# Lecture 7:

## Sequencing strategies

### BAC to BAC Run vs. Whole Genome Shotgun

**Course 485**

**Introduction to Genomics**

# Challenges

1) We can not sequence the entire genome at once.

**Any ideas why?**
**Technological limitations?**
**Natural limitations?**

2) What is after the sequences obtained.

**Genome assembly?**

# Review

**Across sequencing methodologies, what do we need to sequence a fragment/piece of DNA?**

1. Enough identical copies of the DNA fragment.

2. Primers.

# Review

**Across sequencing methodologies, what do we need to sequence a fragment/piece of DNA?**

1. Enough identical copies of the DNA fragment.

Why?

2. Primers.

# Review

**Across sequencing methodologies, what do we need to sequence a fragment/piece of DNA?**

1. Enough identical copies of the DNA fragment.

Why?

What sequencing technology does not require many copies of DNA fragment?

2. Primers.

# Review

**Across sequencing methodologies, what do we need to sequence a fragment/piece of DNA?**

1. Enough identical copies of the DNA fragment.

Why?

What sequencing technology does not require many copies of DNA fragment?

2. Primers.

Why?

# Review

**Across sequencing methodologies, what do we need to sequence a fragment/piece of DNA?**

1. Enough identical copies of the DNA fragment.

Why?

What sequencing technology does not require many copies of DNA fragment?

2. Primers.

Why?

What sequencing technology does not require primers?

# Review

**How can we get enough copies of an identical DNA fragment?**

1. Molecular cloning.

2. PCR.

# Review

**How can we get enough copies of an identical DNA fragment?**

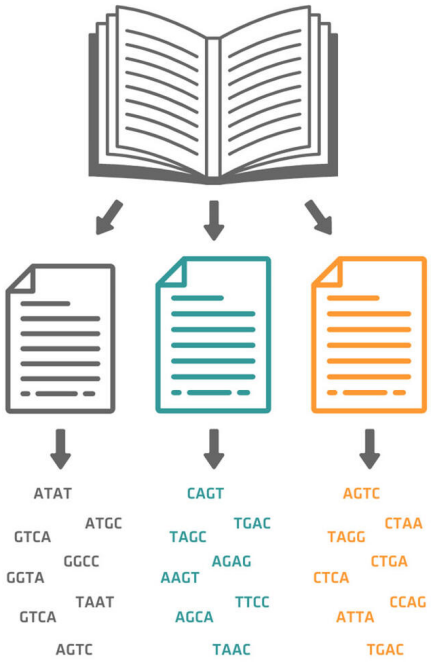
1. Molecular cloning.

2. PCR.

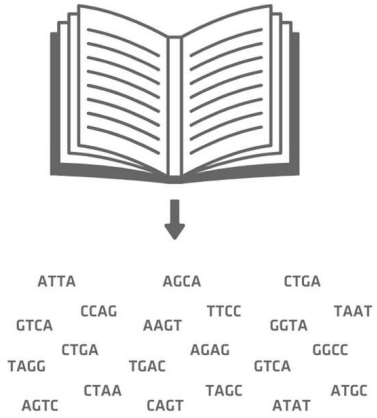
What is the most important requirement for PCR?

# Sequencing strategies

## Hierarchical (BAC to BAC) sequencing




## Whole genome shotgun (WGS) sequencing

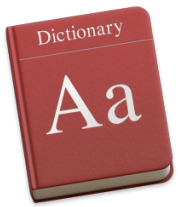


# What is hierarchical?

**hi·er·ar·chi·cal** | ˌhī(ə)ˈrärkək(ə)l |

adjective

of the nature of a hierarchy; arranged in order of rank:
*the hierarchical bureaucracy of a local authority.*



# Hierarchical sequencing

1. Genome

2. Genome fragmentation

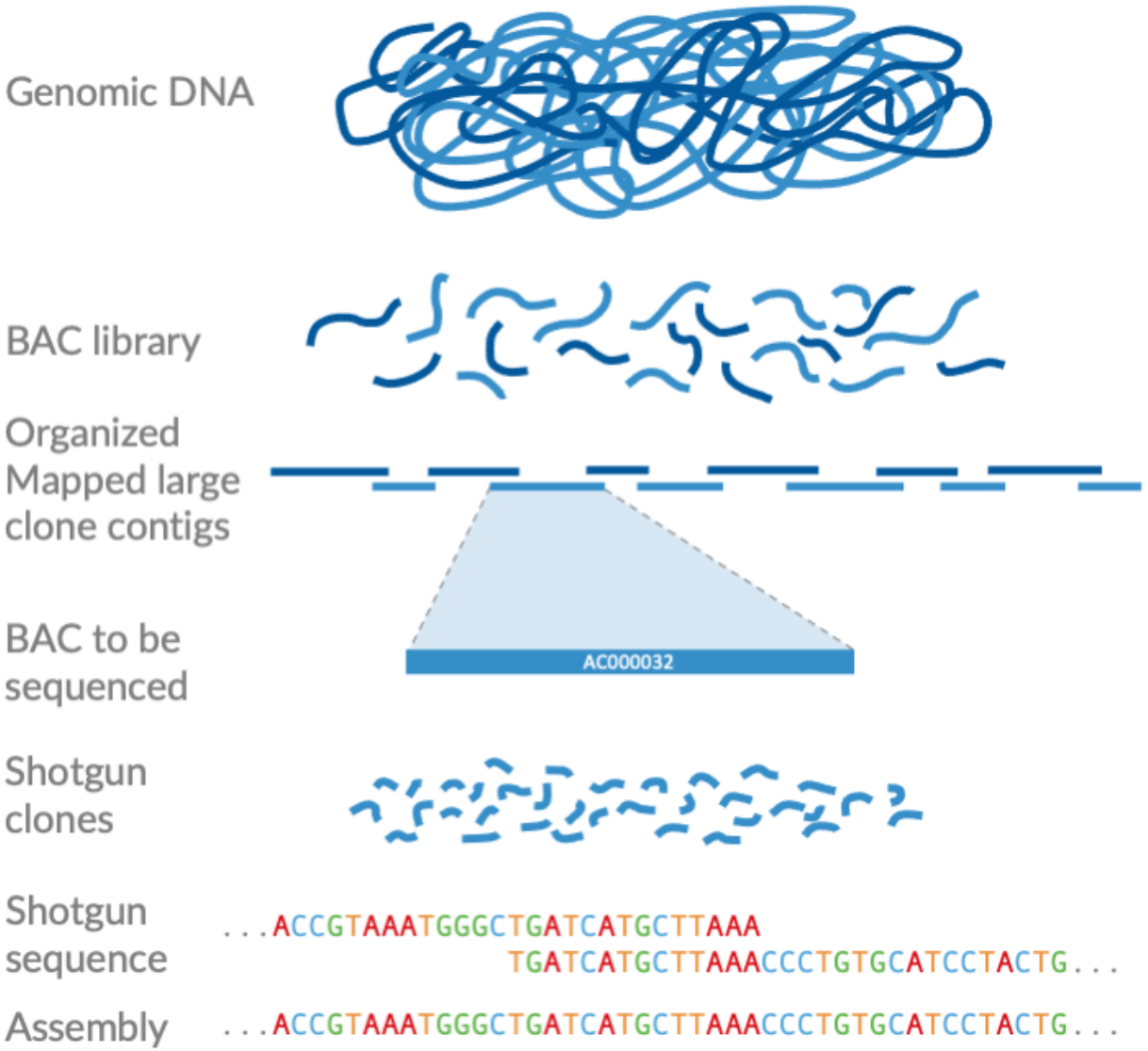
3. Cloning genome fragments into multiple BAC clones

4. Single BAC clone fragmentation

5. Cloning each BAC clone fragment into a sequencing vector

6. Sequencing ends of the DNA fragment

# Hierarchical sequencing



# Hierarchical sequencing

1. Genome

Where do we get the genome from?

How many copies of the genome are in the sample extraction tube?

Can we sequence the genome directly from the extraction tube?

# Hierarchical sequencing

## 2. Genome fragmentation

How can we fragment a genome?

What are restriction enzymes?

Can we predict/estimate the number of restriction sites in an unknown genome?

# Hierarchical sequencing

%AT = 40%

## 2. Genome fragmentation

Let's assume that a genome is **1Mb** in size and a %GC of **60%**. How many *EcoRI* sites are likely to be in the genome?

Genome size =

1Mb = $1x10^6$bp = 1,000,000bp

P(G) = 0.3

P(C) = 0.3

P(A) = 0.2

P(T) = 0.2


P(GAATTC) = 0.3x0.2x0.2x0.2x0.2x0.3 = 0.000144

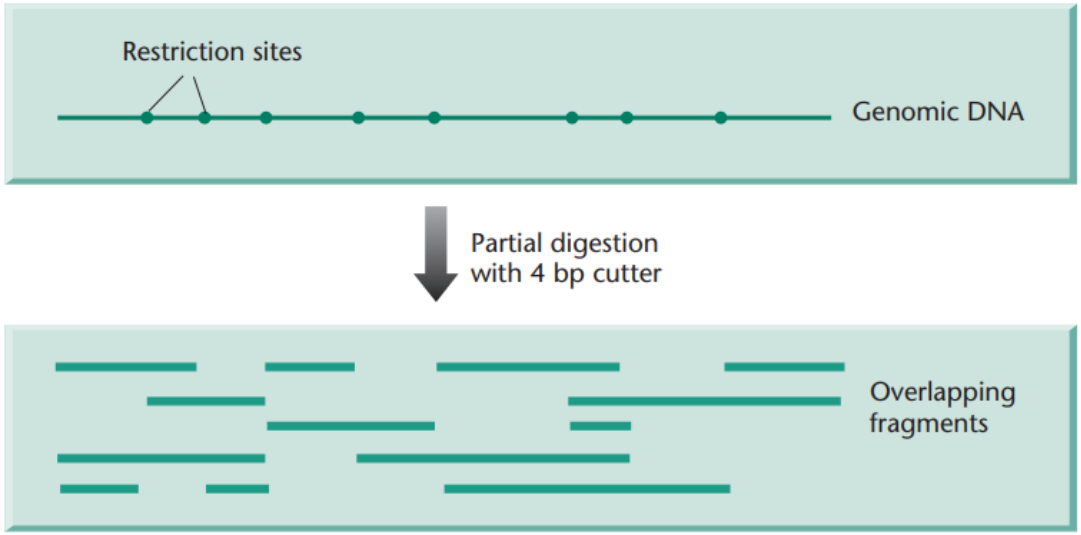
# of bp to find GAATTC = 1/0.000144 = 6944bp

# of GAATTC in 1Mb genome = 1000,000/6944 = 144

# Hierarchical sequencing

## 2. Genome fragmentation

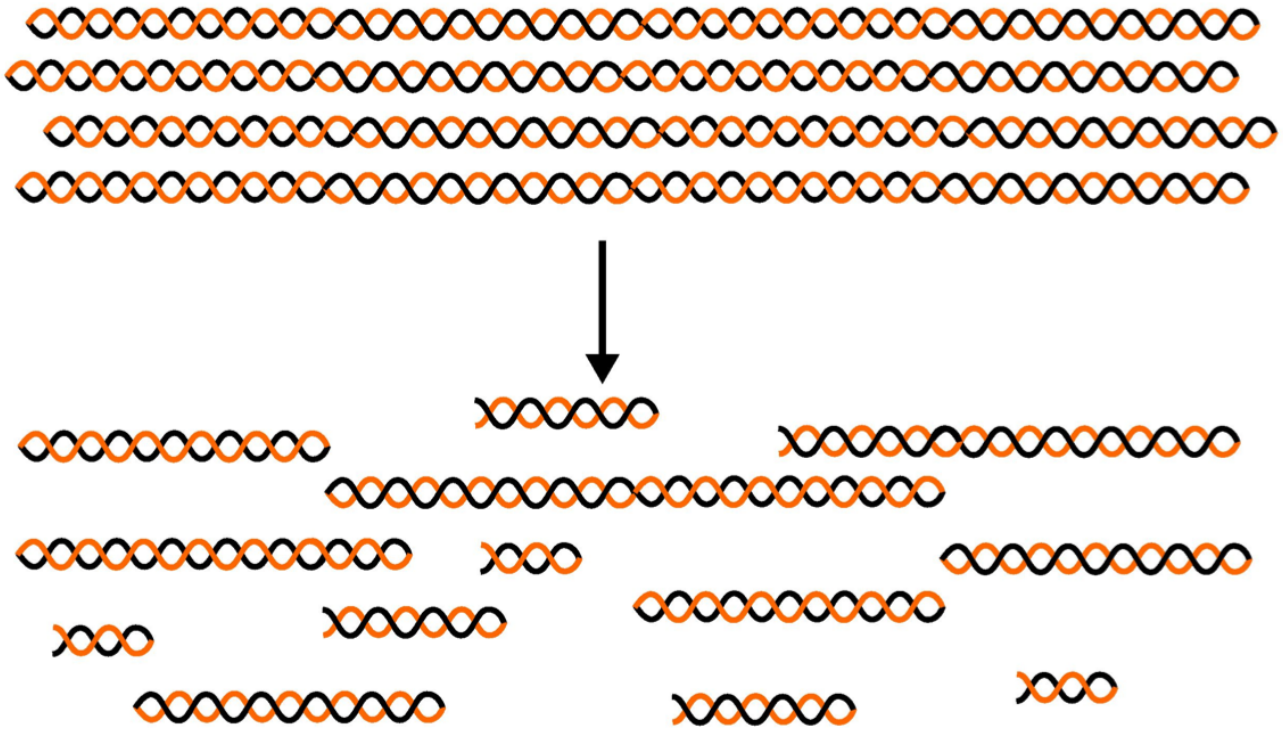
Complete vs. partial digestion
Which approach is better and why?



# Hierarchical sequencing

## 2. Genome fragmentation

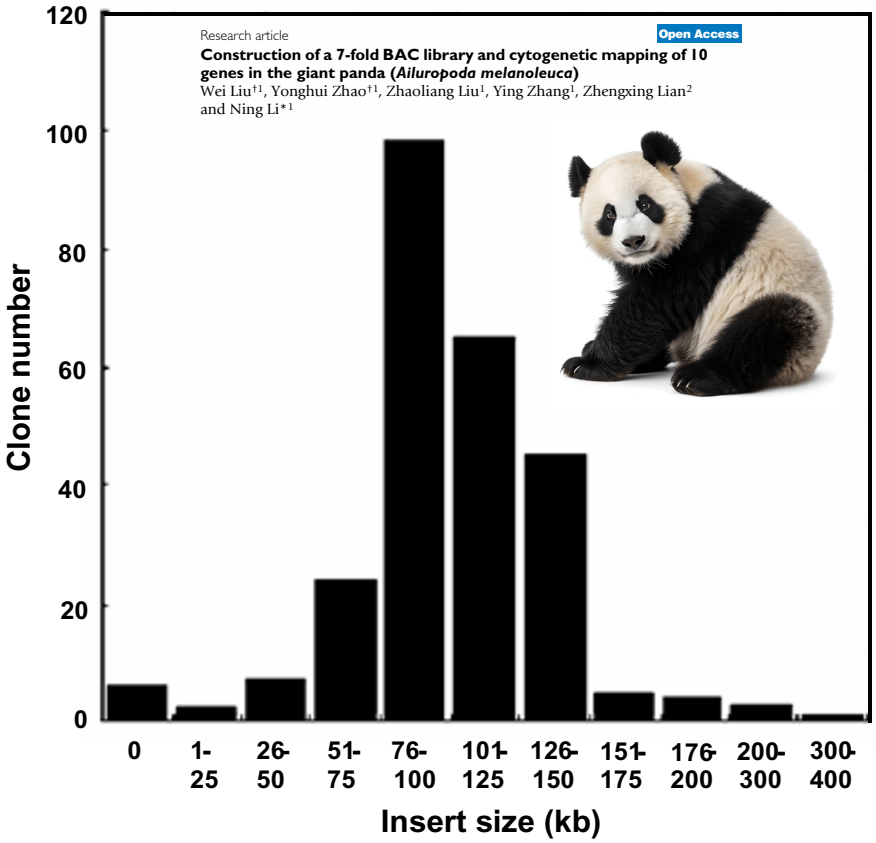
Using multiple different restriction enzymes
why?

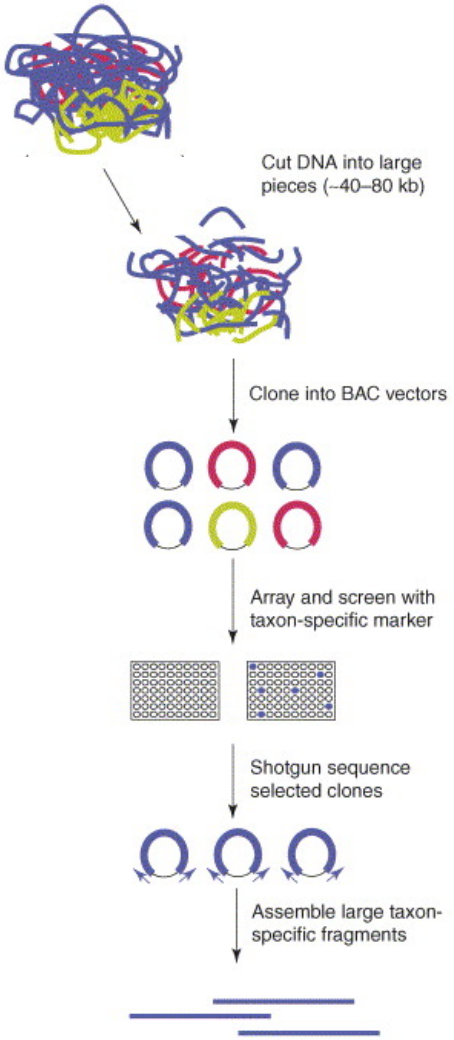


# Hierarchical sequencing

## 2. Genome fragmentation

What is the size of the needed DNA fragment?



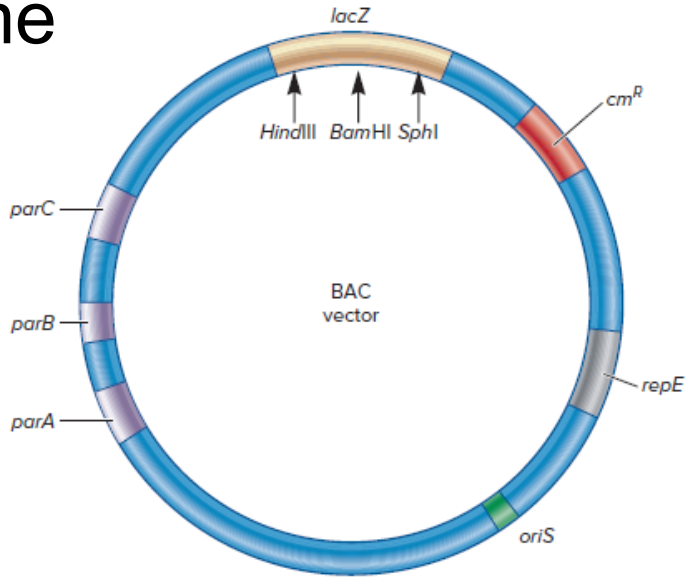



# Hierarchical sequencing

## 3. Cloning genome fragments into multiple BAC clones

What is a BAC?

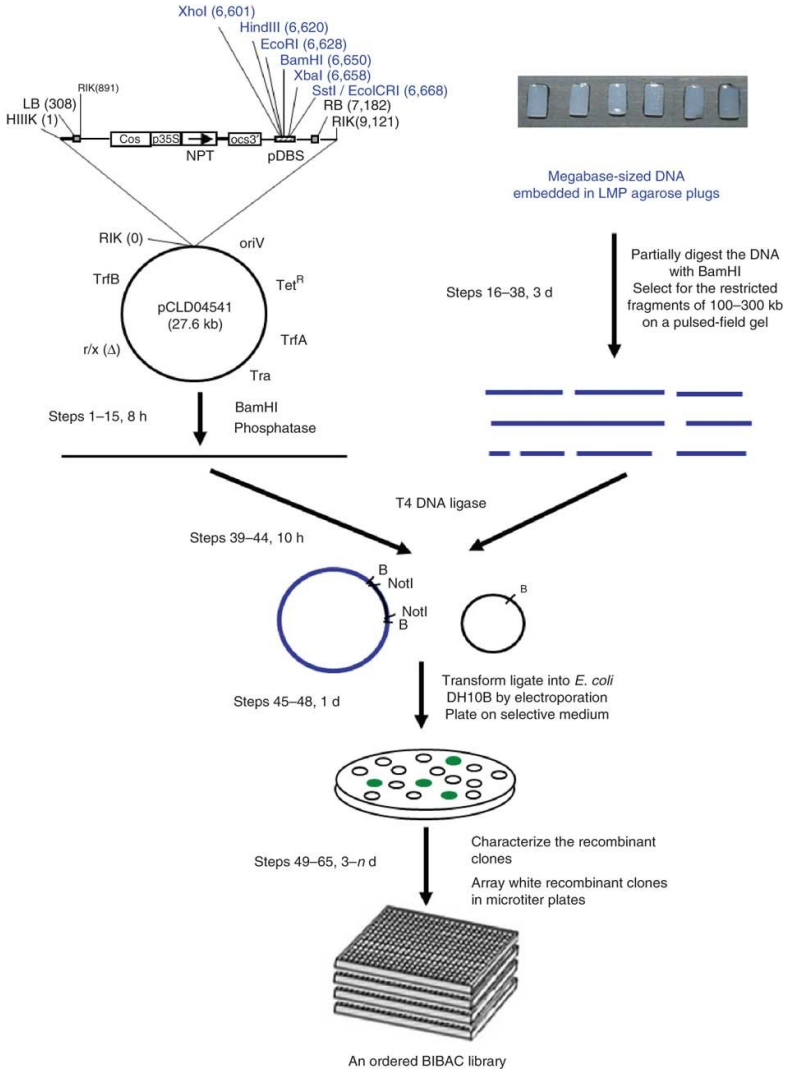
- Bacterial Artificial Chromosome
- Insert size **150-350Kb**
- Used for physical mapping and hierarchical sequencing.
- More stable and easy to manipulate.



# Hierarchical sequencing

## 3. Cloning genome fragments into multiple BAC clones

What is a BAC clone library?



# Hierarchical sequencing

## 3. Cloning genome fragments into multiple BAC clones

Let's assume that a genome is 1Gb in size. How many clones of an average 120Kb fragments are needed to achieve 5 fold coverage of the genome?

Genome size =1Gb = $1x10^9$bp = 1,000,000,000bp

Average insert size = 120Kb = $120x10^3$bp = 120,000bp

Coverage = 5X

$$N = \frac{\text{Genome size x Coverage (X)}}{\text{Average insert size}}$$

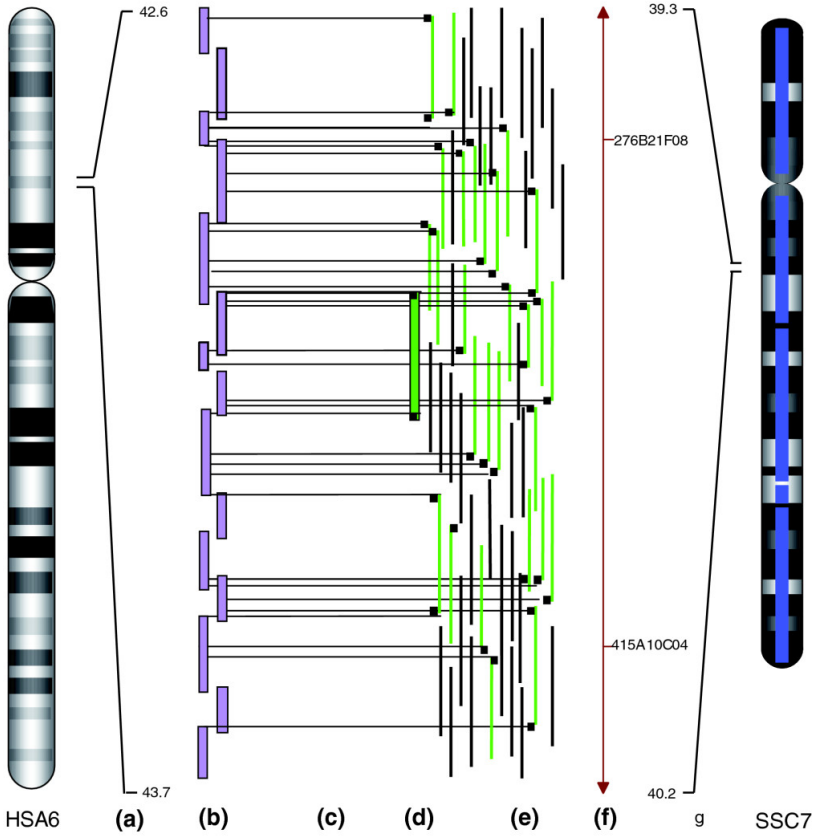
$$N = \frac{1x10^9\text{bp} \times 5}{120x10^3\text{bp}} = \frac{5x10^9\text{bp}}{120x10^3\text{bp}} = 41{,}666 \text{ clones}$$

# Hierarchical sequencing

## 3. Cloning genome fragments into multiple BAC clones

What are the benefits of BAC clones?

- Allows the mapping of the unknown fragments to specific locations of the chromosome.
- This can greatly affect the genome assembly (later).

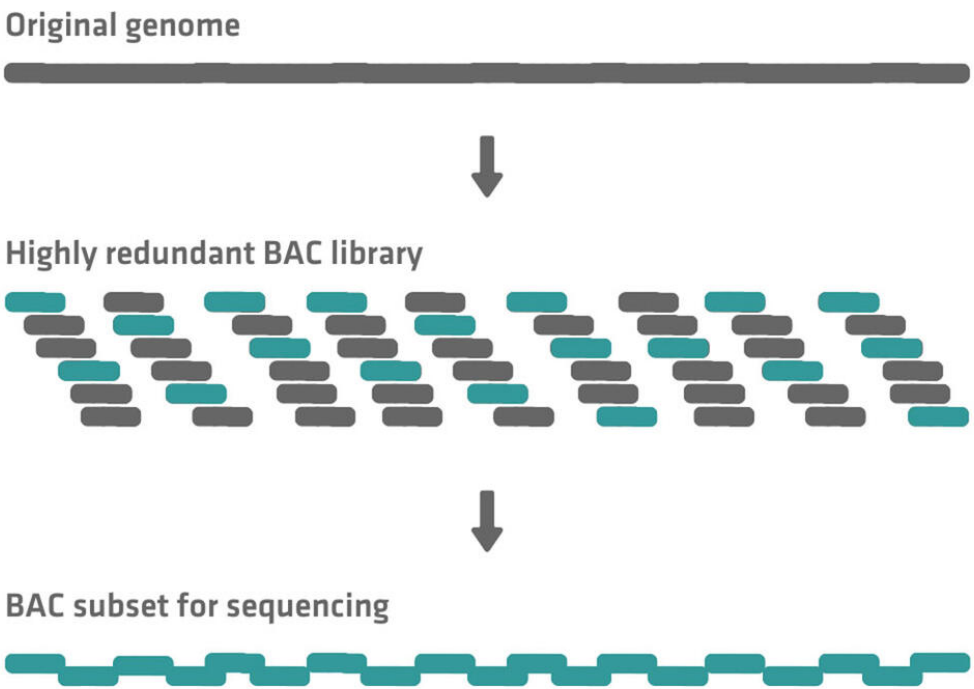


# Hierarchical sequencing

## 3. Cloning genome fragments into multiple BAC clones

Do we need all BAC clones generated?

- Remove redundant clones.
- Select minimal set of **overlapping** clones covering the whole-genome

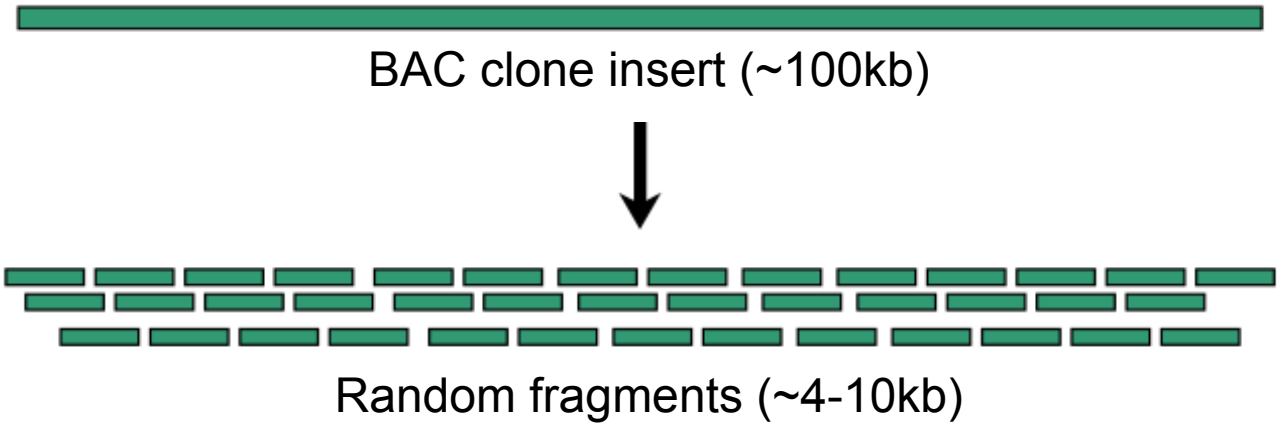


# Hierarchical sequencing

## 4. Single BAC clone-insert fragmentation

Fragmentation of BAC clone-insert by shotgun

Do you what a shotgun is?



# Hierarchical sequencing

## 4. Single BAC clone-insert fragmentation

Aldeehani double trap-Olympic gold medal athlete

# Hierarchical sequencing

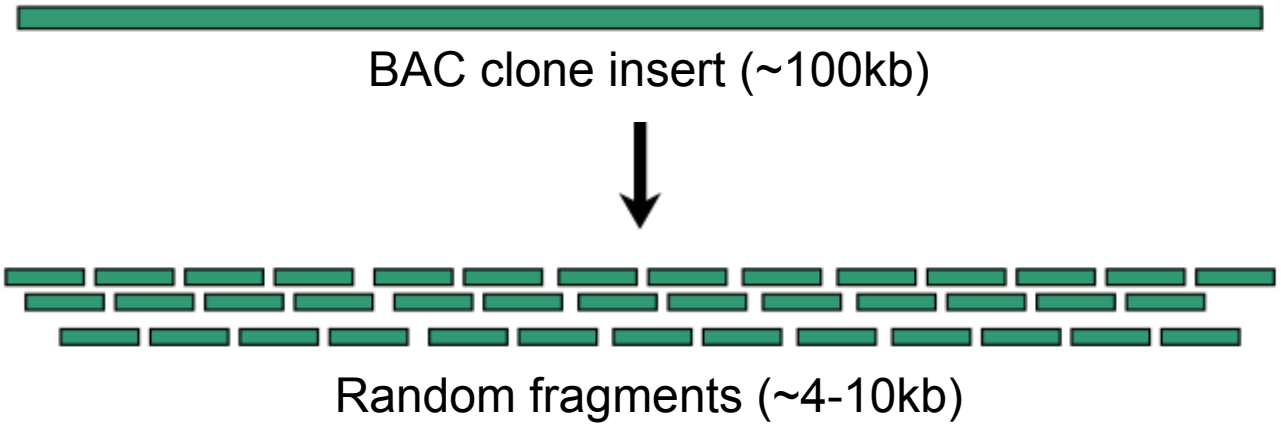
## 4. Single BAC clone-insert fragmentation

Shotgun!

# Hierarchical sequencing

## 4. Single BAC clone-insert fragmentation

How is the shotgun fragmentation is achieved?

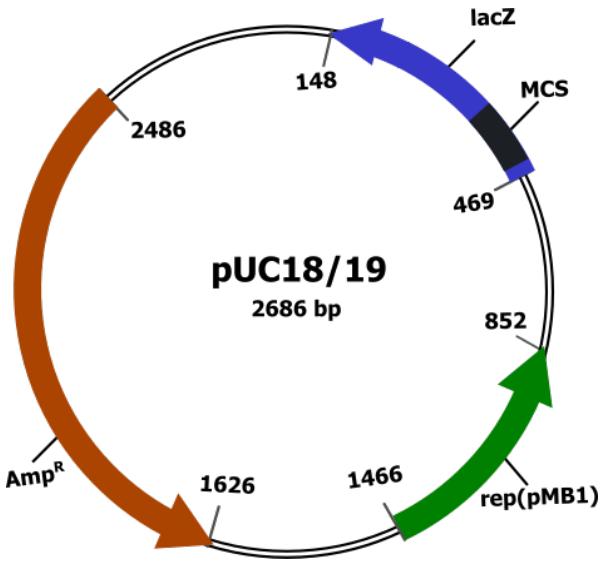


# Hierarchical sequencing

## 5. Cloning each BAC clone fragment into a sequencing vector

What is a sequencing vector?

Why are we cloning smaller fragments?

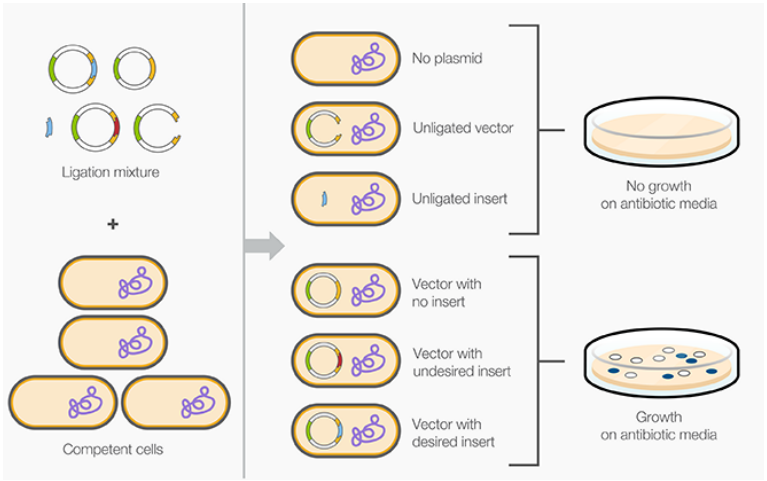


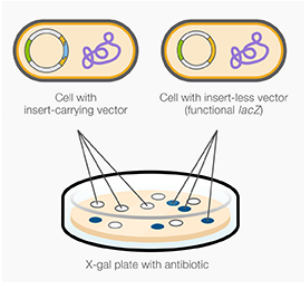

- Smaller than BACs
- Smaller insert size
- Known sequence flanking the multiple cloning site (MCS)

# Hierarchical sequencing

## 5. Cloning each BAC clone fragment into a sequencing vector

How many clones do you need?

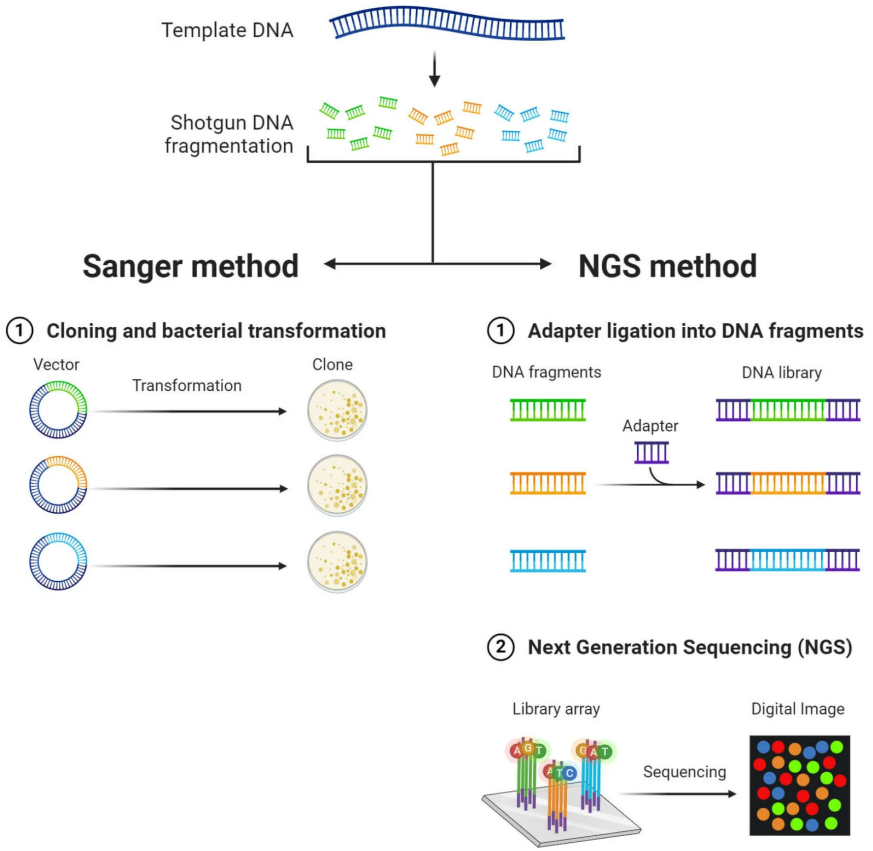


# Hierarchical sequencing

## 5. Cloning each BAC clone fragment into a sequencing vector

Is there an easier way to amplify the random fragments and have known flanking sequences for sequencing primers?



# Hierarchical sequencing

**How do we get enough copies of DNA to start sequencing?**

**Now:**

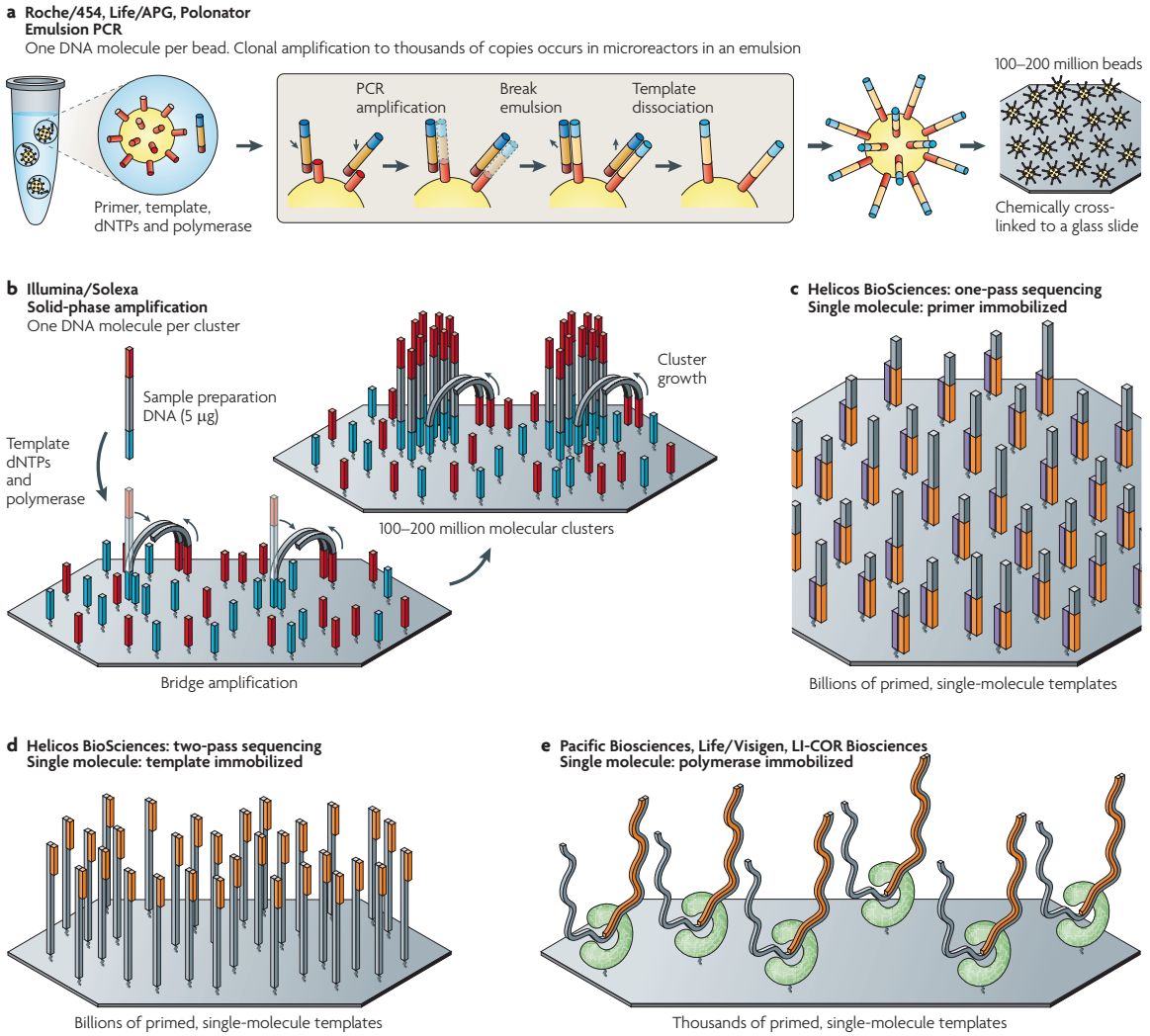
1) Performing PCR using primers of known sequences.
2) Adding synthetic adapters of known sequence to act as primers

**In the past:**

cloning the DNA fragment to vectors (plasmids or Bacterial Artificial Chromosomes (BAC)).

# Hierarchical sequencing

## 5. Cloning each BAC clone fragment into a sequencing vector




APPLICATIONS OF NEXT-GENERATION SEQUENCING

### Sequencing technologies — the next generation

*Michael L. Metzker*[*‡]

Abstract | Demand has never been greater for revolutionary technologies that deliver fast, inexpensive and accurate genome information. This challenge has catalysed the development of next-generation sequencing (NGS) technologies. The inexpensive production of large volumes of sequence data is the primary advantage over conventional methods. Here, I present a technical review of template preparation, sequencing and imaging, genome alignment and assembly approaches, and recent advances in current and near-term commercially available NGS instruments. I also outline the broad range of applications for NGS technologies, in addition to providing guidelines for platform selection to address biological questions of interest.

# Hierarchical sequencing

**What are the requirements to sequence any fragment of DNA?**

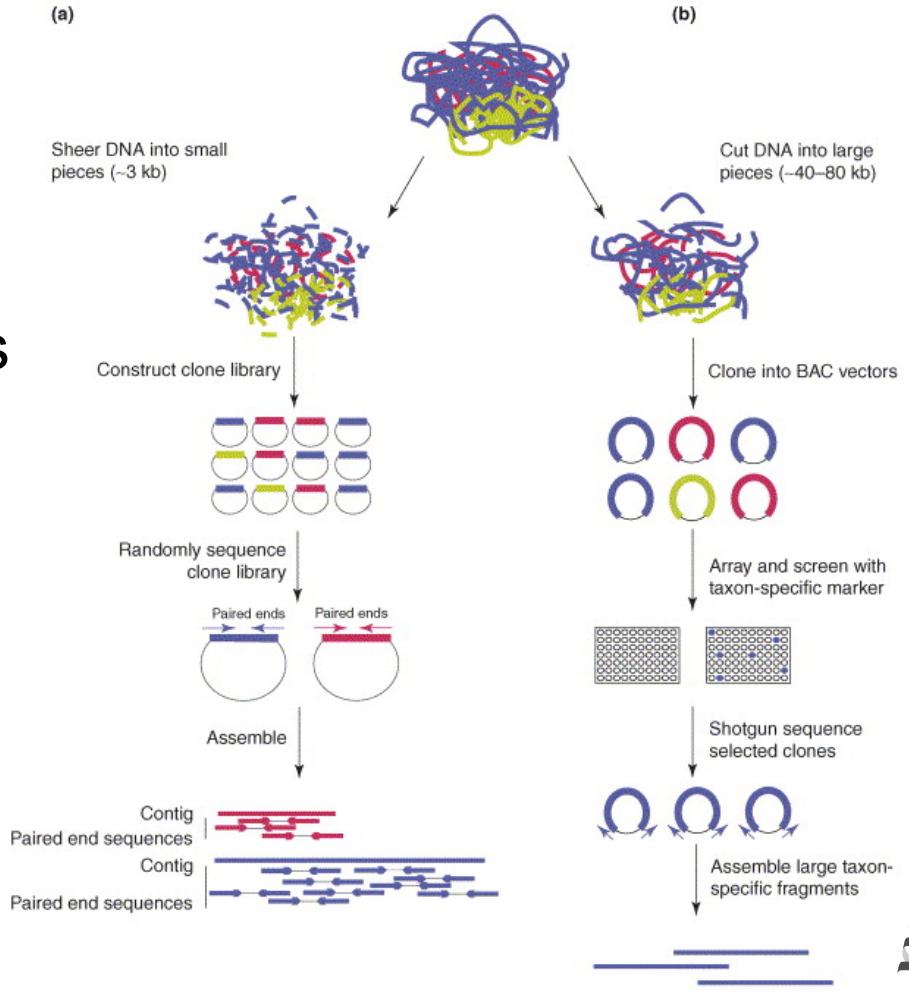
1) Enough copies of the DNA fragment

2) Primers corresponding to known locations

# Hierarchical sequencing

## 6. Sequencing ends of the DNA fragment

Using the known sequences of the vector as primers or the synthetic adapters, sequence the ends of the unknown fragment

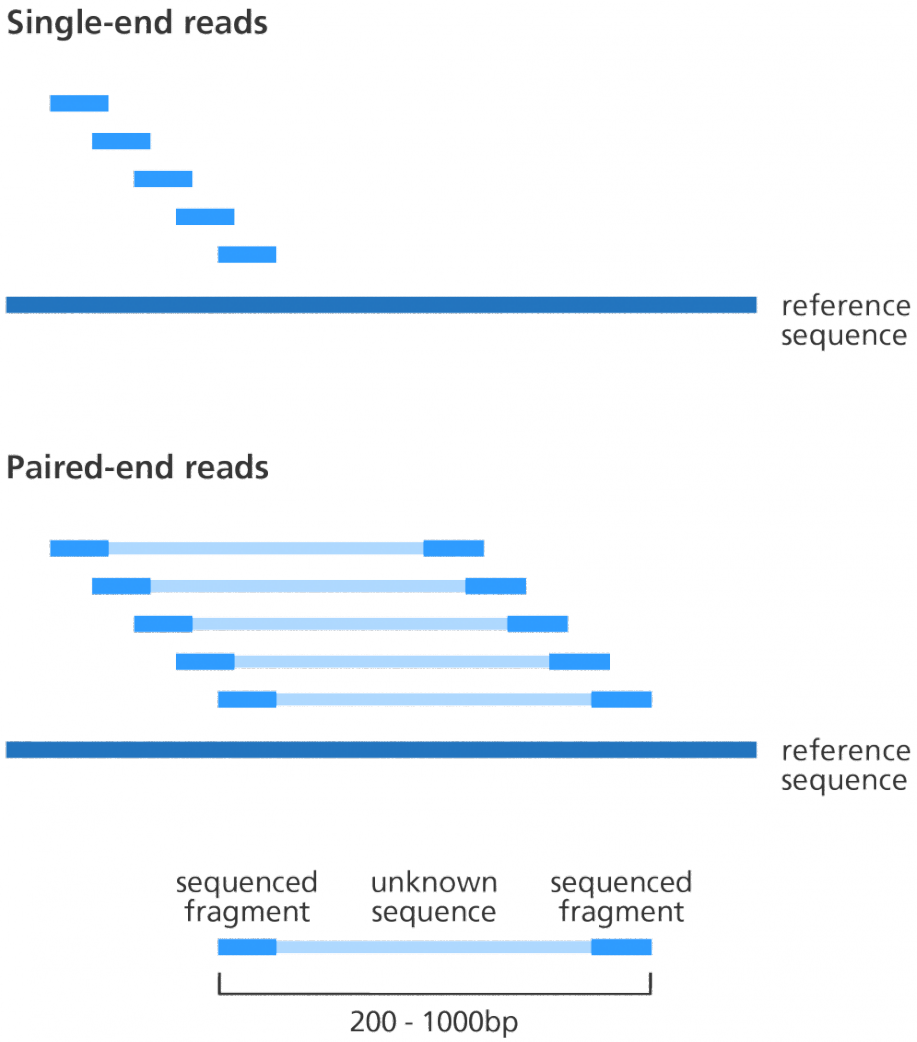


# Hierarchical sequencing

## 6. Sequencing ends of the DNA fragment

When do you use single-end sequencing?

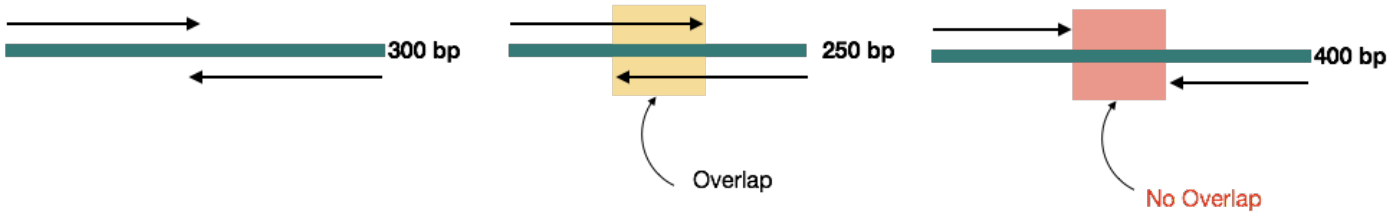
When do you use paired-end sequencing?

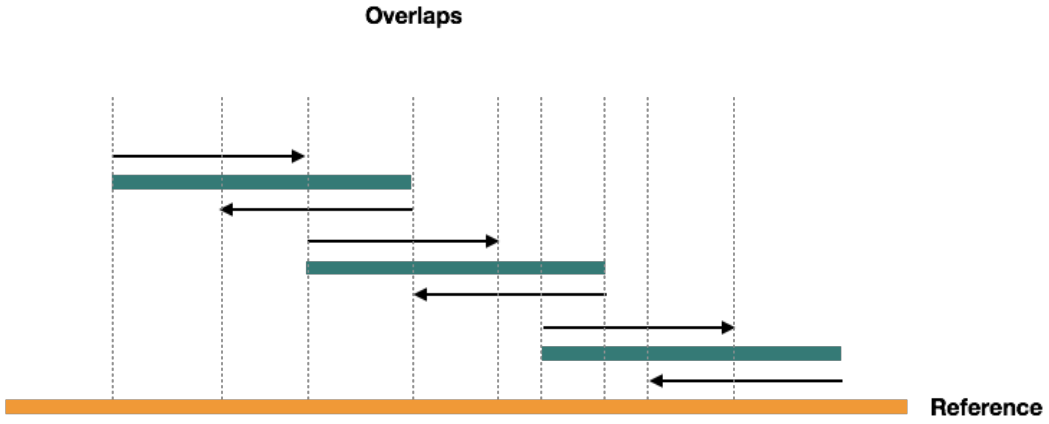


# Hierarchical sequencing

## 6. Sequencing ends of the DNA fragment

Do the sequences of the ends always overlap?

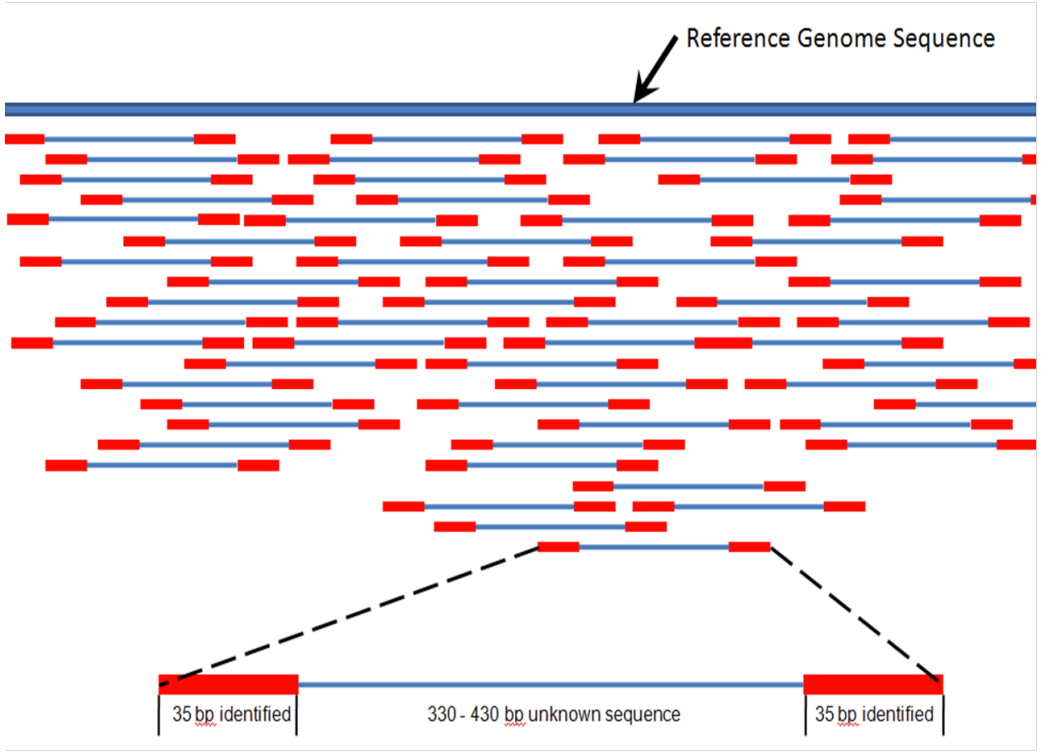


# Hierarchical sequencing

## 6. Sequencing ends of the DNA fragment

The end-sequences of the random fragments overlap to give a complete sequence.



# Whole-genome shotgun sequencing

1. Genome

2. Genome **shotgun** fragmentation

~~3. Cloning genome fragments into multiple BAC clones~~

~~4. Single BAC clone fragmentation~~

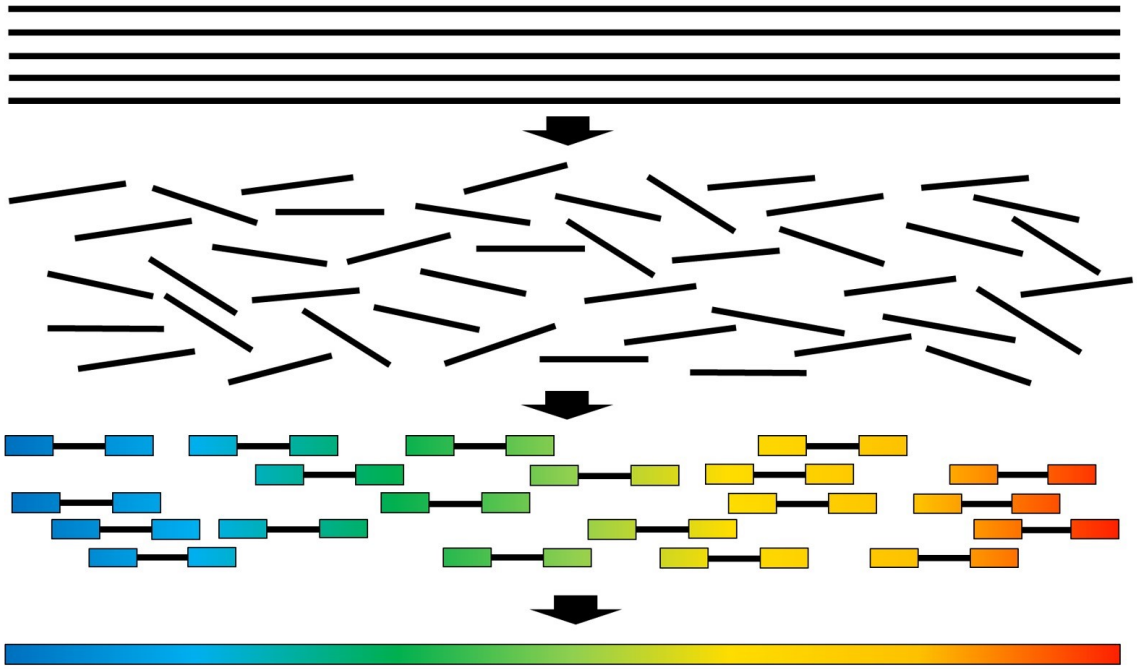
~~5. Cloning each BAC clone fragment into a sequencing vector~~

6. Sequencing ends of the DNA fragment

# Whole-genome shotgun sequencing

1. Genome

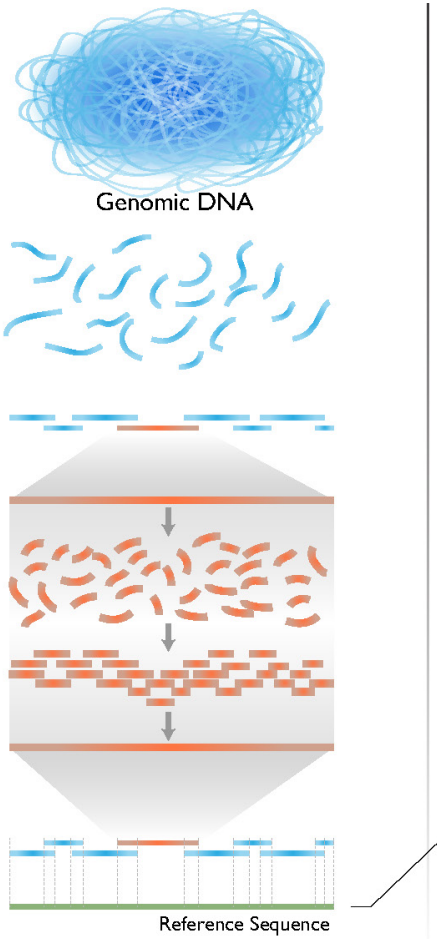
2. Genome **shotgun** fragmentation



3. Sequencing ends of the DNA fragment

# Comparison

| | Clone-by-clone | WGS | |
|---|---|---|---|
| High accuracy | | | Low accuracy |
| Easy to assemble | | | Hard to assemble |
| ↓Computational Demand | | | ↑Computational Demand |
| Very slow | | | Fast |
| Laborious | | | high throughput |
| High cost | | | Low cost |
| Requires mapping | | | Requires mapping |




# Disclaimer

Figures, photos, and graphs in my lectures are collected using google searches. I do not claim to have personally produced the material (except for some). I do cite only articles or books used. I thank all owners of the visual aid that I use and apologize for not citing each individual item. If anybody finds the inclusion of their material in my lectures a violation of their copy rights, please contact me via email.

hhalhaddad@gmail.com