ĐẠI HỌC QUỐC GIA HÀ NỘI

ĐẠI HỌC KHOA HỌC TỰ NHIÊN

KHOA TOÁN - CƠ – TIN HỌC

**TÓM TẮT LUẬN VĂN**

Học viên: **Tạ Văn Nhân.**

Giáo viên hướng dẫn: **PGS. TS. Nguyễn Thị Hồng Minh.**

Tên đề tài: **Áp dụng phương pháp gióng hàng trình tự cho bài toán dự đoán biến thể gen.**

|  |  |
| --- | --- |
| **1 KIẾN THỨC CƠ SỞ** |  |
| **1.1 Một số khái niệm về sinh học phân tử và di truyền** |  |

1.1.1 Các phân tử của một tế bào

|  |  |
| --- | --- |
| 1.1.2 Luận thuyết trung tâm |  |
| 1.1.3 Nhiễm sắc thể |  |
| 1.1.4 Đột biến |  |
| 1.1.5 Bệnh liên quan đến gen |  |
| **1.2 Các công nghệ giải trình tự DNA** |  |
| 1.2.1 Giải trình tự Sanger |  |
| 1.2.2 Giải trình tự thế hệ tiếp theo (NGS ) |  |
| 1.2.3 Các loại trình tự nhận được từ máy giải trình tự |  |
| **1.3 Các bài toán tin sinh học** |  |
| 1.3.1 Một số bài toán phổ biến  1.3.2 Bài toán dự đoán ảnh hưởng của biến thể gen  1.3.2.1 Một số cách tiếp cận và hạn chế |  |
| 1.3.2.2 Giải trình tự bộ gen người |  |

**1.4 Dóng hàng trình tự**   
 1.4.1 Khái niệm   
 1.4.2 Sự phát triển các thuật toán   
  
  
 **2 PHÁT TRIỂN CÁC THUẬT TOÁN DÓNG HÀNG TRÌNH TỰ**

|  |  |
| --- | --- |
| **2.1 Thuật toán dựa trên chuyển dạng Burrows-Wheeler**  2.1.1 Một số cấu trúc dữ liệu |  |
| 2.1.1.1 Mảng hậu tố (Suffix Arrays) |  |
| 2.1.1.2 Ma trận chuyển dạng Burrows-Wheeler  2.1.1.3 Ma trận điểm kiểm tra (Checkpoint Arrays)  2.1.2 Thuật toán |  |
| 2.1.2.1 Thuật toán khớp chính xác |  |
| 2.1.2.2 Thuật toán khớp xấp xỉ |  |

2.1.2.3 Cho điểm dóng hàng

|  |  |
| --- | --- |
| **2.2 Thuật toán Smith-Waterman** |  |
| 2.2.1 Đồ thị Manhattan ba cấp |  |
| 2.2.2 Thuật toán tiết kiệm bộ nhớ |  |
| 2.2.2.1 Giai đoạn chia (bài toán tìm cạnh giữa) |  |
| 2.2.2.2 Giai đoạn trị |  |
| 2.2.3 Thuật toán tham lam cho dóng hàng đa trình tự |  |
| 2.2.4 Tính điểm cho dóng hàng đa trình tự |  |

**2.3 Thực nghiệm thuật toán**

2.3.1 Thuật toán song song với Golang   
 2.3.2 Thực nghiệm   
 2.3.2.1 Dữ liệu   
 2.3.2.2 Tham số và đầu vào

2.3.2.3 Kết quả

**3 ỨNG DỤNG THUẬT TOÁN TRONG DỰ ĐOÁN BIẾN THỂ GEN**

|  |  |
| --- | --- |
| **3.1 Dữ liệu** |  |
| **3.2 Tiền xử lý dữ liệu** |  |
| 3.2.1 Kiểm tra chất lượng  3.2.1.1 Điểm chất lượng trên mỗi vị trí nucleotide  3.2.1.2 Thành phần GC trên các bazơ |  |

3.2.1.3 Phần trăm trình tự trùng lặp

|  |  |
| --- | --- |
| 3.2.2 Loại bỏ các bazơ có điểm chất lượng kém |  |
| 3.2.3 Dóng hàng trình tự |  |

**3.3 Xác định biến thể**

|  |  |
| --- | --- |
| **3.4 Chú thích chức năng** |  |
| **3.5 Kết quả** |  |