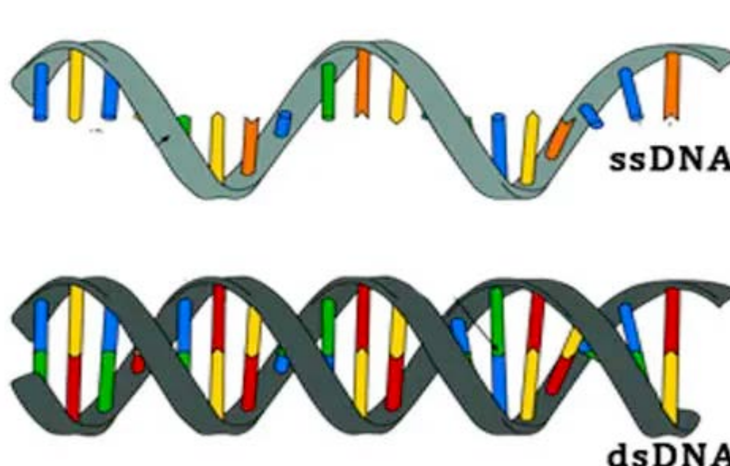
HOW LARGE ARE DNA PROGRAMS?

Species	*T2 phage*	*Escherichia coli*	*Drosophila melanogaster*	*Homo sapiens*	*Paris japonica*
Genome Size	170,000 bp	4.6 million bp	130 million bp	3.2 billion bp	150 billion bp
Common Name	Virus	Bacteria	Fruit fly	Human	Canopy Plant

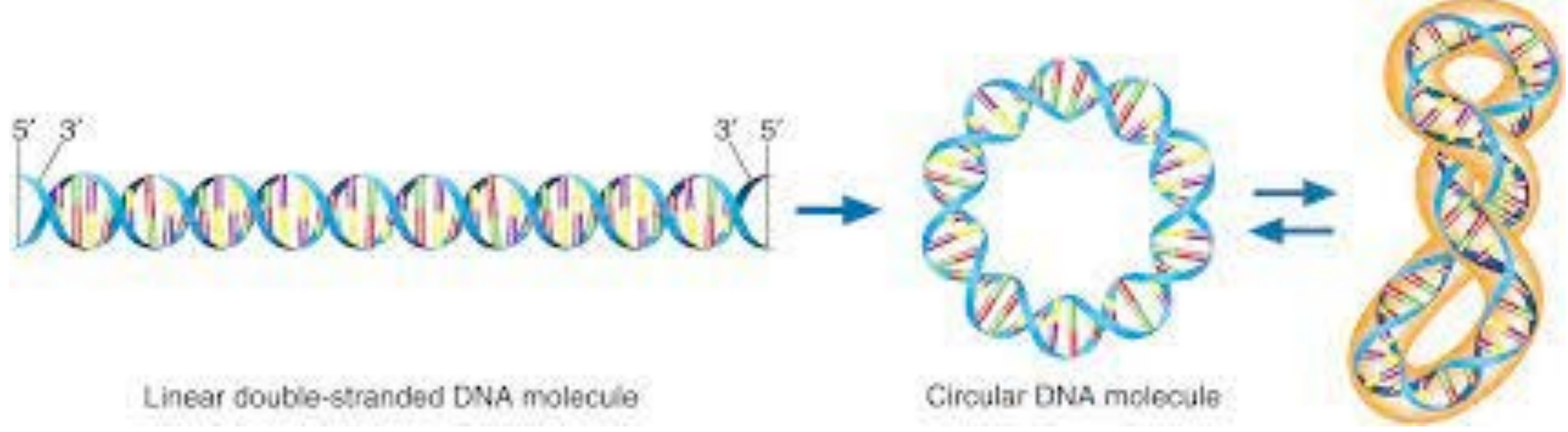
majority

genetic engineering



single stranded DNA (<100bases)

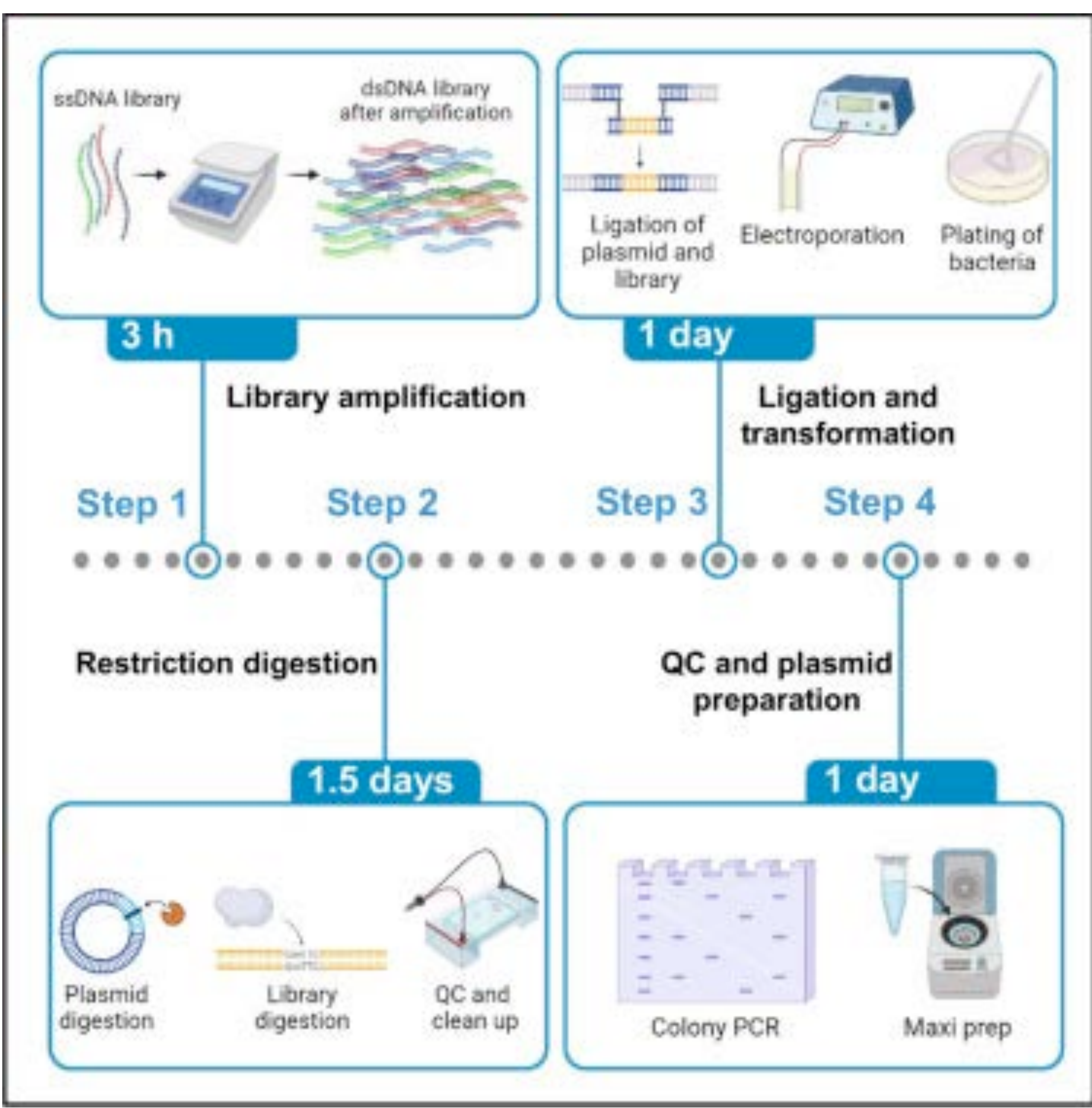
double stranded DNA (linear, <3000bp)



double stranded DNA (circular, <20Kbp)

fragments
(ssDNA-> dsDNA
ssDNA solid chemistry
dsDNA PCR

bigger plasmids
(golden gate assembly)

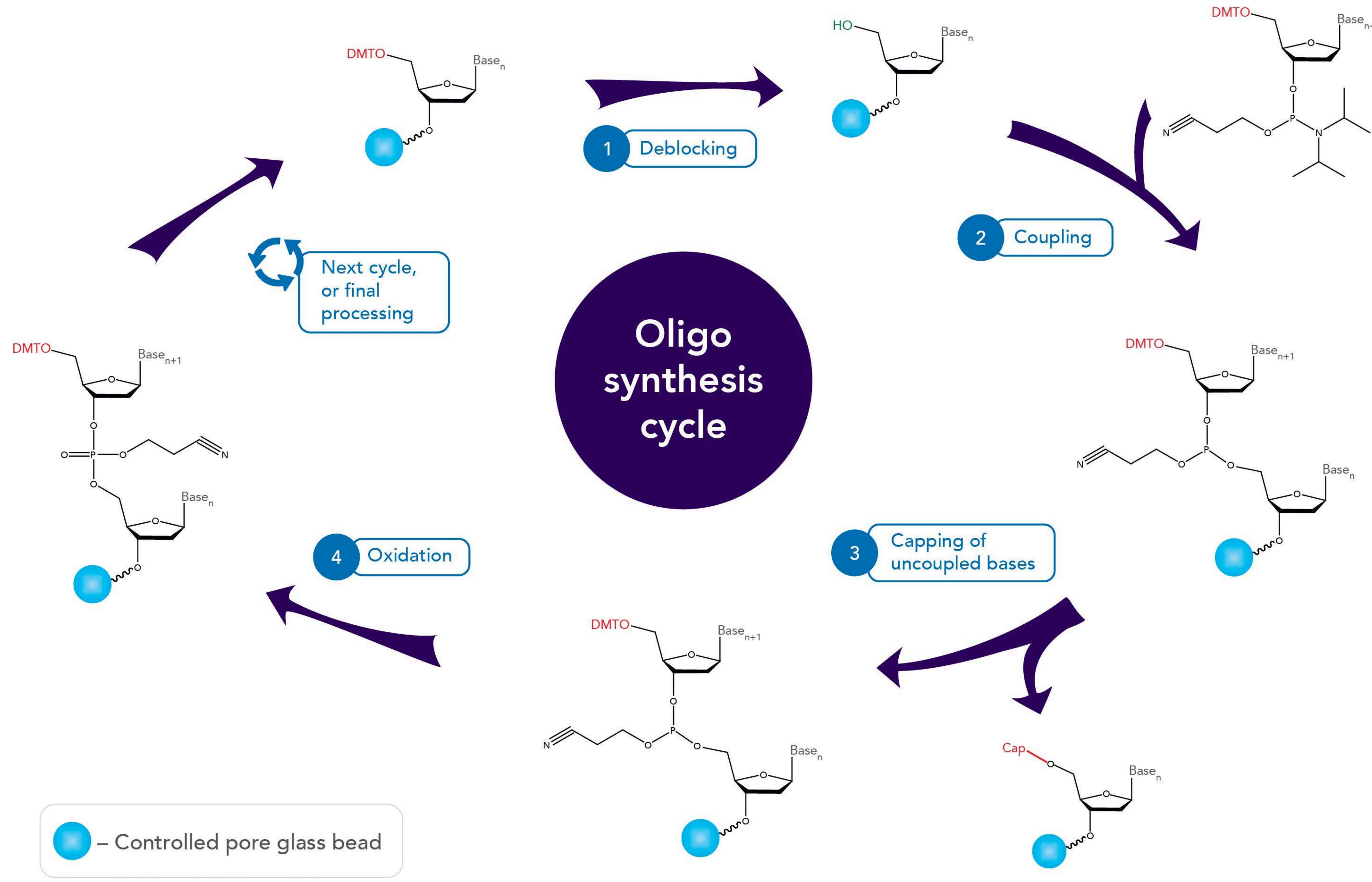


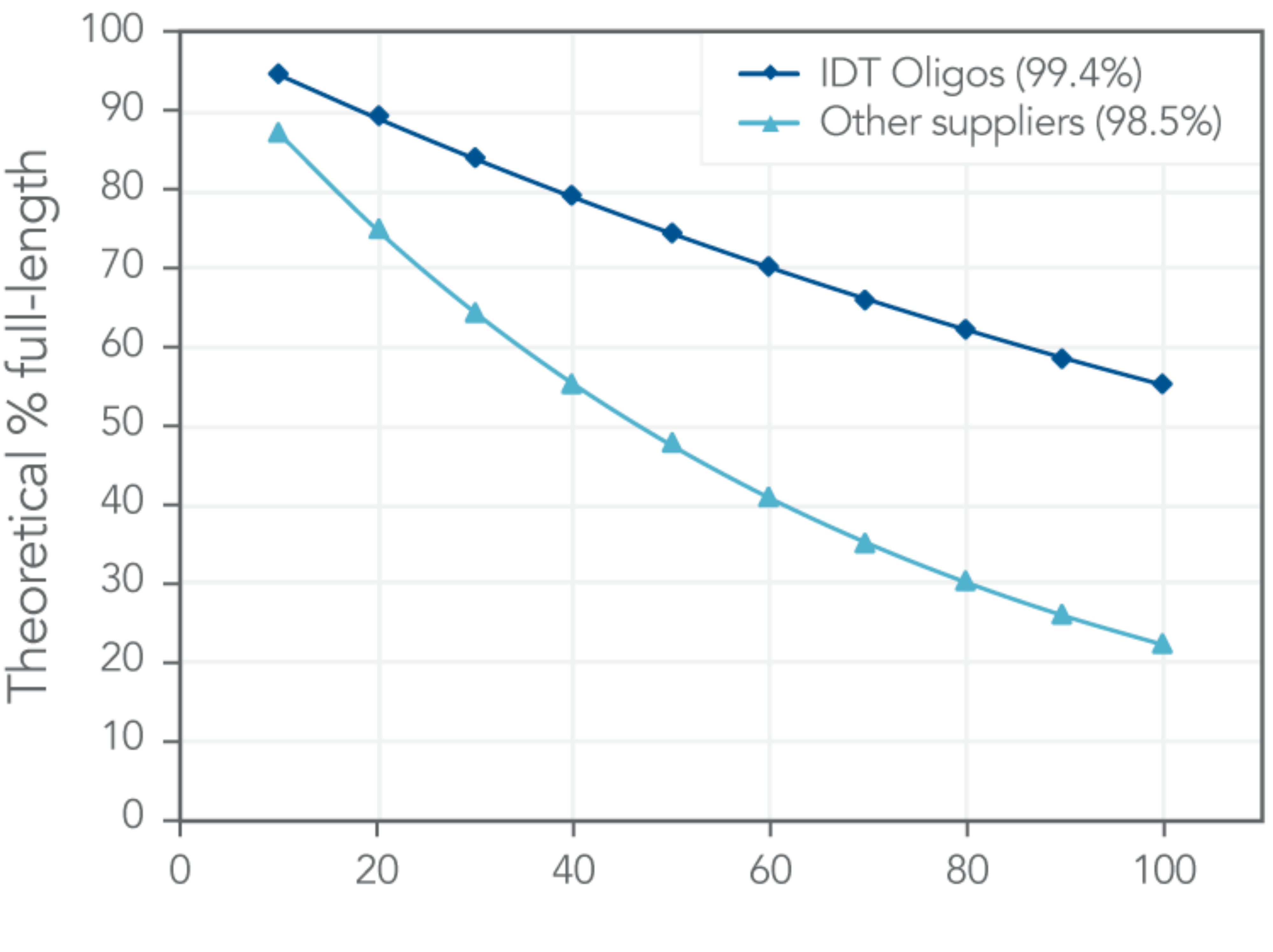
plasmids
(restriction/ligation)

validation
(size based, sequence based,
serial, massively parallel manner)

solid phase chemistry

1. ssDNA (<80bases, de novo, oligos, primers)
2. each cycle is imperfect
3. safe stops
4. 99.5%
5. 99.5%^80 = 60%
6. other 40% are truncated versions of the full sequence





IDT Oligos (99.4%)
Other suppliers (98.5%)
Theoretical % full-length
100
90
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
Oligo length (bases)

oligo -> ds DNA?

below are examples of de novo solutions

synthesize complementary oligos and mix them (limited by 80bases) - possible yourself

5’

5’

synthesize partially complementary oligos that overlap (TWIST bioscience, <1800bp, $.04/bp)

- repeats are a problem
- secondary structure is a problem
- need an algorithm that designs the overlaps
- need pure oligo source

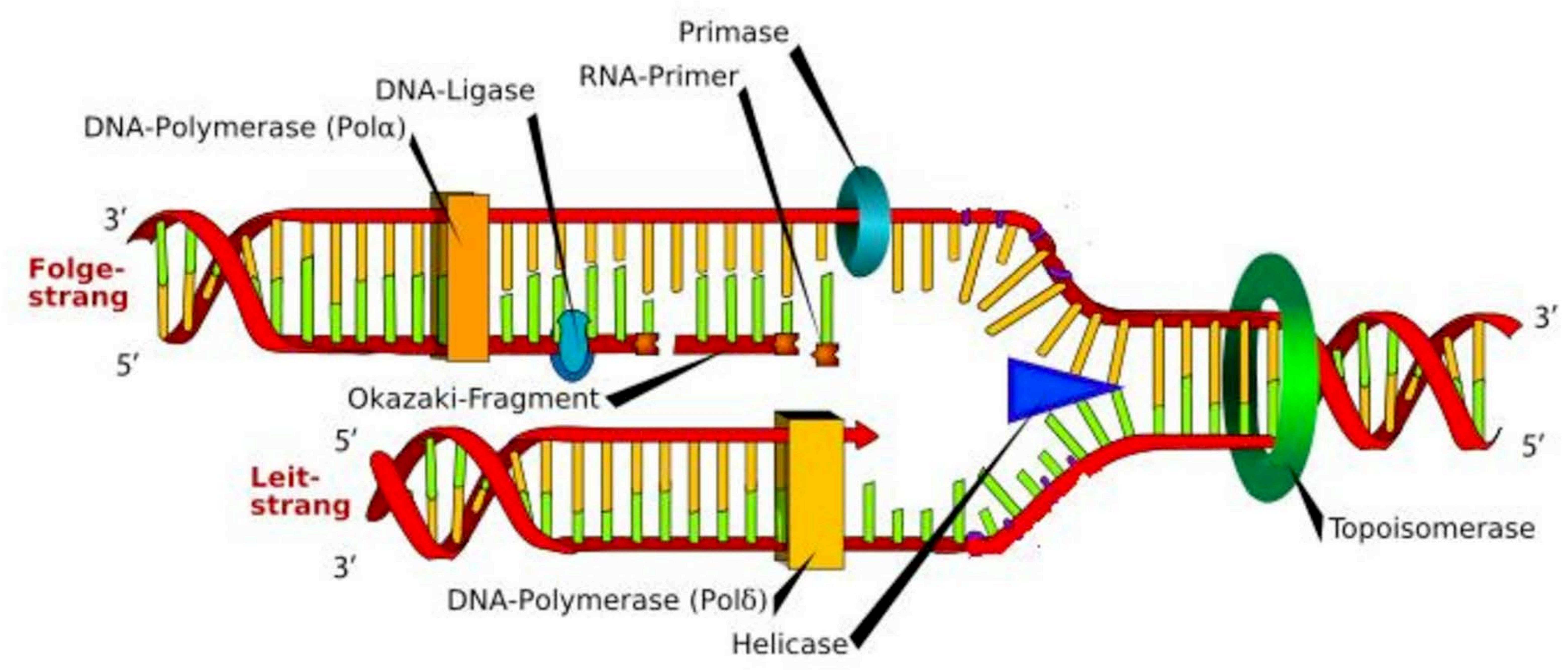
5’

5’

oligo + template ds DNA -> modified and amplified ds DNA

PCR is a method for amplifying DNA that also introduces changes

PCR is more reliable than de novo methods because it is enzymatic (amplifies DNA similarly to nature)



oligo + template ds DNA -> modified and amplified ds DNA

PCR is a method for amplifying DNA that also introduces changes

PCR is more reliable than de novo methods because it is enzymatic (amplifies DNA similarly to nature)

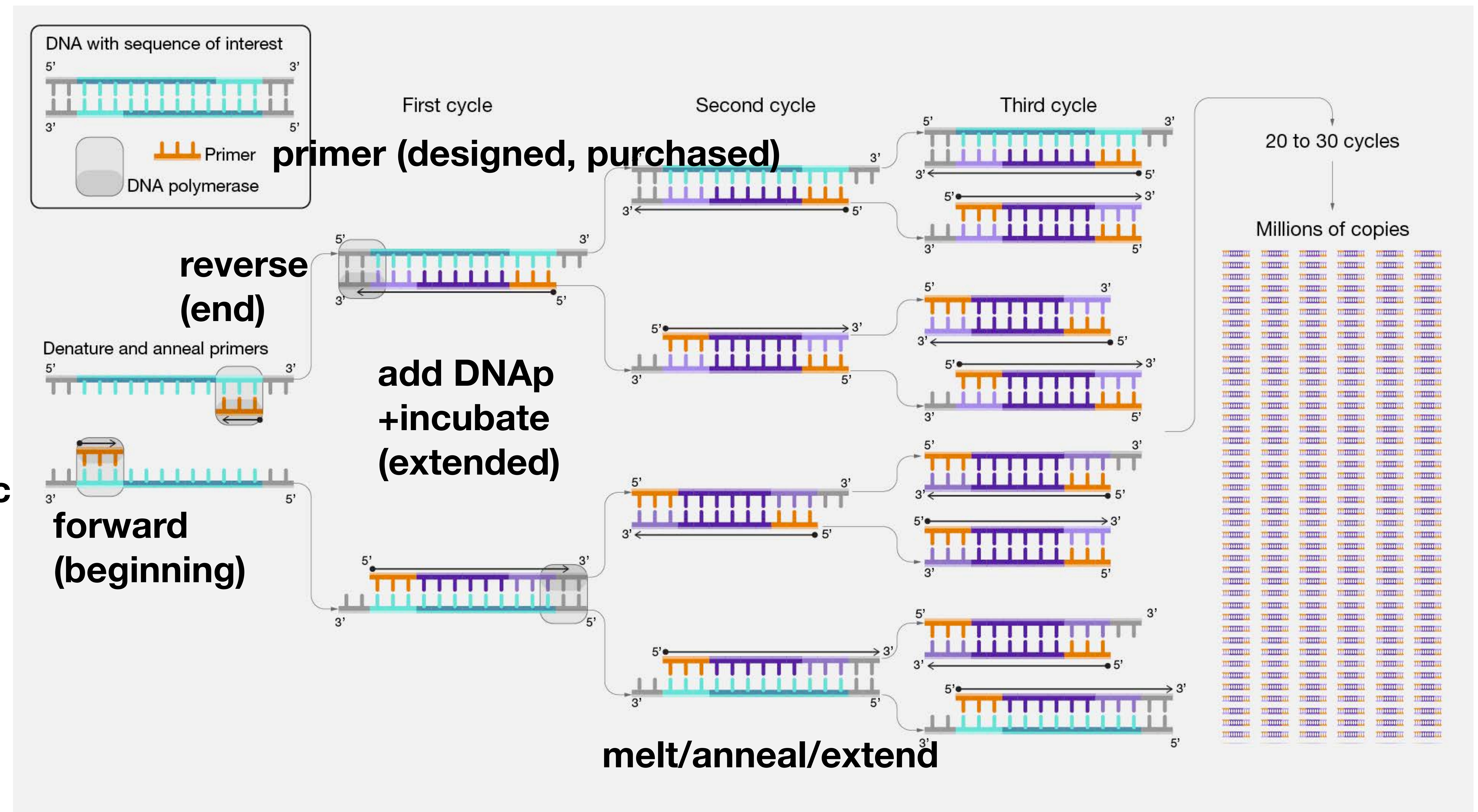
template (existing)

DNAp (purchased)

melt DNA (98C)

anneal DNA (60C)

primers high conc



primer (designed, purchased)

reverse (end)

forward (beginning)

add DNAp +incubate (extended)

melt/anneal/extend

discovery of thermophilic organisms (99C, thermal vents)

have to replicate possess DNAp that are stable up to 99C

oligo + template ds DNA -> modified and amplified ds DNA

PCR is a method for amplifying DNA that also introduces changes

PCR is more reliable than de novo methods because it is enzymatic (amplifies DNA similarly to nature)

template (existing)

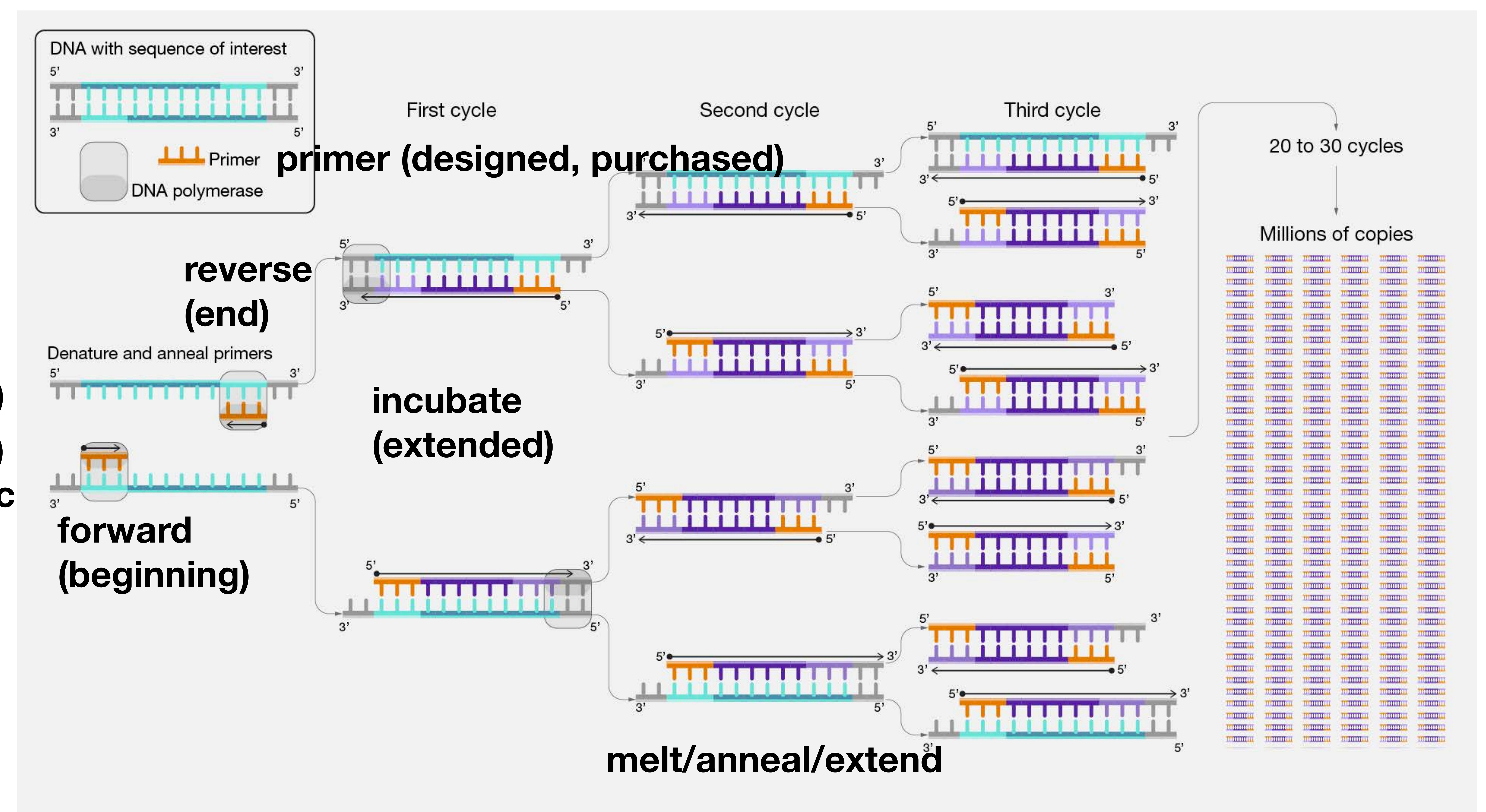
DNAp (purchased)

melt DNA (98C)

anneal DNA (60C)

primers high conc

add DNAp



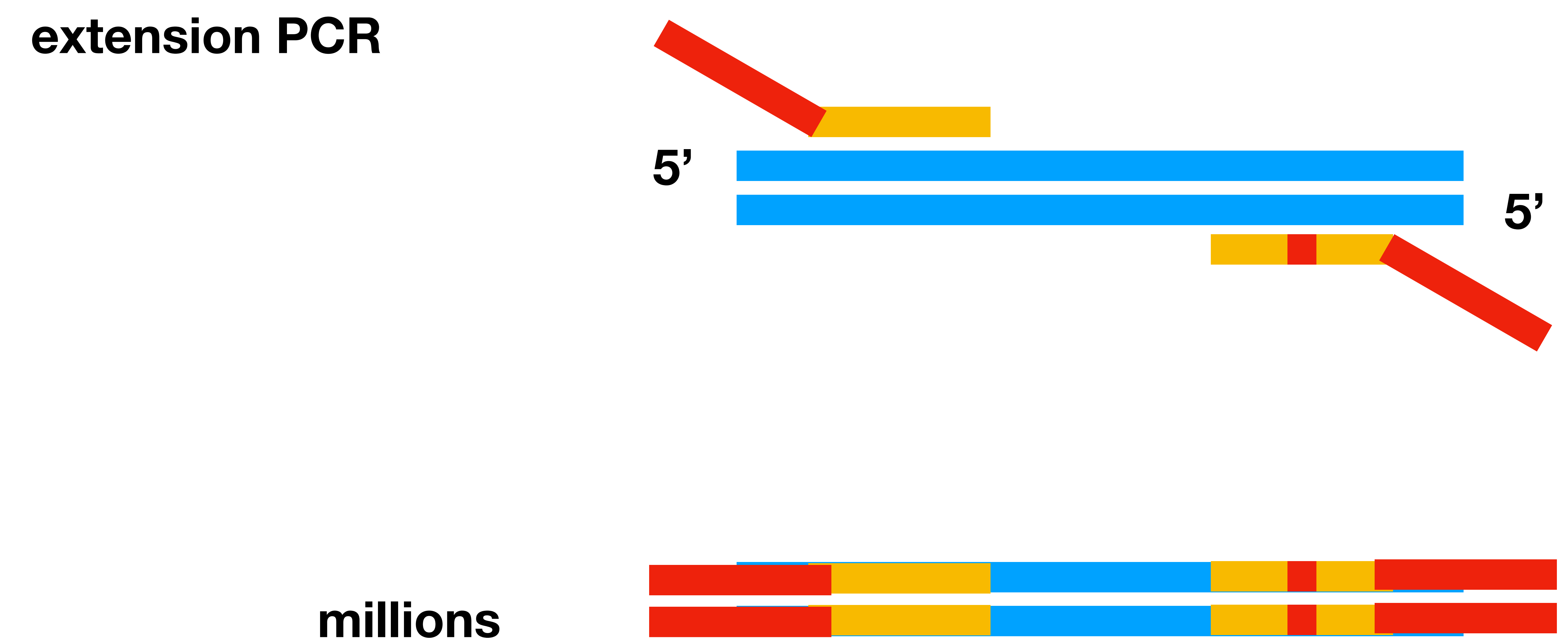
discovery of thermophilic organisms (99C, thermal vents)

have to replicate

possess DNAp that are stable up to 99C

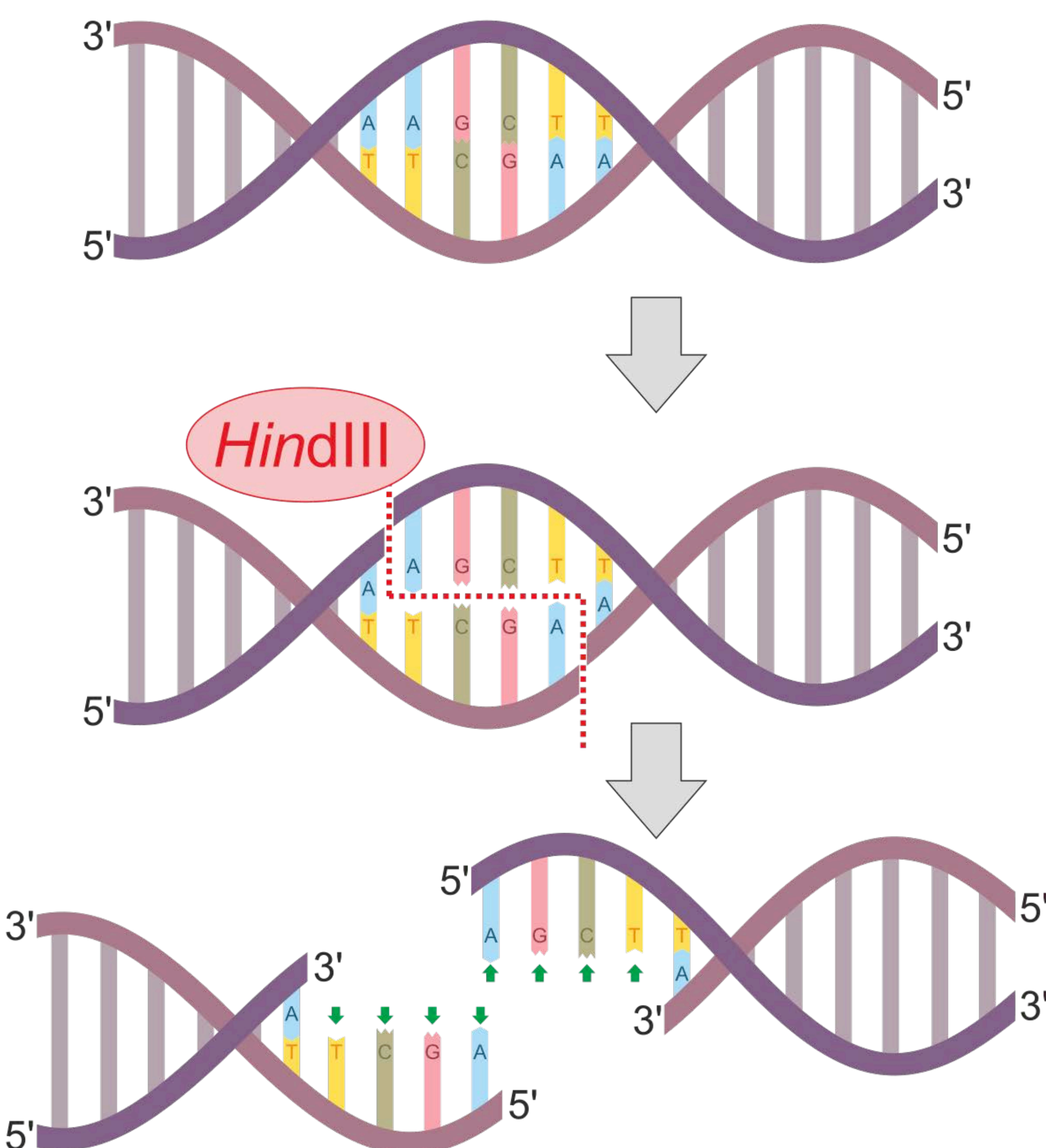
if primer sequence appears elsewhere (repetitive sequence, promoter) then PCR will lead to non-specific amplification

what else can PCR be used for?



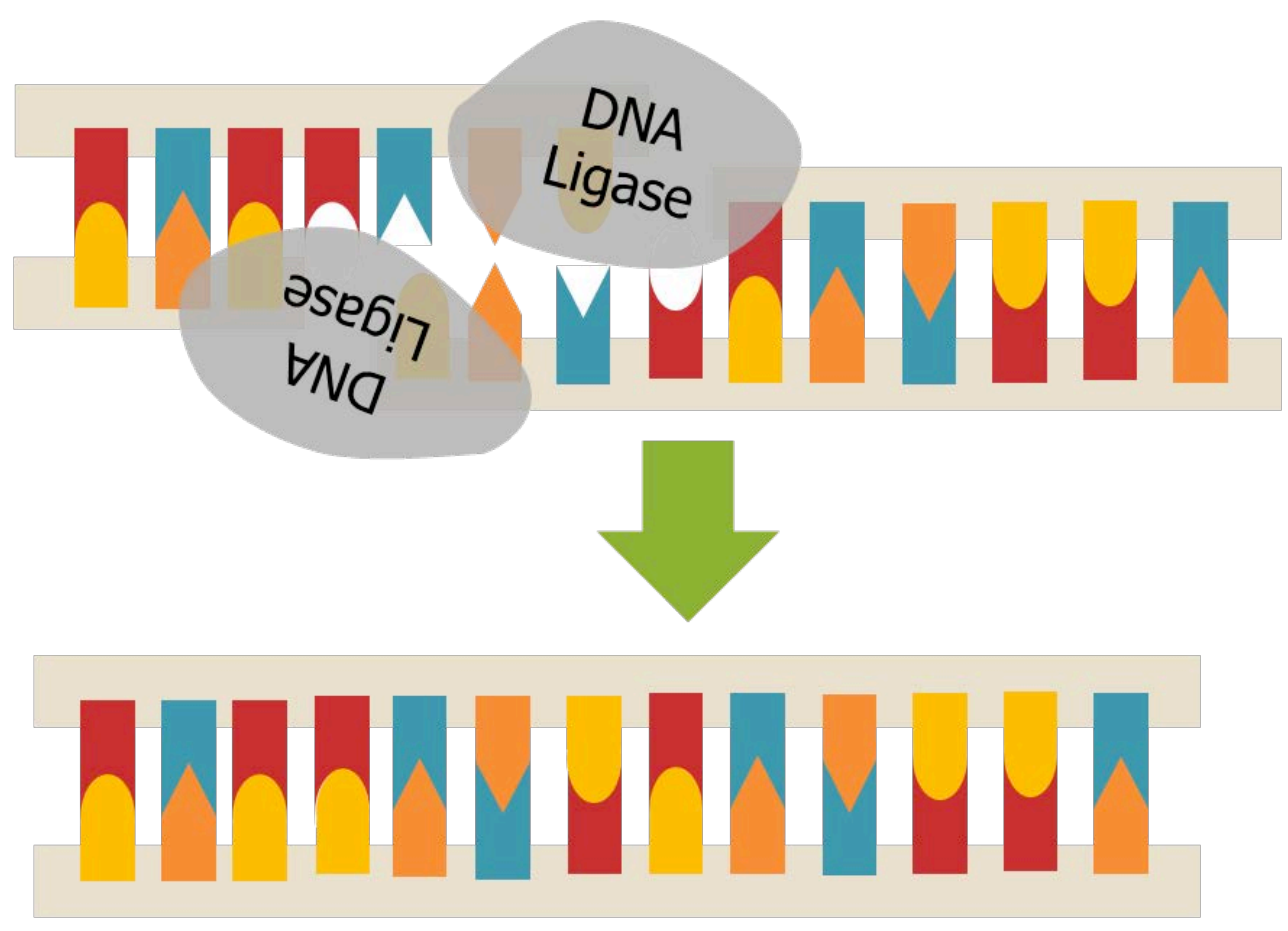
we've gone from oligo to fragment (~2k bp for de novo, ~10k bp for PCR)
how do we get to 20k bp?

endonuclease hydrolyzes DNA at a specific DNA sequence
DNA scissor
there are many different endonucleases used to cut DNA at various sites (EcoRI, HindIII, etc)



endonucleases can make blunt cuts (straight across) and jagged cuts (sticky ends)

endonucleases are also called restriction enzymes



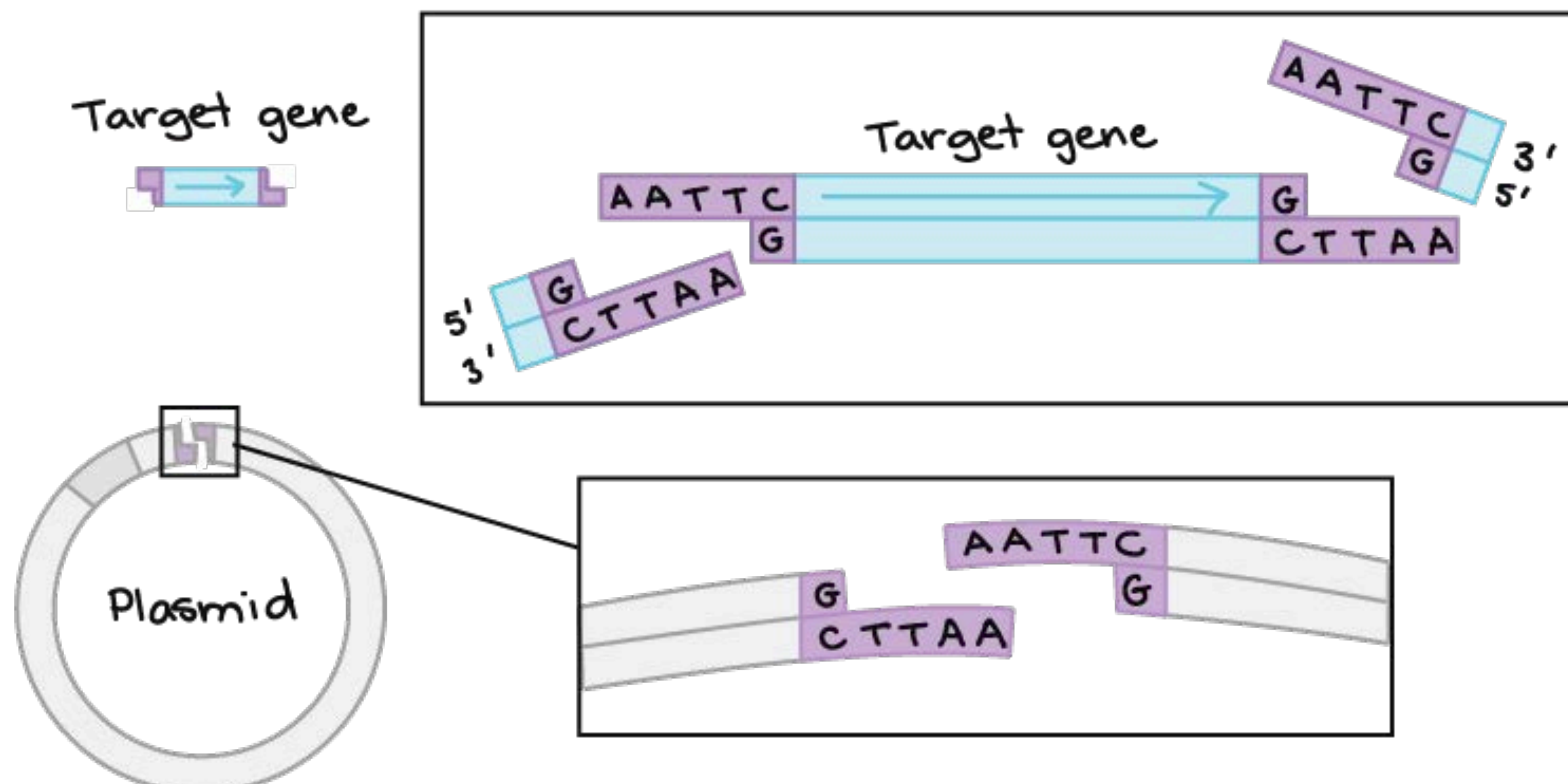
there are blunt end ligases and sticky end ligases

blunt end ligases can ligate any two blunt ends together

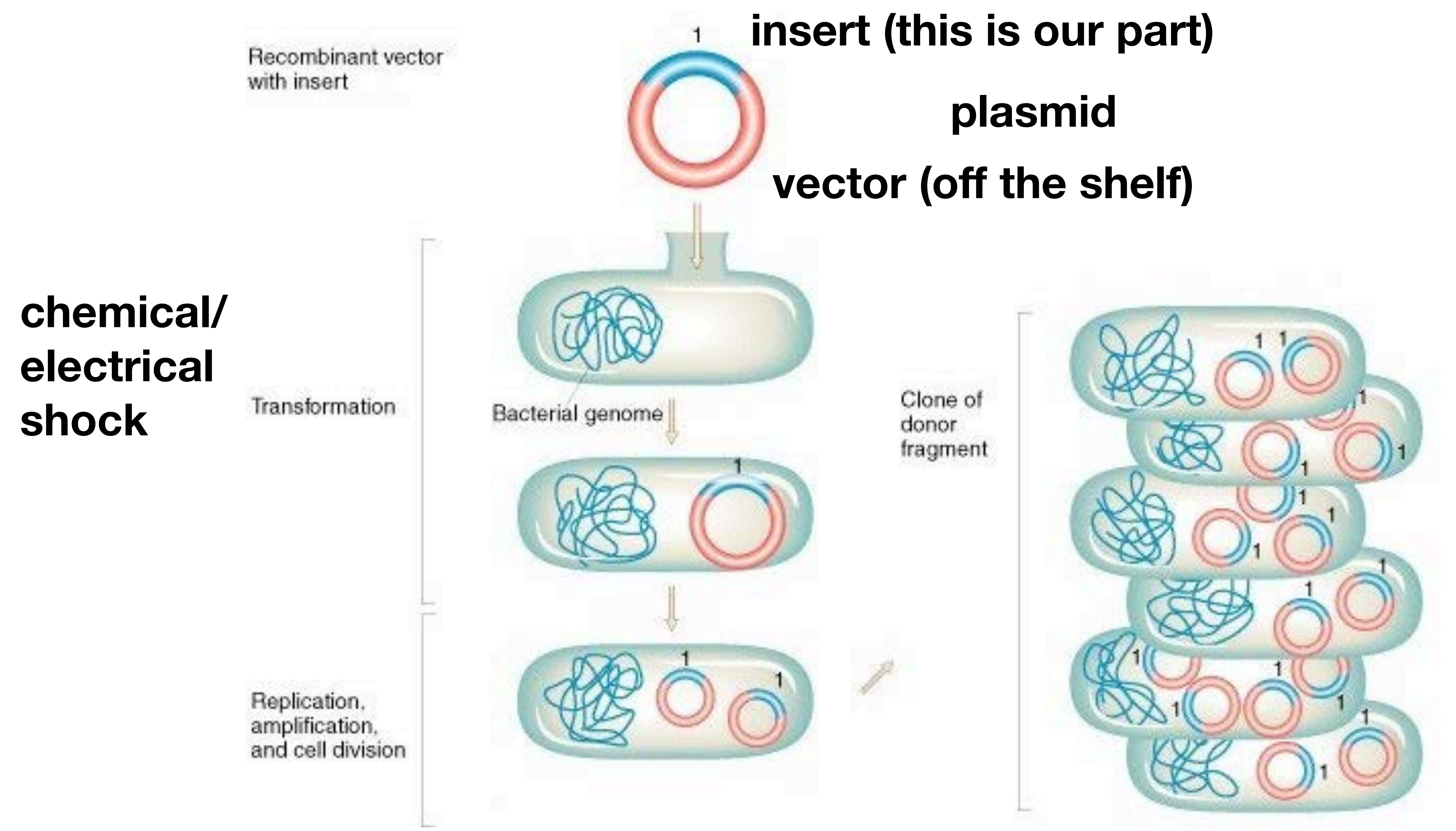
sticky end ligases require the template to be present

restriction/ligation

1. place a DNA fragment in a vector
2. vector is a piece of DNA that can be effectively selected for and replicated in an organism
3. selection refers to the identification of organisms that have the DNA sequence in them
4. replication refers to a sequence that tells the organism to copy the entire sequence during cell division
5. vector is usually in the form of a plasmid
6. plasmid is a circular piece of DNA used in nature for horizontal DNA transfer
7. horizontal DNA transfer is most often employed by microorganism to quickly improve their fitness (for instance to gain antibiotic resistance)
8. vector can be anywhere from 5kbp to 20kbp



bacterial transformation and plasmid amplification

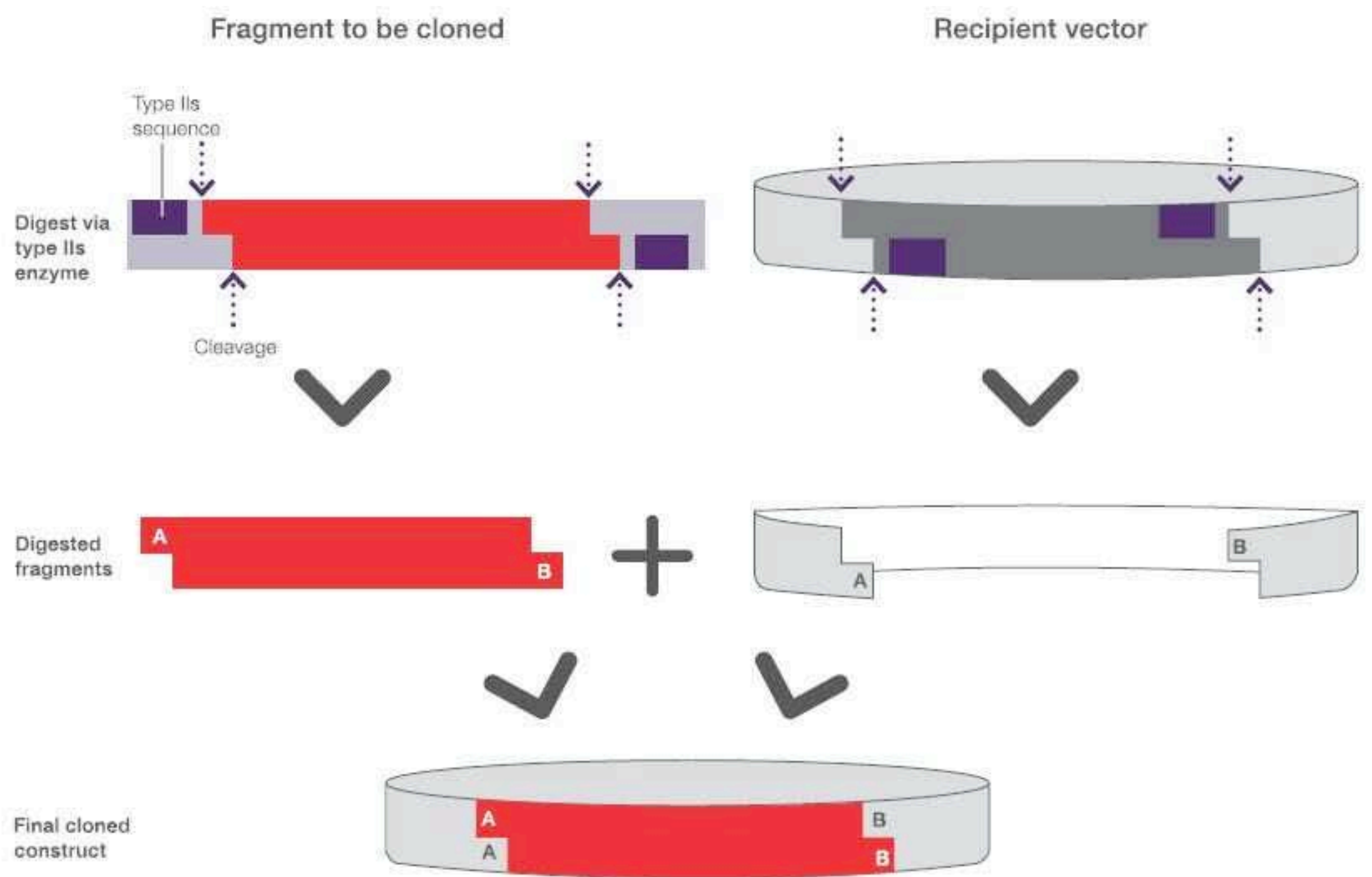


this can also be used to make copies of designed DNA

the most reliable way of making copies of large pieces of DNA

DNA replication in test tube is roughly 2 orders of magnitude less reliable (error prone) than DNA replication in a simple bacterium

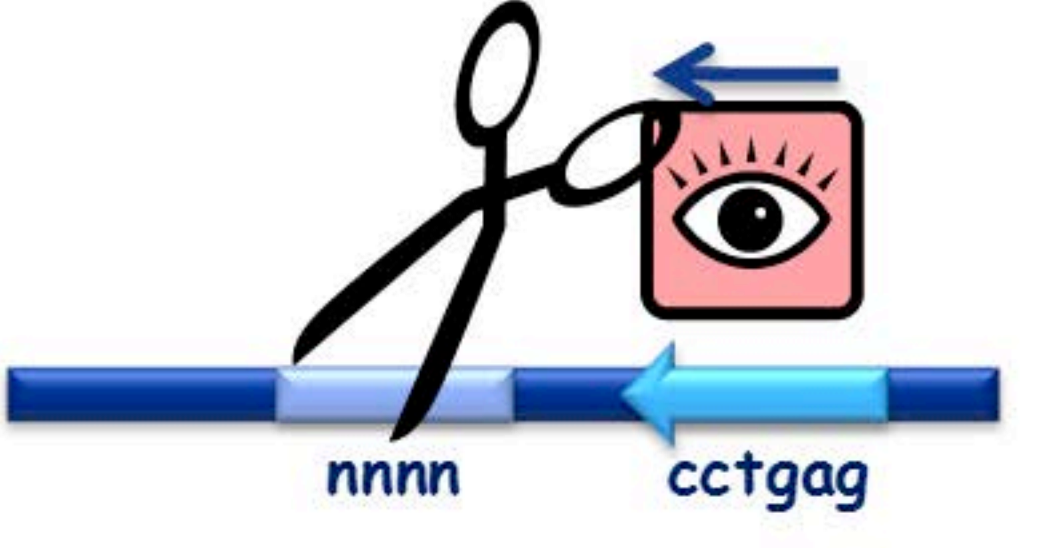
restriction/digestion (cloning)

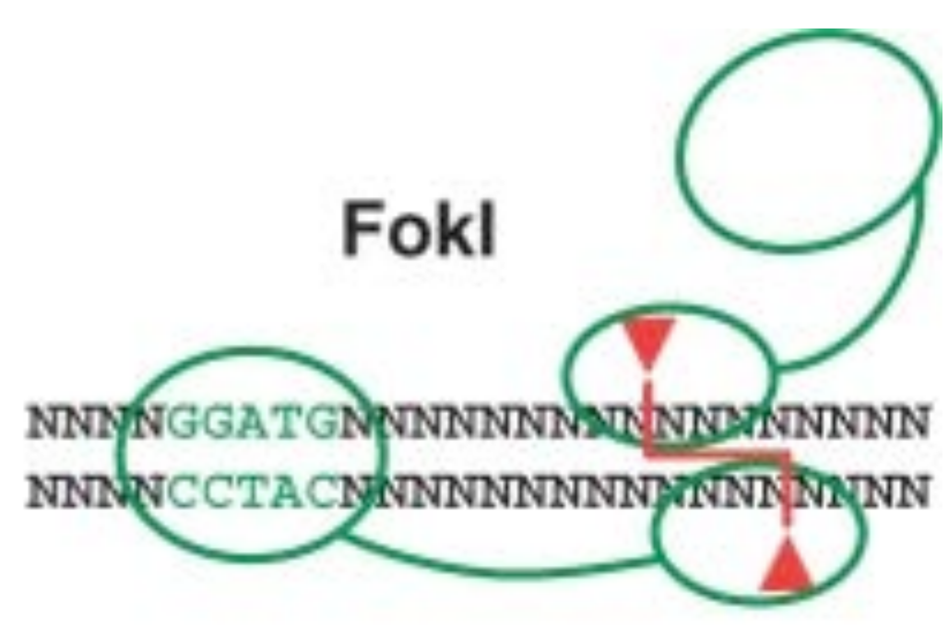


type IIS restriction enzymes

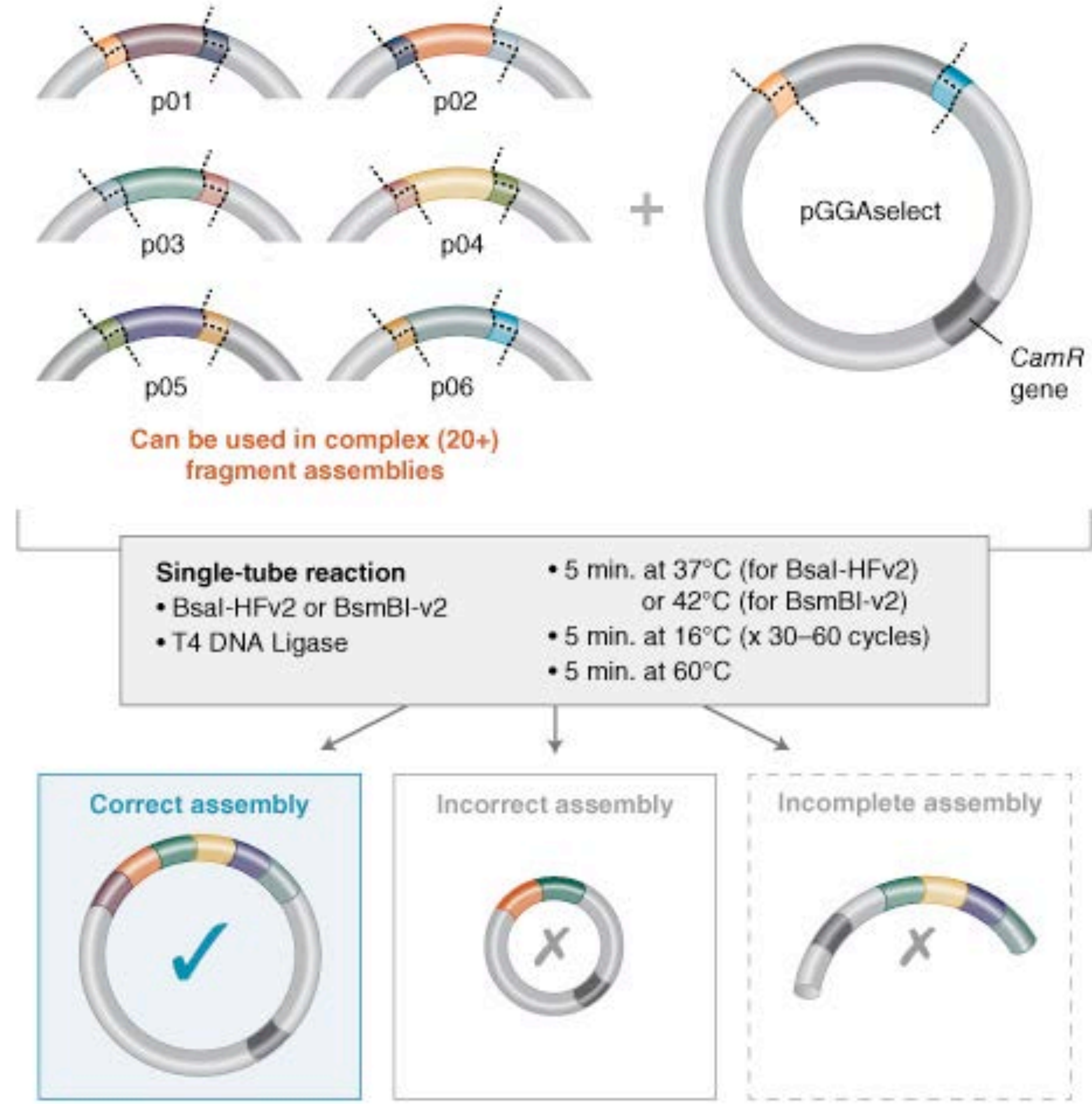
Type IIS restriction enzymes recognize asymmetric DNA sequences and cleave outside of their recognition sequence.

5´... G G T C T C $(N)_1$▼... 3´
3´... C C A G A G $(N)_5$▲... 5´

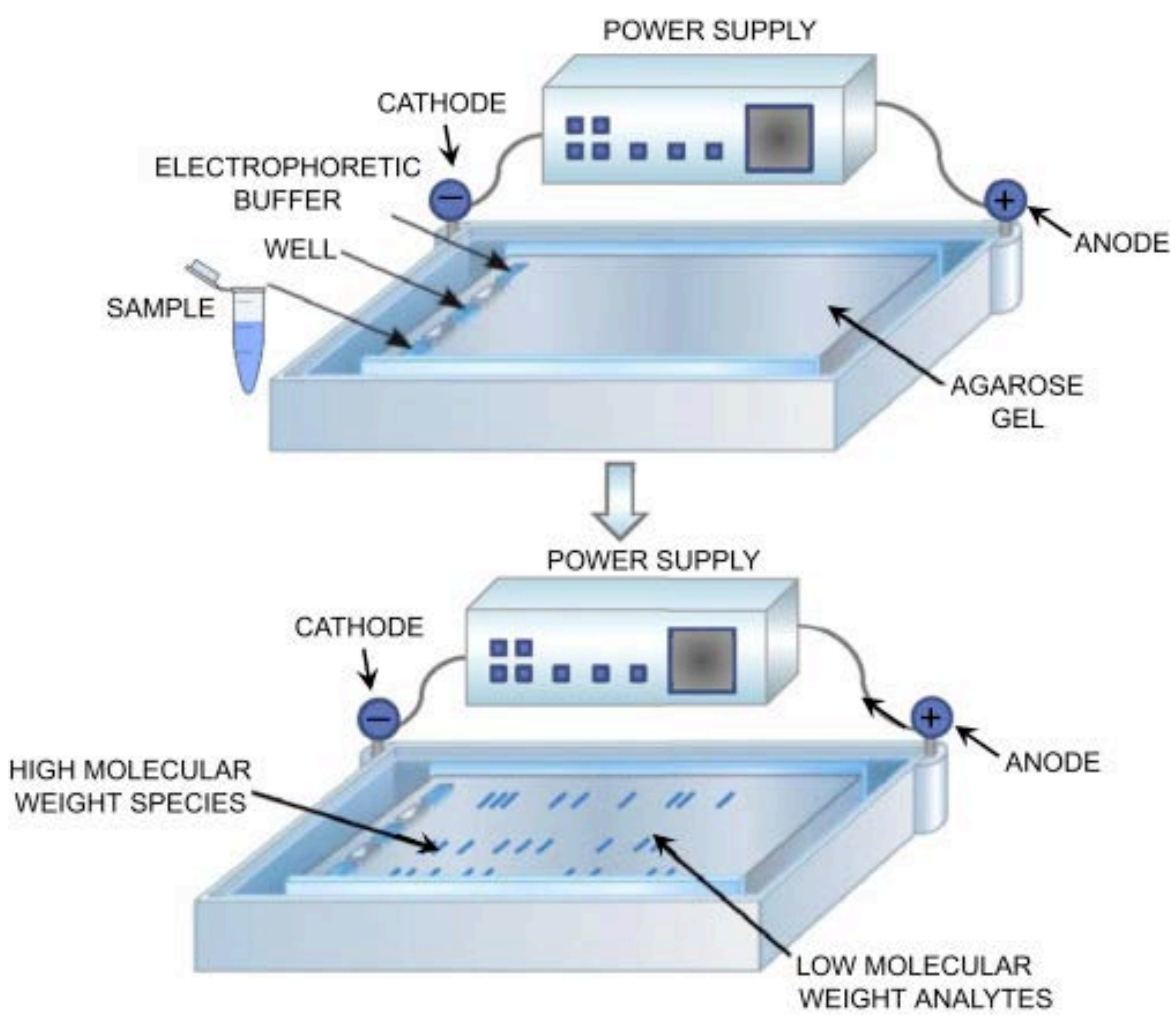




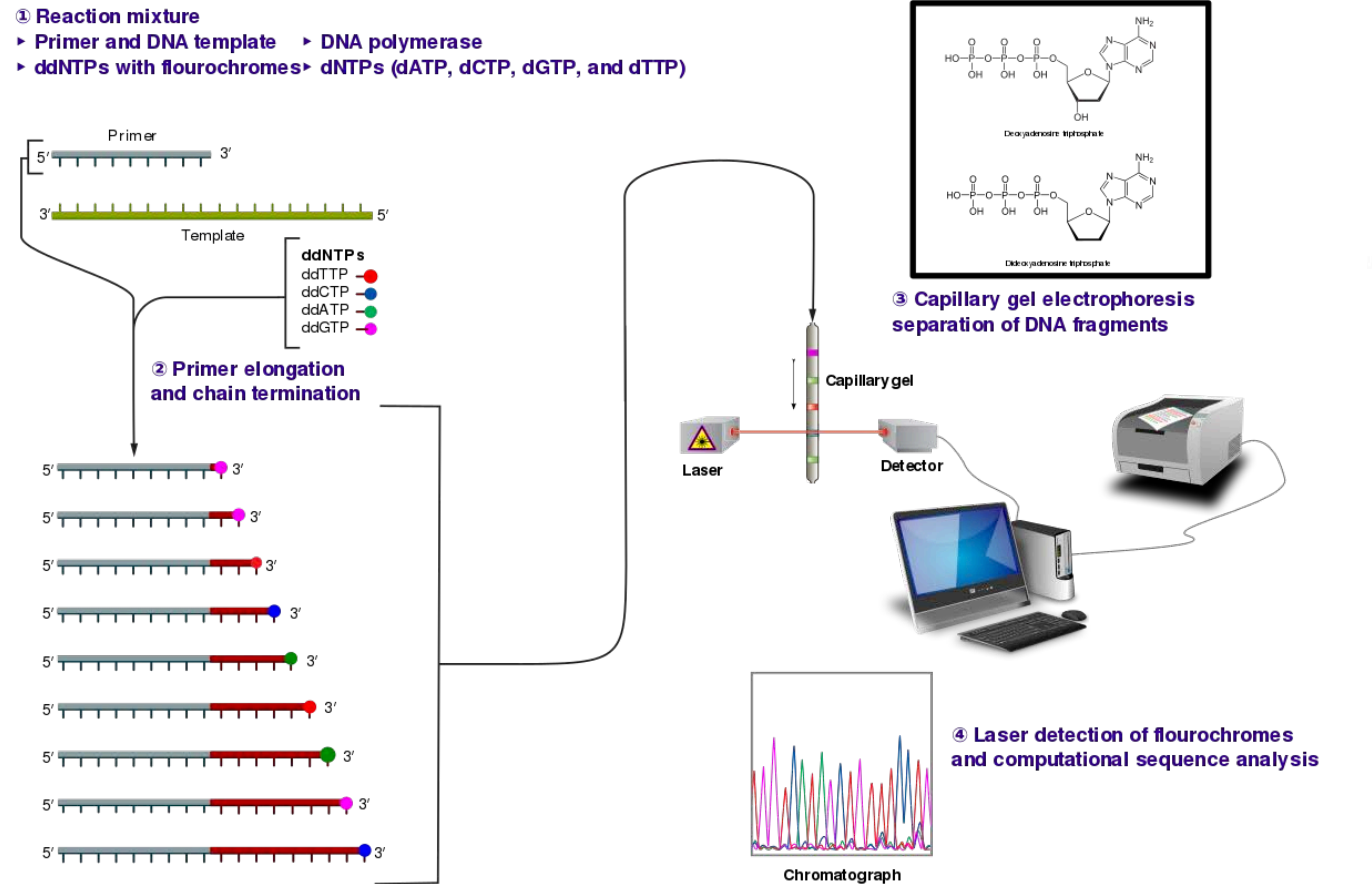
golden gate assembly



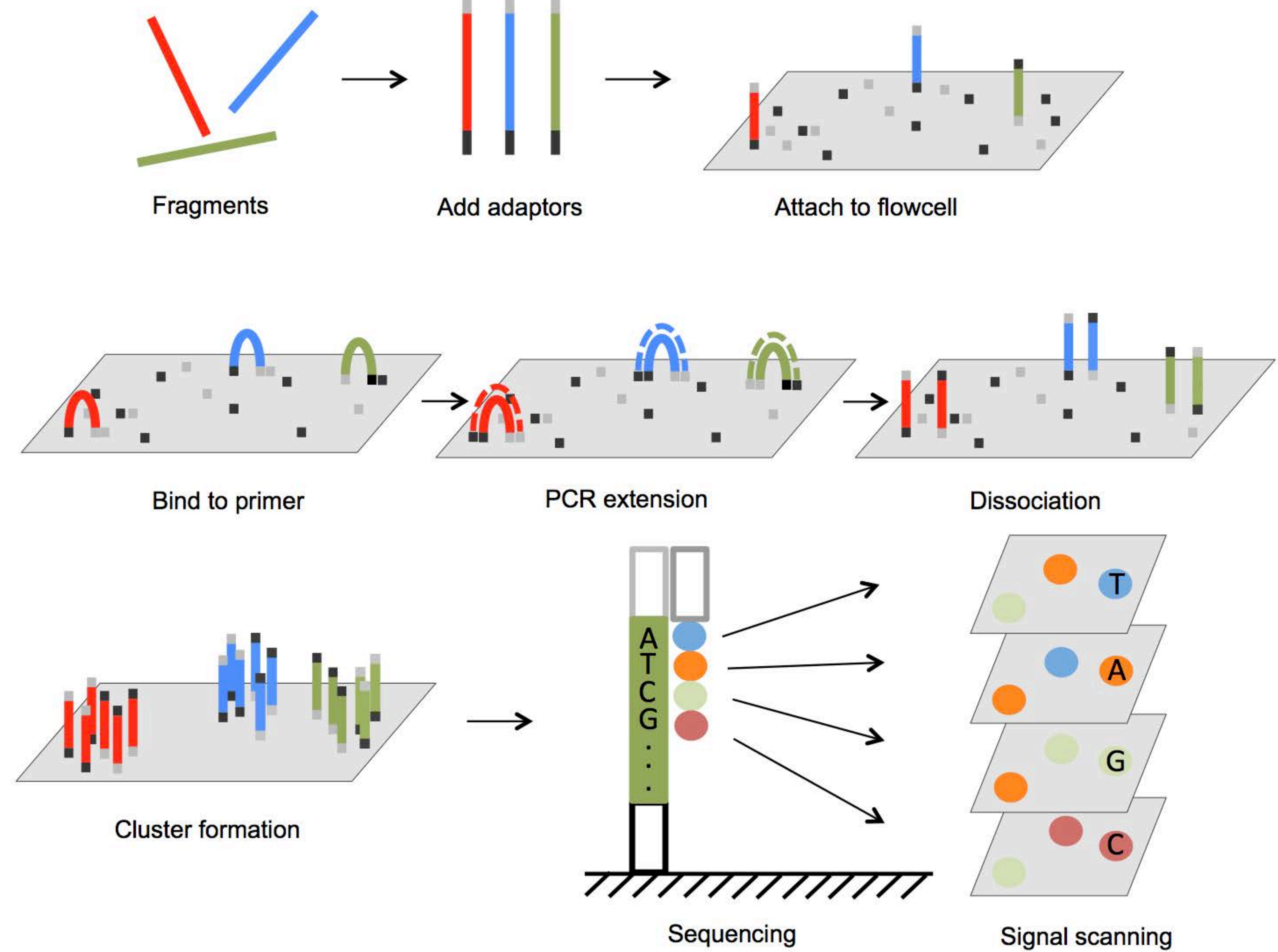
gel electrophoresis



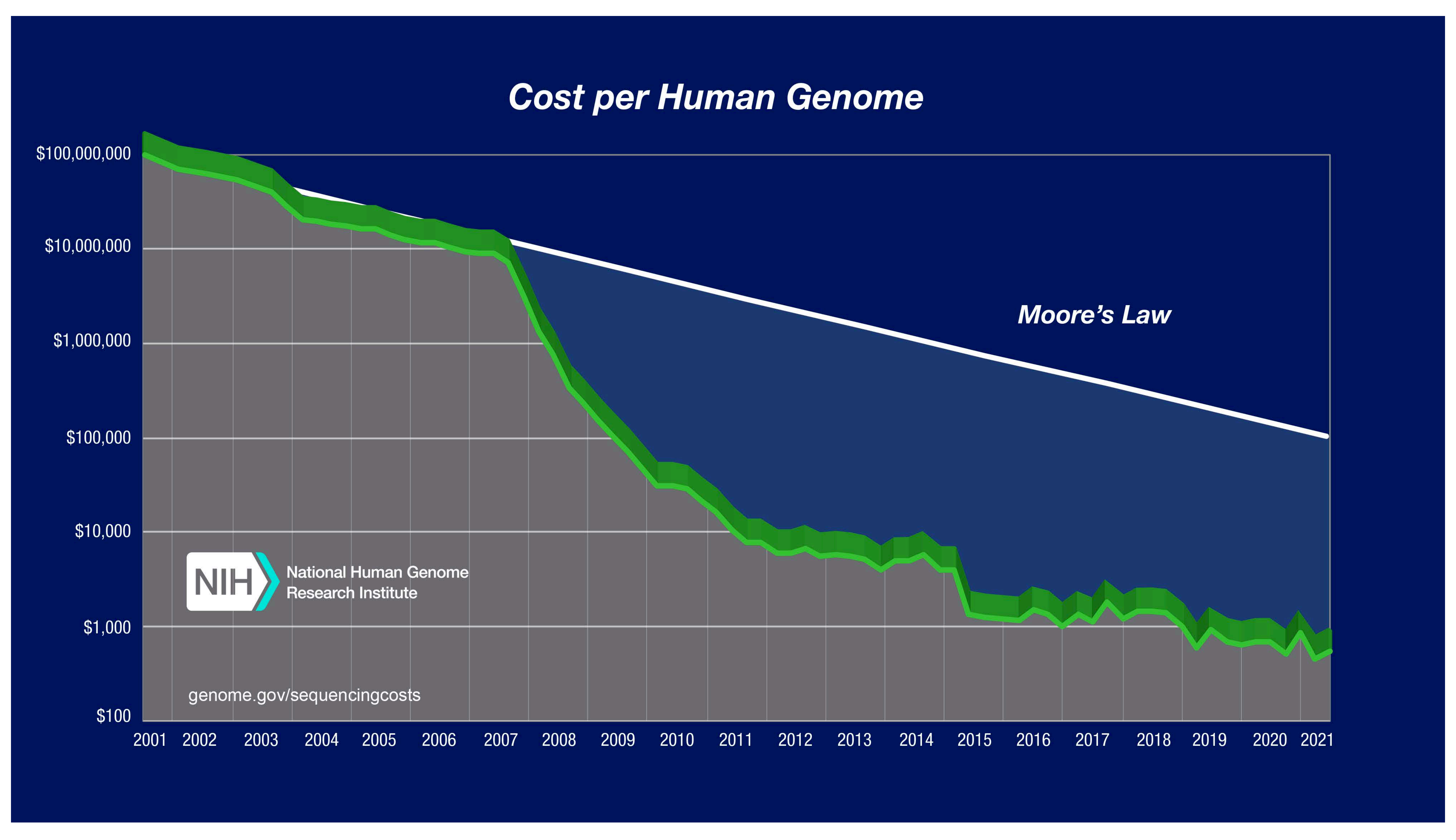
sanger sequencing



massively parallel sequencing



evolution of sequencing costs



Assignment 3

1. **Amplify previously analyzed GFP gene with 5’ prime and 3’ cloning extensions**
    1. **annotate the gene starting 500bp upstream of the start codon and ending at the end of the terminator**
    2. **copy and paste the sequence into a new file**
    3. **design type IIS BsaI extensions with the 5’ insert overhang being CCCT and the 3’ insert overhang being CCGA**
    4. **design forward and reverse primers that anneal to the template at approx. 56C**
    5. **anneal the primers to your template**
    6. **execute the virtual PCR reaction on your template sequence**
    7. **provide link to your template with annealed primers**
    8. **provide links to your forward and reverse primers**
    9. **provide links to your PCR amplicate**
2. **Clone the insert into the provided vector named Vector-AmpR**
    1. **run Golden Gate assembly program with BsaI endonuclease**
    2. **finalize the assembly**
    3. **provide link to your final plasmid**
3. **Validate the sequence**
    1. **locate the NotI restriction sites**
    2. **compute the length of the two fragments flanked by the restriction sites**
    3. **run the virtual digest on benchling**
    4. **provide a printscreen of your restriction digest**