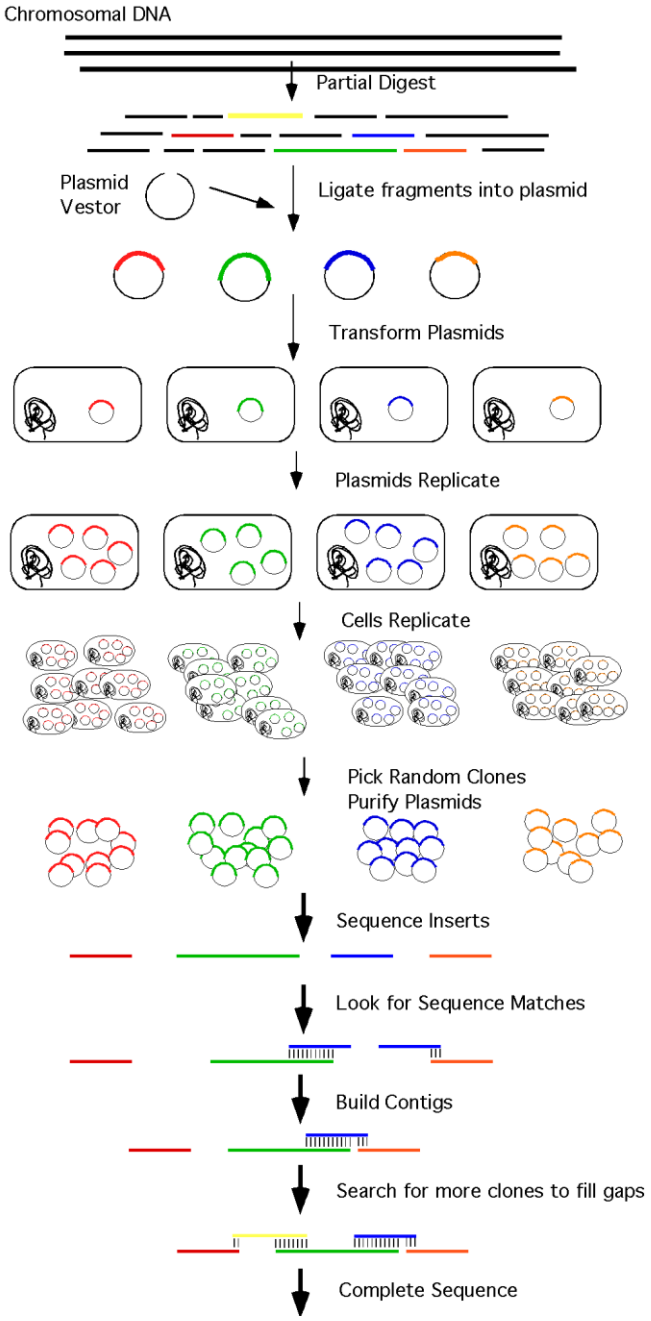
Lecture 3: Genomic and cDNA Libraries

DNA Library

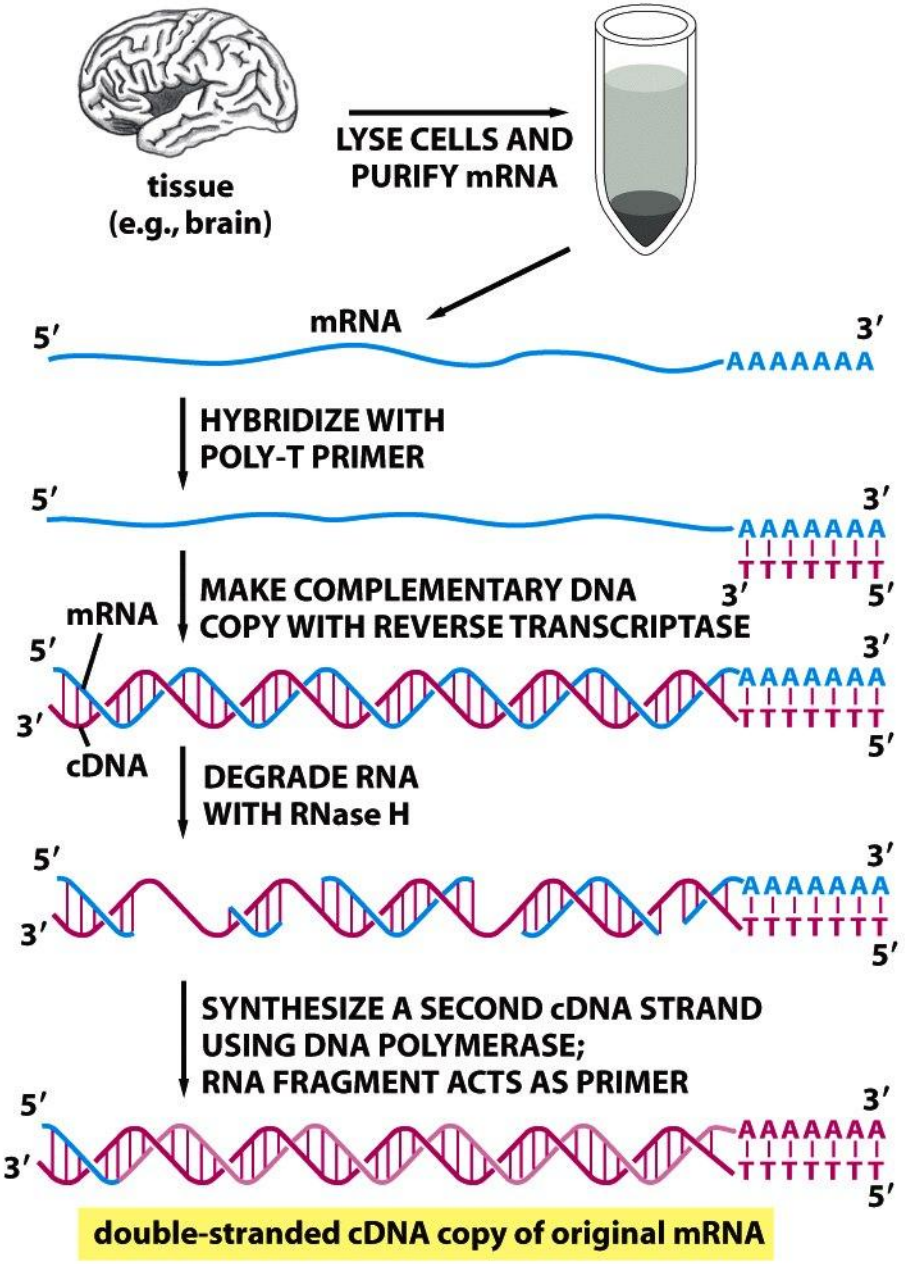
- A DNA library is a collection of DNA fragments that have been cloned into vectors so that DNA fragments of interest can be identifed and isolated for further study.
- There basically are two kinds of libraries: genomic DNA and cDNA libraries.
- Genomic DNA libraries contain large fragments of DNA in either bacteriophages or bacterial or P1-derived artificial chromosomes (BACs and PACs).
- cDNA libraries are made with cloned, reverse-transcribed mRNA, and, therefore, lack DNA sequences corresponding to genomic regions that are not expressed, such as introns and 5′- and 3′-noncoding regions.
- cDNA libraries generally contain much smaller fragments than genomic DNA libraries, and are usually cloned into plasmid vectors.

Construction and Analysis of a Genomic DNA Library

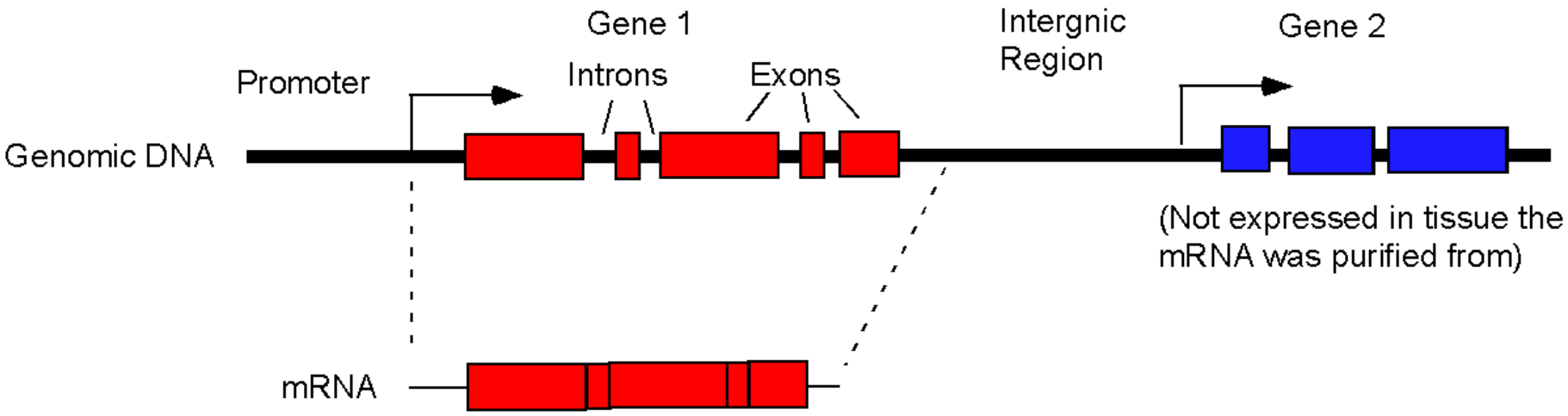


Synthesis of cDNA

- **Total mRNA is extracted from a tissue and reverse transcriptase produces DNA copies (cDNA) of the mRNA molecules.**
- **A short oligonucleotide complementary to the poly-A tail at the 3′ end of the mRNA is first hybridized to the RNA into a cDNA chain, thereby forming a DNA/RNA hybrid helix.**
- **Treating the DNA/RNA hybrid with RNase H creates nicks and gaps in the RNA strand.**
- **RNase H is a non-specific endonuclease that cleaves the 3'-O-P-bond of RNA in a DNA/RNA duplex to produce 3'-hydroxyl and 5'-phosphate terminated products.**
- **DNA polymerase then copies the remaining single-stranded cDNA into double-stranded cDNA.**
- **The fragment of the original mRNA is the primer for this synthesis reaction.**
- **Because the DNA polymerase used to synthesize the second DNA strand can synthesize through the bound RNA molecules, the RNA fragment that is base-paired to the 3′ end of the first DNA strand usually acts as the primer for the final product of the second strand synthesis.**



Differences Between a Genomic and cDNA Library

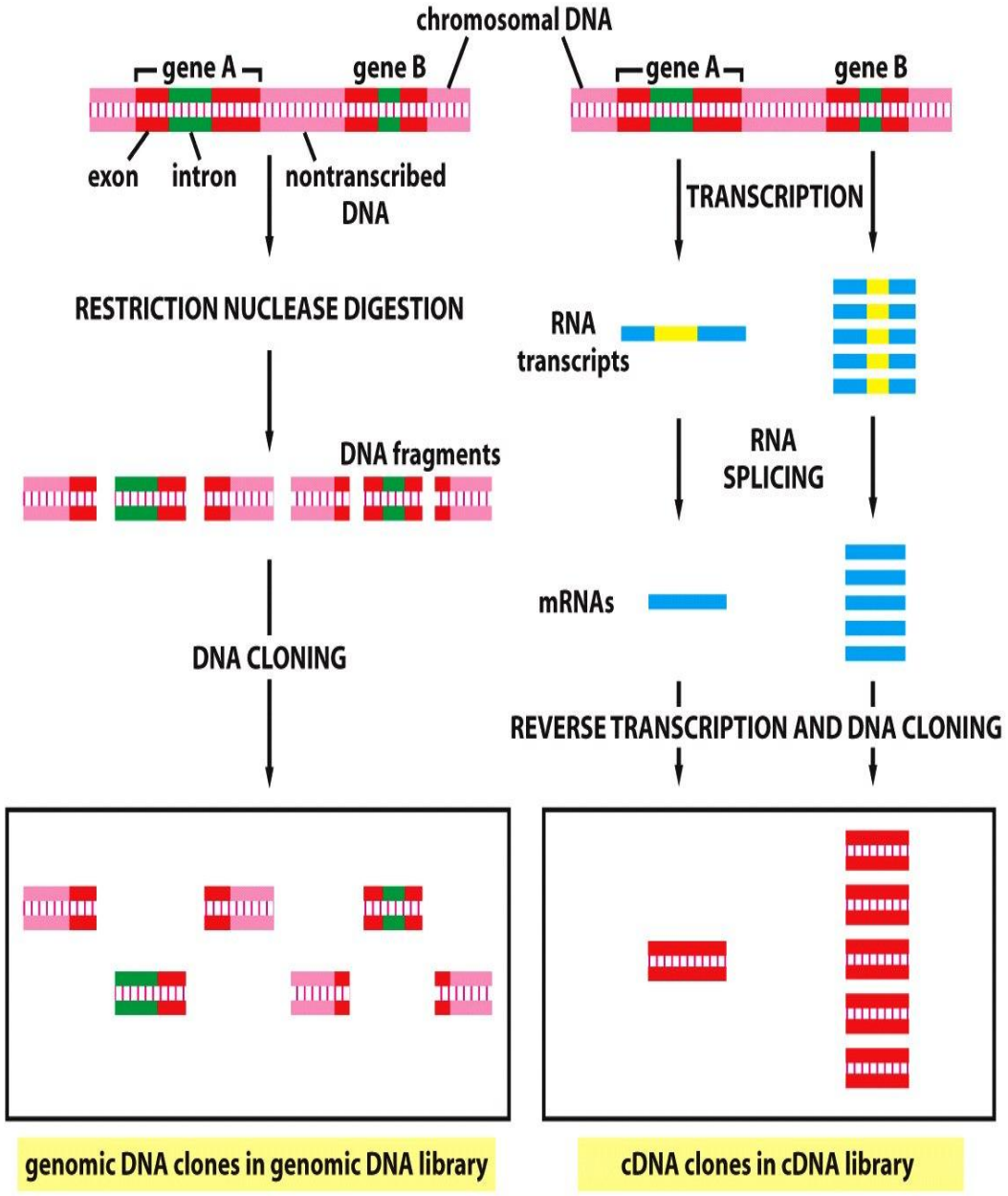


<u>Genomic Library</u>
Promoters
Introns
Intergenic
Non-expressed genes

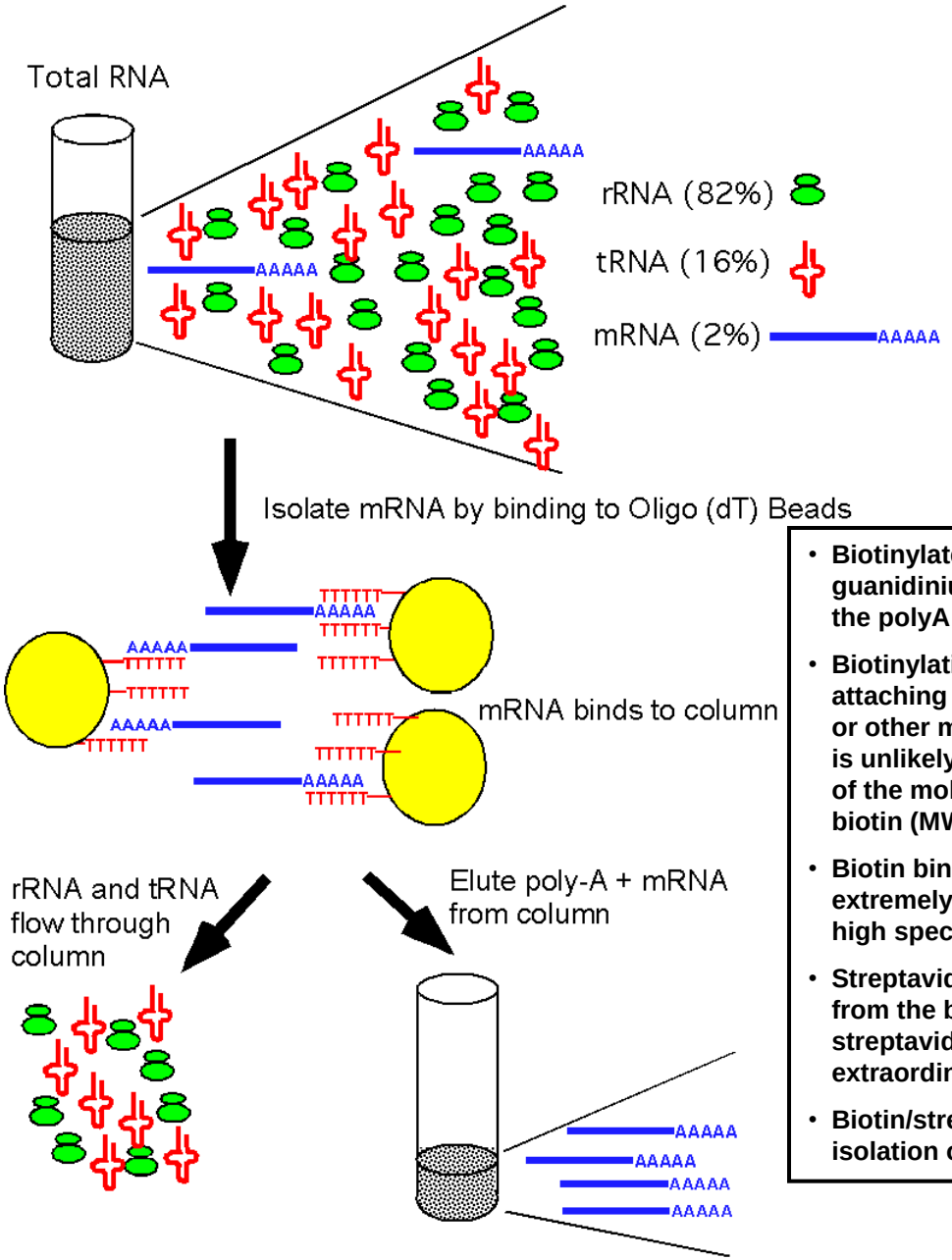
<u>cDNA Library</u>
Expressed genes
Transcription start sites
Open reading frames (ORFs)
Splice points

Genomic DNA and cDNA Clones

- Gene A is infrequently transcribed, whereas gene B is frequently transcribed, and both genes contain introns (green).
- In a genomic DNA library, both the introns and the non-transcribed DNA (pink) are included in the clones, and most clones contain, at most, only part of the coding sequence of a gene (red).
- In the cDNA clones, the intron sequences (yellow) have been removed by RNA splicing during the formation of the mRNA (blue), and a continuous coding sequence is, therefore, present in each clone.
- Because gene B is transcribed more frequently than gene A in the cells from which the cDNA library was made, it is represented much more frequently than A in the cDNA library.
- In contrast, A and B are in principle represented equally in the genomic DNA library.



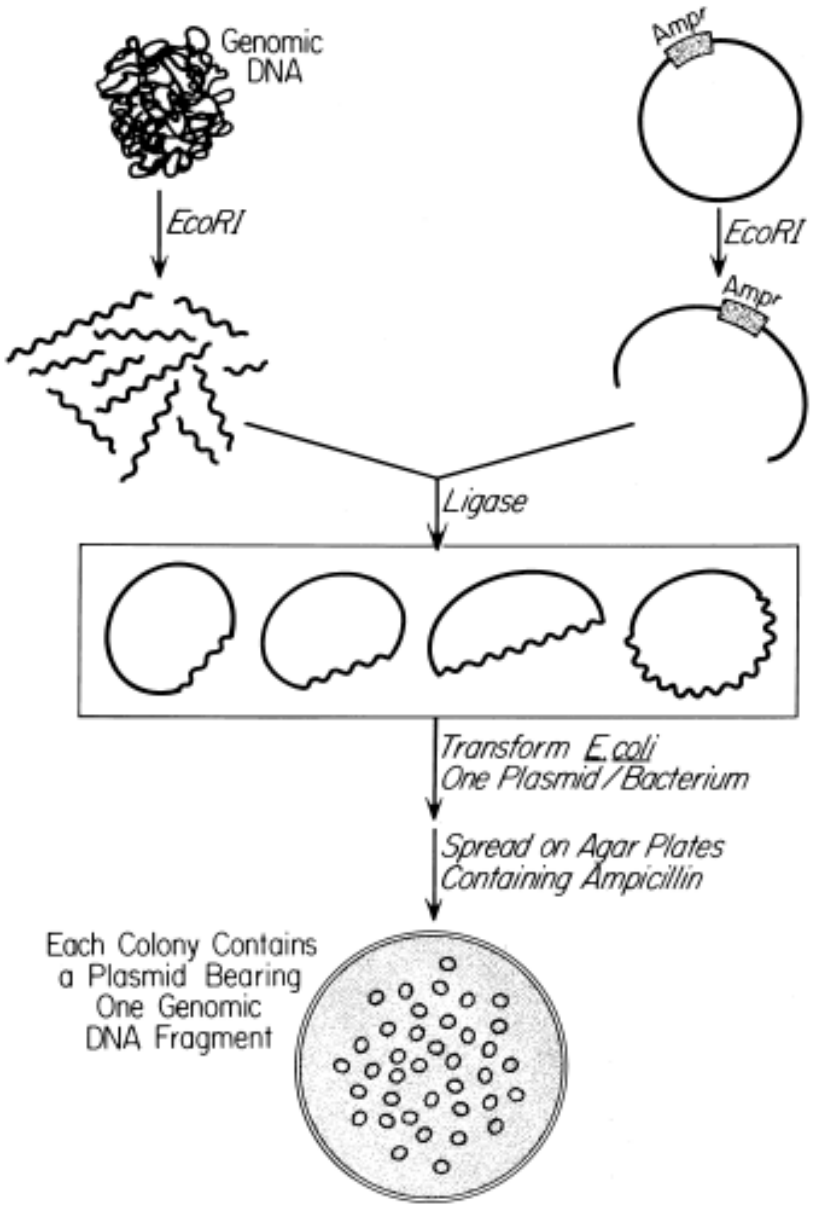
Purification of mRNA



- **Biotinylated oligo dT is added to guanidinium-treated cells, and anneals to the polyA tail of mRNAs.**
- **Biotinylation is the process of covalently attaching biotin to a protein, nucleic acid or other molecule; it is rapid, specific and is unlikely to perturb the natural function of the molecule due to the small size of biotin (MW = 244.31 g/mol).**
- **Biotin binds to streptavidin and avidin with extremely high affinity, fast on-rate and high specificity,**
- **Streptavidin is a 52.8 kDa-protein purified from the bacterium *Streptomyces avidinii*; streptavidin homo-tetramers have an extraordinarily high affinity for biotin.**
- **Biotin/streptavidin interactions permit isolation of the mRNA/oligo dT complexes.**

Amplification of Genomic Library

- **A few nanograms of foreign DNA is digested with EcoRI.**
- **The DNA must contain the same restriction endonuclease recognition sites as the vector.**
- **Plasmid vector is also digested with EcoRI to create a linear DNA molecule.**
- **The "sticky" single-stranded ends of the foreign DNA align and base-pair with the complementary "sticky ends" of the plasmid, after which DNA ligase covalently bonds foreign DNA to plasmid DNA.**
- **This recombinant DNA is introduced into *E. coli* by transformation.**
- **The plasmid contains a bacterial origin of replication so that as the bacterial culture grows, plasmids replicate resulting in several copies in each bacterium.**
- **When the culture has grown to sufficient size, plasmid DNA is isolated and foreign DNA is cut from the plasmid using EcoRI.**
- **The resulting yield will often be milligrams of DNA, i.e., >100-fold amplification.**

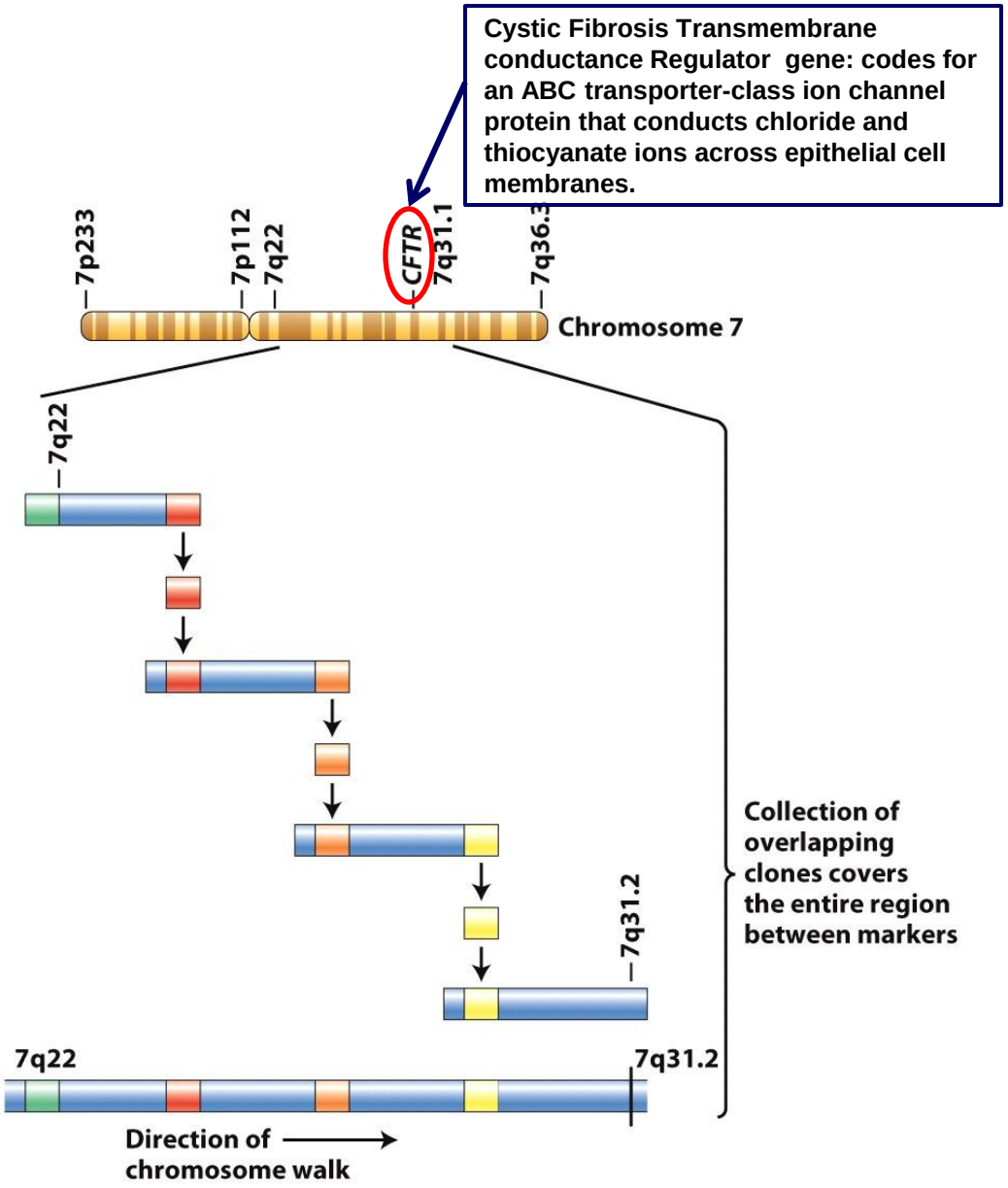


Overlapping and Non-overlapping Fragments

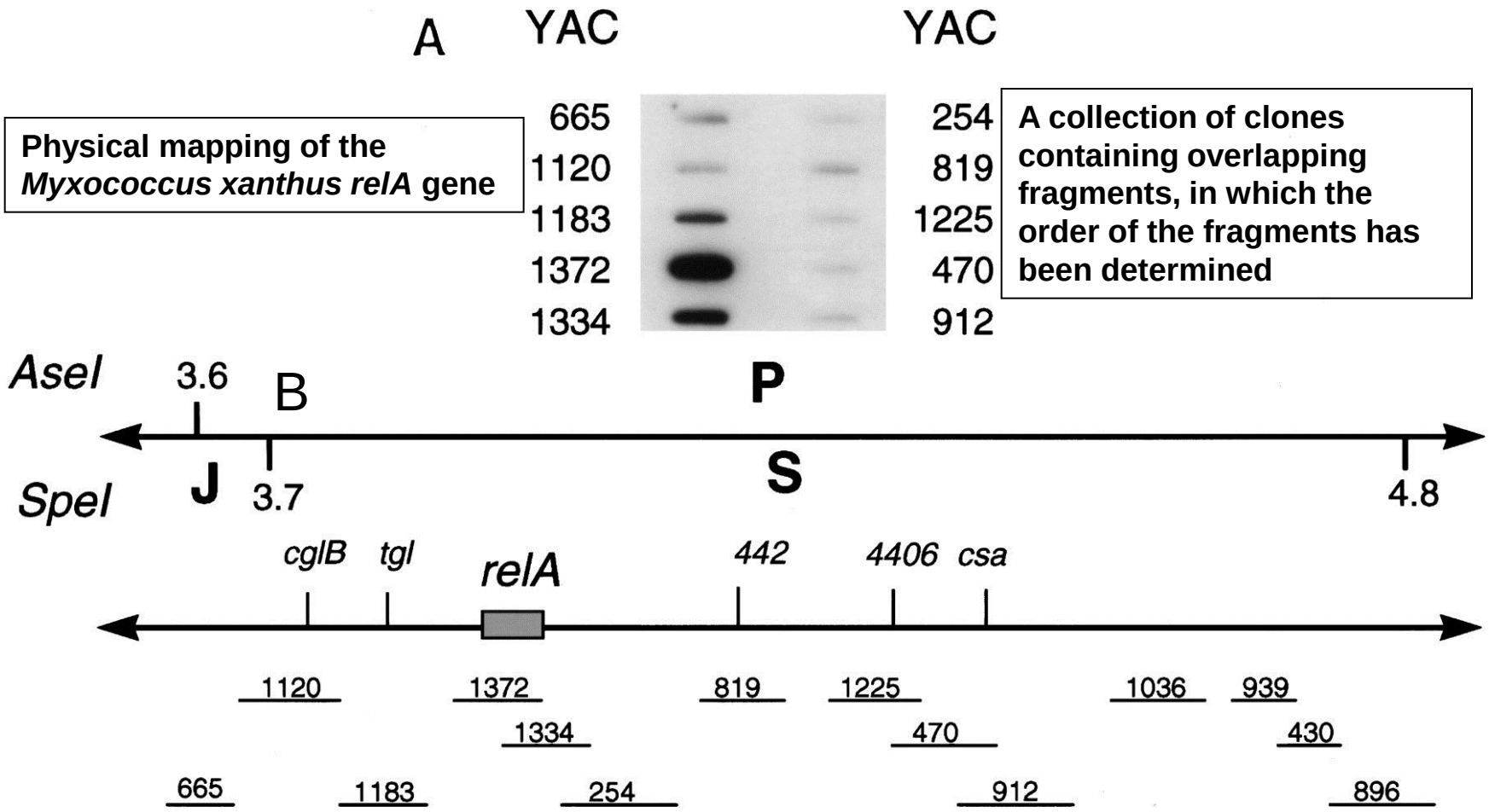
- **Incomplete digestion with an endonuclease will result in a library containing overlapping fragments.**
- **Complete digestion will result in a library containing no overlapping fragments.**
- **Using incomplete digestion, sequence information obtained from one clone will allow the isolation of clones containing neighboring (overlapping) sequence information.**
- **This process can allow large contiguous stretches of sequence information to be obtained by "chromosome walking".**

Using a Chromosomal Walk to Order a Set of Clones

- **Chromosomal walking** is a method of **positional cloning** used to find, isolate and clone a **particular** allele in a gene library.
- It involves **mapping** of the **position** of a DNA site or a gene by using **overlapping** restriction fragments.
- This chromosomal walk begins with a **recombinant** phage or BAC clone obtained from a library that contains large inserts representing an **entire** eukaryotic genome.
- The **molecular marker** 7q22 was used to **probe** a human genomic library.
- **Only** the insert DNAs are shown.
- The insert DNA selected by the probe is then used to **isolate** another recombinant phage or BAC containing a **neighboring segment** of eukaryotic DNA.
- This walk illustrates how to start at molecular landmark 7q22 and get to marker 7q31.1, which is on the other side of the CF gene.

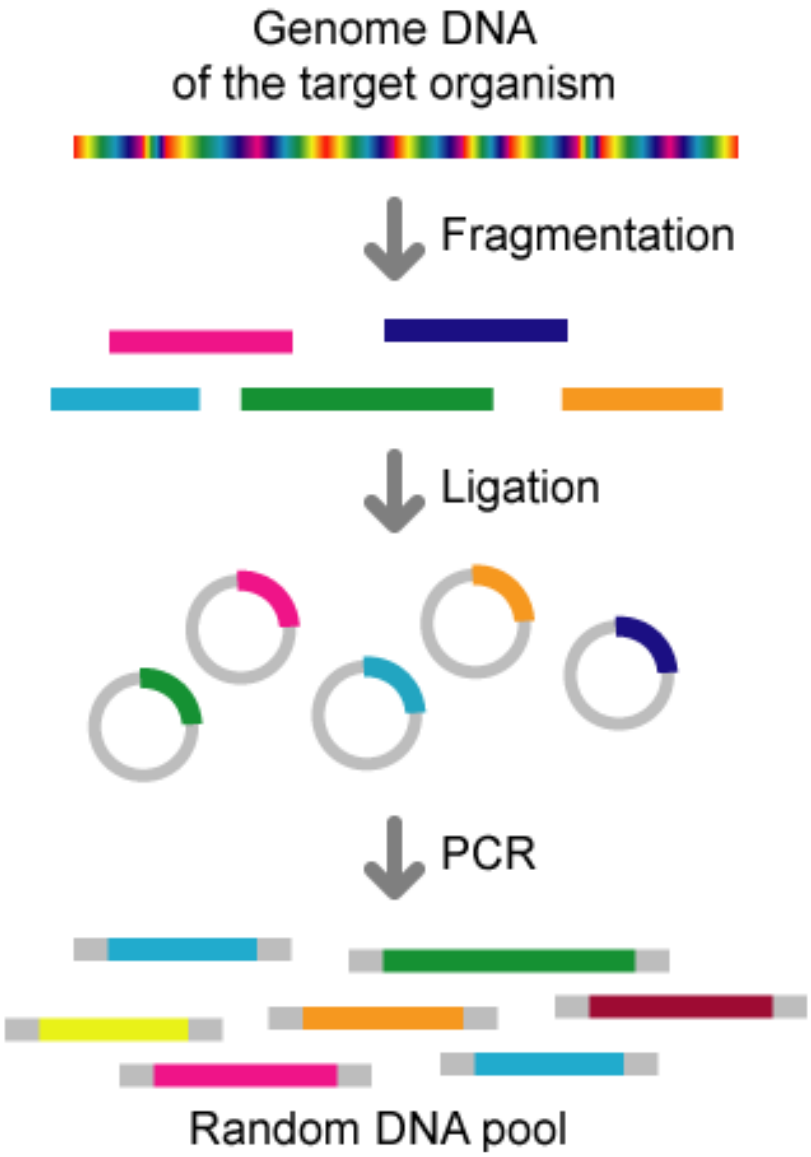


Ordered Library

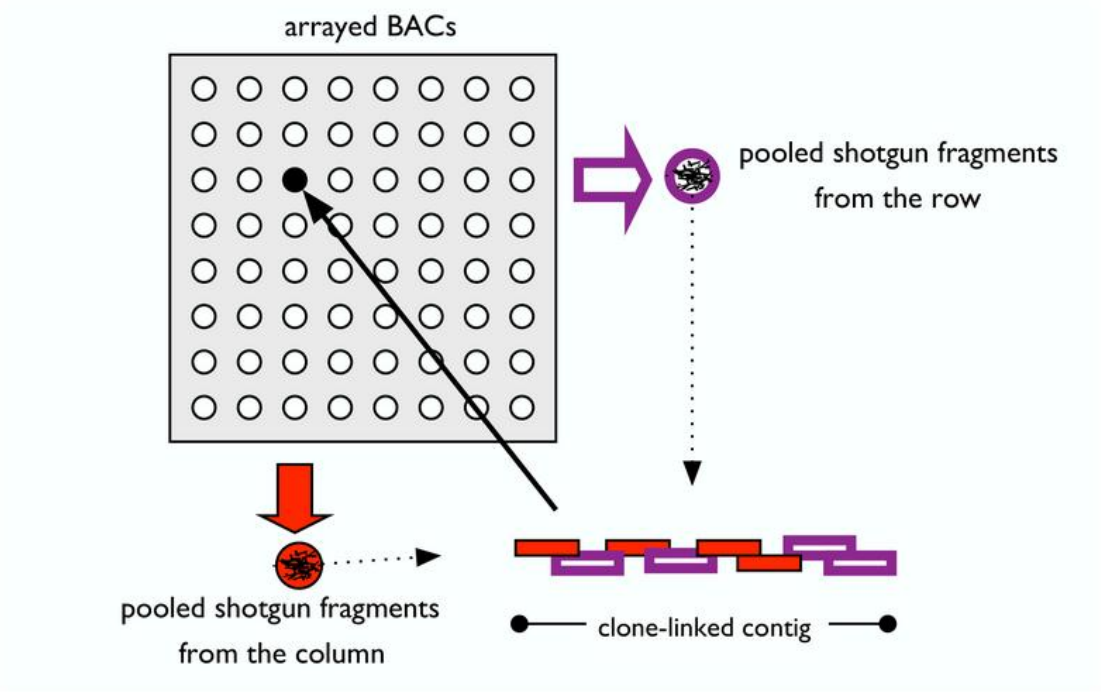


- **A, Slot blots using total YAC DNA from an ordered YAC library, probed with DNA fragment harboring the *relA* gene (*relA* gene product is responsible for the synthesis of guanosine 3',5'-bispyrophosphate (ppGpp) during the stringent response to amino acid starvation).**
- **B, Schematic of the 3.6- to 4.8-Mb region of the physical map of *M. xanthus.* YACs covering this region are indicated and labeled at the *bottom.* The position of *relA* is designated by the shaded box.**
- **The top line represents *Ase*I and *Spe*I fragments as obtained from the literature; the second line represents the position of previously mapped markers also obtained from the literature.**

Random Library



Arrayed Library



- **DNA extracted from each clone is pooled together with other clones in the same row and column.**
- **Subclone libraries are prepared from the pools, and shotgun sequences are collected from the sub-libraries.**
- **Sequences are assembled into contigs.**
- **If a contig contains sequences from a row and a column pool's sub-library, the contig is assigned to the BAC at the intersection of the row and the column.**