**TRƯỜNG ĐẠI HỌC AN GIANG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**TRẦN THỊ MINH THÙY**

**KHÓA LUẬN TỐT NGHIỆP**

**NGÀNH CÔNG NGHỆ THÔNG TIN**

**Nghiên cứu kiểm thử tự động và   
ứng dụng trong giảng dạy tại   
Khoa Công nghệ thông tin**

**GIẢNG VIÊN HƯỚNG DẪN**

**TS. HUỲNH PHƯỚC HẢI**

**an giang, 2020**

LỜI CAM ĐOAN

Tôi cam đoan rằng nội dung của luận văn này là kết quả nghiên cứu của bản thân dưới sự hướng dẫn trực tiếp của thầy TS. Huỳnh Phước Hải. Tất cả những tham khảo đều được trích dẫn nguồn gốc một cách rõ ràng. Những đóng góp trong luận văn là kết quả nghiên cứu của tác giả và chưa được công bố trong bất kỳ công trình khoa học nào khác.

Tác giả luận văn

LỜI CẢM ƠN

Trong quá trình hoàn thành luận văn này, tôi đã được các thầy cô nơi cơ sở đào tạo giúp đỡ tận tình, cơ quan nơi công tác tạo mọi điều kiện thuận lợi và gia đình cùng bạn bè thường xuyên động viên khích lệ.

Luận văn này không thể hoàn thành tốt nếu không có sự tận tình hướng dẫn và sự giúp đỡ quý báu của TS. Huỳnh Phước Hải, người thầy hướng dẫn mà tôi muốn bày tỏ lòng biết ơn sâu sắc nhất.

Tôi cũng muốn bày tỏ lòng biết ơn đối với tập thể thầy cô trong Khoa CNTT – Đại học An Giang đã giúp đỡ và tạo điều kiện cho tôi rất nhiều trong quá trình học tập và nghiên cứu ở Khoa. Cám ơn Bộ môn Công Nghệ Thông Tin về sự hỗ trợ các thủ tục hoàn thành luận văn.

Cuối cùng, tôi cám ơn gia đình và bạn bè đã góp nhiều ý kiến thiết thực và có những lời động viên khích lệ quý báu giúp tôi hoàn thành tốt luận văn.

TÓM TẮT

Sự ra đời của công nghệ giải trình tự gien thế hệ tiếp theo đã tạo ra lượng dữ liệu khổng lồ về gien được lưu trữ trên các ngân hàng gien của thế giới và chi phí để giải trình tự ngày càng thấp. Bước tiến này đã cách mạng hóa nhiều lĩnh vực của khoa học sinh học trong đó có nghiên cứu đa dạng sinh học. Trong đó, Meta barcoding là một dự án dùng để xác định và nhận dạng trình tự ADN dựa trên thông lượng cao từ một mẫu dữ liệu giải trình tự (dữ liệu thô).

Trong đó, dữ liệu gien Chloroplast là cơ sở quan trọng và linh hoạt để xây dựng mã vạch và thiết lập mối quan hệ di truyền giữa các loài thực vật. Vì vậy việc nghiên cứu giải mã và xác định gien Chloroplast là một cơ sở quan trọng để xây dựng các module trong dự án Metabarcoding. Tuy nhiên, phương pháp xác định trình tự bộ gien Chloroplast truyền thống hạn chế phải sử dụng mẫu gien gốc để làm cơ sở xác định. Cách tiếp cận trên không giải quyết được được yêu cầu mẫu dữ liệu đầu vào của dự án Metabarcoding.

Để giải quyết vấn đề trên trong luận văn chúng tôi đề xuất qui trình xác định trình tự bộ gien Chloroplast mới có thể xác định các gien này từ các mẫu dữ liệu thô và không cần sử dụng các gien tham chiếu. Bên cạnh đó, chúng tôi cũng xây dựng các công cụ để thực hiện qui trình mới này. Để đánh giá qui trình, trong thực nghiệm chúng tôi xác định trình tự gien Chloroplast từ bốn tập dữ liệu gien và so sánh kết quả thực nghiệm với các mẫu gien Chloroplast đã được các nhà sinh học xây dựng. Kết quả chúng tôi đã xác định trình tự của bộ gien Chloroplast theo qui trình mới tương đồng với các bộ gien đã công bố.

ABSTRACT

The Next generation sequencing (NGS) technologies have created an enormous genetic data stored on DNA genomic banks and its cost has increased slowly. This step has made an revolution in many aspects of biology and bioinformatics, especially in biodiversity. In particular, Metabarcoding is a project to define and identify DNA sequences based on high-throughput from NGS dataset.

In addition, the chloroplast genomes are an important and flexible fundament for constructing bar-code and setting up genetic relationships among plants. Therefore, assembly of Chloroplast genomes provides an important basis to construct modules for the Metabarcoding project. However, the limitation of the traditional Chloroplast Genome Sequencing is the use of the references chloroplast genomes. This approach didn't solve the input data requirements of Metabarcoding project.

To solve the above problems, in this study, we have proposed a new process to establish Chloroplast genomes which can define this gene from NGS data without using reference genome. We have developed tools to implement this new process. To evaluate our approach, we compare experimental results with four references Chloroplast genomes which are sequenced by biologist. The results show that the Chloroplast genomes determined  by our approach are the same the reference Chloroplast genomes.

MỤC LỤC

[LỜI CAM ĐOAN i](#_Toc56072055)

[LỜI CẢM ƠN ii](#_Toc56072056)

[TÓM TẮT iii](#_Toc56072057)

[ABSTRACT iv](#_Toc56072058)

[MỤC LỤC v](#_Toc56072059)

[DANH MỤC BẢNG BIỂU vii](#_Toc56072060)

[DANH MỤC HÌNH ẢNH viii](#_Toc56072061)

[DANH MỤC THUẬT NGỮ VIẾT TẮT x](#_Toc56072062)

[CHƯƠNG 1 TỔNG QUAN VỀ KIỂM THỬ PHẦN MỀM 1](#_Toc56072063)

[1.1 Kiểm thử phần mềm 1](#_Toc56072064)

[1.2 Phân loại kiểm thử 1](#_Toc56072065)

[1.3 Phạm vi và mục đích nghiên cứu 1](#_Toc56072066)

[1.4 Ứng dụng thực tiễn của đề tài 1](#_Toc56072067)

[1.5 Bố cục luận văn 1](#_Toc56072068)

[CHƯƠNG 2 kiểm thử tự động 3](#_Toc56072069)

[2.1 Lịch sử kỹ thuật giải trình tự gien 3](#_Toc56072070)

[2.2 Các phương pháp giải trình tự gien 3](#_Toc56072071)

[2.3 Dữ liệu giải trình tự 3](#_Toc56072072)

[2.4 Kết luận 3](#_Toc56072073)

[CHƯƠNG 3 QUI TRÌNH XÁC ĐỊNH TRÌNH TỰ BỘ GIEN CHLOROPLAST CỦA CÁC LOÀI CÂY 4](#_Toc56072074)

[3.1 Qui trình xác định trình tự bộ gien Chloroplast hiện tại 4](#_Toc56072075)

[3.2 Kết luận 4](#_Toc56072076)

[CHƯƠNG 4 THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ 5](#_Toc56072077)

[4.1 Mô tả thực nghiệm 5](#_Toc56072078)

[CHƯƠNG 5 KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 6](#_Toc56072079)

[5.1 Kết luận 6](#_Toc56072080)

[5.2 Hướng phát triển 6](#_Toc56072081)

[DANH MỤC TÀI LIỆU THAM KHẢO 7](#_Toc56072082)

[PHỤ LỤC 8](#_Toc56072083)

DANH MỤC BẢNG BIỂU

No table of figures entries found.DANH MỤC HÌNH ẢNH

**No table of figures entries found.**

DANH MỤC THUẬT NGỮ VIẾT TẮT

|  |  |  |
| --- | --- | --- |
| **Từ viết tắt** | **Diễn giải tiếng Anh** | **Diễn giải tiếng Việt** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# TỔNG QUAN VỀ KIỂM THỬ PHẦN MỀM

## Kiểm thử phần mềm

## Phân loại kiểm thử

## Phạm vi và mục đích nghiên cứu

Đề tài tập trung nghiên cứu qui trình xác định trình tự bộ gien Chloroplast của các loài cây từ một mẫu dữ liệu thô (draw data). Nghiên cứu này là một phần trong dự án Metabarcoding DNA, qui trình và các gói chương trình thực hiện sẽ xác định trình tự của bộ gien Chloroplast từ một mẫu dữ liệu có nhiều gien mà không cần thực hiện so sánh mẫu. Để giải quyết đề tài cần thực hiện lần lượt các công việc:

* Tìm hiểu qui trình xác định trình tự gien Chloroplast truyền thống;
* Đề xuất qui trình xác định trình tự bộ gien Chloroplast các loài cây;
* Đánh giá qui trình xác định trình tự bộ gien Chloroplast mới với bốn tập dữ liệu thử nghiệm.

Đồng thời luận văn thực hiện tìm hiểu các kiến thức liên quan đến đề tài như các kiến thức cơ sở về tin sinh học, các kỹ thuật xác định trình tự ADN (DNA Sequencing), các kỹ thuật lắp ráp trình tự (Assembly ), các công cụ sử dụng trong qui trình xây dựng một bộ gien, …Chúng tôi cũng sẽ tiến hành đánh giá tính chính xác và hiệu quả của qui trình mà chúng tôi đề xuất bằng cách thực hiện ánh xạ (mapping) các trình tự gien Chloroplast thu được vào các gien mẫu (references genomes).

## Ứng dụng thực tiễn của đề tài

Qui trình thực hiện sẽ được sử dụng tích hợp trong dự án Metabarcoding để xây dựng barcoding ADN cho các loài cây. Đây là một ứng dụng phục vụ cho rất nhiều nghiên cứu trong đa dạng sinh học, phân tích các loài cây, sinh vật phù du và phân tích cộng đồng sinh vật đáy, xây dựng lại hệ sinh thái trong quá khứ.

## Bố cục luận văn

Nội dung luận văn được chia làm năm chương. Các chương này được tổ chức nội dung như sau:

**Chương 1: Giới thiệu**

Nội dung chương này chúng tôi giới thiệu tổng quan các nội dung cơ sở lý thuyết liên quan đến đề tài. Sau cùng chúng tôi trình bày phạm vi nghiên cứu của đề tài và ứng dụng của đề tài.

**Chương 2: Tổng quan về qui trình xác định trình tự bộ gien**

Nội dung chương này chúng tôi trình bày chi tiết về các giai đoạn trong qui trình xác định trình tự một bộ gien gồm: giải trình tự gien (Sequencing DNA), lắp ráp trình tự gien ( Assembly) , ánh xạ gien mẫu (Mapping).

**Chương 3: Xây dựng qui trình xác định trình tự bộ gien Chloroplast của các loài cây**

Trong chương này chúng tôi đề xuất một qui trình xác định trình tự bộ gien Chloroplast mới dựa trên qui trình truyền thống. Trong mỗi giai đoạn chúng tôi trình bày chi tiết giải thuật, cách cài đặt và xây dựng các công cụ.

**Chương 4. Thực nghiệm và đánh giá kết quả.**

Nội dung chương này trình bày cách thức thực hiện quá trình thực nghiệm để đánh giá qui trình. Chúng tôi trình bày cụ thể cách xử lý dữ liệu trong từng giai đoạn của qui trình và cách thức sử dụng các công cụ đã được xây dựng trong qui trình. Tiếp theo chúng tôi cách đánh giá qui trình trên bốn tập dữ liệu khác nhau.

**Chương 5. Kết luận và hướng phát triển**

Nội dung chương này chúng tôi nêu ra các kết luận và hướng phát triển của đề tài như cải tiến qui trình xác định trình tự bộ gien Chloroplast.

# kiểm thử tự động

## Lịch sử kỹ thuật giải trình tự gien

## Các phương pháp giải trình tự gien

## Dữ liệu giải trình tự

## Kết luận

# 

# ỨNG DỤNG KIỂM THỬ TỰ ĐỘNG TRONG GIẢNG DẠY NGÀNH CÔNG NGHỆ THÔNG TIN

## Qui trình xác định trình tự bộ gien Chloroplast hiện tại

## Kết luận

# THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

## Mô tả thực nghiệm

# KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

## Kết luận

## Hướng phát triển

DANH MỤC TÀI LIỆU THAM KHẢO

PHỤ LỤC