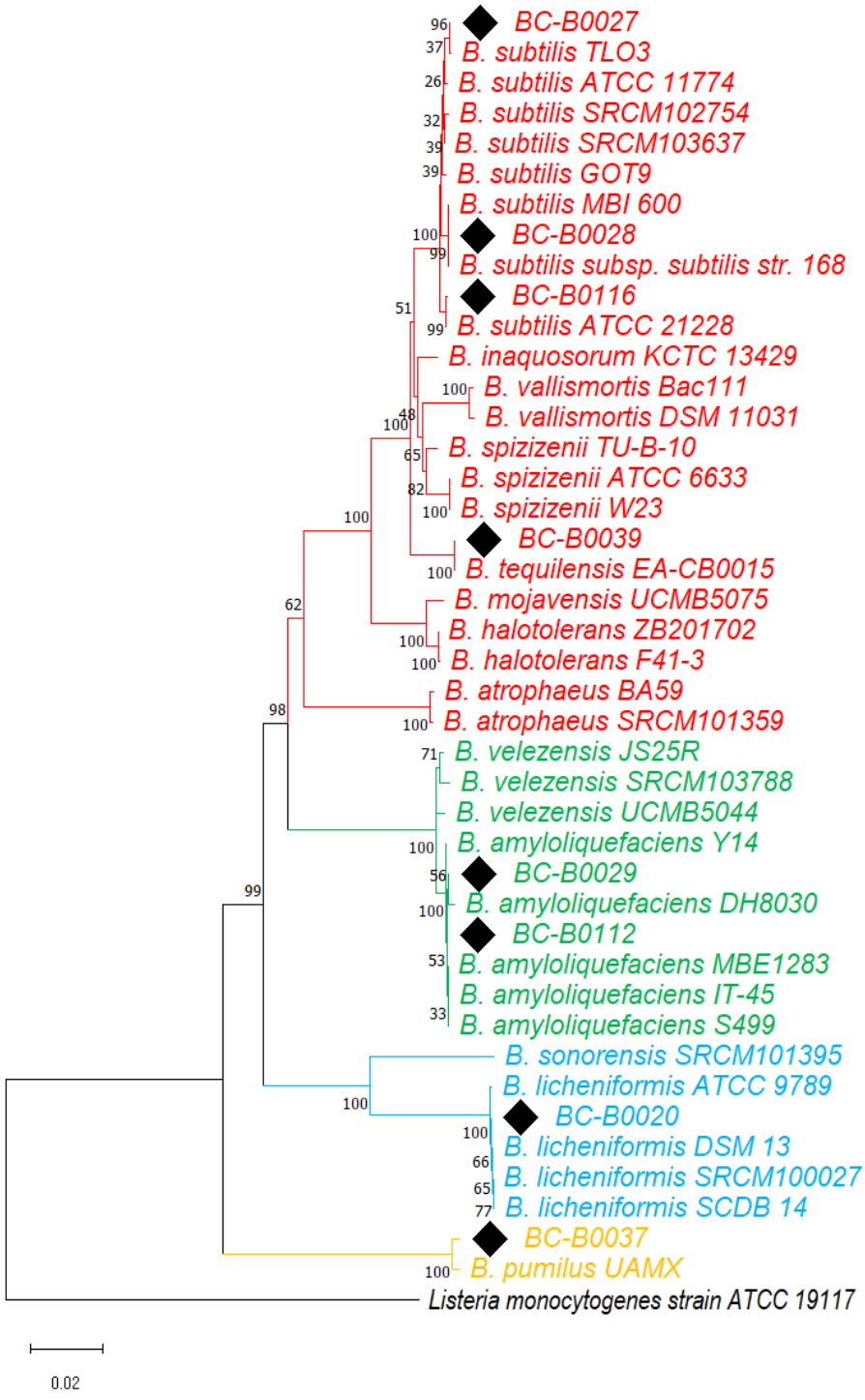
rRNA có thể sử dụng phương pháp MLSA kết hợp gen *rpoD* vào trình tự *16S rRNA* để đưa ra kết luận loài.



Hình 3. 12: Cây phát sinh loài từ trình tự 2 gen *16S rRNA-rpoD*, sử dụng phương pháp Neighbor-Joining.

Tỷ lệ phần trăm bootstrap (1000 lần lặp lại) được hiển thị bên cạnh các nhánh Loài utgroup được sử dụng là *Listeria monocytogenes strain ATCC*. Scale bar = 0,020. Tám chủng mục tiêu được đánh dấu ♦

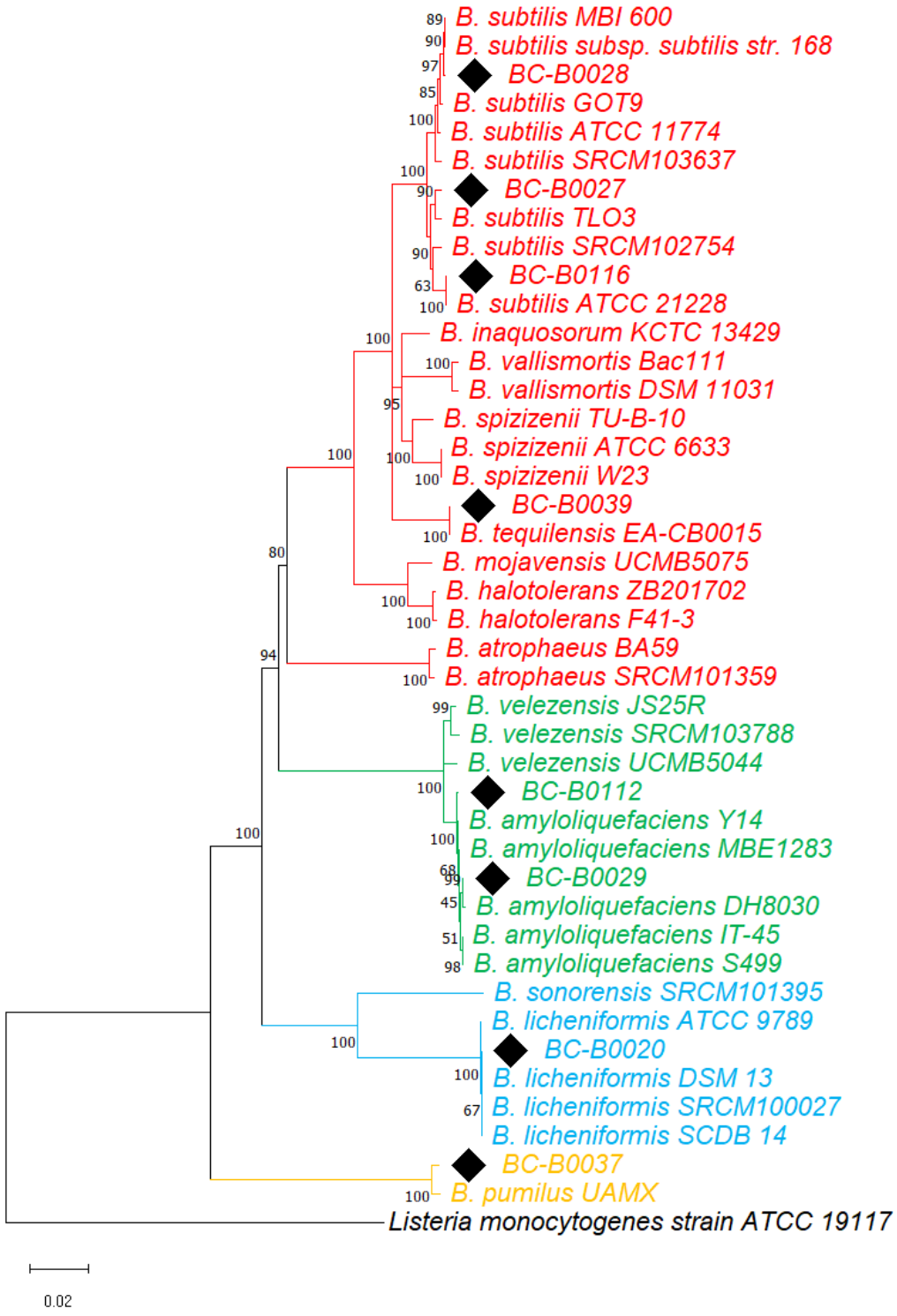
Cây phát sinh loài dựng từ 3 gen *16S rRNA-rpoD-pta* (Hình 3.13) đưa ra một cái nhìn khác về mối quan hệ phát sinh loài trong nhóm loài *B. subtilis* cụ thể *B. atrophaeus* gần với *B. subtilis* hơn nhóm *B. amyloliquefaciens* và *B. velezensis* so với 2 cây phát sinh loài từ gen *16S rRNA* và *16S rRNA-rpoD*, nhóm này cũng tách biệt hẳn với nhóm *B. subtilis* (giá trị bootstrap 100%), kết quả này phù hợp với kết quả nghiên cứu phân loại về phức hợp loài *Bacillus subtilis* của Rooney và cộng sự năm 2009 [56]. Nghiên cứu này cho thấy khi phân tích phát sinh loài bằng phương pháp MLSA với sự kết hợp của 5 gen giữ nhà (*gyrA, rpoB, purH, polC và groEL*) cũng cho ra kết quả *B. atrophaeus* gần với *B. subtilis* hơn nhóm *B. amyloliquefaciens* và *B. velezensis*, và 2 loài *B. subtilis* và *B. amyloliquefaciens* đã được phân biệt rõ. Tuy nhiên chưa thể phân biệt 2 loài *B. amyloliquefaciens* và *B. velezensis*, có sự lẫn lộn giữa 2 loài này với giá trị bootstrap tương đối thấp 78%, kết quả này phù hợp với giá trị khoảng cách giữa 2 loài, chỉ đạt 0.53% cho thấy mức độ tương đồng trình tự của hai loài này là rất cao (bảng 3.4). Dù cùng một loài *B. velezensis* tuy nhiên chủng *B. velezensis* UCMB5044 có sự sai khác trình tự với các chủng còn lại dẫn đến sự phân nhánh, lẫn lộn giữa 2 loài *B. amyloliquefaciens* và *B. velezensis*. Có lẽ đó là lý do dẫn đến giá trị khoảng cách trung bình của 2 loài này là rất thấp.



Hình 3. 13: Cây phát sinh loài từ trình tự 3 gen *16S rRNA-rpoD-pta*, sử dụng phương pháp Neighbor-Joining.

Tỷ lệ phần trăm bootstrap (1000 lần lặp lại) được hiển thị bên cạnh các nhánh Loài outgroup được sử dụng là *Listeria monocytogenes strain ATCC*. Scale bar = 0,020. Tám chủng mục tiêu được đánh dấu ♦

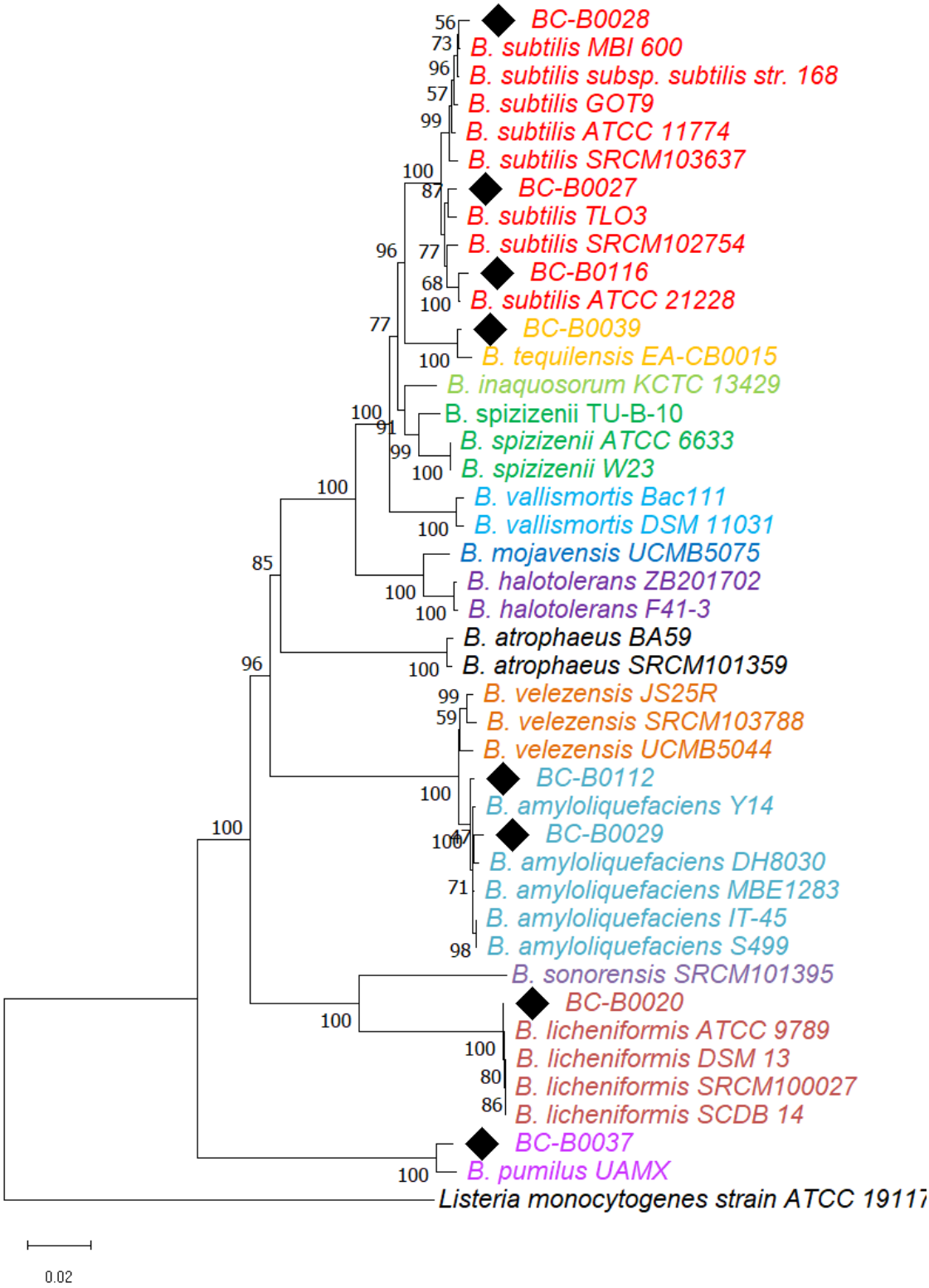
Tiếp tục ghép nối thêm gen *purH* vào kết quả trên có được cây phát sinh loài từ trình tự 4 gen *16S rRNA-rpoD-pta-purH* (hình 3.14). Từ đây ta có thể thấy, nhánh *B. amyloliquefaciens* và *B. velezensis* đã tách biệt rõ ràng hơn, với giá trị bootstrap 100% có thể kết luận 2 chủng BC-0029 và BC-0112 là *B. amyloliquefaciens.* Kết quả này tương đương với kết quả nghiên cứu genome của nhóm loài *B. amyloliquefaciens* được thực hiện bởi Ben Fan và cộng sự năm 2017 [57]. Dựa vào kết quả nghiên cứu genome, tác giả đã đề xuất rằng nhánh *B. amyloliquefaciens* nên được coi là một đơn vị phân loại trên cấp độ loài, bao gồm các loài *B. amyloliquefaciens, B. siamensis* và *B. velezensis*. Trong nhóm *B. subtilis* đã có sự phân nhánh rõ rệt với giá trị bootstrap 100% như loài *B. atrophaeus, B. halotolerans, B. mojavensis*, và có thể kết luận BC-0039 là *B. tequilensis* với giá trị bootstrap đạt 100%. Ở cây phát sinh loài từ trình tự ghép nối của 4 gen này gần như đã phân biệt rõ được các loài trong chi *Bacilus* ở phạm vi nghiên cứu này, tuy nhiên nhánh chính của nhóm loài *B. subtilis* và *B. amyloliquefaciens* có giá trị bootstraps tương đối thấp là 94% và nhánh *B. subtilis* và *B. subtilis subsp subtilis* chỉ đạt 80% cho thấy mức độ tin cậy vẫn chưa cao.



Hình 3. 14: Cây phát sinh loài từ trình tự 4 gen *16S rRNA-rpoD-pta-purH*, sử dụng phương pháp Neighbor-Joining.

Tỷ lệ phần trăm bootstrap (1000 lần lặp lại) được hiển thị bên cạnh các nhánh Loài outgroup được sử dụng là *Listeria monocytogenes strain ATCC*. Scale bar = 0,020. Tám chủng mục tiêu được đánh dấu ♦

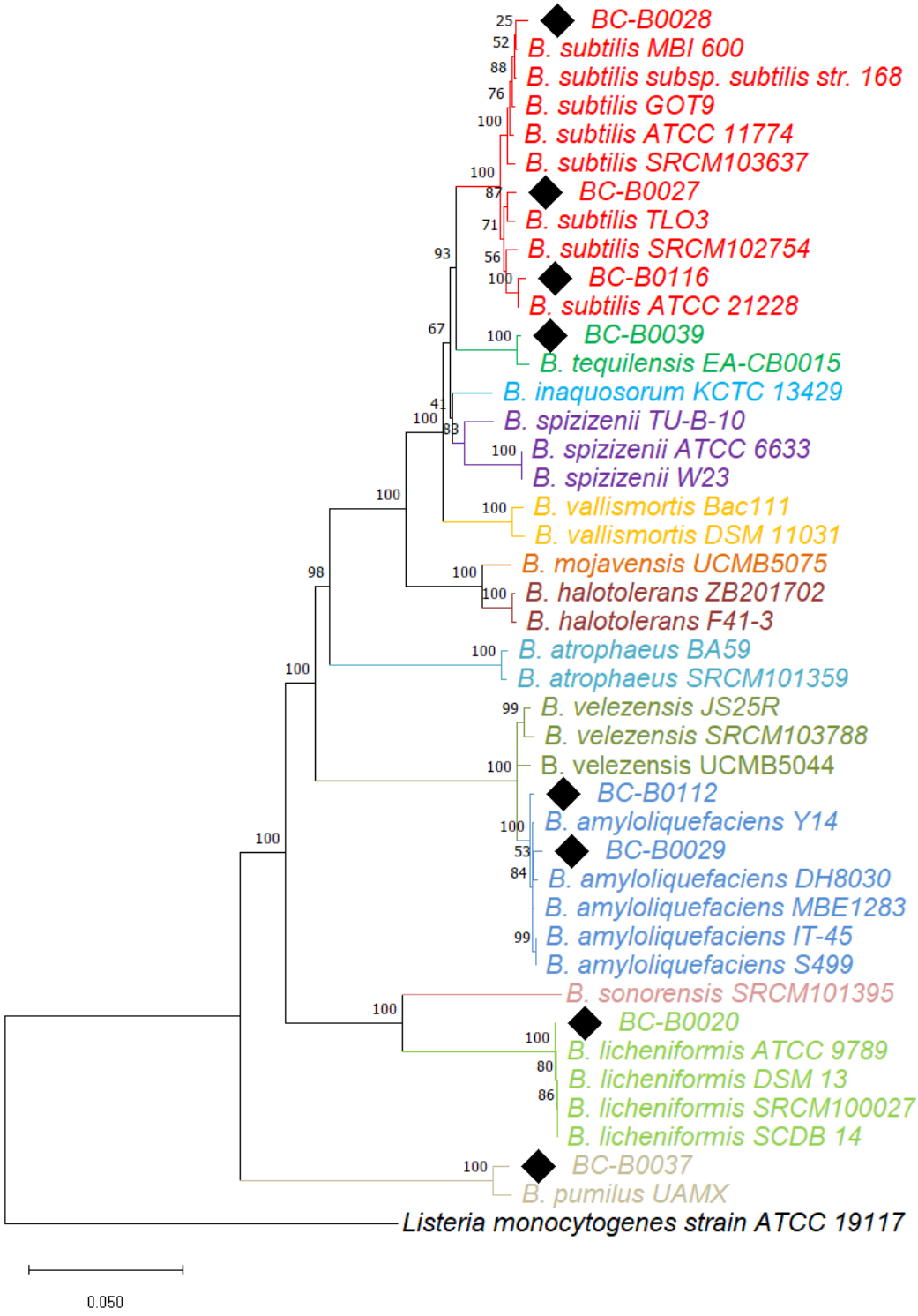
Trình tự gen *pycA* được ghép nối vào trình tự ở trên, xây dựng được cây phát sinh loài của trình tự 5 gen *16S rRNA-rpoD-pta-purH- pycA* (Hình 3.15). Với kết quả này, các nhánh đã được phân biệt rõ ràng với giá trị bootstraps đáng tin cậy >95%. Tuy nhiên cấu trúc cây phát sinh loài tại nhánh các loài *B. subtilis, B. inaquosorum, B. spizizenii, B. tequilensis, B. vallismortis* có sự thay đổi đáng kể cụ thể như loài *B. tequelensis* có mối quan hệ rất gần gũi với nhóm *B. subtilis,* không phân biệt được rõ như ở trình tự 4 gen. Đáng chú ý là nhánh loài *B. spizizenii* đã tách ra thành 1 nhánh riêng với giá trị bootstrap đạt 100% với 2 chủng *B. spizizenii W23* và *B. spizizenii ATCC 6633*, nhưng chủng *B. spizizenii TU-B-10* lại không nằm trong nhánh này mà lại xuất hiện trong nhánh của nhóm loài *B. subtilis*. Với kết quả này có thể đặt vấn đề rằng liệu chủng *B. spizizenii TU-B-10* có thuộc phân loài *B. subtilis subsp spizizenii* hay không. Từ các kết quả này có thể thấy việc lựa chọn gen sử dụng trong phân tích MLSA ảnh hưởng đáng để đến mức độ phát sinh loài, các khảo sát tương tự nên được thực hiện trước khi phân tích phát sinh loài bằng phương pháp MLSA.



Hình 3. 15: Cây phát sinh loài từ trình tự 5 gen *16S rRNA-rpoD-pta-purH- pycA*, sử dụng phương pháp Neighbor-Joining.

Tỷ lệ phần trăm bootstrap (1000 lần lặp lại) được hiển thị bên cạnh các nhánh Loài outgroup được sử dụng là *Listeria monocytogenes strain ATCC*. Scale bar = 0,050. Tám chủng mục tiêu được đánh dấu ♦

Khi ghép nối tiếp gen *glpF* ta được cây phát sinh loài từ 6 gen *16S rRNA-rpoD-pta-purH-pycA-glpF* (Hình 3.16), so với cây phát sinh loài từ 5 gen thì chỉ khác biệt ở nhánh các loài *B. subtilis, B. inaquosorum, B. spizizenii, B. tequilensis, B.* vallismortis, đặc biệt chú ý là nhánh loài *B. vallismortis* đã tách ra thành 1 nhánh riêng biệt với giá trị bootstrap 100%, chủng *B. spizizenii TU-B-10* đã gần gũi với loài *B. spizizenii* hơn với giá trị bootstrap 83%. Các loài còn lại cũng phân tách thành các nhánh dài hơn và có giá trị bootstrap cao hơn (98% - 100%), cụ thể như nhánh nhóm loài B. *subtilis và B. tequelensis* đã có giá trị bootstrap tăng từ 63% lên 93%, và các nhánh loài B. *subtilis* và *B. tequelensis* đạt 100%, có thể kết luận các chủng BC-0027, BC-0028, BC-0116 là loài *B. subtilis,* BC-0039 là loài *B. tequelensis*. Kết quả từ cây phát sinh loài này cho thấy các loài được xếp vào phân loài của *B. subtilis* như *B. spizizenii* và *B. inaquosorum,* đã tách ra hẳn thành các nhánh riêng biệt, dù rằng giá trị bootstrap tại các nhánh này cũng còn tương đối thấp (41%). Với các nghiên cứu có mục đích làm rõ mối quan hệ phát sinh loài của phân loài *B. inaquosorum* và *B. spizizenii* thì cần tăng số lượng gen phân tích MLSA để tăng độ tin cậy cho cây phát sinh loài hoặc tham khảo lựa chọn các gen *gyrA, rpoB, purH, polC* và *groEL* như trong nghiên cứu của Rooney và cộng sự, 2009 [56].



Hình 3. 16: cây phát sinh loài từ trình tự 6 gen *16S rRNA-rpoD-pta-purH-pycA-glpF*, sử dụng phương pháp Neighbor-Joining.

Tỷ lệ phần trăm bootstrap (1000 lần lặp lại) được hiển thị bên cạnh các nhánh Loài outgroup được sử dụng là *Listeria monocytogenes strain ATCC*. Scale bar = 0,020. Tám chủng mục tiêu được đánh dấu ♦

Trong tất cả các trường hợp, các loài không chỉ được xác định dựa vào cấu trúc liên kết cây phát sinh loài và giá trị bootstrap mà còn ở sự tương đồng trình tự giữa các loài được thể hiện trong Bảng 3.4, cho thấy cái nhìn tổng quát về mức độ phân loại khi thay đổi số lượng gen sử dụng trong phương pháp MLSA, cụ thể như để phân biệt 2 loài *B. subtilis* và *B. amyloliquefaciens* có thể sử dụng phương pháp MLSA với chỉ 2 gen ghép nối là *16S rRNA-rpoD* có thể tăng tỉ lệ khác biệt trình tự từ 0,48% lên 6,24%. Hoặc như loài *B. amyloliquefaciens* và *B. velezensis* thì khi sử dụng 6 gen thì vẫn chưa phân biệt tốt với giá trị khoảng cách rất thấp 1.18% tương đương tỉ lệ tương đồng trình tự lên tới 98,82% mặc dù khi quan sát bằng cây phát sinh loài thì cho thấy 2 loài này đã tách biệt và có chỉ số bootstrap đạt 100%, để phân tích phát sinh loài của 2 loài này theo Fan và cộng sự (2017) đã sử dụng phương pháp so sánh trình tự toàn bộ genome [57] nên việc sử dụng 6 gen để phân tích MLSA trong nghiên cứu này cho ra kết quả không cao là phù hợp, tuy nhiên nếu chỉ xét mục đích định danh đến loài 2 loài này thì sử dụng phương pháp MLSA đã chứng minh được ưu điểm của phương pháp này.

Nhìn chung khi tăng số lượng gen thì mức độ khác biệt trình tự các loài tăng. Dựa vào kết quả từ bảng 3.4 kết hợp với kết quả từ cây phát sinh loài của các trình tự ghép nối càng khẳng định thêm ý kiến với 6 gen sử dụng phân tích MLSA thì phân loài *B. inaquosorum* và *B. spizizenii* của *B. subtilis* là chưa có khác biệt rõ ràng. Bên cạnh đó cũng cho thấy tầm quan trọng của việc chọn lựa gen trước khi tiến hành phương pháp MLSA, rõ ràng rằng các gen giữ nhà có chức năng quan trọng nên được lựa chọn vì chúng ổn định và có tính bảo tồn cao. Tuy nhiên, một số gen phổ biến cho phép phân loại phân cấp của tất cả các sinh vật nhân sơ phải phù hợp với gen tốt nhất để phân biệt các đơn vị phân loại có liên quan chặt chẽ [58]. Có nghĩa là việc chọn lựa gen không chỉ dựa vào mức độ phổ biến của gen đó trong nhóm loài mục tiêu mà còn phải là gen phân biệt tốt nhất, như trong nghiên cứu của Mulet và cộng sự năm 2010 về việc sử dụng phương pháp MLSA dựa trên bốn gen giữ nhà (*16S rRNA*, *gyrB*, *rpoB* và *rpoD*) của 107 loài trong chi *Pseudomonas* [59] đã thực hiện khảo sát mức độ bảo tồn cũng như đánh giá khả năng phân biệt các loài mục tiêu của các gen trước khi tiến hành phân tích MLSA. Các gen thường được lựa chọn độc lập cho mỗi đơn vị phân loại mới được điều tra. Sự phát triển của phương pháp MLSA dựa trên các housekeeping gene xuất hiện rộng rãi làm cho việc mở rộng các nghiên cứu MLSA (ví dụ: đối với các loài, chi hoặc họ khác nhau) trở nên dễ dàng hơn tùy thuộc vào mục đích nghiên cứu. Như đã được chỉ ra bởi Gevers và cộng sự [58], các nghiên cứu quy mô lớn về các chi nổi tiếng vẫn cần thiết để cung cấp khuôn khổ cho các nghiên cứu tiếp theo. Nói chung, các gen phải là các trình tự duy nhất trong bộ

gen, cũng như tương đồng và phổ biến trong các đơn vị phân loại được nghiên cứu. Zeigler [60] và De Vos [55] đưa ra thông tin chi tiết hơn về tiêu chí chọn gen để phân tích MLSA, những gen này chủ yếu là thành phần chính của các chức năng trao đổi chất cơ bản nên được lấy từ dữ liệu WGS; các gen phải là các trình tự duy nhất trong bộ gen, đây là điều kiện tiên quyết để tránh giải trình tự các gen cận thân là kết quả của sự sao chép gen và do đó có thể đã tiến hóa sang một chức năng khác; độc lập di truyền so với các gen khác để hạn chế khả năng chuyển gen ngang – việc chuyển gen giữa các tế bào độc lập – xảy ra giữa nhiều gen; các gen được chọn phải có mặt trong tất cả các chủng thuộc về các đơn vị phân loại đang nghiên cứu (tức là thuộc về lõi bộ gen của nhóm sinh vật nhân sơ cụ thể đó); và sự biến đổi trình tự của các gen phải trùng với sự tương đồng bộ gen tổng thể.

Để cung cấp giải pháp tốt hơn cho việc phân tích mối quan hệ phát sinh loài, cách phương pháp MLSA đã đưa ra con đường mới để nghiên cứu sâu hơn và đã được sử dụng rộng rãi [61] [51] [62] . Trong quá trình điều tra về sự tiến hóa và phân loại của các loài *Glycyrrhiza* làm thuốc có nốt sần bản địa ở Trung Quốc, 4 chủng được đề xuất là các loài mới thuộc chi *Mesorhizobium* bởi MLSA [63]. Rooney năm 2009 đã sử dụng phương pháp MLSA để phân tích phát sinh loài của nhóm phức loài *Bacillus subtilis* đã đưa ra một đại diện cho một đơn vị phân loại mới là *B. subtilis subsp. Inaquosorum*[56].

Trong nghiên cứu này, MLSA đã trình bày một số ưu điểm so với các nghiên cứu dựa trên gen *16S rRNA.* Mặc dù cây phát sinh loài dựa trên MLSA cho thấy cấu trúc liên kết tương tự với cây trình tự gen *16S rRNA*, nhưng phương pháp MLSA cho thấy độ phân biệt rõ ràng hơn, các nhánh cũng dài hơn và được hỗ trợ bởi các giá trị bootstrap cao hơn nhiều. Thay vì các loài chồng chéo trong cây phát sinh loài *16S rRNA* hình 3.6, các loài *Bacillus* trong phương pháp MLSA được tách biệt. Chúng ta có thể dễ dàng nhận thấy 13 loài khác nhau thuộc chi *Bacillus* có bất kỳ sự trộn lẫn nào. Cùng với đó, phương pháp MLSA đã đưa ra một cái nhìn mới về mối quan hệ phát sinh loài trong quần thể chi *Bacillus* đặc biệt là nhóm loài *Bacillus subtilis*. Bởi vì sự phân biệt hiện nay đã rõ ràng hơn nhiều, một số loài dường như có mối quan hệ khác, chẳng hạn như *B. atrophaeus* gần với *B. subtilis* hơn *B. amyloliquefaciens* hoặc *B. velezensis* hiện thuộc các nhánh khác với *B. amyloliquefaciens.*

Bảng 3. 4 Giá trị khoảng cách trình tự (%) giữa các loài khi thay đổi số lượng gen phân tích MLSA. Ma trận khoảng cách được tính toán bằng phương pháp p-distance, bootstraps 1000 (số lần lặp lại), bằng chương trình MEGA11

Gen / Loài	*16S rRNA*	*16S rRNA-ropD*	*16S rRNA-ropD-pta*	*16S rRNA-ropD-pta-purH*	*16S rRNA-ropD-pta-purH-glpF*	*16S rRNA-ropD-pta-purH-glpF-pycA*
B.amyloliquefaciens						
B.subtilis	0,48	6,24	8,98	11,78	13,04	13,63
B.tequilensis	0,40	6,34	9,11	11,71	12,92	13,48
B.licheniformis	2,04	9,14	11,31	14,01	16,01	16,69
B.velezensis	0,36	0,53	0,69	1,10	1,12	1,18
B.spizizenii	0,59	6,44	9,25	11,95	13,26	13,83
B.atrophaeus	0,14	4,89	7,85	11,17	12,46	13,28
B.inaquosorum	0,74	6,53	9,16	11,79	13,02	13,63
B.sonorensis	2,37	8,97	11,26	13,85	15,84	16,71
B.vallismortis	0,57	6,28	9,19	11,88	13,16	13,96
B.mojavensis	0,67	6,41	9,06	11,74	13,14	13,78
B.halotolerans	0,66	6,33	8,96	11,70	13,07	13,82
B.pumilus	2,91	9,97	13,19	16,95	18,39	19,06
B.subtilis	*16S rRNA*	*16S rRNA-ropD*	*16S rRNA-ropD-pta*	*16S rRNA-ropD-pta-purH*	*16S rRNA-ropD-pta-purH-glpF*	*16S rRNA-ropD-pta-purH-glpF-pycA*
B.tequilensis	0,12	1,60	2,33	3,70	4,21	4,23
B.licheniformis	1,83	9,02	11,48	13,50	15,33	16,19
B.velezensis	0,18	6,20	8,86	11,63	12,84	13,44
B.spizizenii	0,21	1,35	1,90	3,34	4,07	4,25
B.atrophaeus	0,44	5,53	7,78	10,46	11,17	11,82
B.inaquosorum	0,33	1,27	1,62	3,02	3,43	3,59
B.sonorensis	2,11	9,19	11,71	13,88	15,80	16,58
B.vallismortis	0,49	1,80	2,72	4,06	4,62	5,11
B.mojavensis	0,29	2,93	4,20	5,72	6,65	7,07
B.halotolerans	0,27	2,78	4,03	5,87	6,68	7,20
B.pumilus	2,85	9,95	12,82	15,62	16,95	17,50
B.tequilensis	*16S rRNA*	*16S rRNA-ropD*	*16S rRNA-ropD-pta*	*16S rRNA-ropD-pta-purH*	*16S rRNA-ropD-pta-purH-glpF*	*16S rRNA-ropD-pta-purH-glpF-pycA*
B.licheniformis	1,81	8,62	11,48	13,60	15,23	16,12
B.velezensis	0,14	6,18	8,94	11,57	12,78	13,35
B.spizizenii	0,24	1,46	2,25	3,51	4,35	4,55
B.atrophaeus	0,36	5,54	7,90	10,47	11,34	12,01
B.inaquosorum	0,36	1,43	2,10	3,37	3,80	3,98
B.sonorensis	2,08	8,76	11,52	13,65	15,56	16,37
B.vallismortis	0,47	2,09	3,00	4,25	4,69	5,18
B.mojavensis	0,29	3,19	4,48	6,05	6,91	7,29
B.halotolerans	0,29	2,89	4,41	6,38	7,13	7,58
B.pumilus	2,80	9,93	12,86	15,62	17,24	17,77
B.licheniformis	*16S rRNA*	*16S rRNA-ropD*	*16S rRNA-ropD-pta*	*16S rRNA-ropD-pta-purH*	*16S rRNA-ropD-pta-purH-glpF*	*16S rRNA-ropD-pta-purH-glpF-pycA*
B.velezensis	1,86	8,84	11,17	13,77	15,79	16,48
B.spizizenii	1,81	8,87	11,59	13,78	15,61	16,46
B.atrophaeus	2,04	9,64	11,49	13,47	15,06	15,91
B.inaquosorum	1,95	8,99	11,51	13,37	15,13	15,96
B.sonorensis	0,62	4,97	6,92	8,54	9,56	10,22
B.vallismortis	2,04	8,84	11,39	13,52	15,32	16,17
B.mojavensis	1,88	9,29	11,82	13,77	15,55	16,24
B.halotolerans	1,88	9,23	11,69	13,91	15,69	16,40
B.pumilus	3,84	11,83	14,44	17,51	18,93	19,65

Gen / Loài	*16S rRNA*	*16S rRNA-ropD*	*16S rRNA-ropD-pta*	*16S rRNA-ropD-pta-purH*	*16S rRNA-ropD-pta-purH-glpF*	*16S rRNA-ropD-pta-purH-glpF-pycA*
B.velezensis						
B.spizizenii	0,29	6,38	9,17	11,88	13,15	13,71
B.atrophaeus	0,31	4,99	7,92	11,27	12,48	13,20
B.inaquosorum	0,41	6,46	9,11	11,73	12,91	13,51
B.sonorensis	2,15	8,81	11,22	13,74	15,73	16,61
B.vallismortis	0,47	6,41	9,23	11,86	13,12	13,92
B.mojavensis	0,33	6,32	9,03	11,72	13,05	13,68
B.halotolerans	0,34	6,25	8,90	11,59	12,92	13,67
B.pumilus	2,85	10,06	13,22	16,92	18,27	18,93
B.spizizenii	***16S rRNA***	***16S rRNA-ropD***	***16S rRNA-ropD-pta***	***16S rRNA-ropD-pta-purH***	***16S rRNA-ropD-pta-purH-glpF***	***16S rRNA-ropD-pta-purH-glpF-pycA***
B.atrophaeus	0,57	5,46	7,54	9,99	10,90	11,58
B.inaquosorum	0,19	0,85	1,46	2,28	3,23	3,27
B.sonorensis	2,08	9,04	11,62	13,52	15,57	16,40
B.vallismortis	0,37	1,39	2,13	3,22	4,11	4,66
B.mojavensis	0,22	2,99	4,19	5,62	5,74	6,23
B.halotolerans	0,14	2,70	4,06	5,73	5,81	6,46
B.pumilus	2,88	10,30	12,96	15,59	17,10	17,64
B.atrophaeus	***16S rRNA***	***16S rRNA-ropD***	***16S rRNA-ropD-pta***	***16S rRNA-ropD-pta-purH***	***16S rRNA-ropD-pta-purH-glpF***	***16S rRNA-ropD-pta-purH-glpF-pycA***
B.inaquosorum	0,72	5,66	7,81	10,17	10,95	11,55
B.sonorensis	2,37	9,32	11,58	13,60	15,45	16,43
B.vallismortis	0,54	5,64	7,74	10,29	11,21	11,87
B.mojavensis	0,65	5,06	7,34	9,89	10,86	11,64
B.halotolerans	0,65	5,18	7,31	9,85	10,71	11,53
B.pumilus	2,88	10,39	13,11	15,85	17,07	17,58
B.inaquosorum	***16S rRNA***	***16S rRNA-ropD***	***16S rRNA-ropD-pta***	***16S rRNA-ropD-pta-purH***	***16S rRNA-ropD-pta-purH-glpF***	***16S rRNA-ropD-pta-purH-glpF-pycA***
B.sonorensis	2,08	9,20	11,66	13,61	15,58	16,42
B.vallismortis	0,47	1,18	2,14	2,97	3,58	4,09
B.mojavensis	0,36	2,95	4,02	5,28	6,31	6,68
B.halotolerans	0,29	2,69	3,92	5,47	6,42	6,95
B.pumilus	2,88	10,09	12,72	15,41	16,94	17,46
B.sonorensis	***16S rRNA***	***16S rRNA-ropD***	***16S rRNA-ropD-pta***	***16S rRNA-ropD-pta-purH***	***16S rRNA-ropD-pta-purH-glpF***	***16S rRNA-ropD-pta-purH-glpF-pycA***
B.vallismortis	2,37	9,34	12,02	13,92	15,95	16,75
B.mojavensis	2,15	9,32	11,57	13,75	15,80	16,54
B.halotolerans	2,15	9,07	11,45	13,69	15,74	16,47
B.pumilus	3,95	11,22	14,23	17,54	19,01	19,76
B.vallismortis	***16S rRNA***	***16S rRNA-ropD***	***16S rRNA-ropD-pta***	***16S rRNA-ropD-pta-purH***	***16S rRNA-ropD-pta-purH-glpF***	***16S rRNA-ropD-pta-purH-glpF-pycA***
B.mojavensis	0,54	3,39	4,77	6,21	7,12	7,40
B.halotolerans	0,47	3,21	4,73	6,39	7,27	7,58
B.pumilus	2,88	10,29	12,92	15,57	17,19	17,91
B.mojavensis	***16S rRNA***	***16S rRNA-ropD***	***16S rRNA-ropD-pta***	***16S rRNA-ropD-pta-purH***	***16S rRNA-ropD-pta-purH-glpF***	***16S rRNA-ropD-pta-purH-glpF-pycA***
B.halotolerans	0,14	0,54	0,88	1,86	1,98	2,06
B.pumilus	3,09	10,13	13,06	15,84	17,37	17,98
B.halotolerans	***16S rRNA***	***16S rRNA-ropD***	***16S rRNA-ropD-pta***	***16S rRNA-ropD-pta-purH***	***16S rRNA-ropD-pta-purH-glpF***	***16S rRNA-ropD-pta-purH-glpF-pycA***
B.pumilus	2,99	9,98	12,82	15,68	17,27	17,95

KẾT LUẬN VÀ KIẾN NGHỊ

KẾT LUẬN

- Đã xác định được nhiệt độ bắt cặp tối ưu đối với các cặp mồi sử dụng cho phản ứng PCR khuếch đại các gen *rpoD, pta, purH, glpF và pycA.* Các gen *16S rRNA, rpoD, pta, purH, pycA và glpF* của 8 chủng *Bacillus spp*. đã được giải trình tự thành công và các trình tự này được sử dụng để phân tích phát sinh loài
- Gen *16S rRNA, rpoD, pta, purH, pycA và glpF* riêng lẻ thì không thích hợp để phân tích phát sinh loài của các chủng có quan hệ gần gũi và độ tin cậy của cây phát sinh loài không cao.
- Thứ tự các gen ghép nối phân tích MLSA: *16S rRNA-rpoD, 16S rRNA-rpoD-pta, 16S rRNA-rpoD-pta-purH, 16S rRNA-rpoD-pta-purH-glpF, 16S rRNA-rpoD-pta-purH-glpF-pycA*
- Có thể định danh loài *B. amyloliquefaciens* và *B. subtilis* bằng phương pháp MLSA với trình tự gen ghép nối *16S rRNA-rpoD.*
- BC-B0020 là *Bacillus licheniformis;* BC-B0027, BC-B0028, BC-B0116 là *Bacillus subtilis;* BC-B0029, BC-B0112 là *Bacillus amyloliquefaciens;* BC-B0037 là *Bacillus pumilus,* BC-B0039 là *Bacillus tequelensis*

KIẾN NGHỊ

- Khảo sát thêm các housekeeping gen khác để phân tích phát sinh loài bằng phương pháp MLSA

- So sánh kết quả phát sinh loài bằng phương pháp MLSA với dữ liệu toàn bộ genome

DANH MỤC CÔNG TRÌNH CỦA TÁC GIẢ

STT	Tên bài báo khoa học	Tác giả	Tên tạp chí, kỷ yếu, hội nghị	Năm công bố	Abstract/ Fulltext	Oral/ Poster
1	Phylogenetic analysis of *Bacillus* genus by multilocus sequencing analysis (MLSA) scheme	Nguyen Thi Thuy Tien, Nguyen Thi Thuy Duong, Trinh Thi Bich Huyen, Dang Vu Bich Hanh, Le Thi Huynh Tram	The International Workshop on Advanced Environmental Microbiology	2023	Abstract	Poster
2	Ứng dụng kỹ thuật MultiLocus Sequencing Analysis (MLSA) để phân biệt hai loài vi khuẩn *Bacillus subtilis* và *B. amyloliquefaciens*	Lê Thị Huỳnh Trâm, Nguyễn Thị Thủy Tiên, Trang Hoàng Long, Trần Thị Phấn, Đinh Anh Hòa, Đạo Nữ Diệu Hồng	Báo cáo tóm tắt Hội nghị Công nghệ sinh học toàn quốc năm 2022	2022	Abstract	Oral-poster
3	A Multi Locus Sequence Analysis Scheme for Phylogeny of the *Bacillus subtilis* species complex and its advantages over *16S rRNA* genes	Dao Nu Dieu Hong, Trang Hoang Long, Nguyen Thi Thuy Tien, Dinh Anh Hoa, Tran Thi Phan, Le Thi Huynh Tram	Ho Chi Minh City Open University Journal Of Science - Engineering And Technology, 12 (2) 2022, 22-33, ISSN: 2734-9330	2022	Fulltext	

DANH MỤC TÀI LIỆU THAM KHẢO

[1] C. Brady, I. Cleenwerck, S. Venter, M. Vancanneyt, J. Swings, and T. Coutinho, "Phylogeny and identification of *Pantoea* species associated with plants, humans and the natural environment based on multilocus sequence analysis (MLSA)," *Syst Appl Microbiol*, vol. 31, no. 6–8, pp. 447–460, Dec. 2008, doi: 10.1016/j.syapm.2008.09.004.

[2] C. L. Brady, S. N. Venter, I. Cleenwerck, K. Vandemeulebroecke, P. De Vos, and T. A. Coutinho, "Transfer of *Pantoea citrea*, *Pantoea punctata* and *Pantoea terrea* to the genus *Tatumella emend.* as *Tatumella citrea* comb. nov., *Tatumella punctata* comb. nov. and *Tatumella terrea* comb. nov. and description of *Tatumella morbirosei* sp. nov.," *Int J Syst Evol Microbiol*, vol. 60, no. 3, pp. 484–494, Mar. 2010, doi: 10.1099/IJS.0.012070-0.

[3] C. Brady, I. Cleenwerck, S. Venter, T. Coutinho, and P. De Vos, "Taxonomic evaluation of the genus Enterobacter based on multilocus sequence analysis (MLSA): Proposal to reclassify *E. nimipressuralis* and *E. amnigenus* into *Lelliottia* gen. nov. as *Lelliottia nimipressuralis* comb. nov. and *Lelliottia amnigena* comb. nov., respectively, *E. gergoviae* and *E. pyrinus* into *Pluralibacter* gen. nov. as *Pluralibacter gergoviae* comb. nov. and *Pluralibacter pyrinus* comb. nov.," *Syst Appl Microbiol*, vol. 36, no. 5, pp. 309–319, Jul. 2013, doi: 10.1016/j.syapm.2013.03.005.

[4] X. Rong and Y. Huang, "Multi-locus Sequence Analysis: Taking Prokaryotic Systematics to the Next Level," *Methods in Microbiology*, vol. 41, pp. 221–251, Jan. 2014, doi: 10.1016/BS.MIM.2014.10.001.

[5] H. Meghdadi, A. D. Khosravi, M. Hashemzadeh, and M. R. Tabandeh, "New design of multilocus sequence analysis of *rpoB, ssrA, tuf, atpE, ku*, and *dnaK* for identification of *Mycobacterium* species," *Mol Biol Rep*, vol. 49, no. 8, pp. 7967–7977, Aug. 2022, doi: 10.1007/S11033-022-07638-0.

[6] K. S. Ko *et al.*, "Population structure of the *Bacillus cereus* group as determined by sequence analysis of six housekeeping genes and the *plcR* gene," *Infect Immun*, vol. 72, no. 9, pp. 5253–5261, Sep. 2004, doi: 10.1128/IAI.72.9.5253-5261.2004/ASSET/85E0261D-F4D2-4106-A868-0464FD95D725/ASSETS/GRAPHIC/ZII0090442310003.JPEG.

[7] F. G. Priest, M. Barker, L. W. J. Baillie, E. C. Holmes, and M. C. J. Maiden, "Population structure and evolution of the *Bacillus cereus* group," *J Bacteriol*,

vol. 186, no. 23, pp. 7959–7970, Dec. 2004, doi: 10.1128/JB.186.23.7959-7970.2004.

[8] B. Soufiane, M. Baizet, and J.-C. Côté, "Multilocus sequence analysis of *Bacillus thuringiensis serovars navarrensis*, *bolivia* and *vazensis* and *Bacillus weihenstephanensis* reveals a common phylogeny," *Antonie Van Leeuwenhoek*, vol. 103, no. 1, pp. 195–205, Jan. 2013, doi: 10.1007/s10482-012-9800-5.

[9] Y. Liu *et al.*, "Phylogenetic diversity of the *Bacillus pumilus* group and the marine ecotype revealed by multilocus sequence analysis," *PLoS One*, vol. 8, no. 11, Nov. 2013, doi: 10.1371/JOURNAL.PONE.0080097.

[10] K. H. Schleifer, "Classification of Bacteria and Archaea: Past, present and future," *Syst Appl Microbiol*, vol. 32, no. 8, pp. 533–542, Dec. 2009, doi: 10.1016/j.syapm.2009.09.002.

[11] B. J. Tindall, R. Rosselló-Móra, H. J. Busse, W. Ludwig, and P. Kämpfer, "Notes on the characterization of prokaryote strains for taxonomic purposes," *Int J Syst Evol Microbiol*, vol. 60, no. Pt 1, pp. 249–266, 2010, doi: 10.1099/IJS.0.016949-0.

[12] E. Stackebrandt *et al.*, "Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology," *Int J Syst Evol Microbiol*, vol. 52, no. Pt 3, pp. 1043–1047, May 2002, doi: 10.1099/00207713-52-3-1043.

[13] J. He, "Phylogenetic methods," in *Encyclopedia of Microbiology*, Elsevier, 2019, pp. 538–550. doi: 10.1016/B978-0-12-801238-3.99224-9.

[14] N. Saitou and M. Nei, "The neighbor-joining method: a new method for reconstructing phylogenetic trees," *Mol Biol Evol*, vol. 4, no. 4, pp. 406–425, 1987, doi: 10.1093/OXFORDJOURNALS.MOLBEV.A040454.

[15] H. Kishino and M. Hasegawa, "Journal of Molecular Evolution Evaluation of the Maximum Likelihood Estimate of the Evolutionary Tree Topologies from DNA Sequence Data, and the Branching Order in Hominoidea," 1989.

[16] Z. Yang and B. Rannala, "Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo Method.," *Mol Biol Evol*, vol. 14, no. 7, pp. 717–724, Jul. 1997, doi: 10.1093/OXFORDJOURNALS.MOLBEV.A025811.

[17] W. M. Fitch, “Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology,” *Syst Biol*, vol. 20, no. 4, pp. 406–416, Dec. 1971, doi: 10.1093/SYSBIO/20.4.406.

[18] J. Minnaard, I. S. Rolny, and P. F. Pérez, “*Bacillus*”, *Laboratory Models for Foodborne Infections*, pp. 131–154, Jan. 1996, doi: 10.1201/9781315120089-8.

[19] “*Bacillus* sp., izolated from the ecosystem on nutritious substratum MPA... | Download Scientific Diagram.” https://www.researchgate.net/figure/Bacillus-sp-izolated-from-the-ecosystem-on-nutritious-substratum-MPA-Zivanovic-et-al_fig2_336116050 (accessed Apr. 12, 2023).

[20] D. R. Zeigler and J. B. Perkins, “The Genus *Bacillus*,” *Practical Handbook of Microbiology*, pp. 249–278, May 2021, doi: 10.1201/9781003099277-24.

[21] “Taxonomy of genus *Bacillus* (Source: Bergey’s manual of systematic... | Download Scientific Diagram.” https://www.researchgate.net/figure/Taxonomy-of-genus-Bacillus-Source-Bergeys-manual-of-systematic-bacteriology-volume-3_fig3_319046518 (accessed Apr. 11, 2023).

[22] W. Ludwig and H.-P. Klenk, “Overview: A Phylogenetic Backbone and Taxonomic Framework for Procaryotic Systematics,” *Bergey’s Manual® of Systematic Bacteriology*, pp. 49–65, 2001, doi: 10.1007/978-0-387-21609-6_8.

[23] M. C. J. Maiden, “Multilocus sequence typing of bacteria,” *Annu Rev Microbiol*, vol. 60, pp. 561–588, 2006, doi: 10.1146/ANNUREV.MICRO.59.030804.121325.

[24] T. Sawabe, K. Kita-Tsukamoto, and F. L. Thompson, “Inferring the Evolutionary History of *Vibrios* by Means of Multilocus Sequence Analysis,” *J Bacteriol*, vol. 189, no. 21, p. 7932, Nov. 2007, doi: 10.1128/JB.00693-07.

[25] E. Stackebrandt *et al.*, “Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology.,” *Int J Syst Evol Microbiol*, vol. 52, no. 3, pp. 1043–1047, May 2002, doi: 10.1099/00207713-52-3-1043.

[26] T. Tamura *et al.*, “A genome sequence-based approach to taxonomy of the genus *Nocardia*,” *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, vol. 102, no. 3, pp. 481–491, Oct. 2012, doi: 10.1007/S10482-012-9780-5/METRICS.

[27] X. T. Le *et al.*, "Exploration of genetic diversity of *Bacillus* spp. from industrial shrimp ponds in Vietnam by multi-locus sequence typing," *Fish Aquatic Sci*, vol. 22, no. 1, pp. 1–9, Aug. 2019, doi: 10.1186/S41240-019-0132-5/FIGURES/2.

[28] C. J. Joshi, W. Ke, A. Drangowska-Way, E. J. O'Rourke, and N. E. Lewis, "What are housekeeping genes?" *PLoS Comput Biol*, vol. 18, no. 7, Jul. 2022, doi: 10.1371/JOURNAL.PCBI.1010295.

[29] L. Zhang and W. H. Li, "Mammalian Housekeeping Genes Evolve More Slowly than Tissue-Specific Genes," *Mol Biol Evol*, vol. 21, no. 2, pp. 236–239, Feb. 2004, doi: 10.1093/MOLBEV/MSH010.

[30] Z. Tu, L. Wang, M. Xu, X. Zhou, T. Chen, and F. Sun, "Further understanding human disease genes by comparing with housekeeping genes and other genes," *BMC Genomics*, vol. 7, Feb. 2006, doi: 10.1186/1471-2164-7-31.

[31] C. R. Woese, O. Kandler, and M. L. Wheelis, "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya," *Proc Natl Acad Sci U S A*, vol. 87, no. 12, pp. 4576–4579, 1990, doi: 10.1073/PNAS.87.12.4576.

[32] D. Lane, "*16S/23S rRNA* sequencing." John Wiley & Sons, pp. 115–175, 1991.

[33] K. Fukuda, M. Ogawa, H. Taniguchi, and M. Saito, "Molecular approaches to studying microbial communities: Targeting the *16S ribosomal RNA* gene," *J UOEH*, vol. 38, no. 3, pp. 223–232, 2016, doi: 10.7888/JUOEH.38.223.

[34] M. M. S. M. Wösten, "Eubacterial sigma-factors," *FEMS Microbiol Rev*, vol. 22, no. 3, pp. 127–150, Sep. 1998, doi: 10.1111/j.1574-6976.1998.tb00364.x.

[35] T. M. Gruber and C. A. Gross, "Multiple sigma subunits and the partitioning of bacterial transcription space," *Annu Rev Microbiol*, vol. 57, pp. 441–466, 2003, doi: 10.1146/annurev.micro.57.030502.090913.

[36] J. D. Helmann and M. J. Chamberlin, "Structure and function of bacterial sigma factors," *Annu Rev Biochem*, vol. 57, pp. 839–872, 1988, doi: 10.1146/annurev.bi.57.070188.004203.

[37] "glpF - Glycerol uptake facilitator protein - *Escherichia coli* (strain K12) | UniProtKB | UniProt."

https://www.uniprot.org/uniprotkb/P0AER0/entry#function (accessed Apr. 12, 2023).

[38] "pta - Phosphate acetyltransferase - *Escherichia coli* (strain K12) | UniProtKB | UniProt." https://www.uniprot.org/uniprotkb/P0A9M8/entry (accessed Apr. 12, 2023).

[39] A. H. Klein, A. Shulla, S. A. Reimann, D. H. Keating, and A. J. Wolfe, "The intracellular concentration of acetyl phosphate in *Escherichia coli* is sufficient for direct phosphorylation of two-component response regulators," *J Bacteriol*, vol. 189, no. 15, pp. 5574–5581, Aug. 2007, doi: 10.1128/JB.00564-07/ASSET/E47DAAB2-D67E-41D4-B7ED-0594AB5BBF9C/ASSETS/GRAPHIC/ZJB0150770030006.JPEG.

[40] "Genomic organization of the PA5436-5435 (*pycAB*) operon and the PA5437... |Download Scientific Diagram." https://www.researchgate.net/figure/Genomic-organization-of-the-PA5436-5435-pycAB-operon-and-the-PA5437-pycR-gene-The_fig1_5253250 (accessed Apr. 12, 2023).

[41] Y. Zhang, M. Morar, and S. E. Ealick, "Structural biology of the purine biosynthetic pathway," *Cellular and Molecular Life Sciences*, vol. 65, no. 23, pp. 3699–3724, Nov. 2008, doi: 10.1007/S00018-008-8295-8/METRICS.

[42] D. W. Wolan, S. E. Greasley, M. J. Wall, S. J. Benkovic, and I. A. Wilson, "Structure of avian AICAR transformylase with a multisubstrate adduct inhibitor β-DADF identifies the folate binding site," *Biochemistry*, vol. 42, no. 37, pp. 10904–10914, Sep. 2003, doi: 10.1021/BI030106H.

[43] S. P. Glaeser and P. Kämpfer, "Multilocus sequence analysis (MLSA) in prokaryotic taxonomy," *Syst Appl Microbiol*, vol. 38, no. 4, pp. 237–245, Jun. 2015, doi: 10.1016/J.SYAPM.2015.03.007.

[44] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Res*, vol. 32, no. 5, pp. 1792–1797, 2004, doi: 10.1093/NAR/GKH340.

[45] J. Rozas and X. Messeguer, "DnaSP, DNA Polymorphism Analyses by the Coalescent and Other Methods Mammillaria Genome Sequencing View project Conservation Genomics applied to the Balearic Shearwater View project," 2014, doi: 10.1079/9780851994758.0139.

[46] J. Felsenstein, “Confidence limits on phylogenies: an approach using the bootstrap,” *Evolution*, vol. 39, no. 4, pp. 783–791, Jul. 1985, doi: 10.1111/J.1558-5646.1985.TB00420.X.

[47] R. H. Thomas, “Molecular Evolution and Phylogenetics,” *Heredity 2001 86:3*, vol. 86, no. 3, pp. 385–385, Mar. 2001, doi: 10.1046/j.1365-2540.2001.0923a.x.

[48] K. Tamura, G. Stecher, and S. Kumar, “MEGA11: Molecular Evolutionary Genetics Analysis Version 11,” *Mol Biol Evol*, vol. 38, no. 7, pp. 3022–3027, Jul. 2021, doi: 10.1093/MOLBEV/MSAB120.

[49] P. Kämpfer, S. P. Glaeser, M. W. Raza, S. A. Abbasi, and J. D. Perry, “*Pseudocitrobacter* gen. nov., a novel genus of the *Enterobacteriaceae* with two new species *Pseudocitrobacter faecalis* sp. nov., and *Pseudocitrobacter anthropi* sp. nov, isolated from fecal samples from hospitalized patients in Pakistan,” *Syst Appl Microbiol*, vol. 37, no. 1, pp. 17–22, Feb. 2014, doi: 10.1016/j.syapm.2013.08.003.

[50] P. Yarza *et al.*, “Uniting the classification of cultured and uncultured bacteria and archaea using *16S rRNA* gene sequences,” *Nat Rev Microbiol*, vol. 12, no. 9, pp. 635–645, 2014, doi: 10.1038/NRMICRO3330.

[51] S. Das, H. R. Dash, N. Mangwani, J. Chakraborty, and S. Kumari, “Understanding molecular identification and polyphasic taxonomic approaches for genetic relatedness and phylogenetic relationships of microorganisms,” *J Microbiol Methods*, vol. 103, pp. 80–100, 2014, doi: 10.1016/J.MIMET.2014.05.013.

[52] Y. Liu *et al.*, “Genomic insights into the taxonomic status of the *Bacillus cereus* group,” *Sci Rep*, vol. 5, Sep. 2015, doi: 10.1038/SREP14082.

[53] S. P. Glaeser and P. Kämpfer, “Multilocus sequence analysis (MLSA) in prokaryotic taxonomy,” *Syst Appl Microbiol*, vol. 38, no. 4, pp. 237–245, Jun. 2015, doi: 10.1016/j.syapm.2015.03.007.

[54] K. T. Konstantinidis, A. Ramette, and J. M. Tiedje, “The bacterial species definition in the genomic era,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 361, no. 1475, pp. 1929–1940, Nov. 2006, doi: 10.1098/RSTB.2006.1920.

[55] P. De Vos, “17 - Multilocus Sequence Determination and Analysis,” in *Methods in Microbiology*, F. Rainey and A. Oren, Eds., Academic Press, 2011, pp. 385–407. doi: https://doi.org/10.1016/B978-0-12-387730-7.00017-6.

[56] A. P. Rooney, N. P. J. Price, C. Ehrhardt, J. L. Sewzey, and J. D. Bannan, “Phylogeny and molecular taxonomy of the *Bacillus subtilis* species complex and description of *Bacillus subtilis subsp. inaquosorum subsp*. nov,” *Int J Syst Evol Microbiol*, vol. 59, no. Pt 10, pp. 2429–2436, 2009, doi: 10.1099/IJS.0.009126-0.

[57] B. Fan, J. Blom, H. P. Klenk, and R. Borriss, “*Bacillus amyloliquefaciens*, *Bacillus velezensis*, and *Bacillus siamensis* Form an ‘Operational Group *B. amyloliquefaciens*’ within the *B. subtilis* species complex,” *Front Microbiol*, vol. 8, no. JAN, p. 22, Jan. 2017, doi: 10.3389/FMICB.2017.00022/FULL.

[58] D. Gevers *et al.*, “Re-evaluating prokaryotic species,” *Nat Rev Microbiol*, vol. 3, no. 9, pp. 733–739, Sep. 2005, doi: 10.1038/NRMICRO1236.

[59] M. Mulet, J. Lalucat, and E. García-Valdés, “DNA sequence-based analysis of the *Pseudomonas* species,” *Environ Microbiol*, vol. 12, no. 6, pp. 1513–1530, Jun. 2010, doi: 10.1111/J.1462-2920.2010.02181.X.

[60] D. R. Zeigler, “Gene sequences useful for predicting relatedness of whole genomes in bacteria,” *Int J Syst Evol Microbiol*, vol. 53, no. Pt 6, pp. 1893–1900, Nov. 2003, doi: 10.1099/IJS.0.02713-0.

[61] P. Vinuesa, “Multilocus Sequence Analysis and Bacterial Species Phylogeny Estimation,” *Molecular Phylogeny of Microorganisms*, pp. 41–64, 2010, Accessed: Apr. 03, 2023. [Online]. Available: https://www.researchgate.net/publication/284383416_Multilocus_sequence_analysis_and_bacterial_species_phylogeny_estimation

[62] M. Martens, P. Dawyndt, R. Coopman, M. Gillis, P. De Vos, and A. Willems, “Advantages of multilocus sequence analysis for taxonomic studies: A case study using 10 housekeeping genes in the genus *Ensifer* (including former Sinorhizobium),” *Int J Syst Evol Microbiol*, vol. 58, no. 1, pp. 200–214, Jan. 2008, doi: 10.1099/IJS.0.65392-0.

[63] O. Y. Ampomah, S. A. Mousavi, K. Lindström, and K. Huss-Danell, “Diverse *Mesorhizobium* bacteria nodulate native *Astragalus* and *Oxytropis* in arctic and

subarctic areas in Eurasia,” *Syst Appl Microbiol*, vol. 40, no. 1, pp. 51–58, Jan. 2017, doi: 10.1016/J.SYAPM.2016.11.004.

10th April, 2023

Acceptance Letter

Dear Authors **Nguyen Thi Thuy Tien, Nguyen Thi Thuy Duong, Trinh Thi Bich Huyen, Dang Vu Bich Hanh, Le Thi Huynh Tram**
Biotechnology Center of Ho Chi Minh City
Abstract Topic: **Phylogenetic analysis of Bacillus genus by multilocus sequencing analysis (MLSA) scheme**

It is with great pleasure to inform that your abstract has been accepted as **Oral Presentation** in the International Workshop on Advanced Environmental Microbiology co-hosted by the Vietnamese – German University and Technical University of Darmstadt at VGU Binh Duong, Vietnam on April 24th, 2023.

The workshop aims to provide a space for sharing knowledge, experience, and thinking as well as a fostering the coalition among the researchers, practitioners, students from different backgrounds toward the environmental microbiology. We would appreciate it if you address our audience so they can benefit from your knowledge and experience.

We look forward to greeting you at the **Advanced Environmental Microbiology Workshop** at VGU Binh Duong on April 24th, 2023.

Yours sincerely,

On behalf of Workshop Organizer

Tran Le Luu, Assoc. Prof., PhD

Email: luu.tl@vgu.edu.vn

TRƯỜNG ĐẠI HỌC TÂY NGUYÊN

BÁO CÁO TÓM TẮT

HỘI NGHỊ CÔNG NGHỆ SINH HỌC TOÀN QUỐC 2022

ABSTRACT

OF VIETNAM NATIONAL COFERENCE ON BIOTECHNOLOGY 2022

Đắk Lắk, ngày 4 tháng 11 năm 2022

MỤC LỤC

LỜI NÓI ĐẦU i

BAN TỔ CHỨC HỘI NGHỊ iii

BAN BIÊN TẬP iv

CÁC TIỂU BAN CHUYÊN MÔN v

CHƯƠNG TRÌNH HỘI THẢO vii

CHƯƠNG TRÌNH BÁO CÁO CÁC TIỂU BAN CHUYÊN MÔN x

TÓM TẮT ORAL

I. TIỂU BAN CÔNG NGHỆ SINH HỌC GENE VÀ PROTEIN

OR-GP01. BIỂU HIỆN VÀ TINH SẠCH TIỂU PHẦN N CỦA PROTEIN GLYPICAN 3 (GPC3-Nsub) DẠNG THỂ VÙI TRÊN *Escherichia coli*.

Ngô Lương Đăng Thức, Hoàng Ngọc Phương Mai, Nguyễn Đăng Quân, Võ Nguyễn Thanh Thảo 2

OR-GP02. SỬA LỖI ADN: TỪ VI KHUẨN ĐẾN NGƯỜI.

Hồ Ngọc Hân, Antoine van Oijen, Harshad Ghodke and Stephen West 3

OR-GP03. NGHIÊN CỨU TỔNG HỢP VÀ CỐ ĐỊNH LIÊN TỤC BETA-MANNANASE TRÊN BỀ MẶT *Lactobacillus plantarum* TRONG ĐIỀU KIỆN ĐIỀU CHỈNH pH.

Nguyễn Hoàng Minh, Nguyễn Thu Hà 4

OR-GP04. KỶ NGUYÊN MỚI CỦA XÉT NGHIỆM DI TRUYỀN TIỀN LÀM TỔ: PHƯƠNG PHÁP KHÔNG XÂM LẤN SỬ DỤNG DNA TỰ DO ĐƯỢC PHÓNG THÍCH TỪ PHÔI.

Đỗ Thị Linh, Nguyễn Ngọc Bích, Giang Hoa 5

OR-GP05. ĐẶC ĐIỂM HỆ GEN TY THỂ VÀ ĐƠN VỊ MÃ HÓA RIBOSOME CỦA CÁC LOÀI SÁN LÁ PHỔI THUỘC HỌ *Paragonimidae* Ở CHÂU Á.

Nguyễn Thị Khuê, Đỗ Thị Roan, Lê Thị Kim Xuyến, Phạm Thị Khánh Linh, Đoàn Thị Thanh Hương, Lê Thanh Hòa 7

OR-GP06. NGHIÊN CỨU THU NHẬN LYSOZYME TỪ LÒNG TRẮNG TRỨNG GÀ SỬ DỤNG HỆ MIXEN ĐẢO AOT/ISOOCTANE.

Phạm Thị Minh Hải, Trương Lê Na, Lê Cao Kỳ 8

OR-GP07. NHÂN DÒNG VÀ BIỂU HIỆN 3 CHYMOTRYPSIN-LIKE PROTEASE CỦA SARS-COV-2 Ở *E. coli* BẰNG VECTOR PET28A.

Nguyễn Thị Hồng Loan, Vũ Thị Duy Ly, Trần Thị Xuân Giang, Ngô Thị Trang, Nguyễn Thị Như Quỳnh, Phạm Thị Xuân Diệu, Lê Thế Thái, Đinh Nho Thái, Phan Tuấn Nghĩa 9

OR-GP08. TẠO DÒNG VI KHUẨN *Bacillus subtilis* MANG GEN MÃ HOÁ UREASE A TỪ CHỦNG LÂM SÀNG *Helicobacter pylori.*

Nguyễn Thị Kim Cúc, Phạm Thị Lan, Phạm Thu Thủy, Nguyễn Văn Duy 10

OR-GP09. KNOCKOUT GEN NRF2 CÁ NGỰA VĂN BẰNG CRISPR–Cas9: TIẾT LỘ CON ĐƯỜNG KHÁNG OXY HÓA CỦA CÁC LOẠI CAO CHIẾT THẢO DƯỢC.

Nguyễn Thành Vũ, Nguyễn Thị Tâm Thanh, Vũ Hoàng Huy, Lưu Thu Hiền, Lê Lưu Phương Hạnh, Nguyễn Phương Phi, Dương Nguyễn Mai Quế 11

OR-GP010. CÔNG NGHỆ GIẢI TRÌNH TỰ ĐOẠN DÀI - LONG READ SEQUENCING VÀ ỨNG DỤNG TRONG NGHIÊN CỨU.

Phạm Thanh Tâm 12

OR-GP11. KHẢO SÁT SỰ CÓ MẶT CỦA GEN $pirA^{VP}$/$pirB^{VP}$ VÀ ĐÁNH GIÁ TÍNH KHÁNG KHÁNG SINH CỦA *Vibrio parahaemolyticus* PHÂN LẬP TỪ TÔM THẺ NUÔI TRONG KHU VỰC TỈNH KHÁNH HÒA.

Lê Mai Nhả, Hoàng Ngô Minh Khôi, Nguyễn Thị Anh Thư, Văn Hồng Cầm 13

II. TIỂU BAN CÔNG NGHỆ SINH HỌC TẾ BÀO

OR-TB01. ĐÁNH GIÁ TÁC DỤNG CHỐNG LÃO HÓA CỦA SERUM NỀN DẦU DỪA KẾT HỢP VỚI CHIẾT XUẤT TẾ BÀO GỐC NHUNG HƯƠU TRÊN MÔ HÌNH CHUỘT LÃO HÓA DA.

Phạm Lê Bửu Trúc, Trương Thị Thuý, Nguyễn Thị Thương Huyền, Lao Đức Thuận, Nguyễn Trọng Bình, Nguyễn Đăng Quân..... 14

OR-TB02. SỰ BIỂU HIỆN HOẠT TÍNH ALKALINE PHOSPHATASE Ở TẾ BÀO NANG TÓC NGƯỜI.

Văn Đức Huy, Đặng Thị Tùng Loan, Nguyễn Thị Ngọc Như, Nguyễn Văn Ngọc Thuận, Đào Thị Thanh Thuỷ, Phan Văn Tiến, Lê Quang Trí..... 15

OR-TB03. NANO KIM LOẠI KHẮC PHỤC MỘT SỐ HIỆN TƯỢNG BẤT THƯỜNG VÀ GIA TĂNG CHẤT LƯỢNG CÂY HOA ĐỒNG TIỀN (*Gerbera jamesonii*) NUÔI CẤY IN VITRO*.*

Hoàng Thanh Tùng, Phan Lê Hà Nguyễn, Trần Văn Lịch, Hà Thị Mỹ Ngân, Đỗ Mạnh Cường, Vũ Quốc Luận, Hoàng Đắc Khải, Nguyễn Thị Như Mai, Bùi Văn Thế Vinh, Dương Tấn Nhựt 16

OR-TB04. THEO DÕI HÀM LƯỢNG ESTRADIOL, PROGESTERONE CỦA DÊ BÁCH THẢO VÀ DÊ BOER TRONG GIAI ĐOẠN MANG THAI.

Nguyễn Thị Thu Hiền, Nguyễn Thanh Bình..... 17

OR-TB05. VI NHÂN GIỐNG CÂY DỪA CAO TA XANH (*Cocos nucifera* L.) TỪ MÔ TẾ BÀO PHÁT HOA NON.

Trần Bình Minh, Nguyễn Thiên Quang, Hoàng Thị Lan Xuân, Nguyễn Phương Thảo..... 18

OR-TB06. ĐÁNH GIÁ HOẠT TÍNH KHÁNG NẤM CỦA DỊCH CHIẾT QUẢ BỒ HÒN *(Sapindus mukorossi)* TRONG ĐIỀU KIỆN IN VITRO.

Đinh Trường Sơn, Nông Thị Huệ, Trần Bảo Lâm, Vũ Thanh Hiền, Đặng Thị Thanh Tâm..... 19

OR-TB07. XÂY DỰNG HỆ GEN TI THỂ CỦA CÁ SỬU *Boesemania microlepis* VÀ KHẢO SÁT MỐI QUAN HỆ PHÁT SINH LOÀI CỦA HỌ CÁ LÙ ĐÙ (*Sciaenidae*).

Vũ Đặng Hạ Quyên, Trần Quang Sáng, Trương Thị Oanh, Mai Tùng Quân, Đặng Thúy Bình 20

OR-TB08. NGHIÊN CỨU KHẢ NĂNG KHÁNG BỆNH HÉO RŨ GỐC MỐC TRẮNG CỦA CÂY LẠC (*Arachis hypogaea*) ĐƯỢC CHUYỂN GEN *Chi42.*

Phùng Thị Bích Hòa, Nguyễn Hoàng Tuệ, Nguyễn Ngọc Lương, Lê Thị Hà Thanh, Nguyễn Quang Đức Tiến, Nguyễn Xuân Huy, Nguyễn Hoàng Lộc 21

OR-TB09. HOẠT TÍNH ỨC CHẾ TẾ BÀO UNG THƯ GAN HEPG2 CỦA SeNPs/OLIGOFUCOIDAN CHẾ TẠO BẰNG PHƯƠNG PHÁP CHIẾU XẠ TIA GAMMA (Co-60).

Nguyễn Văn Linh, Trần Đức Trọng, Nguyễn Trọng Nghĩa, Nguyễn Thanh Vũ, Lê Quang Luân 22

OR-TB10. KHẢO SÁT HÀM LƯỢNG FLAVONOID, ALKALOID VÀ HOẠT TÍNH KHÁNG OXY HÓA CỦA CAO CHIẾT RỄ TƠ CÂY BÁ BỆNH *(Eurycoma longifolia* Jack*)* Ở CÁC PHA SINH TRƯỞNG KHÁC NHAU.

Phan Tường Lộc, Nguyễn Huỳnh Cẩm Tú, Đỗ Thị Kim Loan, Hoàng Văn Dương, Trần Thị Ngọc Hà..... 23

III. TIỂU BAN CÔNG NGHỆ SINH HỌC NÔNG NGHIỆP

OR-NN01. TẠO DÒNG, BIỂU HIỆN, VÀ TÁI GẤP CUỘN TOLL-LIKE RECEPTOR 22 TỪ CÁ TRA *Pangasianodon hypophthalmus.*

Mai Quốc Gia, Lê Thị Xuân Trang, Trần Văn Hiếu..... 24

OR-NN02. ĐÁNH GIÁ MỘT SỐ CHỈ THỊ PHÂN TỬ LIÊN QUAN ĐẾN HÀM LƯỢNG KẼM VÀ SẮT THẤP TRONG HẠT Ở MỘT SỐ GIỐNG LÚA (*Oryza sativa* L.).

Trần Hà Linh, Đỗ Minh Ân, Lê Ngọc Diệp, Lê Trần Tuyết Mai, Lại Thị Yến, Lê Quỳnh Mai 25

OR-NN03. NGHIÊN CỨU BIỂU HIỆN ĐẶC THÙ CỦA CÁC GEN LẶP MÃ HÓA TREHALOSE-6-PHOSPHATE PHOSPHATASE TRONG GIAI ĐOẠN SINH TRƯỞNG VÀ PHÁT TRIỂN CỦA CÂY LẠC (*Arachis hypogaea*).

Trần Thị Hải Yến, Vũ Trung Đức, Nguyễn Toàn Thắng, Lê Thị Ngọc Quỳnh, Tống Văn Hải, Chu Đức Hà, Trần Thị Thanh Huyền 26

OR-NN04. ĐÁNH GIÁ SỰ LÂY NHIỄM CỦA Deformed wing virus VÀ Sacbrood virus GÂY BỆNH TRÊN ONG MẬT *Apis cerana* Ở MỘT SỐ TỈNH THÀNH TẠI VIỆT NAM.

Bùi Thị Thùy Dương, Phạm Thị Lành, Hà Thị Thu, Nguyễn Thị Hoa, Nguyễn Đình Duy, Mẫn Hồng Phước, Phạm Thị Phương Thảo, Nguyễn Thị Lan Anh , Phạm Hồng Thái, Đồng Văn Quyền 27

OR-NN05. ẢNH HƯỞNG CỦA ĐIỀU KIỆN SẤY PHUN ĐẾN TÍNH CHẤT HÓA LÝ VÀ KHẢ NĂNG KHÁNG OXY HÓA CỦA DỊCH ÉP TỪ THÂN LÚA NON.

Phan Thị Ngọc Hạnh, Trần Thanh Thảo Nguyên, Đặng Minh Hiền, Nguyễn Tấn Hùng..... 28

OR-NN06. SÀNG LỌC VÀ NGHIÊN CỨU ĐẶC ĐIỂM SINH HỌC CỦA CHỦNG XẠ KHUẨN CÓ KHẢ NĂNG ĐỐI KHÁNG VỚI MỘT SỐ NẤM GÂY BỆNH TRÊN CÂY TRỒNG.

Nguyễn Xuân Cảnh, Nguyễn Thị Thanh Mai, Nguyễn Thanh Hải, Nguyễn Thị Thu, Trần Thị Đào..... 29

OR-NN07. ĐIỀU TRA, ĐÁNH GIÁ VÀ GIẢI PHÁP QUẢN LÝ SỰ Ô NHIỄM VI SINH VẬT NGUY HẠI Ở ỚT SAU THU HOẠCH.

Nguyễn Bảo Hưng, Võ Thị Ngọc Trai, Hoàng Tấn Quảng, Nguyễn Đức Huy, Kim Se-Ri 30

OR-NN08. QUY TRÌNH CÔNG NGHỆ SẢN XUẤT MANNO-OLIGOSACCHARIDES, MANNOSE, VÀ BIOETHANOL TỪ BÃ CÀ PHÊ.

Nguyễn Quỳnh Anh, Trịnh Thị Phi Ly, Hyeun-Jong Bae..... 31

OR-NN09. ỨNG DỤNG KỸ THUẬT GIẢI TRÌNH TỰ THẾ HỆ MỚI TRONG NGHIÊN CỨU BỆNH HỌC THỦY SẢN Ở TRUNG TÂM CÔNG NGHỆ SINH HỌC THÀNH PHỐ HỒ CHÍ MINH.

Ngô Huỳnh Phương Thảo, Vũ Thị Thanh Hương, Lê Thị Thu Thảo, Nguyễn Hoàng Thụy Vy, Bùi Nguyễn Chí Hiếu, Nguyễn Đăng Quân..... 32

OR-NN10. MỘT SỐ CHỦNG VI KHUẨN NỘI SINH RỄ LÚA NƯỚC CÓ KHẢ NĂNG ỨC CHẾ SINH TRƯỞNG *in vitro* CỦA NẤM ĐẠO ÔN *Magnaporthe oryzea* VÀ NẤM *Fusarium* SPP.

Nguyễn Văn Phượng, Nguyễn Phạm Cẩm Trà, Nguyễn Thị Hà Trang, Nguyễn Khánh Linh, Tăng Khánh Linh, Lê Quỳnh Liên, Nguyễn Thị Hoan, Khúc Khánh Linh, Mai Thị Phương Nga, Lê Thị Vân Anh, Phạm Hoàng Nam, Nguyễn Quang Huy..... 33

OR-NN11. ẢNH HƯỞNG CỦA ĐIỀU KIỆN SẤY ĐẾN TÍNH CHẤT HÓA LÝ VÀ HỢP CHẤT SAPONINS CỦA RỄ CÂY DƯỢC LIỆU SÂM CAU (*Curculigo orchioides*).

Nguyễn Văn Tặng, Trần Thanh Giang, Nguyễn Thị Vân, Mai Quốc Kiệt, Nguyễn Thị Minh Thương, Trần Thị Thúy An, Nguyễn Văn Hòa 34

IV. TIỂU BAN CÔNG NGHỆ SINH HỌC Y DƯỢC

OR-YD01. CẢI TIẾN BỀ MẶT NẤM MEN *Saccharomyces cerevisiae* ĐỂ LÀM THỂ MANG VACCINE ĐƯỜNG UỐNG.

Huỳnh Thị Thu Hà, Trần Quý Đức, Hồ Thị Len, Nguyễn Ngọc Lương 35

OR-YD02. CAO CHIẾT LÁ ĐẮNG *Vernonia amygdalina* Del. ỨC CHẾ CON ĐƯỜNG DẪN TRUYỀN TÍN HIỆU CỦA FLT3 VÀ CẢM ỨNG APOPTOSISTRÊN TẾ BÀO BẠCH CẦU CẤP DÒNG TỦY.

Nguyễn Trung Quân, Phan Thị Minh Tâm, Hoàng Kim Sơn, Bùi Thị Kim Lý, Hoàng Thành Chí 36

OR-YD03. PHÁT HIỆN SỚM *Streptococcus* NHÓM B BẰNG REAL-TIME PCR VÀ SỬ DỤNG KHÁNG SINH DỰ PHÒNG Ở THAI PHỤ KHI CHUYỂN DẠ.

Hà Mạnh Tuấn, Trần Thị Bích Huyền, Bùi Thị Kim Thảo, Nguyễn Tuấn Anh..... 37

OR-YD04. SỰ TƯƠNG TÁC GIỮA PROTEIN TIF-90 VÀ GTP THÚC ĐẨY QUÁ TRÌNH SINH TỔNG HỢP RNA RIBOSOME VÀ LÀM TĂNG CÁC HOẠT ĐỘNG ÁC TÍNH CỦA TẾ BÀO UNG THƯ RUỘT KẾT NGƯỜI.

Nguyễn Đăng Quân , Hoàng Đình Hoà, Michael Nelson, Lokesh Nigam, Võ Nguyễn Thanh Thảo, Lianjun Zhang, Phạm Thị Kim Trâm, Hồ Hữu Đức, Nguyễn Thị Đại Đồng, Lâm Trung Quốc, Tất Tố Trinh, Yasmin Elhajmoussa, Lý Quốc Trung, Flavia Pichiorri, Vinod Pullarkat, Bin Zhang, Ya-Huei Kuo, Guido Marcucci, Nguyễn Lê Xuân Trường..... 38

OR-YD05. NGHIÊN CỨU HỆ VI SINH VẬT (MICROBIOME) THEO HƯỚNG PHÁT TRIỂN PROBIOTICS, BACTERIOCIN VÀ VẮC XIN.

Nguyễn Văn Duy 40

OR-YD06. PHÁT TRIỂN ĐIỂM NGUY CƠ METHYL HÓA DNA COVID-19 (COMER) CHO VIỆC PHÂN LOẠI BỆNH NHÂN NẶNG.

Trần Bá Thiên, Trần Nguyễn Trọng Phú, Trần Thị Thanh Khương và Lưu Phúc Lợi 41

OR-YD07. ẢNH HƯỞNG CỦA YẾU TỐ VÔ SINH NAM NẶNG ĐẾN TÌNH TRẠNG LỆCH BỘI NHIỄM SẮC THỂ PHÔI Ở NHÓM BỆNH NHÂN TRẺ TUỔI CÓ CHỈ ĐỊNH PGT-A DO SẨY THAI LIÊN TIẾP CHƯA RÕ NGUYÊN NHÂN.

Lưu Thị Minh Tâm, Nguyễn Thị Cẩm Nhung, Nguyễn Thị Minh Anh, Phan Thị Kim Anh, Trần Tú Cầm, Nguyễn Cao Trí, Hồ Mạnh Tường..... 43

OR-YD08. SỬ DỤNG CÔNG NGHỆ MASTER MIX SẤY KHÔ CHO NGHIÊN CỨU PHÁT TRIỂN KIT REALTIME PCR.

Mai Hoàng Khánh Chi, Ngô Hồng Gấm, Nguyễn Duy Khánh..... 44

OR-YD09. ASCORBIC ACID NHẮM ĐÍCH CÁC TẾ BÀO GỐC UNG THƯ DẠ DÀY.

Nguyễn Phú Hùng..... 45

OR-YD10. TIỀM NĂNG CHỐNG ĐÁI THÁO ĐƯỜNG CỦA CAO CHIẾT VỎ THÂN TRÂM VỎ ĐỎ (*Syzygium zeylanicum* (L.).

Nguyễn Minh Trung, Bùi Thị Bích Huyên, Nguyễn Quang Vinh..... 46

V. TIỂU BAN CÔNG NGHỆ SINH HỌC VI SINH VÀ MÔI TRƯỜNG

OR-VM01. KHẢO SÁT KHẢ NĂNG PHÂN HỦY NHÓM THUỐC TRỪ NẤM TRIAZOLE CỦA MỘT SỐ CHỦNG VI KHUẨN PHÂN LẬP ĐƯỢC TỪ ĐẤT CANH TÁC NÔNG NGHIỆP TẠI ĐÀ LẠT.

Lương Thị Thắm, Nguyễn Thị Hồng Thắm 47

OR-VM02. NGHIÊN CỨU LÊN MEN PHỤ PHẨM NÔNG NGHIỆP ĐỂ TẠO HOẠT CHẤT PRODIGIOSIN, NANO/MICRO-PRODIGIOSIN ỨNG DỤNG KIỂM SOÁT SINH HỌC TRONG NÔNG NGHIỆP.

Nguyễn Văn Bốn, San-Lang Wang, Phan Tứ Quý, Nguyễn Thị Huyền, Nguyễn Thị Hạnh, Nguyễn Đại Nam, Trần Thị Hà Trang, Đoàn Mạnh Dũng, Hồ Nhật Được, Ngô Văn Anh và Nguyễn Anh Dũng 48

OR-VM03. PHÂN LẬP VÀ ỨNG DỤNG CỦA VI KHUẨN TÍA QUANG HỢP TRONG XỬ LÝ NƯỚC THẢI DỆT NHUỘM.

Nguyễn Tiến Đạt, Đỗ Thị Liên, Trần Thị Huyền Nga, Nguyễn Mạnh Khải, Cung Thị Ngọc Mai, Nguyễn Trọng Gia Khánh, Lê Thị Nhi Công..... 49

OR-VM04. TIỀM NĂNG KHÁNG KHUẨN VÀ PROBIOTIC CỦA VI KHUẨN *Lacticaseibacillus rhamnosus* TNU_BD001 PHÂN LẬP TỪ NEM CHUA TẠI THÀNH PHỐ BUÔN MA THUỘT.

Trần Thị Ngọc, Huỳnh Thị Hoa, Wang San-Lang, Đoàn Chiến Thắng 50

OR-VM05. PHÂN LẬP VÀ SÀNG LỌC MỘT SỐ CHỦNG VI KHUẨN LACTIC CÓ KHẢ NĂNG LOẠI BỎ CHOLESTEROL.

Nguyễn Văn Toàn, Nguyễn Thị Liễu, Nguyễn Thị Thùy Trang, Huỳnh Lâm Diễm My, Trần Thị Phương Nhi, Đạo Nữ Diệu Hồng, Phạm Hải Sơn, Nguyễn Thị Dung. 51

OR-VM06. NGHIÊN CỨU PHÁT TRIỂN CHẾ PHẨM VI KHUẨN *Bacillus velezensis* ĐỐI KHÁNG VỚI NẤM *Cercospora arachidicola* GÂY BỆNH ĐỐM NÂU TRÊN LẠC.

Lê Thanh Khang và Trần Yên Thảo 52

OR-VM07. KHẢO SÁT VI SINH VẬT TIỀM NĂNG ỨNG DỤNG TRONG CẢI TẠO NỀN ĐẤT SỬ DỤNG MÔ HÌNH HẠT THỦY TINH.

Nguyễn Phạm Hương Huyền, Nguyễn Ngọc Trí Huỳnh, Nguyễn Khánh Sơn, Phạm Minh Tuấn 53

OR-VM08. ỨNG DỤNG KỸ THUẬT MULTILOCUS SEQUENCING ANALYSIS (MLSA) ĐỂ PHÂN BIỆT HAI LOÀI VI KHUẨN *Bacillus subtilis* VÀ *B. amyloliquefaciens.*

Lê Thị Huỳnh Trâm, Nguyễn Thị Thủy Tiên, Trang Hoàng Long, Trần Thị Phấn, Đinh Anh Hòa, Đạo Nữ Diệu Hồng..... 54

OR-VM08. ỨNG DỤNG KỸ THUẬT MULTILOCUS SEQUENCING ANALYSIS (MLSA) ĐỂ PHÂN BIỆT HAI LOÀI VI KHUẨN *Bacillus subtilis* VÀ *B. amyloliquefaciens*

Lê Thị Huỳnh Trâm[1,2*], Nguyễn Thị Thủy Tiên[2], Trang Hoàng Long[3], Trần Thị Phấn[2], Đinh Anh Hòa[2], Đạo Nữ Diệu Hồng[2]

[1].Khoa Môi trường và Tài nguyên, Trường Đại học Bách khoa - ĐH Quốc gia TP.HCM

[2].Trung tâm Công nghệ sinh học Thành phố Hồ Chí Minh

[3]. Khoa Công nghệ Sinh học, Đại học Quốc tế - ĐH Quốc gia TP.HCM

TÓM TẮT

Bacillus subtilis và *B. amyloliquefaciens* thuộc nhóm loài *Bacillus subtilis*. Sự tương đồng về đặc điểm hình thái, sinh hóa và trình tự 16S rRNA ở hai loài này là rất cao, lên đến 99%. Nghiên cứu này được thực hiện nhằm ứng dụng kỹ thuật Multilocus Sequence Analysis (MLSA) dựa trên sự kết hợp của trình tự của 5 gen giữ nhà (housekeeping genes): *glpF, pta, purH, pycA* và *rpoD* để phân tích phát sinh loài của 2 loài *B. subtilis* và *B. amyloliquefaciens.* Sau khi khuếch đại bằng phản ứng PCR và giải trình tự 5 gen, cây phát sinh loài được xây dựng từ trình tự ghép nối của 5 gen giữ nhà và so sánh với cây phát sinh loài được xây dựng chỉ dựa trên trình tự vùng16S rRNA. Kết quả cho thấy rằng, cây phát sinh loài được xây dựng bằng kỹ thuật MLSA có sự phân nhánh rõ rệt hơn giữa *B. subtilis* và *B. amyloliquefaciens*, các nhánh cũng dài hơn và có giá trị tin cậy cao hơn so với cây phát sinh loài từ 16S rRNA. Các loài khác thuộc chi *Bacillus* cũng có sự tách biệt rõ hơn. Sử dụng kỹ thuật MLSA để xây dựng cây phát sinh loài đã đưa ra một cái nhìn mới về mối quan hệ về tiến hóa trong nhóm loài *B. subtilis*. Bởi vì sự phân tách giữa các nhóm loài rõ hơn, một số loài thuộc chi *Bacillus* dường như có mối quan hệ khác nhau trong quá trình tiến hóa, chẳng hạn như *B. atrophaeus* gần với *B. subtilis* hơn *B. amyloliquefaciens* hoặc *B. velezensis* hiện thuộc các nhánh khác với *B. amyloliquefaciens*.

Từ khóa: *Bacillus amyloliquefaciens, Bacillus subtilis*, housekeeping gene, Multilocus Sequence Analysis, phát sinh loài.

CREATE A NEW MULTI LOCUS SEQUENCE ANALYSIS (MLSA) SCHEME IN DISTINGUISHING THE TWO BACTERIA *Bacillus subtilis* AND *B. amyloliquefaciens*

Le Thi Huynh Tram[1,2*] , Nguyen Thi Thuy Tien[2], Trang Hoang Long[3], Tran Thi Phan[2], Dinh Anh Hoa[2], Dao Nu Dieu Hong[2]

[1].Faculty of Environment and Natural Resources, Ho Chi Minh City University of Technology - VNU HCM

[2].Biotechnology Center of Ho Chi Minh City

3.School of Biotechonology, International University - VNU HCM

SUMMARY

Bacillus subtilis and *B. amyloliquefaciens* belong to the subgenus, namely *Bacillus subtilis* species complex. The colonical morphology, biochemical characteristics and 16S rRNA sequences of the two species are more than 99 percent indentical to each other. In this study, we created the MLSA scheme based on the concatenation of the five protein-coding housekeeping genes, *glpF, pta, purH, pycA*, and *rpoD*, to generate phylogentic tree and subsequently deduce phylogenies. After PCR amplification and sequencing, phylogenetic tree from the five-concatenated sequences is constructed and compared with 16S rRNA gene-based phylogenetic tree. The results show that the phylogenetic tree created by MLSA scheme displays higher resolution between the two species, *B. subtilis* and *B. amyloliquefaciens*, the phylogenetic branches are also longer and supported by much higher bootstrap values compared to the 16S rRNA gene-based phylogenetic tree. Other *Bacillus* species were also obviously distinguished. Using MLSA scheme to generate phylogentic tree has given a new look for evolution relationship in the *B. subtilis* species complex. Because of the more discrimination between *Bacillus* species, some species seem to have different relationship, such as *B. atrophaeus* are closer to *B. subtilis* than *B. amyloliquefaciens* or *B. velezensis* are belong other branches with *B. amyloliquefaciens*.

Keywords: *Bacillus amyloliquefaciens, Bacillus subtilis*, housekeeping gene, Multilocus Sequence Analysis, phylogeny.

* Author for correspondence: Tel: +84-911471646; Email: lthtram@hcmut.edu.vn

A multi locus sequence analysis scheme for phylogeny of the *Bacillus subtilis* species complex and its advantages over 16S rRNA genes

Dao Nu Dieu Hong[1*], Trang Hoang Long[2], Nguyen Thi Thuy Tien[1], Dinh Anh Hoa[1], Tran Thi Phan[1], Le Thi Huynh Tram[1,2]

[1]Biotechnology Center of Ho Chi Minh City, Ho Chi Minh City, Vietnam
[2]International University, Vietnam National University-HCMC, Vietnam
[*]Corresponding author: dieuhong1791@gmail.com

ARTICLE INFO

DOI:10.46223/HCMCOUJS.tech.en.12.2.2255.2022

Received: April 25th, 2022

Revised: May 18th, 2022

Accepted: May 27th, 2022

Keywords:

a Multi Locus Sequence Analysis (MLSA approach); *Bacillus subtilis;* housekeeping genes; 16S rRNA

ABSTRACT

A multi locus sequence analysis (MLSA approach) was studied on the *Bacillus* genus, or the *Bacillus subtilis* species complex for specific, including 08 strains from four species (*B. subtilis, B. pumilus, B. licheniformis* and *B. amyloliquefaciens*) were provided by Biotechnology Center of Ho Chi Minh City. The research was based on sequences of 16S rRNA genes, the concatenation of five protein-coding housekeeping genes: *glpF, pta, purH, pycA,* and *rpoD*. After PCR amplification and sequencing the phylogenetic tree of 16S rRNA sequences, concatenate sequences (as well as the phylogenetic tree of each housekeeping gene) are constructed for comparison and discussion. The aim of this study is reach for better resolution and differentiation of strains and species within the *B. subtilis* species and to determine whether MLSA scheme show advantages in 16S rRNA gene-based studies.

1. Introduction

The genus *Bacillus* was named in 1835 by Christian Gottfried Ehrenberg, to contain rod-shaped (bacillus) bacteria (Ehrenberg, 1835). *Bacillus* was later amended by Ferdinand Cohn to further describe them as spore-forming, Gram-positive, aerobic, or facultatively anaerobic bacteria (Cohn, 1872). Since identified, *Bacillus* species have been studied worldwide for their uses in many medical, pharmaceutical, agricultural, and industrial processes that take advantage of their wide range of physiology characteristics and capacity to produce a variety of enzymes, antibiotics, and other metabolites (Turnbull & Baron, 1966). At the time of writing, there are 318 species have been named in total, with *Bacillus subtilis* as the type species (Skerman, McGowan, & Sneath, 1980). The *Bacillus subtilis* species complex is a group that presents phylogenetically and phonetically homogenous species, consisting of the important species *B. subtilis*, *B. licheniformis*, and *B. pumilus* (Fritze, 2004). Besides the three original species has been mentioned above, which were described in 1973 (Gordon, Haynes, & Pang, 1973), many novel species have been identified that belong to the *B. subtilis* species complex, which is: *B. atrophaeus* (Nakamura, 1989), *B. mojavensis* (Roberts, Nakamura, & Cohan, 1994), *B. vallismortis* (Roberts et al., 1994), *B. sonorensis* (Palmisano, Nakamura, Duncan, Istock, & Cohan, 2001), *B. velezensis* (Ruiz-García, Béjar, Martinez-Checa, Llamas, & Quesada, 2005), *B.*

tequilensis (Gatson et al., 2006), *B. halotolerans* (Bacon & Hinton, 2002) and *B. amyloliquefacien* (Priest, Goodfellow, Shute, & Berkeley, 1987). Also, in 1999 (Nakamura, Roberts, & Cohan, 1999) *B. subtilis* was subdivided into two subspecies: *B. subtilis* subsp. *subtilis, B. subtilis* subsp. *spizizenii* and in 2009 (Rooney, Price, Ehrhardt, Swezey, & Bannan, 2009), *B. subtilis* subsp. *inaquosorum* has been classified as the third subspecies of *B. subtilis*. For the *B. subtilis* species complex, it has been recognized that these species are hard to distinguish on the basis of traditional phenotypic methods for many years.

For over 40 years, in perspective of phylogenetic analyses of bacteria, the 16S rRNA gene is most popular genetic marker for identification of bacteria (Yarza et al., 2014). Liu's study that we use 16S rRNA gene sequence is unable to distingguish different strains (Liu et al., 2013).

The Multi Locus Sequence Analysis (MLSA) method has been increasingly used in phylogeny and taxonomy (Gevers et al., 2005) (Peeters et al., 2016). In general, MLSA concatenates the sequence of gene fragments of the housekeeping genes and utilize this concatenated sequence to determine phylogenetic relationships.

Many housekeeping genes are *rpoD* (RNA polymerase sigma factor), *pycA* (pyruvate carboxylase), *pta* (phosphate acetyltransferase), *glpF* (glycerol uptake facilitator) and *purH* (phosphoribosyl aminoimidazole carboxamide formyl transferase), because (Glaeser & Kämpfer, 2015). The concatenated housekeeping genes could minimize the weight, combination and position of closely strains (Glaeser & Kämpfer, 2015; Timilsina et al., 2015). This analysis MLSA assigns a high similarity between sequences differing only a single nucleotide. As a result, this analysis is more suitable for organisms with clonal evolution and can determine phylogenic relationships between closely related species.

In this study, the new MLSA scheme was created based on five housekeeping genes with *glpF, pta, purH, pycA, rpoD* and 16S rRNA, then construct the phylogenetic tree to compare between 16S rRNA gene-based studies and MLSA scheme to show that whether MLSA approach truly better resolution than 16S rRNA sequences.

2. Materials and methods

2.1. Strains and primers

Eight strains from the *Bacillus subtilis* species complex whose names have been validly published belonging to the genus *Bacillus* were analyzed in this study (Table 1). All strains belonged to the HCMBiotech Collection of Microorganisms (HBCM) in the Biotechnology Center of Ho Chi Minh City, Vietnam.

These strains were grown aerobically on TSA agar. An addition of 35 *Bacillus* species (34 species belong to the *B. subtilis* species complex, one used as an outgroup) was used to reconstruct the phylogenetic tree. All 35 *Bacillus* species were retrieved from National Center for Biotechnology Information (NCBI) accession number can be found in Supplementary Table 1. All oligonucleotide primers used in this study are listed in Table 2. Primers's names include all information on the region amplified for each gene.

Table 1

The information of all strains used in this study include strains ID, Bacterium's name

Strains number	Strains ID	Bacterium
1	HBCM-B0029	*B. amyloliquefaciens*
2	HBCM-B0037	*B. pumilus*
3	HBCM-B0027	*B. subtilis*
4	HBCM-B0028	*B. subtilis*
5	HBCM-B0039	*B. tequilensis*
6	HBCM-B0020	*B. licheniformis*
7	HBCM-B0116	*B. subtilis*
8	HBCM-B0112	*B. amyloliquefaciens*

Source: HBCM data bacteria of collecting species

Table 2

List of primers were used in this study

Primers	Sequence (5'-3')[1]	Target genes	Source
20F	AGA GTT TGA TCM TGG CTC AG	16S rRNA	Integrated DNA Technologies
1500R	CGA TCC TAC TTG CGT AG		
glpF-F	WTG ACA GCA TTT TGG GG	glycerol uptake facilitator (*glpF*)	
glpF-R	GTA AAA TAC RCC GCC GA		
pycA-F	AAA TCA GAR GCG AAA GC	pyruvate carboxylase (*pycA*)	
pycA-R	CCT GAG CGG TAA GCC AT		
pta-F	ATA CAT AYG AAG GVA TGG AAG A	phosphotransacet ylase (*pta*)	
pta-R	TAG CCG ATR TTY CCY GCT		
purH-F	AYA TTC ACG GMG GNC TBC T	phosphoribosyl aminoimidazole carboxamide formyl transferase (*purH*)	
purH-R	TGY TCK BTC GGY TCY CTT TT		
rpoD-F	GCY GAA GAA GAA TTT GAC CTB AA	RNA polymerase sigma factor (*rpoD*)	
rpoD-R	CGT TTR CTT CTG CTH GGA TGT CT		

[1]Degenerate base codes: R = A+G; Y = C+T; M = A+C; K = G+T; W = A+T; H = A+T+C; B = G+T+C; N = A+C+G+T; V = A+C+G

Source: The researcher's data analysis

2.2. DNA extraction and Annealing temperature determining

Total DNA extraction were completed by using chemical and enzymatic lysis protocol to extract DNA from bacterial biomass (Elkins, 2013). The quantity and purity of the total DNA

were measured using a Nanodrop spectrophotometer at 260/280nm. The optimum annealing temperature of each primer pair was determined by gradient PCR, using a range of eight different temperatures with Tm - 5 degrees as the middle temperature. The PCR results were analysis by 1% agarose gel electrophoresis with a DNA ladder from Thermo Fisher Scientific. The annealing temperature displaying a good amplification was selected for later PCR reactions. The base Tm of each primer was taken from the production, and was shown in Supplementary Table 2. Five different temperatures for each primer are shown in Supplementary Table 3.

2.3. Amplification and Sequencing

In each strain, five gene fragments *glpF*, *pycA*, *pta*, *purH* and *rpoD* were amplified and sequenced. The 16S rRNA region was also amplified and sequenced for comparison. PCR amplification was performed on total DNA extraction. The amplification recipe followed Thermal Scientific DreamTaq DNA Polymerase according to the manufacturer's recommendation. All reactions were performed in an Eppendorf Mastercycler EP gradient thermocycler. After amplification step, PCR products were visualized by 1% electrophoresis agarose gel and stained with GelRed® Nucleic Acid Gel Stain (Biotium). Their concentration is verified by a Nanodrop spectrophotometer at 260/280nm. Then, PCR products were purified using the PCR purification kit (Axygen Scientific, Inc., USA) according to the manufacturer's instructions and collected DNA purified PCR products were. Next, purified PCR products were sequenced using the Sanger sequencing method with BigDye™ Direct Cycle Sequencing Kit (Thermo Fisher Scientific).

2.4. Individual gene analyses

Some characteristics of the individual gene was analyzed such as parsimony informative sites, GC content and sequence similarities for further discussion and comparison.

2.5. Multisequence alignment

PCR sequencing products were then imported into machine to get fully nucleotide sequences. The nucleotide sequences were edited by using ATGC software (GENETYX CORPORATION). After that, all gene sequences were analyzed using the software MEGA-X to compare. Sequences from the 16S region and five other housekeeping genes *glpF*, pycA, *pta*, *purH* and *rpoD* from different *Bacillus* isolates strains and one species from outside of the *B. subtilis* species complex are obtained from NCBI. Multisequence alignment using ClustalW with default settings to align the same region on different strains obtained. Phylogenic analysis from individual genes was conducted, then gene fragments of each species were concatenated. Phylogenic reconstructions were generated in each strain in this study using Neighbor-Joining Analysis in MEGA-X.

3. Results

3.1. Primers optimal annealing temperature

The goal of this experiment was to determine the optimal annealing temperature of five pairs of primers: *glpF, pycA, pta, purH, rpoD*. Gradient PCR was used to determine the optimal annealing temperature for PCR reactions to amplify the five genes' fragments. PCR products were visualized under agarose gel electrophoresis. The result shows that in the *glpF* fragment the optimal temperature is 48˚C. Similarly, the optimal temperature for four genes fragments *pycA*; *pta purH, rpoD* is 46˚C, 49˚C, 54˚C, and 51˚C, respectively.

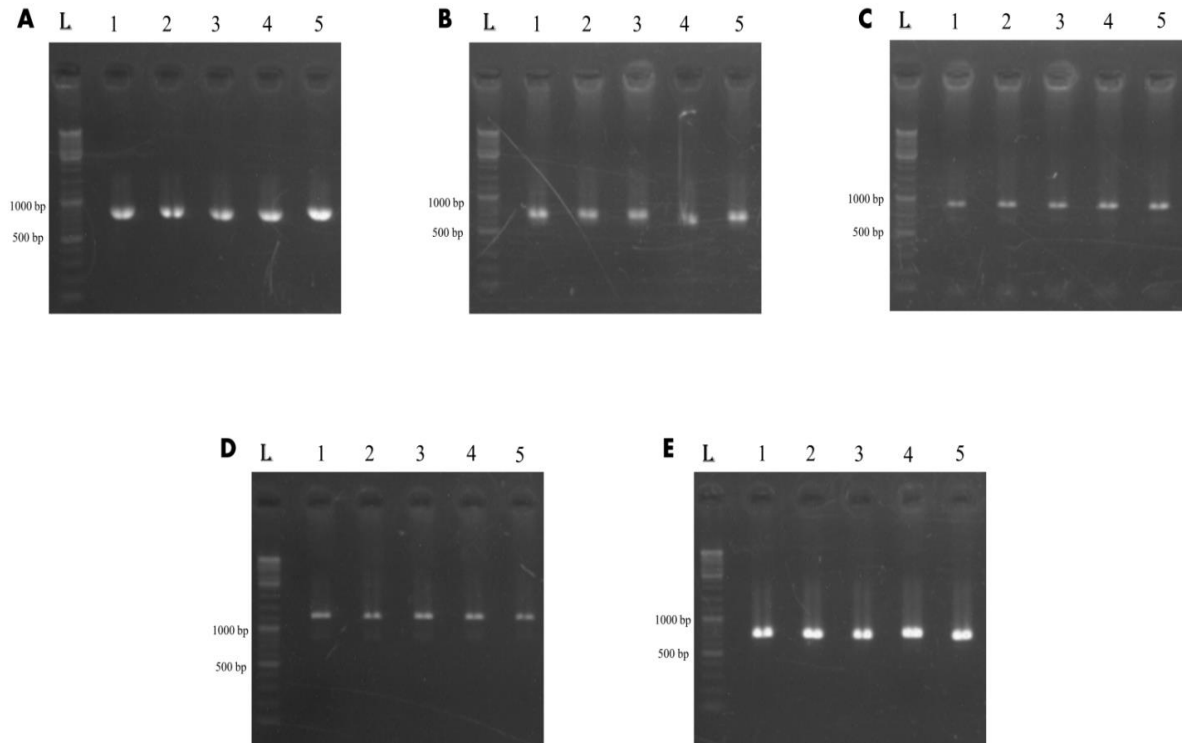


Figure 1. Gradient PCR results for determination of optimal annealing temperature (46˚C, 49˚C, 54˚C, 51˚C) for primer pairs of the five genes fragments. (A) is *glpF*, (B) is *pycA*, (C) is *pta*, (D) is *purH*, (E) is *rpoD*. The "L" is short for DNA ladder

3.2. PCR amplification and sequencing

The five gene fragments and 16S region were amplified and sequenced on eight strains.

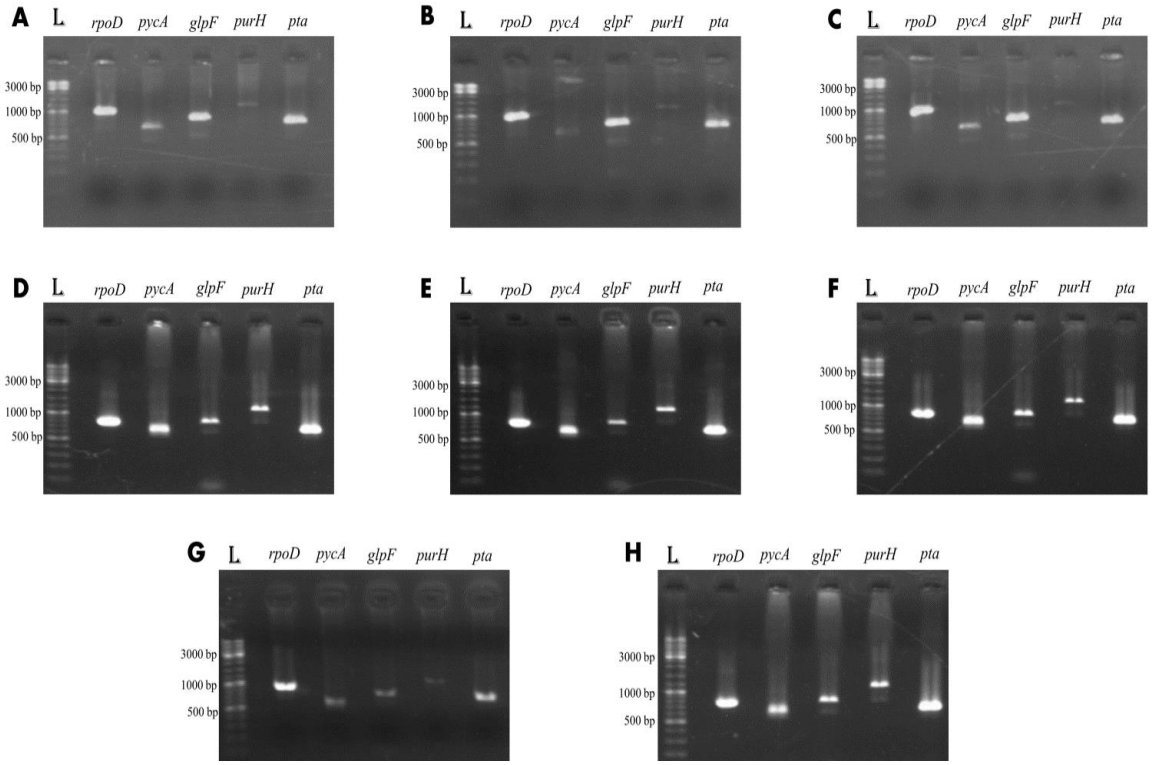


Figure 2. PCR amplification on eight strains. The "L" is short for DNA ladder. (A) to (H) are strain 1 to strain 8, respectively

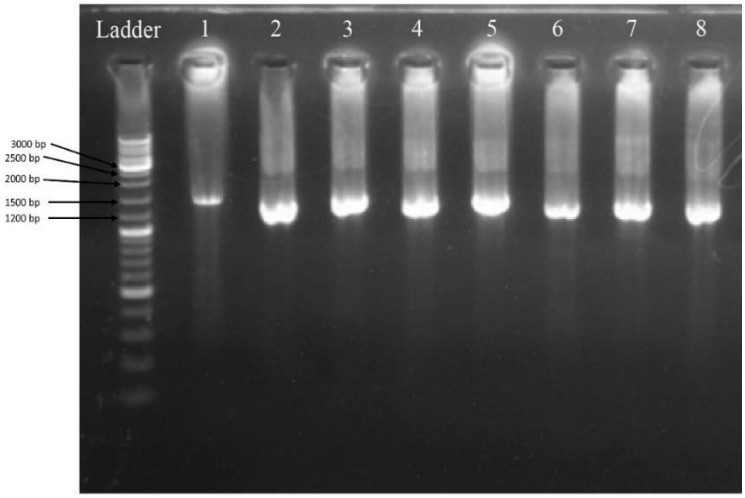


Figure 3. PCR amplification on eight strains for 16S region. Lanes 1 to 8: strain 1 to strain 8, respectively

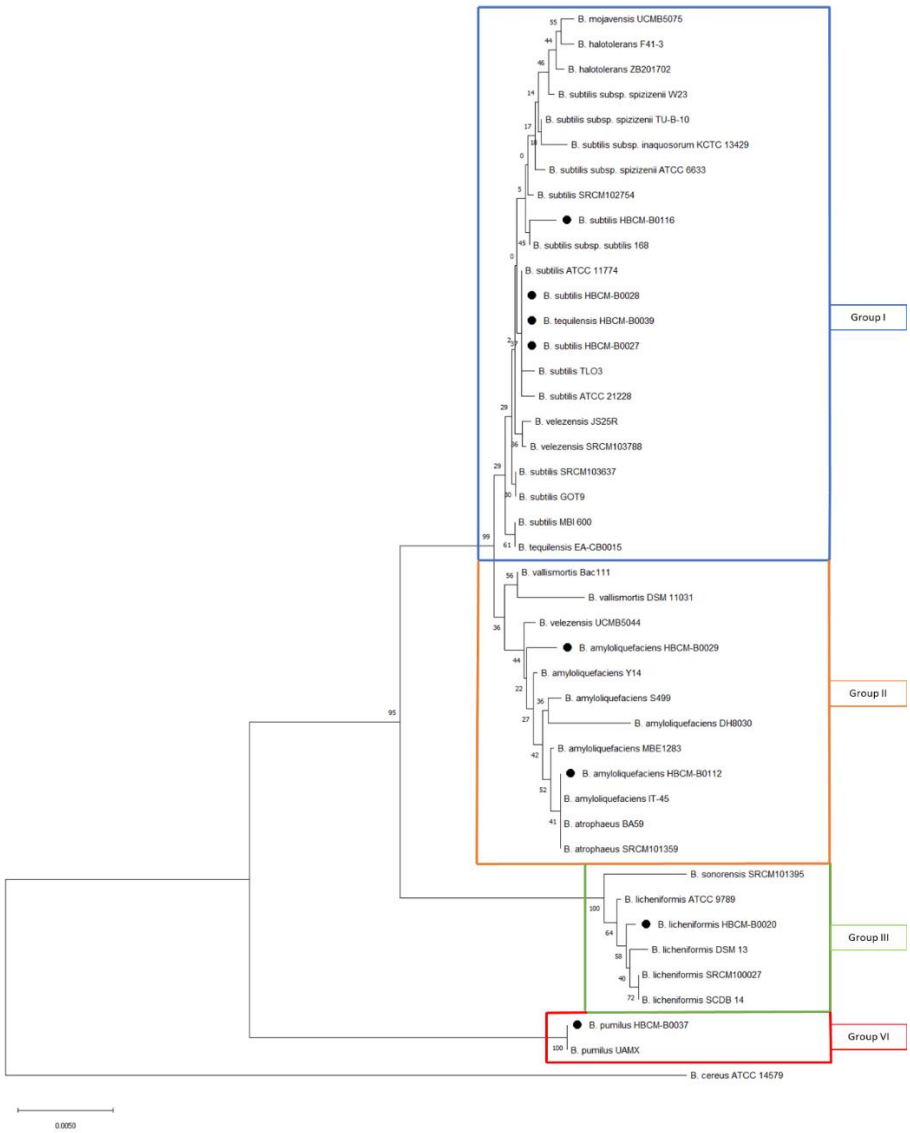


Figure 4. Phylogenetic tree of *Bacillus subtilis* species complex using 16S rRNA sequence analyzed by Neighbor-Joining method. *Bacillus cereus* ATCC 14579 was used as an outgroup. All phylogenetic trees were based on 1,000 replications (bootstraps). Scale bar = 0.0050 substitutions per nucleotide position. Eight strains which used for analysis are labeled with black dot
Source: Gordon et al. (1973)

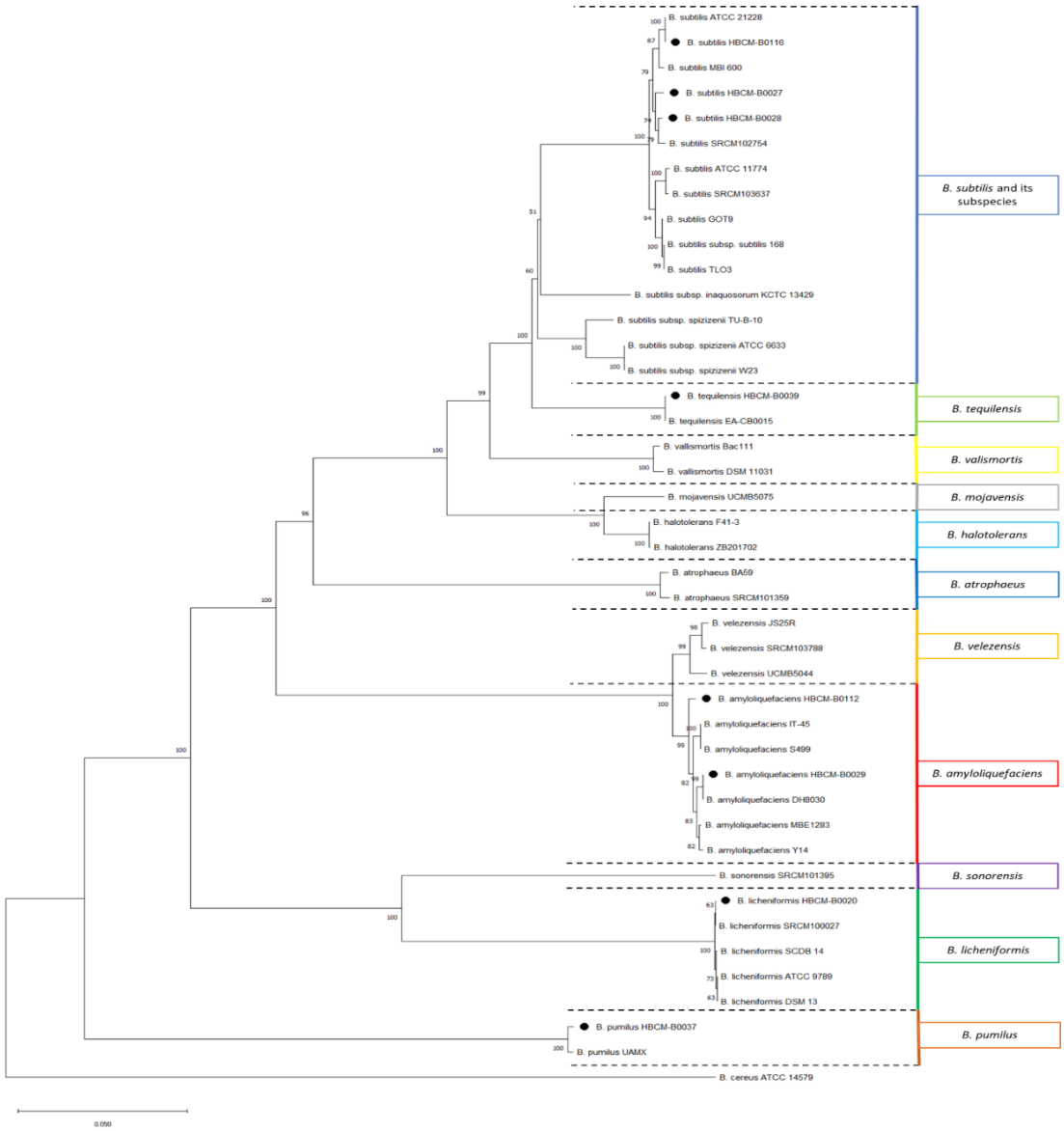


Figure 5. Phylogenetic tree of *Bacillus subtilis* species complex using MLSA scheme analyzed by Neighbor-Joining method. *Bacillus cereus* ATCC 14579 was used as an outgroup. All phylogenetic trees were based on 1,000 replications (bootstraps). Scale bar = 0.050 substitutions per nucleotide position. Eight strains dwhich used for analysis are labeled with black dot

Source: Gordon et al. (1973)

Table 3

Characteristics of the 16S rRNA gene, single housekeeping gene and the concatenated genes from all strains. (Konstantinidis & Tiedje, 2005)

Locus	Length (approximate)	Parsimony informative site		GC content (%)	Similarities (%)
		No.	%		
16S rRNA	1475	63	4.27	54.9	96.95 - 100
glpF	820	375	45.2	46.9	73.41 - 100
pta	960	329	33.8	46.4	78.42 - 100

purH	1200	484	38.7	47.9	77.91 - 99.92
pycA	720	309	42.7	46	75.29 - 99.91
rpoD	1100	351	31.1	46.3	81.02 - 100
MLSA	4800	1902	38.7	47.1	78.77 - 99.66

4. Discussion

All 16S rRNA region of eight analyzed strains is visualized under gel electrophoresis. The length was easy to determine as it goes from 1,500 base pairs to 1,600 base pairs, which match with the common length of 16S rRNA region in *Bacillus* species. All sequences are reverse complement (if needed) for the same plus strand. To maintain the right order for a further experiment, we came to the decision that the order of gene fragments concatenate would follow alphabet order. The final concatenate gene will be *glpF-pta-purH-pycA-rpoD*, and the length of the concatenate gene will be approximately 4,700 base pairs.

We suggest that within the *B. subtilis* species complex four groups can be distinguished based on the result of the 16S rRNA phylogenetic tree:

- Group (I) consisting of *B. subtilis* including its three subspecies: *B. subtilis subsp subtilis, B. subtilis subsp spizizenii*, and *B. subtilis subsp inaquosorum, B. tequilensis, B. mojavensis, B. halotolerans*
- Group (II) consisting of *B. amyloliquefaciens* and *B. vallismortis* and *B. atrophaeus*
- Group (III) consisting of *B. licheniformis* and *B. sonorensis*
- Group (IV) consisting of *B. pumilus*

B. velezensis was appeared in two group, Group (I) and Group (II).

As we had mention above, 16S rRNA gene-based studies often provides low resolution in differentiation between species that are very closely related. Many previous papers have shown that some species share over 97% or even 100% of 16S rRNA gene sequences (Das, Dash, Mangwani, Chakraborty, & Kumari, 2014). In this study, we found that there are two problems in dealing with 16S rRNA gene-based studies. The first thing is that the strains in the same species have almost 99.9% similarities between 16S rRNA gene sequences. For instance, in Figure 4, in Group (I), there are 02 strains of *B. subtilis* retrieved from NCBI (SCRM 103637 and GOT9) sharing 100% sequence similarities. This also occurred in three other groups. The second thing is 16S rRNA region between species to species also shown no significant difference. Look at Group (I) when *B. subtilis* MBI 600 shares more similarities with *B. tequilensis* EA-CB0015 than other *B. subtilis* species. Also, the sequence identity means between *B. subtilis* and *B. amyloliquefaciens* is higher than 99% (approximate 99.35%). This is a really high identity considering these two species are very hard to distinguish. Furthermore, the bootstrap value is low in many branches, the tree is not considered reliable. Combine all the observation, we can say that although 16S rRNA gene-based studies has been widely recognized for microorganisms' phylogenetic relationships identification for a very long time, it still has a downside and not good for further research to examine the accurate taxonomic position or to obtain higher resolution for phylogenetic relationships, for example, the strains of the *B. subtilis* species complex in this study.

In order to provide better resolution for relationship phylogenetic resolution, the MLSA approach has given a new path for further research and has been extensively used (Martens et al., 2008). Some species used the MLSA method based on housekeeping genes such as *Pseudomonas, Glycyrrhiza, Mesorhizobium* (Ampomah, Mousavi, Lindstrom, & Huss-Danell,

2017; Mulet, Lalucat, & Garcia-Valdes, 2010).

In this study, MLSA has presented some advantages over 16S rRNA gene-based studies. Although the phylogenetic tree based on MLSA shown a similar topology to the 16S rRNA genes sequences tree, the MLSA approach demonstrated more clear resolution, the branches are also longer and supported by much higher bootstrap values. Instead of the disorders species in 16S rRNA phylogenetic tree, *Bacillus* species in the MLSA approach are separated. We can easily see that 13 different species belong to the *Bacillus subtilis* species complex without any mixing. Along with that, the MLSA approach has given a new look at phylogeny relationship in the *Bacillus subtilis* species complex. Because the resolution is much clearer now, some species seem to have a different relationship, such as *B. atrophaeus* are closer to *B. subtilis* than *B. amyloliquefaciens* or *B. velezensis* now belong to other branches with *B. amyloliquefaciens.*

There are some materials that support the MLSA approach to making phylogenetic trees. The similarities between sequences in the MLSA scheme shown significantly different than sequences in 16S rRNA. In 16S rRNA, some sequences share 100% similarities, which is make researchers harder to distinguish new strains or even new species if not conduct further research. Parsimony informative sites also state that 16S rRNA gene-based studies are inadequate to discriminate closely related strains and species with only 4.27%. MLSA approach with 38.7% parsimony informative site is much more supported for better resolution in phylogeny. A phylogenetic tree of each housekeeping gene was also conducted to see if one housekeeping gene can also show better resolution than 16S rRNA. All five trees (Supplementary Figure 1 to Supplementary Figure 5) demonstrated a similar topology tree to the MLSA tree, but bootstrap values are lower, with some species still showning high similarities in sequences. The order of phylogenetic species is also not homologous with the MLSA tree, for instance, the phylogenetic tree of *pta* genes (Supplementary Figure 2) and *purH* genes (Supplementary Figure 5). Despite many reasons to support the MLSA approach, there still cannot completely reject 16S rRNA gene-base studies. GC content of 16S rRNA is higher than MLSA approach, which means 16S rRNA sequences are more likely conserved than MLSA sequences, also higher GC content contains more information about phylogenetics although in this study GC content of 16S rRNA and MLSA are considered low (< 60%). Besides that, there is something in the MLSA tree that strongly consider. B. subtilis since 2009 (Nakamura et al., 1999) as been *B. subtilis* has been subdivided into three subspecies: *B. subtilis* subsp. *subtilis*, *B. subtilis* subsp. *spizizenii* and *B. subtilis* subsp. *inaquosorum.* This study shown something otherwise. *B. subtilis* subsp. *subtilis* and *B. subtilis* subsp. *inaquosorum* are still in the same branch with *B. subtilis*, but MLSA approach has separated *B. subtilis* subsp. *spizizenii* into new branches, suggesting the species is not subspecies of *B. subtilis* anymore. Unfortunately, our study cannot provide enough materials to strongly prove that hypothesis, so it is still the hypothesis for considering the MLSA can give new look for phylogeny analysis.

MLSA is a practical, easy-to-do, and reliable technique for replacing 16S rRNA gene-based studies for better phylogenetic relationship resolution of bacteria species (Glaeser & Kämpfer, 2015) it still has some disadvantages. The MLSA approach does not based on one particular canonical scheme. Many housekeeping genes from different strains were used and with analytical methodologies make it difficult to unify one scheme that can access to all bacteria genera. In the formation about a public database combination are limited (Liu, Lai, & Shao, 2017), the study about genome sequencing of Dddh AND and will contribute to compensate for the shortcomings in the MLSA method (Hahnke et al., 2016).

5. Conclusion

In this study, we successfully described the better resolution and differentiation of species within the *B. subtilis* species complex, including 08 strains from four species (*B. subtilis*, *B.*

amyloliquefaciens, *B. licheniformis*, and *B. pumilus*) were provided by Biotechnology Center of Ho Chi Minh City. A multi locus sequence analysis (MLSA approach with the concatenation of five protein-coding housekeeping genes: glpF, purH, pycA and rpoD) was established and performed on the *Bacillus* genus, MLSA scheme shown more advantages upon 16S rRNA gene-based studies.

Funding

This research received funding sponsored by the Biotechnology Center of Ho Chi Minh City, Vietnam.

ACKNOWLEDGMENTS

We are very grateful for the eight *Bacillus* strains used in this study from the Microbiology Department from Biotechnology Center of Ho Chi Minh City.

References

Ampomah, O. Y., Mousavi, S. A., Lindstrom, K., & Huss-Danell, K. (2017). Diverse *Mesorhizobium* bacteria nodulate native *Astragalus* and *Oxytropis* in arctic and subarctic areas in Eurasia. *Systematic and Applied Microbiology, 40*(1), 51-58. doi:10.1016/j.syapm.2016.11.004

Bacon, C. W., & Hinton, D. M. (2002). Endophytic and biological control potential of *Bacillus mojavensis* and related species. *Biological Control, 23*(3), 274-284. doi:10.1006/bcon.2001.1016

Cohn, F. (1872). Untersuchungen über Bakterien. *Beitrage zur Biologie der Pflanzen Heft 2, 1*(2), 127-224.

Das, S., Dash, H. R., Mangwani, N., Chakraborty, J., & Kumari, S. (2014). Understanding molecular identification and polyphasic taxonomic approaches for genetic relatedness and phylogenetic relationships of microorganisms. *Journal Microbiological Methods, 103,* 80-100. doi:10.1016/j.mimet.2014.05.013

Ehrenberg, C. G. (1835). Physikalische abhandlungen der koeniglichen akademie der wissenschaften zu berlin aus den jahren 1833-1835. *Königlich Bayerische Akademie,* 145-336.

Elkins, K. M. (2013). DNA extraction. *Forensic DNA Biology*, 39-52. doi:10.1016/B978-0-12-394585-3.00004-3

Fritze, D. (2004). Taxonomy of the genus Bacillus and related genera: The aerobic endospore-forming bacteria. *Phytopathology, 94*(11), 1245-1248. doi:10.1094/PHYTO.2004.94.11.1245

Gatson, J. W., Benz, B. F., Chandrasekaran, C., Satomi, M., Venkateswaran, K., & Hart, M. E. (2006). *Bacillus tequilensis* sp. nov., isolated from 2000-year-old Mexican shaft-tomb, is closely related to *Bacillus subtilis*. *International Journal of Systematic Bacteriology, 56*(7), 1475-1484. doi:10.1099/ijs.0.63946-0

Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., ... Swings, J. (2005). Opinion: Re-evaluating prokaryotic species. *Nature Review Microbiology, 3*(11), 733-739. doi:10.1038/nrmicro1236

Glaeser, S. P., & Kämpfer, P. (2015). Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Systematic and Applied Microbiology, 38*(4), 237-245. doi:10.1016/j.syapm.2015.03.007

Gordon, R. E., Haynes, W. C., & Pang, C. H.-N. (1973). *The genus bacillus*. Washington, D.C.: United States Department of Agriculture.

Hahnke, R. L., Meier-Kolthoff, J. P., Garcia-Lopez, M., Mukherjee, S., Huntemann, M., Ivanova, N. N., … Goker, M. (2016). Genome-based taxonomic classification of *bacteroidetes*. *Frontiers in Microbiology, 7*, Article 2003. doi:10.3389/fmicb.2016.02003

Kitahara, K., & Miyazaki, K. (2013). Revisiting bacterial phylogeny: Natural and experimental evidence for horizontal gene transfer of 16S rRNA. *Mobile Genetic Element, 3*(1), Article e24210. doi:10.4161/mge.24210

Konstantinidis, K. T., & Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America, 102*(7), 2567-2572. doi:10.1073/pnas.0409727102

Liu, Y., Lai, Q., & Shao, Z. (2017). A multilocus sequence analysis scheme for phylogeny of thioclava bacteria and proposal of two novel species. *Frontiers in Microbiology, 8*, Article 1321. doi:10.3389/fmicb.2017.01321

Liu, Y., Lai, Q., Dong, C., Sun, F., Wang, L., Li, G., & Shao, Z. (2013). Phylogenetic diversity of the *Bacillus pumilus* group and the marine ecotype revealed by multilocus sequence analysis. *PLoS ONE, 8*(11), Article 80097. doi:10.1371/journal.pone.0080097

Liu, Y., Lai, Q., Göker, M., Meier-Kolthoff, J. P., Wang, M., Sun, Y., … Shao, Z. (2015). Genomic insights into the taxonomic status of the *Bacillus cereus* group. *Scientific Reports, 5*(1), Article 14082. doi:10.1038/srep14082

Macheras, E., Roux, A. L., Bastian, S., Leão, S. C., Palaci, M., Sivadon-Tardy, V., … Heym, B. (2011). Multilocus sequence analysis and rpoB sequencing of Mycobacterium abscessus (sensu lato) strains. *Journal Clinical Microbiology, 49*(2), 491-499. doi:10.1128/JCM.01274-10

Martens, M., Dawyndt, P., Coopman, R., Gillis, M., De Vos, P., & Willems, A. (2008). Advantages of multilocus sequence analysis for taxonomic studies: A case study using 10 housekeeping genes in the genus *Ensifer* (including former *Sinorhizobium*). *International Journal of Systematic Bacteriology, 58*(1), 200-214. doi:10.1099/ijs.0.65392-0

Mulet, M., Lalucat, J., & Garcia-Valdes, E. (2010). DNA sequence-based analysis of the *Pseudomonas* species. *Environmental Microbiology, 12*(6), 1513-1530. doi:10.1111/j.1462-2920.2010.02181

Nakamura, L. K. (1989). Taxonomic relationship of black-pigmented *Bacillus subtilis* strains and a proposal for *Bacillus atrophaeus* sp. nov. *International Journal of Systematic Bacteriology, 39*(3), 295-300. doi:10.1099/00207713-39-3-295

Nakamura, L. K., Roberts, M. S., & Cohan, F. M. (1999). Relationship of *Bacillus subtilis* clades associated with strains 168 and W23: A proposal for *Bacillus subtilis* subsp. *subtilis* subsp. nov. and *Bacillus subtilis* subsp. *spizizenii subsp*. nov. *International Journal of Systematic Bacteriology, 49*(3), 1211-1215. doi:10.1099/00207713-49-3-1211

Palmisano, M. M., Nakamura, L. K., Duncan, K. E., Istock, C. A., & Cohan, F. M. (2001). *Bacillus sonorensis* sp. nov., a close relative of *B. licheniformis*, isolated from soil in the Sonoran Desert, Arizona. *International Journal of Systematic Bacteriology, 51*(5), 1671-1679. doi:10.1099/00207713-51-5-1671

Peeters, C., Meier-Kolthoff, J. P., Verheyde, B., De Brandt, E., Cooper, V. S., & Vandamme, P. (2016). Phylogenomic study of *Burkholderia glathei*-like organisms, proposal of 13 novel *Burkholderia* species and emended descriptions of *Burkholderia sordidicola, Burkholderia zhejiangensis,* and *Burkholderia grimmiae*. *Frontier in Microbiology, 7*, Article 877. doi:10.3389/fmicb.2016.00877

Priest, F. G., Goodfellow, M., Shute, L. A., & Berkeley, W. (1987). *Bacillus amyloliquefaciens* sp. nov., nom. rev. *International Journal of Systematic Bacteriology, 37*(1), 69-71. doi:10.1099/00207713-37-1-69

Roberts, M. S., Nakamura, L. K., & Cohan, F. M. (1994). *Bacillus mojavensis* sp. nov., distinguishable from *Bacillus subtilis* by sexual isolation, divergence in DNA sequence, and differences in fatty acid composition. *International Journal of Systematic Bacteriology, 44*(2), 256-264. doi:10.1099/00207713-44-2-256

Roberts, M. S., Nakamura, L. K., & Cohan, F. M. (1996). *Bacillus vallismortis* sp. nov., a close relative of *Bacillus subtilis*, isolated from soil in death valley, California. *International Journal of Systematic Bacteriology, 46*(2), 470-475. doi:10.1099/00207713-46-2-470

Rooney, A. P., Price, N. P., Ehrhardt, C., Swezey, J. L., & Bannan, J. D. (2009). Phylogeny and molecular taxonomy of the *Bacillus subtilis* species complex and description of *Bacillus subtilis* subsp. *inaquosorum* subsp. nov. *International Journal of Systematic Bacteriology, 59*(10), 2420-2436. doi:10.1099/ijs.0.009126-0

Ruiz-García, C., Béjar, V., Martinez-Checa, F., Llamas, I., & Quesada, E. (2005). *Bacillus velezensis* sp. nov., a surfactant-producing bacterium isolated from the river Velez in Malaga, southern Spain. *International Journal of Systematic Bacteriology, 55*(1), 191-195. doi:10.1099/ijs.0.63310-0

Skerman, V. B. D., McGowan, V., & Sneath, P. H. A. (Eds.). (1980). Approved lists of bacterial names. *International Journal of Systematic Bacteriology, 30*(3), 225-420. doi:10.1099/00207713-30-1-225

Timilsina, S., Jibrin, M. O., Potnis, N., Minsavage, G. V., Kebede, M., Schwartz, A., … Goss, E. M. (2015). Multilocus sequence analysis of xanthomonads causing bacterial spot of tomato and pepper plants reveals strains generated by recombination among species and recent global spread of Xanthomonas gardneri. *Applied Enviroment Microbiology, 81*(4), 1520-1529. doi:10.1128/AEM.03000-14

Turnbull, P. C. B., & Baron, S. (Eds.) (1996). *Medical microbiology* (4th ed.). Galveston, TX: University of Texas Medical Branch at Galveston.

Vinuesa, P. (2010). Multilocus sequence analysis and bacterial species phylogeny. In O. Aharon & R. T. Papke (Eds.), *Molecular phylogeny of microorganisms* (pp. 45-64). Norfolk, VA: Caister Academic Press.

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K. H., … Amann, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Review Microbiology, 2*(9), 635-645. doi:10.1038/nrmicro3330

Creative Commons Attribution-NonCommercial 4.0 International License.