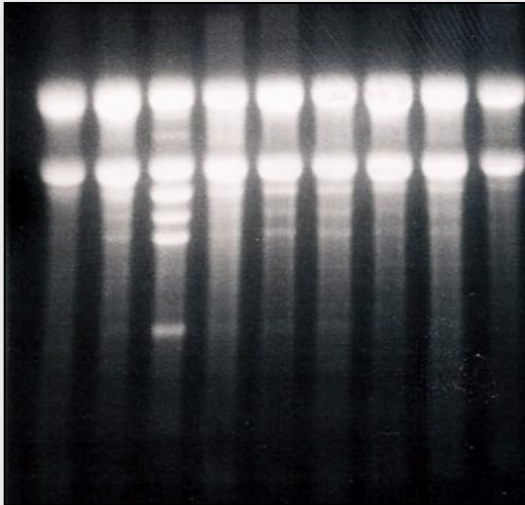


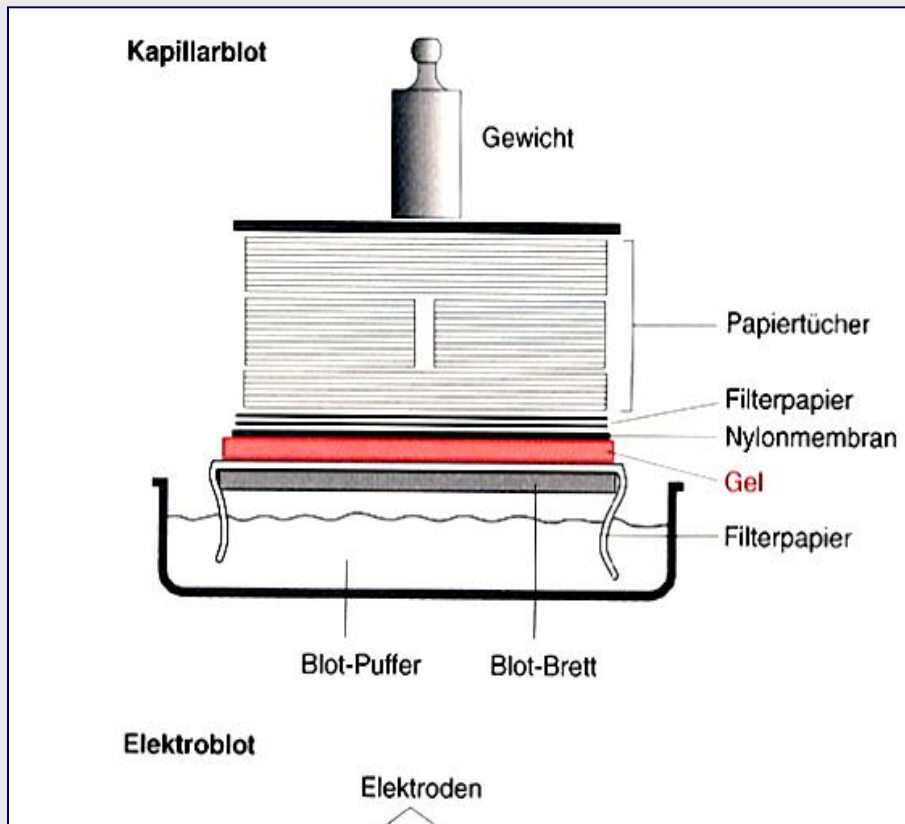
Isolation von intakter RNA und Auftrennung im Agarosegel (denaturierend, Formaldehyd)



→Formaldehyd: reagiert mit Aminogruppen der Purin- und Pyrimidin-Basen unter Bildung einer Schiff'schen Base, verhindert damit die Ausbildung von Wasserstoffbrückenbindungen

→Formamid: Doppelhelix destabilisierendes Agens, verhindert die Reassoziaton von RNA-Molekülen

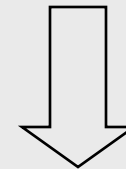
Transfer der getrennten RNA auf Nylonmembranen durch Kapillarblotting



Bindung der RNA an die Nylonmembran

→ Backen: 2 h 75-80°C

→ Cross linking: UV-Strahlen



Hybridisierung der Membran mit der markierten Gensonde

Markierungsmethoden von „Sonden“

1. 5'-Endmarkierung mit T4-Polynukleotid Kinase
2. 3'-Endmarkierung mit Terminaler Transferase
3. Nick-Translation
4. PCR-Labeling
5. Random Priming

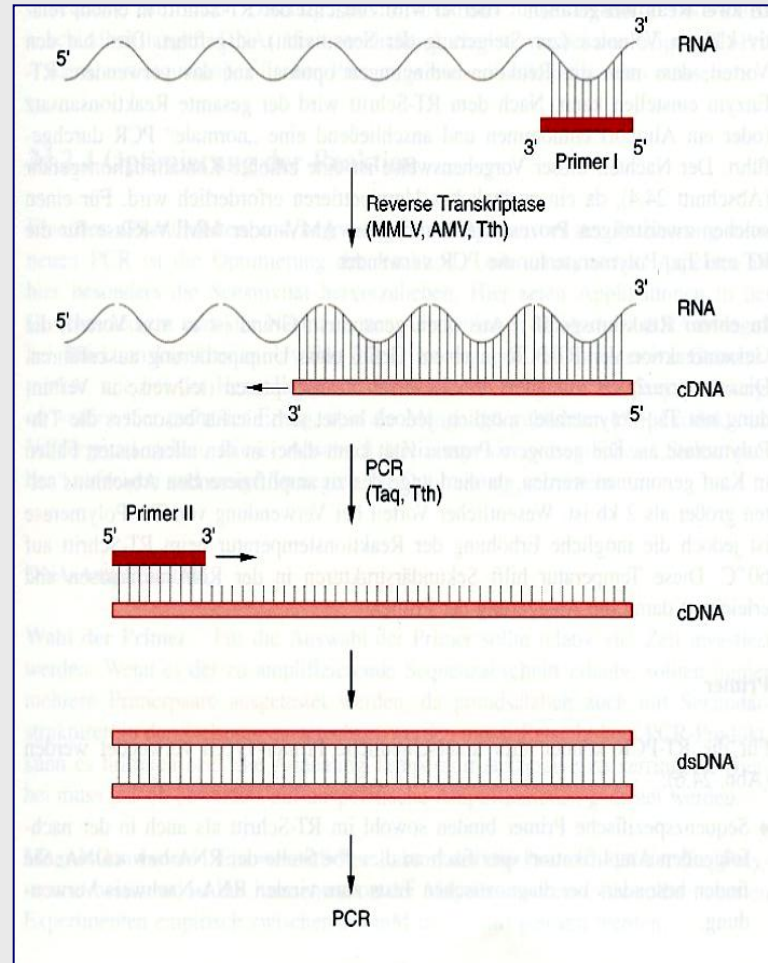


+Fe -Fe

Northern-Blot-Hybridisierung

RT-PCR

→ Kombination von reverser Transkription und Polymerase Ketten Reaktion (PCR)



RT-PCR

Vorteile:

- Sehr sensitiv durch Einsatz der PCR-Technik
- Sehr spezifisch, Unterscheidung von sehr ähnlichen Genen einer Gen-Familie möglich

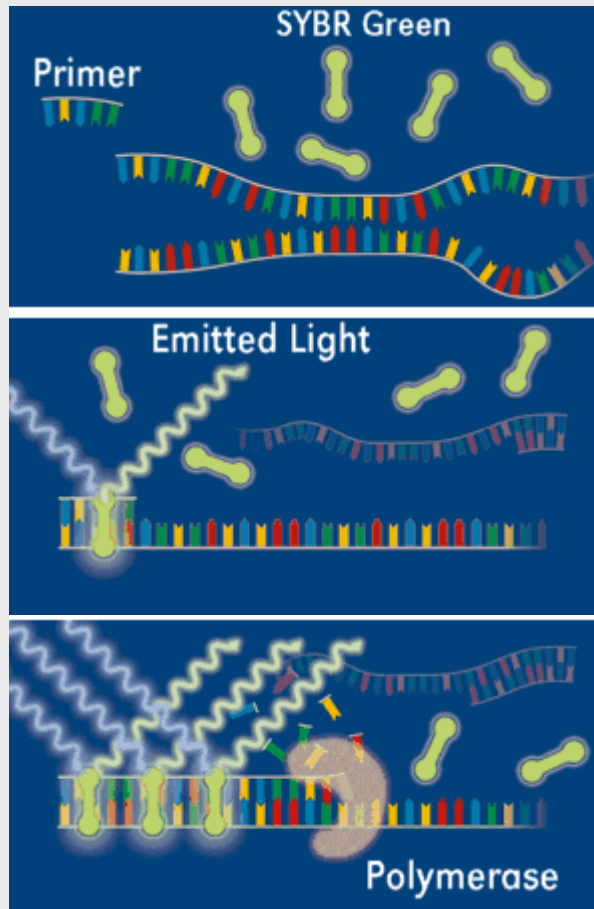
Problem:

- Nicht wirklich quantitativ! Nur qualitative Aussagen möglich!!

Voraussetzung:

- Hochreine RNA, keine Kontamination mit genomischer DNA
- Mitführen einer konstitutiven Genkontrolle

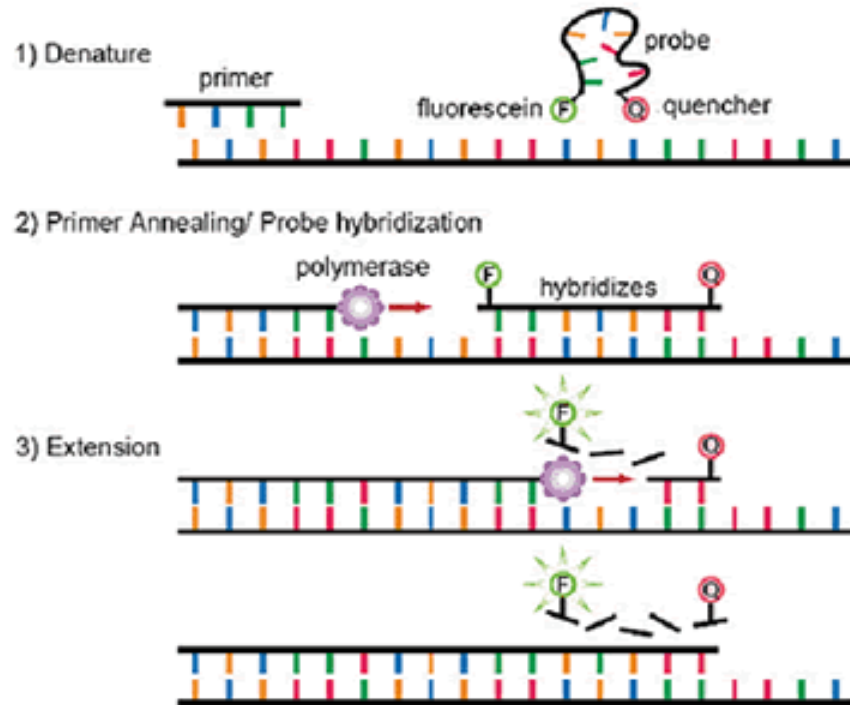
Real time-PCR (qPCR)



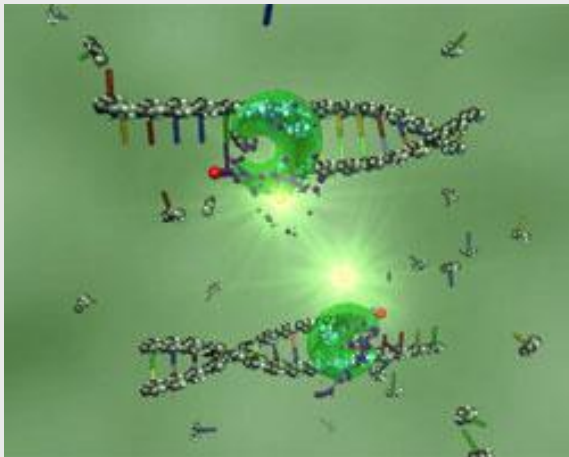
- Methode zur genauen Quantifizierung von mRNAs
 - Beruht auf Einbau eines Farbstoffs, der bei Interkalation in DNA-Doppelstränge fluoresziert
- „SYBR-Green“

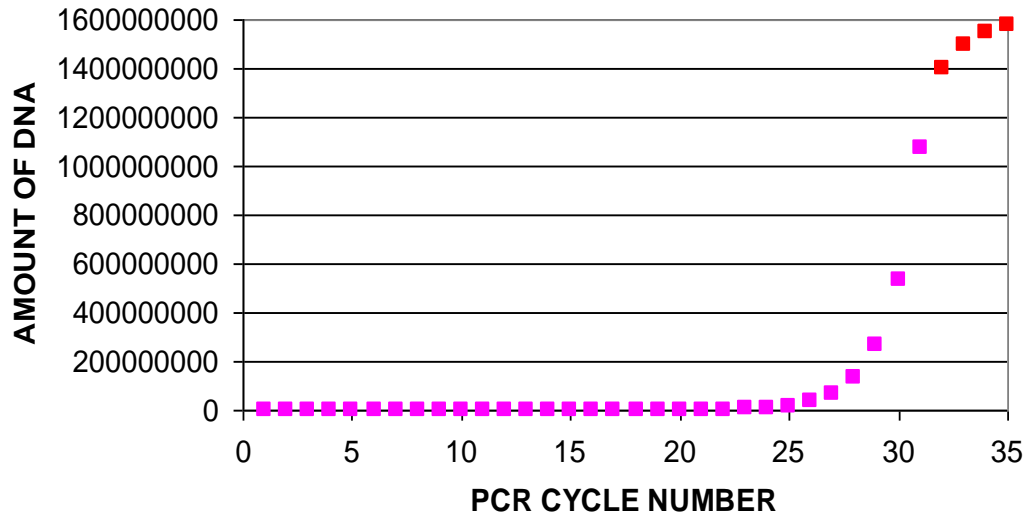


Real time - PCR (TaqMan^R)



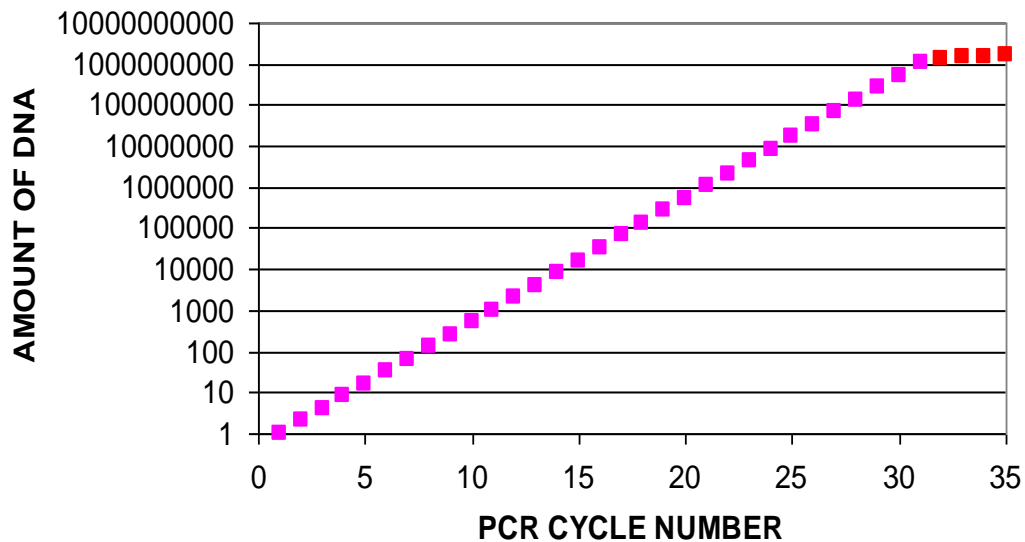
TaqMan[®] Probe Method

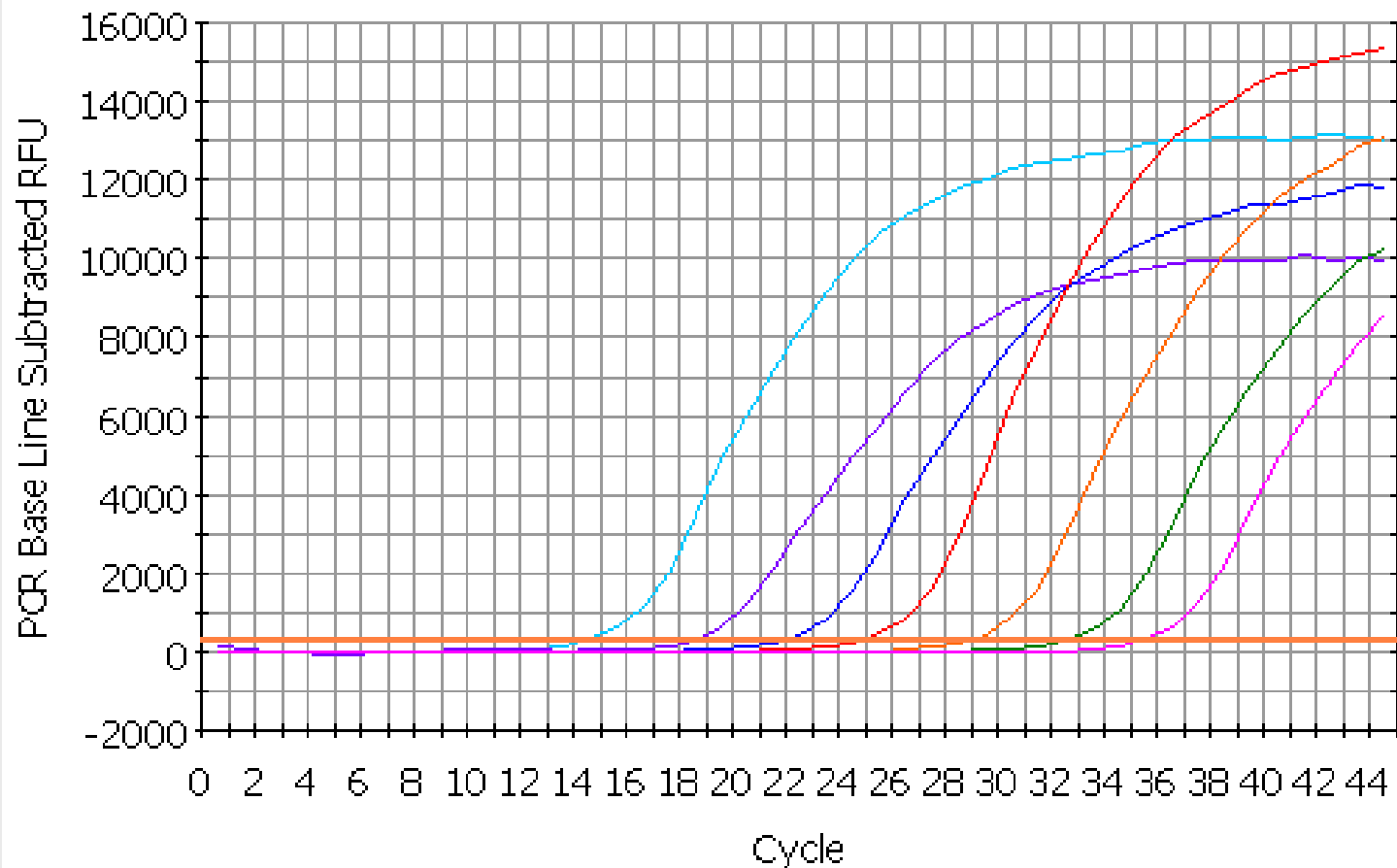




- bis zur Sättigung der PCR erfolgt DNA-Amplifikation logarithmisch

- Messung der DNA-Menge im „linearen“ Bereich





(Reihe von 10-fach Verdünnung.)

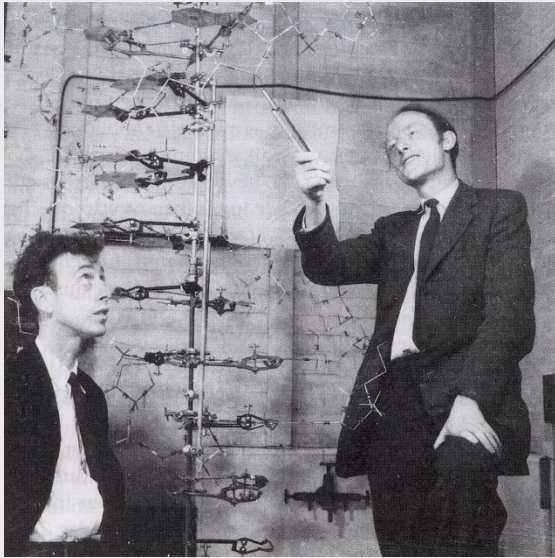
- anhand der Amplifikationskurven kann die eingesetzte DNA-Konzentration berechnet werden
- Die Kurve gibt auch Aufschluss über die Spezifität der Reaktion

Verfahren zur Sequenzierung von DNA

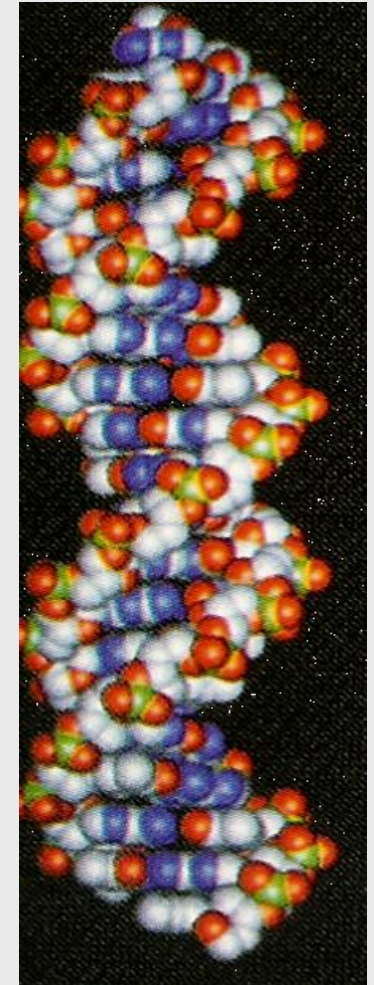
Prof. Dr. Martin Hagemann



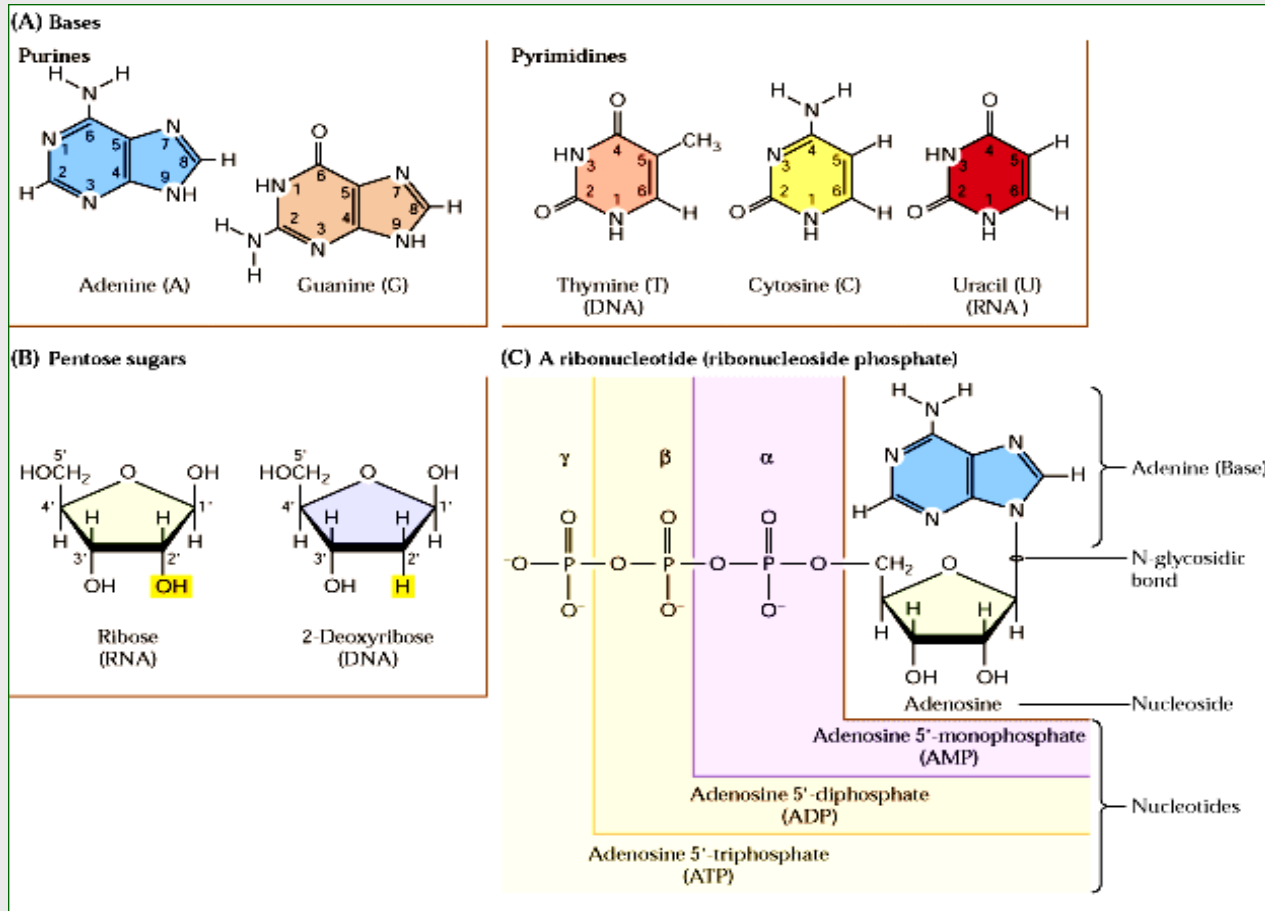
Milestone



Watson und Crick 1952/53
Auflösung der DNA-Struktur
Doppelhelix-Modell



Aufbau der DNA



Die millionenhafte Wiederholung dieser chemisch wenig variablen Bausteine machte eine Sequenzierung zur Herausforderung.



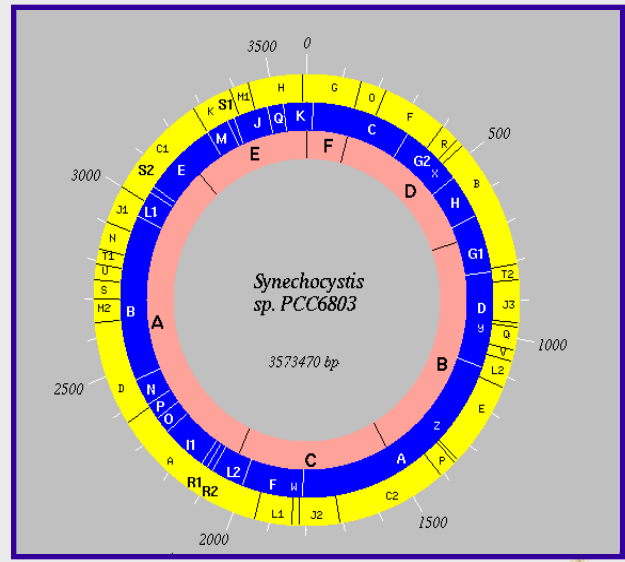
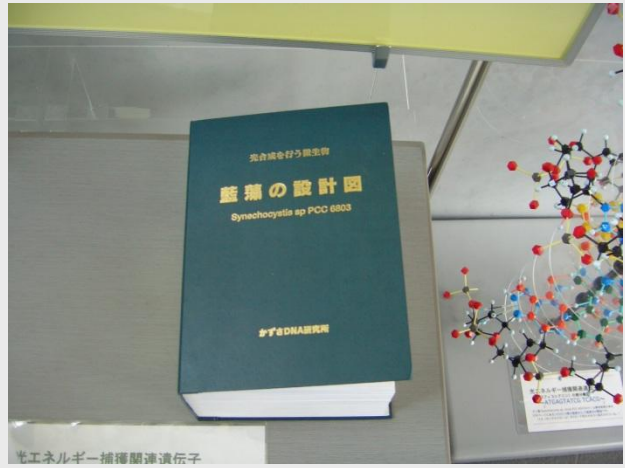
Genomsequenzierung ist heute eine Industrie

Bis heute (11/2015) wurden/werden Genome von 2.099 Archaea, 231.972 Bacteria, & 14.637 Eukaryoten sequenziert, die wenigsten sind komplett!

<http://www.genomesonline.org/>



D/J Meeting in Kazusa 2005



<http://genome.kazusa.or.jp/cyanobase/>

Humangenomprojekt – HUGO vs. Craig Venter



Die vollständige Sequenzierung des *Arabidopsis* Genoms

The Arabidopsis Genome Initiative

Nature, 2000

articles



Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*

The Arabidopsis Genome Initiative*

* Authorship of this paper should be cited as 'The Arabidopsis Genome Initiative'. A full list of contributors appears at the end of this paper

The flowering plant *Arabidopsis thaliana* is an important model system for identifying genes and determining their functions. Here we report the analysis of the genomic sequence of *Arabidopsis*. The sequenced regions cover 115.4 megabases of the 125-megabase genome and extend into centromeric regions. The evolution of *Arabidopsis* involved a whole-genome duplication, followed by subsequent gene loss and extensive local gene duplications, giving rise to a dynamic genome enriched by lateral gene transfer from a cyanobacterial-like ancestor of the plastid. The genome contains 25,498 genes encoding proteins from 11,000 families, similar to the functional diversity of *Drosophila* and *Caenorhabditis elegans*— the other sequenced multicellular eukaryotes. *Arabidopsis* has many families of new proteins but also lacks several common protein families, indicating that the sets of common proteins have undergone differential expansion and contraction in the three multicellular eukaryotes. This is the first complete genome sequence of a plant and provides the foundations for more comprehensive comparison of conserved processes in all eukaryotes, identifying a wide range of plant-specific gene functions and establishing rapid systematic ways to identify genes for crop improvement.

Pflanzenphysiologie

PUR

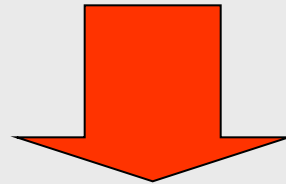
Universität Rostock



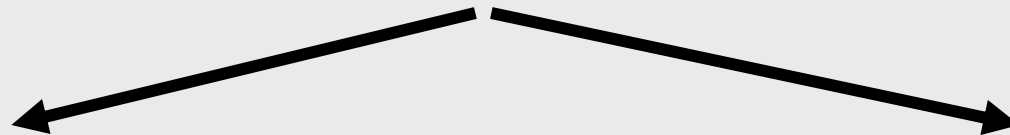
Die klassischen Sequenzierverfahren



DNA-Sequenzierung



In den 70-iger Jahren wurden zwei grundsätzliche Methoden zur Sequenzierung entwickelt



Kettenabbruch-Methode
F.Sanger und
A.R.Coulson in
Großbritannien

**Chemische Abbau-
Methode** A.Maxam und
W.Gilbert in USA



Das Maxam-Gilbert-Verfahren- beruht auf einem chemischen Abbau der DNA

Besonderheiten:

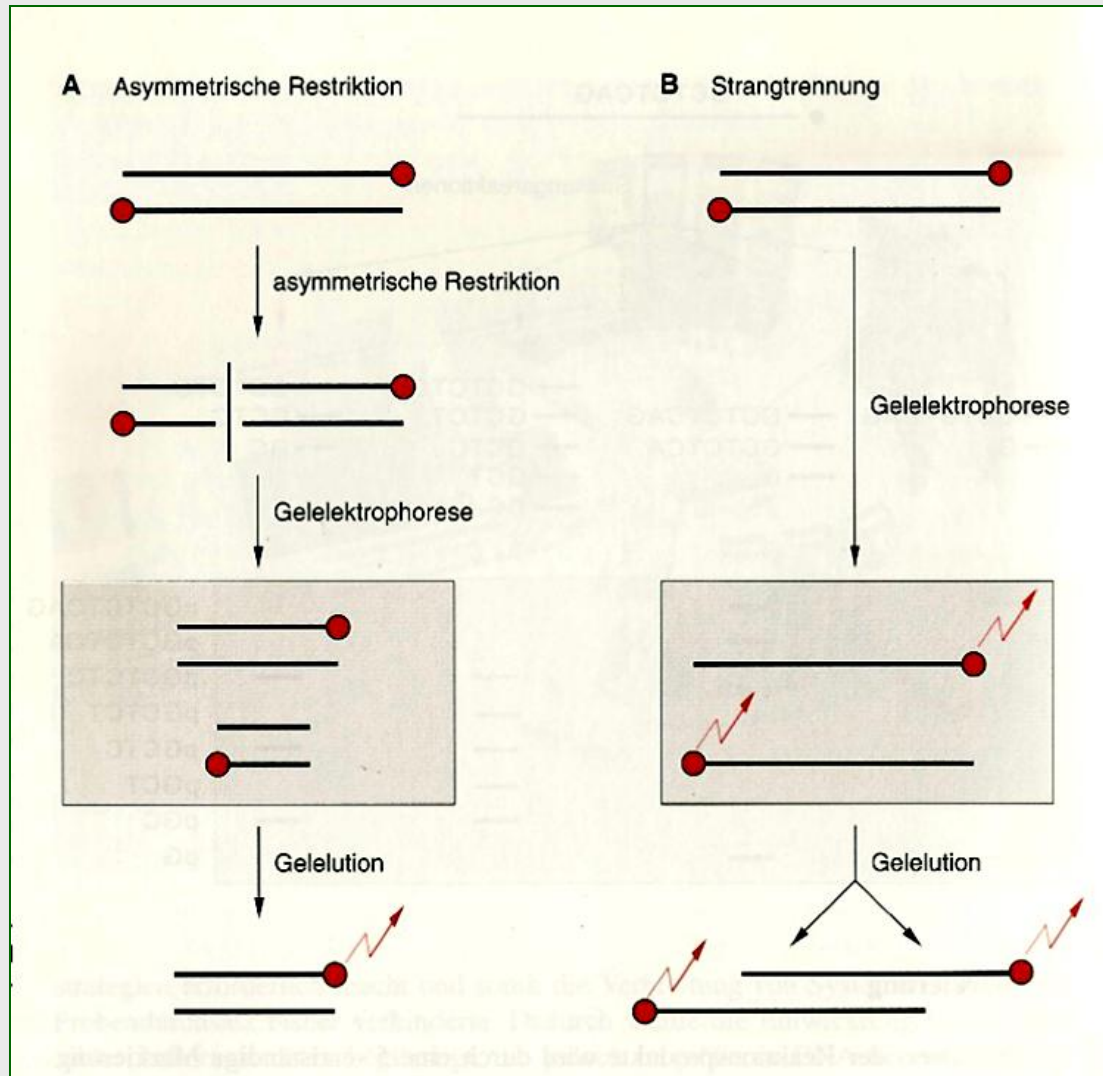
- Doppelsträngige DNA-Fragmente
- Wird ohne einen Primer durchgeführt
- Prinzip beruht auf der chemischen Spaltung eines bereits vorhandenen DNA-Moleküls, die spezifisch an Nucleotiden erfolgt.

Prinzip:

- 1. Endständige DNA-Markierung, z.B. durch radioaktives γ -ATP eines doppelsträngigen DNA-Moleküls.
- 2. Partielle chemische Spaltung der DNA in vier unabhängigen Reaktionen
- 3. Gelelektrophoretische Trennung der Spaltprodukte/ Analyse der Sequenz



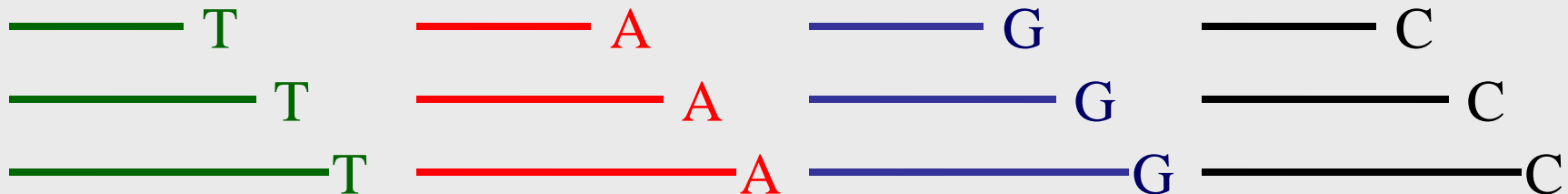
Isolierung an einem Ende markierter DNA-Fragmente



Maxam-Gilbert-Verfahren – Basen-spez. Spaltung

Spezifische chemische Spaltung der markierten DNA nach deren vorheriger Basen-“spezifischer“ Derivatisierung, partielle Reaktion!!!

A: Spaltung in vier spezifischen Reaktionen (G-Reaktion, A-Reaktion, T-Reaktion und C-Reaktion) z.B. mit Dimethylsulfat und Behandlung mit heißem Piperidin oder Hydrazin mit anschließender Piperidin-Behandlung führt zu Spaltprodukten mit jeweils einem endständigem A, T, C oder G



Sequenz 5'-³²P-GCTACGTA-3'

Spaltung bei A: ³²P- GCT + ACGTA + A

³²P- GCTACGT + A

Spaltung bei G: ³²P- GCTAC + GTA

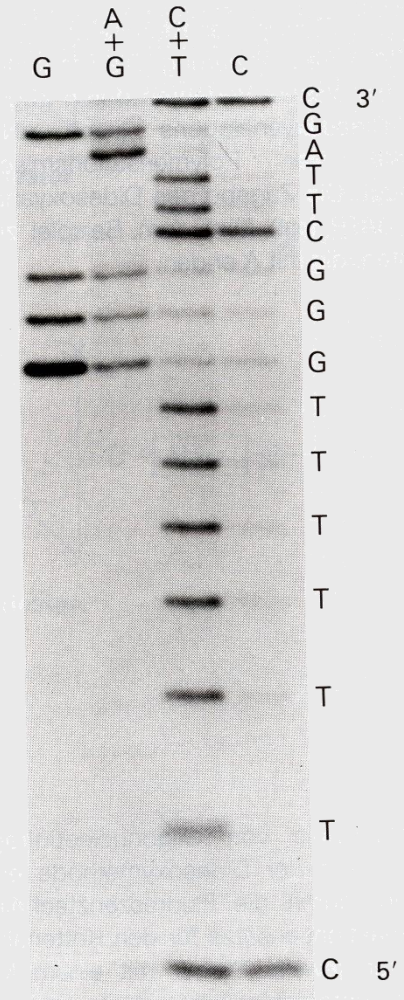
Spaltung bei C: ³²P- G + CTAC + GTA

³²P- GCTA + CGTA

Spaltung bei T: ³²P- GC + TACGTA

³²P- GCTACG + TA

5' CTTTTTTTGGGCTTAGC-3'

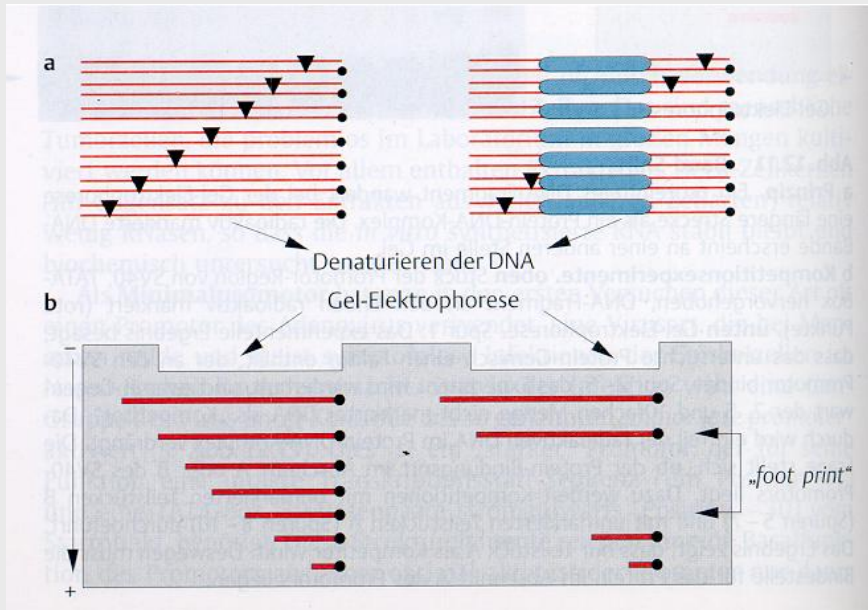


Wird heute z.T. noch für „foot-print“ Analysen genutzt!

Extrem gut bei GC-reichen Sequenzen!



Einschub – „Footprint“analysen zur Bindung von Proteinen an definierte DNA-Sequenzmodife



Endmarkierte DNA-Stücke mit einer vermuteten Proteinbindestelle werden mit und ohne hypothetischem Bindeprotein inkubiert – durch Nachbehandlung wird der durch Proteinbindung geschützte Bereich direkt sichtbar!

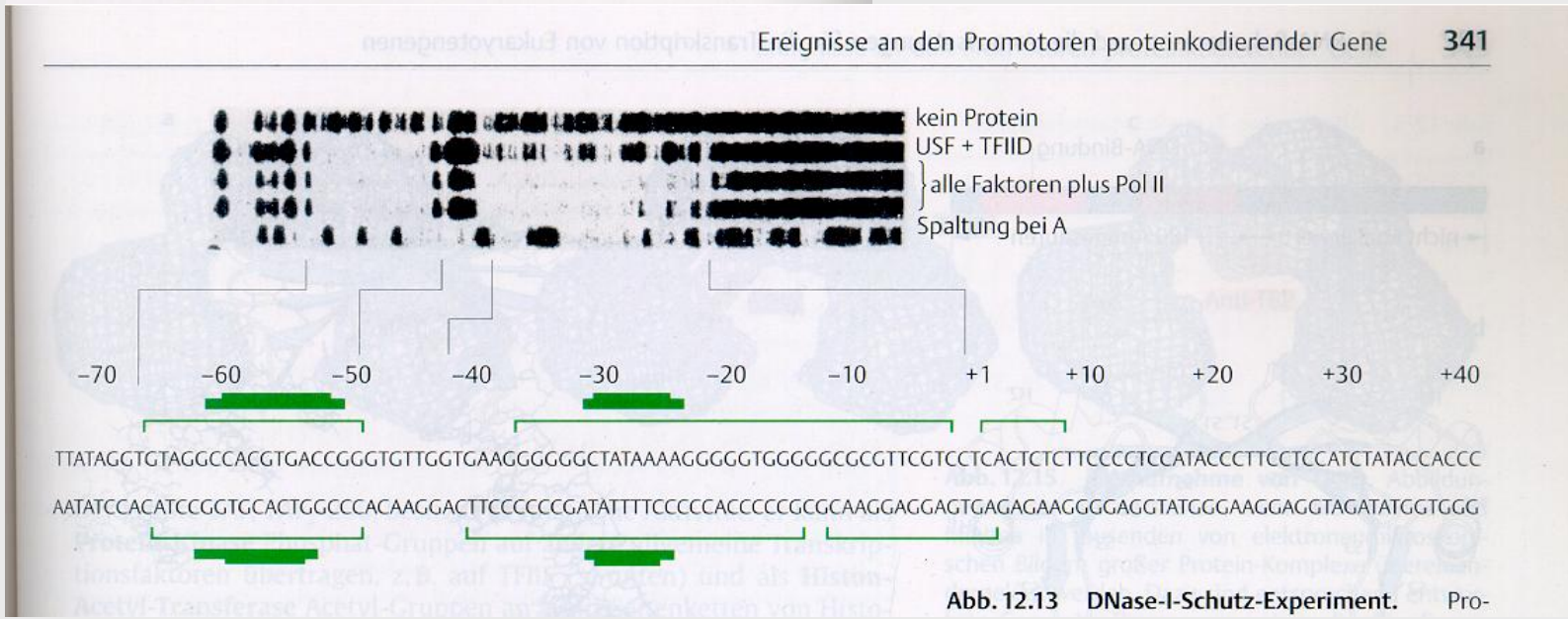
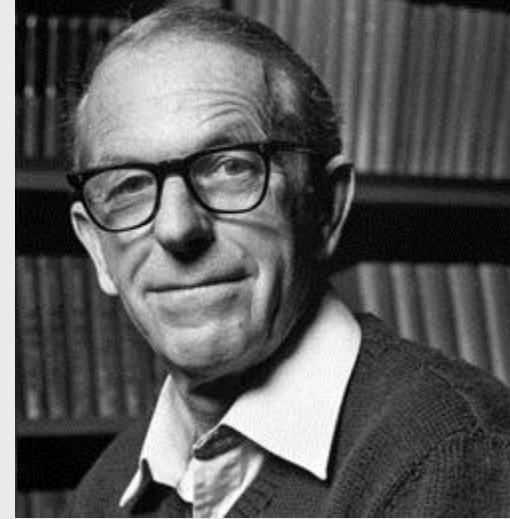


Abb. 12.13 DNase-I-Schutz-Experiment. Pro-

Das Sanger-Coulson Verfahren zur Sequenzierung

Heute das noch dominierende Verfahren für Einzelsequenzen!

Sanger-, M13-, Didesoxy- etc. Sequenzierung



An einer einzelsträngigen DNA-Matrize wird von einem definierten Startpunkt (Primer) ausgehend der komplementäre Strang durch eine DNA-Polymerase synthetisiert.

Der Syntheseprozess bricht an jeder möglichen Position mindestens einmal ab, wobei das Ende des Abbruchs bekannt ist, so dass die Sequenz rekonstruiert werden kann.



Das Sanger-Coulson Verfahren zur Sequenzierung nach der Kettenabbruch-Methode

Für dieses Verfahren wird einzelsträngige DNA benötigt



Klonierung des zu sequenzierenden DNA-Moleküls in einen M13-Vektor bzw. pUC-Plasmid (Denaturierung)



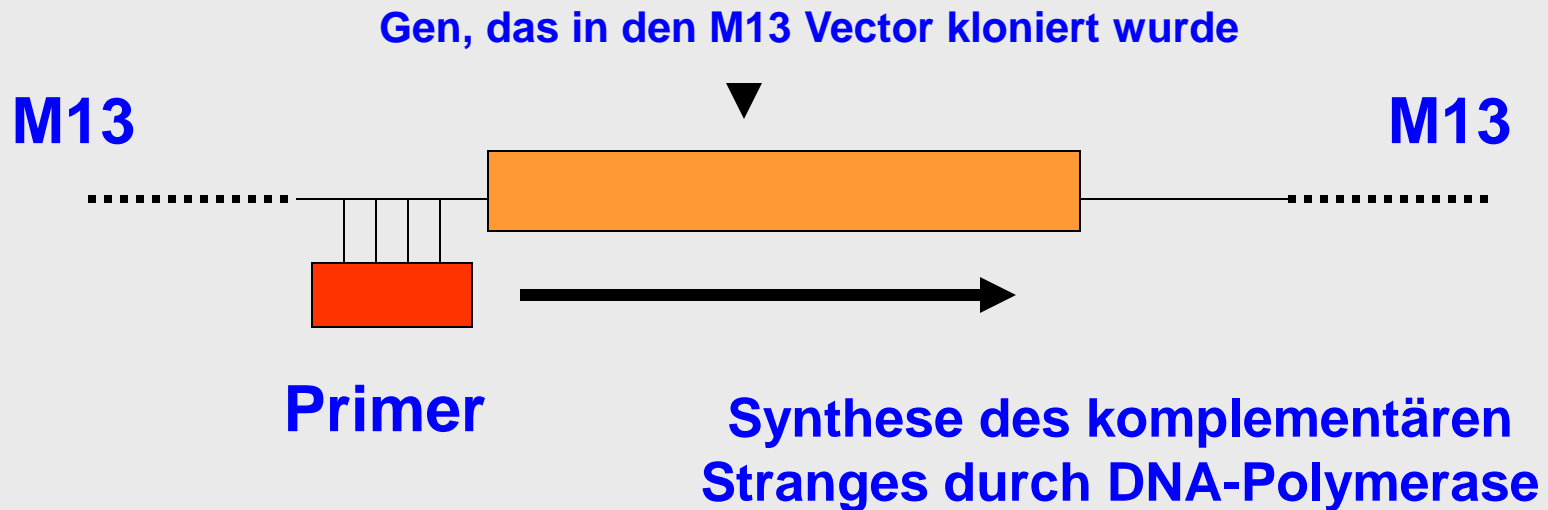
Enzymatische Synthese des komplementären Stranges von der vorhandenen Matrize durch Einsatz z.B. des Klenow Fragmentes der DNA-Polymerase



Hierzu ist es notwendig, einen doppelsträngigen Bereich zu erzeugen, der als Startpunkt der DNA-Polymerase dient. Das wird durch Hybridisierung eines kurzen Oligonucleotides (Primer) erreicht, der als Ausgangspunkt für die Synthese des komplementären Stranges dient



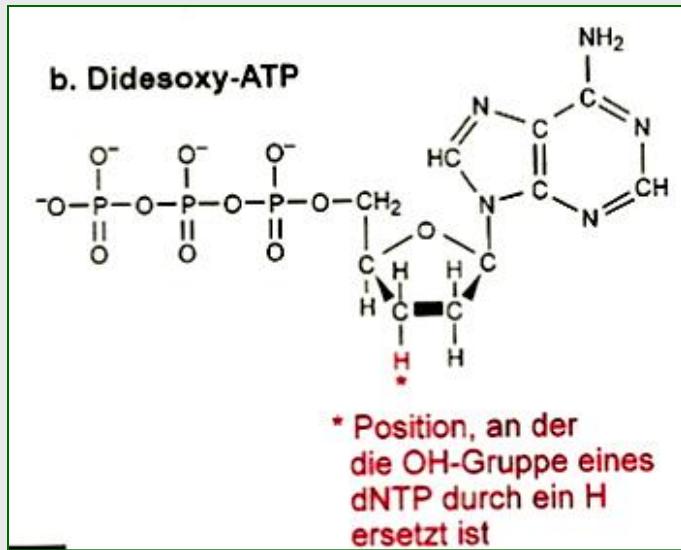
1. Schritt des Sanger Verfahrens: Anlagerung eines Primers



Wie kann man durch Synthese des zweiten Stranges die Sequenz der DNA ermitteln?



2. Schritt: definierter Abbruch der Synthese-Reaktion

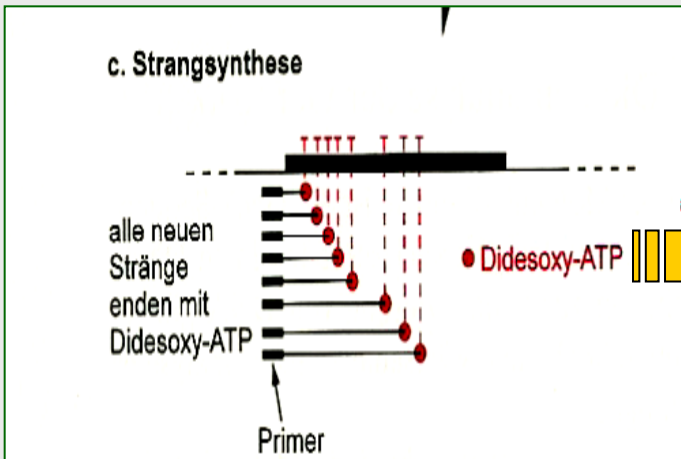


Substrate für die Synthesereaktion

DNA-Polymerase + dATP, dTTP, dGTP, dCTP

Zusatz von Dideoxynucleotiden führt zum Kettenabbruch, da OH-Gruppe in der 3'-Position fehlt, die für das Anheften des nächsten Nucleotids notwendig ist

Wird z.B. dem Reaktionsansatz dd-ATP hinzugesetzt, erhält man einen Kettenabbruch an allen Stellen, an der sich in der Matrize ein Thymidin befindet (A-T).

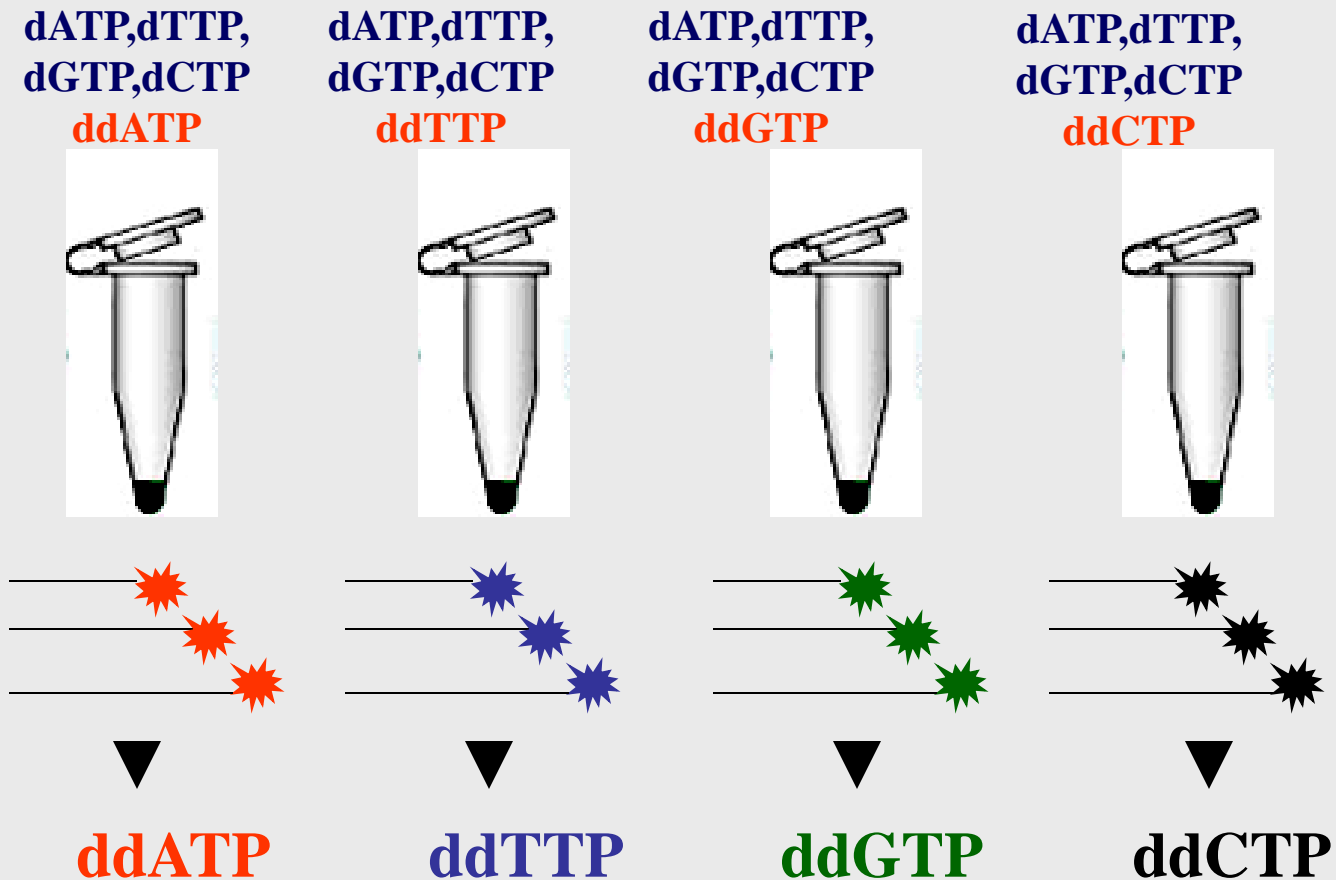


Resultat

Populationen von unterschiedlich langen DNA-Strängen, die alle mit einem ddATP enden.



Um alle vier Nucleotidarten zu erhalten, wird die Synthesereaktion in vier parallel ablaufenden Ansätzen durchgeführt



Vier Populationen neusynthetisierter Oligonucleotide unterschiedlicher Länge, die jeweils mit ddATP, ddTTP, ddCTP bzw. ddGTP enden und sich nur um ein Nucleotid unterscheiden

3. Schritt: Gelelektrophoretische Auftrennung der jeweiligen Ansätze



groß



klein

DNA-Fragmente

- Polyacrylamidgele, hauchdünn 0,5 mm
- Enthalten Harnstoff-Denaturierung der DNA
- Hochspannung, 60°C, verhindert Renaturierung

- Jede Bande im Gel entspricht einem DNA Fragment, das mit dem entsprechenden dd-NTP endet und sich jeweils um ein Nucleotid unterscheidet.

- Da hier sehr wenig DNA in jeder Bande enthalten ist, werden die aufgetrennten DNA-Fragmente mit Hilfe der Autoradiographie sichtbar gemacht. Das erfolgt durch Zusatz von ^{35}S - oder ^{32}P -dATP

Wie wird die Sequenz aus dem Autoradiogramm abgelesen?



Welche Bande ist am weitesten gewandert?

Diese Bande repräsentiert das kleinste DNA - Fragment, also dasjenige, das durch den Einbau des Didesoxinukleotides (hier ddATP) an der ersten Position der Matrize abgebrochen ist (200 bp waren im Durchschnitt lesbar).

In welcher Spur beginnt die Sequenz?

ddNTP Sequenz

→ ddG → ATTGCGAT

→ ddG → ATTGCGA

→ ddG → ATTGCG

→ ddC → ATTGC

→ ddG → ATTG

→ ddT → ATT

→ ddT → AT

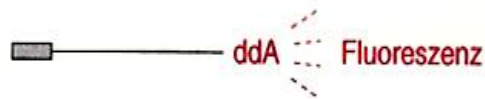
→ ddA → A

„Frühe Automaten“ nutzten Fluoreszenzmarkierte Primer – 1000 bp lesbar.

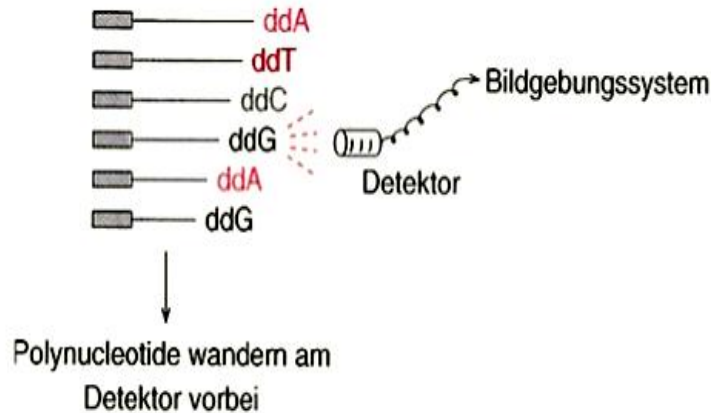


Automatische DNA-Sequenzierung

a. Ein fluoreszenzmarkiertes Polynucleotid mit abgebrochener Kette



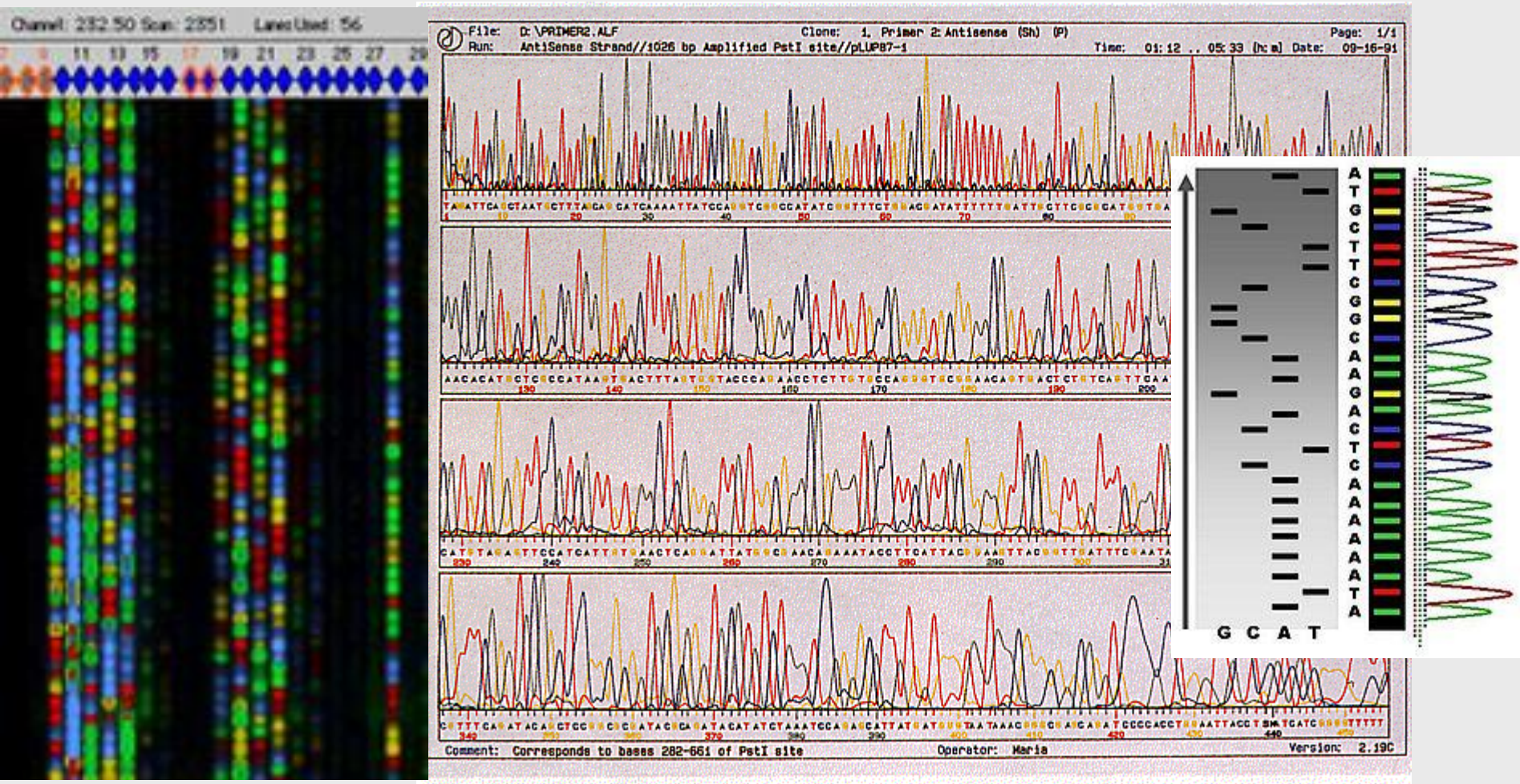
b. Nachweis von Polynucleotiden mit abgebrochener Kette



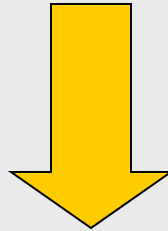
- Prinzip des Sanger Verfahrens
- Fluoreszenzmarkierung
- Jedes ddNTP wird mit einem anderen Fluorozenzfarbstoff markiert
- keine getrennten Ansätze
- Trennung durch Kapillarelektrophorese
- Ca. 500-1300 bp lesbar
- Computer gestützte Bildgebungssysteme



Automatische DNA-Sequenzierung - Auswertung



Kann die PCR-Reaktion mit der Sequenzierungsreaktion gekoppelt werden?



thermal cycle sequencing

oder

zyklische Sequenzierung

Merkmale:

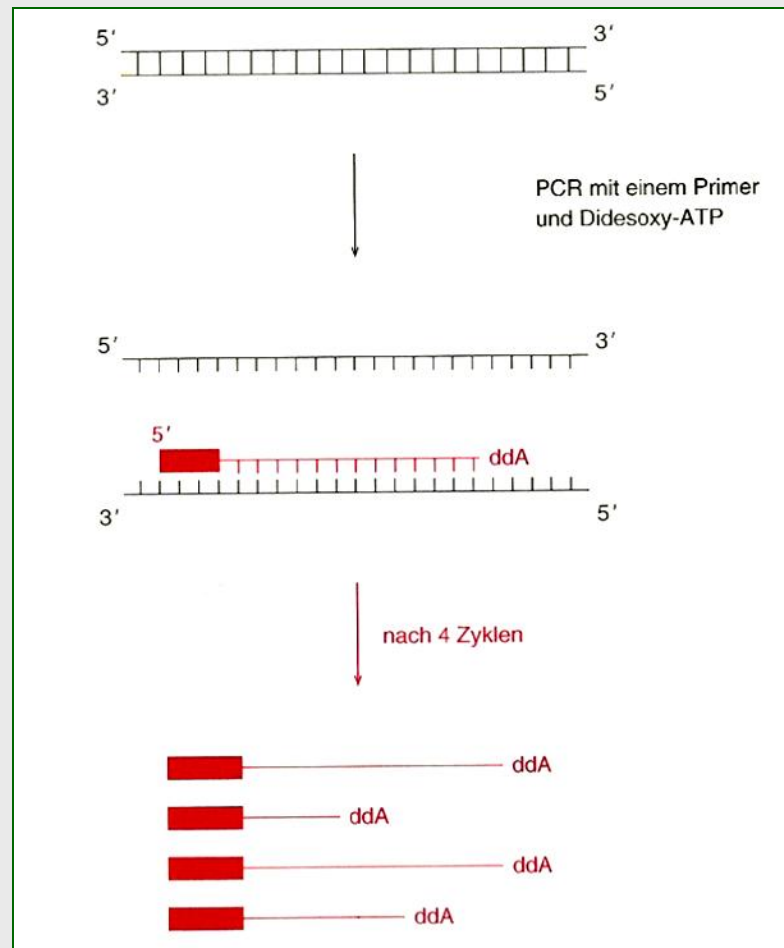
PCR-Reaktion nur mit einem Primer, d.h. es wird nur ein Strang kopiert

Aber in vielen Wiederholungen (Anzahl der Zyklen der PCR-Reaktion)

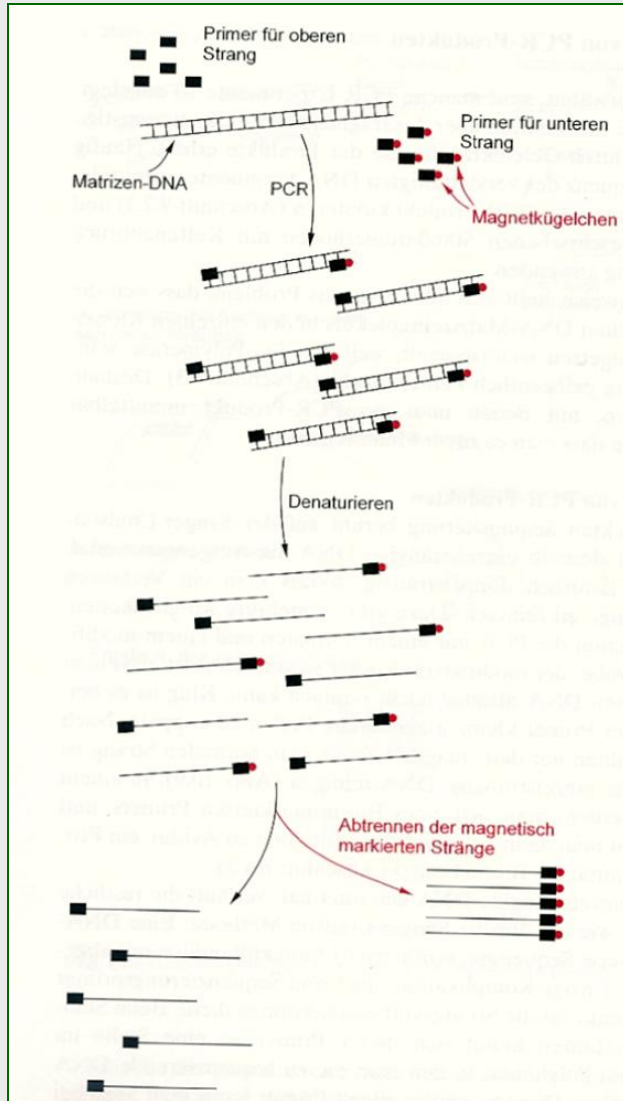
Vier parallel – Reaktionen immer unter Zusatz der dNTP's + einem ddNTP



Zyklische Sequenzierung Kopplung von PCR-Reaktion und DNA Sequenzierung



Direkte Sequenzierung von PCR-Produkten



- Beruht auf dem Sanger-Verfahren
- Einzelsträngige DNA
- Doppelsträngiges PCR-Produkt muss in Einzelstränge getrennt werden
- Primer A ohne Modifikation
- Primer B + magnetische Partikel
- oder Biotin markiert (Avidin)
- Denaturierung des PCR-Produktes
- Trennung des modifizierten Stranges vom nicht modifizierten Strang



Automatische DNA-Sequenzierung

Universität Rostock, Inst. Biowiss. – Dr. Ralf Bastrop übernimmt Sequenzierung (5 Euro)

Automat heißt nicht ohne Arbeit!!

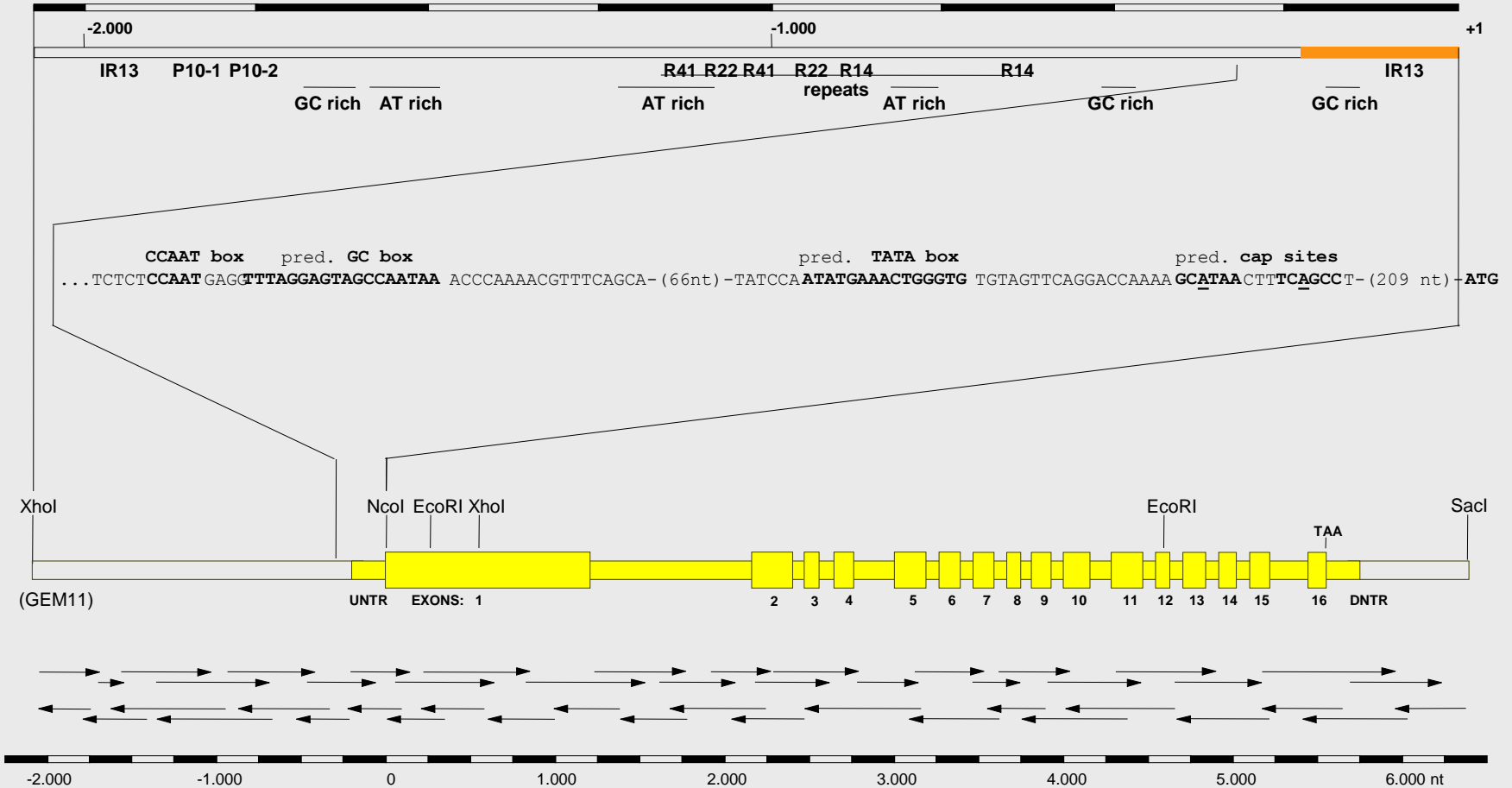
Bitte vernünftig damit umgehen:

- Sequenzen haben Fehler!!!
- Vor der Postulierung von SNPs u.ä. vertrauenswürdige Daten erheben!!!
- Fehler ausmerzen durch Kontrolle der erhaltenen Sequenzauswertung
- Fehler ausmerzen durch Sequenzierung beider Stränge
- Fehler ausmerzen durch Sequenzierung mehrerer Klone
- Abschneiden von Primersequenzen

Diese Sequenzierung kann pro „Automat“ im Jahr maximal 350 Mio. bp Sequenz generieren, allerdings mit sehr guter Qualität. Die neuen Verfahren bringen das am Tag.

Sequenzierungsstrategien

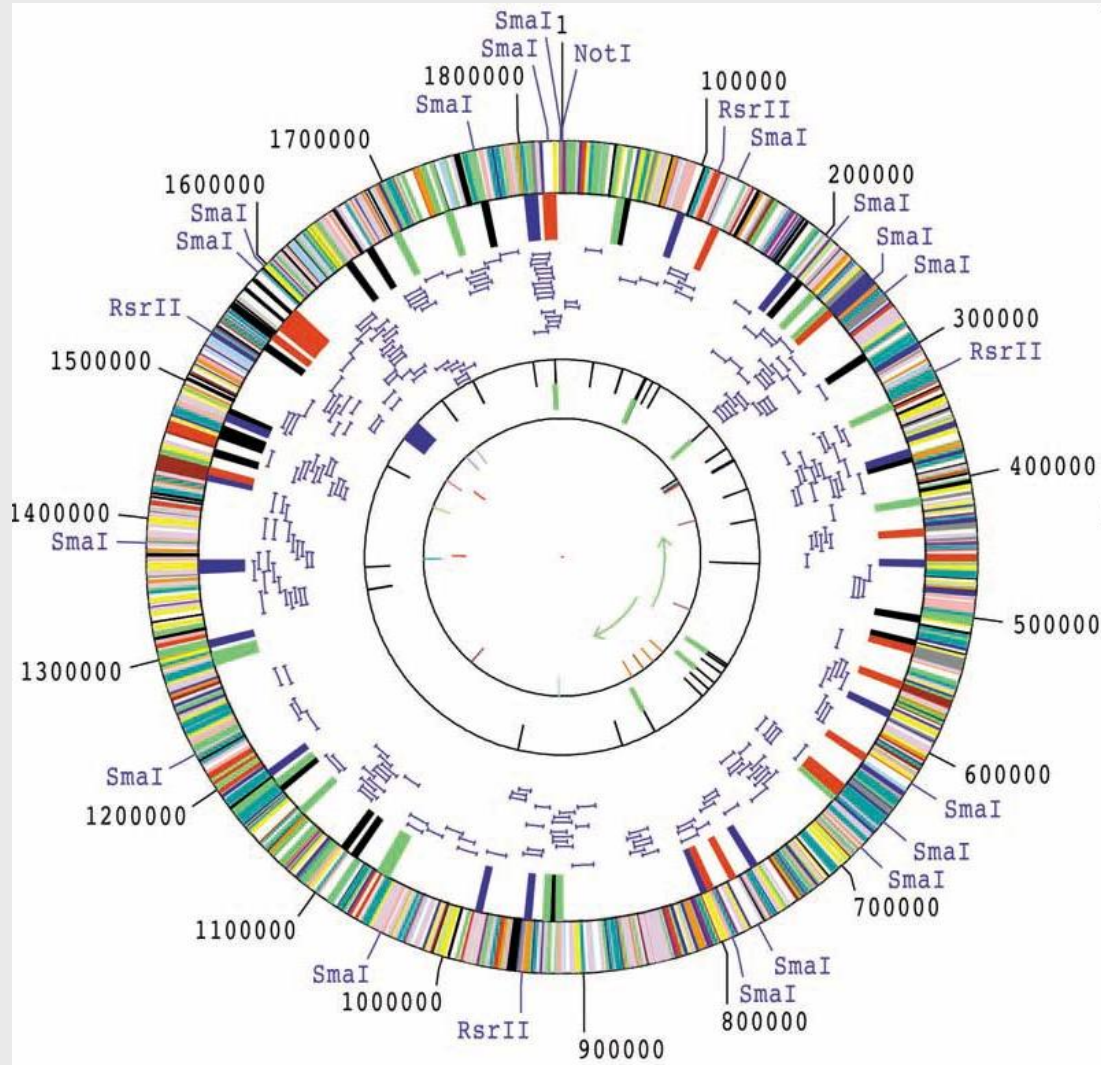
Schrittweise Sequenzierung des Inserts nach Subklonierung oder „chromosome walking“ es entsteht eine kontinuierliche Sequenz - **contig**



Sequenzierungsstrategien

Shot-gun Sequenzierung von Zufallsfragmenten (Venter)

Zerlegen großer DNA-Fragmente in viele kleine Fragmente (partieller Verdau, mechanisch), Klonierung in pUC-Vektoren, Sequenzierung, Computer (coverage – 6-10mal, entstehen viele „Contigs“, trotzdem Lücken!!)



1995

Craig Venter, Hamilton Smith, Claire Fraser, and colleagues at TIGR elucidate the first complete genome sequence of a microorganism - *Haemophilus influenzae* Rd.
1.830.137 bp

Since that time, the genome sequencing was mainly done using this strategy. (The Institute for Genome Research - TIGR)

Humangenomprojekt – HUGO vs. Craig Venter





Pflanzenphysiologie

PUR

Universität Rostock



Sequenzierungsstrategien - Metagenomics

Shot-gun Sequenzierung von Zufallsfragmenten aus Umweltproben bzw. von Lebensgemeinschaften
bis hin zur Zusammenstellung kompletter bakterieller Genome

THE METAGENOMICS PROCESS



Extract all DNA from
microbial community in
sampled environment

DETERMINE WHAT THE GENES ARE

(Sequence-based metagenomics)

- Identify genes and metabolic pathways
- Compare to other communities
- and more...

DETERMINE WHAT THE GENES DO

(Function-based metagenomics)

- Screen to identify functions of interest, such as vitamin or antibiotic production
- Find the genes that code for functions of interest
- and more...

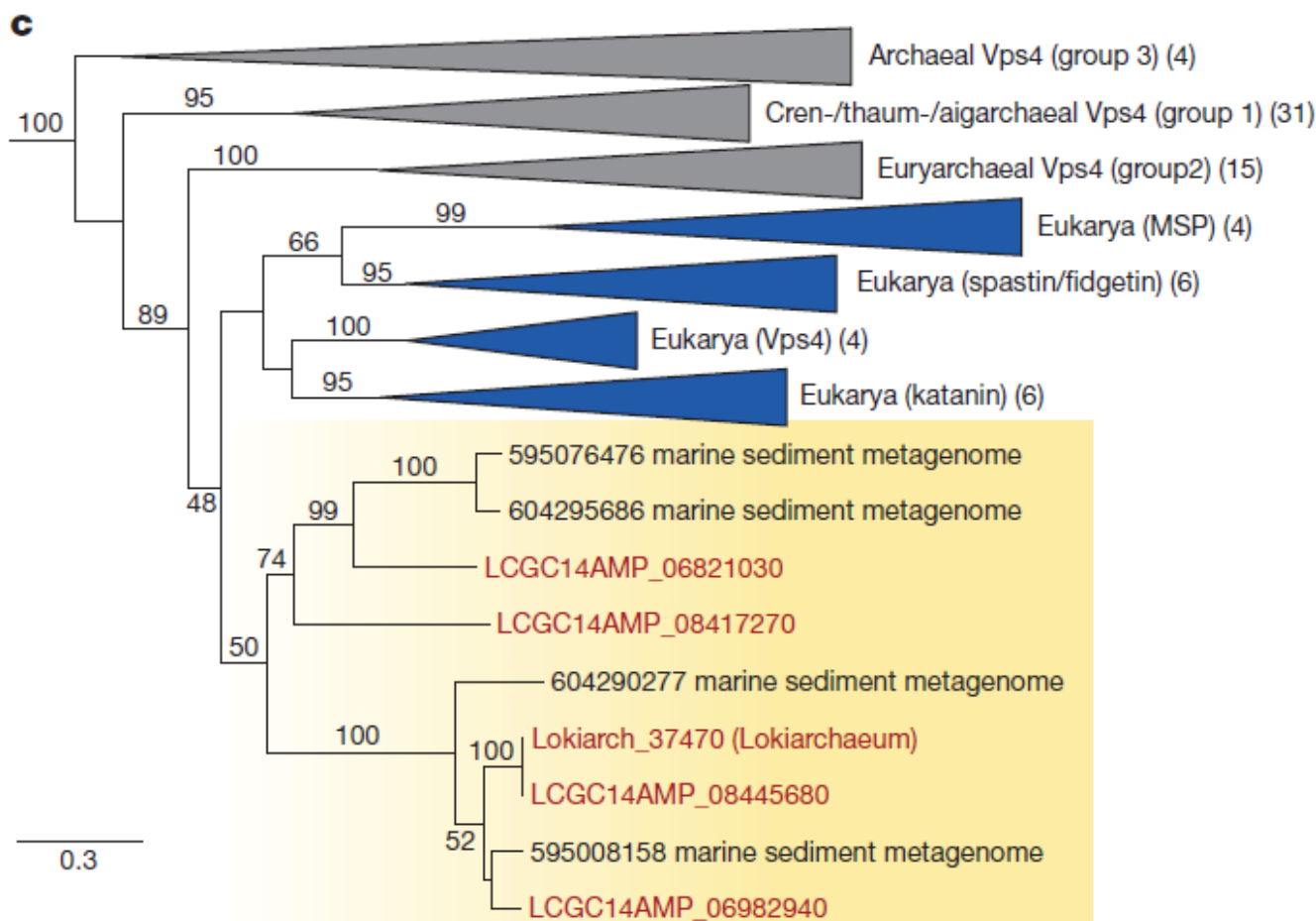
Complex archaea that bridge the gap between prokaryotes and eukaryotes

Anja Spang¹
Roel van Eijl

J. Lind¹,

The origin of eukaryotes has long been a subject of debate. Recent studies have provided the identification of 'Lokiarchaeum' in a phylogenetic tree, suggesting a relationship between which the eukaryotic 'starter-kit' genes are found.

These studies suggest that life, but discovery of prokaryotes in that are eukaryotes in underpin genomic studies.



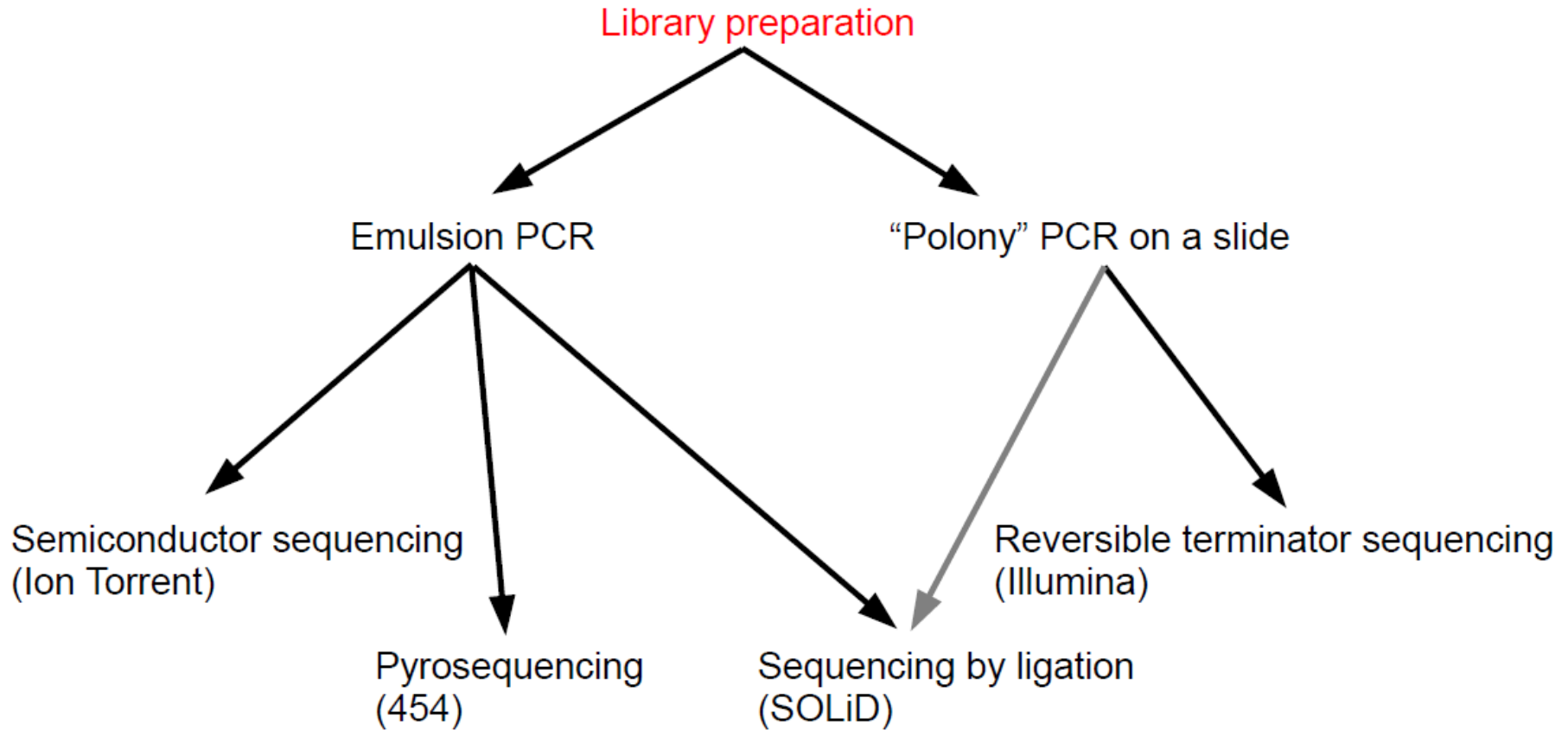
Next generation sequencing (ab 2005)

- 454 Sequencing / Roche
 - GS Junior System
 - GS FLX+ System
- Illumina (Solexa)
 - HiSeq System
 - Genome analyzer Iix
 - MySeq
- Applied Biosystems - Life Technologies
 - SOLiD 5500 System
 - SOLiD 5500xl System
- Ion Torrent - Life Technologies
 - Personal Genome Machine (PGM)
 - Proton
- Helicos
 - Helicos Genetic Analysis System
- Pacific Biosciences
 - PacBio RS
- Oxford Nanopore Technologies
 - GridION System
 - MinION

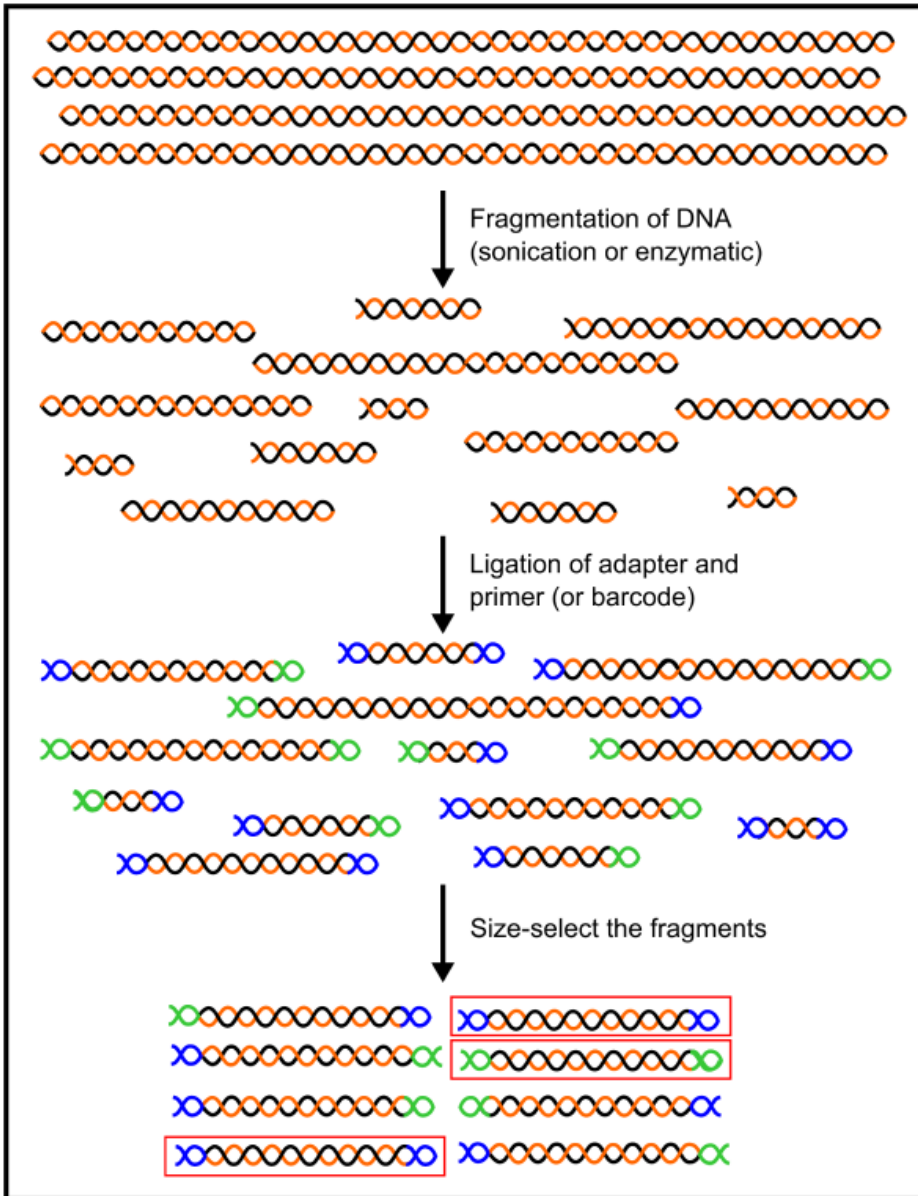
Next Generation Sequencing
Amplified Single Molecule Sequencing

Third Generation Sequencing,
Next Next Generation Sequencing,
Single Molecule Sequencing

Next Generation Sequencing : Amplified Single Molecule Sequencing

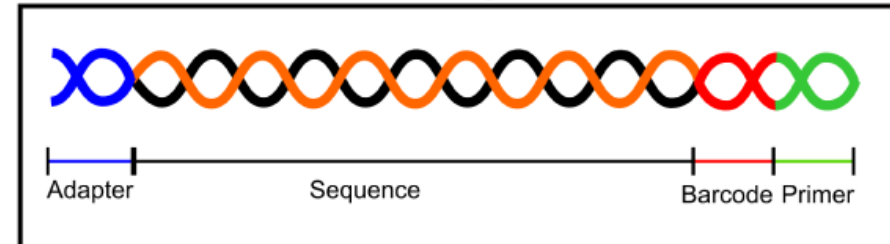


Next Generation Sequencing : Amplified Single Molecule Sequencing

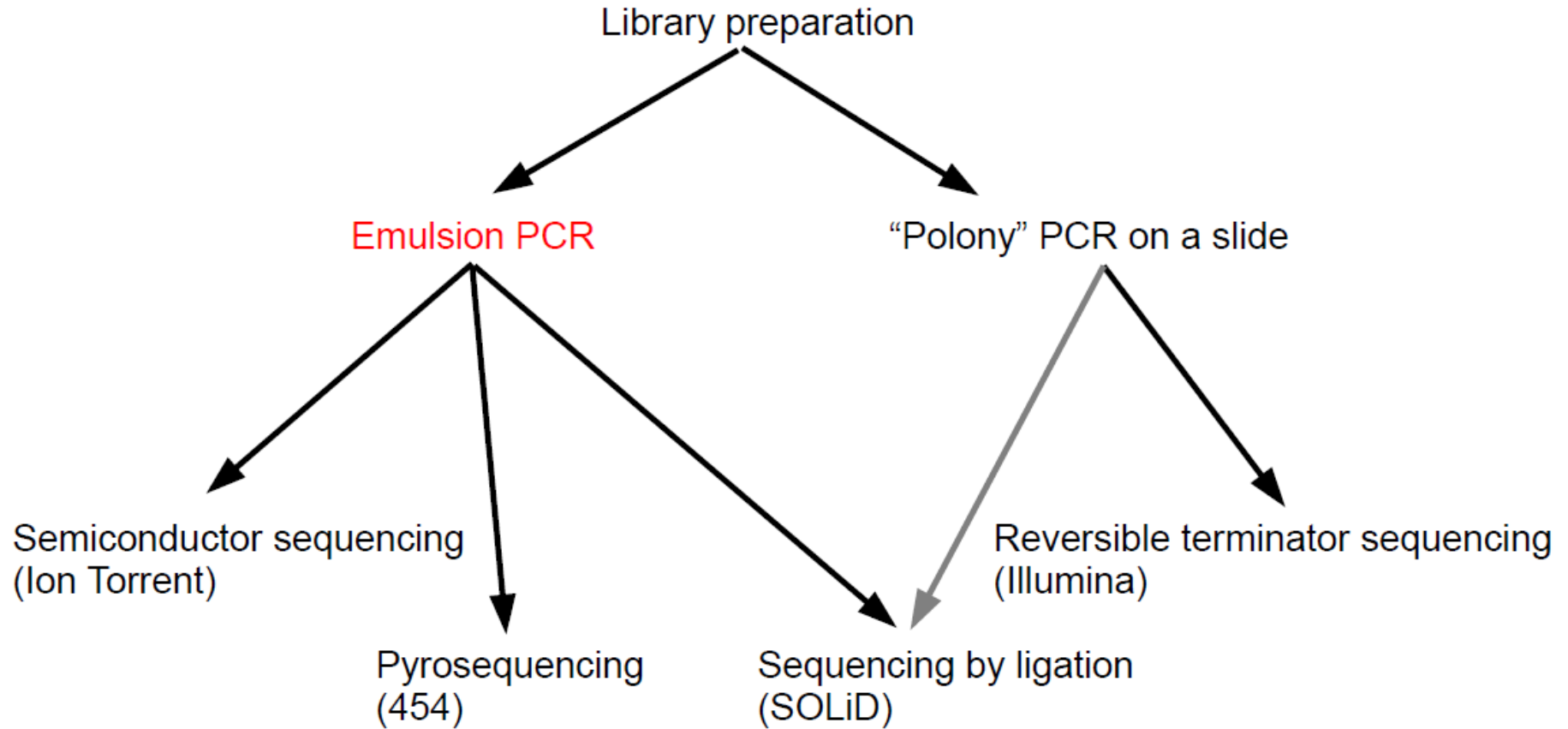


Library preparation

Good fragments :



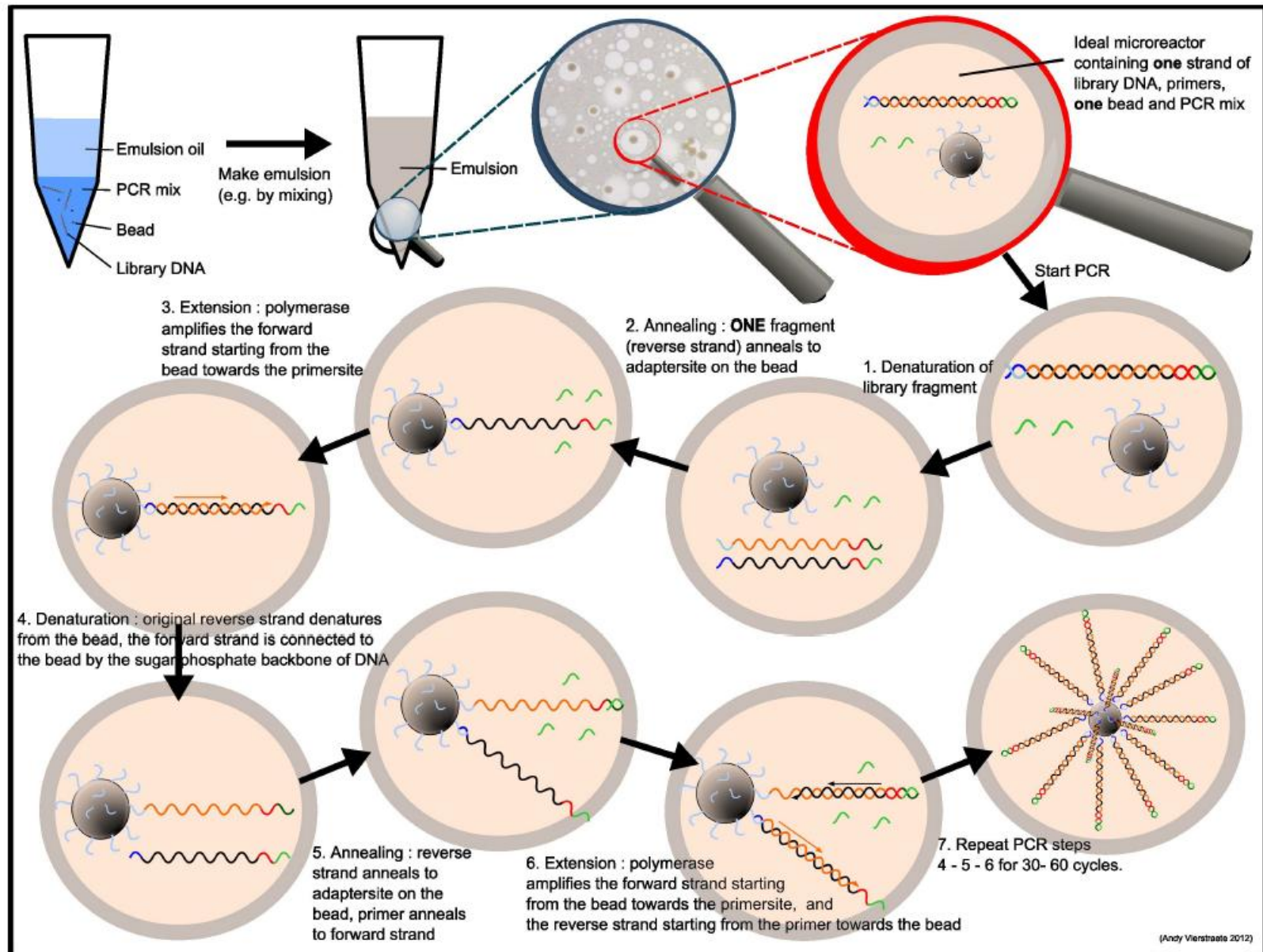
Next Generation Sequencing : Amplified Single Molecule Sequencing



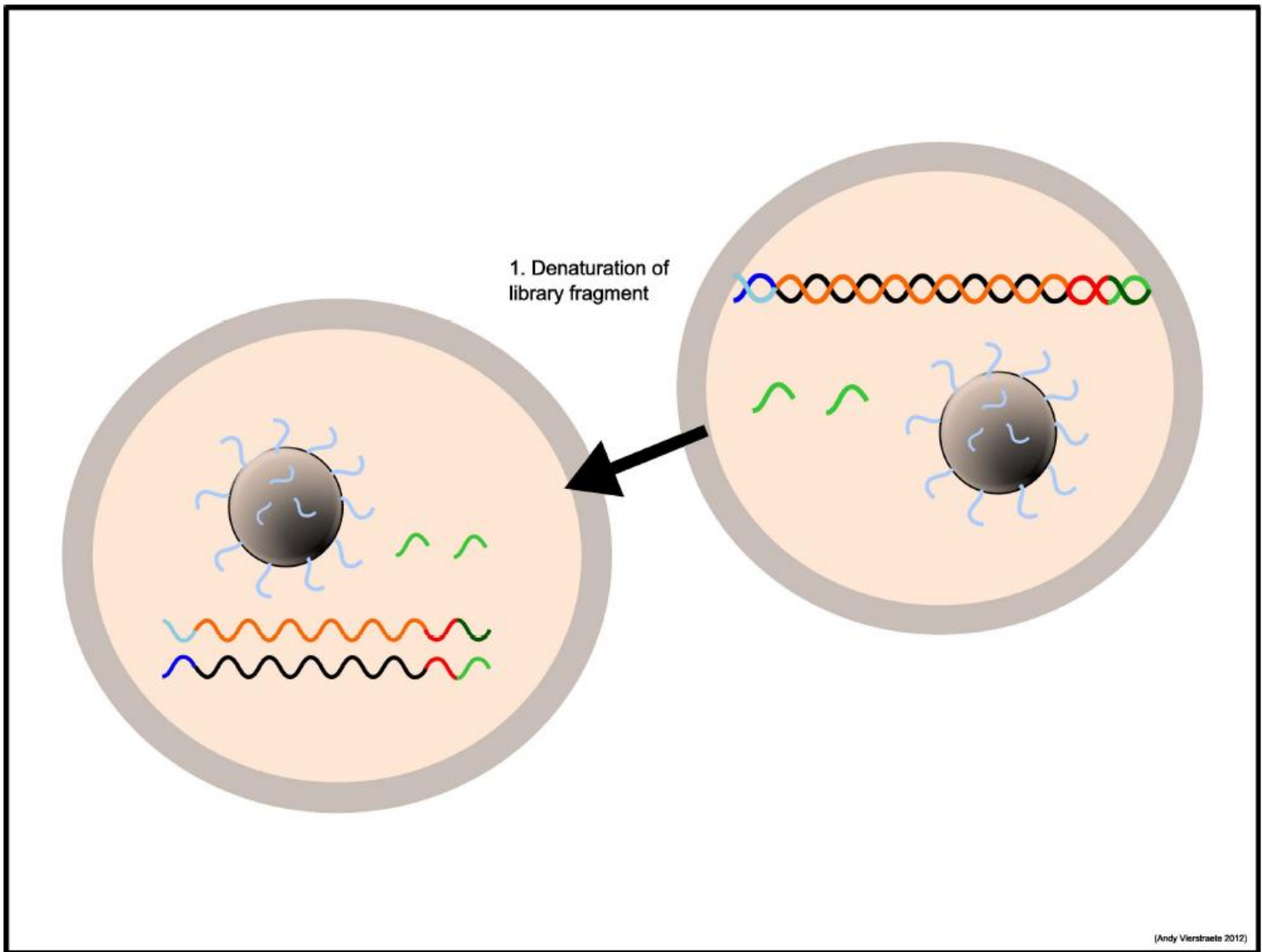
Next Generation Sequencing Workflow

Next Generation Sequencing : Amplified Single Molecule Sequencing Emulsion PCR

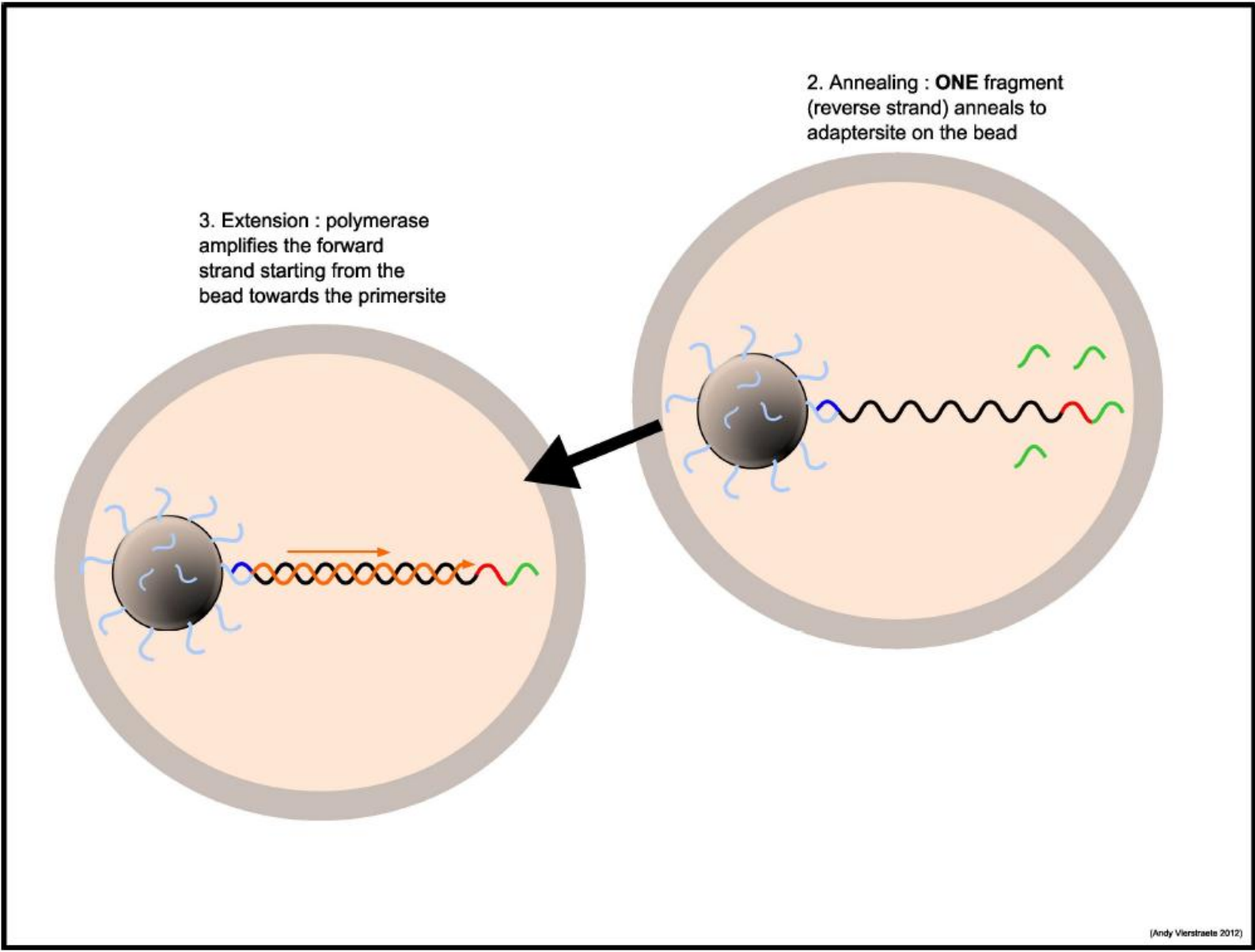
11/132



Next Generation Sequencing : Amplified Single Molecule Sequencing Emulsion PCR 12/132

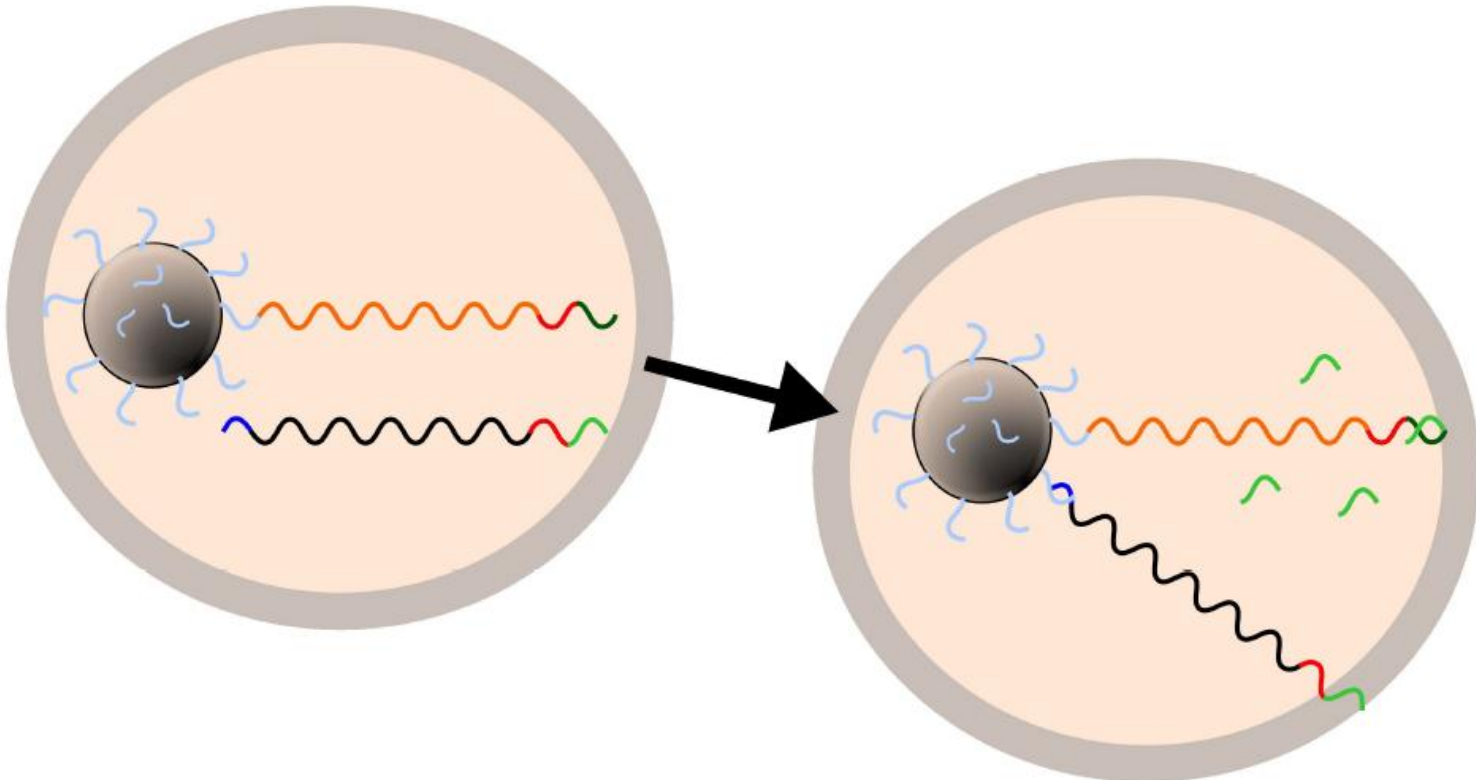


Next Generation Sequencing : Amplified Single Molecule Sequencing Emulsion PCR 13/132



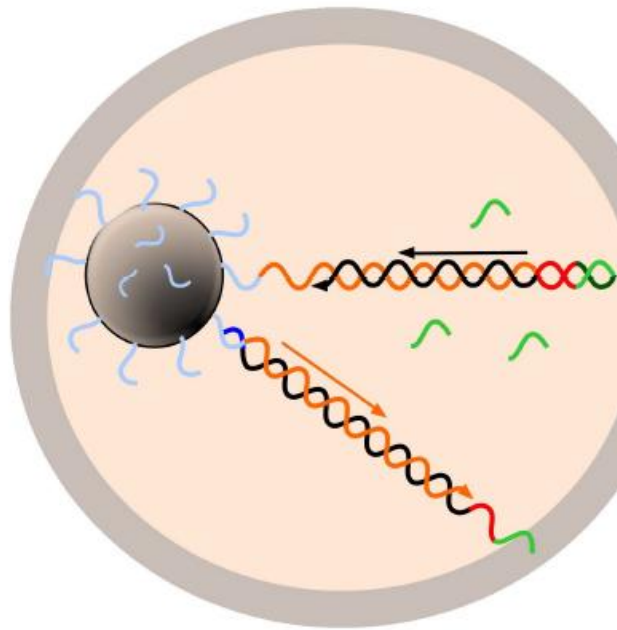
Next Generation Sequencing : Amplified Single Molecule Sequencing Emulsion PCR ^{14/132}

4. Denaturation : original reverse strand denatures from the bead, the forward strand is connected to the bead by the sugar phosphate backbone of DNA

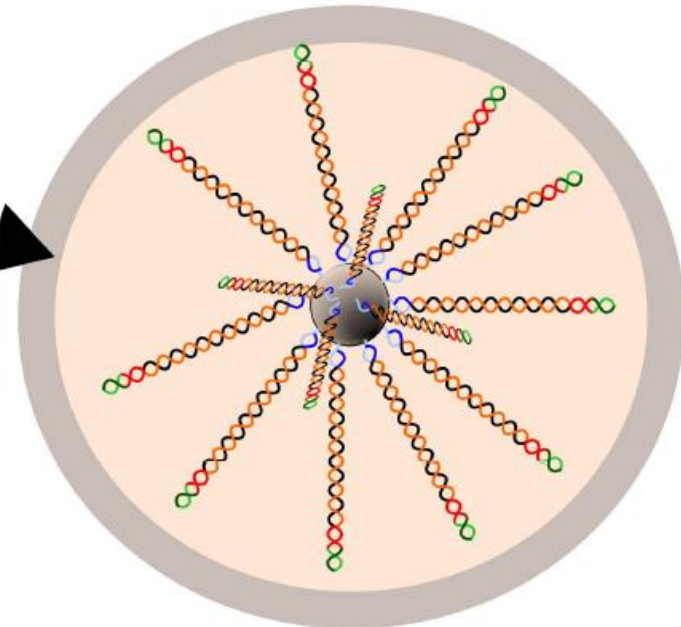


5. Annealing : reverse strand anneals to adaptersite on the bead, primer anneals to forward strand

Next Generation Sequencing : Amplified Single Molecule Sequencing Emulsion PCR ^{15/132}



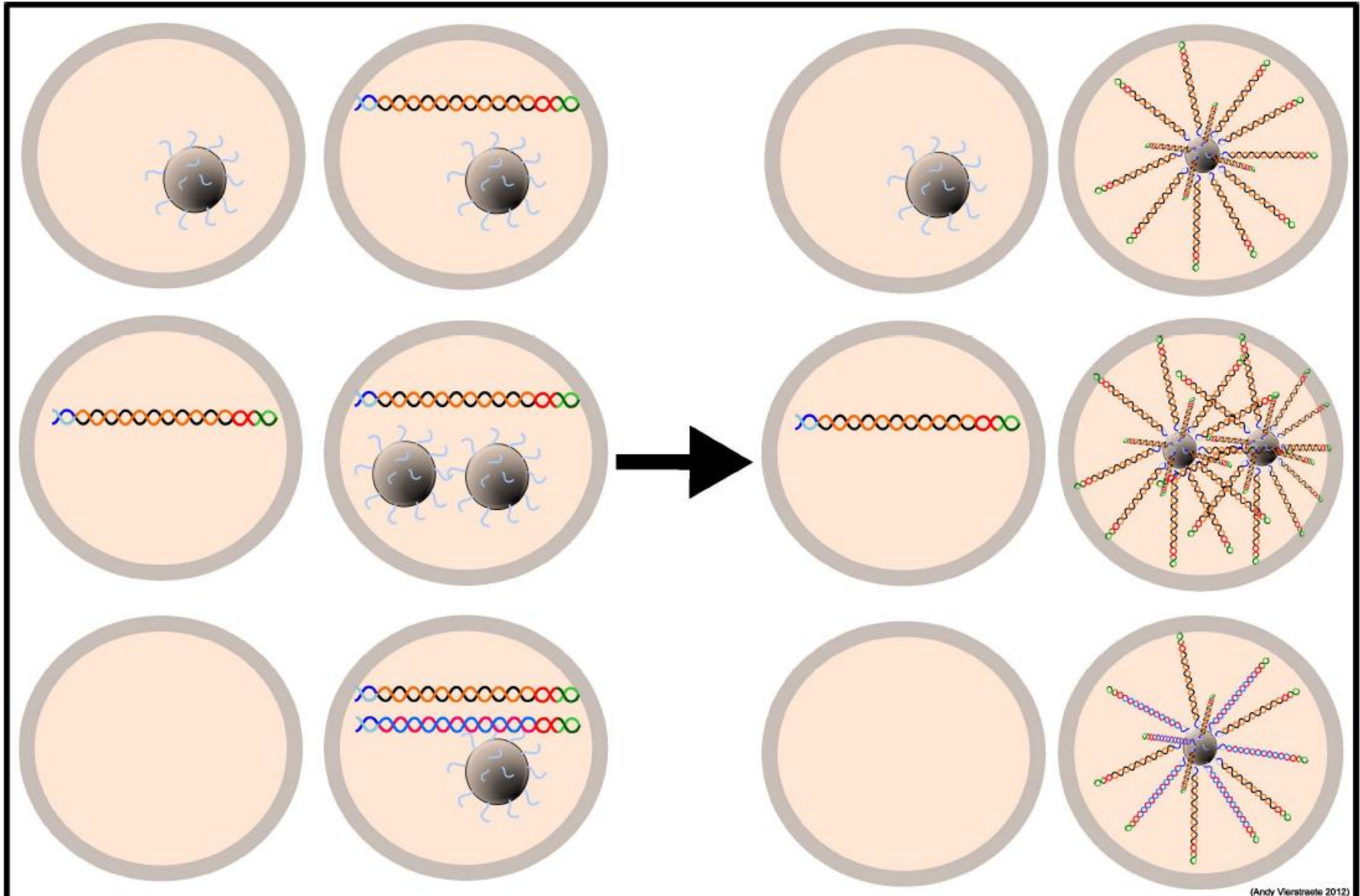
6. Extension : polymerase amplifies the forward strand starting from the bead towards the primersite, and the reverse strand starting from the primer towards the bead



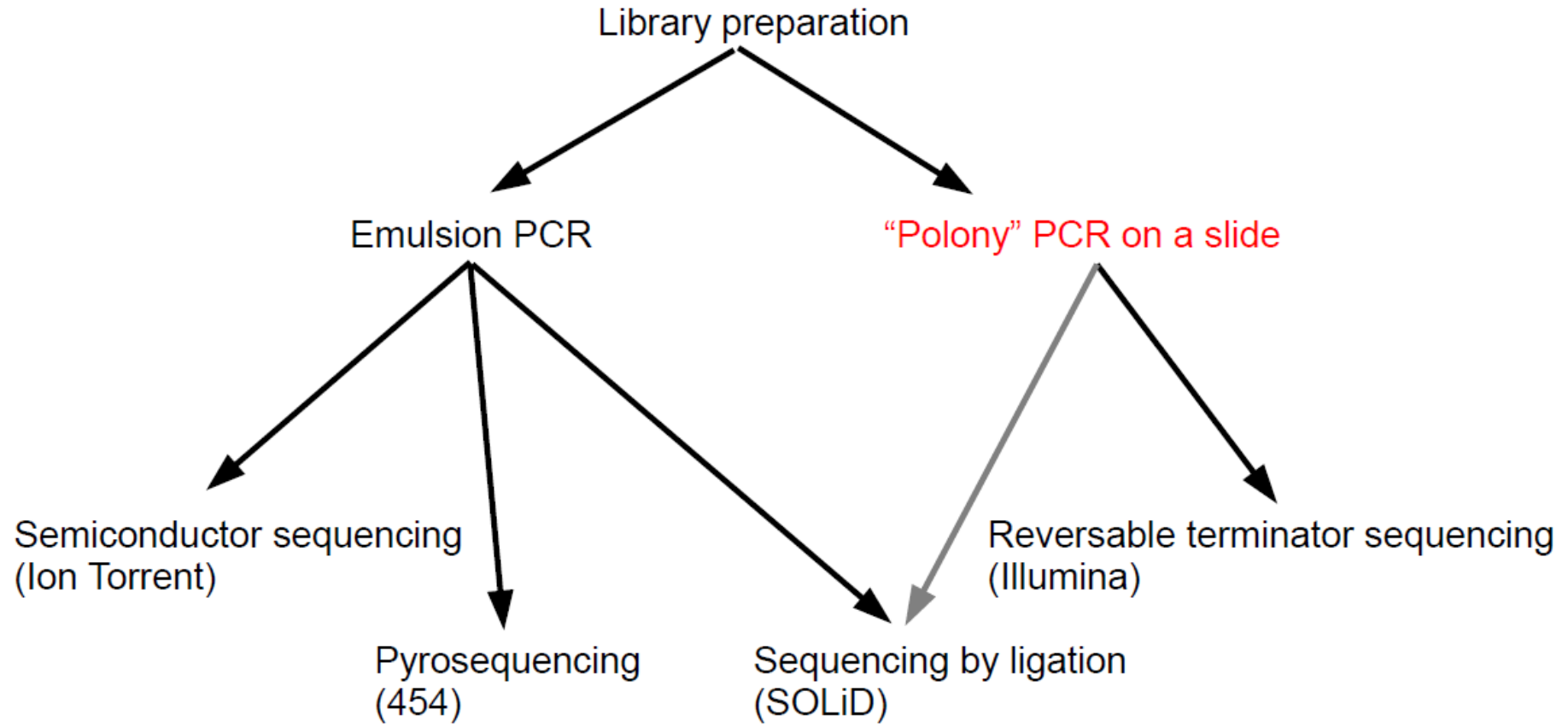
7. Repeat PCR steps
4 - 5 - 6 for 30- 60 cycles.

Next Generation Sequencing : Amplified Single Molecule Sequencing Emulsion PCR

different micro reactors : only 15 % are good ones

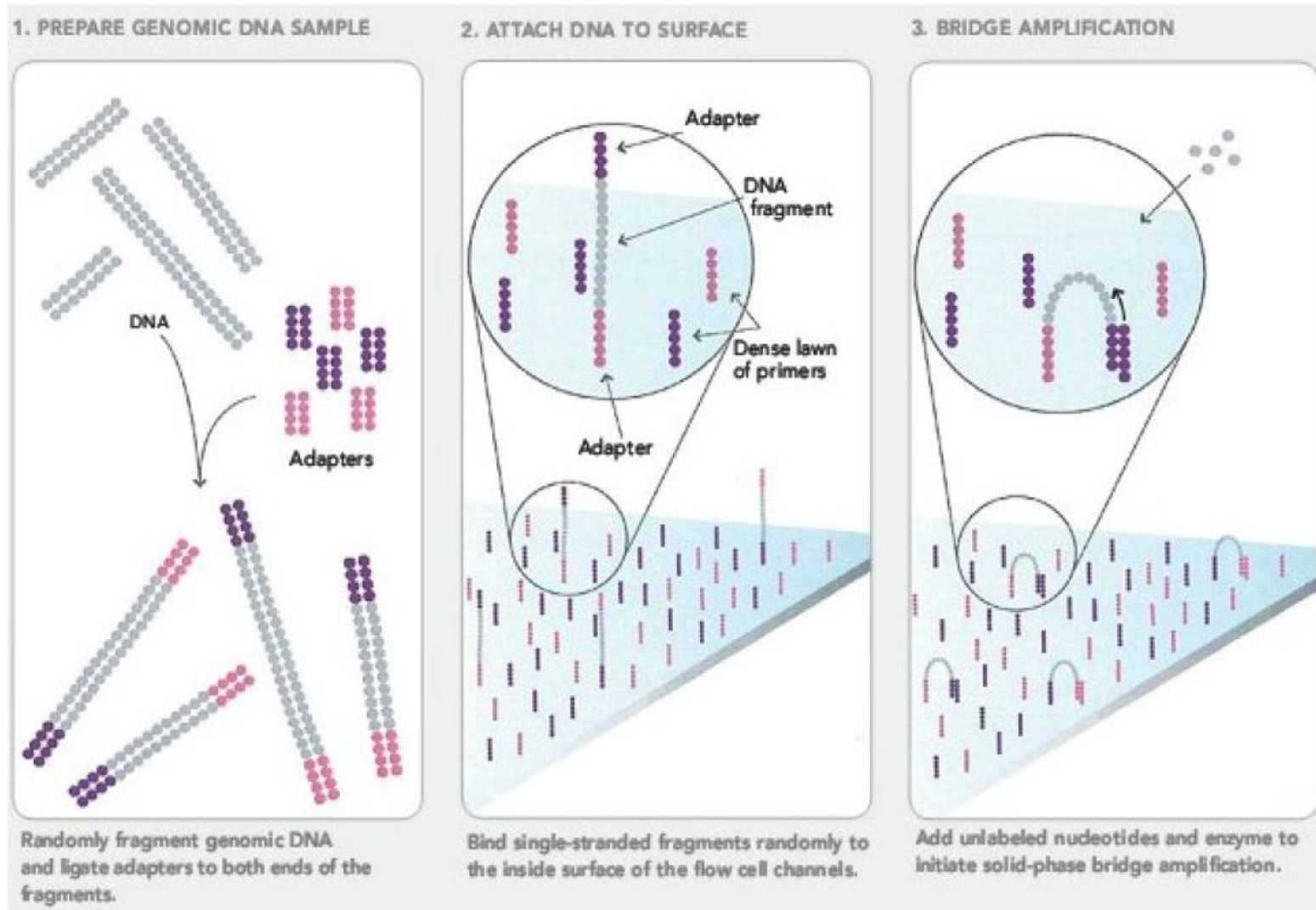


Next Generation Sequencing : Amplified Single Molecule Sequencing



Next Generation Sequencing : Amplified Single Molecule Sequencing “Polony” PCR

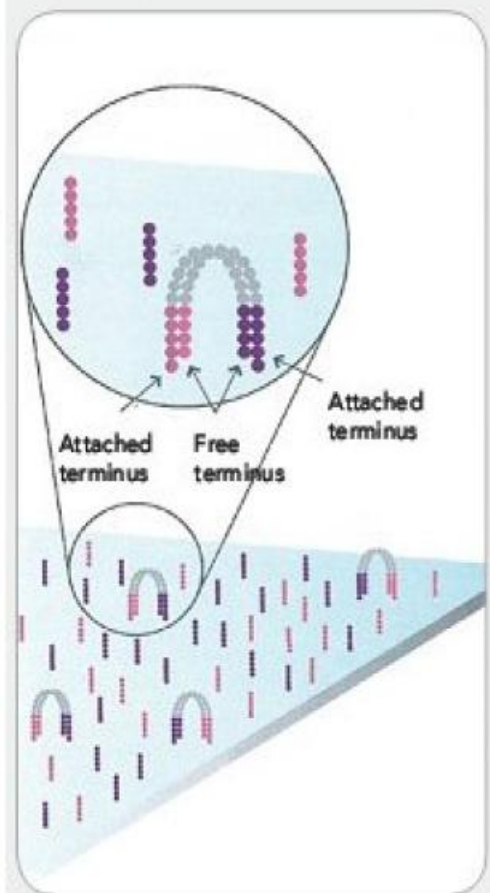
Bridge amplification : Illumina



Next Generation Sequencing : Amplified Single Molecule Sequencing “Polony” PCR

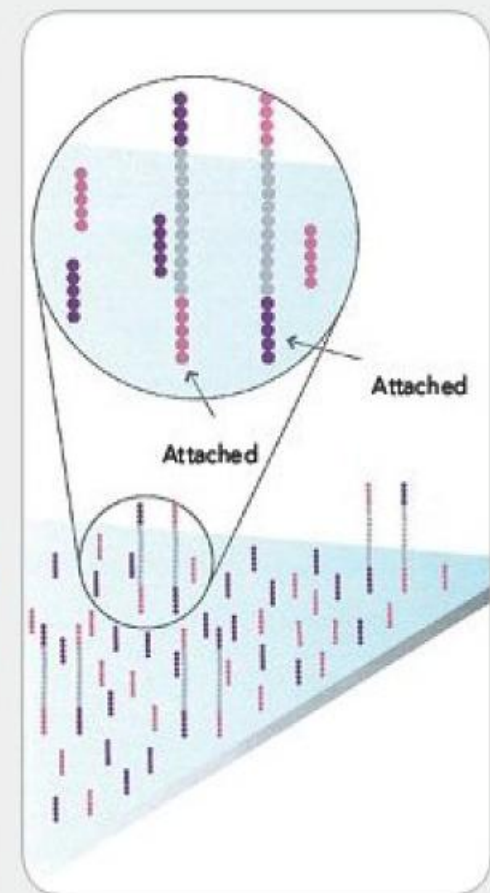
Bridge amplification : Illumina

4. FRAGMENTS BECOME DOUBLE STRANDED



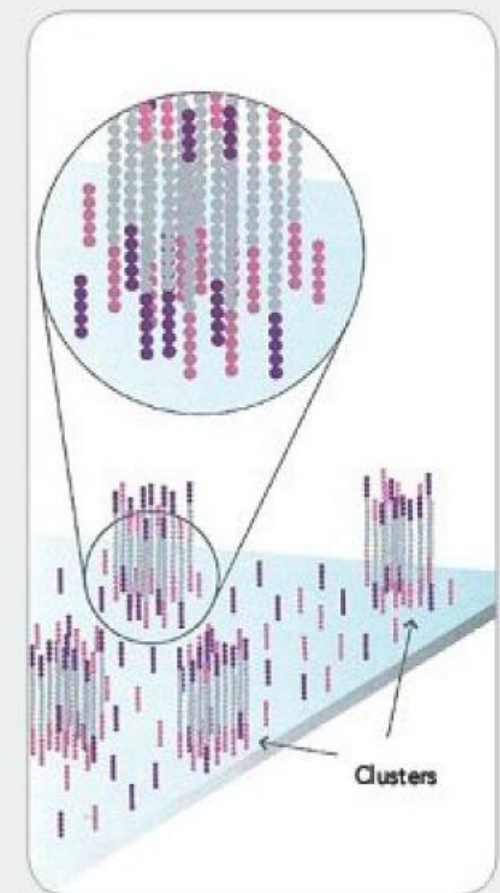
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

5. DENATURE THE DOUBLE-STRANDED MOLECULES



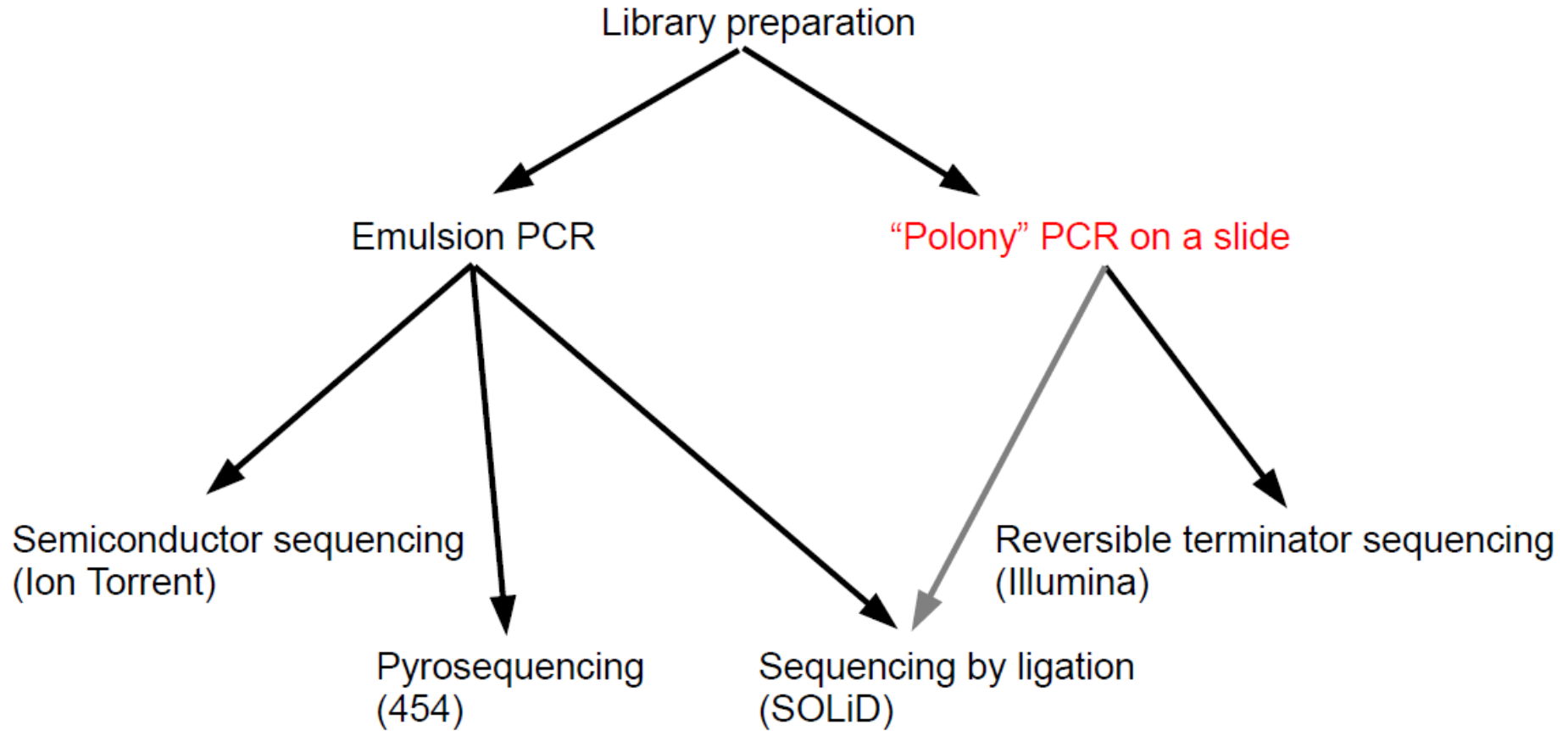
Denaturation leaves single-stranded templates anchored to the substrate.

6. COMPLETE AMPLIFICATION



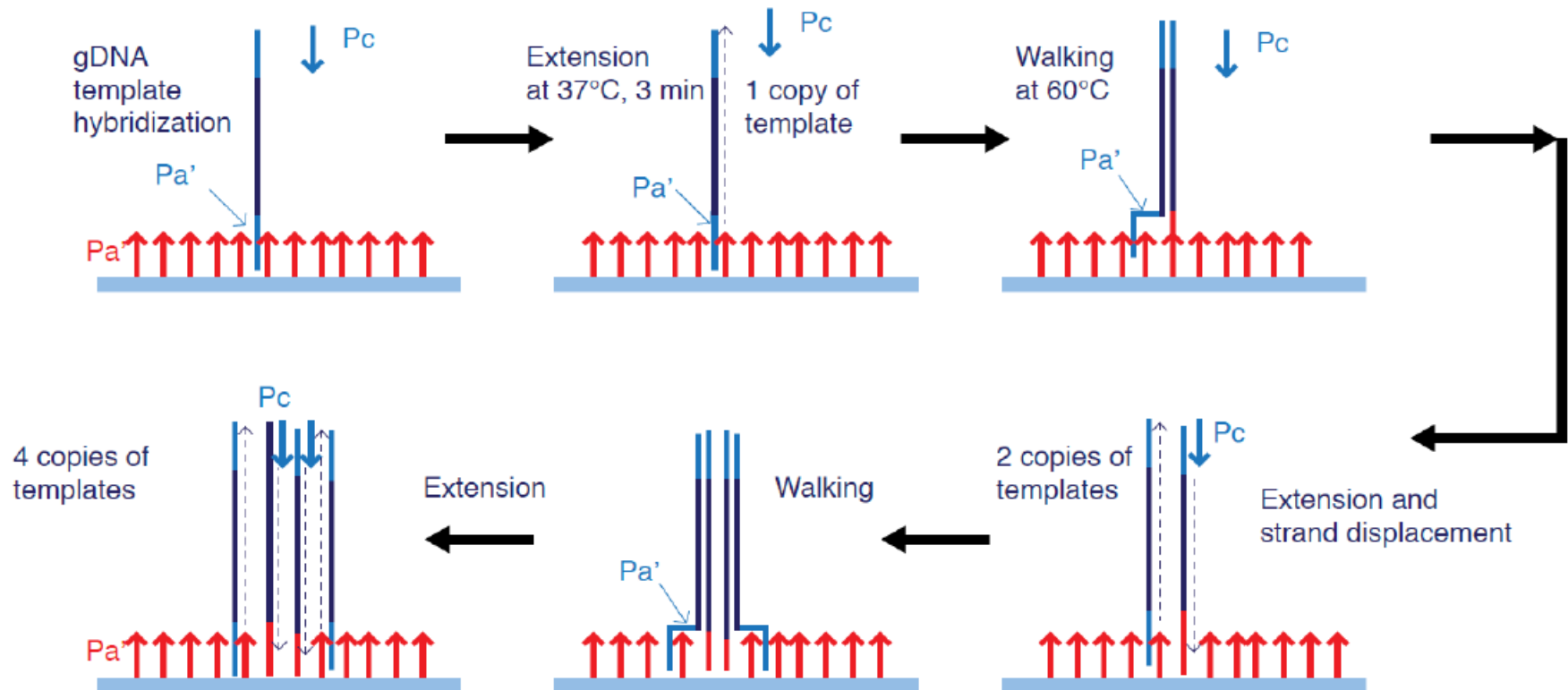
Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

Next Generation Sequencing : Amplified Single Molecule Sequencing



Next Generation Sequencing : Amplified Single Molecule Sequencing “Polony” PCR 21/132

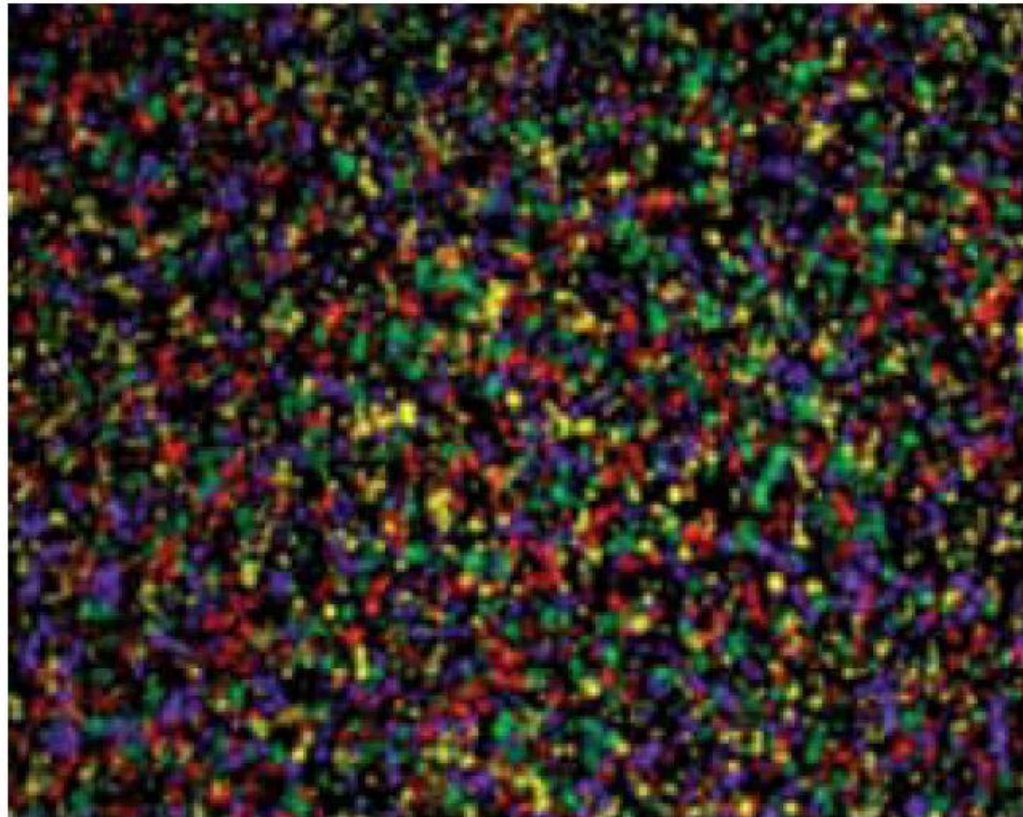
Wildfire amplification : SOLiD



Wildfire chemistry schematic.

Next Generation Sequencing : Amplified Single Molecule Sequencing “Polony” PCR

Wildfire amplification : SOLiD



One million colonies/mm² per FlowChip surface.

Quality scores in sequencing :
Q17, Q20, Q30, ...

Quality score	Probability of incorrect bases	Base call accuracy
10	1 in 10	90 %
17	1 in 50	98 %
20	1 in 100	99 %
30	1 in 1000	99,9 %
40	1 in 10.000	99,99 %
50	1 in 100.000	99,999 %
60	1 in 1.000.000	99,9999%

1 Gb genome : 1 time coverage :

Q20 : 10.000.000 errors

Q30 : 1.000.000 errors

More coverage reduce the errors