Universidade Tiradentes

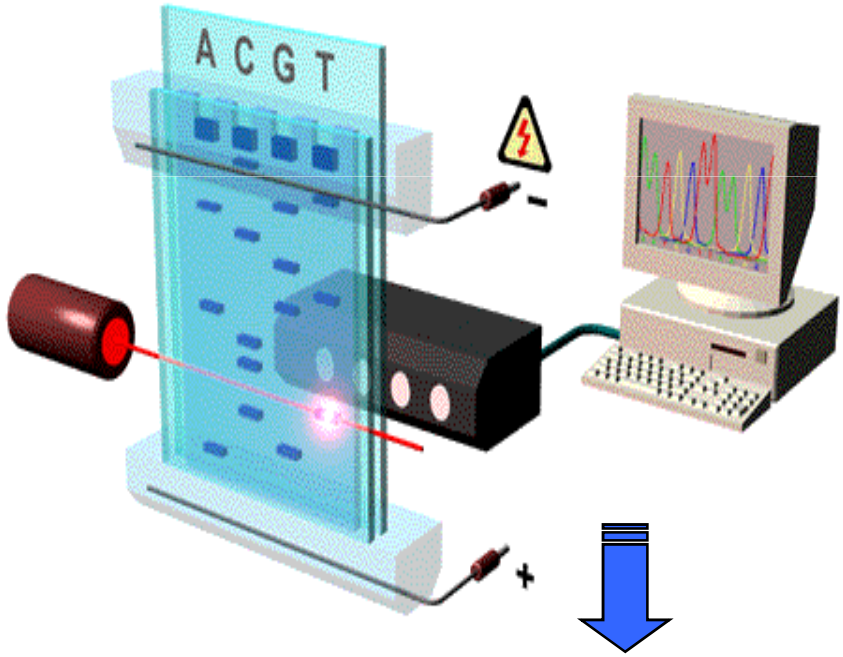
Mestrado em Biotecnologia Industrial

Seqüenciamento de DNA

Sibele Borsuk

sibele@ufpel.tche.br

Sequenciamento de DNA em MegaBACE DNA Analysis Systems



TGTGAACACACGTGTGGATTGG...

Seqüenciamento de DNA

Definição

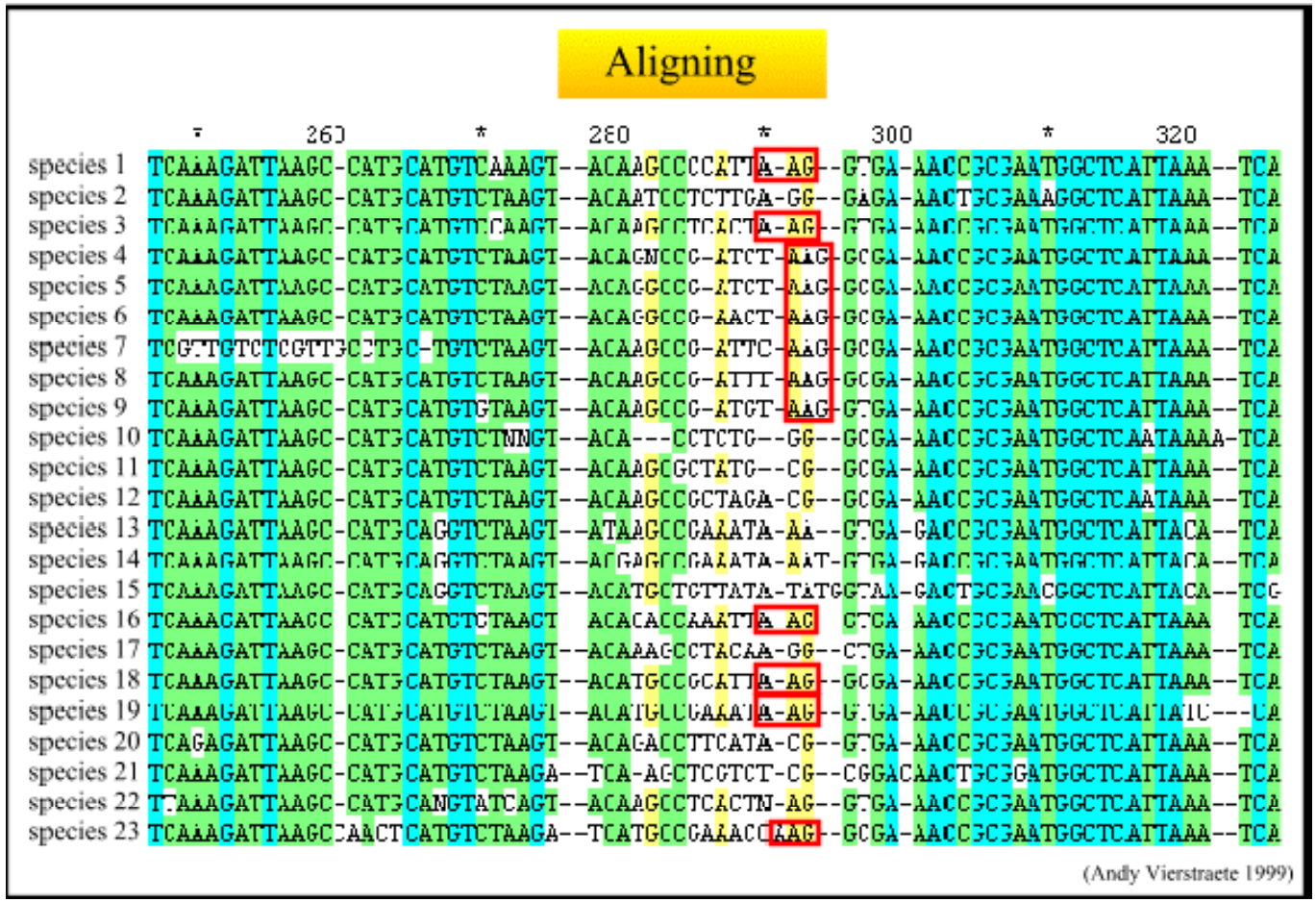
- Identificação da ordem exata dos pares de bases (A, T, G e C) em um segmento de DNA.

Importância

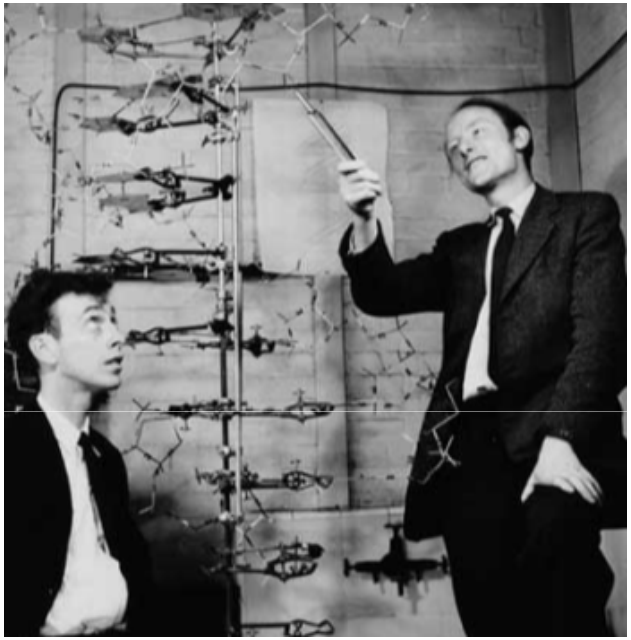
- O conhecimento da seqüência de bases de um gene fornece importantes informações sobre sua estrutura, função e **relação evolutiva** com outros genes (de um mesmo organismo ou de organismos diferentes).

Os organismos possuem padrões

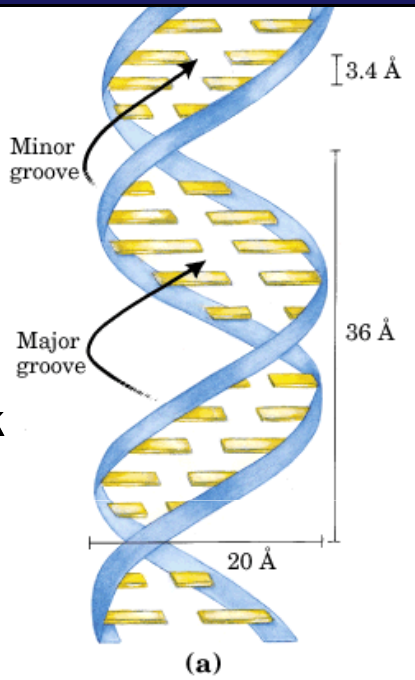
As moléculas também.

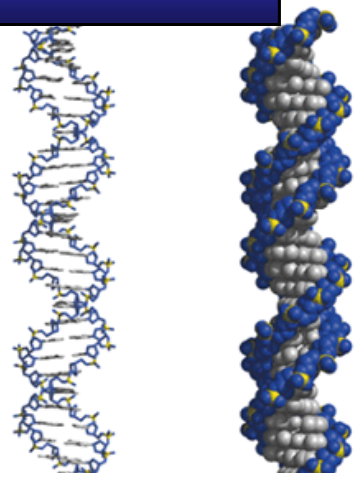


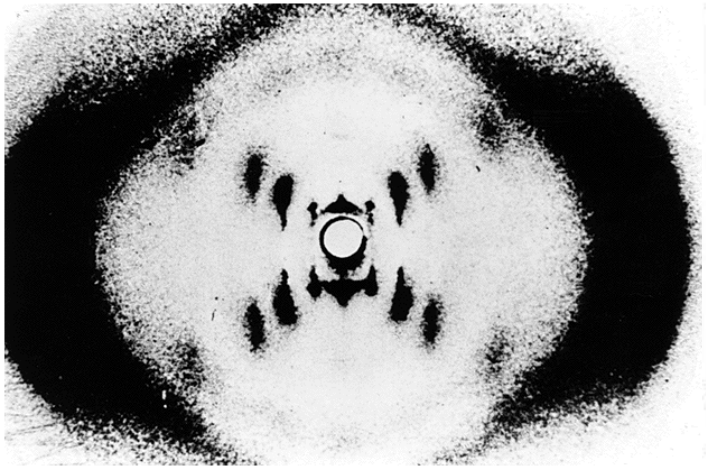
Histórico – O sequenciamento de DNA no tempo



Watson & Crick
Nobel 1962

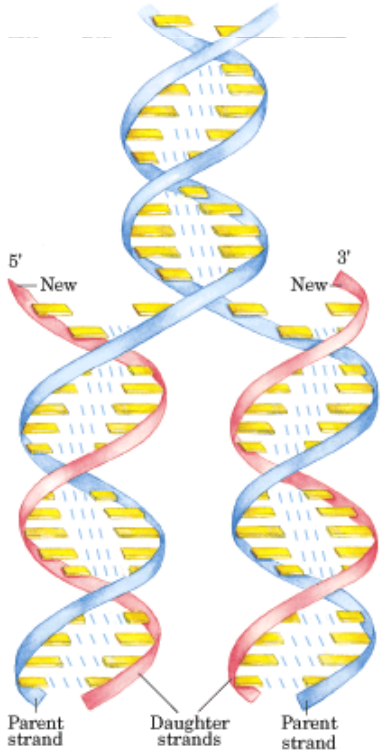






Rosalind Franklin,
1920–1958

Maurice Wilkins



Histórico – O sequenciamento de DNA no tempo

A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase

F. Sanger and A. R. Coulson

Medical Research Council
Laboratory of Molecular Biology
Hills Road, Cambridge CB2 2QH, England

Frederick Sanger
Prêmio Nobel de medicina e fisiologia em 1980
J. Mol. Biol. v.94, p. 441-448, 1975

A new method for sequencing DNA

(DNA chemistry/dimethyl sulfate cleavage/hydrazine/piperidine)

ALLAN M. MAXAM AND WALTER GILBERT

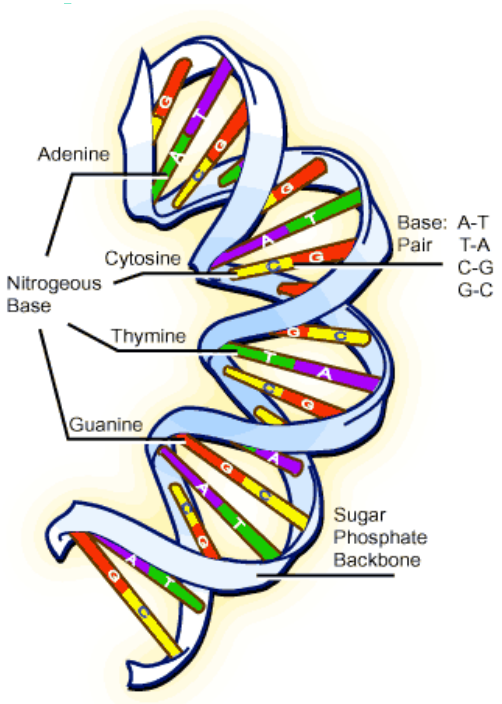
Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138

Contributed by Walter Gilbert, December 8, 1976

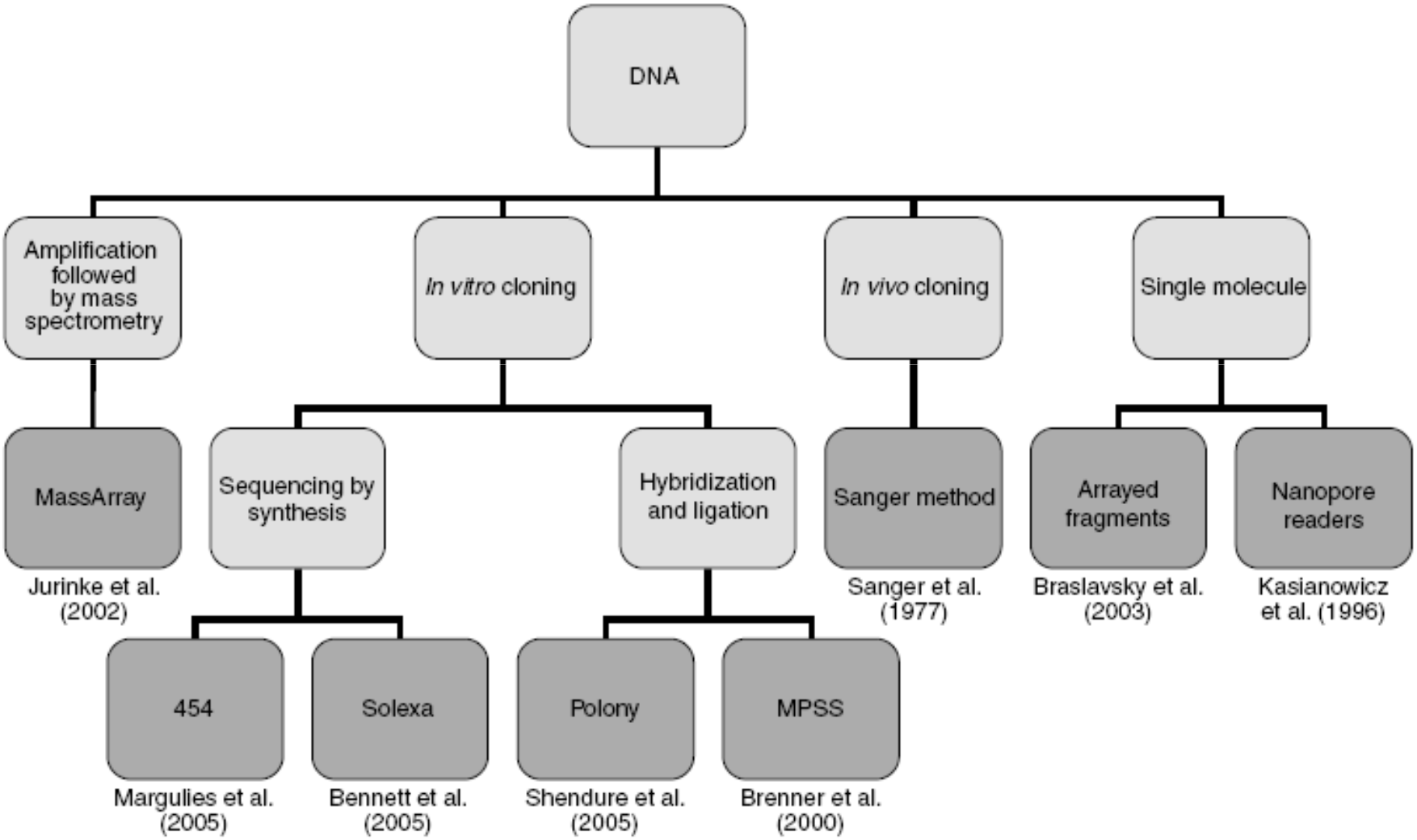
Walter Gilbert
Prêmio Nobel de medicina e fisiologia em 1980
PNAS, vol. 74 No. 2 p. 560-564, 1977

Métodos de seqüenciamento

- Maxam & Gilbert, método químico- 1972
- Sanger sequencing
  - PNAS 74 (1977), n. 12, 5463-5467
  - Sequenciador MegaBACE (1Mpb/24 ho
- Pirosequenciamento
  - Science 281 (1998), n. 5375, 363-365
  -Nature 437 (2005), 362-7
  - Sequenciador 454 (150Mpb/24 horas)

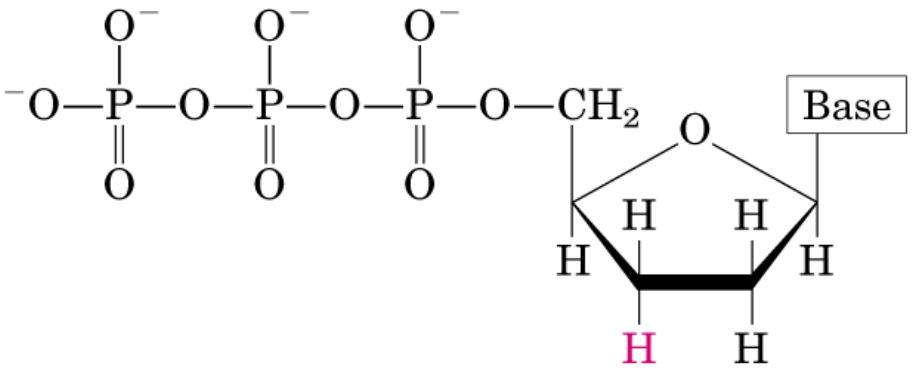


Métodos de seqüenciamento



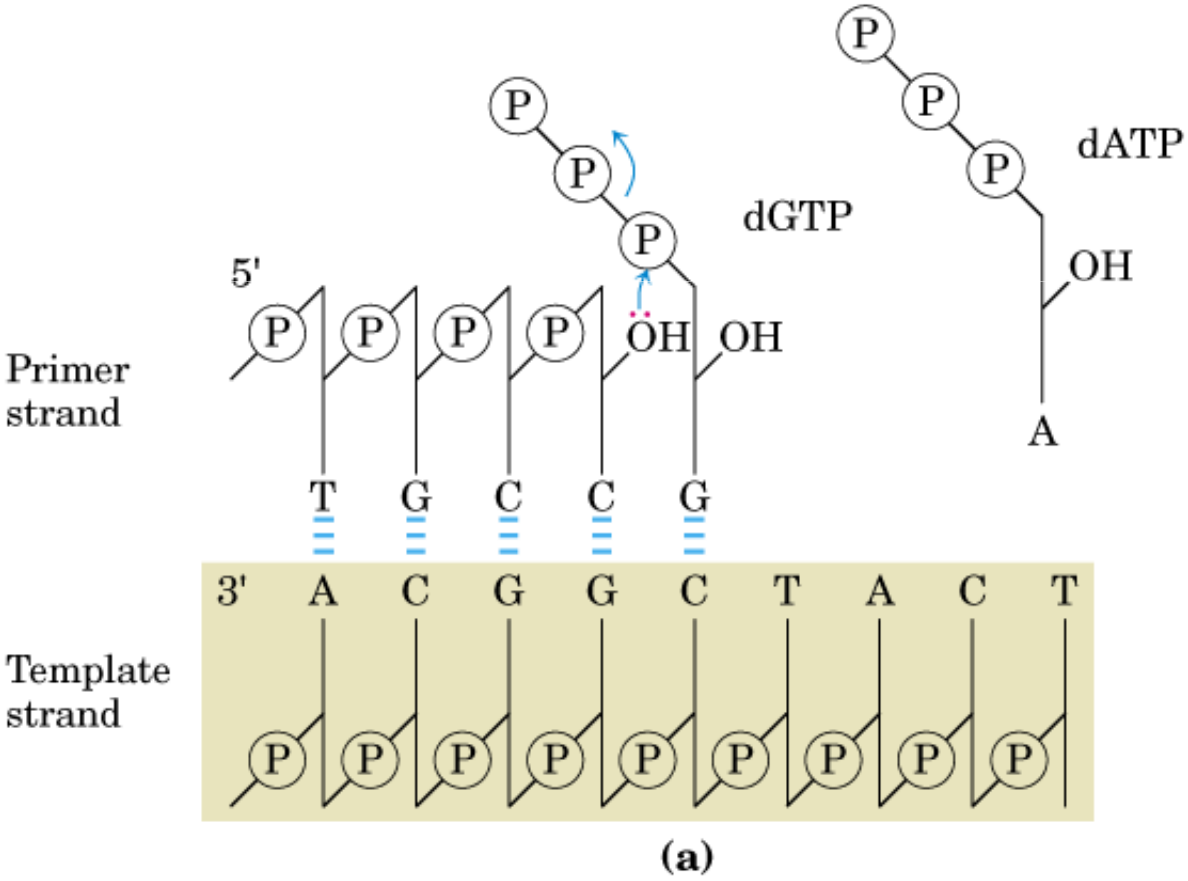
Método "*dideoxi*" de F. Sanger

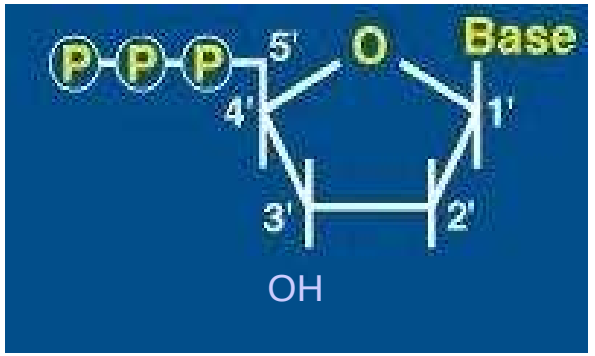
- **Dezembro de 1977:**
  - Também chamado de Método de Terminação da Cadeia;
  - Baseado na utilização de um análogo ao dNTPs, o ddNTP:

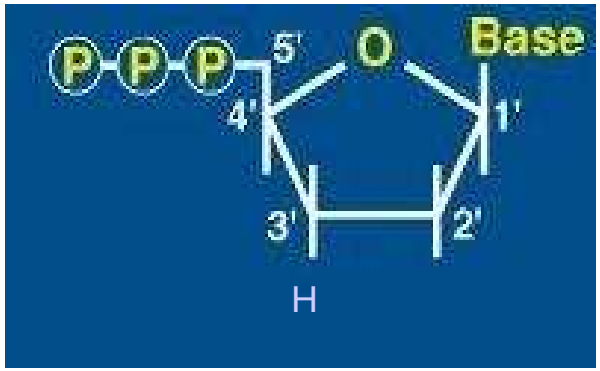


ddNTP analog

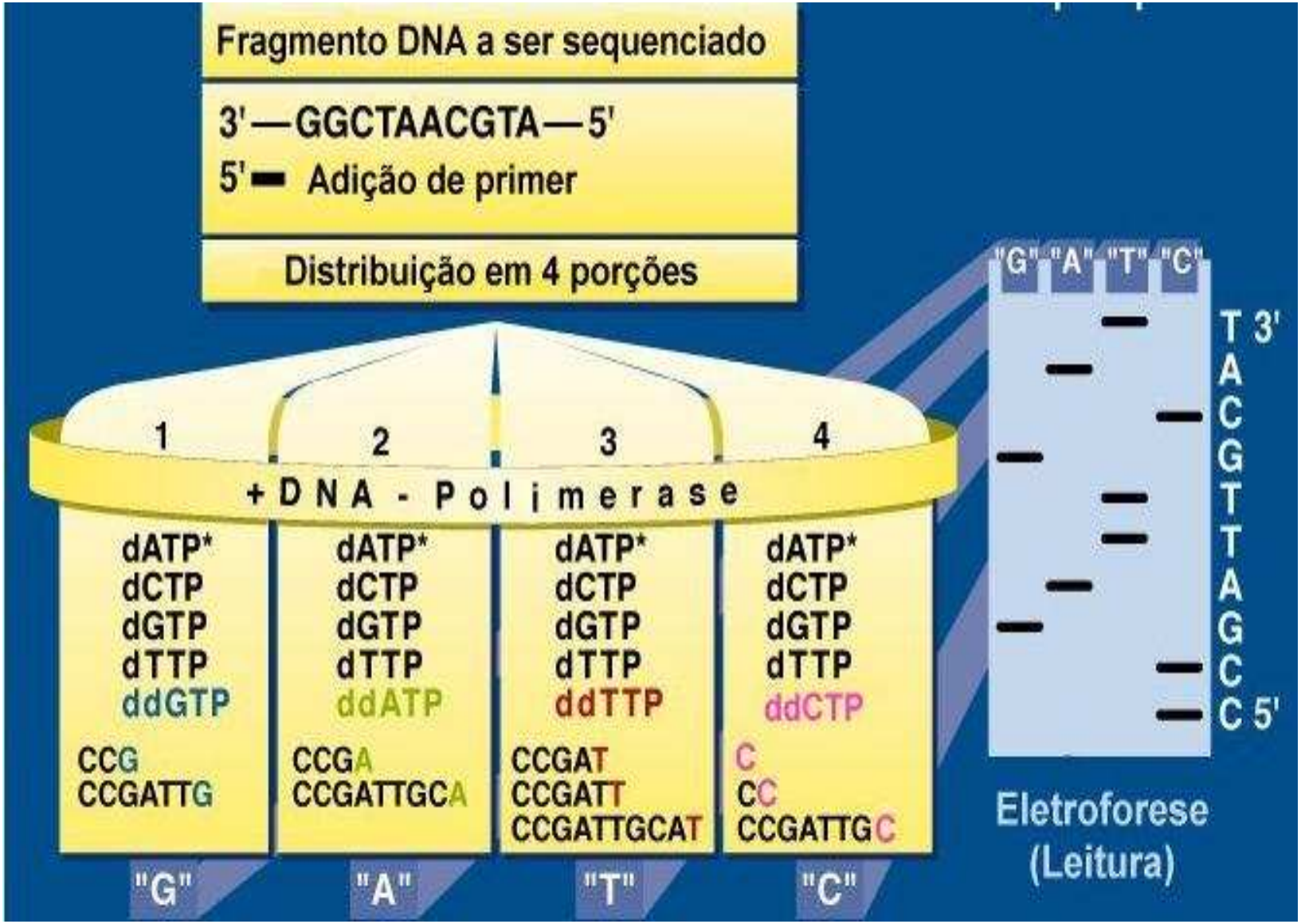
Reação de polimerização







Reação de seqüenciamento com radioisotopos



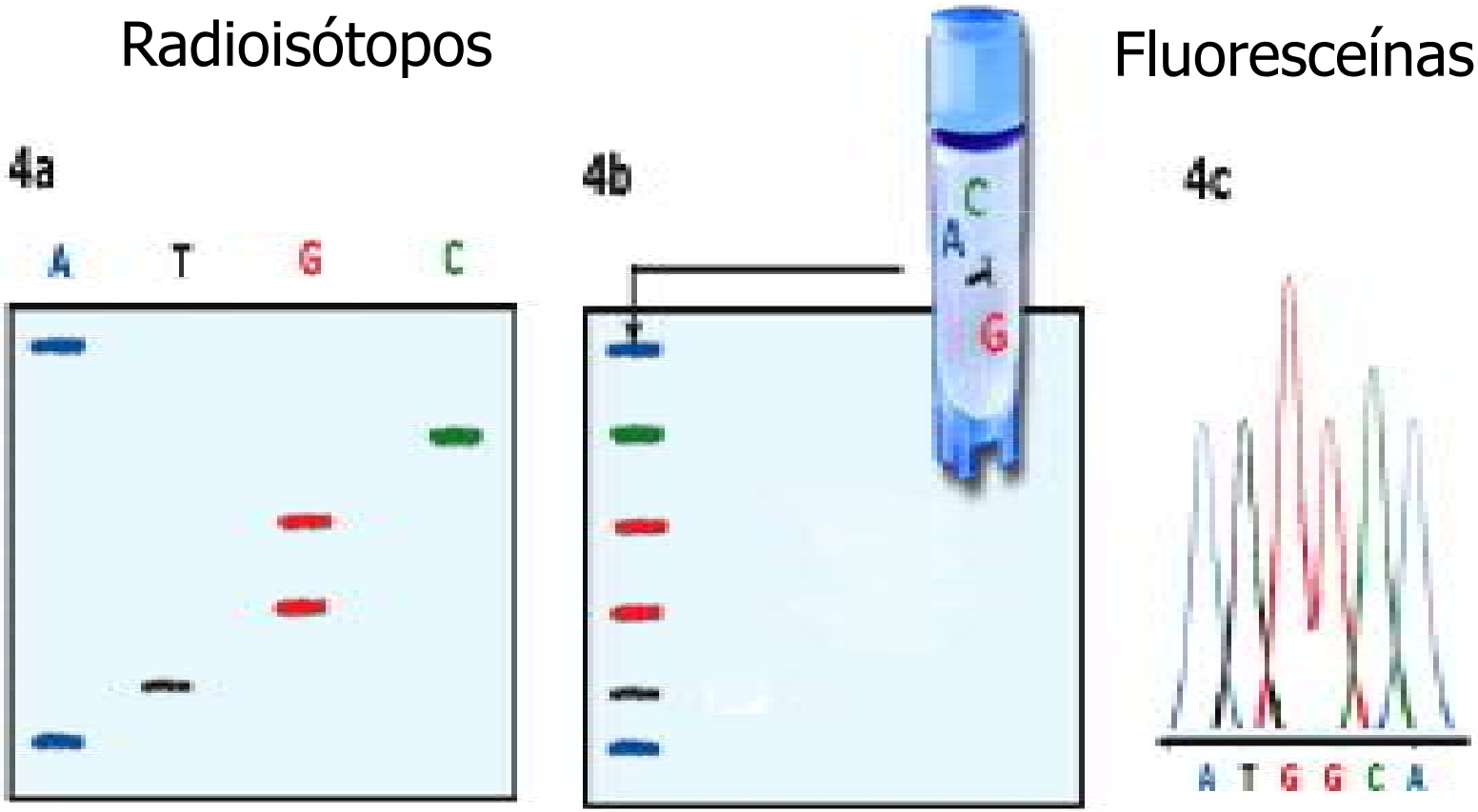
Reação de sequenciamento e marcação do DNA

Leitura das sequências - Manual

Leitura das sequências - Automática

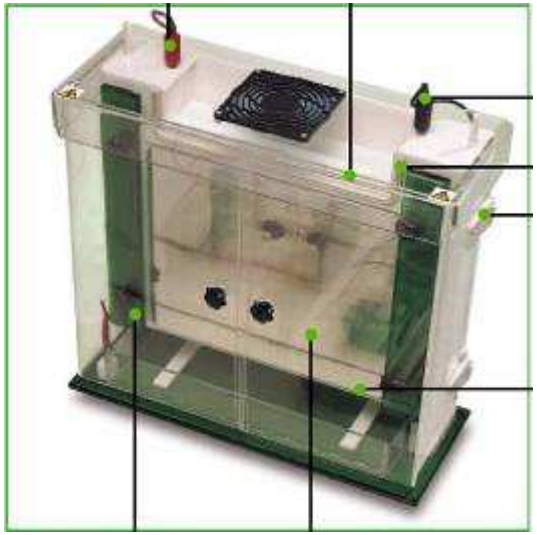
Radioisótopos

Fluoresceínas

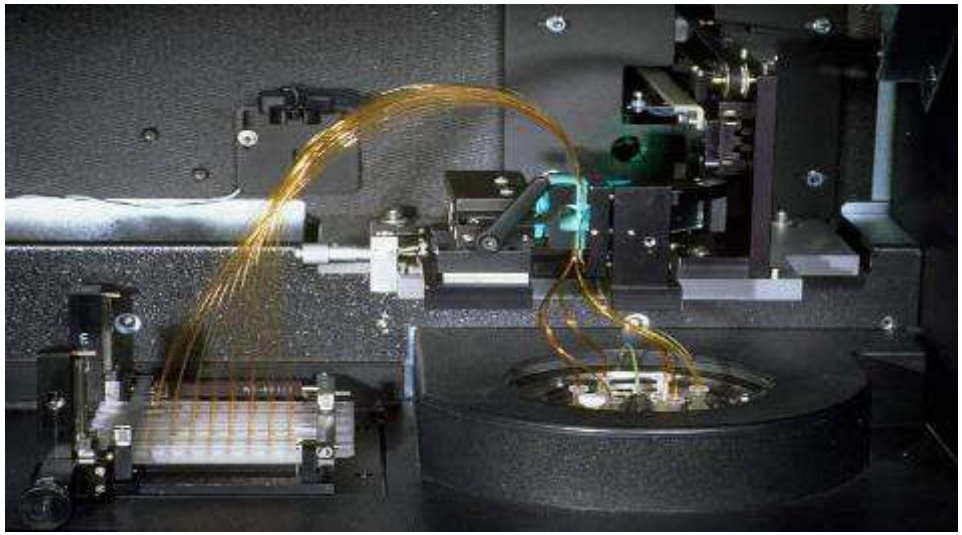


Eletroforese

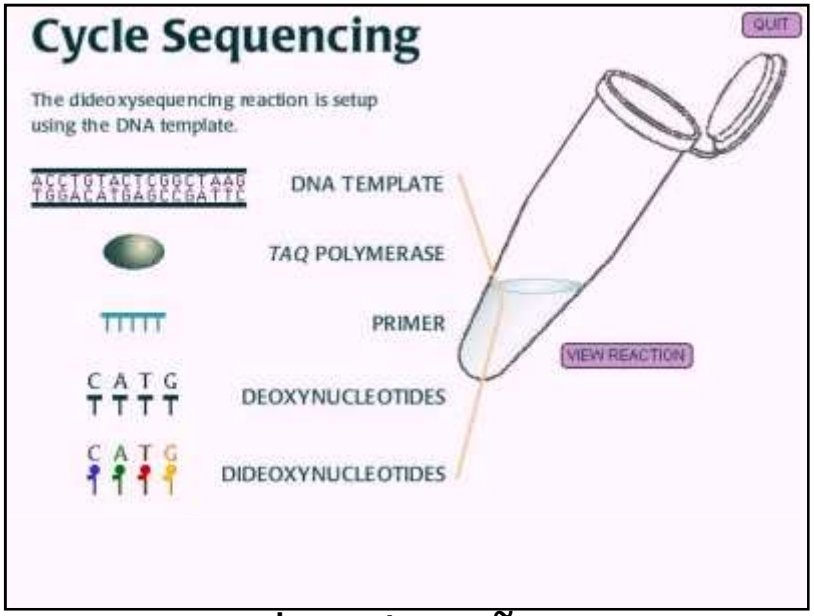
- Placa



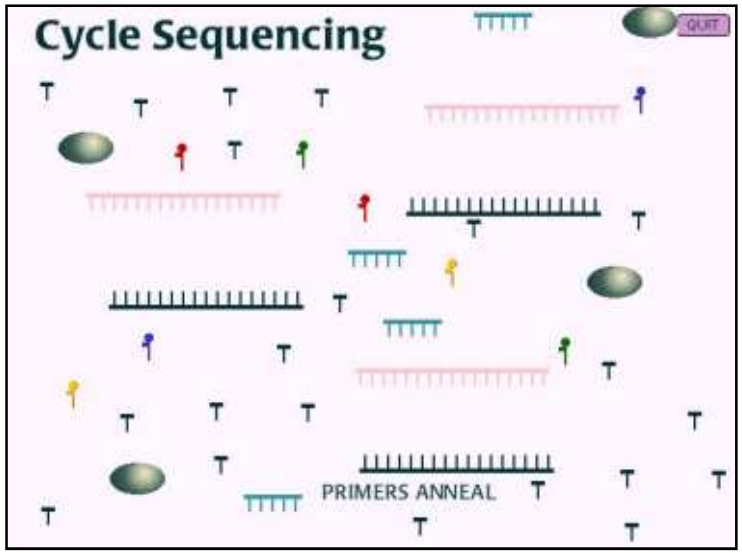
- Capilar



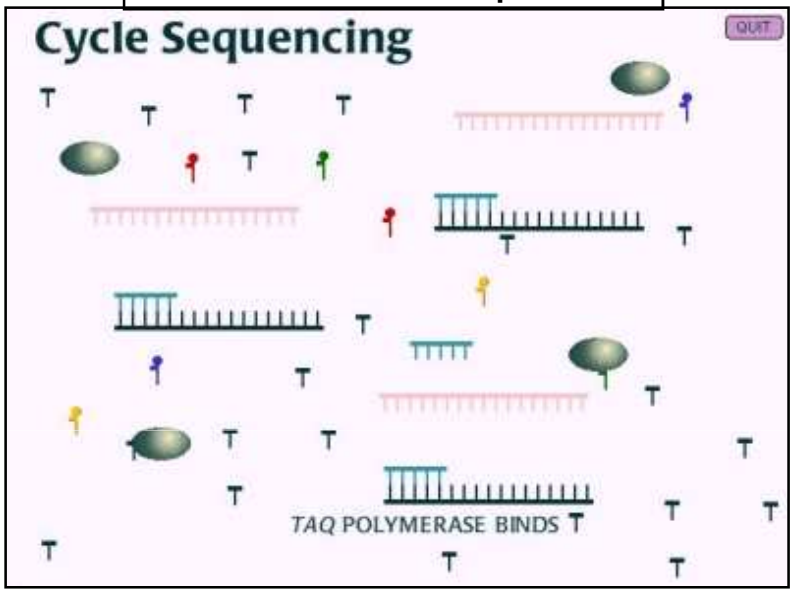
Reação de sequenciamento e leitura automatizada

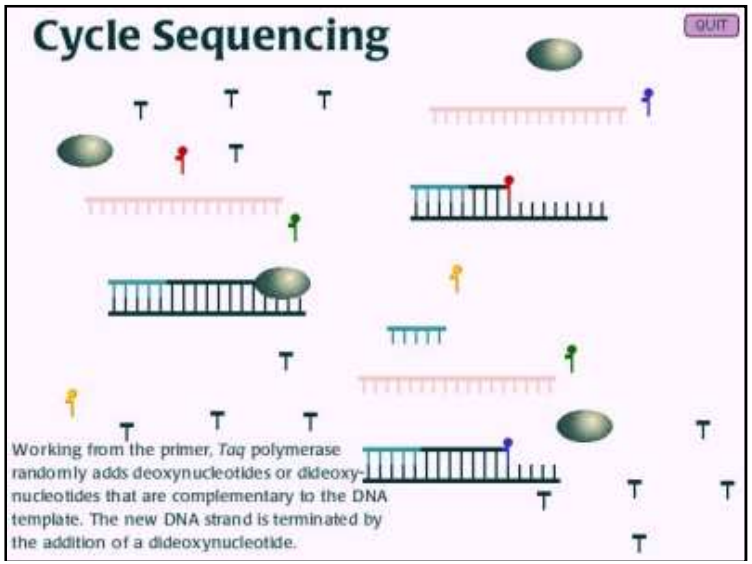


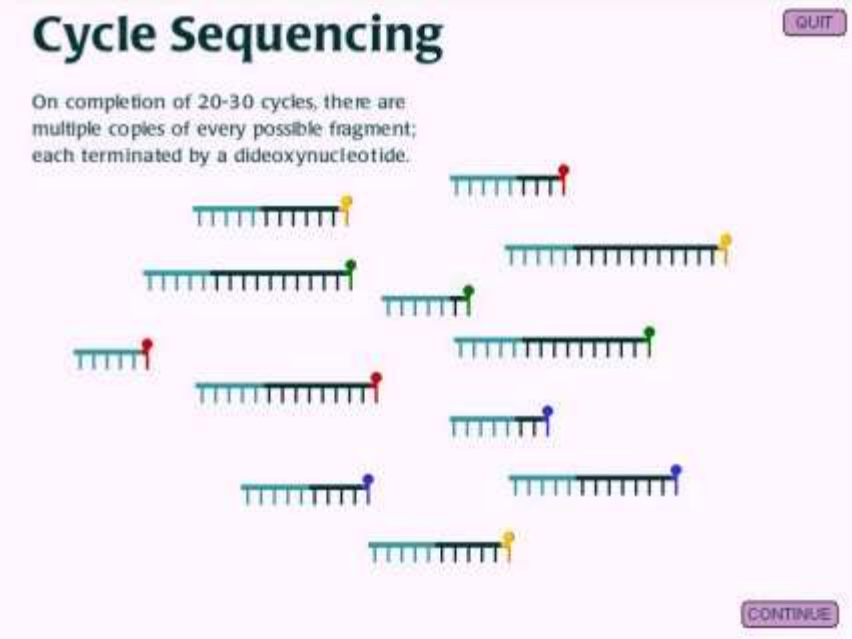
desnaturação

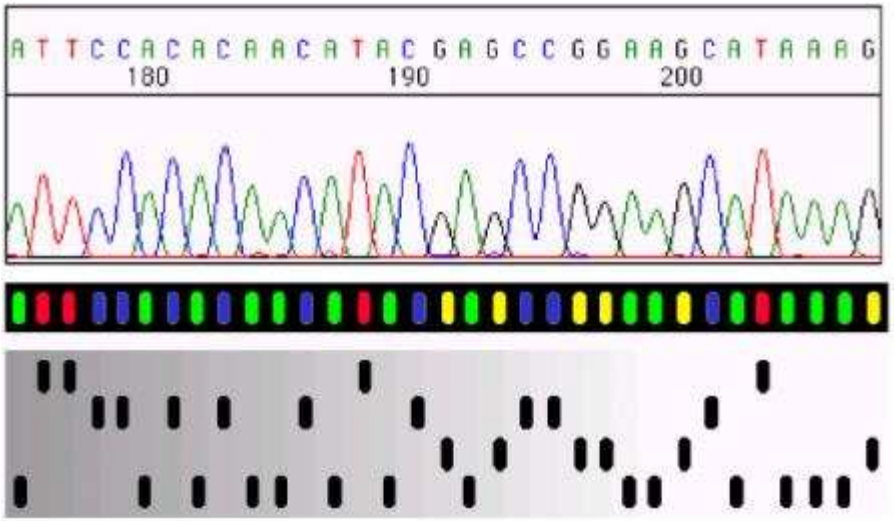


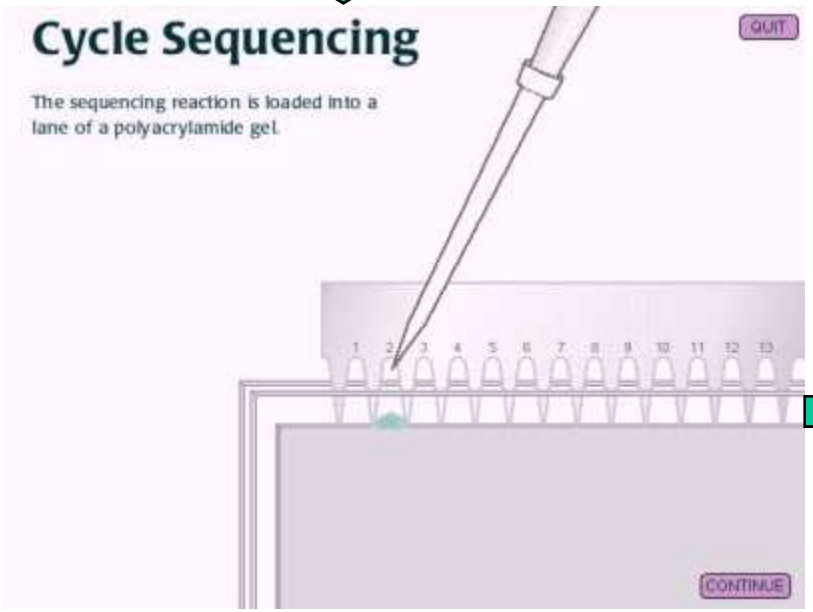
anelamento dos primers

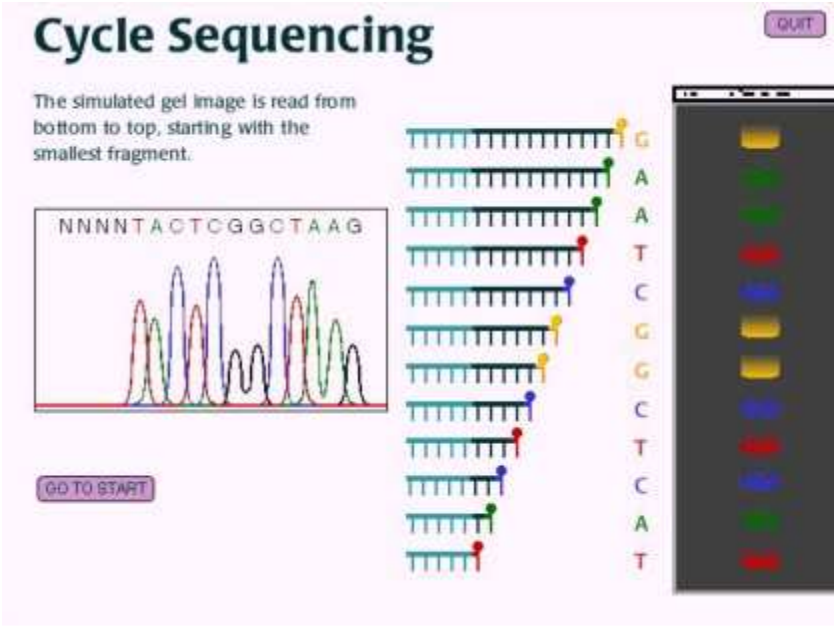


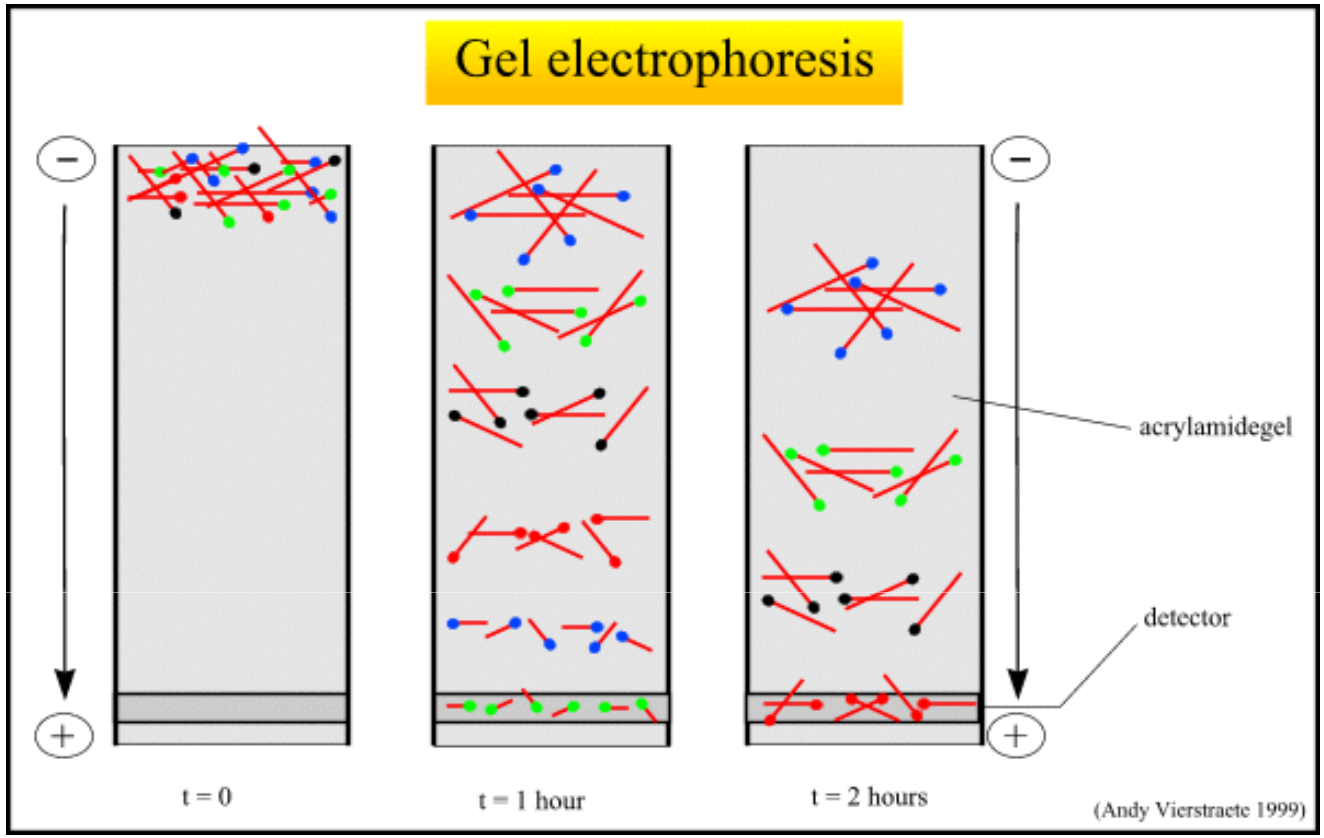



Cycle Sequencing
QUIT
On completion of 20-30 cycles, there are multiple copies of every possible fragment; each terminated by a dideoxynucleotide.
CONTINUE


A T T C C A C A C A A C A T A C G A G C C G G A A G C A T A A A G
180
190
200


Cycle Sequencing
QUIT
The sequencing reaction is loaded into a lane of a polyacrylamide gel.
1 2 3 4 5 6 7 8 9 10 11 12 13
CONTINUE

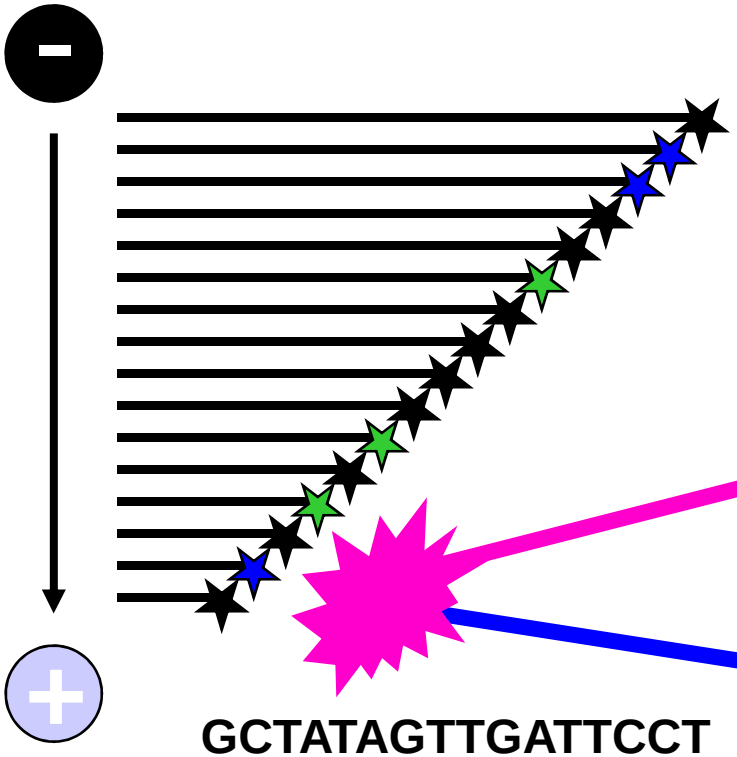

Cycle Sequencing
QUIT
The simulated gel image is read from bottom to top, starting with the smallest fragment.
NNNNTACTCGGCTAAG
GO TO START
G
A
A
T
C
G
G
C
T
C
A
T



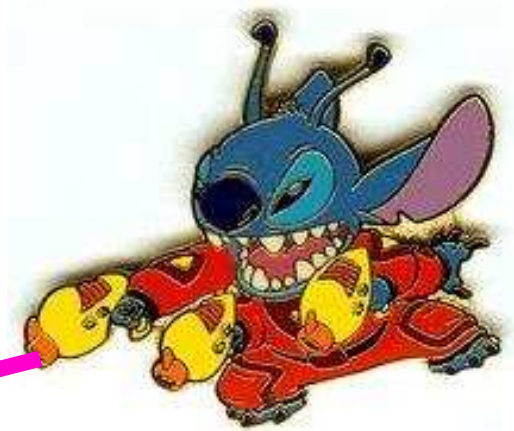
Exemplo de gel utilizado nos seqüenciadores de placa (ex.: 377). A diferença de tamanho permite a separação dos grupos de fragmentos, e esta "distribuição normal" da passagem dos fragmentos é representada pelo eletroferograma (ou cromatograma) de cada seqüência (read).

Generic Sequencing Instrument

Electrical Field

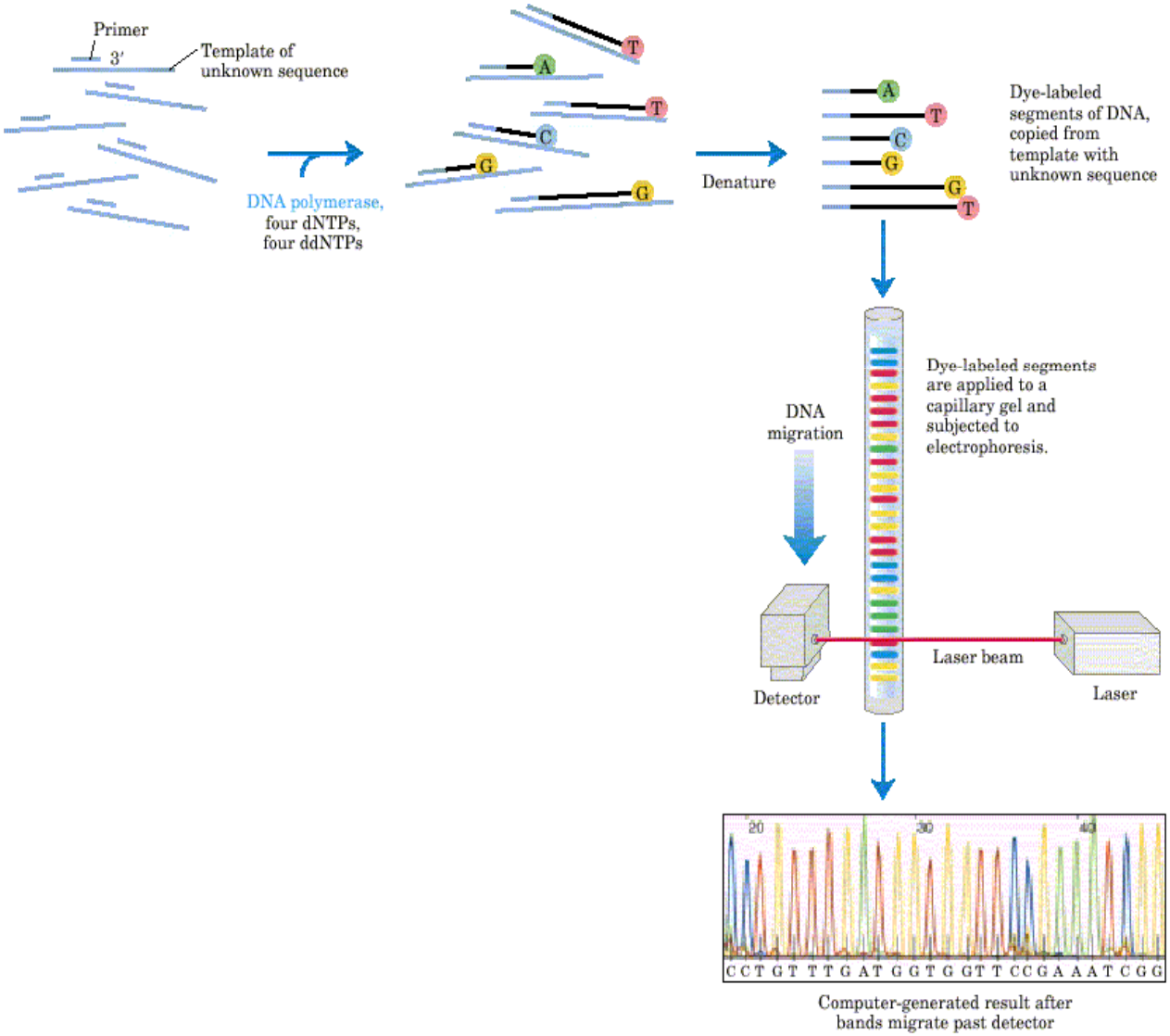


GCTATAGTTGATTCCT



Laser

Sequenciamento- Automático



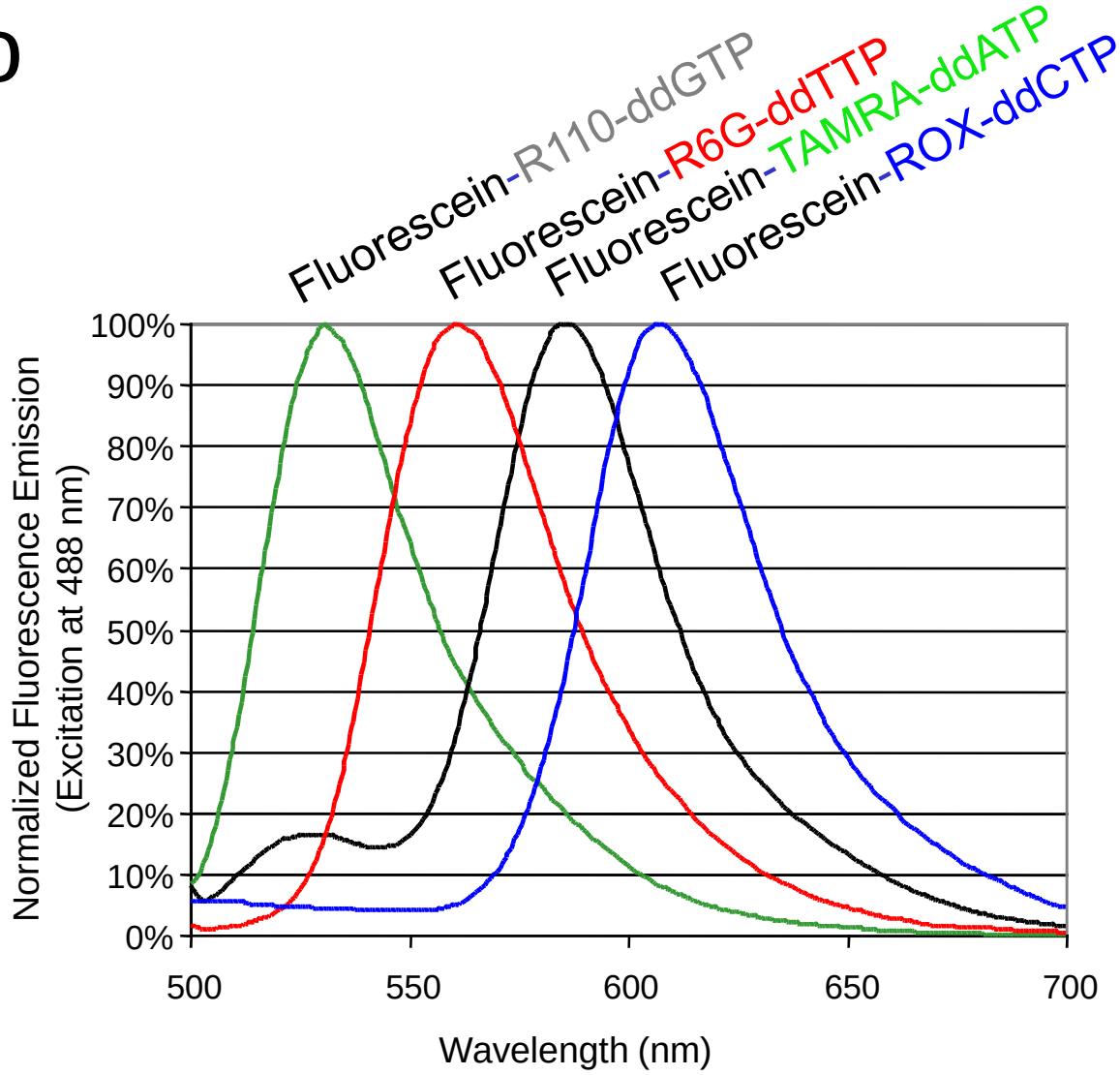
DYEnamic ET Terminators – Pré Mix

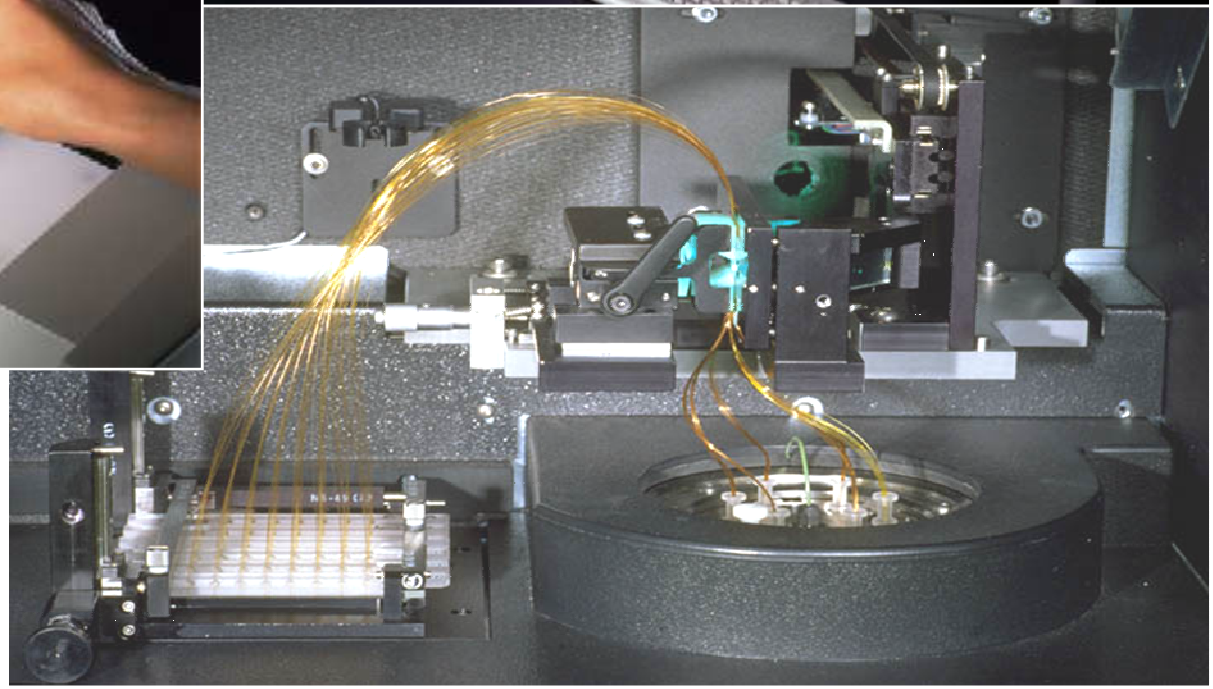
Energy Transfer Dyes

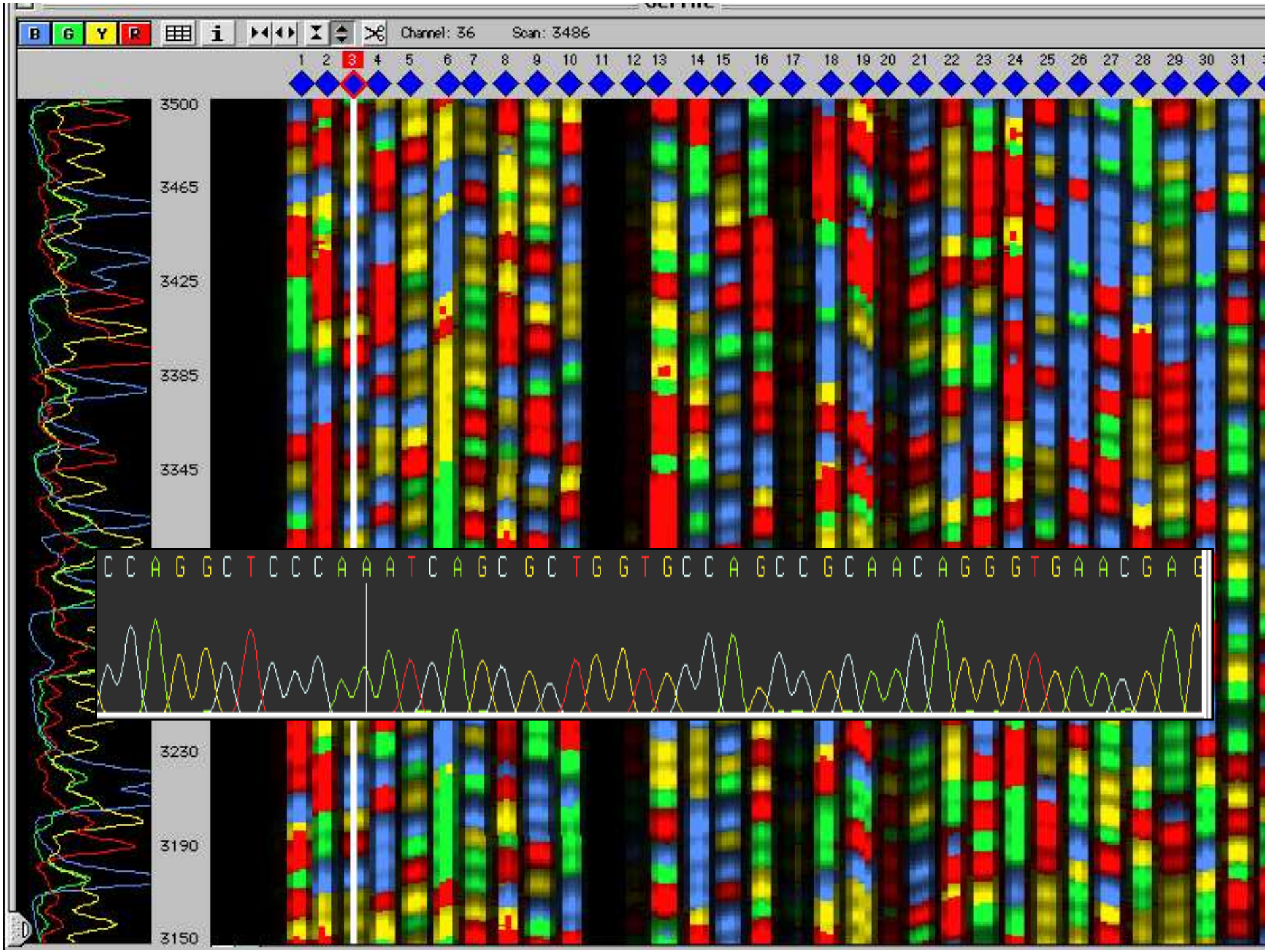
- Fluorescein Donor Dye
- Standard Rhodamine Acceptor Dyes

Fluorescein-R110-**ddGTP**
Fluorescein-R6G-**ddTTP**
Fluorescein-TAMRA-**ddATP**
Fluorescein-ROX-**ddCTP**

DYEnamic ET terminator - espectro de emissão






B G Y R
Channel: 36
Scan: 3486
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
3500
3465
3425
3385
3345
C C A G G C T C C C A A A T C A G C G C T G G T G C C A G C C G C A A C A G G G T G A A C G A G
3230
3190
3150

Análise dos resultados por Bioinformática

Organizando as Seqüências de DNA

5′ 3′

ATGCTTCTGGCAGATCTGAACAGTGTTACTGAT

ATGCTTCTGGCAGATCTGAACAGTGT

ATGCTTCTGGCAGATCTGAACAGTGTTACTGATATTGCTT

ATGCTTCTGGCAGATCTGAACAGTGTTACTGATAT

ATGCTTCTGGCAGATCTGAACAGTGTTAC

ATGCTTCTGGCAGATCTGAACAGT

ATGCTTCTGGCAGATCTGAACAGTGTTACTGATATTGC

ATGCTTCTGGCAGATCTGAACAGTGTTACTGATATT

ATGCTTCT

ATGCTTCTGGCAGATCT

ATGCTTCTGGCAGATCTGAACAGTGTTACTGATATTGCTT

ATGCTTCTGGCAGATCTGAACAGTGTT

ATGCTTCTGGCAGAT

||||||||||||||||||||||||||||||||||||||||

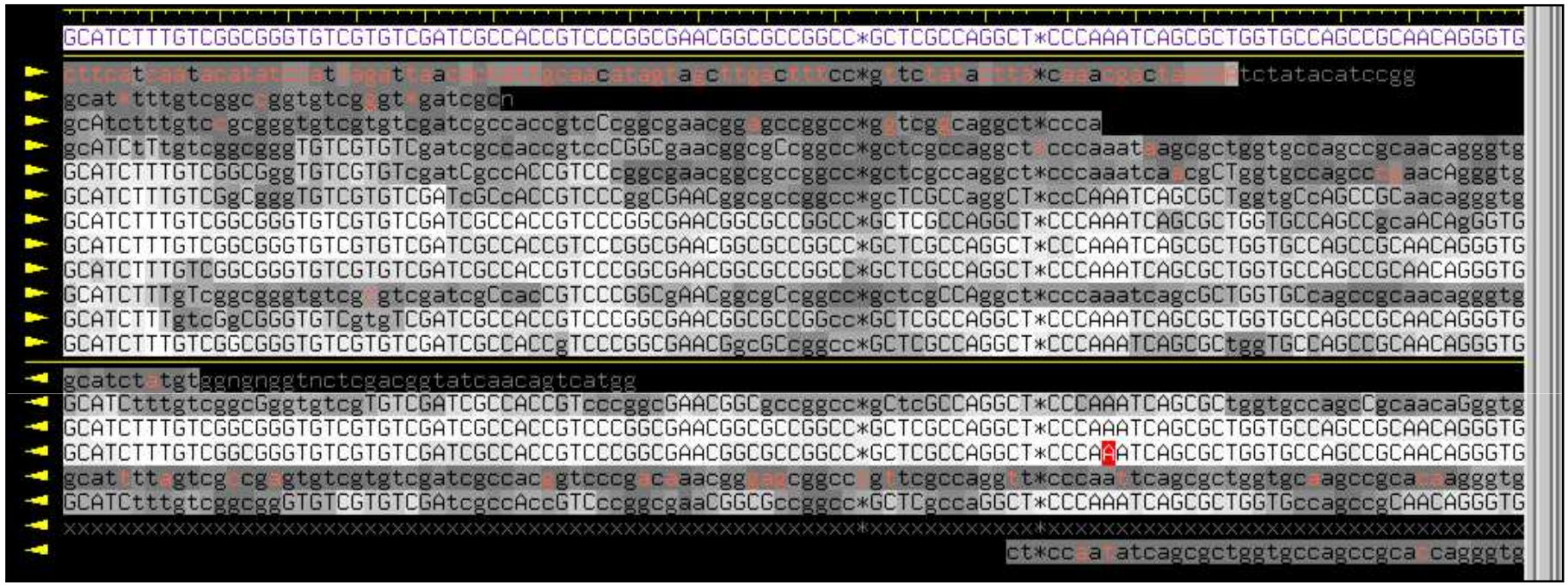
TACGAAGACCGTCTAGACTTGTCACAATGACTATAACGAA

Leitura da seqüência de DNA

Gel:

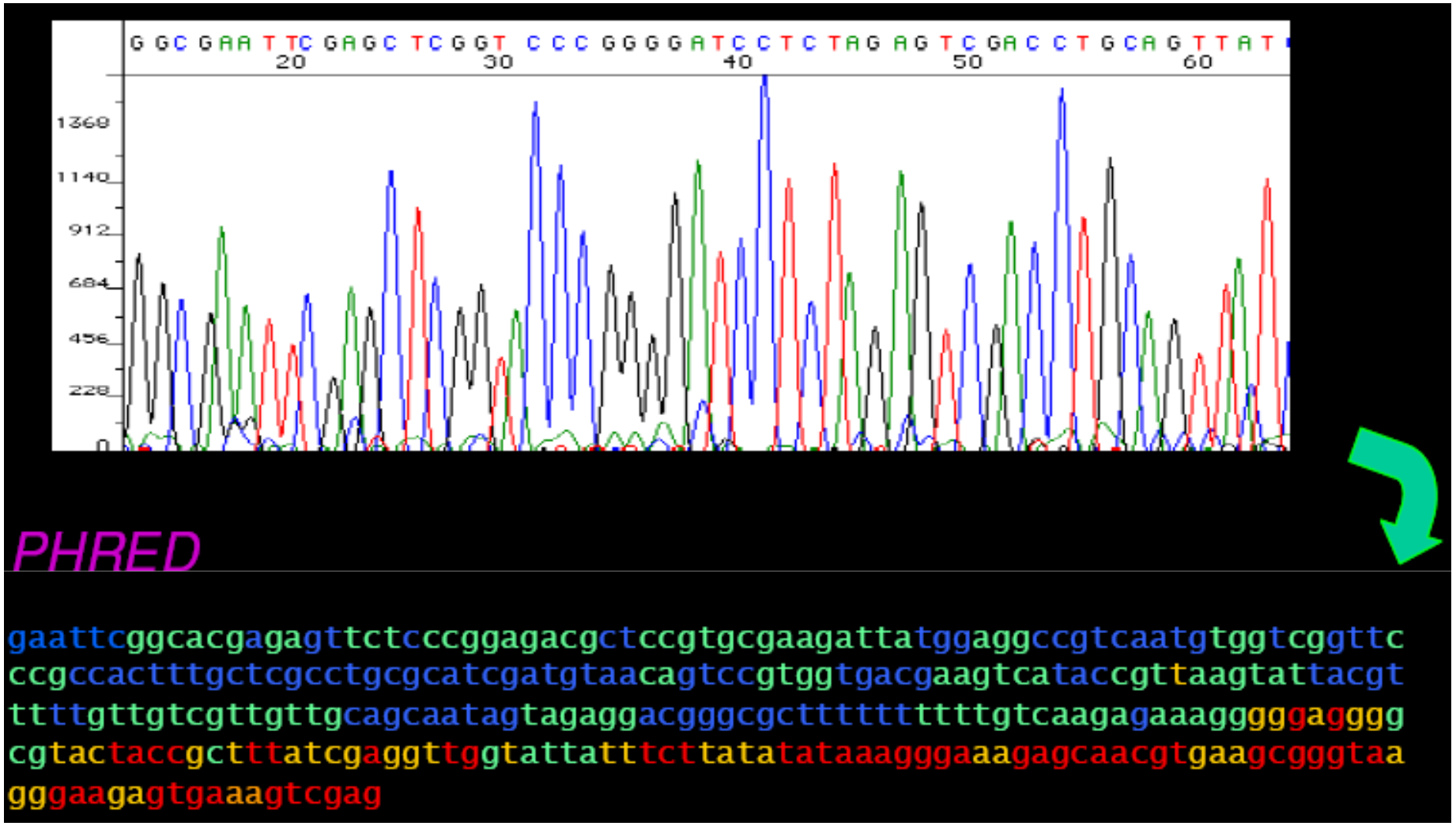
G	GCGAATGCGTCCACAACGCTACAGGT**G**
T	GCGAATGCGTCCACAACGCTACAGG**T**
G	GCGAATGCGTCCACAACGCTACAG**G**
G	GCGAATGCGTCCACAACGCTACA**G**
A	GCGAATGCGTCCACAACGCTAC**A**
C	GCGAATGCGTCCACAACGCTA**C**
A	GCGAATGCGTCCACAACGCT**A**
T	GCGAATGCGTCCACAACGC**T**
C	GCGAATGCGTCCACAACG**C**
G	GCGAATGCGTCCACAACG
C	GCGAATGCGTCCACAA**C**
A	GCGAATGCGTCCACA**A**
A	GCGAATGCGTCCAC**A**
C	GCGAATGCGTCCA**C**
A	GCGAATGCGTCC**A**
C	GCGAATGCGTC**C**
C	GCGAATGCGT**C**
T	GCGAATGCG**T**
G	GCGAATGC**G**
C	GCGAATG**C**
G	GCGAAT**G**
T	GCGAA**T**

Montagem



Limpeza das seqüências:

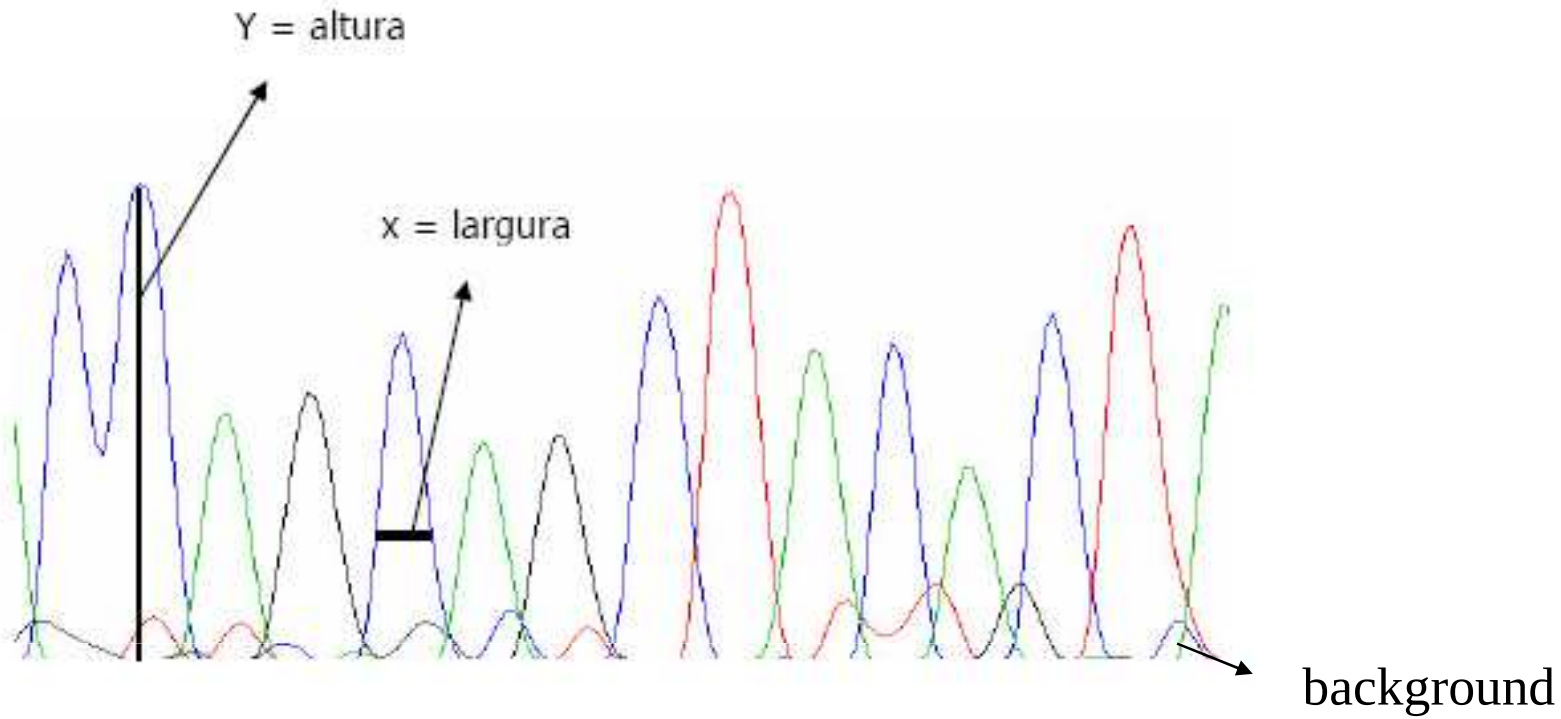
- remoção de seqüências ribossômicas,
- remoção de seqüências de vetor,
- remoção da região de poliA,
- corte por qualidade,
- eliminação das derrapagens



O programa PHRED lê o chromatograma identificando e dando uma nota para cada base que forma a sequência :

0 0 5 6 7 10 10 9 12 15 20 20 30 30 35 40 41 45 50 **56 56** 50 40 ...

Genome Research 8 (3) (1998), 175-185



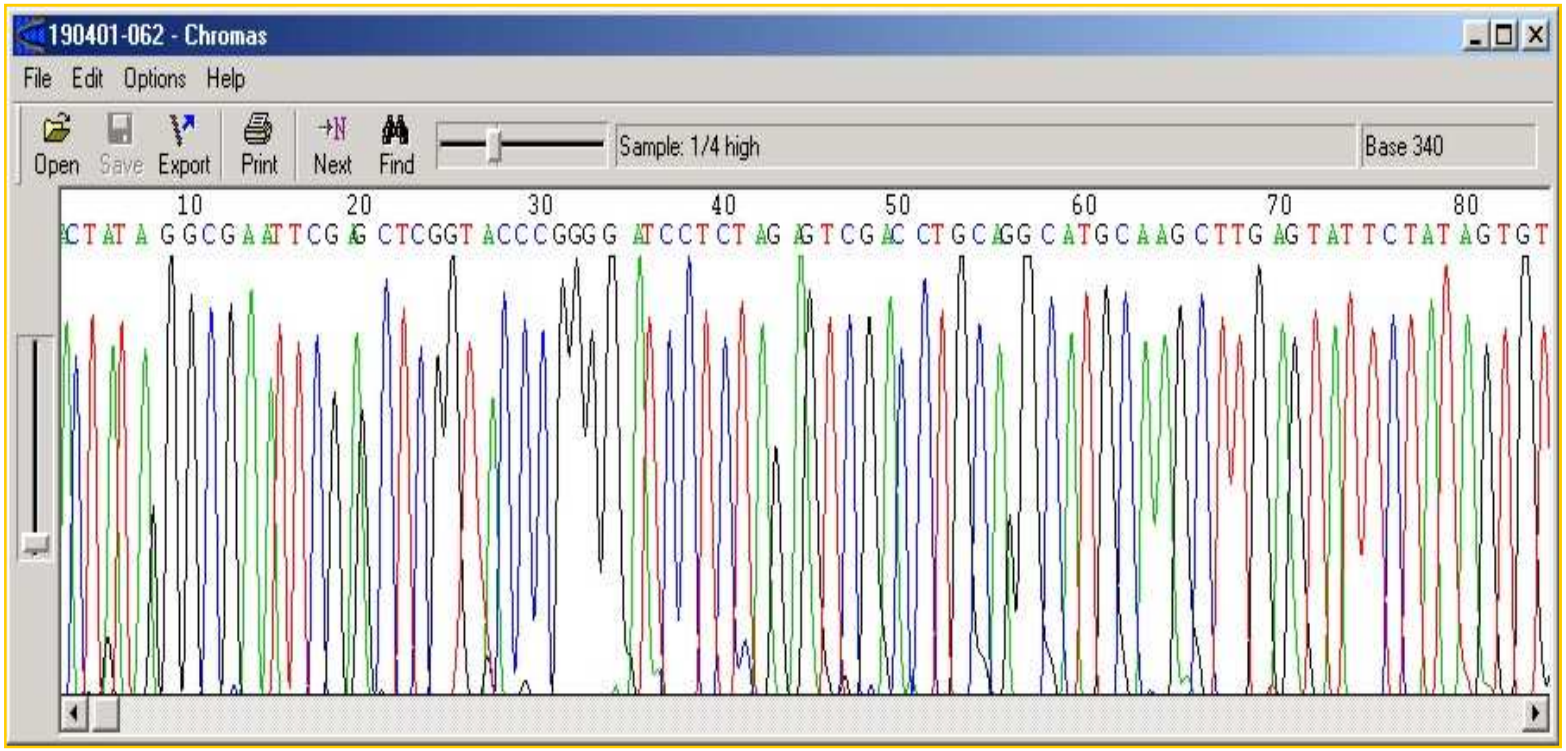
Cromatograma gerado pelo Sequenciador

- A identificação dos picos é feita através de uma transformada de fourier do sinal

- A nota é ligada com a resolução entre os picos vizinhos e a altura do background

Analisando o cromatograma

Região de qualidade alta



- Picos bem definidos e grandes.
- Linha de base boa.
- Distância entre picos anterior e posterior constante.

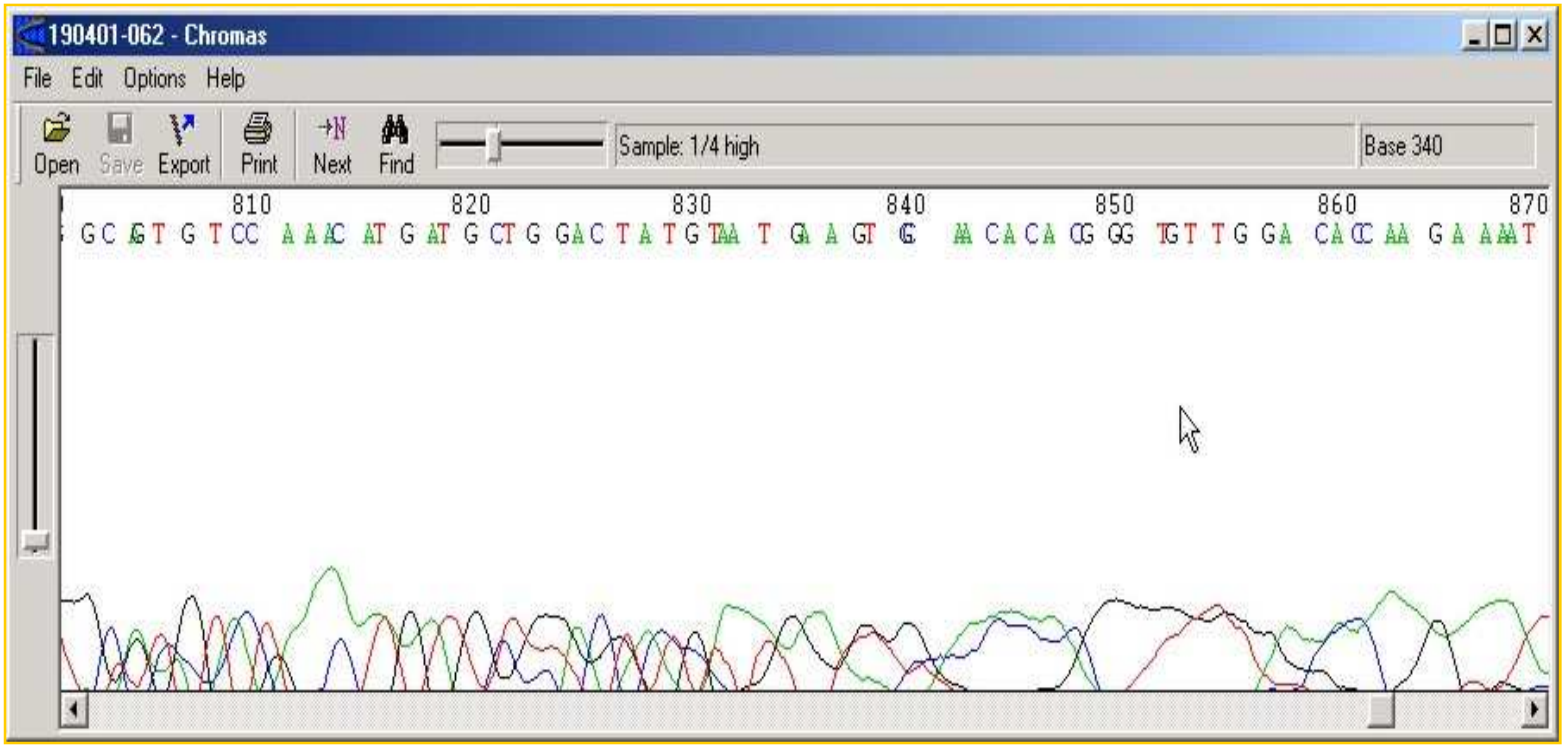
Analisando o cromatograma

Região de qualidade média – poucas ambigüidades

- Picos razoavelmente bem definidos e de tamanho médio.
- Linha de base boa a razoável.
- Distância entre picos anterior e posterior razoável.

Analisando o cromatograma

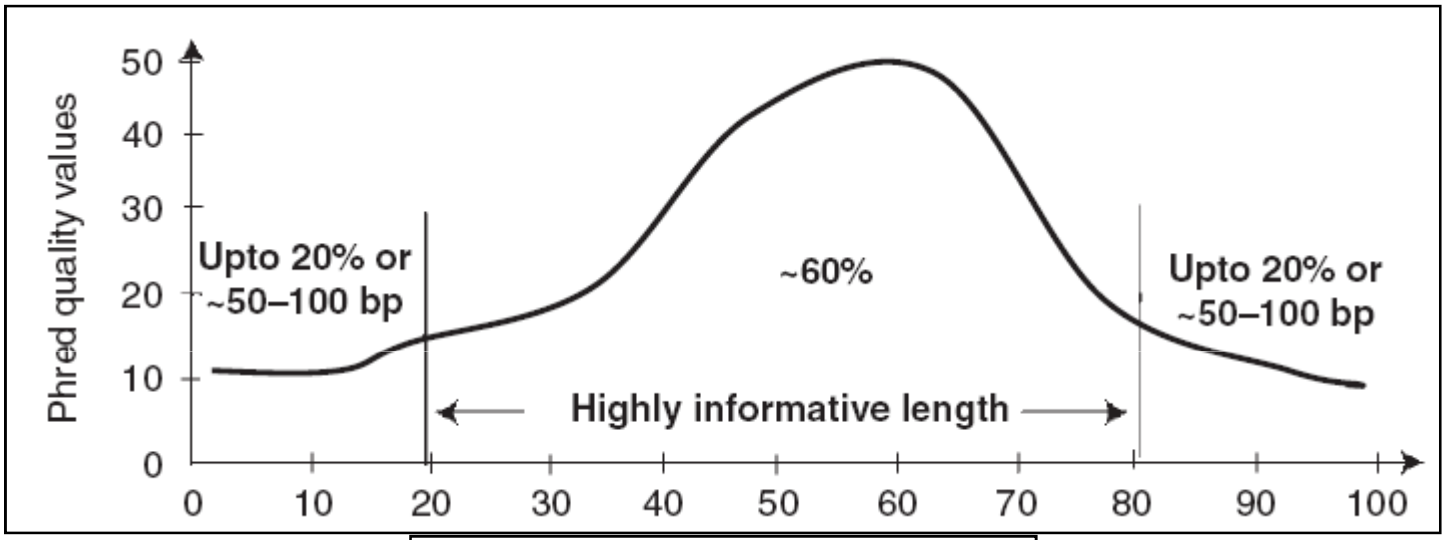
Região de qualidade baixa – baixa confiabilidade



- Picos mal definidos e de tamanho pequeno.
- Linha de base confusa.
- Distância entre picos anterior e posterior inconstante.

Analisando o cromatograma

Sequenciamento produz seqüências da ordem de 500 pb

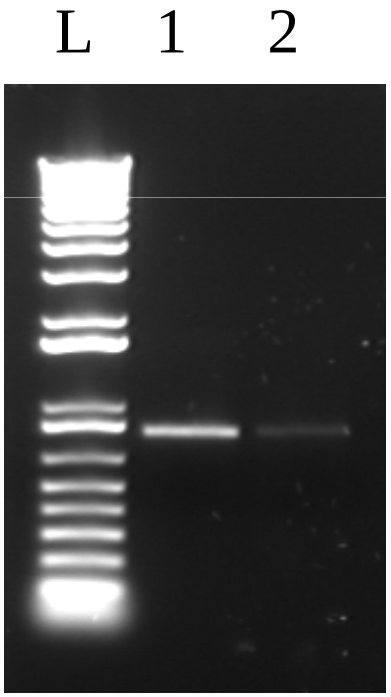


$$q = -10 \times \log_{10}(P)$$

Onde q é a nota phred e P é a probabilidade encontrar uma base errada :

- Nota phred = 20 => 1 base errada a cada 100 (99%)
- Nota phred = 30 => 1 base errada a cada 1000 (99.9%)

PCR sequenciamento



100 - 400 ng DNA / reação

L – 1Kb plus

1 – PCR purificada em coluna (conc. adequada para MegaBACE)

2 – PCR purificada em coluna (pouco DNA para sequenciar no MegaBACE)

Gel de agarose 1%, 3ul de PCR

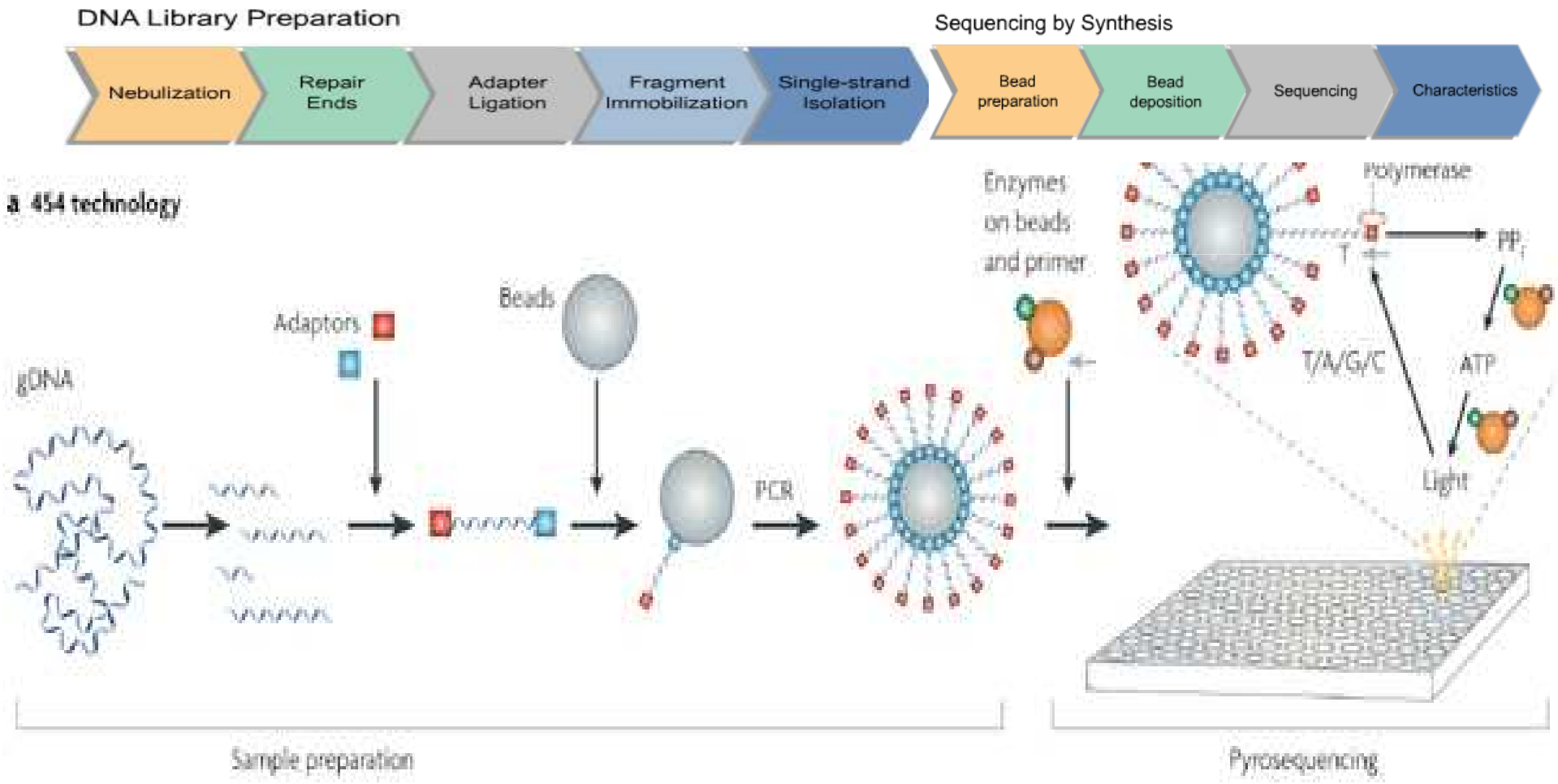
Novas Metodologias de Seqüenciamento

Pirosequenciamento

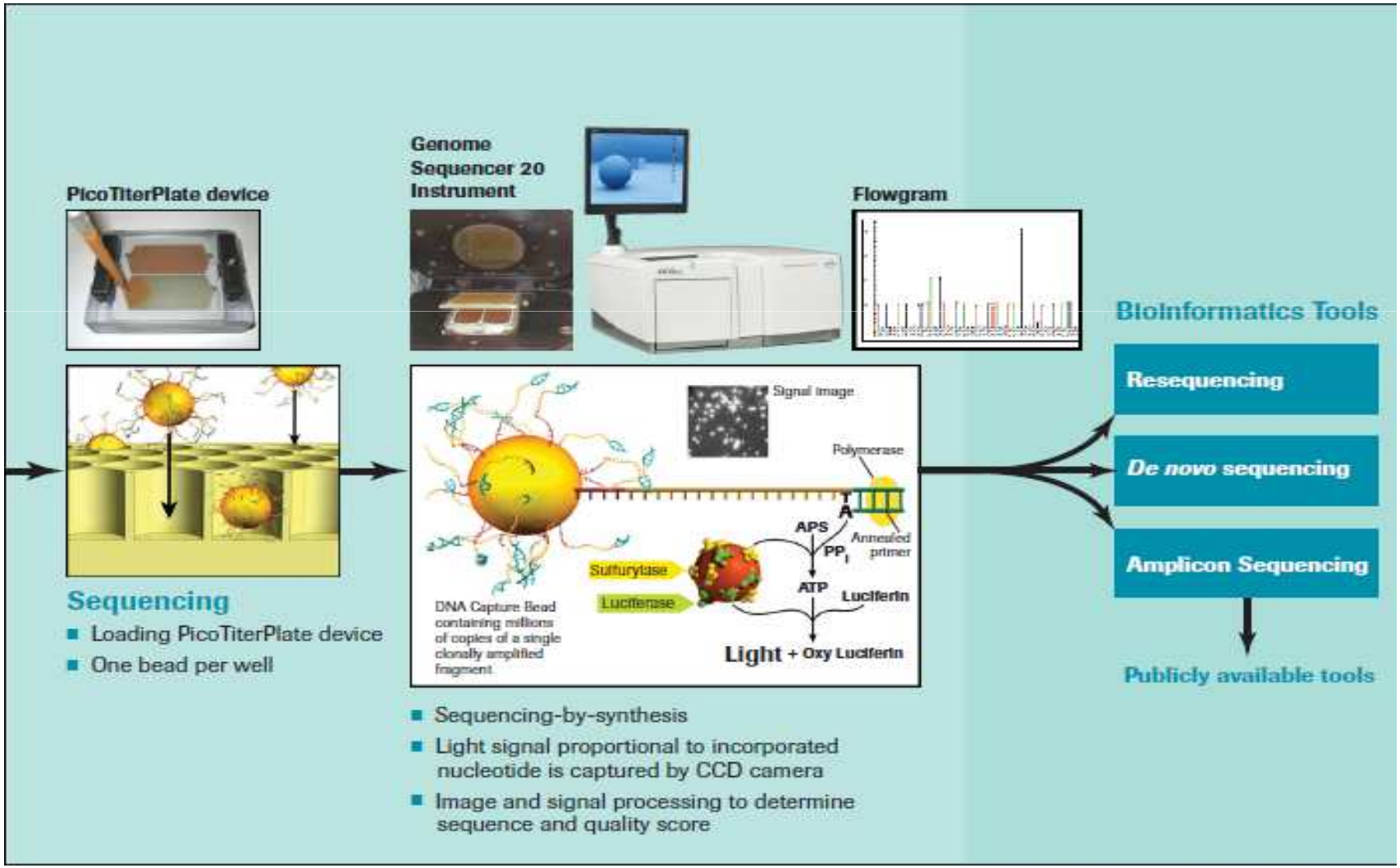
Pirosequenciamento

- Ronaghi *et al.*, 1996, 1998.
- Monitoramento em tempo real da síntese de DNA:
  - Síntese com primers pela DNA polimerase;
  - A incorporação é monitorada pela detecção luminosa da liberação do PPi, em uma reação envolvendo a enzima Luciferase;
  - Complexo com 4 Enzimas incluídas:
    - Fragmento Klenow da DNA Polimerase I;
    - ATP Sulfurilase;
    - Luciferase;
    - Apyrase.
  - Os Substratos:
    - Fosfosulfato de adenosina (APS);
    - D-Luciferina;
    - O DNA molde de seqüenciamento com um *primer* anelado.
  - Os 4 nucleotídeos são adicionados, um de cada vez, iterativamente, em uma maneira cíclica.

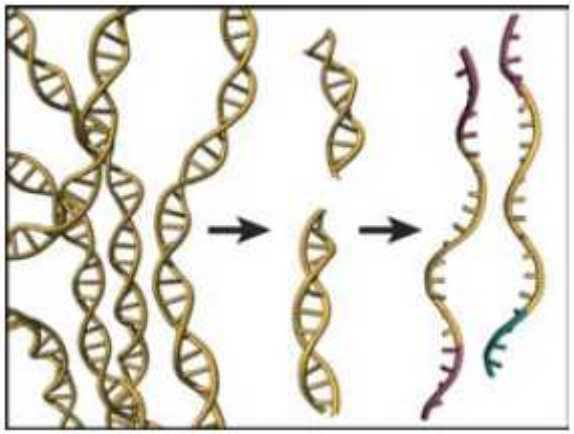
Sequenciamento- 454 Roche



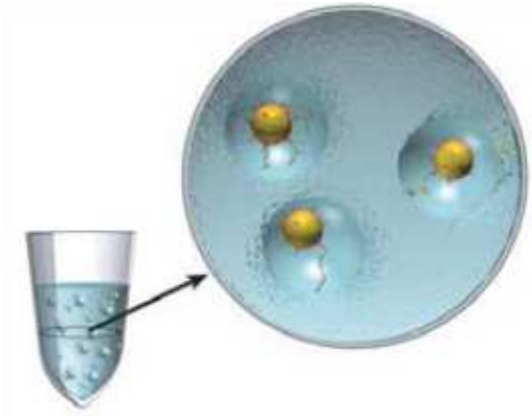
Sequenciamento- 454 Roche



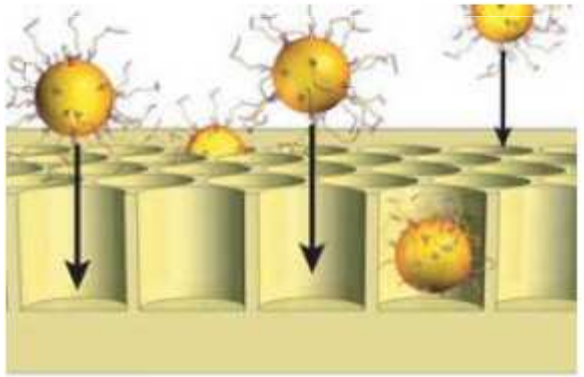
Seqüenciamento 454



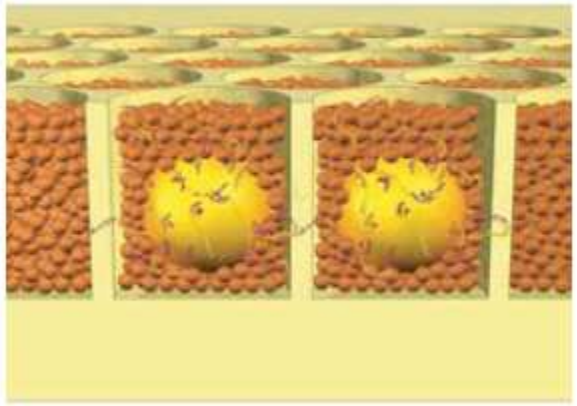
DNA genômico é isolado, fragmentado, ligado a adaptadores e separado em ssDNA



Ligação dos fragmentos a pequenas esferas (*beads*), e realização de uma PCR em emulsão, resultando em ≅ 10 milhões de cópias de DNA molde em cada *bead*.

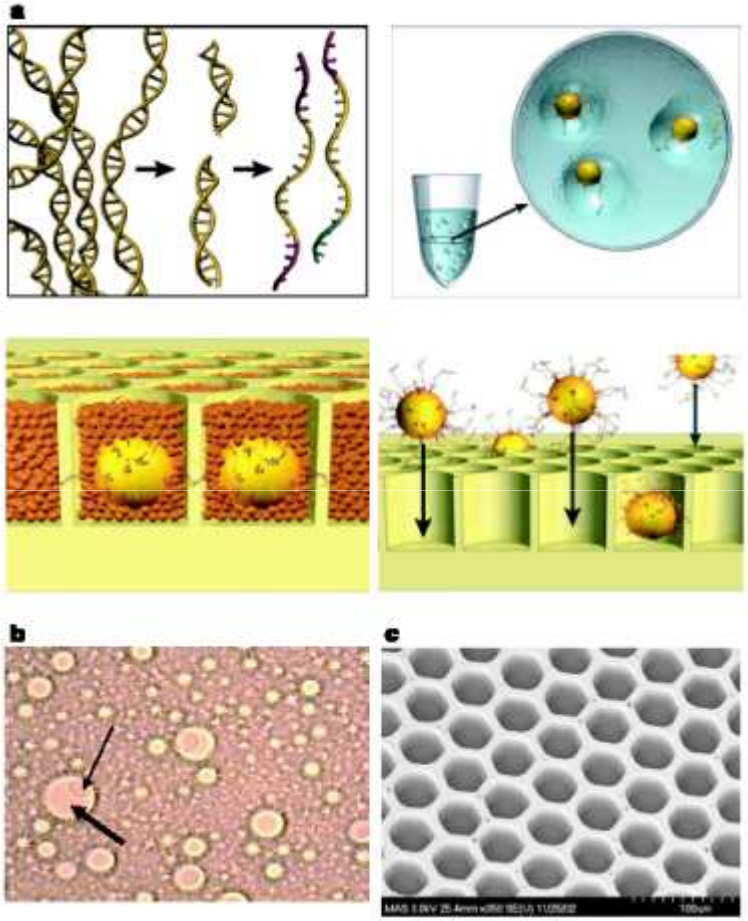


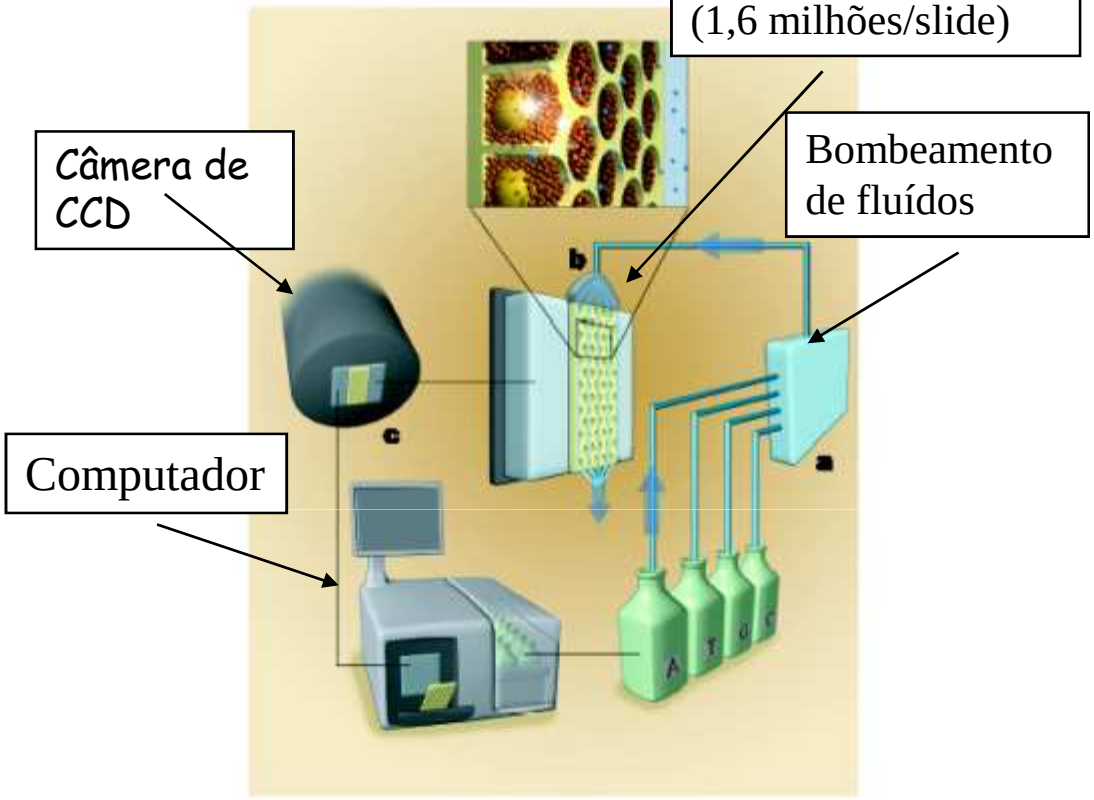
Quebra da emulsão, desnaturação das fitas de DNA e deposição das *beads* em slide de fibra ótica



Pequenas *Beads*, com as enzimas necessárias ao piroseqüenciamento imobilizadas, são depositadas em cada poço.

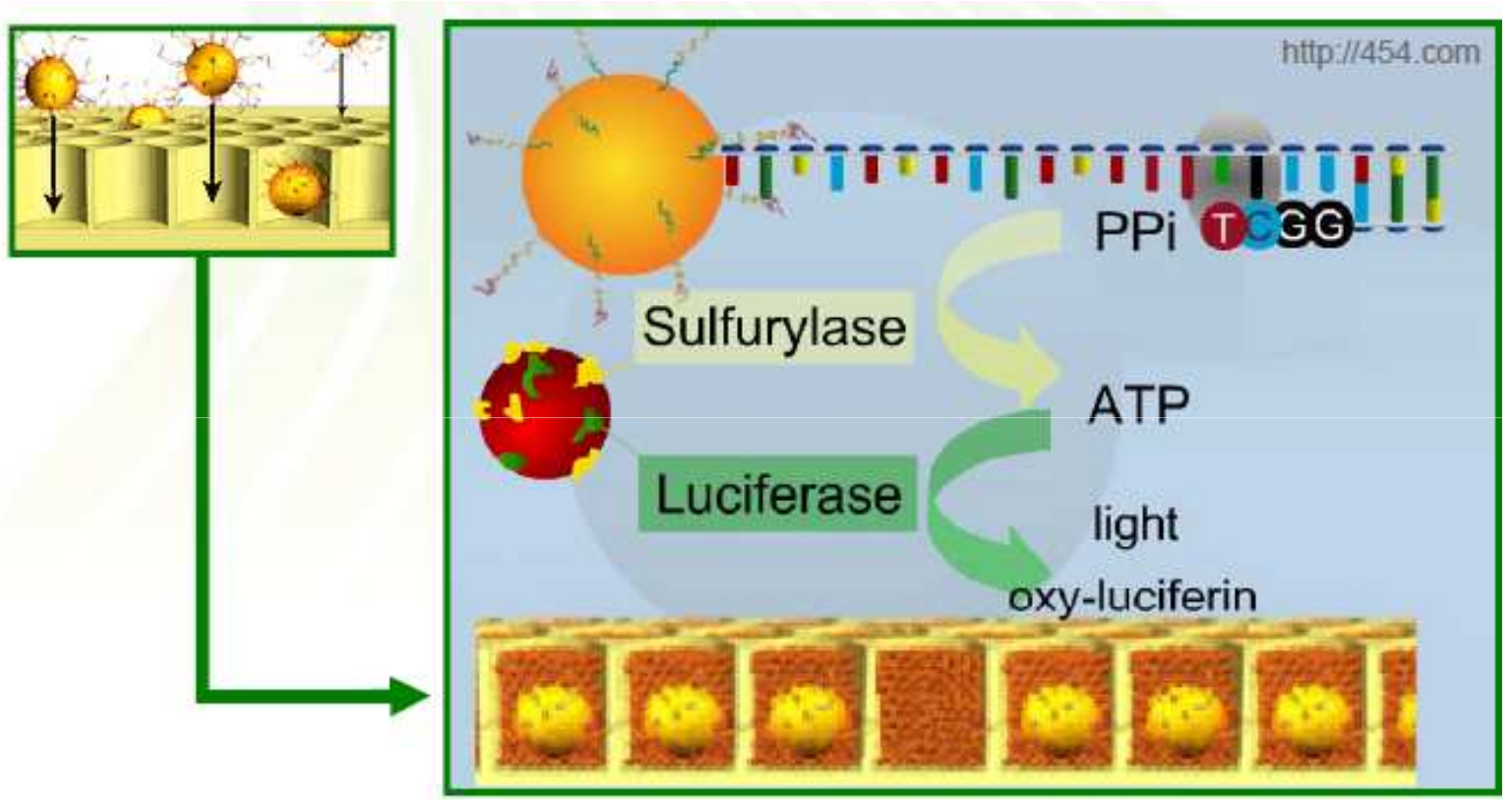
Seqüenciamento 454





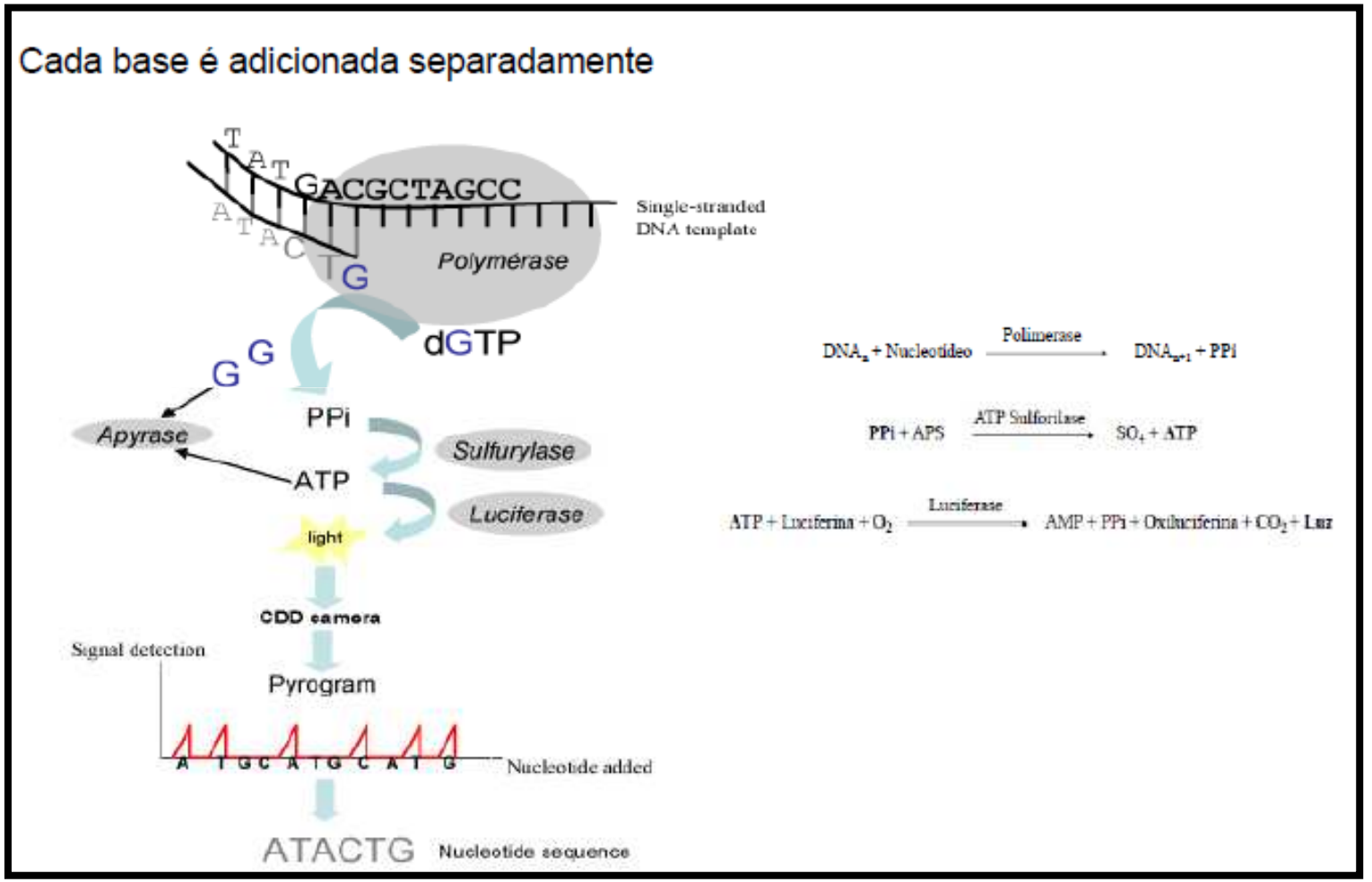
Um pmol de DNA numa reação de pirosequenciamento produz 10^{11} moléculas de ATP gerando mais de 10^{9} fótons, com comprimento de onda de 560 nm, em um período de 3-4 segundos. Facilmente detectado por uma câmera de CCD.

Pirosequenciamento



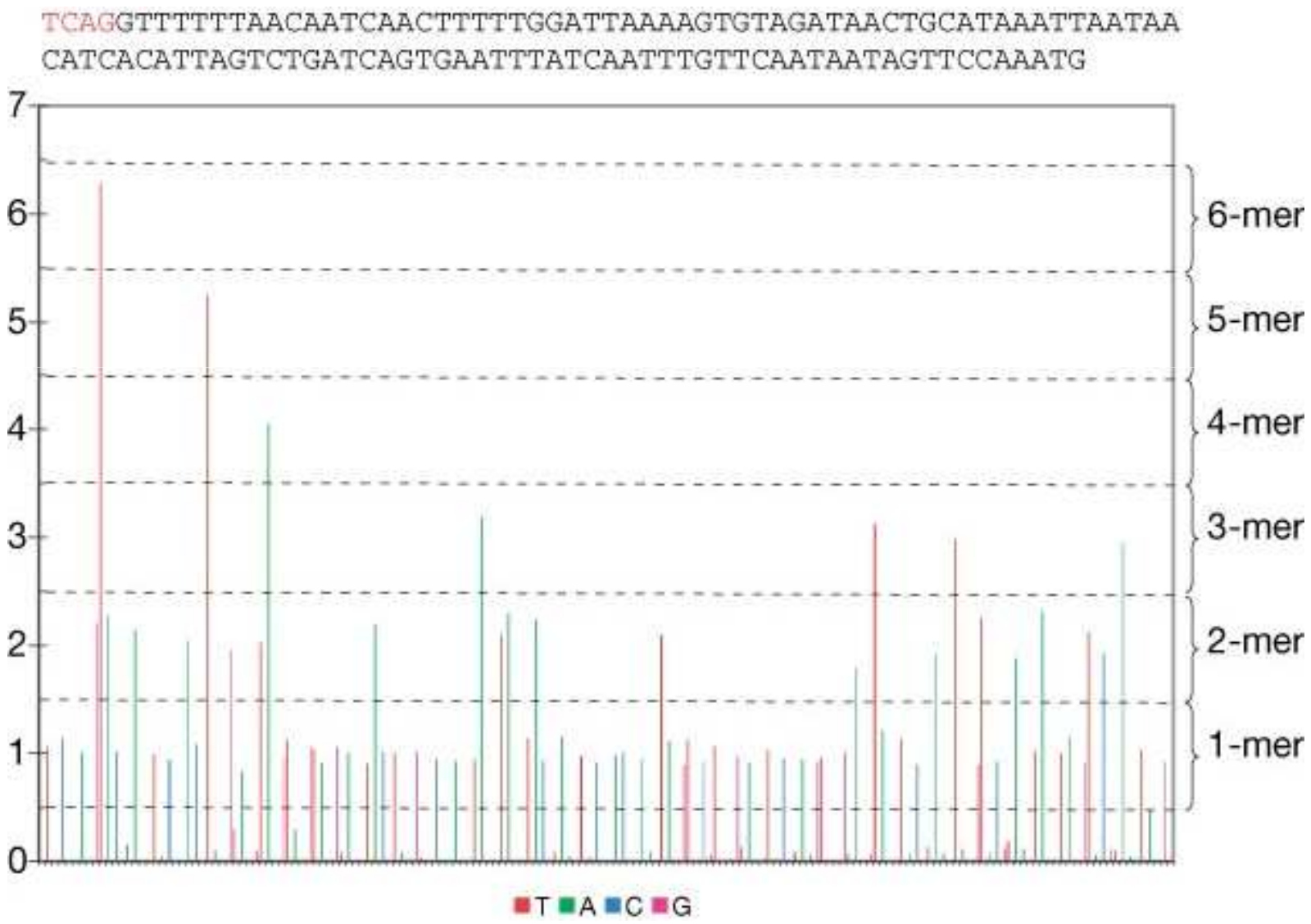
http://www.roche-applied-science.com/publications/multimedia/genome_sequencer/flx_multimedia/wbt.htm

Pirosequenciamento



http://www.roche-applied-science.com/publications/multimedia/genome_sequencer/flx_multimedia/wbt.htm

Pirograma


TCAGGTTTTTTAACAATCAACTTTTTGGATTAAAAGTGTAGATAACTGCATAAATTAATAA
CATCACATTAGTCTGATCAGTGAATTTATCAATTTGTTCAATAATAGTTCCAAATG
7
6
5
4
3
2
1
0
6-mer
5-mer
4-mer
3-mer
2-mer
1-mer
T A C G

Sanger vs Pirosequenciamento

SANGER

- Depende de clonagem em bactéria (2 semanas de trabalho)
- 1 milhão de pb em 24 horas
- Reads de ~700 bp
- Clones de fita dupla permitem seqüenciamento em ambas direções (facilita orientação e montagem)
- 6 meses de sequenciamento, 24 horas por dia, para sequenciar o genoma de um fungo

Outros tipos

- Não há clonagem
- 25 milhões de bp em 4 horas (100x mais rápido)
- Reads de ~100 bp
- Fragmentos fita simples não permitem seqüenciamento em ambas direções
- 24 horas para sequenciar o genoma de um fungo

Conclusão : a união faz a força

PNAS 103 (2006), 11240

Animações

- http://www.dnalc.org/ddnalc/resources/sangerseq.html
- http://www.dnalc.org/ddnalc/resources/cycseq.html
- http://www.biomolweb.kit.net/pages_html/sequencing.html