Lecture 6: Sequencing a Cloned Gene – Principles of Sequencing

Sanger Sequencing Method

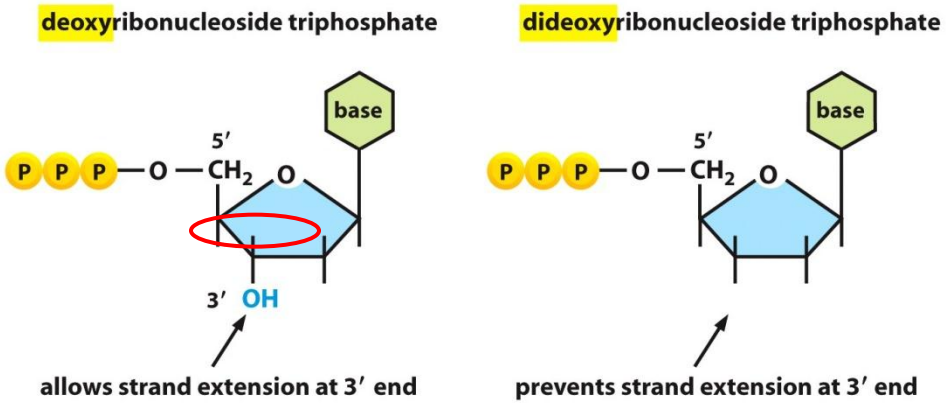
A DNA template is used to synthesize new DNA molecules terminated by a specific dideoxyribonucleotide.

- **Uses polymerase and nucleotides**
  - **a mixture of regular deoxyribonucleotides (dATP, dGTP, dCTP and dTTP) plus one dideoxyribonucleotide**

- **Addition of regular nucleotides allows continued DNA formation.**

- **Addition of a dideoxyribonucleotide stops DNA formation.**

- **A dideoxyribonucleotide lacks the chemical group required to add new nucleotides to the DNA chain.**



Sanger Sequencing Method

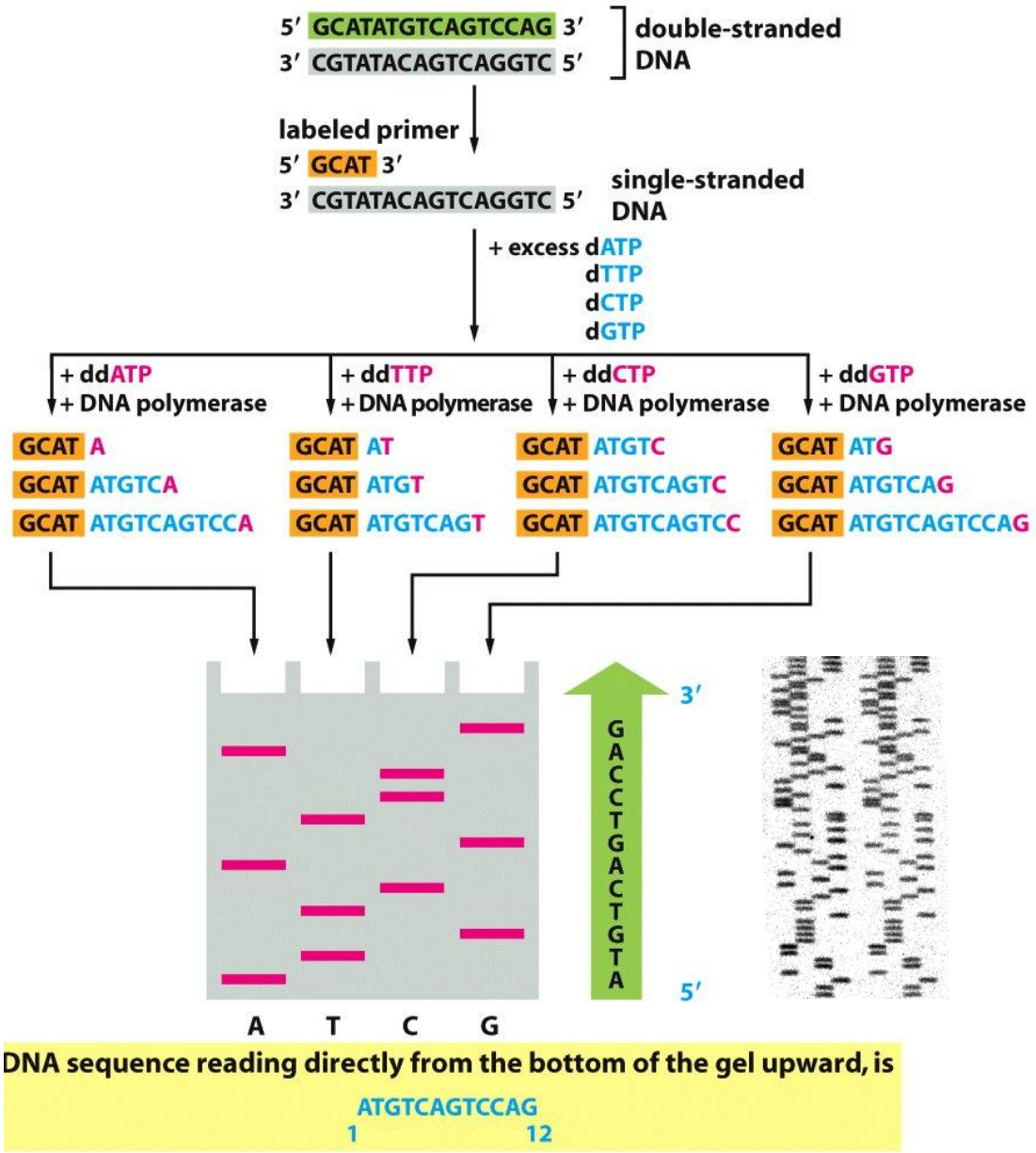
- If dideoxyadenosine is used as the chain terminator,
  - Multiple DNA molecules are made, all ending in adenosine
  - If the template does not have adenosine at a specific length, there will be no molecules of that length in the test tube.

- This process is repeated separately with G, C and T dideoxynucleotides.

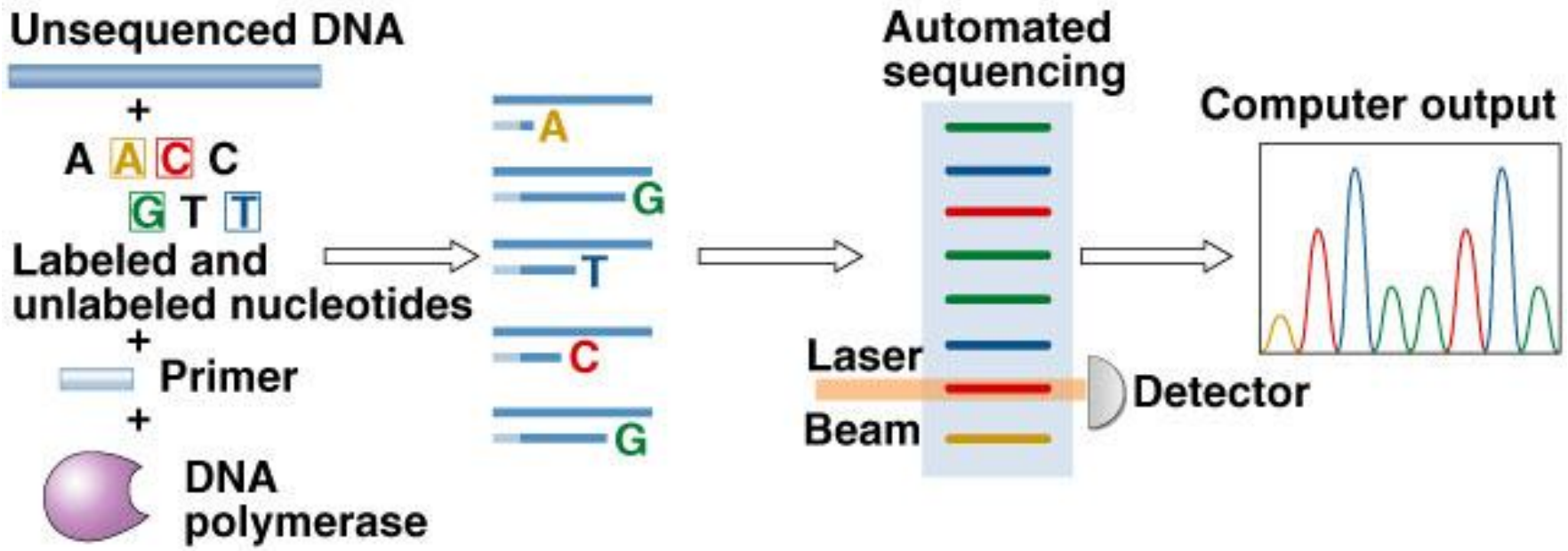
- Computer compares fragments and sequences the DNA.

Production of DNA Copies Terminated at Different Points in the Sequence

- The products of the four Rx's are separated by electrophoresis in four parallel lanes of a polyacrylamide gel (labeled A, T, C and G.
- The newly synthesized fragments are detected by a radioactive or fluorescent label incorporated either into the primer or into one of the deoxyribonucleoside triphosphates used to extend the DNA chain.
- The bands in each lane represent fragments that have terminated at a given nucleotide.
- By reading off the bands in order, starting at the bottom of the gel and working across all lanes, the DNA sequence of the newly synthesized strand can be determined.



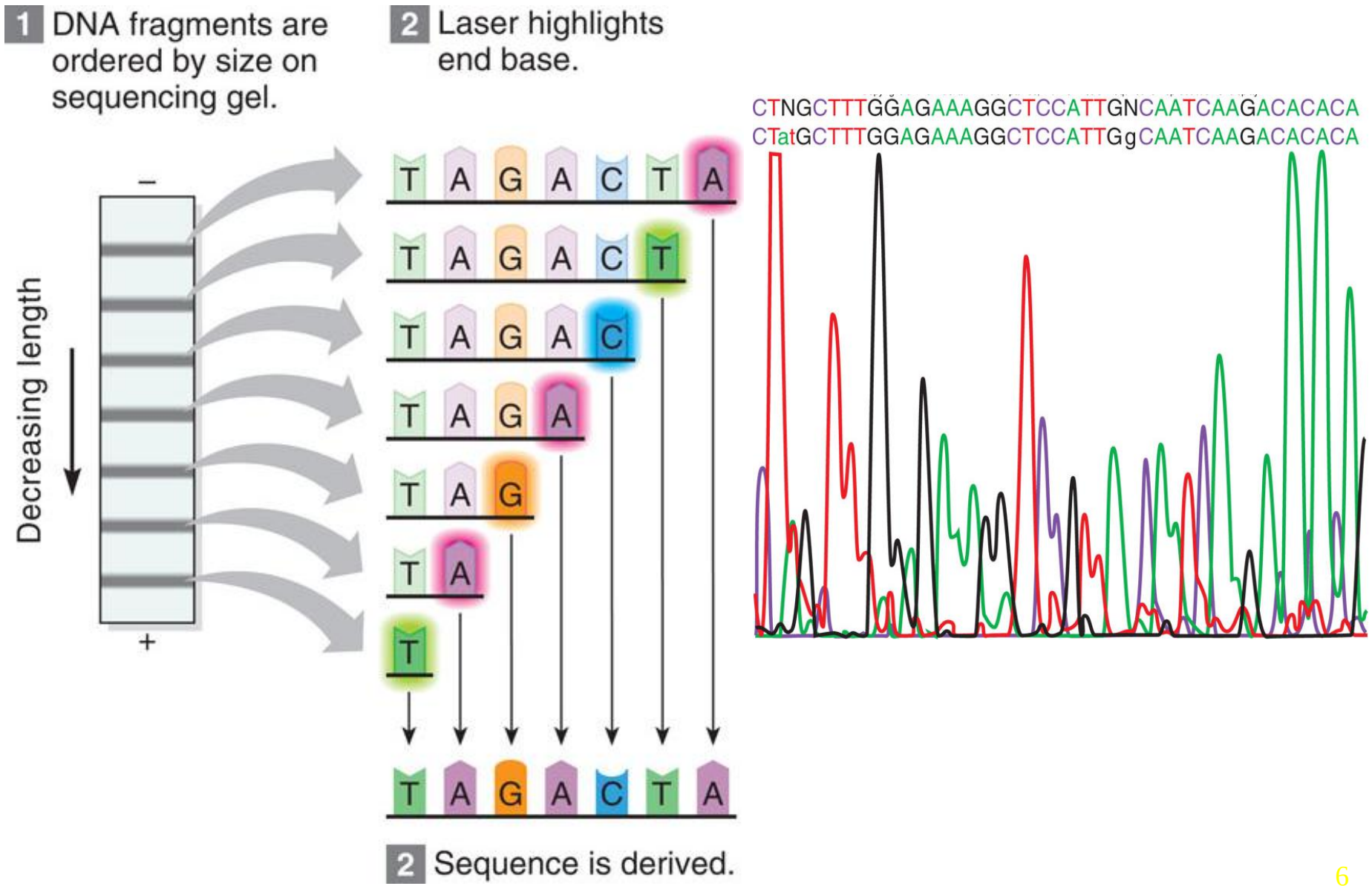
Sequencing DNA



primer – a short nucleic acid sequence that provides a starting point for DNA synthesis

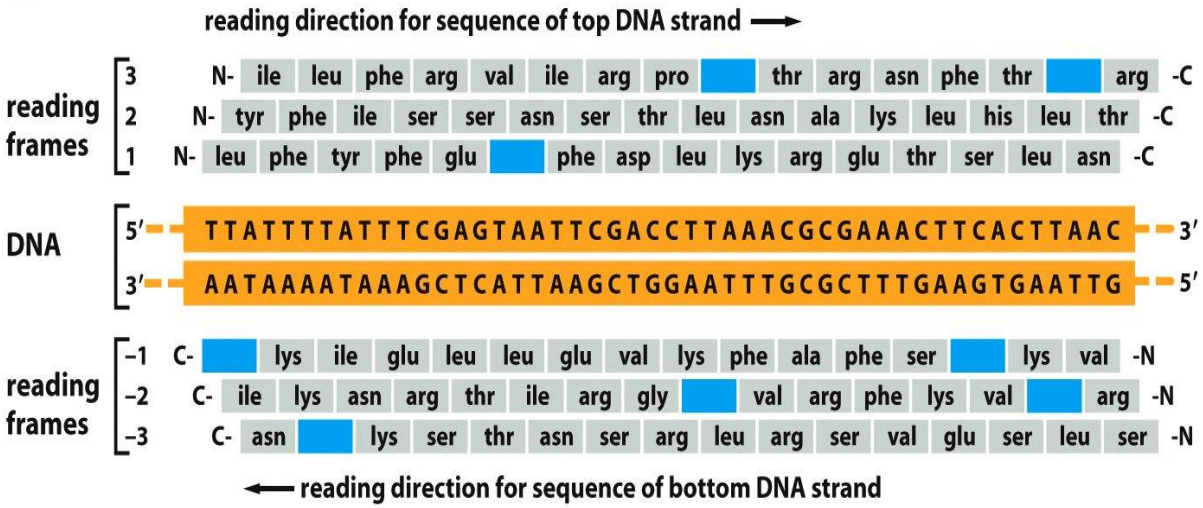
- **In living organisms, primers are short strands of RNA.**
- **A primer must be synthesized by an enzyme called primase, which is a type of RNA polymerase, before DNA replication can occur.**
- **The synthesis of a primer is necessary because the enzymes that synthesize DNA, which are called DNA polymerases, can only attach new DNA nucleotides to an existing strand of nucleotides.**
- **The primer, therefore, serves to prime and lay a foundation for DNA synthesis.**

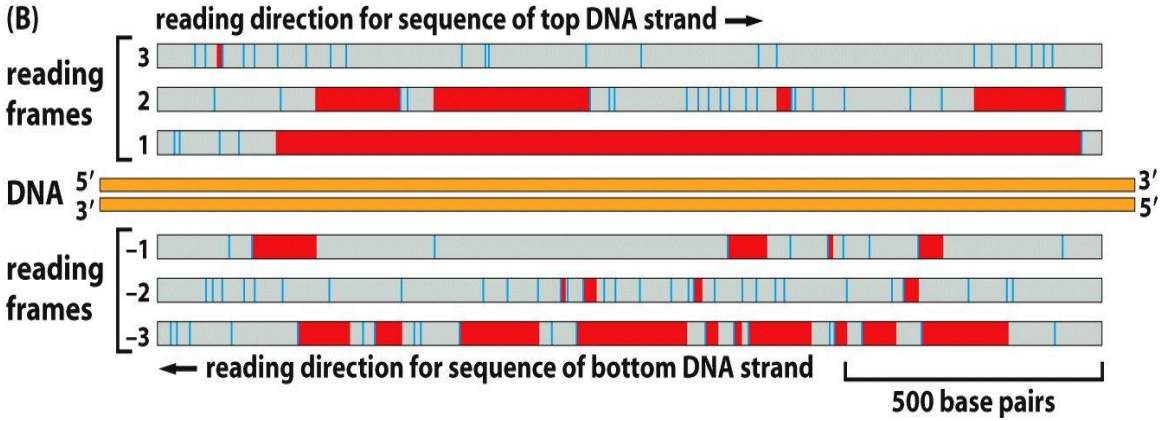
Reading a DNA Sequence



Finding the Regions in a DNA Sequence That Encode a Protein

- Any region of the DNA sequence can, in principle, code for six different amino acid sequences, because any one of the three different reading frames can be used to interpret the nucleotide sequence on each strand.
- An open reading frame contains no stop codons.
- A long open reading frame most likely is part of a gene.
- Search of a DNA sequence for possible protein-encoding sequence can be accomplished by finding start and stop signals (blue bars and lines) .
- Only reading frame 2 lacks a stop signal.
- All of the regions between possible start and stop signals for protein synthesis are displayed as red bars.
- Only reading frame 1 actually encodes a protein, which is 475 amino acid residues long.





- Stop signal for protein synthesis – encountered, on average, about once every 20 amino acids
- Regions between possible start and stop signals for protein synthesis

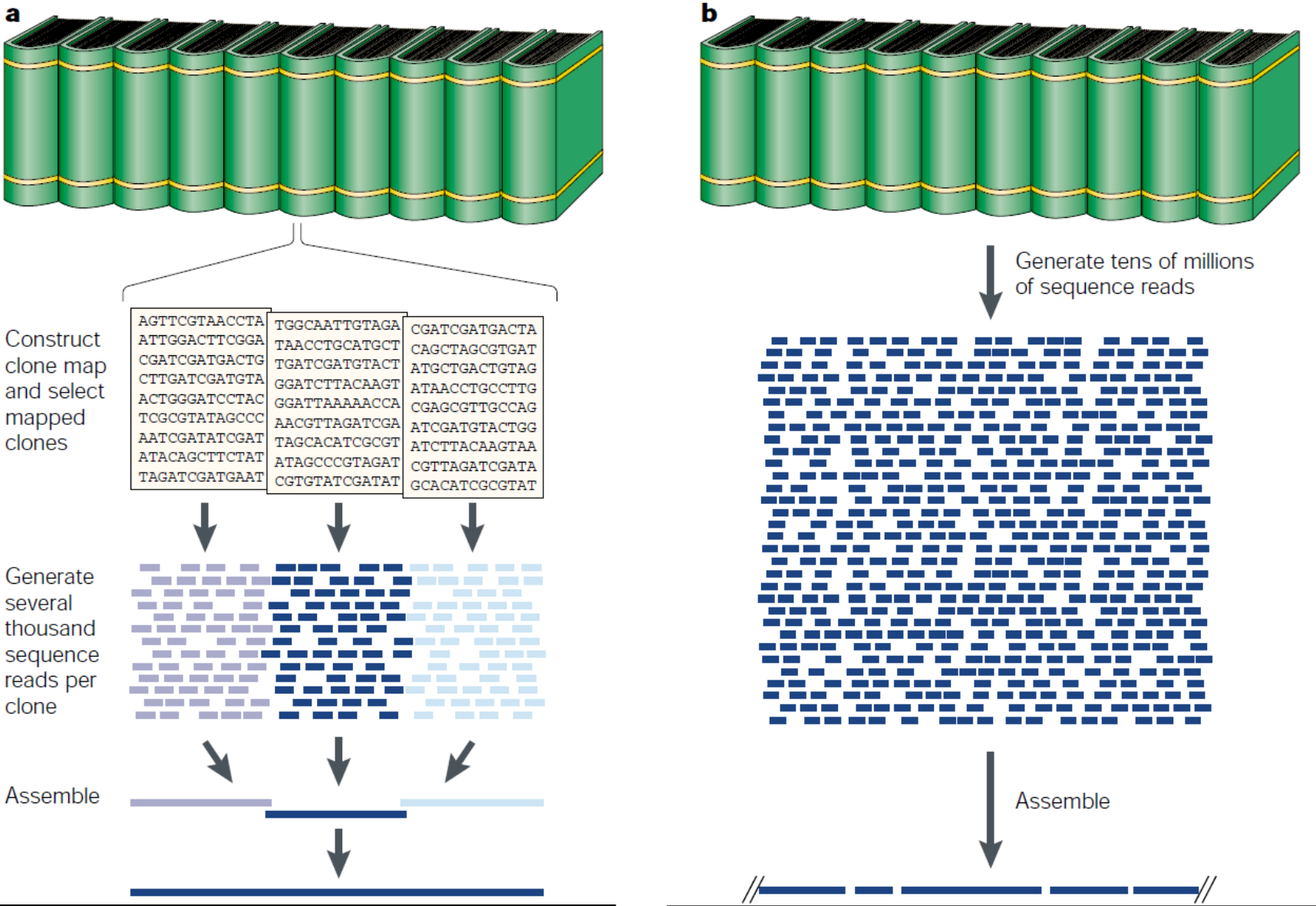
Genome Sequencing

- **Sequencing by whole genomes**
  - ✓ **clone-by-clone sequencing – cloning larger inserts in YAC requires construction of a physical map, then marking the site of YAC clones for later sequencing**

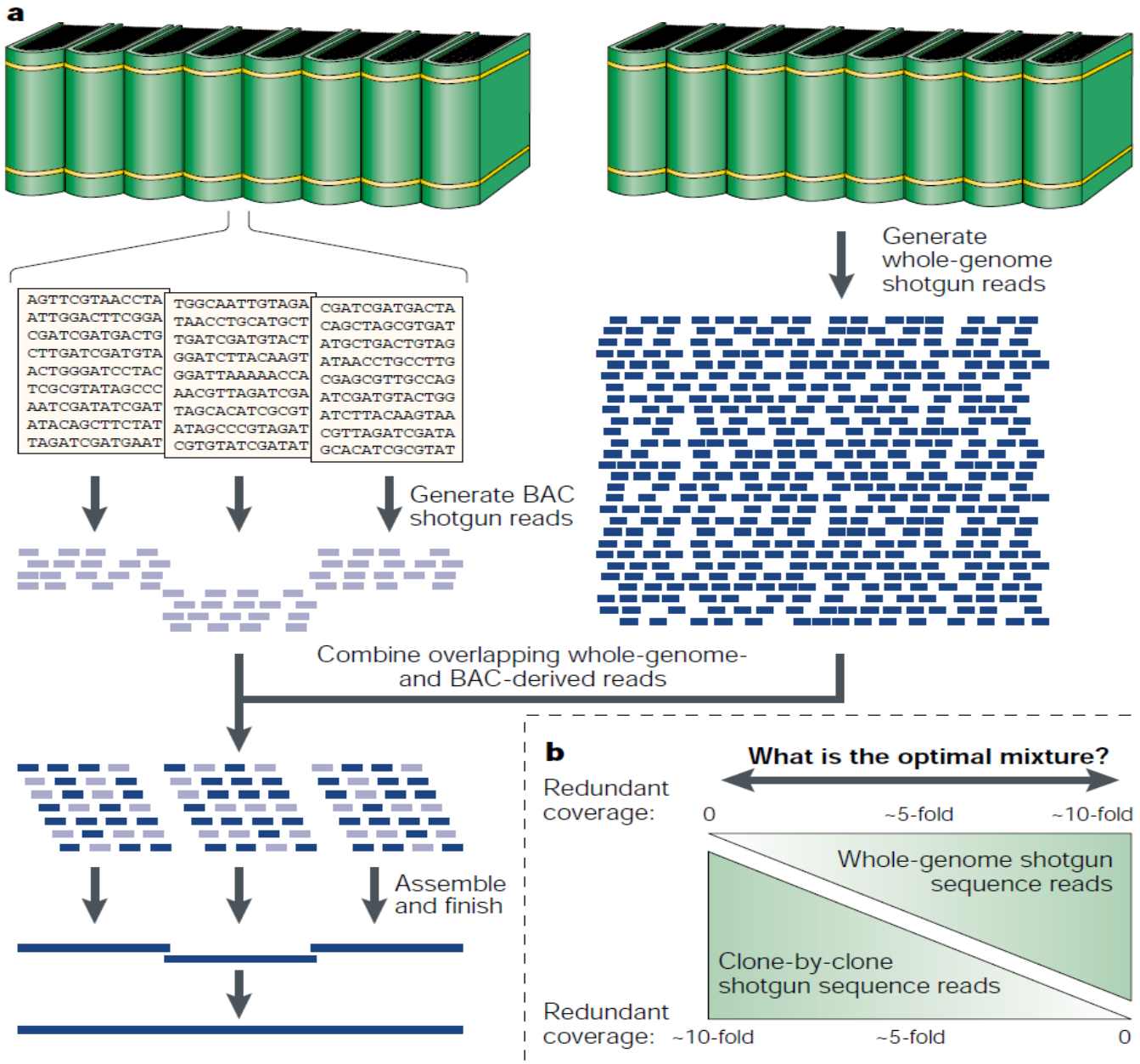
Genome Sequencing

- **Shotgun sequencing – sequence all cloned fragments and use a computer to put together overlaps**
  - ✓ **requires abundant computing power**
  - ✓ **does not tie the sequence to any other information about the genome**
  - ✓ **assembler programs assemble a consensus sequence**

Method Comparison

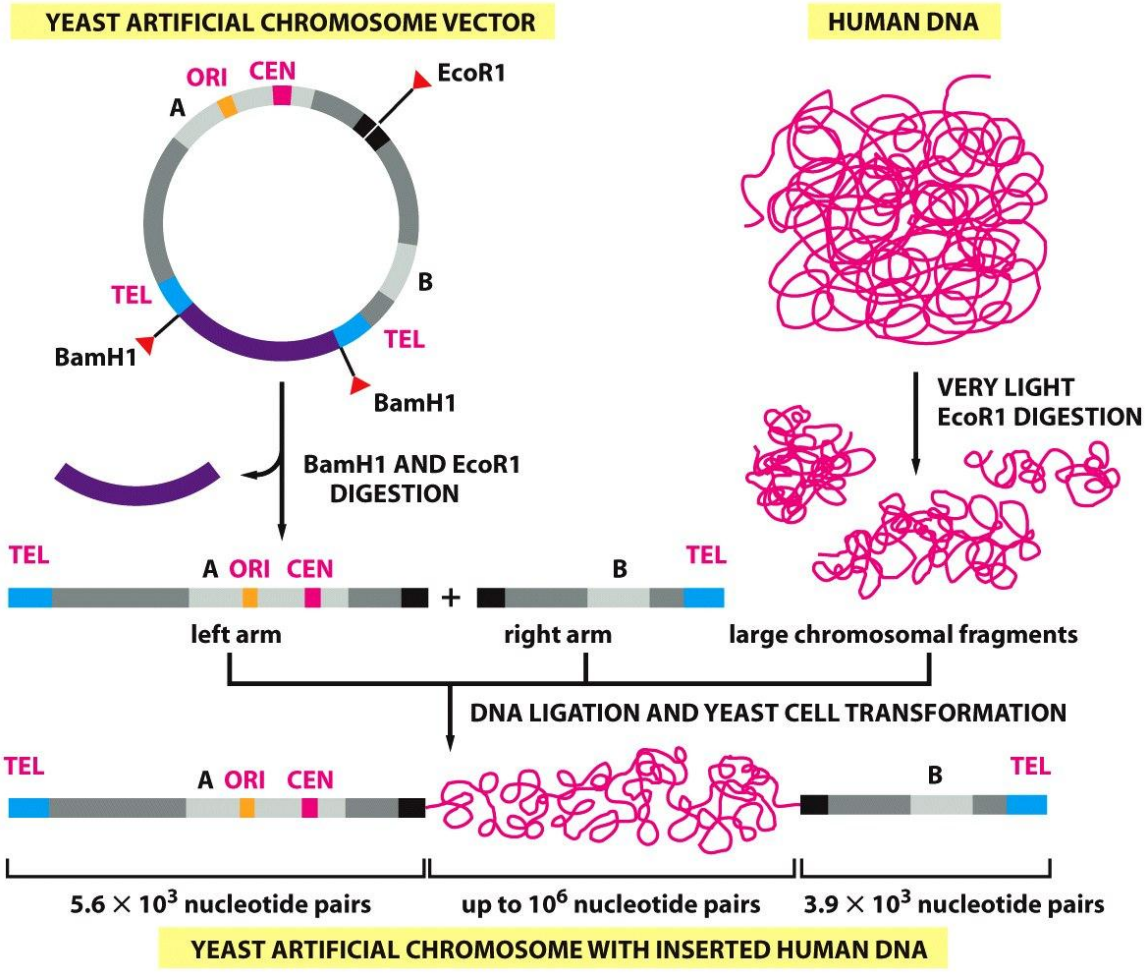


Hybrid Shotgun Sequencing Approach



Yeast Artificial Chromosome

- A YAC vector allows the cloning of very large DNA molecules
- TEL, CEN and ORI are the telomere, centromere and origin of replication sequences, respectively, for *S. cerevisiae*, which are required to propagate the YAC.
- The sequences denoted A and B encode enzymes that serve as selectable markers to allow easy isolation of yeast cells that have taken up the artificial chromosome.
- Because bacteria divide more rapidly than yeasts, most large-scale cloning projects now use *E. coli* as the means for amplifying DNA.



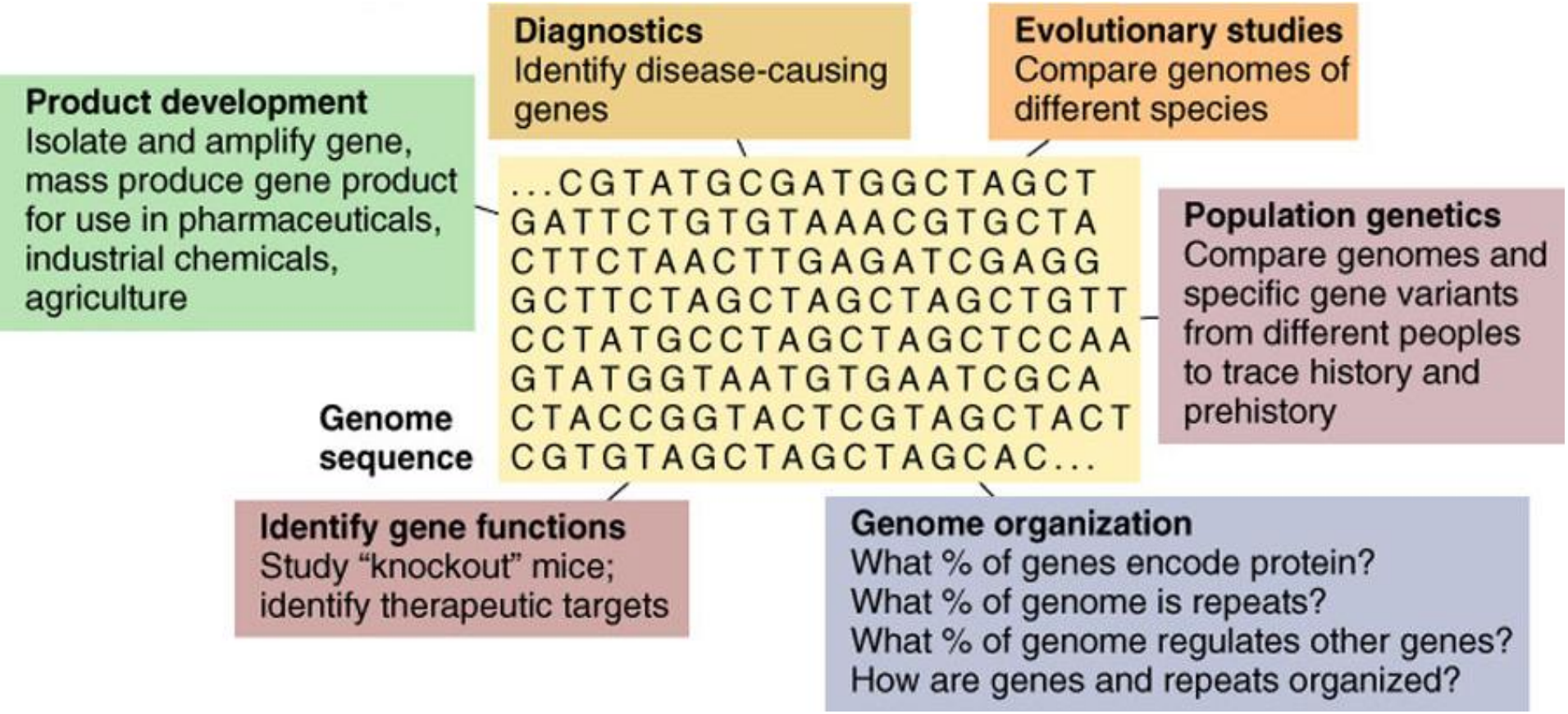
The Human Genome Project

- **Decade-long project to sequence the human genome or determine the order of the nucleotides present in each of the chromosomes end-to-end.**

- **The draft of the human genome was announced in February, 2001. Project was completed in 2003.**

- **This project represented the work of thousands of researchers in an international collaboration.**

Impact of the Human Genome Project: Uses of Genetic Information



Bioinformatics

- **Rapid automated DNA sequencing was instrumental in the success of the Human Genome Project, an international effort begun in 1990 to sequence the human genome and that of a number of organisms.**

- **However, a genomic sequence is like a book using an alphabet of only four letters, without spaces or punctuation.**

- **Identifying genes and their functions is a major challenge.**

- **The annotation of genomic sequences at this level is one aspect of bioinformatics, defined broadly as the use of computers in the interpretation and management of biological data**