**ĐẠI HỌC BÁCH KHOA HÀ NỘI**

A red and yellow sign

Description automatically generated with low confidence

**Báo cáo tìm hiểu chủ đề:**

**Dịch tễ học bệnh truyền nhiễm**

**Họ tên: Nguyễn Thanh Hoàn**

**MSSV: 20173126**

|  |
| --- |
| **HÀ NỘI, 8/2022** |

**Báo cáo tìm hiểu chủ đề: Dịch tễ học bệnh truyền nhiễm**

**Epidemiology**

Genetic epidemiology nguyên cứu vai trò của các yếu tố di truyền trong việc xác định tình trạng sức khỏe và bệnh tật trong quần thể người và xã hội, cũng như sự tương hỗ của các yếu tố đó với các yếu tố từ môi trường.

**Module 1: Giới thiệu**

1. **Sự nổi lên của bệnh truyền nhiễm**

* Mầm bệnh bên trong cơ thể người tiến hóa
* Các bệnh lây từ động vật sang người
* Xác định các điểm nóng về bệnh truyền nhiễm trên toàn cầu
* Xác định các yếu tố gây bệnh: Rừng mưa nhiệt đới, mật độ dân số, biến đổi khí hậu, sự đa dạng của các loài động vật

**Dịch Ebola**

* Bùng nổ ở Tây Phi từ 12/2013 đến 05/2016
* Ảnh hưởng đến du lịch toàn cầu
* Phân tích dịch tễ cho kết quả:

+ Người bệnh bị phơi nhiễm với các nguồn mang bệnh ngoài tự nhiên (con dơi)

+ Sự bùng dịch sảy ra do lây nhiễm từ người sang người

* Thiệt hại do chưa có kinh nghiệm đối phó với đại dịch (thiếu các cơ sở cách ly hiệu quả)

1. **Kiểm soát dựa vào hệ gen từ môi trường**

- Cô lập và làm giàu RNAs cúm từ các mẫu trầm tích ở các vùng ngập nước và sử dụng phương pháp giải trình gene thông lượng cao

- Sử dụng các phân tích tin sinh để để phân loai các gene

- Sử dụng phương pháp này có được tỉ lệ dương tính: 30.4%

* Có thể xác định đc mầm bênh từ các mẫu ngoài môi trường

1. **AMR and Mobile Elements**

Các gen kháng thuốc có thể di chuyển xung quanh vật chủ

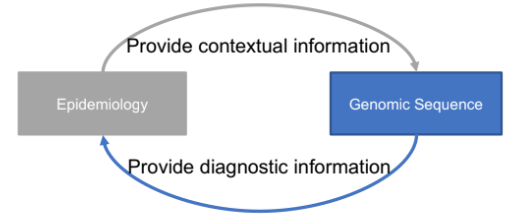
+ Phát hiện gen bằng phương pháp PCR không đủ để hiểu đc sự lây truyền của gen kháng thuốc

**Tại sao phải giải trình tự gen của SARS-CoV-2**

+ Theo dõi sự truyền nhiễm theo quy mỗ vùng, quốc gia và quốc tế

+ Phân cụm việc điều tra sự lây lan

**Genomic Epidemiology**: Kết hợp từ dữ liệu giải trình tự toàn bộ hệ gen từ mầm bệnh với các điều tra về dịch tễ học để theo dõi sự lây lan của một bệnh truyền nhiễm



Dịch tễ học cung cấp các thông tin về ngữ cảnh cho việc giải trình tự gene

Trình tự hệ gene cung cấp dữ liệu chuẩn đoán giúp xác định hướng điều tra cho dịch tễ học

1. **High-throughput Sequencing (HTS)**

Dữ liệu có nhiều lợi ích cho y học thực hành và thí nhiệm

Có thế dùng cho việc nghiên cứu

1. **Hệ gen của mầm bệnh**

Vi khuẩn: Thường xuất hiện bên trong các nhiễm sắc thể vòng đơn. Kích thước 3-5Mb

Vi rút: Có thể là DNA hoặc RNA, chuỗi đơn hoặc chuỗi đôi. Kích thước 1-2Mb

Kí sinh trùng hoặc nấm: Gồm nhiều nhiễm sắc thể. Kích thước vài trăm Mb

1. **Phân tích dữ liệu giải trình tự**

Quá trình giải trình tự sinh ra các đoạn trình tự chồng chéo 1 phần lên nhau

**Các bước phân tích:**

* **Tổng hợp hệ gen**: Khôi phục toàn bộ hệ gen từ những đoạn DNA rời rạc.

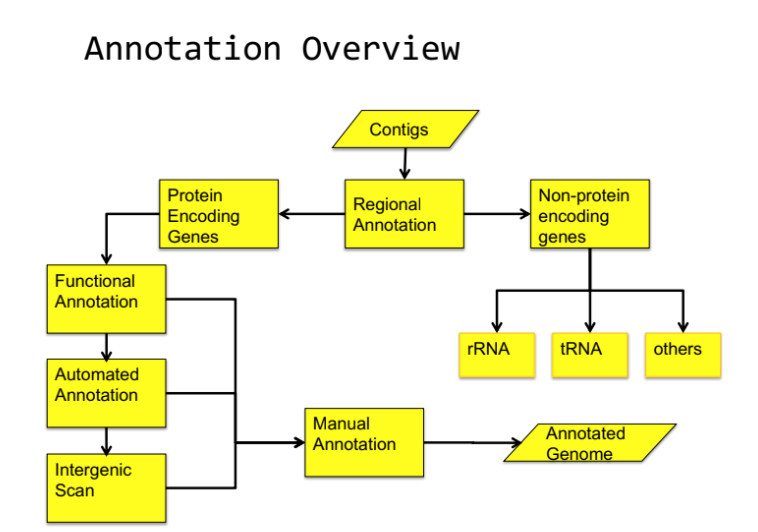
+ Denovo: Sử dụng thuật toán máy tính để xác định đoạn bị chồng trên từng đoạn DNA, từ đó gộp chúng lại thành hệ gen hoàn chỉnh.

+ Mapping: Ghép các đoạn DNA lại dựa vào một hệ gen đã biết, chất lượng phụ thuộc vào sự tương đồng giữa 2 gen.

->Vẫn sẽ có lỗi trong quá trình tổng hợp. Có thể thực hiện nhiều lần để chọn ra kết quả tối ưu nhất

* **Đánh dấu hệ gen**

Xác định chức năng và vị trí của từng gen trong chuỗi giải trình tự



**+ Xác định vị trí:**

Vị trí của các gene mã hóa thông tin có thể được xác định bởi chương trình máy tính bởi vì các gene này có tần số xuất hiện của các nucleotide khác với các gene không mã hóa thông tin di truyền.

**+ Xác định chức năng**

Dựa vào sự tương đồng. Các gene tương tự nhau được cho là kế thừa từ cùng tổ tiên và vì vậy có thể có chức năng giống nhau

Được gọi là transitive annotation

Cần database của các gen đã biết (GenBank)

1. **Đối chiếu các hệ gene**

* Để xác định các biến thể liên quan đến đặc điểm kiểu hình của tế bào
* Có thể sử dụng các biến thể này để theo dõi sự lây lan của mầm bệnh
* Các biển có thể sảy ra ở các cấp độ khác nhau:

+ Theo vị trí: Một đoạn DNA có thể xuất hiện ở một loại và không có ở những loài khác

+ Theo gene: Một gene bị thiếu hoặc mã hóa cho một amino acid khác

+ Theo Nucleic Acid: Một nucleotide khác với các biến thể khác

**Module 2: Phylogenetic trees**

Phylogenetics: Nghiên cứu về mặt tiến hóa mối quan hệ giữa các thực thể sinh học. Dựa vào trình tự trong phân tử (DNA, protein) hoặc đặc điểm hình thái (cấu trúc xương, bộ phận cơ thể)

* Kết quả được thể hiện bởi cây phả hệ.

**Cây phả hệ:**

+ Các sinh vật sống được thể hiện bởi các thành phần lá trên cây (không có các nút con)

+ Các nút có nút con thể hiện giả thuyết về tổ tiên trong quá khứ của các nút lá

+ Quá trình xây dựng cây được thực hiện với một số giả định về sự phân nhánh, sự đa dạng sinh học, sự đồng nhất về vật chất di truyền.

**Các đặc điểm:**

+ Hướng của cây hay thứ tự các nút là không ảnh hướng đến mối quan hệ của các thành phần thể hiện bởi cây

+ Nút gốc là tổ tiên của tất cả các sinh vật trong cây, cây có nút gốc chỉ ra thứ tự của các nút con (thể hiện chiều của sự tiến hóa)

+ Xây dựng được cây phả hệ phụ thuộc vào các yếu tố: phương pháp căn chỉnh chuỗi, mô hình tiến hóa và thuật toán xây dựng cây

**Xây dựng cây phả hệ:** xác định cây thể hiện quan hệ của các dữ liệu theo một mô hình tiến hóa

+ Xây dựng hệ thống căn chỉnh chuỗi

+ Xác định mô hình tiến hóa

+ Xây dựng cây phả hệ

+ Đánh giá cây

**Mô hình tiến hóa**

+ Transitions: Đột biến bằng cách trao đổi nucleotide ở purines(A <-> G) hoặc pyrimidines( C <-> T)

+ Transversions: Đột biến bằng cách trao đổi nucleotide ở purines với pyrimidines với nhau

**Hai phương pháp phổ biến để xây dựng cây phả hệ:**

1. Dựa vào sự khác biệt (Distance Method):

+ Xây dựng cây dựa vào việc tính toán ma trận khoảng cách M(i,j) của các trình tự gen

+ Hàm khoảng cách đơn giản nhất là số lượng thành phần không giống nhau của hai trình tự

+ Mục tiêu là xây dựng một cây mà khoảng cách các nút trong cây d(i,j) tương ứng với giá trị M(i,j)

+ Hoàn thành mục tiêu trên thường là không thể, vì vậy ta thường tìm cây mà tổng sự chênh lệch của hai khoảng cách bên trên là nhỏ nhất:

-> Vẫn là bài toán NP khó

1. Dựa vào trình tự căn chỉnh (Character Methods): Sử dụng trực tiếp kết quả của việc căn chỉnh chuỗi để xây dựng cây

* Phương pháp Maximum Parsimony (MP) : Xây dựng cây mô tả quá trình tiến hóa xảy ra với số bước tiến hóa nhỏ nhất (chiều cao của cây là thấp nhất hay ít đột biến nhất)
* Phương pháp Maximum Likelihood (ML)

+ Tính toán xác suất xuất hiện của chuỗi trình tự trong tập dữ liệu và xây dựng cây có giá trị xác suất lớn nhất theo một mô hình tiến hóa nào đó.

+ Yêu cầu lựa chọn mô hình tiến hóa và tính xác suất có điều kiện của mỗi trình tự trên toàn tập dữ liệu.

**Đánh giá cây: Phương pháp bootstrap**

+ Sử dụng để kiểm xem tập dữ liệu có tương thích với cây được tạo ra hay không

+ Áp dụng được với cây tạo bởi Distance Method và Character Method

+ Thực hiện bằng cách tạo ra nhiều tập dữ liệu phụ (bootstrap replicates) và với mỗi tập phụ, xây dựng cây bằng phương pháp tương tự với tập dữ liệu ban đầu.

+ Cây đầu tiên có sự đồng thuận với từ 70% trở lên với các cây được tạo mới được cho là có thể chấp nhận để mô tả sự tiến hóa trong tập dữ liệu.

**Module 3: WGS-based subtyping**

Vài trò của molecular epidemiology, molecular subtyping

Phân tích biến thể

Kiểm soát hệ gen

1. **Dịch tễ học phân tử**

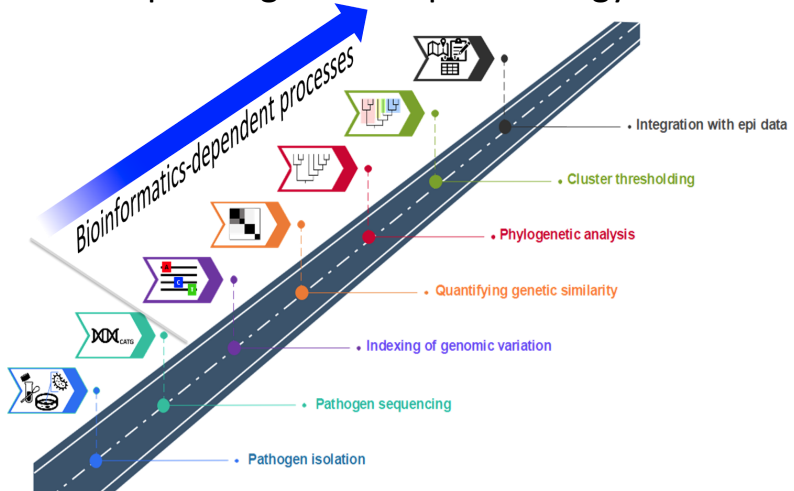
* Dưa ra các vấn đề về dịch tễ học mà khó có thể tiếp cận hoặc sẽ tốn công sức, tài nguyên và thời gian để tiếp cận bằng các kĩ thuật thông thường.
* Việc sử dụng cách tiếp cận về mặt phân tử để xác định đặc tính của các yếu tố gây bệnh truyền nhiễm, ta có thể phân tích được sự phân bổ cũng như cách thức lây truyền của chúng.
  1. Mô hình dịch tễ học phân tử

Tập các quy tắc:

+ Nếu các mẫu có liên quan đến nhau về mặt dịch tễ học thì cũng sẽ liên quan về gen và ngược lại

+ Có sự tương đồng giữa nhưng quan sát về gen và dịch tễ học

+ Sử dụng phân loại phân tử để xác định sự tương đồng về gen giữa các mẫu



**Phân tích các biến thể:**

1. SNV(Single Nucleotide Variant) analysis

Phương pháp dựa vào việc so sánh các biến thể của từng nucleotide trong từng hệ gene.

Biến thể của chuỗi DNA xuất hiện khi có 1 nucleotide bị thay đổi trong trình tự gene

Thực hiện bằng việc so sánh với hệ gen tham chiếu để tìm ra nucleotide bị biến đổi

1. MLST(Multi Locus Sequence Typing) analysis

**Module 5: Antimicrobial Resistant (AMR) Gene Analysis**

Các loại AMR và cách thức lan truyền

Ưu nhược điểm của các phương pháp nghiên cứu

Xác định AMR từ các dữ liệu giải trình tự

1. **Khái niệm**

Anitimicrobial

* Là tác nhân giết chết hoặc ngăn chặn sự phát triển của vi sinh vật
* Được phân nhóm theo loại sinh vật chịu ảnh hưởng:

+ Kháng sinh - Antibiotics: chống lại vi khuẩn

+ Kháng virus – Antivirals: chống lại vi rút

+ Kháng nấm – Antifungals: Chống nấm

Antimicrobials Resistance – Kháng thuốc kháng vi sinh

* Là khả năng các vi sinh vật tiếp tục phát triển khi có sự hiển diện các tác nhân kháng khuẩn mà bình thường có tác dụng đe dọa sự sinh trưởng của chúng
* Được cho là hiểm họa nghiệm trọng đối với sức khỏe cộng đồng

**Các biện pháp:**

* Nâng cao hiểu biết về AMR
* Nghiên cứu các thuốc kháng sinh mới
* Sử dụng kháng sinh một cách hiệu quả
* Ngăn ngừa bệnh bằng vắc-xin, hệ thống vệ sinh, chăm sóc sức khỏe cộng đồng

**Module 6: Emerging Pathogen Detection**

Quy trình và kĩ thuật dùng để phát hiện bệnh truyền nhiễm

Phát hiện các mầm bệnh mới dựa vào dữ liệu giải trình tự gen

**Bệnh truyền nhiễm mới nổi:**

* Là loại bệnh truyền nhiễm có tỷ lệ mắc tăng lên trong thời gian gần đây và có thể gia tăng trong tương lai gần.
* Chiếm 12% trong số các mầm bệnh ở người
* Gây ra bởi các loài mới được phát hiện hoặc có thể đã tiến triển từ một bệnh nhiễm trùng đã biết (ví dụ như cúm) hoặc lây lan sang một dân số mới (ví dụ Sốt Tây sông Nile) hoặc đến một khu vực đang trải qua quá trình biến đổi sinh thái (ví dụ Bệnh Lyme), hoặc tái nhiễm trùng, như bệnh lao kháng thuốc.
* Các phương pháp chuẩn đoán hiện tại không thành công trong việc phát hiện ra các mầm bệnh này

**Thực hành**

1. **Vẽ cây phân loài từ dữ liệu giải trình tự gen**

Bước 1: Chuẩn bị dữ liệu

* Sử dụng dữ liệu giải trình tự gen ở file sharks.fasta bên trong thư mục phylogenetic\_tree

Bước 2: Thực hiện alignment

* Truy cập <https://www.ebi.ac.uk/Tools/msa/muscle/> sử dụng trình tự ở B1 để hệ thống thực hiện việc alignment.
* Đợi và download file kết quả (tương ứng với file sharks.aln trong phylogenetic\_tree)

Bước 3: Xây dựng cây bằng phương pháp dựa vào khoảng cách:

* Thêm thư viện biopython

Text

Description automatically generated

* Đọc dữ liệu alignment:

Text

Description automatically generated

* Tính ma trận khoảng cách:

Text

Description automatically generated

* Xây dựng cây từ ma trận khoảng cách:

Text

Description automatically generated

* Kết quả:

A screenshot of a computer

Description automatically generated with medium confidence