

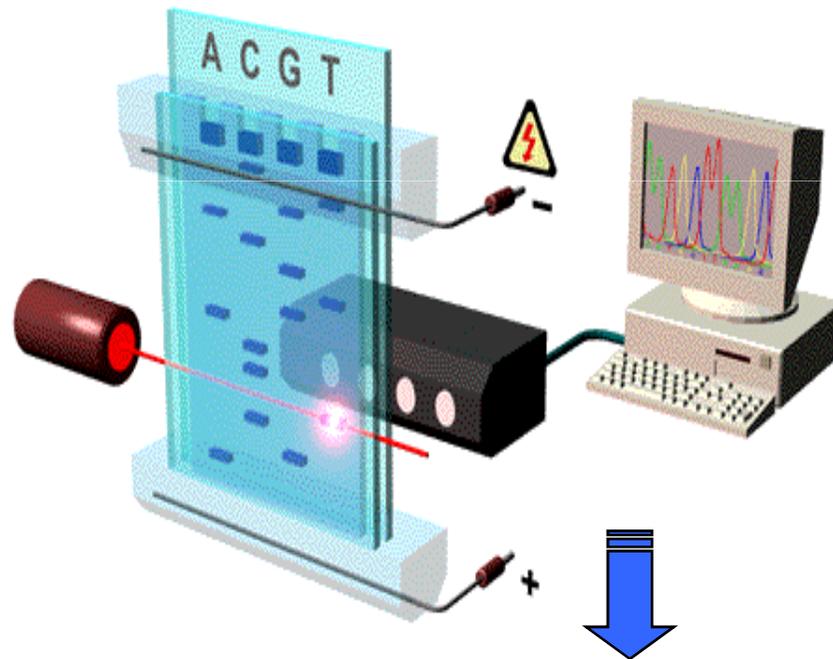


Universidade Tiradentes
Mestrado em Biotecnologia Industrial

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).

As moléculas também.

Aligning

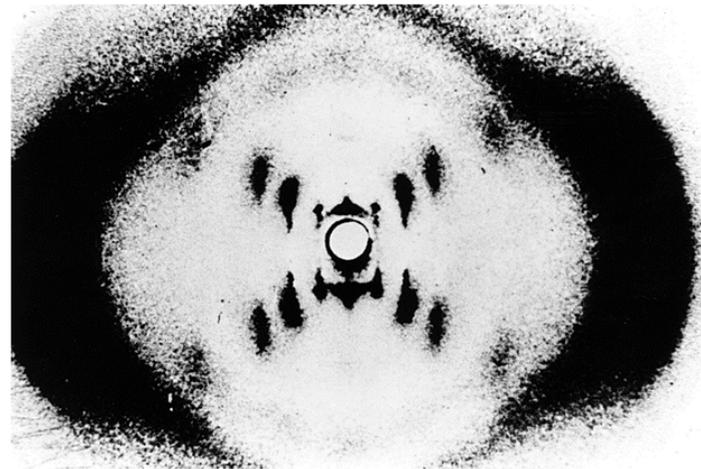
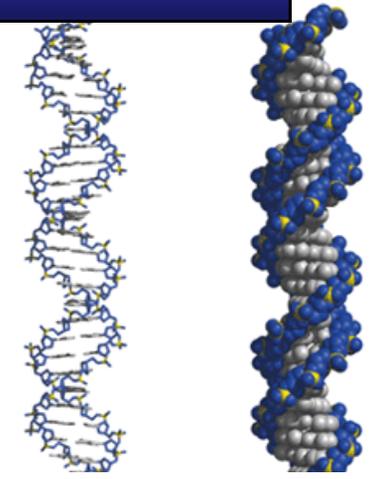
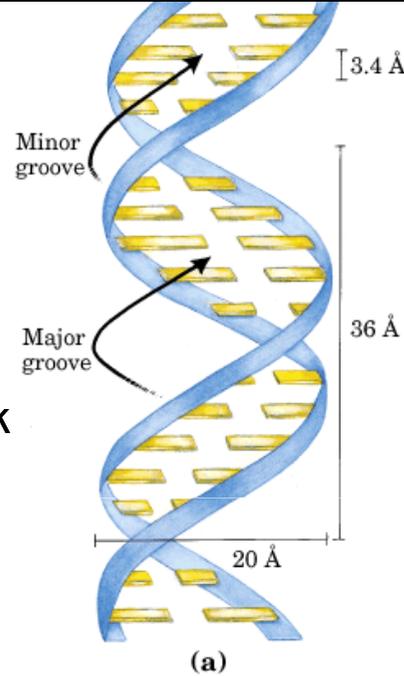
	260	*	280	*	300	*	320	
species 1	TCAAAGATTAAGC	CAT3CATGTCAAAGT	ACAAGCCCACTA	A-AG	G7GA	AAACC3CAATGGCTCAITAAA	TCA	
species 2	TCAAAGATTAAGC	CAT3CATGTCTAAGT	ACAATCCTCTTGA	GG	GAGA	AACT3C3AAAGGCTCAITAAA	TCA	
species 3	TCAAAGATTAAGC	CAT3CATGTCAAAGT	ACAAGCCCACTA	A-AG	G7GA	AAACC3CAATGGCTCAITAAA	TCA	
species 4	TCAAAGATTAAGC	CAT3CATGTCTAAGT	ACAGNCCG	ATCT	AAG	GCGA	AAACC3CAATGGCTCAITAAA	TCA
species 5	TCAAAGATTAAGC	CAT3CATGTCTAAGT	ACAGGCCG	ATCT	AAG	GCGA	AAACC3CAATGGCTCAITAAA	TCA
species 6	TCAAAGATTAAGC	CAT3CATGTCTAAGT	ACAGGCCG	AACT	AAG	GCGA	AAACC3CAATGGCTCAITAAA	TCA
species 7	TCGTTGTCTCGTT3CCT3C	T3TCTAAGT	ACAAGCCG	ATTC	AAG	GCGA	AAACC3CAATGGCTCAITAAA	TCA
species 8	TCAAAGATTAAGC	CAT3CATGTCTAAGT	ACAAGCCG	ATTT	AAG	GCGA	AAACC3CAATGGCTCAITAAA	TCA
species 9	TCAAAGATTAAGC	CAT3CATGTGTAAGT	ACAAGCCG	ATGT	AAG	G7GA	AAACC3CAATGGCTCAITAAA	TCA
species 10	TCAAAGATTAAGC	CAT3CATGTCTNNGT	ACA---	CCTCTG	GG	GCGA	AAACC3CAATGGCTCAATAAAA	TCA
species 11	TCAAAGATTAAGC	CAT3CATGTCTAAGT	ACAAGCGCTATG	CG	GCGA	AAACC3CAATGGCTCAITAAA	TCA	
species 12	TCAAAGATTAAGC	CAT3CATGTCTAAGT	ACAAGCCGCTAGA	CG	GCGA	AAACC3CAATGGCTCAATAAA	TCA	
species 13	TCAAAGATTAAGC	CAT3CAGGTCTAAGT	ATAAGCCGAAATA	AA	G7GA	GACC3C3AATGGCTCAITACA	TCA	
species 14	TCAAAGATTAAGC	CAT3CAGGTCTAAGT	ACGAGCCGAAATA	AAT	G7GA	GACC3C3AATGGCTCAITACA	TCA	
species 15	TCAAAGATTAAGC	CAT3CAGGTCTAAGT	ACATGCTCTTATA	TATGGTAA	GACT3C3AACGGCTCAITACA	TCC		
species 16	TCAAAGATTAAGC	CAT3CATCTCTAAGT	ACACACCAAATTA	AG	G7GA	AAACC3CAATGGCTCAITAAA	TCA	
species 17	TCAAAGATTAAGC	CAT3CATGTCTAAGT	ACAAGCCCTACAA	GG	CTGA	AAACC3CAATGGCTCAITAAA	TCA	
species 18	TCAAAGATTAAGC	CAT3CATGTCTAAGT	ACATGCCGCAITTA	A-AG	GCGA	AAACC3CAATGGCTCAITAAA	TCA	
species 19	TCAAAGATTAAGC	CAT3CATGTCTAAGT	ACATGCCGCAITTA	A-AG	G7GA	AAACC3CAATGGCTCAITAAA	TCA	
species 20	TCAGAGATTAAGC	CAT3CATGTCTAAGT	ACAGACCTTCATA	CG	G7GA	AAACC3CAATGGCTCAITAAA	TCA	
species 21	TCAAAGATTAAGC	CAT3CATGTCTAAGA	TCA	AGCTCGTCT	CG	GCGA	AACT3C3GATGGCTCAITAAA	TCA
species 22	TCAAAGATTAAGC	CAT3CANGTATCAGT	ACAAGCCCTCACTN	AG	G7GA	AAACC3CAATGGCTCAITAAA	TCA	
species 23	TCAAAGATTAAGC	CAACTCATGTCTAAGA	TCATGCCGAAACC	AAG	GCGA	AAACC3CAATGGCTCAITAAA	TCA	

(Andy Vierstraete 1999)

Histórico – O sequenciamento de DNA no tempo



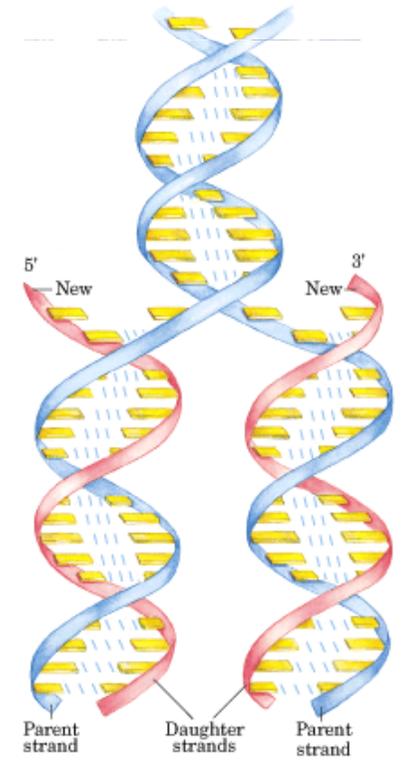
Watson & Crick
Nobel 1962



Rosalind Franklin,
1920-1958



Maurice Wilkins



Histórico – O sequenciamento de DNA no tempo



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

F. SANGER AND A. R. COCHRAN

Medical Research Council
Laboratory of Molecular Biology
Hills Road, Cambridge CB2 2QJ, 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 9, 1976

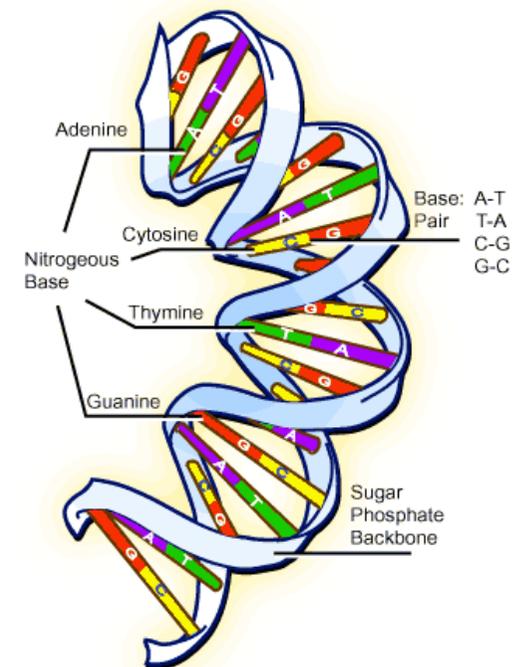
Walter Gilbert

Prêmio Nobel de medicina e fisiologia em 1980

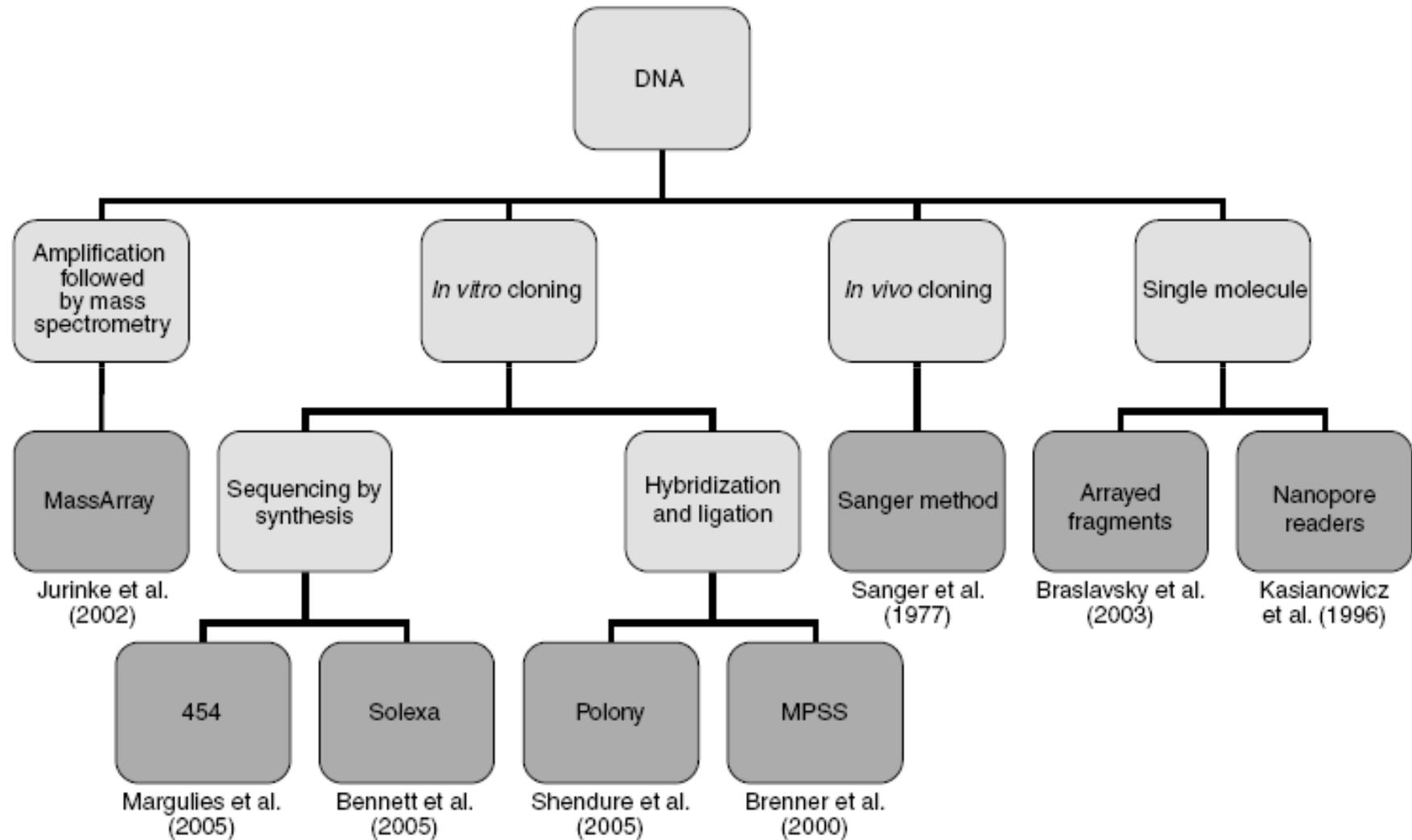
PNAS, vol. 74 No. 2 p. 560-564, 1977

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)



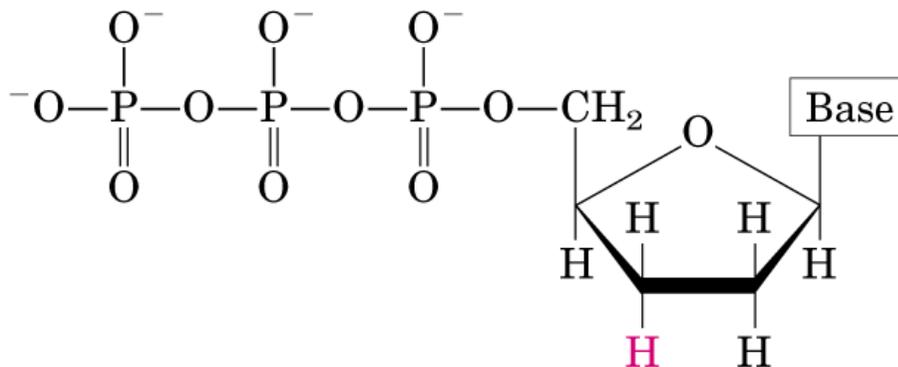
Métodos de seqüenciamento



Método "dideoxi" de F. Sanger

- **Dezembro de 1977:**

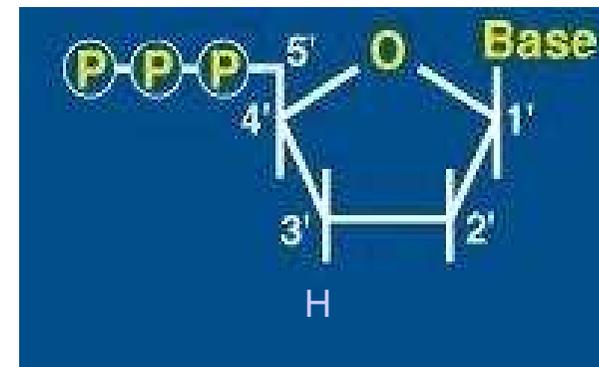
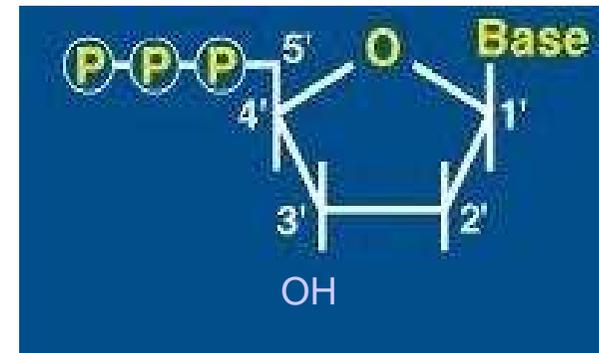
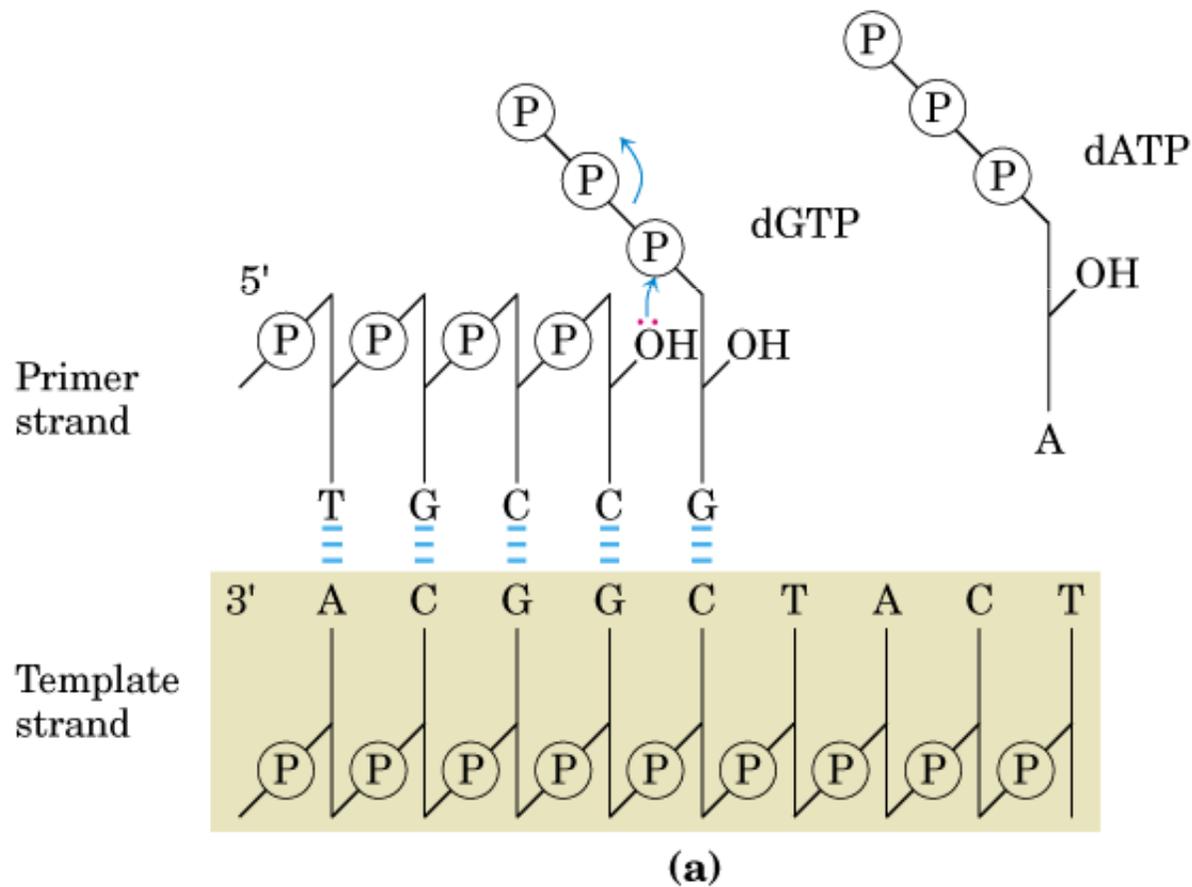
- Também chamado de Método de Terminação da Cadeia;
- Baseado na utilização de um análogo ao dNTPs, o ddNTP:



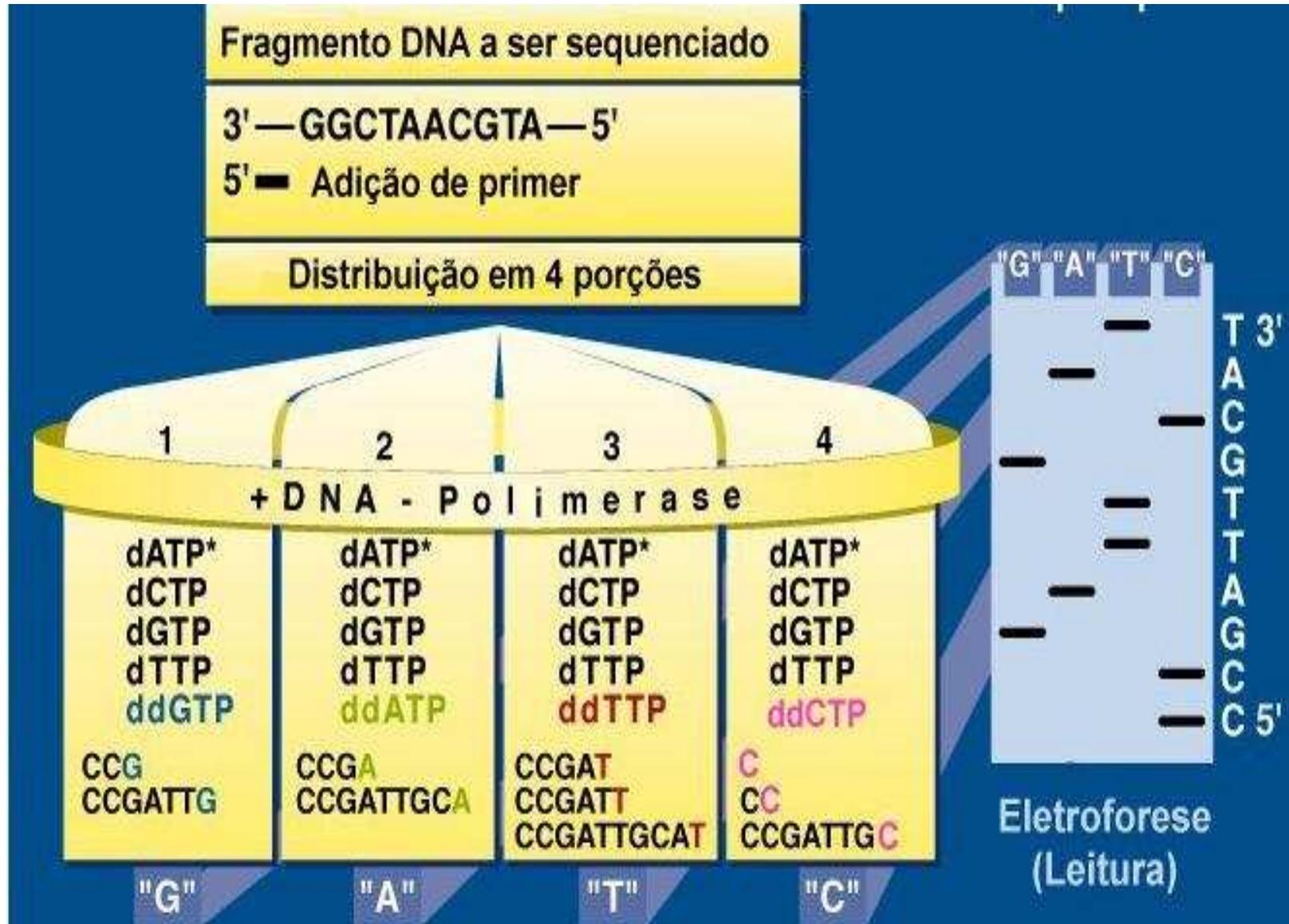
ddNTP analog



Reação de polimerização



Reação de seqüenciamento com radioisotopos

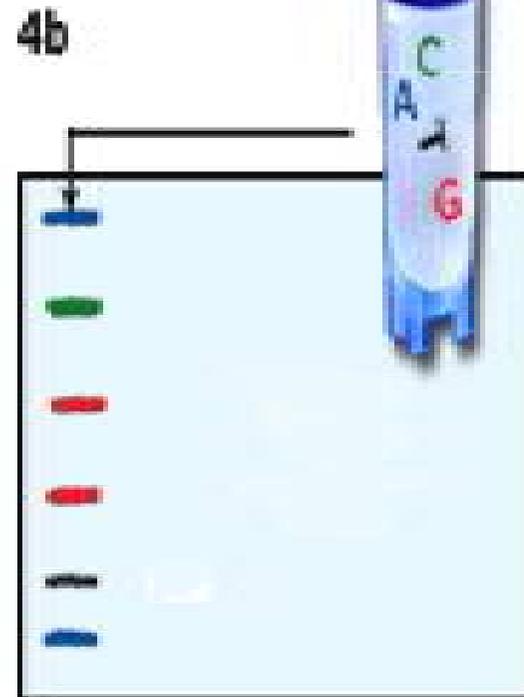
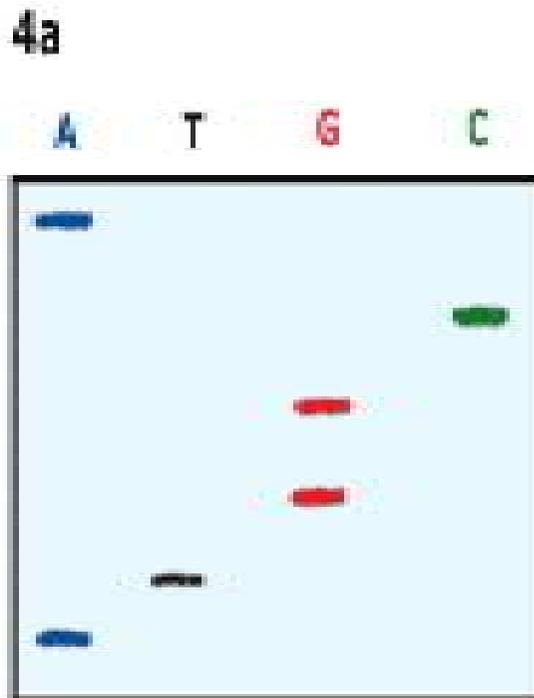


Reação de sequenciamento e marcação do DNA

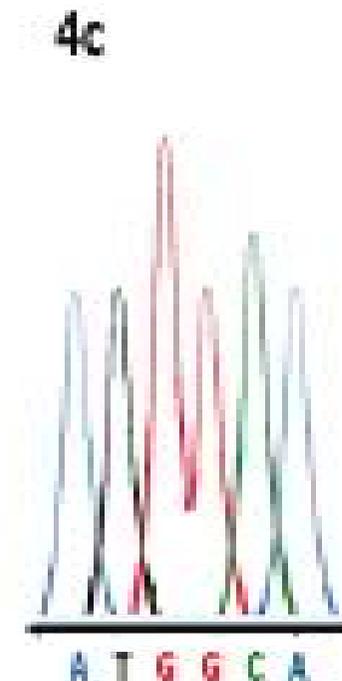
Leitura das sequências -
Manual

Leitura das sequências -
Automática

Radioisótopos

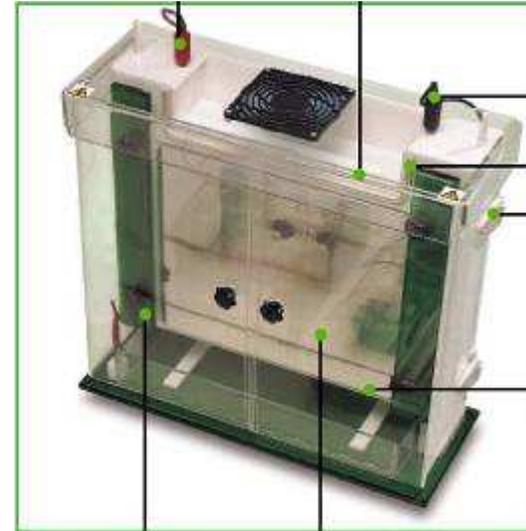


Fluoresceínas

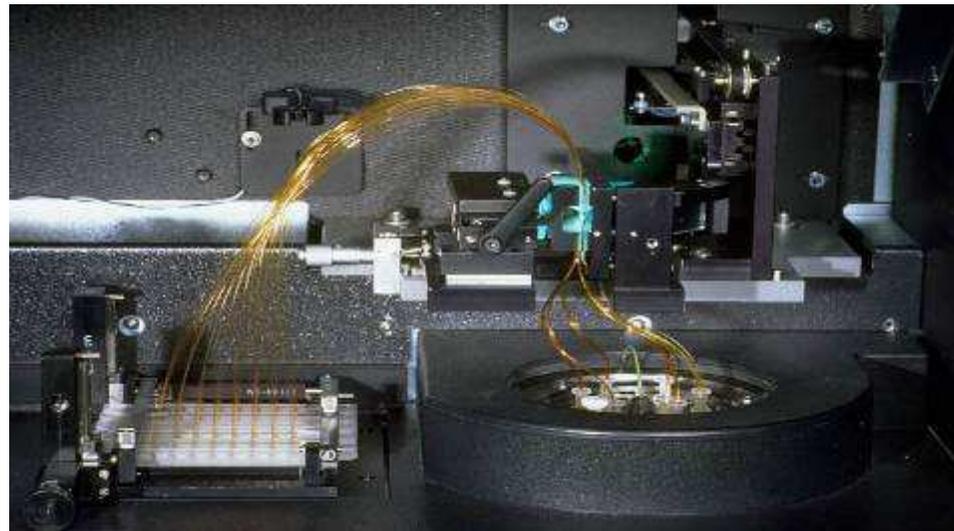


Eletroforese

- Placa



- Capilar



Reação de sequenciamento e leitura automatizada

Cycle Sequencing

The dideoxysequencing reaction is setup using the DNA template.

ACCTGTACTGGGCTAAG
TGGACATGAGCCGATTC

DNA TEMPLATE

TAQ POLYMERASE

TTTTT PRIMER

C A T G
T T T T DEOXYNUCLEOTIDES

C A T G
T T T T DIDEOXYNUCLEOTIDES

VIEW REACTION

QUIT

desnaturação

Cycle Sequencing

TTTTT

QUIT

PRIMERS ANNEAL

anelamento dos primers

Cycle Sequencing

QUIT

TAQ POLYMERASE BINDS T

Cycle Sequencing

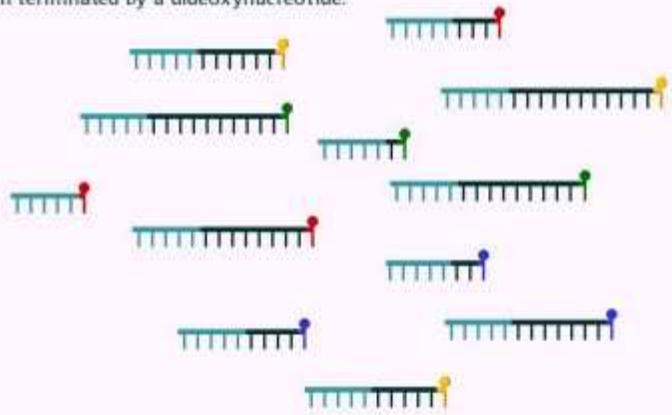
QUIT

Working from the primer, Taq polymerase randomly adds deoxynucleotides or dideoxynucleotides that are complementary to the DNA template. The new DNA strand is terminated by the addition of a dideoxynucleotide.

Cycle Sequencing

QUIT

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



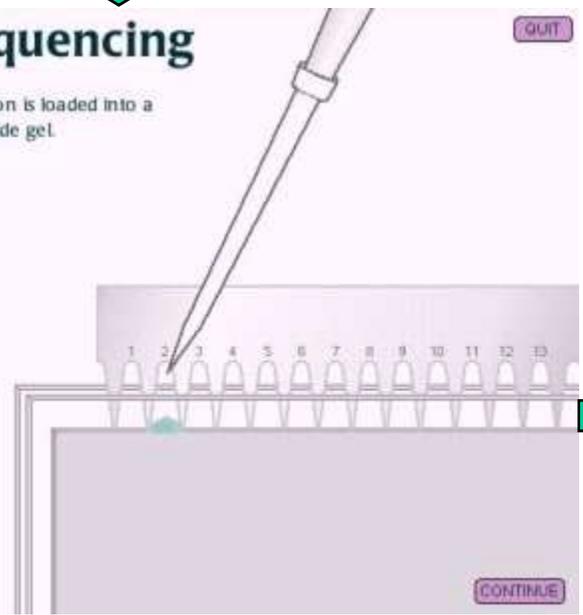
CONTINUE



Cycle Sequencing

QUIT

The sequencing reaction is loaded into a lane of a polyacrylamide gel.



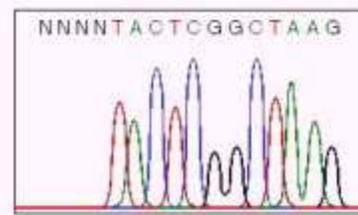
CONTINUE



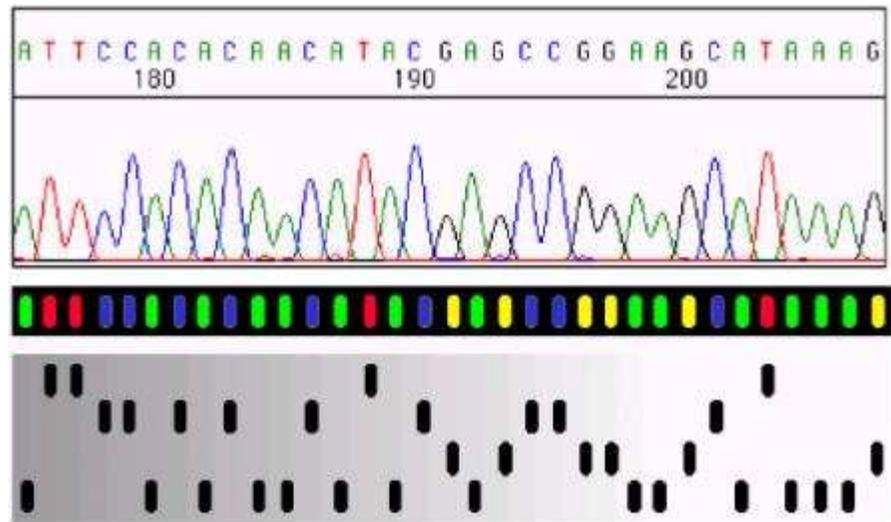
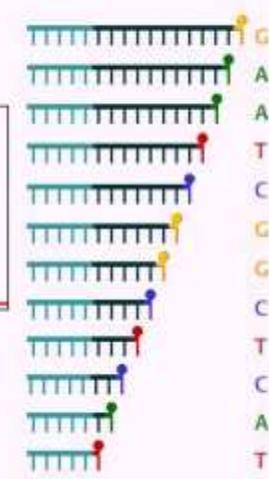
Cycle Sequencing

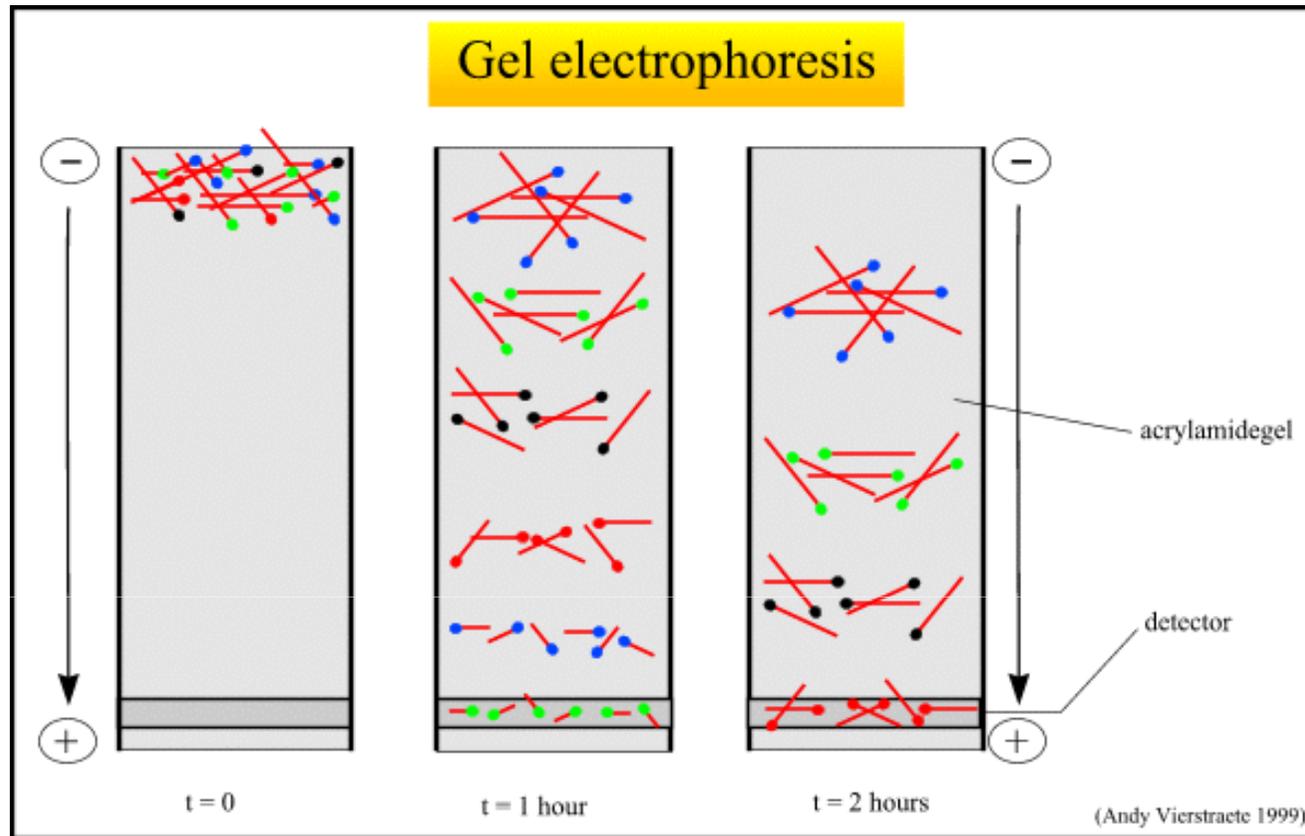
QUIT

The simulated gel image is read from bottom to top, starting with the smallest fragment.



GO TO START

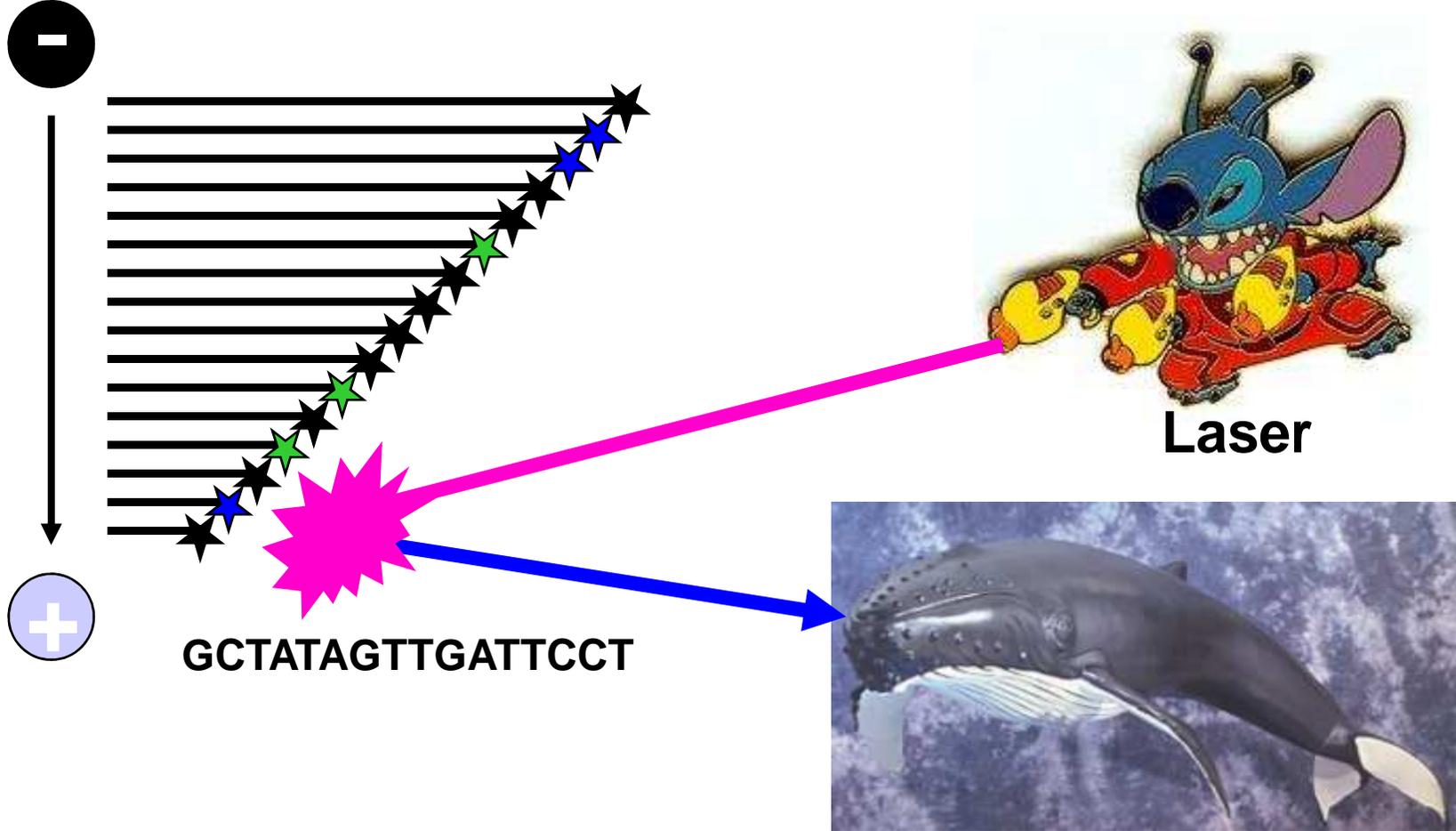




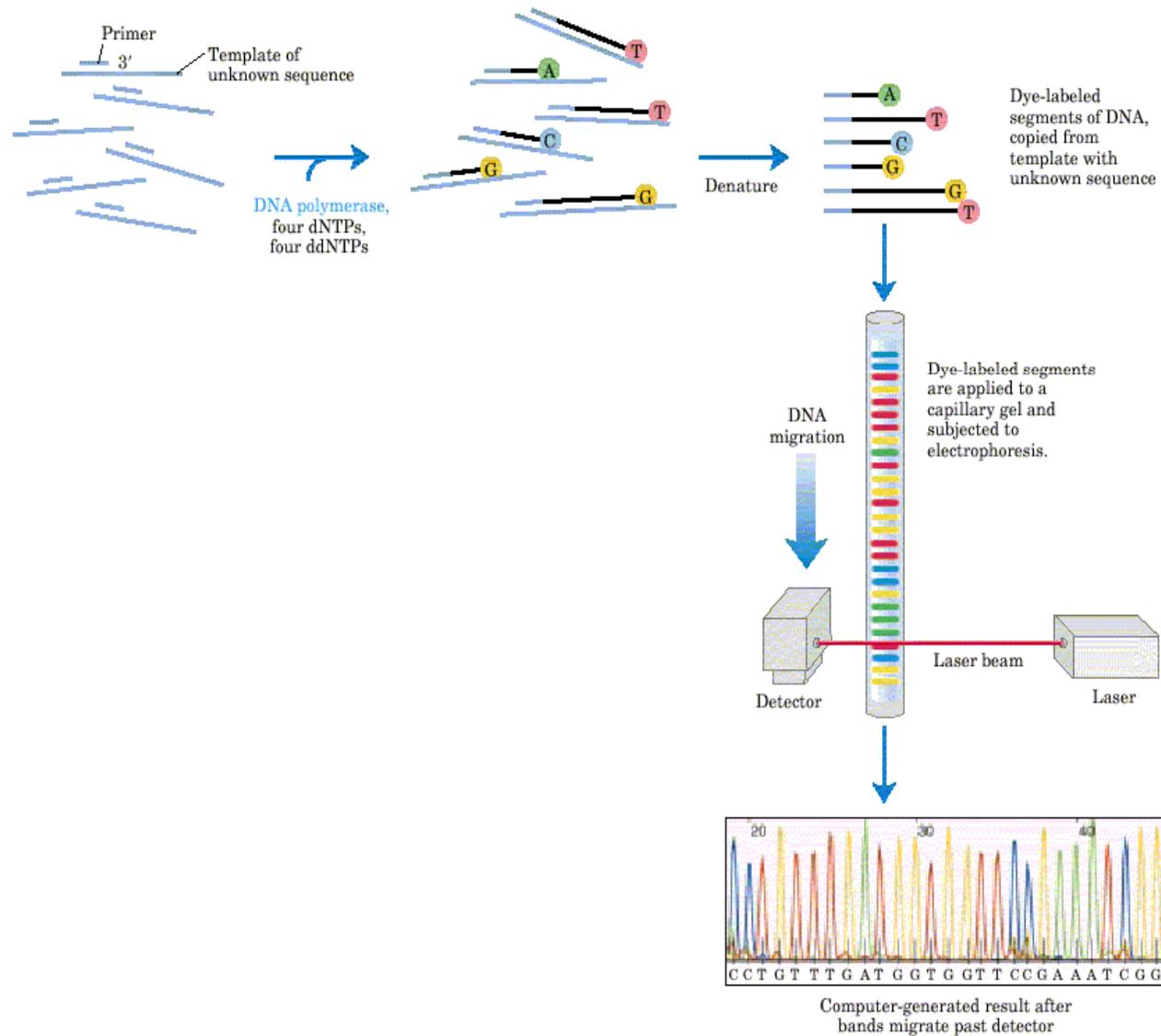
Exemplo de gel utilizado nos seqüenciadores de placa (ex.: 377). A diferença de tamanho permite a separação dos grupos de fragmentos, e esta “distribuição normal” da passagem dos fragmentos é representada pelo eletroferograma (ou cromatograma) de cada seqüência (read).

Generic Sequencing Instrument

Electrical Field



Sequenciamento- Automático



DYEnamic ET Terminators – Pré Mix

Energy Transfer Dyes

- Fluorescein Donor Dye
- Standard Rhodamine Acceptor Dyes

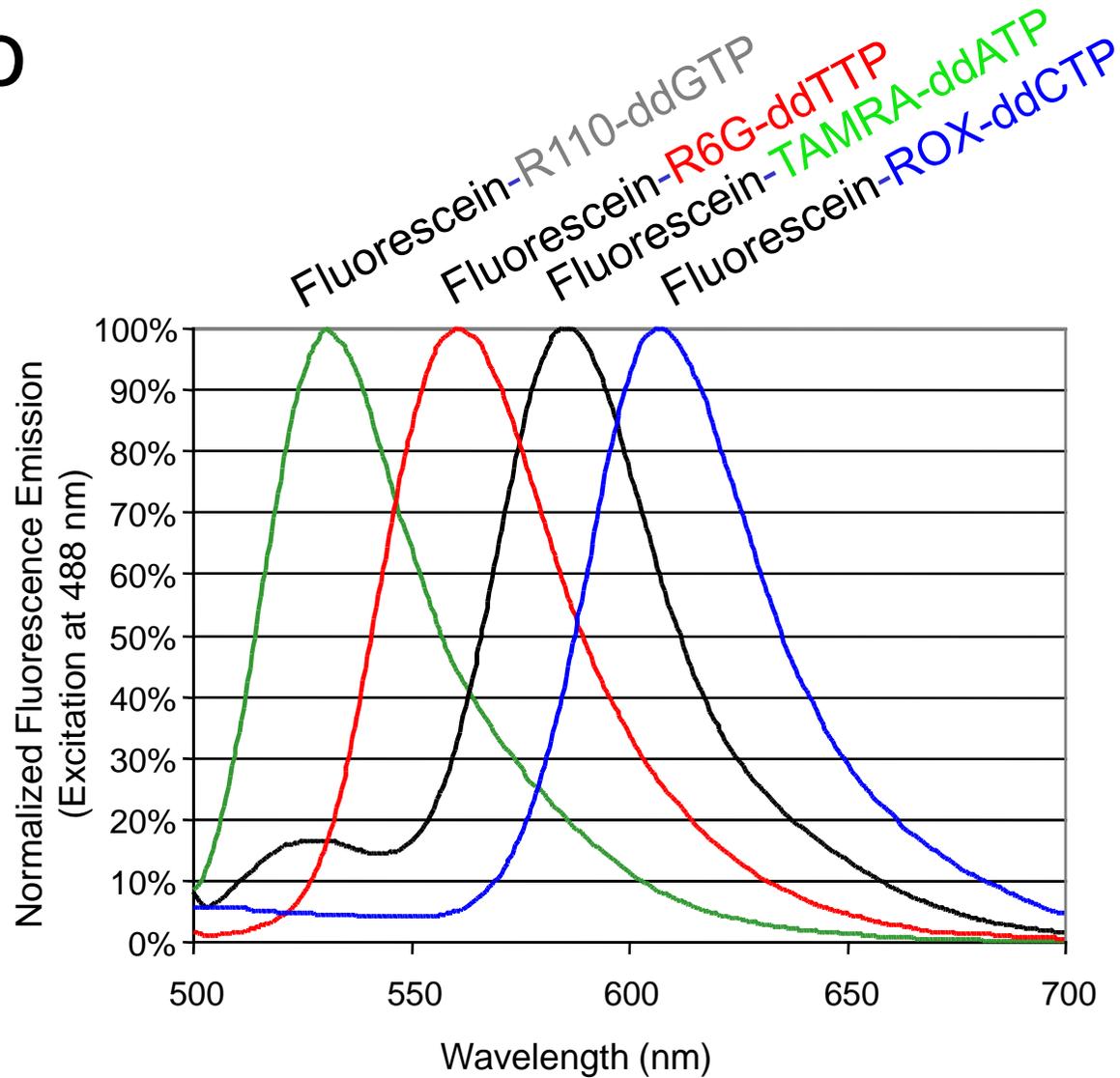
Fluorescein-R110-ddGTP

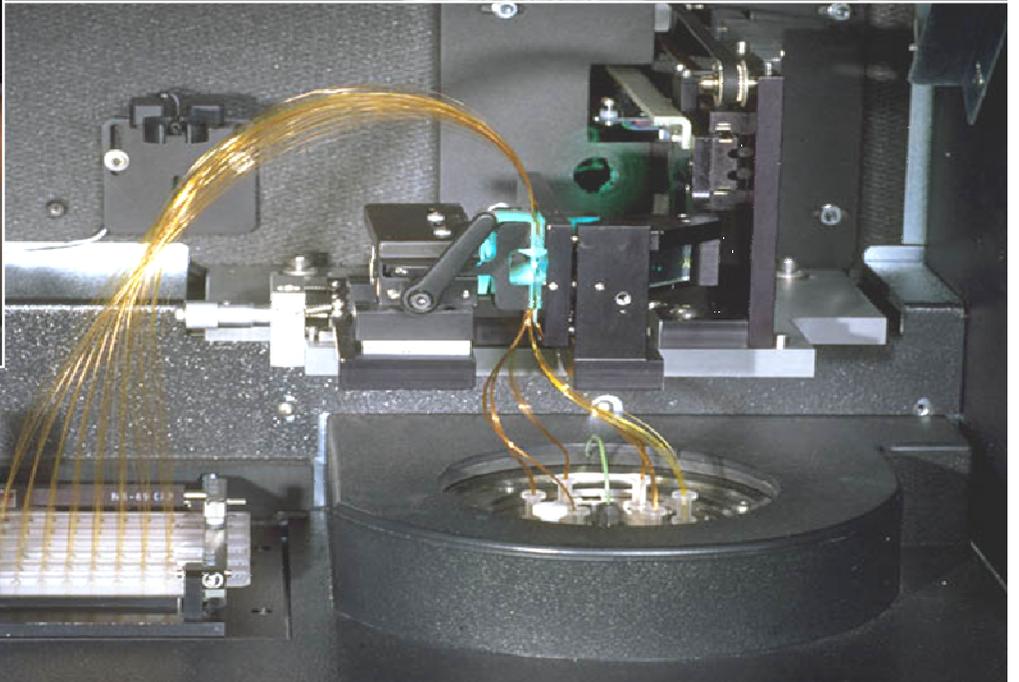
Fluorescein-R6G-ddTTP

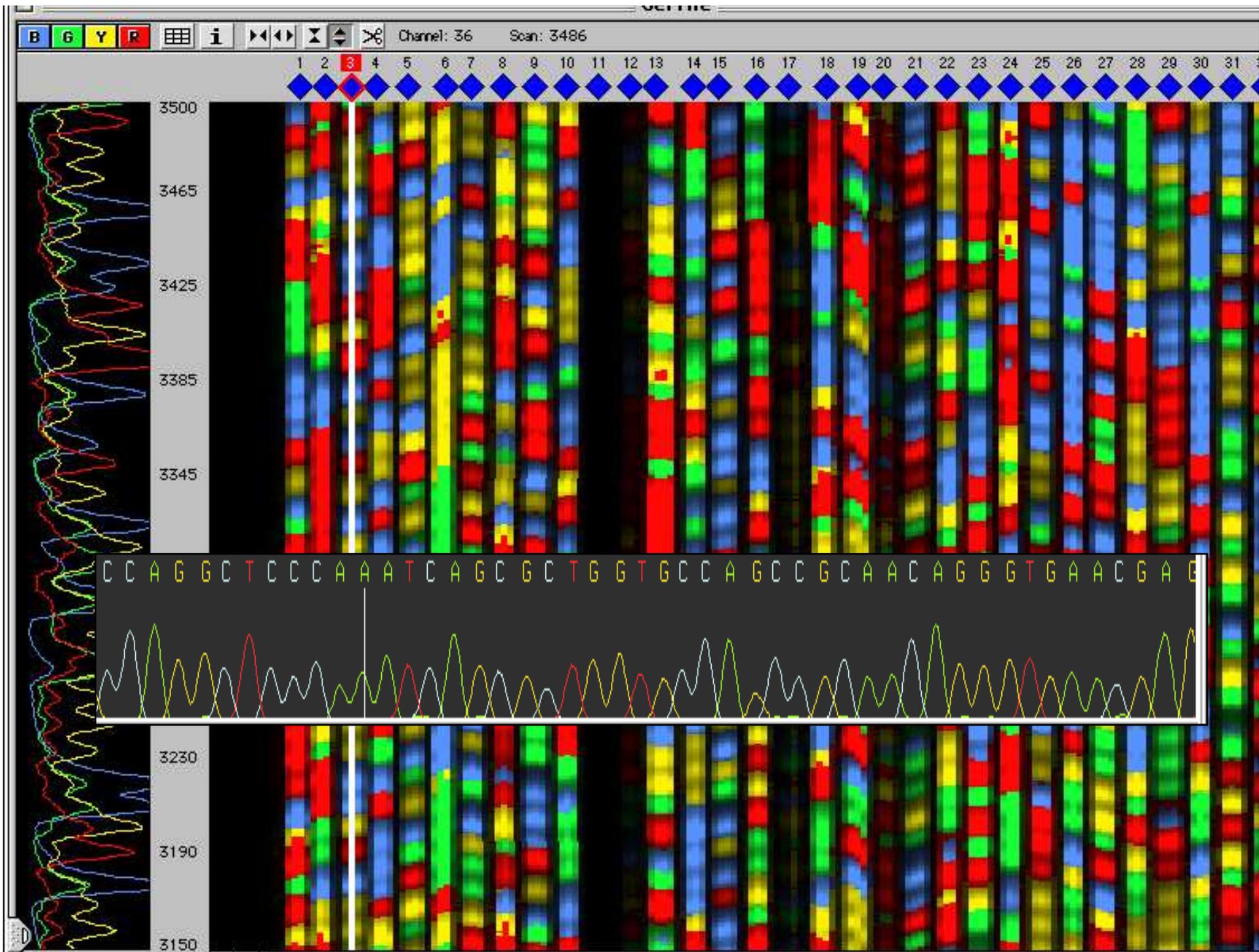
Fluorescein-TAMRA-ddATP

Fluorescein-ROX-ddCTP

DYEnamic ET terminator - espectro de emissão







Análise dos resultados por Bioinformática

Leitura da seqüência de DNA

Gel:

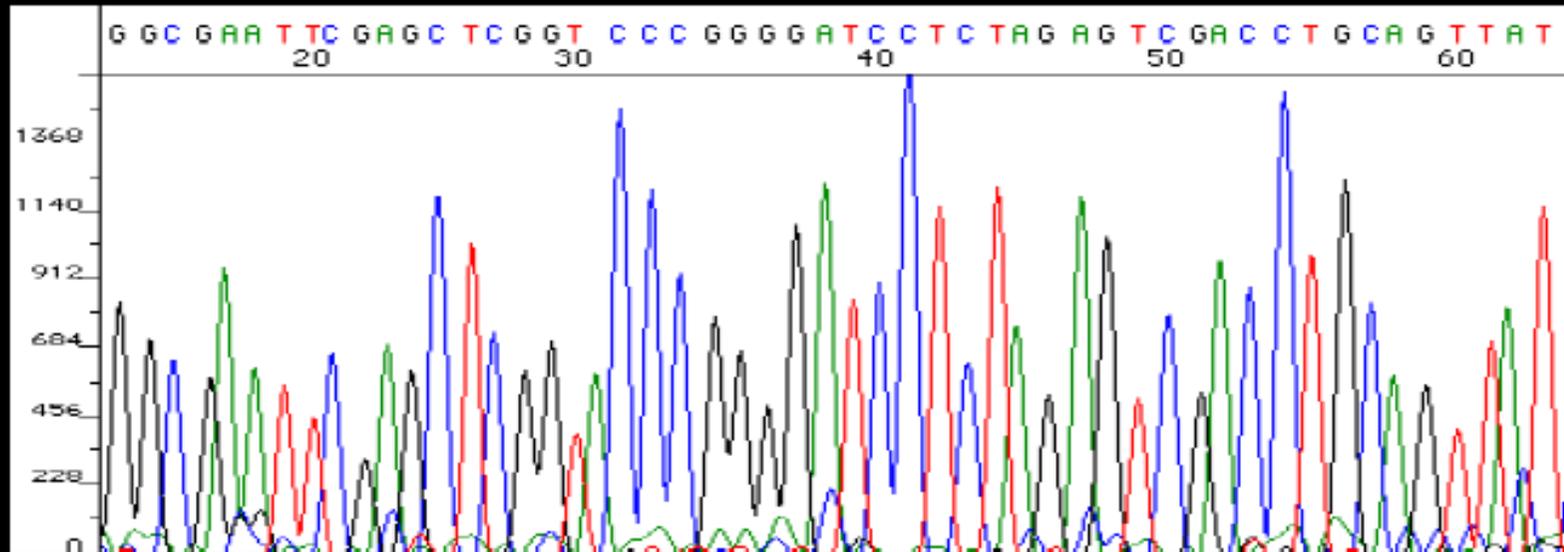
	G	GCGAATGCGTCCACACGCTACAGGTG
	T	GCGAATGCGTCCACACGCTACAGGT
	G	GCGAATGCGTCCACACGCTACAGG
	G	GCGAATGCGTCCACACGCTACAG
	A	GCGAATGCGTCCACACGCTACA
	C	GCGAATGCGTCCACACGCTAC
	A	GCGAATGCGTCCACACGCTA
	T	GCGAATGCGTCCACACGCT
	C	GCGAATGCGTCCACACGC
	G	GCGAATGCGTCCACACG
	C	GCGAATGCGTCCACAC
	A	GCGAATGCGTCCACAA
	A	GCGAATGCGTCCACA
	C	GCGAATGCGTCCAC
	A	GCGAATGCGTCCA
	C	GCGAATGCGTCC
	C	GCGAATGCGTC
	T	GCGAATGCGT
	G	GCGAATGCG
	C	GCGAATGC
	G	GCGAATG
	T	GCGAAT

Montagem

```
GCATCTTTGTCGGCGGGTGTCTGTGTCGATCGCCACCGTCCCGGCGAACGGCGCCGGCC*GCTCGCCAGGCT*CCCAAATCAGCGCTGGTGCCAGCCGCAACAGGGTG
t t t t t a a a c c a c c c c g t c a * c a c tctatacatccgg
gcat tttgtcggc ggtgtcg gt gatcgcn
gcAtctttgtc gcgggtgtcgtgtcgcgacccgctcCgggogaacgg gccggcc*g tcg caggct*ccca
gcATCtTgtcggcgggTGTCTGTGTCGATCGCCACCGTCCCgggogaacggcgCgggoc*gctcggcaggct cccaaat agcgctggtgocagccgcaacagggtg
GCATCTTTGTCGGCGGGTGTCTGTGTCGATCGCCACCGTCCCgggogaacggcgCgggoc*gctcggcaggct*cccaaatca cgCTggtgocagcc aacAggggtg
GCATCTTTGTCGGCGGGTGTCTGTGTCGATCGCCACCGTCCCgggogaacggcgCgggoc*gctcggcaggct*cccaaatCAGCGCTggtgCcAGCCGCaacagggtg
GCATCTTTGTCGGCGGGTGTCTGTGTCGATCGCCACCGTCCCGGCGAACGGCGCCGGCC*GCTCGCCAGGCT*CCCAAATCAGCGCTGGTGCCAGCCgcaACAgGGTG
GCATCTTTGTCGGCGGGTGTCTGTGTCGATCGCCACCGTCCCGGCGAACGGCGCCGGCC*GCTCGCCAGGCT*CCCAAATCAGCGCTGGTGCCAGCCGCAACAGGGTG
GCATCTTTGTCGGCGGGTGTCTGTGTCGATCGCCACCGTCCCGGCGAACGGCGCCGGCC*GCTCGCCAGGCT*CCCAAATCAGCGCTGGTGCCAGCCGCAACAGGGTG
GCATCTTTgTcggcggggtgtcg gtcgatcgCcacCGTCCCGGCGAACgggocCgggoc*gctcggcaggct*cccaaatcagcGCTGGTGccagccgcaacagggtg
GCATCTTTgtcGgCGGGTGTcgtTcGATCGCCACCGTCCCGGCGAACGGCGCCGGcc*GCTCGCCAGGCT*CCCAAATCAGCGCTGGTGCCAGCCGCAACAGGGTG
GCATCTTTGTCGGCGGGTGTCTGTGTCGATCGCCACCGTCCCGGCGAACGgcGccggcc*GCTCGCCAGGCT*CCCAAATCAGCGctggTGCCAGCCGCAACAGGGTG
gcatct tgtggngnggttctcgcacggtatcaacagtcgatgg
GCATCtttgtcggcggggtgtcgTGTCTGTGTCGATCGCCACCGTcccggcGAACGGCGccggcc*gCtcGCCAGGCT*CCCAAATCAGCGctggtgocagcCgcaacaGgggtg
GCATCTTTGTCGGCGGGTGTCTGTGTCGATCGCCACCGTCCCGGCGAACGGCGCCGGCC*GCTCGCCAGGCT*CCCAAATCAGCGCTGGTGCCAGCCGCAACAGGGTG
GCATCTTTGTCGGCGGGTGTCTGTGTCGATCGCCACCGTCCCGGCGAACGGCGCCGGCC*GCTCGCCAGGCT*CCCAATCAGCGCTGGTGCCAGCCGCAACAGGGTG
gcat tt gtcg cg gttgtcgtgtcgcgac gtcccg c aacgg cggcc g tccgagg t*cccaa tcagcgctggtgc agccgca aggggtg
GCATCtttgtcggcgggTGTCTGTGTCGATcgccAccGTcccggcgaacGGCGccggcc*GCTcgcaGGCT*CCCAAATCAGCGCTGGTGccagccgCAACAGGGTG
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
ct*cc a atcagcgctggtgocagccgca cagggtg
```

Limpeza das seqüências:

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



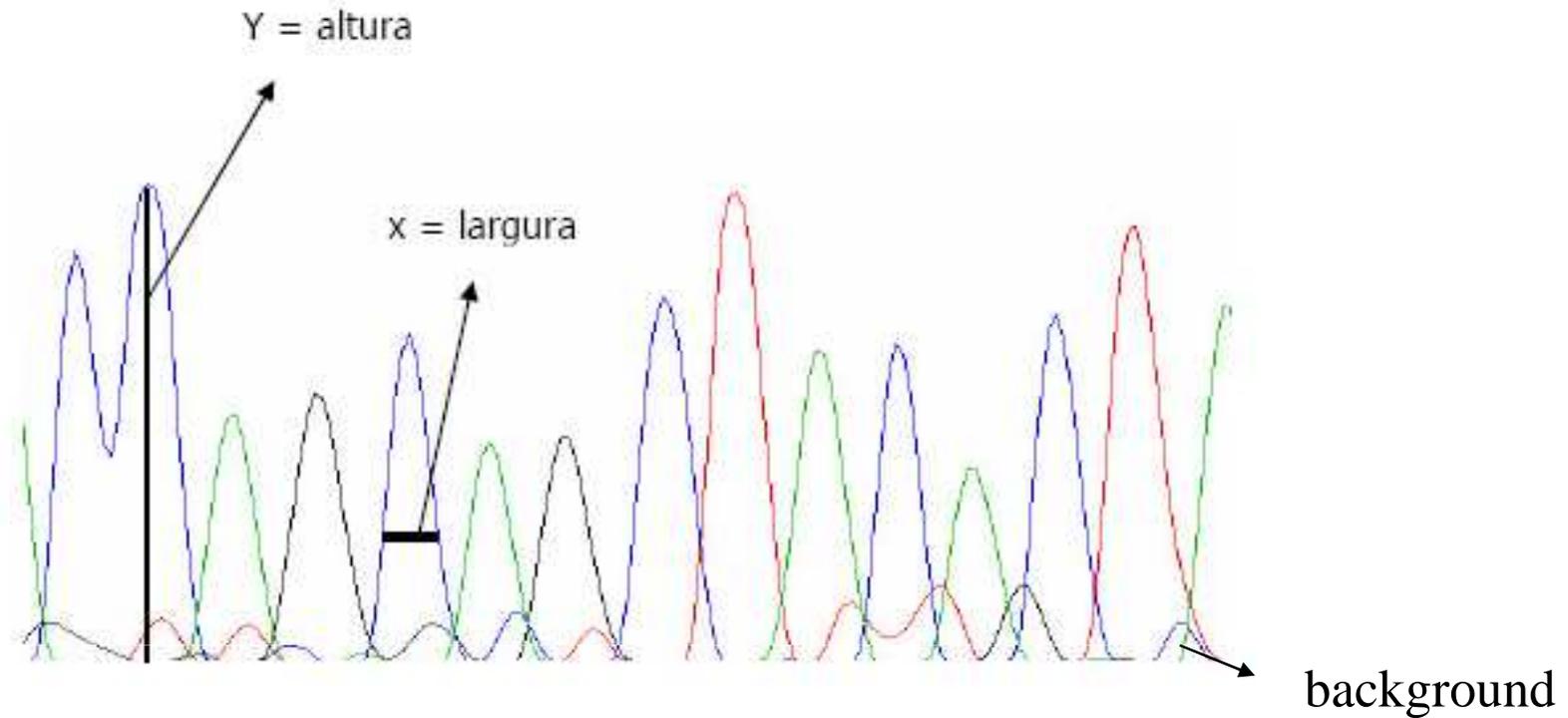
PHRED

```

gaattcggcaccgagagttctccggagacgctccgtgcgaagattatggaggccgtcaatgtggtcggttc
ccgccactttgctcgctgcgcatcgatgtaacagtcctggtgacgaagtcataccgtaagtattacgt
ttttgttgcgttggttcagcaatagtagaggacgggcgctttttttttgtcaagagaaaggggagggg
cgtactaccgctttatcgagggttggtattattcttatataaaagggaaagagcaacgtgaagcgggtaa
gggaagagtgaaagtcgag
  
```

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 ...

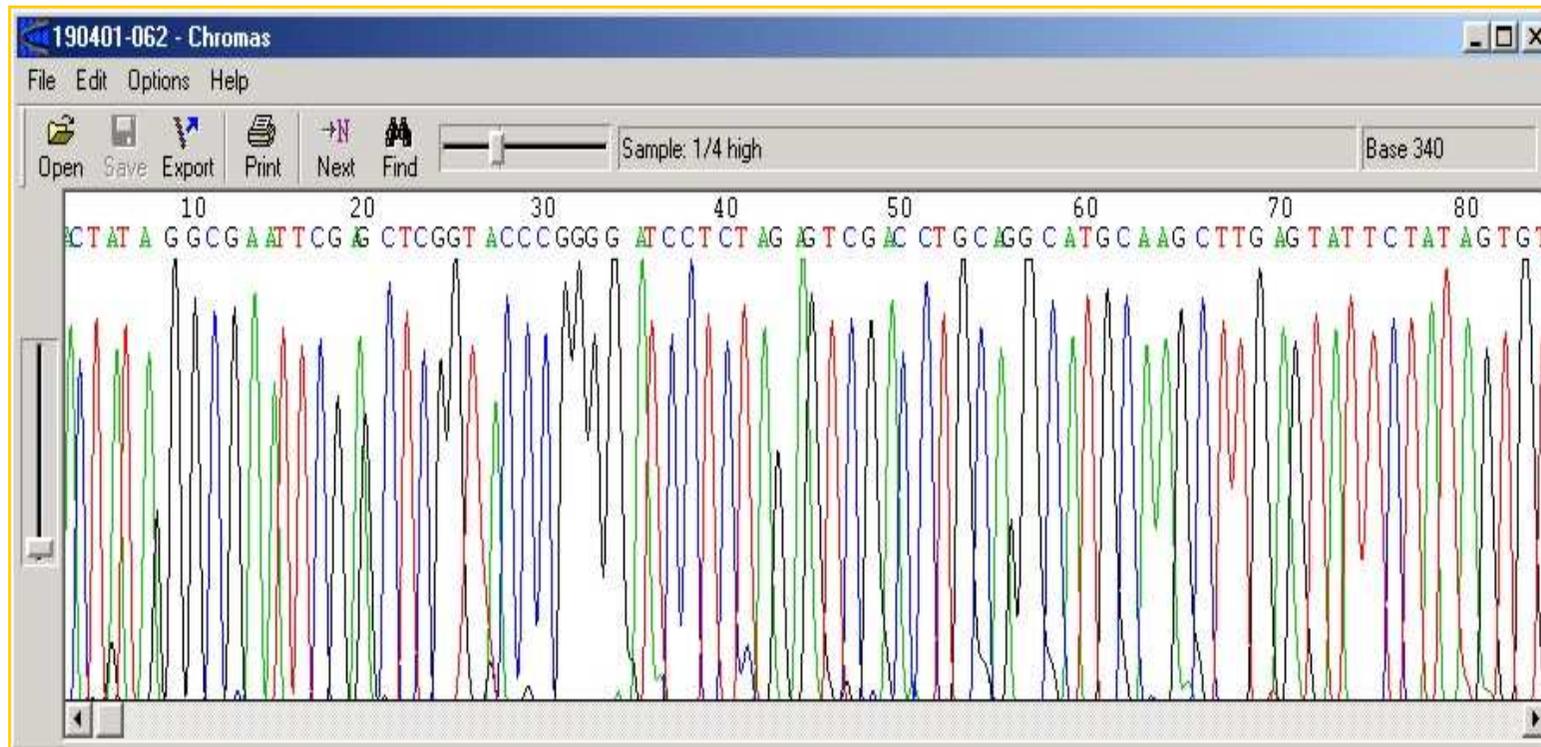


Cromatograma gerado pelo Sequenciador

- 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

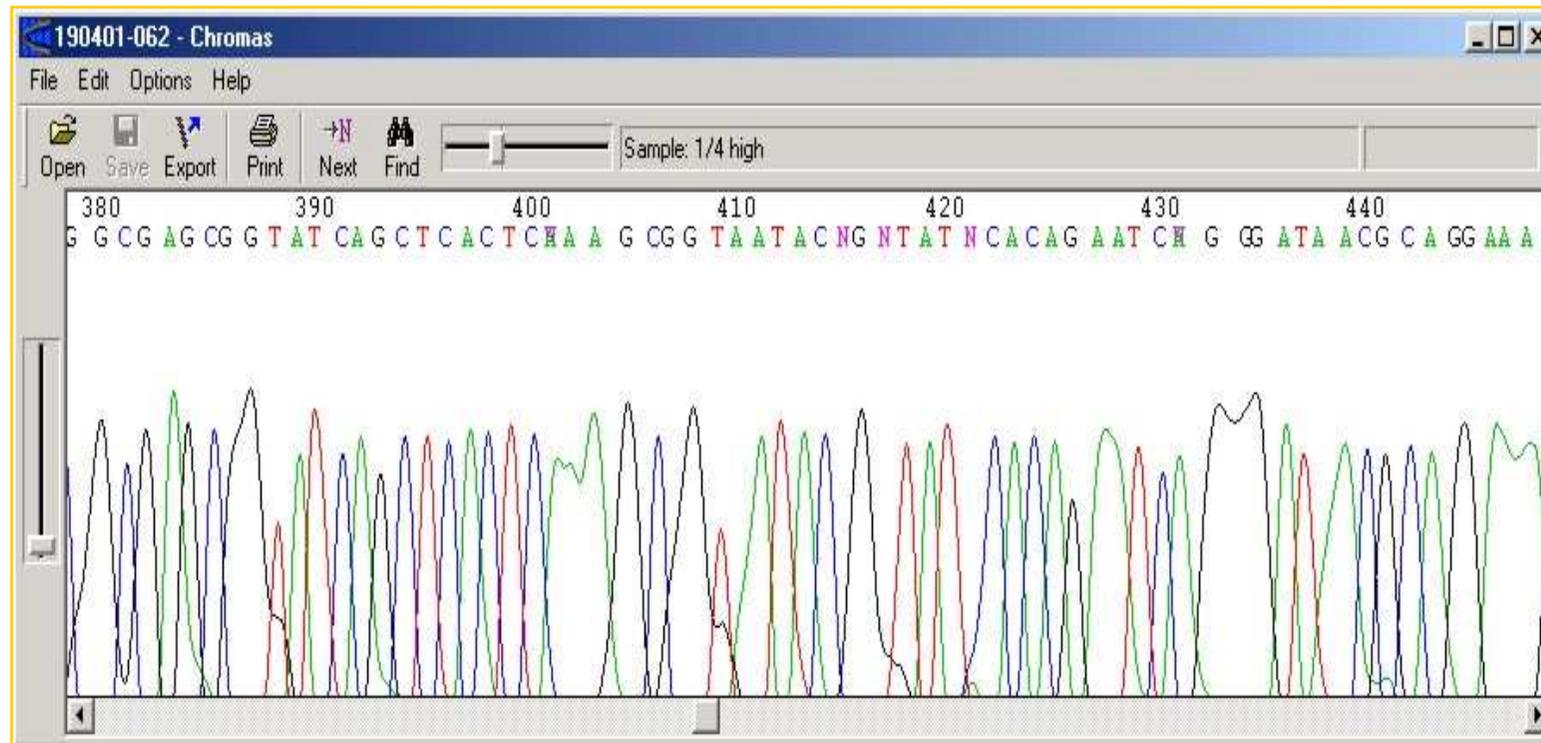
Região de qualidade alta



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

Analisando o cromatograma

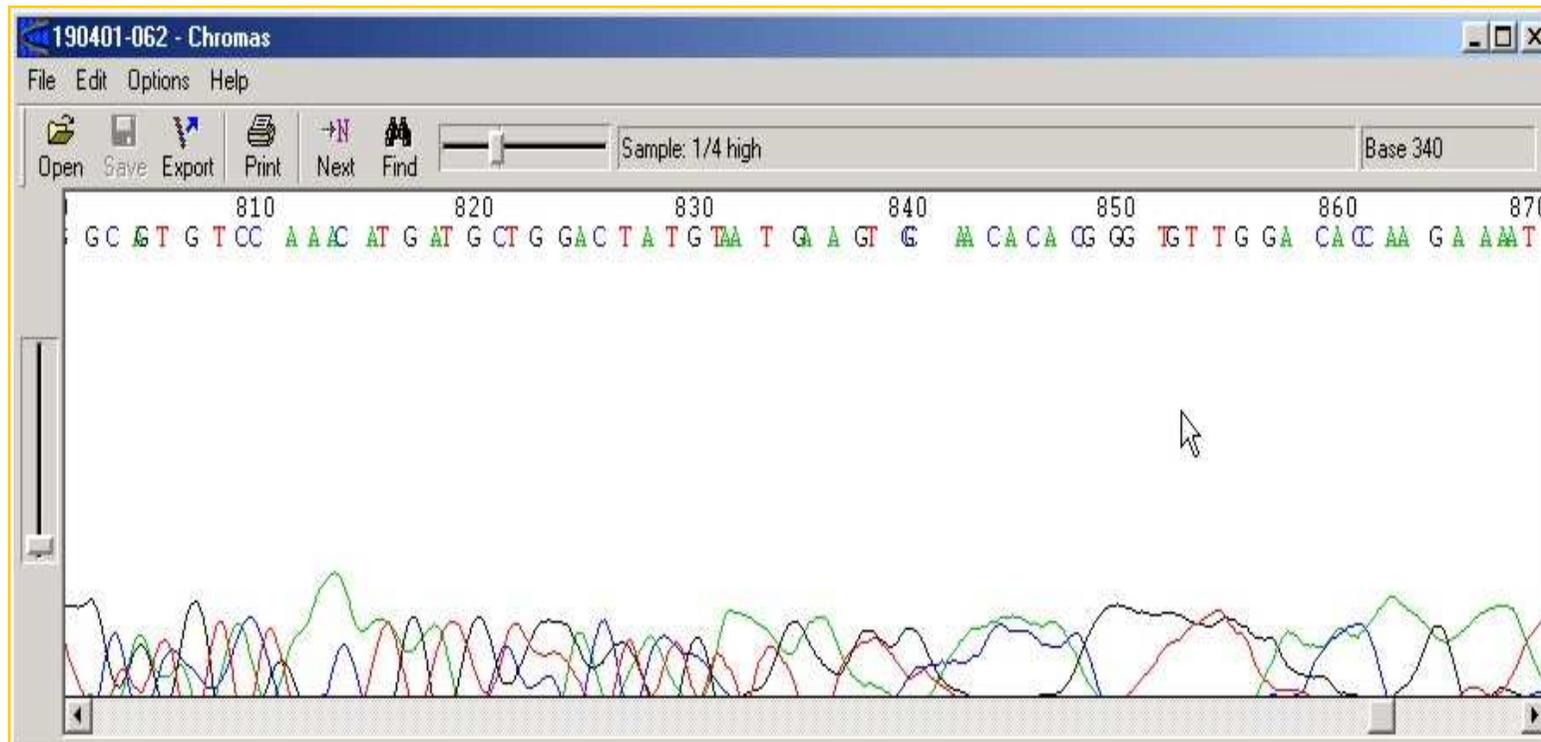
Região de qualidade média – poucas ambigüidades



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

Analisando o cromatograma

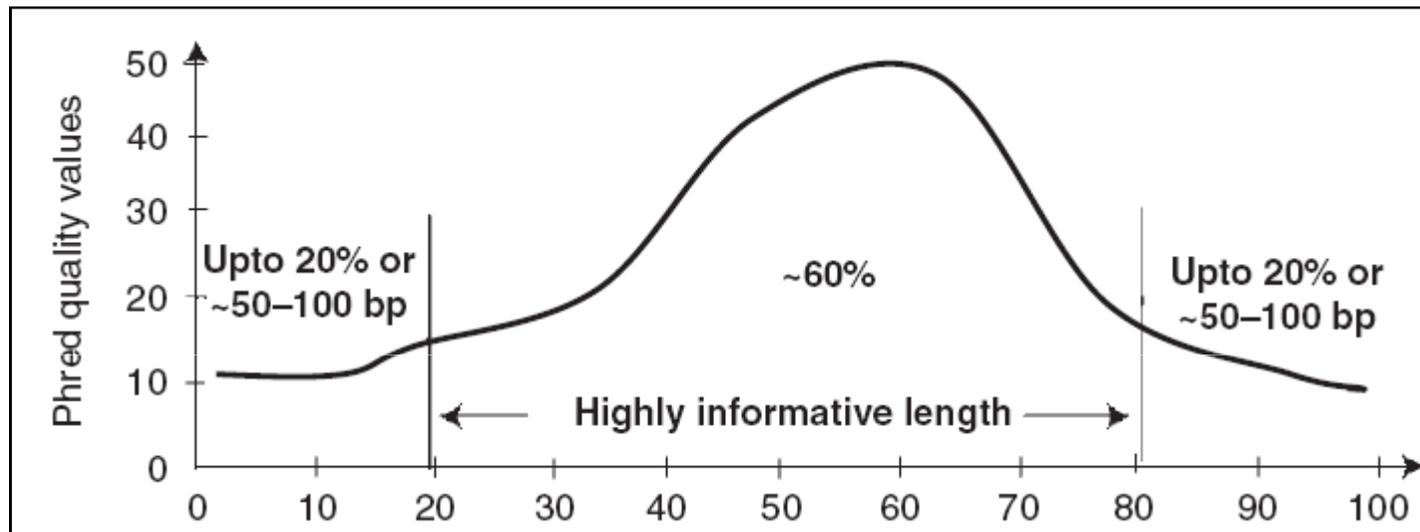
Região de qualidade baixa – baixa confiabilidade



- 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

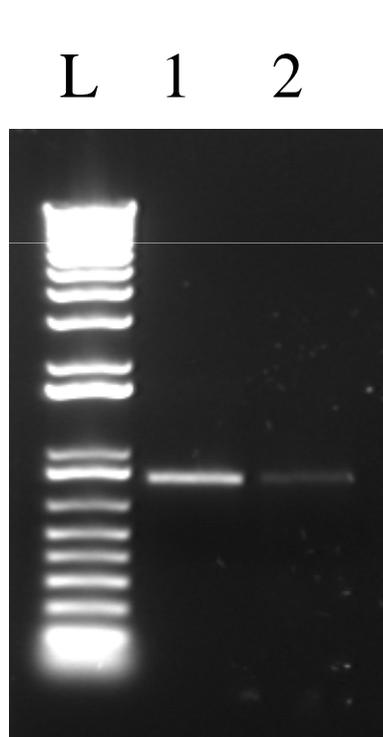


$$q = -10 \times \log_{10}(P)$$

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

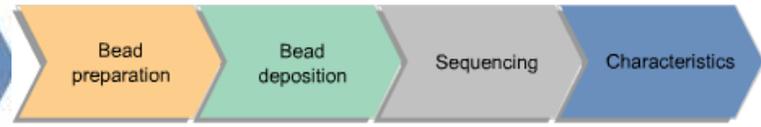
- Ronaghi *et al.*, 1996, 1998.
- Monitoramento em tempo real da síntese de DNA:
 - Síntese com primers pela DNA polimerase;
 - A incorporação é monitorada pela detecção luminosa da liberação do PPI, em uma reação envolvendo a enzima Luciferase;
 - Complexo com 4 Enzimas incluídas:
 - Fragmento Klenow 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

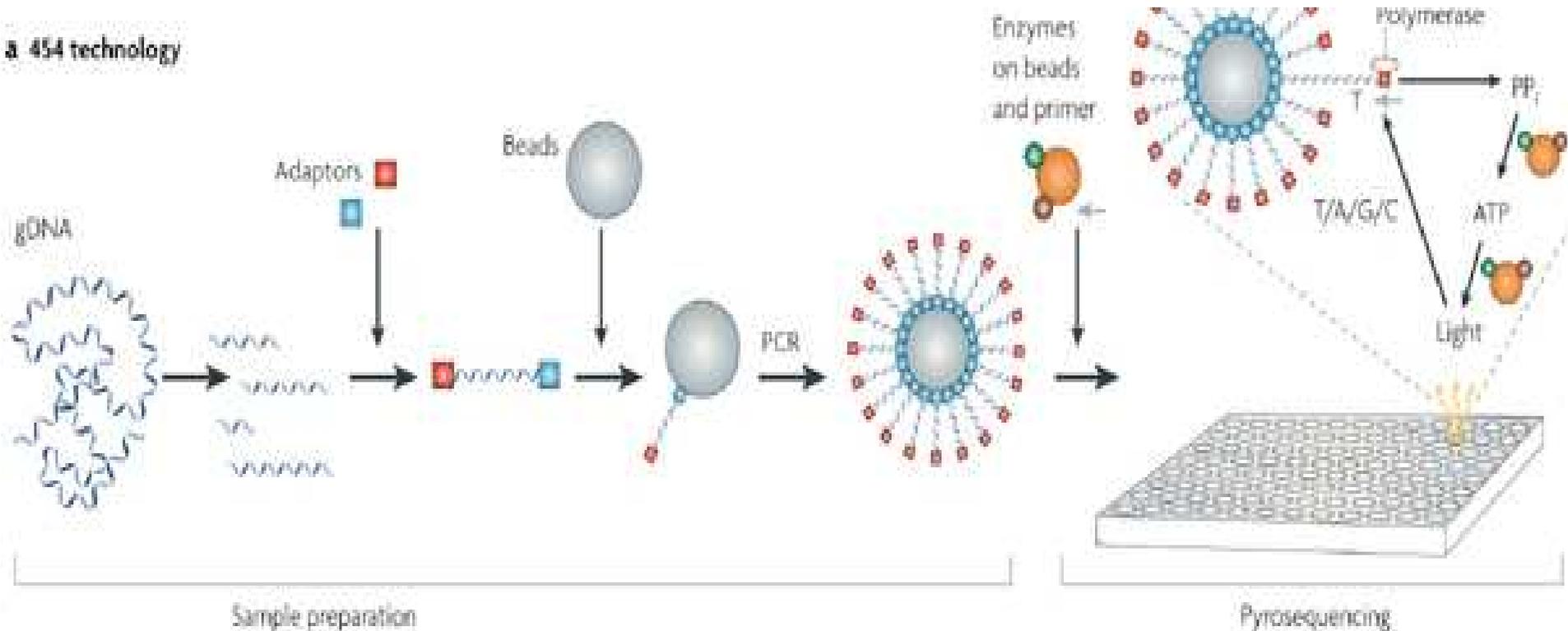
DNA Library Preparation



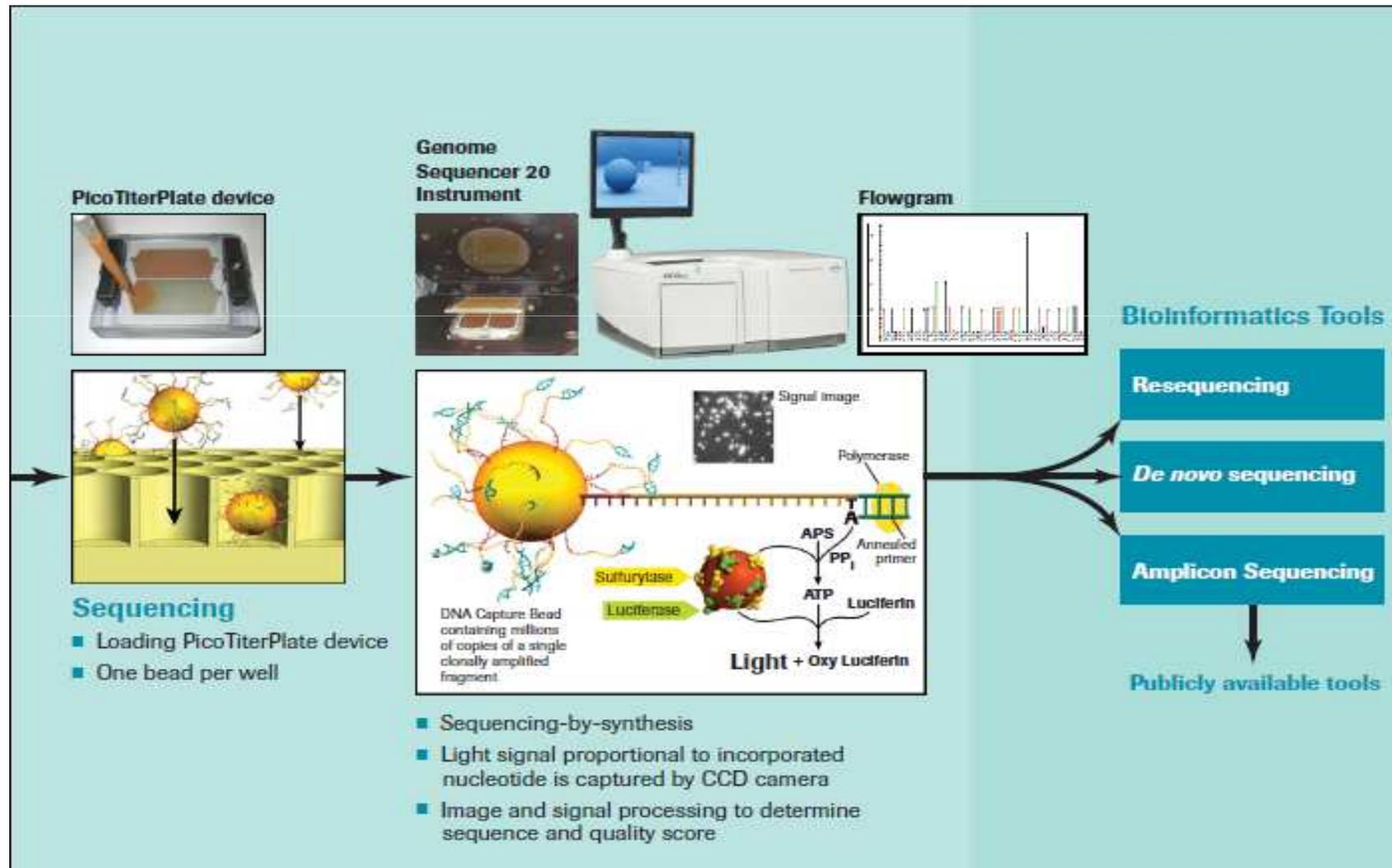
Sequencing by Synthesis



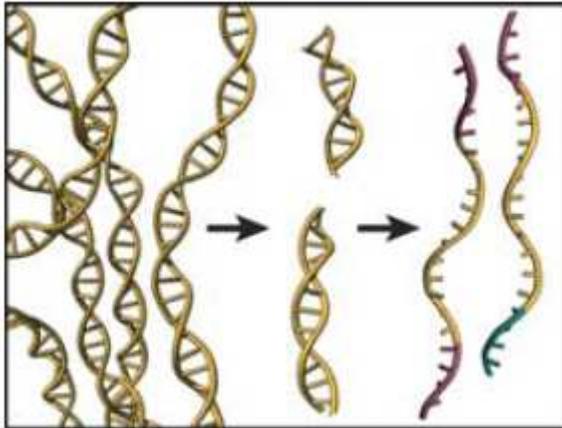
a 454 technology



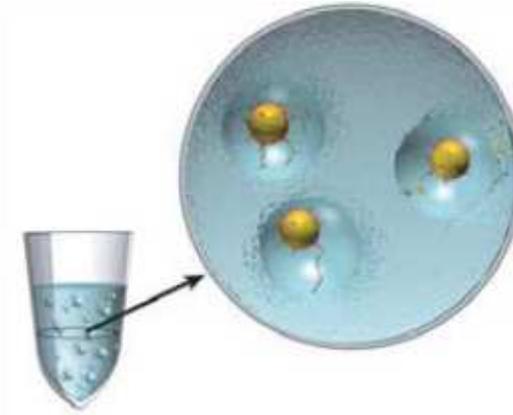
Sequenciamento- 454 Roche



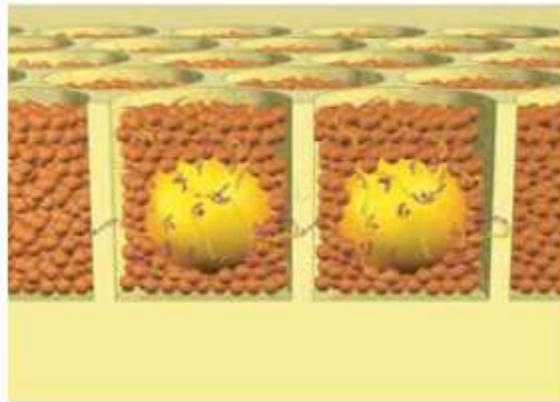
Seqüenciamento 454



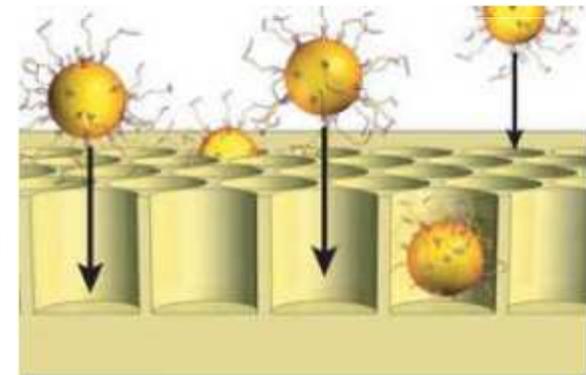
DNA genômico é isolado, fragmentado, ligado a adaptadores e separado em ssDNA



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

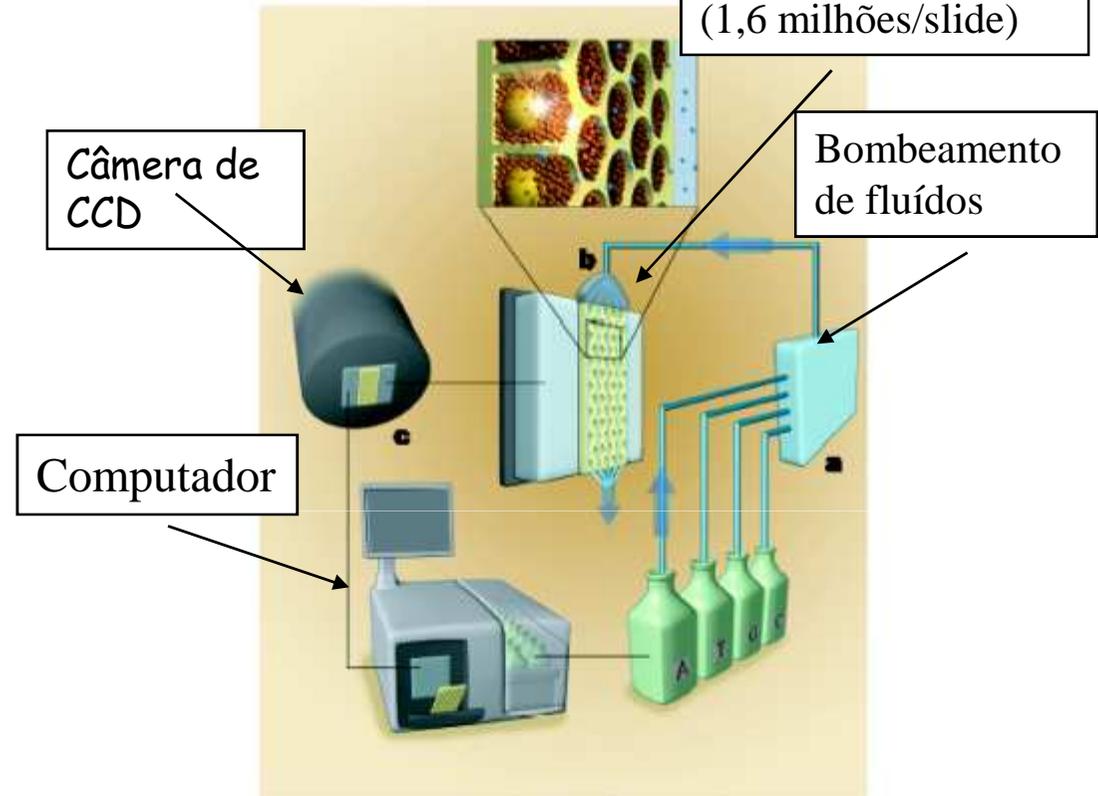
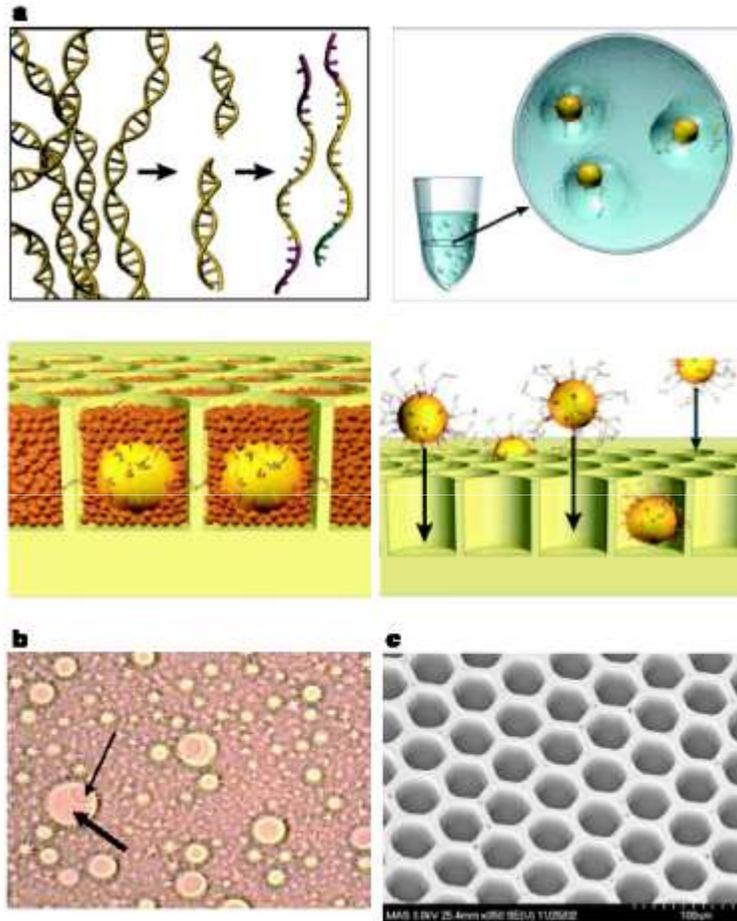


Pequenas *Beads*, com as enzimas necessárias ao piroseqüenciamento imobilizadas, são depositadas em cada poço.



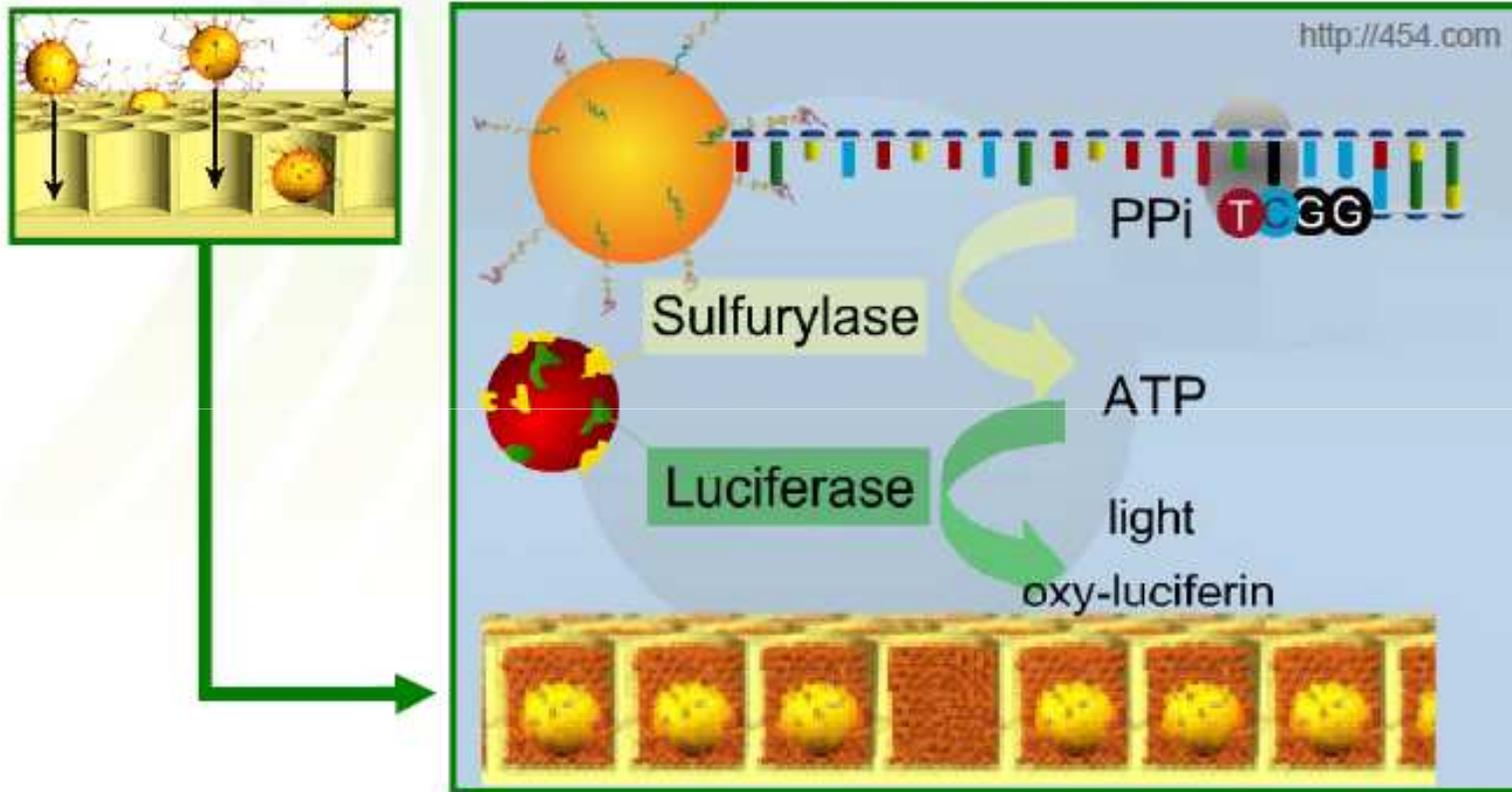
Quebra da emulsão, desnaturaçãõ das fitas de DNA e deposição das *beads* em slide de fibra ótica

Seqüenciamento 454



Um pmol de DNA numa reação de piroseqüenciamento 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 câmara de CCD.

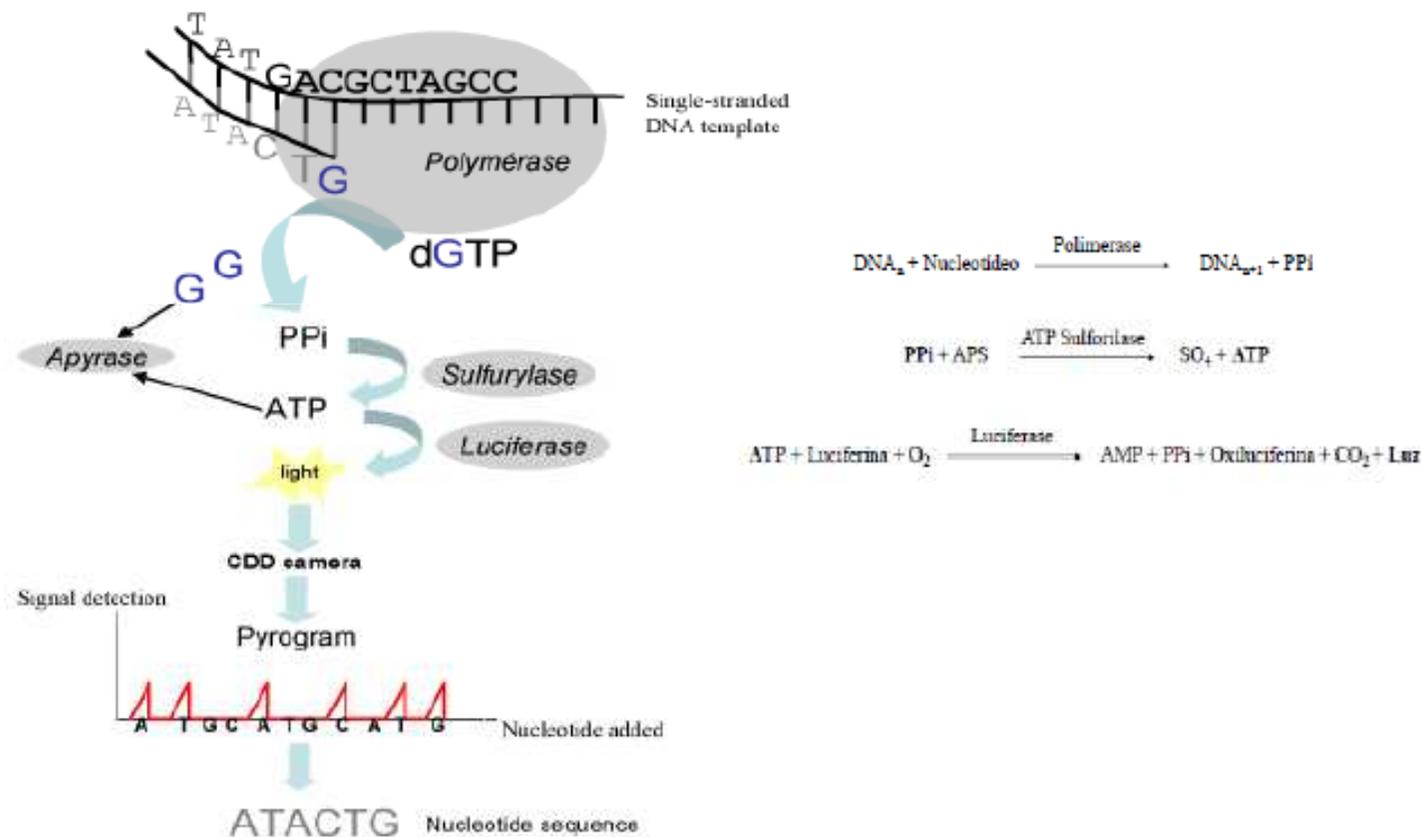
Pirosequenciamento



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

Pirosequenciamento

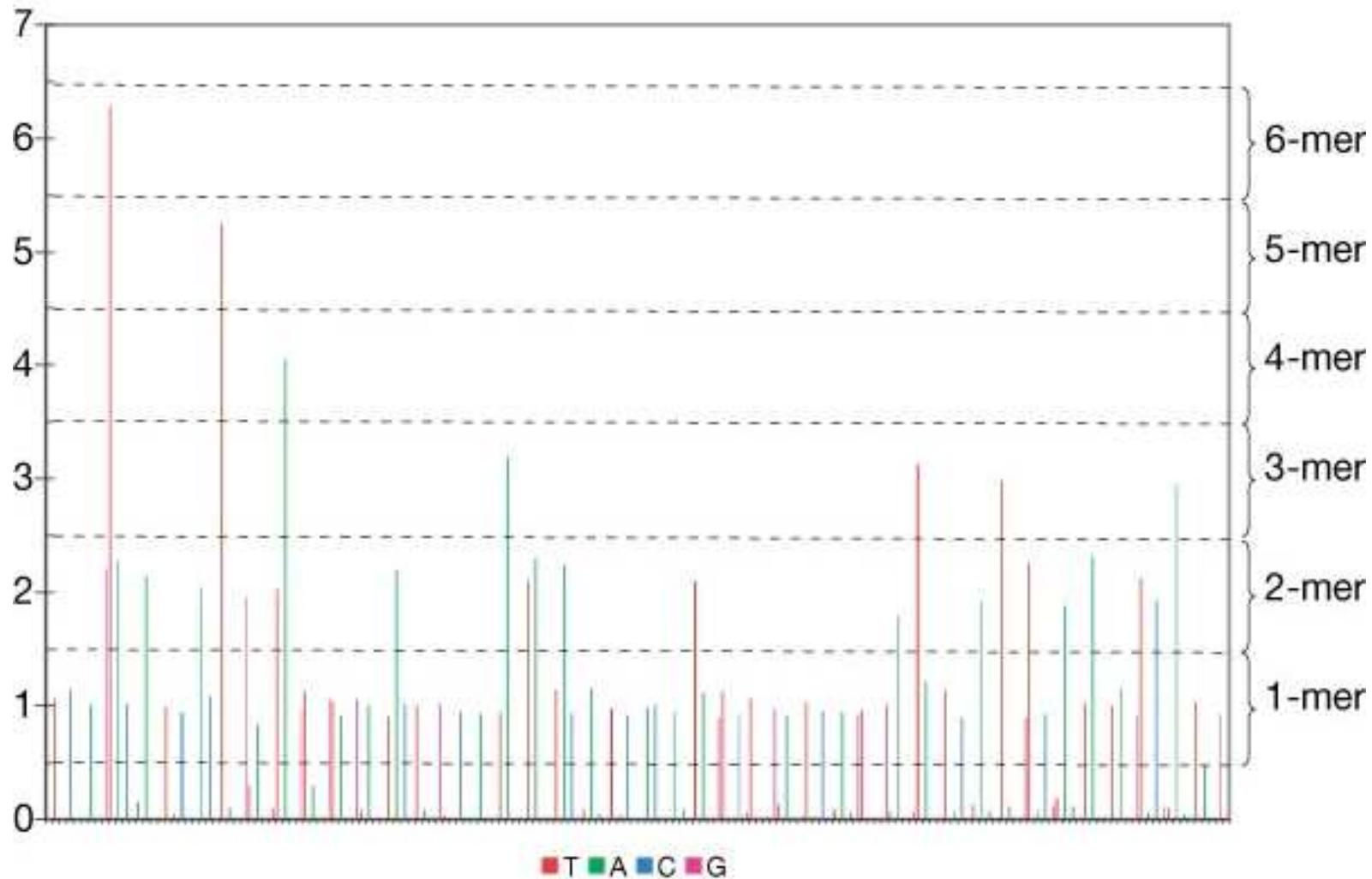
Cada base é adicionada separadamente



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

Pirograma

TCAGGTTTTTTAACAATCAACTTTTTGGATTAAAAGTGTAGATAACTGCATAAATTAATAA
CATCACATTAGTCTGATCAGTGAATTTATCAATTTGTTCAATAATAGTTCCAAATG



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