

# Genoma y Ensamblado

Seminario de Modelos y Métodos cuantitativos (Bioinformática)



Rodrigo Fernández  
Cristián Maureira  
Gabriel Zamora

Universidad Técnica  
Federico Santa María

18 de noviembre de 2010

Genoma

Secuenciación

Shotgun Sequencing

Hierarchical Shotgun Sequencing

Comparación de las técnicas

Conclusiones



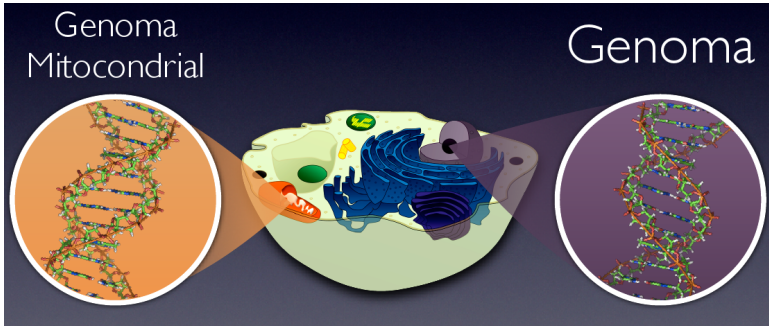
# Genoma



# Genoma

¿Qué es?

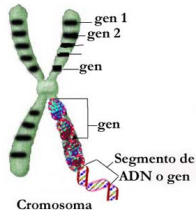
- **Totalidad** de la información genética que posee un organismo en particular.
  - No olviden al genoma Mitocondrial!



# Genoma

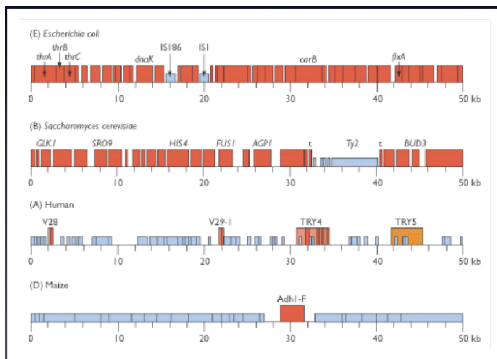
¿Qué contiene?

- **Genes** que codifican **proteínas**.
- “Genes” que codifican **RNA's estructurales**: rRNA, tRNA, etc.
- Secuencias de **control**.
- En general, mucha **basura**.
  - Repeticiones de secuencias.
  - Hartas A y T (en relación a la estructura de la cromatina).
  - Duplicados en desuso.
  - Minisatélites, Microsatélites.
  - Largo etc.



# Genoma

¿Qué contiene?



*E. coli* 90%

*S. cerevisiae* 50%

Human 2%

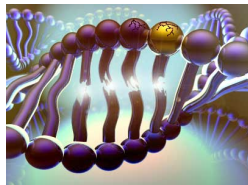
Maize 1%



# Genoma

¿Para que sirve?

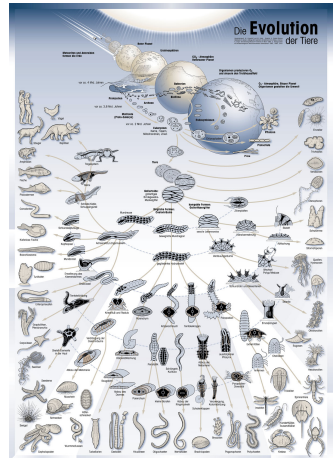
- **Diagnóstico** y **prevención** de enfermedades
- Estudio de **susceptibilidad** en las enfermedades
- Intervención (**tratamiento**) sobre la enfermedad.
  - Enfermedades hereditarias.
  - La “terapia génica”.
  - Fármacos a la medida.



# Genoma

Otros Beneficios...

- Medicina molecular.
- Genómica microbiana.
- Valoraciones de riesgo.
- Bioarqueología, antropología, **evolución...**
- **Identificación ADN.**
- **Agricultura** y **bioprocesamiento.**



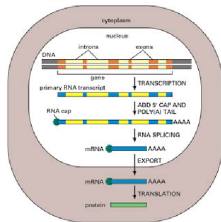


# Genoma

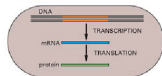
## Eucariotas

- Uno o más  **cromosomas**  de DNA lineal.
- De cada cromosoma, un par en cada célula.
- Todos contenidos en el  **núcleo** .
- Los  **genes**  (codificando proteínas) ocupan un porcentaje pequeño del DNA.
- Los genes suelen estar interrumpidos por  **intrones** .
- **Mayor tamaño**  que en los procariotas.

(A) EUKARYOTES



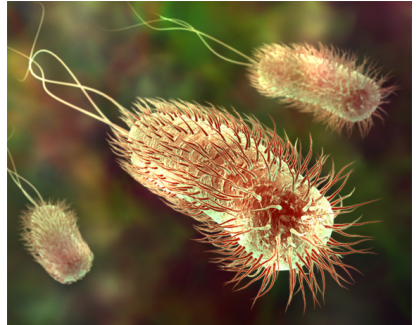
(B) PROCARYOTES



# Genoma

## Procariotas

- Doble hélice Circular.
- **Flota** “por ahí” .
- Suelen haber **plásmidos**.



# Genoma Humano

## Algunos datos

- **38.000** genes (app. **2 %** de todo el genoma).
- 98 % idéntico al de los **chimpancés** y otros primates. 99,99 % de similitud entre seres humanos.
- Se han encontrado 223 genes humanos que resultan similares a los genes **bacterianos**. (hasta ahora)
- **5 %** del genoma codifica **proteínas**.



# Genoma Humano

## Algunos datos

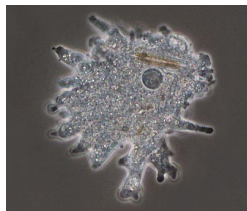
- El **25 %** del genoma humano está casi **desierto**, existiendo largos espacios libres entre un gen y otro.
- **35 %** + del genoma contiene secuencias **repetidas** (ADN basura)
- **250-300.000** proteínas distintas.
  - Por tanto cada gen podría estar implicado por término medio en la síntesis de unas **diez** proteínas.
- Entre **1,4 a 2,1** millones de **SNP** en el genoma. (hasta ahora)



# Genoma

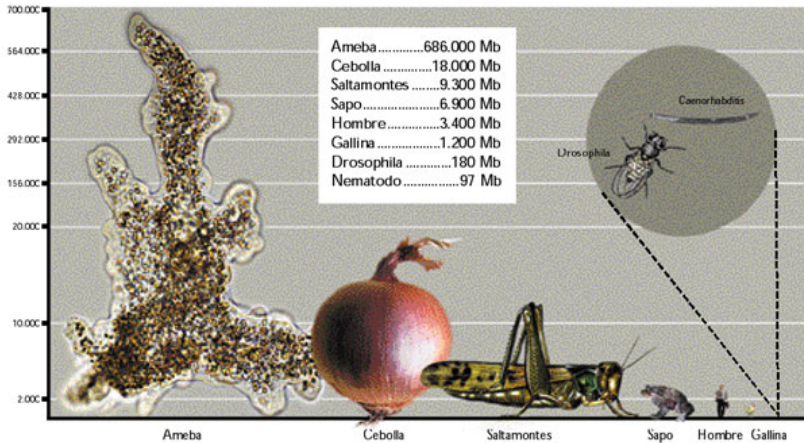
## Información General

- El récord en tamaño de genoma lo tiene la **ameba**, con **670 Gbp** (versus 3 Gbp del humano).
- El récord de **cantidad de genes** va para un cierto **protozoo**, con **60.000** (el triple que nosotros).



# Genoma

## Información General



# Genoma

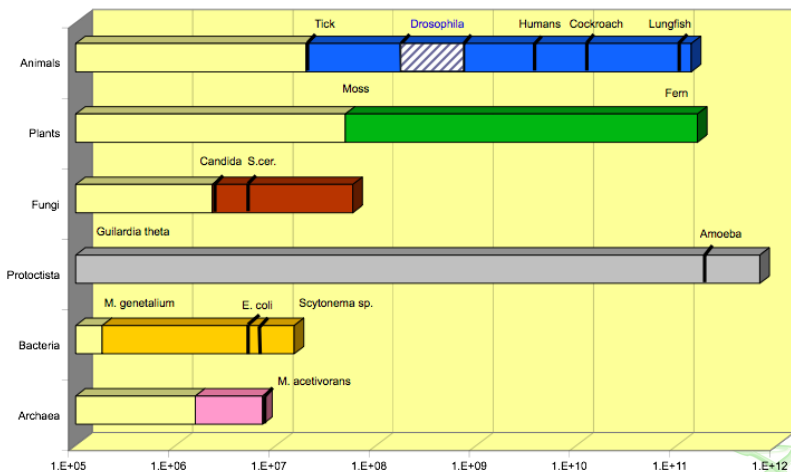
## Información General 2

- Las **vacas** tienen **60** cromosomas; el **pez dorado**, alrededor de **100**.
- Lo único que sí es cierto es que en **procariotas** el tamaño oscila **entre 1 y 100** Mbp.
- Y que las cosas que **vuelan** suelen tener genomas relativamente **pequeños**.



# Genoma

## Información General 3





# Secuenciación de ADN

¿Qué es?



# Secuenciación de ADN

¿Qué es?

Video



# Secuenciación de ADN

¿Qué es?

- Conjunto de métodos y técnicas bioquímicas, cuya **finalidad** es la **determinación del orden** de los nucleótidos (A, C, G y T) en un oligonucleótido de ADN.
- La secuencia de ADN constituye la **información genética heredable** (en núcleo celular, plásmidos, mitocondria, y cloroplastos) **base de desarrollo** de los seres vivos.



# Secuenciación de ADN

¿Para qué sirve?



# Secuenciación de ADN

¿Para qué sirve?

- Comprensión.
- **Atajo!** ayuda a los científicos a encontrar a los genes más fácil y rápidamente.
- Hay **pistas** para encontrar los genes.
- Se espera que sirva para comprender el **todo**.

*“...cómo los genes trabajan juntos para dirigir el crecimiento, desarrollo y mantenimiento de un organismo entero.”*



# Secuenciación de ADN

¿Para qué sirve?

- Los genes son el 5 % del ADN, la secuencia **ayudará** a estudiar la parte del genoma por afuera de los genes.
  - Sitios **reguladores** que controlan la activación/desactivación de genes.
  - ADN **sin** sentido.



# Secuenciación de ADN

¿Cómo se hace?



# Secuenciación de ADN

¿Cómo se hace?

- Se secuencia en *partes*.
- **No** se puede completo (**limitación** métodos sólo con fragmentos)
- Fragmentar en pedazos *pequeños*, secuenciarlos y reconstruirlos en orden.
- Dos técnicas más utilizadas (*con ventajas y desventajas*).
  - Shotgun.
  - Hierarchical Shotgun.
- La secuenciación del genoma humano ha involucrado una combinación de **ambas** técnicas.





# Secuenciación de ADN

¿Cómo funciona?



# Secuenciación de ADN

¿Cómo funciona?

1. Electroforesis para **separar** los fragmentos de ADN.
  - *Técnica para la separación de moléculas según movilidad en un campo eléctrico.*



# Secuenciación de ADN

¿Cómo funciona?

1. Electroforesis para **separar** los fragmentos de ADN.
  - *Técnica para la separación de moléculas según movilidad en un campo eléctrico.*
2. Por lo que el ADN a secuenciar se coloca en un **extremo** del gel.



# Secuenciación de ADN

¿Cómo funciona?

1. Electroforesis para **separar** los fragmentos de ADN.
  - *Técnica para la separación de moléculas según movilidad en un campo eléctrico.*
2. Por lo que el ADN a secuenciar se coloca en un **extremo** del gel.
3. Electrodo en los extremos (aplicamos **corriente**).



# Secuenciación de ADN

¿Cómo funciona?

1. Electroforesis para **separar** los fragmentos de ADN.
  - *Técnica para la separación de moléculas según movilidad en un campo eléctrico.*
2. Por lo que el ADN a secuenciar se coloca en un **extremo** del gel.
3. Electrodo en los extremos (aplicamos **corriente**).
4. Migración de fragmentos de ADN por **tamaño** (a través del gel).



# Secuenciación de ADN

¿Cómo funciona?

- Electroforesis es capaz **sólo** de separar fragmentos de 500 pares de bases (justificación anterior!).
- **Dato:** Hasta 1980, las electroforesis eran leídas por una **persona**.
- Hoy en día el proceso es *automático*.
- En producir una secuencia de 20.000 a 50.000 bases (Persona → 1 año, Máquina → horas).
- Máquinas con diseño **basado** en el proceso manual.



# Secuenciación de ADN

¿Cómo funciona?

- Usando máquinas.

1. Se ubica el gel en un espacio entre dos vidrios de medio milímetro.



# Secuenciación de ADN

¿Cómo funciona?

- Usando máquinas.
  1. Se ubica el gel en un espacio entre dos vidrios de medio milímetro.
  2. Se espera que el gel se endurezca





# Secuenciación de ADN

¿Cómo funciona?

- Usando máquinas.
  1. Se ubica el gel en un espacio entre dos vidrios de medio milímetro.
  2. Se espera que el gel se endurezca
  3. El ADN se siembra en cada una de las 96 calles que corren a lo largo del gel.



# Secuenciación de ADN

¿Cómo funciona?

- Usando máquinas.
  1. Se ubica el gel en un espacio entre dos vidrios de medio milímetro.
  2. Se espera que el gel se endurezca
  3. El ADN se siembra en cada una de las 96 calles que corren a lo largo del gel.
  4. Se aplica corriente eléctrica.



# Secuenciación de ADN

¿Cómo funciona?

- Usando máquinas.

1. Se ubica el gel en un espacio entre dos vidrios de medio milímetro.
2. Se espera que el gel se endurezca
3. El ADN se siembra en cada una de las 96 calles que corren a lo largo del gel.
4. Se aplica corriente eléctrica.
5. Los fragmentos de ADN migran de acuerdo a su tamaño.



# Secuenciación de ADN

¿Cómo funciona?

## ■ Usando máquinas.

1. Se ubica el gel en un espacio entre dos vidrios de medio milímetro.
2. Se espera que el gel se endurezca
3. El ADN se siembra en cada una de las 96 calles que corren a lo largo del gel.
4. Se aplica corriente eléctrica.
5. Los fragmentos de ADN migran de acuerdo a su tamaño.
6. La máquina lee el orden de las bases de ADN y guarda la información.



## Secuenciación de ADN

¿Cómo se reconoce una base?



# Secuenciación de ADN

¿Cómo se reconoce una base?

- Secuenciadores automáticos **no** ven al ADN (hay que **prepararlo**)
- El ADN debe estar
  - fragmentado
  - copiado
  - modificado químicamente
  - unido a fragmentos fluorescentes (por bases)
- Luego del procedimiento, se reconocen por la **fluorescencia** (láser).



## Secuenciación de ADN

¿Qué ocurre después?



# Secuenciación de ADN

¿Qué ocurre después?

- Salida del secuenciador automático.
  - Llegan a una **secuencia cruda** (agujeros, errores y ambigüedades).
  - Proceso de **limpieza** y **ordenamiento** (cuidado con el tiempo)





# Secuenciación de ADN

¿Qué ocurre después?

- Ordenamiento.
  - Mediante un programa.
  - El programa busca y analiza **superposiciones** o secuencias **idénticas** de ADN en diferentes fragmentos.
  - Los fragmentos que contienen superposiciones se encuentran **juntos** en la secuencia final.



# Secuenciación de ADN

¿Qué ocurre después?

- Aceptar o rechazar.
  - Genoma no secuenciado **no** garantiza ser correcto.
  - Pueden ocurrir **errores** (fragmenta el ADN, copia, secuencia u ordena).
  - **Truco**: secuenciar el genoma más de una vez.
  - Luego del ordenamiento, sigue el proceso de terminación. (no tenemos intuición)



## Secuenciación de ADN

¿Qué hace la secuenciación del genoma humano distinto?



# Secuenciación de ADN

¿Qué hace la secuenciación del genoma humana distinto?

- El genoma humano es mucho más **grande** que los genomas que se habían secuenciado anteriormente.
- Los anteriores secuenciados eran virus, bacterias u otros organismos simples.
- Genoma humano posee 25 % a 50 % de ADN repetitivo, bacterias y virus no mucho.
- El ADN repetitivo no sólo es difícil de ordenar sino también de secuenciar.



# Técnicas Shotgun Sequencing



# Shotgun Sequencing

¿Qué es?

- Método de secuenciado de ADN, desarrollado y preferido por “Celera Genomics”, una empresa estadounidense.
- Conocido también como “shotgun cloning”



# Shotgun Sequencing

## Ventajas y Desventajas

- Ventajas:
  - Moderadamente rápido en procesamiento
- Desventajas:
  - Problemas con secuencias con repetición.
  - No detecta secuencias repetidas. Para aquello se basa en datos públicos.



# Shotgun Sequencing

Descripción

Video

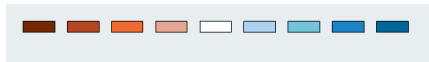




# Shotgun Sequencing

## Descripción

Tomando como ejemplo un genoma trozado en 9 partes:



# Shotgun Sequencing

Descripción

¿Cuántas posibles combinaciones  
podemos formar?



# Shotgun Sequencing

Descripción

$$9 * 8 * 7 * \dots * 2 * 1 = 9! = 362880$$



# Shotgun Sequencing

## Descripción

Dejando de lado la cantidad.

¿Como saber cual es la combinación  
correcta?



# Shotgun Sequencing

## Descripción

A C T C C T G C T A C



# Shotgun Sequencing

## Descripción

A C T C C T G C T A C

A C T / C C T / G C T / A C round 1

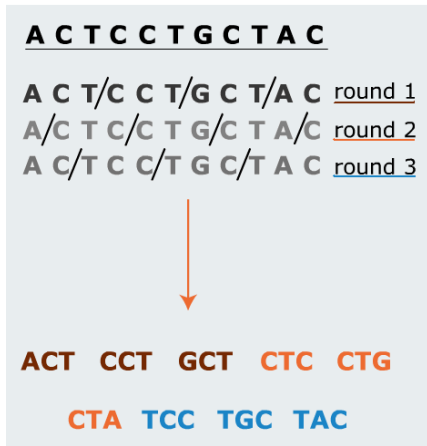
A / C T C / C T G / C T A / C round 2

A C / T C C / T G C / T A C round 3



# Shotgun Sequencing

## Descripción



# Shotgun Sequencing

## Descripción

Luego se crea un grafo dirigido:

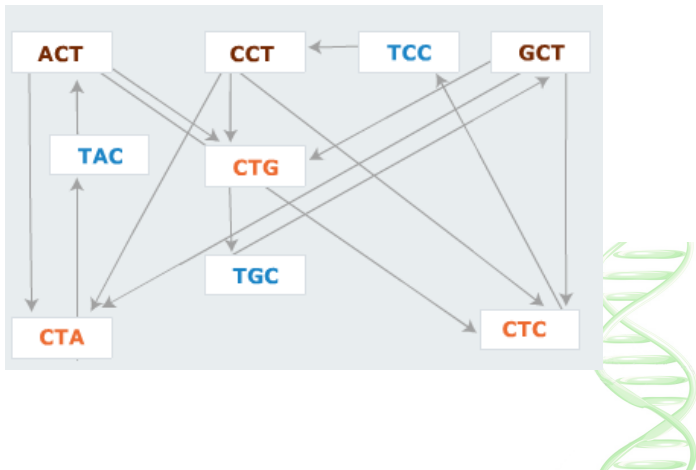
- Sea  $x_i = c_{i1}c_{i2}...c_{in-1}c_{in}$  el contenido de un estado  $i$  en el grafo
- Para ir de un estado  $i$  a un estado  $j$ , es necesario que:  
$$x_{in-1}x_{in} = x_{j1}x_{j2}$$
- Finalmente se busca un ciclo hamiltoneano en dicho grafo, correspondiente a la secuencia buscada.





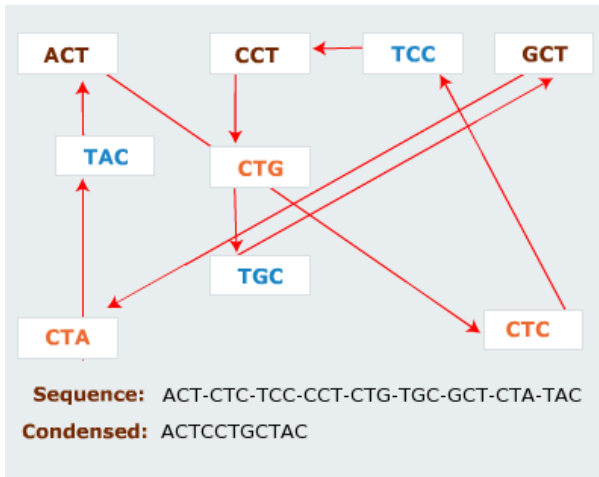
# Shotgun Sequencing

## Descripción



# Shotgun Sequencing

## Descripción



# Técnicas

## Hierarchical Shotgun Sequencing



# Hierarchical Shotgun Sequencing

¿Qué es?

- También conocida como técnica del “*clon-por-clon*”
- Es el método **preferido** por el Proyecto del Genoma Humano.
- Por lo general es utilizado en genomas con **repeticiones**.
- Es **recomendable** utilizar en Eucariontes en general.
- Utiliza enzimas en su proceso.



# Hierarchical Shotgun Sequencing

## Ventajas y Desventajas

### ■ Ventajas:

- Detecta secuencias repetidas.
- No tiene problemas de regiones poco representadas.

### ■ Desventajas:

- Demora mucho tiempo.
- Posee mucho costo computacional.



# Hierarchical Shotgun Sequencing

¿Cómo funciona?

1. Construir una **genoteca** por fragmentación del ADN (utilizando BACs)
  - BAC : Bacterial Artificial Chromosomes.



# Hierarchical Shotgun Sequencing

¿Cómo funciona?

1. Construir una **genoteca** por fragmentación del ADN (utilizando BACs)
  - BAC : Bacterial Artificial Chromosomes.
2. Los fragmentos de ADN representados en la misma son organizados en un **mapa físico**.



# Hierarchical Shotgun Sequencing

¿Cómo funciona?

1. Construir una **genoteca** por fragmentación del ADN (utilizando BACs)
  - BAC : Bacterial Artificial Chromosomes.
2. Los fragmentos de ADN representados en la misma son organizados en un **mapa físico**.
3. Elección al azar de colonias y se **replican**.





# Hierarchical Shotgun Sequencing

¿Cómo funciona?

1. Construir una **genoteca** por fragmentación del ADN (utilizando BACs)
  - BAC : Bacterial Artificial Chromosomes.
2. Los fragmentos de ADN representados en la misma son organizados en un **mapa físico**.
3. Elección al azar de colonias y se **replican**.
4. Los clones individuales son **seleccionados al azar** y secuenciados.



# Hierarchical Shotgun Sequencing

¿Cómo funciona?

1. Construir una **genoteca** por fragmentación del ADN (utilizando BACs)
  - BAC : Bacterial Artificial Chromosomes.
2. Los fragmentos de ADN representados en la misma son organizados en un **mapa físico**.
3. Elección al azar de colonias y se **replican**.
4. Los clones individuales son **seleccionados al azar** y secuenciados.
5. Para el paso anterior es necesario hacer una **nueva genoteca** de trozos más pequeños



# Hierarchical Shotgun Sequencing

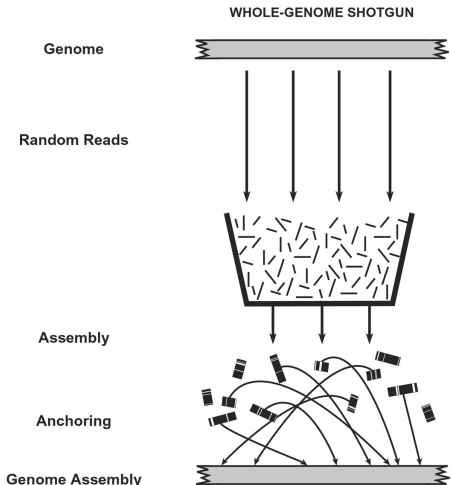
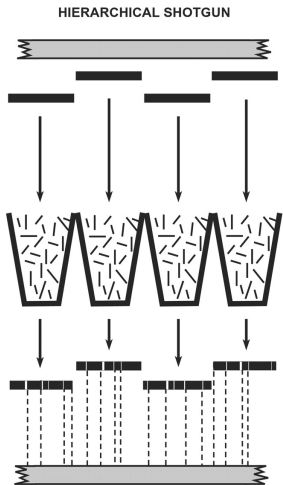
¿Cómo funciona?

1. Construir una **genoteca** por fragmentación del ADN (utilizando BACs)
  - BAC : Bacterial Artificial Chromosomes.
2. Los fragmentos de ADN representados en la misma son organizados en un **mapa físico**.
3. Elección al azar de colonias y se **replican**.
4. Los clones individuales son **seleccionados al azar** y secuenciados.
5. Para el paso anterior es necesario hacer una **nueva genoteca** de trozos más pequeños
6. La secuencia de los clones se **ensambla finalmente** para reconstruir la secuencia del genoma.



# Hierarchical Shotgun Sequencing

¿Cómo funciona?



# Hierarchical Shotgun Sequencing

Video

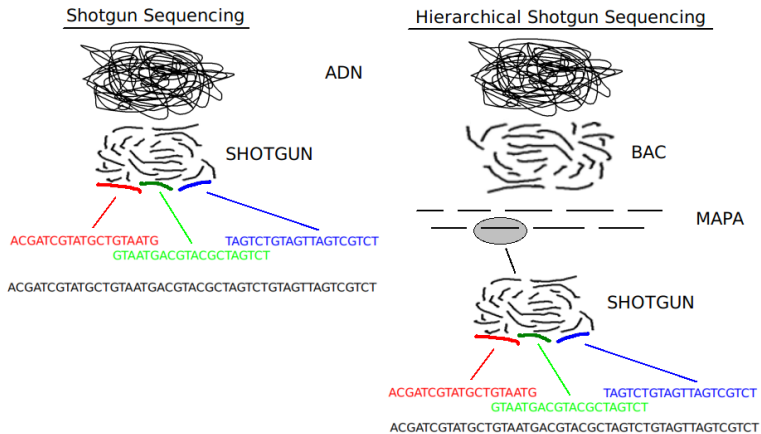


# Comparación



# Hierarchical Shotgun vs Shotgun

Principales diferencias



# Hierarchical Shotgun vs Shotgun

¿Cuál es mejor?

- Depende del tamaño del genoma
- Depende de la complejidad del genoma
- Depende si se desea exactitud o velocidad.





¿Preguntas?

