

**Tóm tắt khóa luận tốt nghiệp**

**XÂY DỰNG CƠ SỞ DỮ LIỆU SSRs (SIMPLE SEQUENCE REPEATS) TỪ ESTs  
(EXPRESSED SEQUENCE TAGS) CỦA CÂY DỨA (*Ananas comosus*)**

**Sinh viên: Trần Nguyễn Minh Đăng**

**Khóa: 2002 – 2006**

Trong những năm qua sinh học không ngừng phát triển, đã tạo ra những kho dữ liệu rất lớn về trình tự gene, protein,... của thực vật, động vật,... Và với các thành tựu vốn có của công nghệ thông tin, những trình tự gene này đã và đang được lưu trữ trong cơ sở dữ liệu sinh học lớn như NCBI, EMBL, DDBj,... Vì các cơ sở dữ liệu này quá lớn và chứa rất nhiều thông tin khác nhau, không tập trung thành từng gene cụ thể nên khó có thể thực hiện việc truy xuất các thông tin phục vụ trực tiếp cho một nghiên cứu chuyên biệt, trong đó có phương pháp microsatellite. Do vậy, mục tiêu của chúng tôi là tiến hành xây dựng cơ sở dữ liệu SSRs từ ESTs của cây dứa *Ananas comosus* được lấy ở cơ sở dữ liệu sinh học NCBI.

Để đạt được mục tiêu trên, khóa luận cần đảm bảo thực hiện nội dung như sau:

- ❖ Dùng Perl script để thu nhận trình tự các nucleotide của gene từ trang cơ sở dữ liệu GenBank NCBI.
- ❖ Tìm và tách các đoạn microsatellite có thể có trong mỗi đoạn gen.
- ❖ Tìm hiểu về mô hình dữ liệu quan hệ, sử dụng mô hình này vào việc lưu trữ dữ liệu các trình tự nucleotide và trình tự SSRs của *Ananas comosus*, và tạo cơ sở dữ liệu chứa những trình tự này. Sau đó chuyển các dữ liệu này vào cơ sở dữ liệu chính.
- ❖ Kết hợp các phần mềm quản lý cơ sở dữ liệu và phần mềm tạo web, thiết kế trang web chia sẻ thông tin với người dùng.