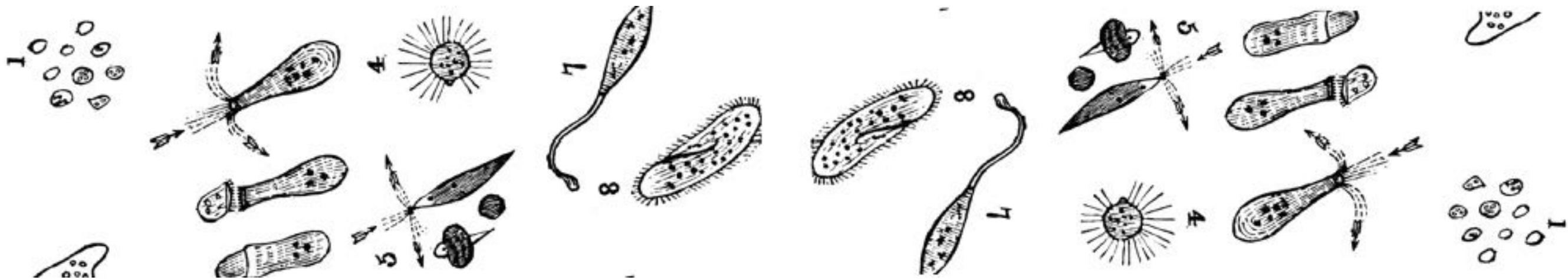




# Alineamiento de secuencias - parte II

[www.castrolab.org](http://www.castrolab.org)  
[www.cbib.cl](http://www.cbib.cl)

Eduardo Castro, PhD  
Universidad Andrés Bello  
4 de septiembre de 2018



# Alineamiento de reads o read mapping

- Fragmentos de DNA típicamente pequeños ( $< 300$  bp)
- Mapear reads = determinar su posición dentro de una secuencia genómica de referencia
- Necesitamos “mapear” las reads en contra de una referencia, de tal forma de identificar qué genes son variables, qué tipo de cambio inducen, etc.

# Alineamiento de reads o read mapping

- El primer desafío es práctico = qué tan rápido podemos mapear millones de reads en una referencia

# Alineamiento de reads o read mapping

- El segundo desafío es estratégico = si una read viene de un elemento repetitivo en la referencia, el programa necesita reportar todas las posibles posiciones o escoger una por sobre las otras

# La clave es usar un índice

- Manera rápida de encontrar la ubicación de cierto contenido (read) en un libro (genoma)

## Index

### A

About cordless telephones 51  
Advanced operation 17  
Answer an external call during an intercom call 15  
Answering system operation 27

### B

Basic operation 14  
Battery 9, 38

### C

Call log 22, 37  
Call waiting 14  
Chart of characters 18

### D

Date and time 8  
Delete from redial 26  
Delete from the call log 24  
Delete from the directory 20  
Delete your announcement 32  
Desk/table bracket installation 4  
Dial a number from redial 26

Dial type 4, 12  
Directory 17  
DSL filter 5

### E

Edit an entry in the directory 20  
Edit handset name 11

### F

FGC, ACTA and IC regulations 53  
Find handset 16

### H

Handset display screen messages 36  
Handset layout 6

### I

Important safety instructions 39  
Index 56-57  
Installation 1  
Install handset battery 2  
Intercom call 15  
Internet 4

# Alineamiento tipo “hash table” o “seed index”

- Los primeros algoritmos utilizaban **read hashing**
- Construir un índice de todos los posibles fragmentos de una read de longitud determinada (**kmers**)
- Alinear **kmers** en contra de la referencia y extender usando algoritmo local Smith-Waterman
- **Menor costo de memoria RAM (variable)**

# Alineamiento tipo “hash table” o “seed index”

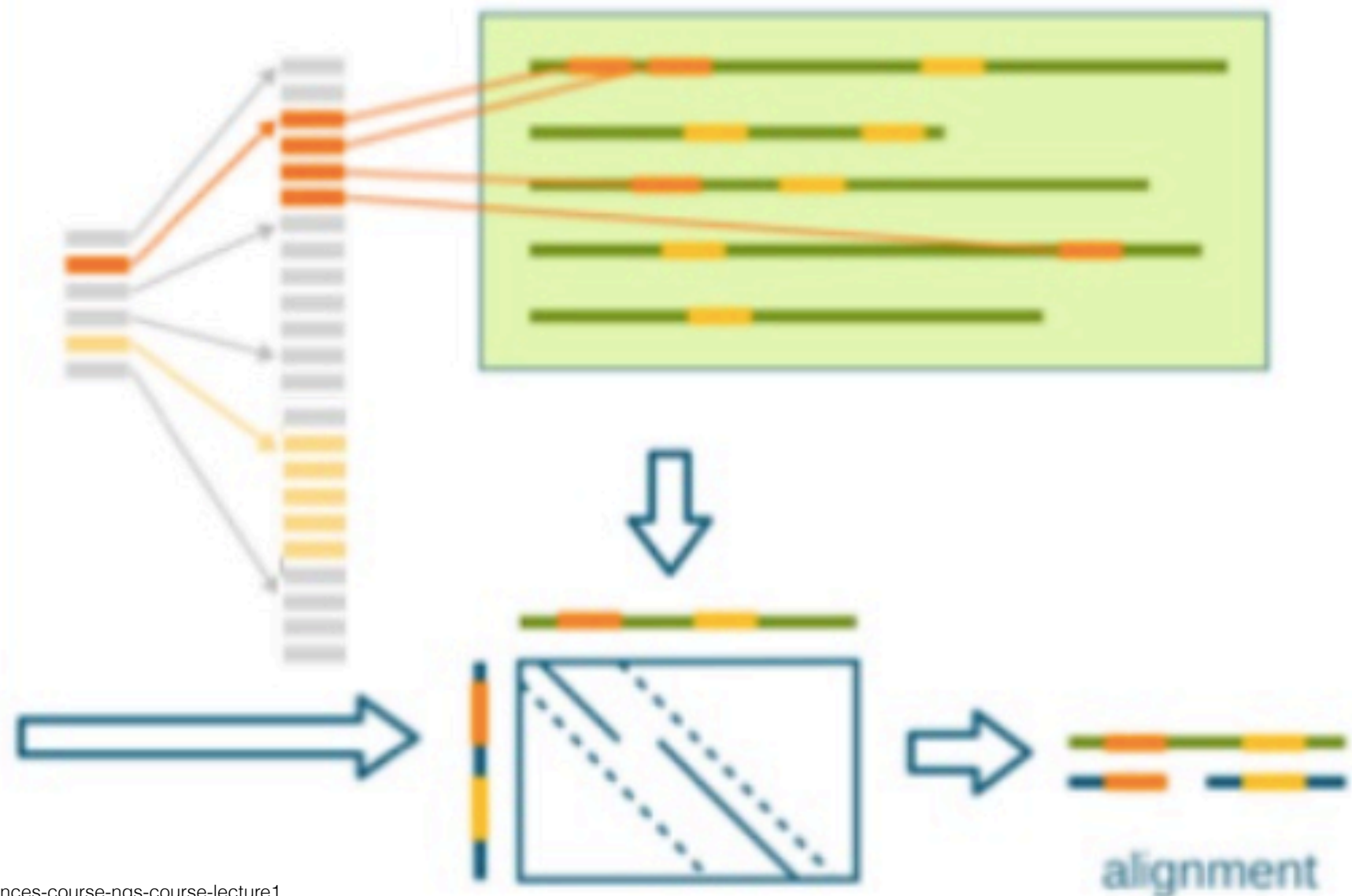
- **Reference hashing**
- Construir un índice de todos los posibles kmers de una referencia y su ubicación relativa (**kmers**)
- Alinear reads en contra del índice de la referencia
- **Constante costo de memoria RAM**

# Alineamiento tipo “hash table” o “seed index”

Sequencing reads

Kmer hash

Reference Genome

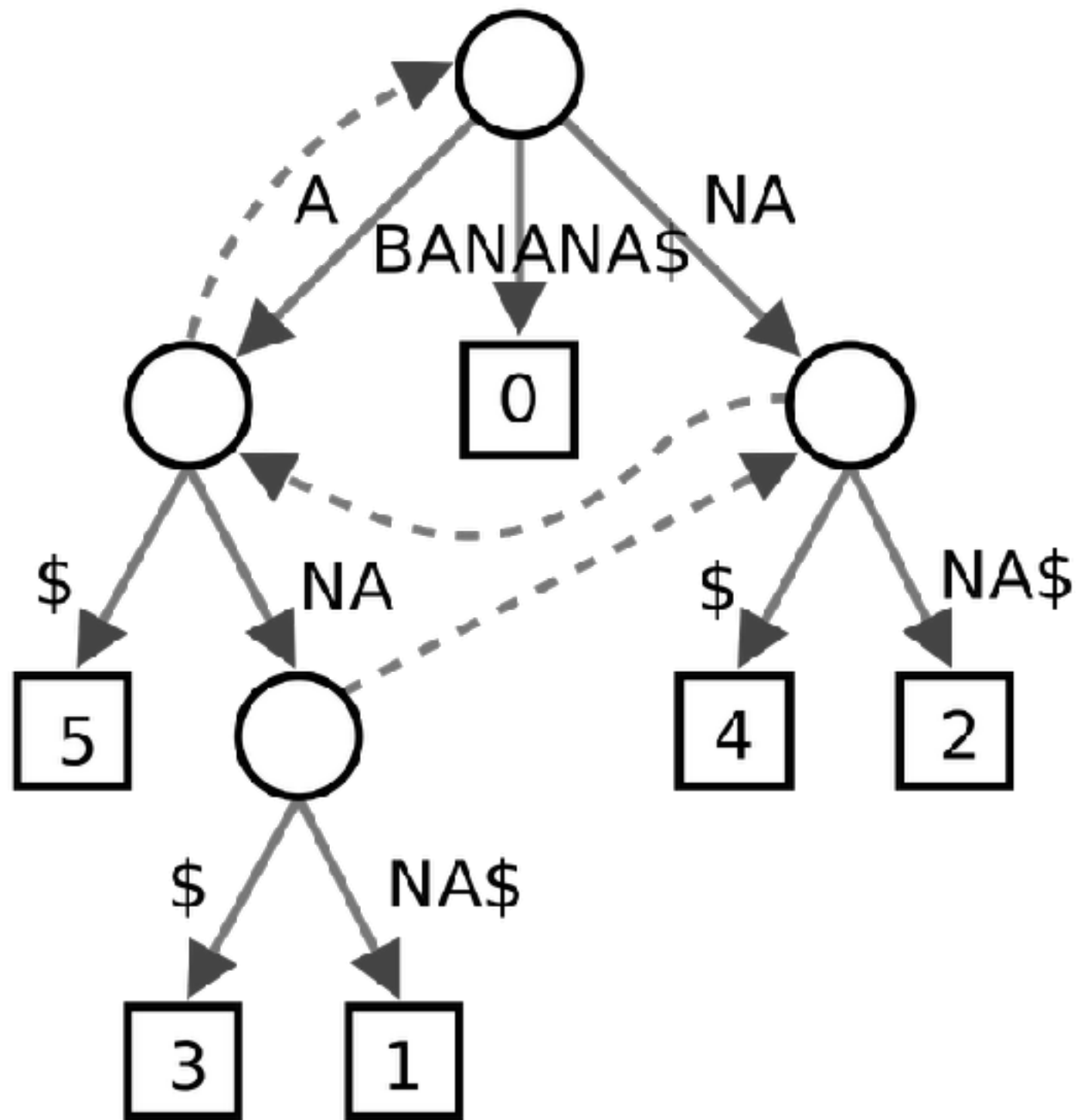




# Alineamiento basado en árbol de sufijos o “Burrows-Wheeler”

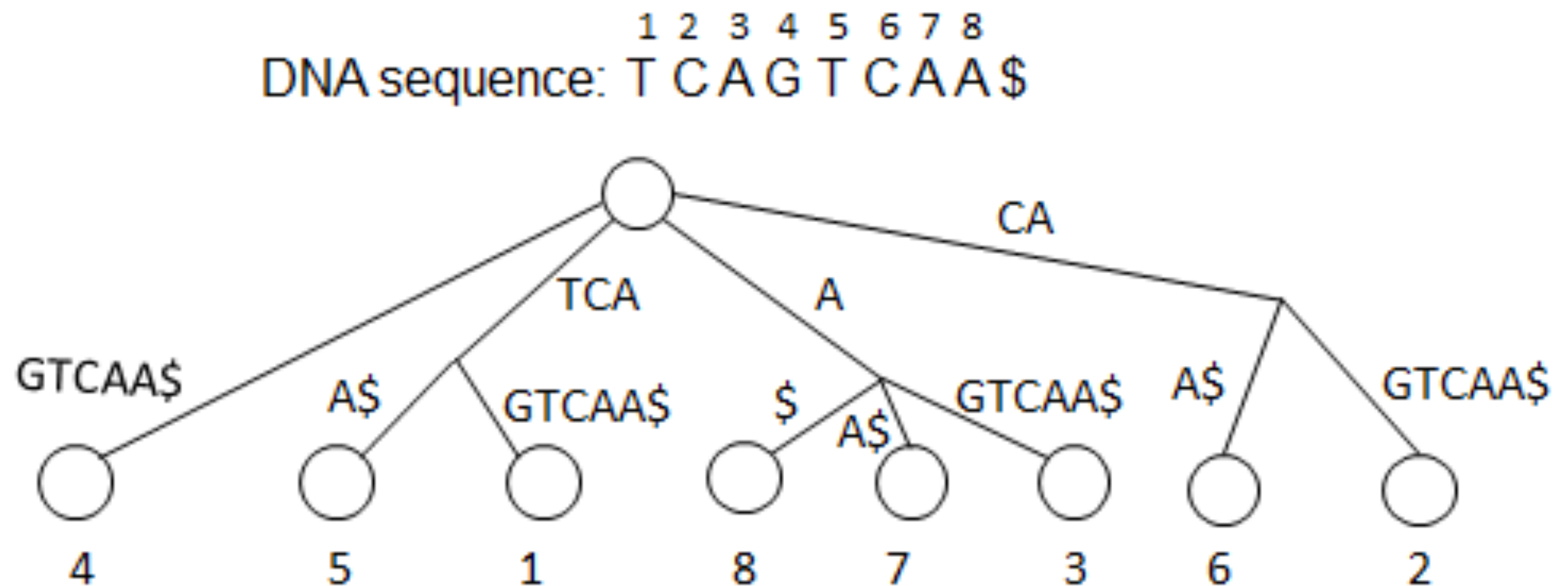
- Formar un índice con la referencia (FM index)
- Formar un árbol de sufijos para una secuencia en particular
- Posicionar el sufijo en contra de la referencia
- Si es un match único sin ambigüedad estamos OK
- Si hace match parcial, se alinea con Smith-Waterman
- **Usa poca memoria RAM**

# Alineamiento basado en árbol de sufijos o “Burrows-Wheeler”

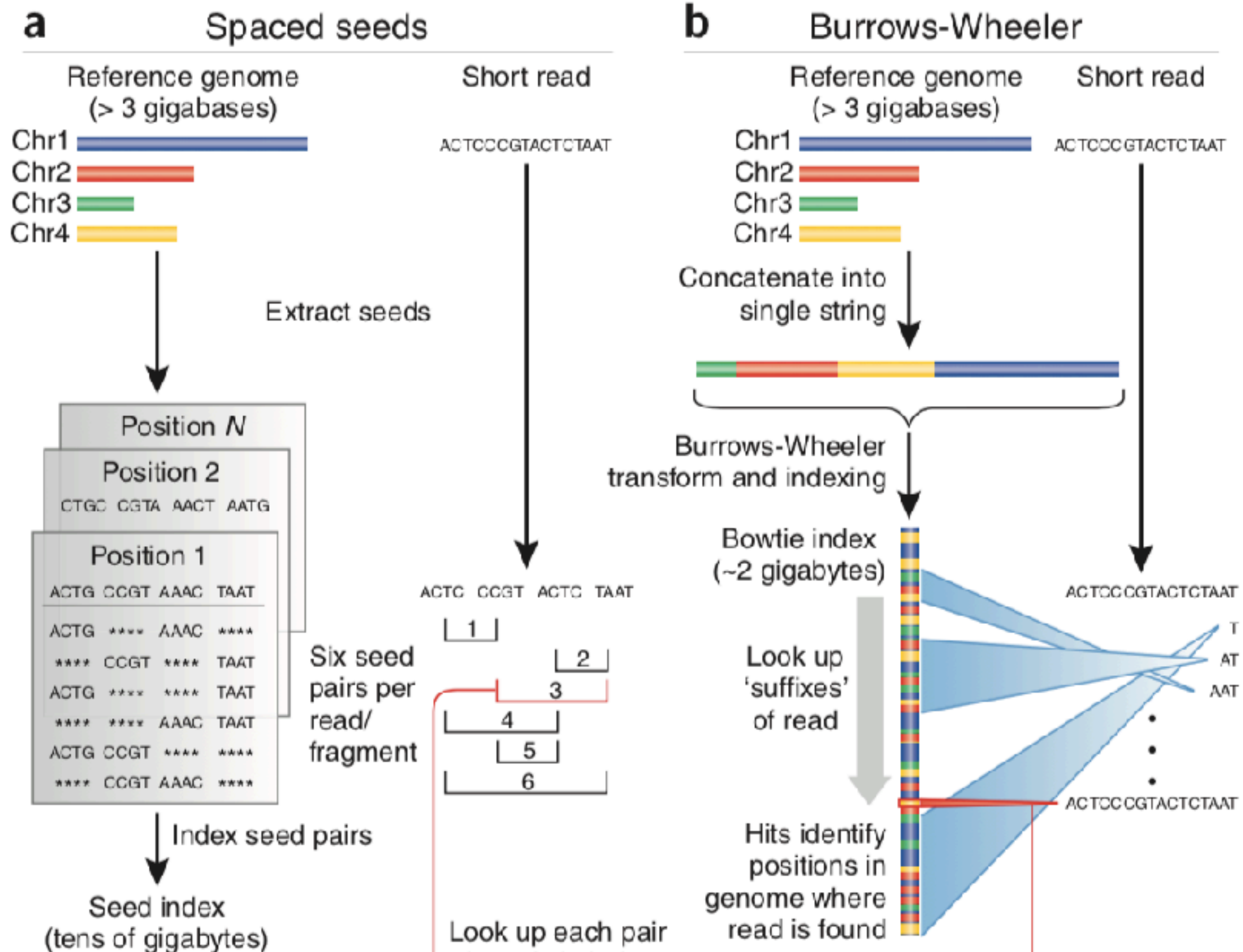


- Árbol de sufijos para la secuencia BANANA
- A\$
- NA\$
- NANA\$
- ANANA\$
- BANANA\$

# Alineamiento basado en árbol de sufijos o “Burrows-Wheeler”



# En resumen



# En resumen

