Genomic Analysis Roadmap

DNA

Mapping

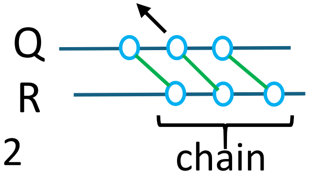
(找read在reference上的對應位置)

路線一: BWT

- 建立BW matrix & FM index
  (reference genome)
- Backward search on FM index
  (query reads)
- 適合short reads (NGS)

路線二: seed & chain

- 取k-mer/minimizer
  (reference genome & query reads)
- 建hash table 紀錄k-mer及出現位置
  (reference genome)
- 找matched k-mer/minimizer → anchor
- 把數個anchor串成chain
  (dynamic programming)
- 較成功的program: minimap2



Alignment

(比對read跟reference的差異)

- Mapping
- Find best alignment
  (dynamic programming)

Assembly (de novo)

(將reads組合成contigs)

路線一: DBG

- 取k-mer
- 用k-mer建De Bruijn graph
- Traverse graph
  (Euler path → contig)
- 適合short reads (NGS)

路線二: OLC

- Mapping (raw read overlap)
- Error correction
- Build assembly graph
- Traverse graph
- 適合long reads (TGS)

RNA

Quantification

(透過統計每個transcript對應的read數量計算各個transcript的表現量)

路線一: Pseudoalignment

- 取k-mer
  (reference transcripts & query reads)
- 建hash table 紀錄k-mer及transcript id
  (reference transcripts)
- 找matched k-mer
- 取得matched k-mer對應的transcript id
- 統計每個transcript對應到的read數量
  (更新similarity table)
- 較成功的program: kallisto

路線二: Exact Alignment

- Mapping
- 統計每個transcript對應到的read數量
  (更新similarity table)
- 速度較慢