Sequenciamento do DNA



http://professor.pucgoias.edu.br/SiteDocente/admin/arquivosUpload/18497/material/Sequ%C3% AAnciamdo%20genomas.pdf

https://kasvi.com.br/sequenciamento-de-dna-desvendando-o-codigo-da-vida/

Bianca Zingales zingales@iq.usp.br

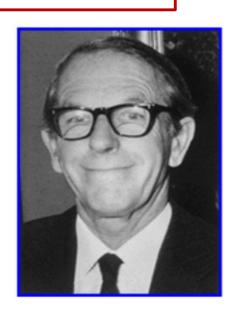
Histórico

1977. Duas técnicas de sequenciamento ALAN MAXAM - WALTER GILBERT – Método Químico

FREDERICK SANGER – Método Enzimático



Walter Gilbert



Frederick Sanger

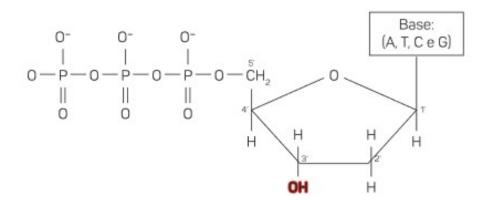
Prêmio Nobel de Química de 1980

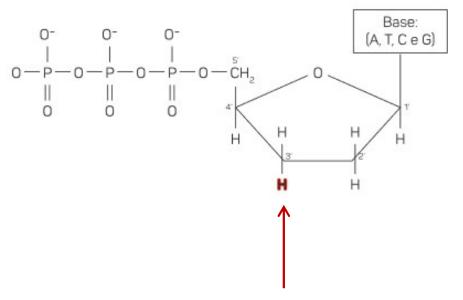
Sequenciamento do DNA – Método de Sanger

Sequenciamento do DNA com inibidores de terminação de cadeia

DESOXINUCLEOTÍDEOS (dNTP)

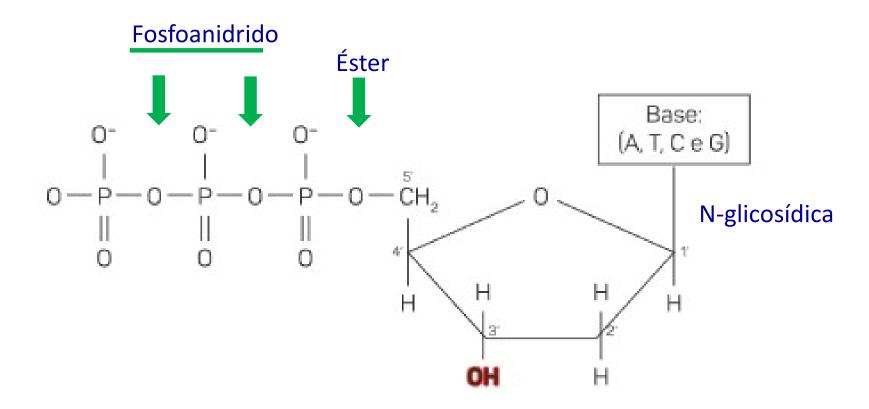
Inibidores: DIDESOXINUCLEOTÍDEOS (ddNTP)





Notar a ausência da hidroxila na posição 3' da **DESOXIRIBOSE**

Revisão das ligações nos **DESOXINUCLEOTÍDEOS (dNTP)**



Nota: Cai na Prova

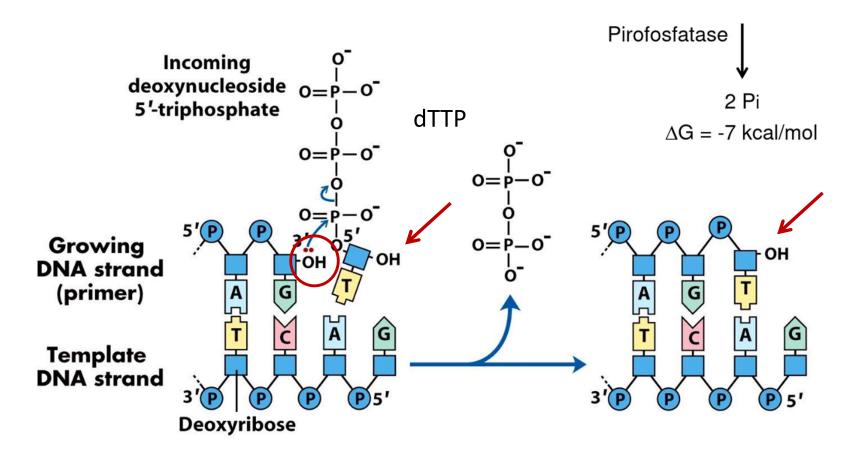
Sequenciamento do DNA – Método de Sanger

REAGENTES

- DNA dupla fita desnaturado
- Um iniciador (primer) complementar à extremidade 3' da fita de DNA (fita molde)
- DNA polimerase
- 4 desoxinucleosídeos trifosfato (dATP, dCTP, dGTP, dTTP)
- 4 didesoxinucleosídeos trifosfato (ddATP, ddCTP, ddGTP, ddTTP) (em menor concentração)

Princípio do método de Sanger

1. A DNA polimerase vai "copiando" a fita molde, a partir do primer, e adicionando os dNTPs complementares.



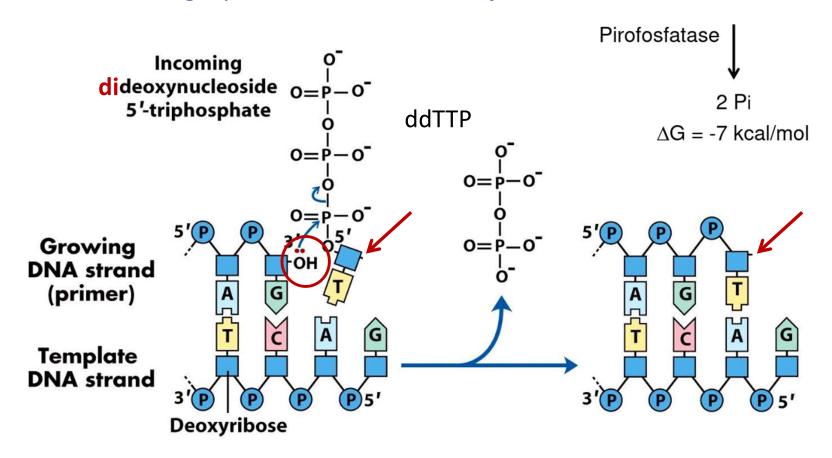
Notar a orientação das fitas!

Notar o papel da 3' OH do nucleotídeo terminal

Princípio do método de Sanger

2. Em algum momento um didesoxinucleotídeo é incorporado pela DNA polimerase na fita nova de DNA.

A ausência do grupo 3'-OH determina a parada da síntese de DNA.



Notar que a fita nova termina com o ddnucleotídio

Neste caso com ddTMP

Primeiros ensaios de sequenciamento

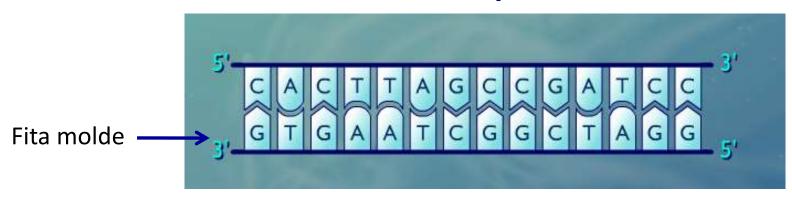
Sequenciamento manual

4 tubos separados

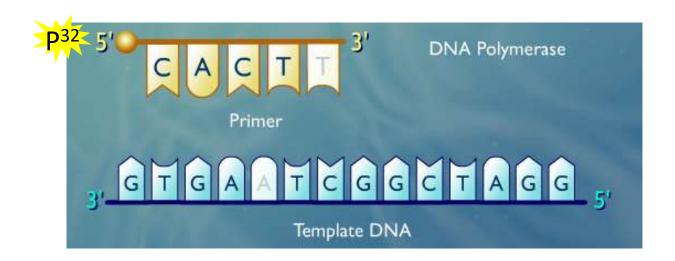
Tubos	1	2	3	4	
DNA desnaturado	+	+	+	+	
Iniciador marcado (α P ³²) +	+	+	+	
DNA polimerase	+	+	+	+	
os 4 dNTPs	+	+	+	+	
1 ddNTP/tubo	ddGTP	ddCTP	ddATP	ddTTP	



DNA dupla fita



Desnaturação do DNA e anelamento do primer

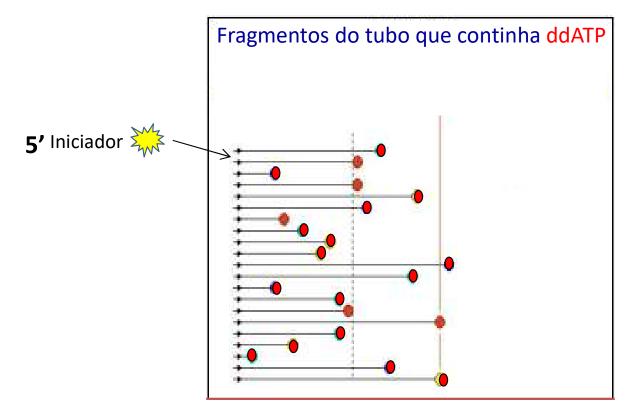


- Iniciador complementar à extremidade 3' da fita de DNA
- Sentido da síntese 5' -> 3'
- A marca radioativa corresponde à fita de DNA que está sendo sintetizada

Resultado

Em cada tubo é sintetizada **uma série de fragmentos de DNA** de tamanhos diferentes (terminação aleatória)

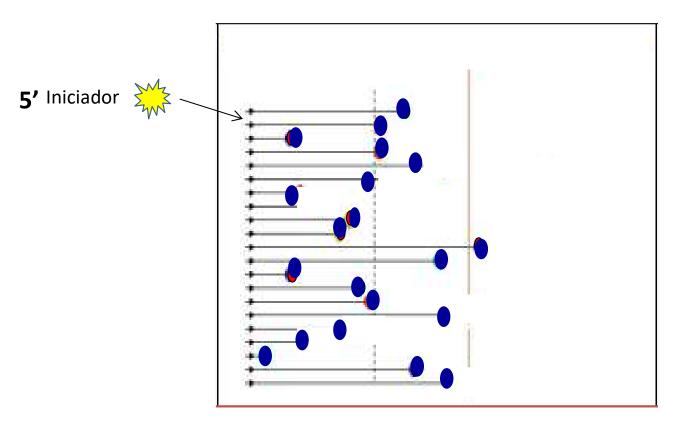
Todos os fragmentos começam com o mesmo iniciador marcado com P³²
Cada fragmento termina ao acaso com um dos didesoxinucleotídeos.



Série terminada com ddAMP

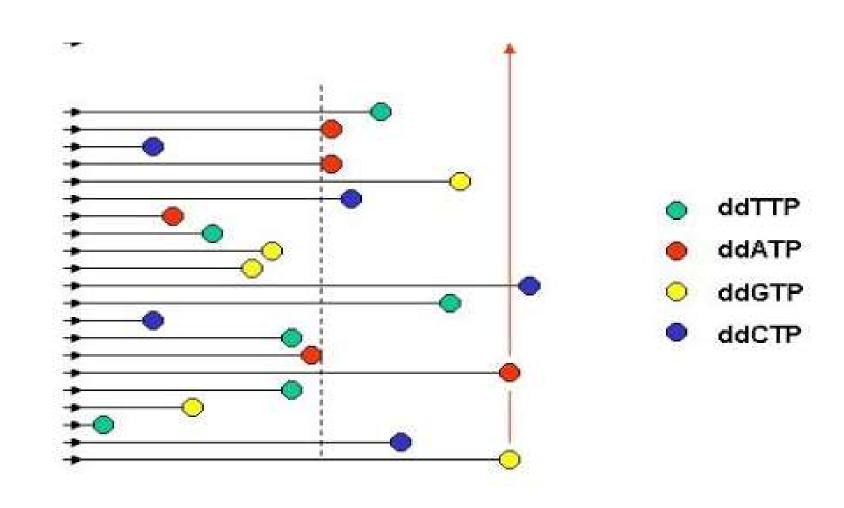
Fragmentos do tubo que continha ddCTP

Fragmentos de tamanhos diferentes – Todos terminados com ddCMP

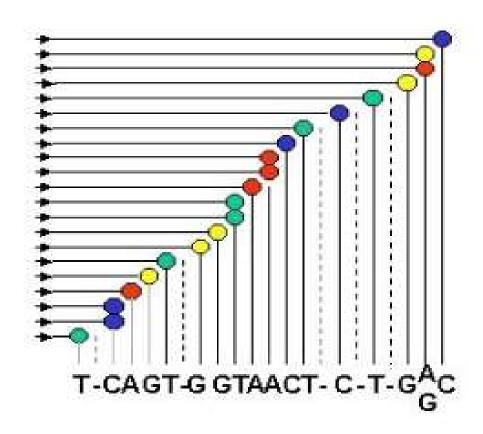


Série terminada com ddCMP

Variação do tamanho do conjunto dos fragmentos sintetizados nos 4 tubos



O conjunto dos fragmentos cobre toda a extensão de uma fita do gene



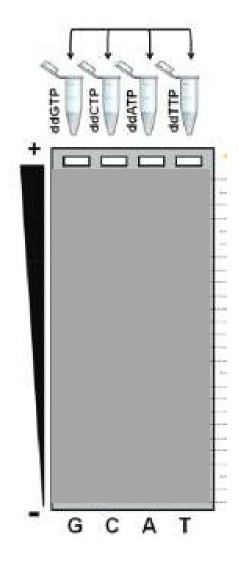
Terminada a reação, a **amostra de cada tubo** é desnaturada e submetida a eletroforese em gel de **poliacrilamida para DNA**

Gel de poliacrilamida permite separar fragmentos com variações de tamanho de 1 nucleotídeo. Gel de agarose não separa.

Eletroforese em gel de poliacrilamida dos fragmentos sintetizados



Uma canaleta para cada tubo de reação



Como "vejo" as bandas no gel?

Uma vez que o primer é marcado com P³²



Todos os fragmentos estarão radioativos

Faz-se uma autoradiografia do gel cobrindo-o com um filme de raio X

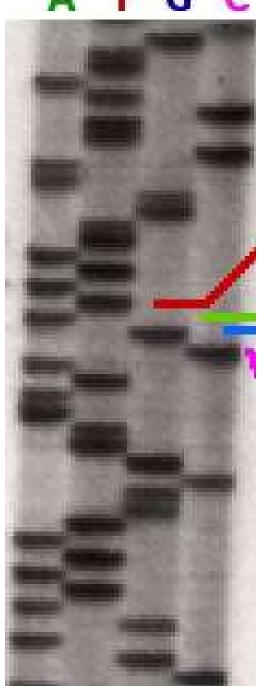
Depois revela-se o filme e tem-se a imagem

Amostras ddNTP









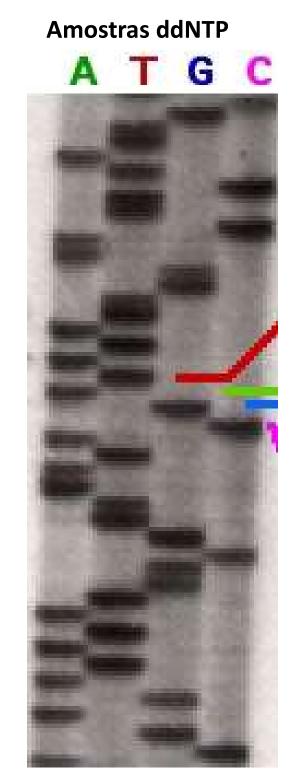


Leitura visual.

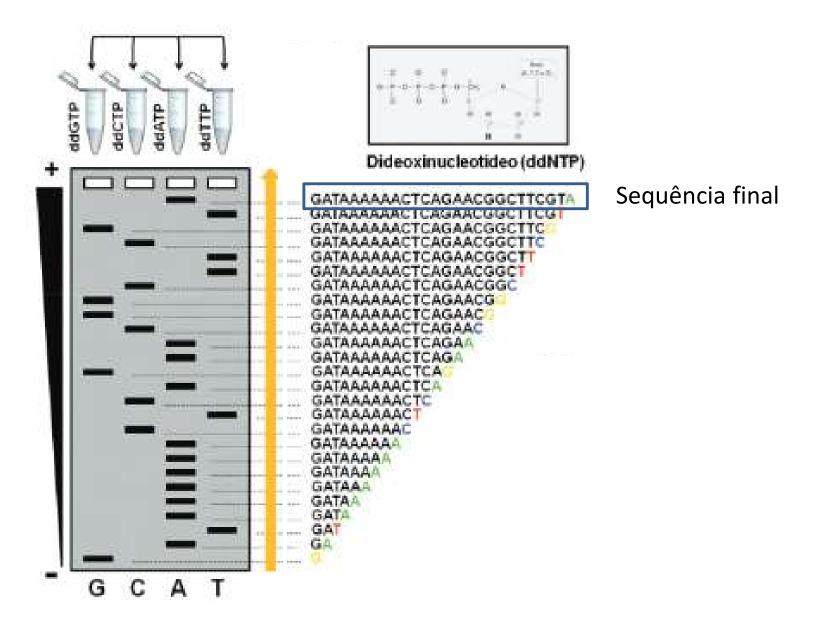
Cada fragmento tem diferença de UM nucleotídeo

5'ACGAGATATATGGCGTTAATACGATT.....

Esta é a sequência da fita que está sendo sintetizada sentido de 5' para 3'

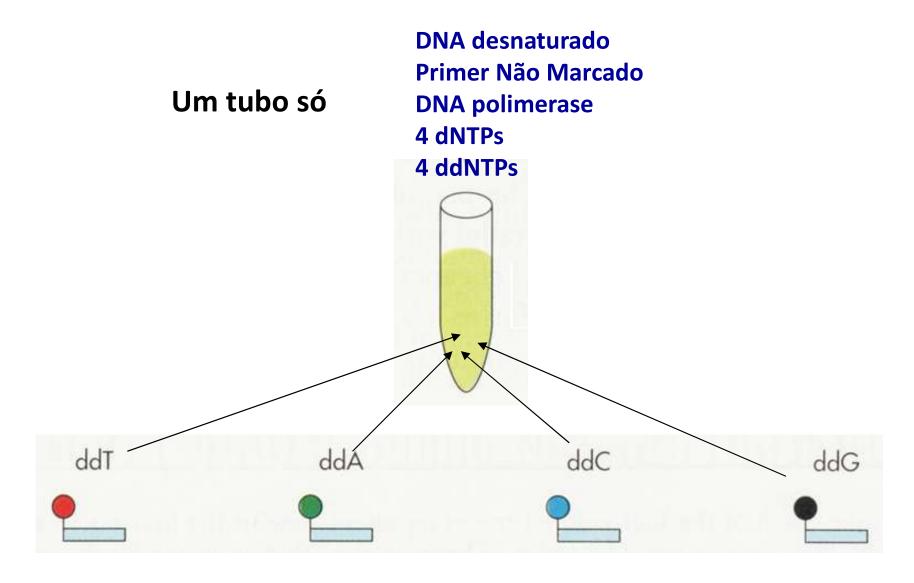


Exemplo 2. Leitura da sequência



Do menor fragmento para o maior

EVOLUÇÃO: Reação de sequenciamento com didesoxinucleotídeos fluorescentes. **Sequenciamento automático**



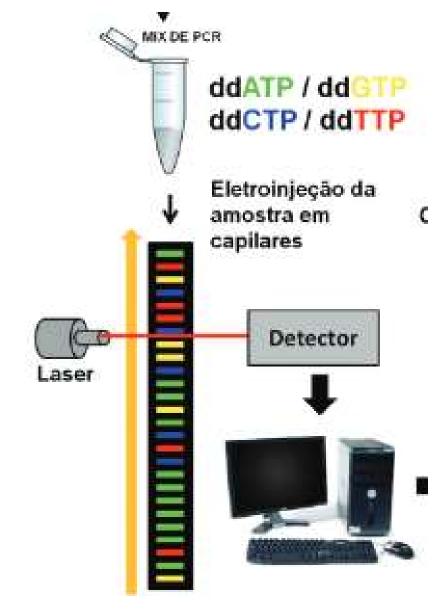
Cada ddNTP é marcado com uma sonda fluorescente diferente, que emite luz em comprimento de onda próprio. Reação feita em termociclador com Taq polimerase

Terminada a reação, a amostra é desnaturada e levada ao sequenciador (ABI Prism 3700).

No sequenciador ocorre a eletroforese

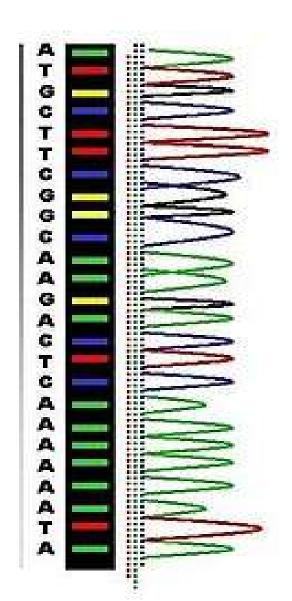


Durante a corrida os fluoróforos são excitados, lidos e registrados



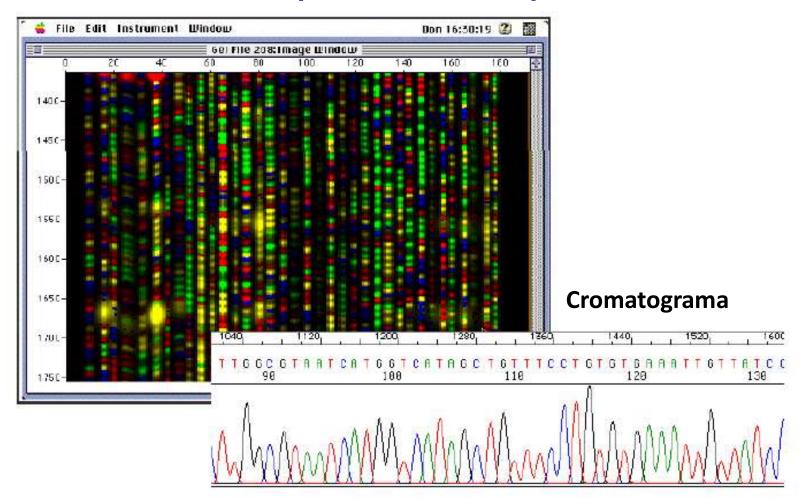
A fluorescência emitida pela passagem do fragmento de DNA pela janela de medição é registrada por um sistema de microcâmeras sensoras.

O sinal é transformado num **gráfico**, conhecido como **cromatograma**.



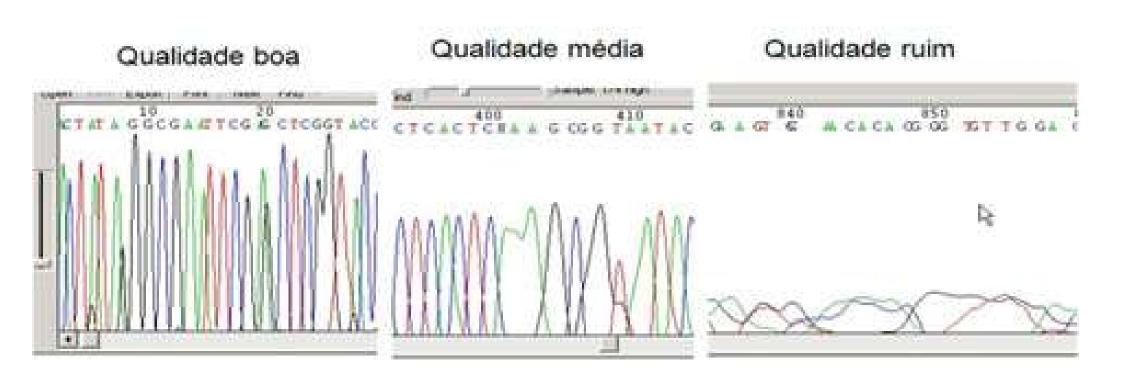
No sequenciador são analisadas muitas amostras de cada vez

Cada coluna corresponde a uma sequência



O pesquisador recebe o cromatograma de cada sequência e verifica sua qualidade

Qualidade das sequências



Dependendo do tipo de sequenciador **~600 bases** são determinadas em cada corrida/reação

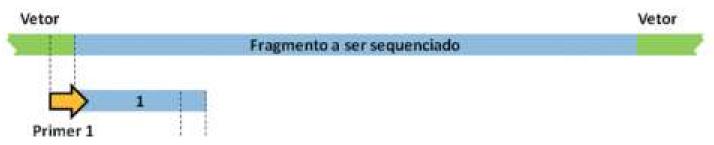
Como sequenciar um gene de 3 kb?

A amostra a ser sequenciada está clonada em um vetor



Etapa 1: Desenhar primer para a sequência do VETOR que está próxima do gene a ser sequenciado.

A sequência de todo o vetor é conhecida



Sequenciar 600 nucleotídeos

Sequenciamento por "primer walking"

Etapa 2: Desenhar primer para a sequência 3' terminal do fragmento sequenciado

Primer 2 Sequencia-se mais 600 nucleotideos.

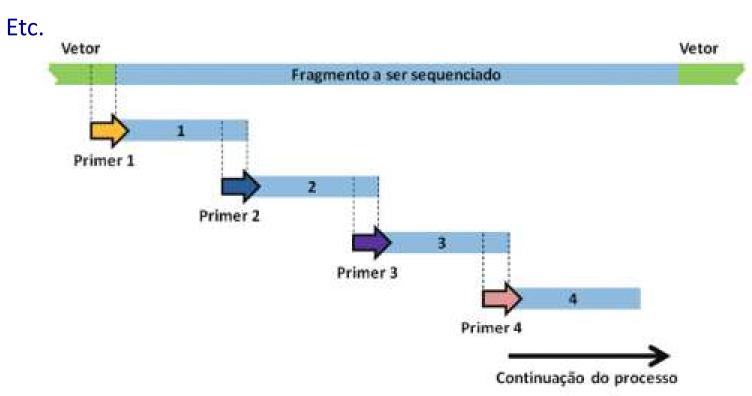
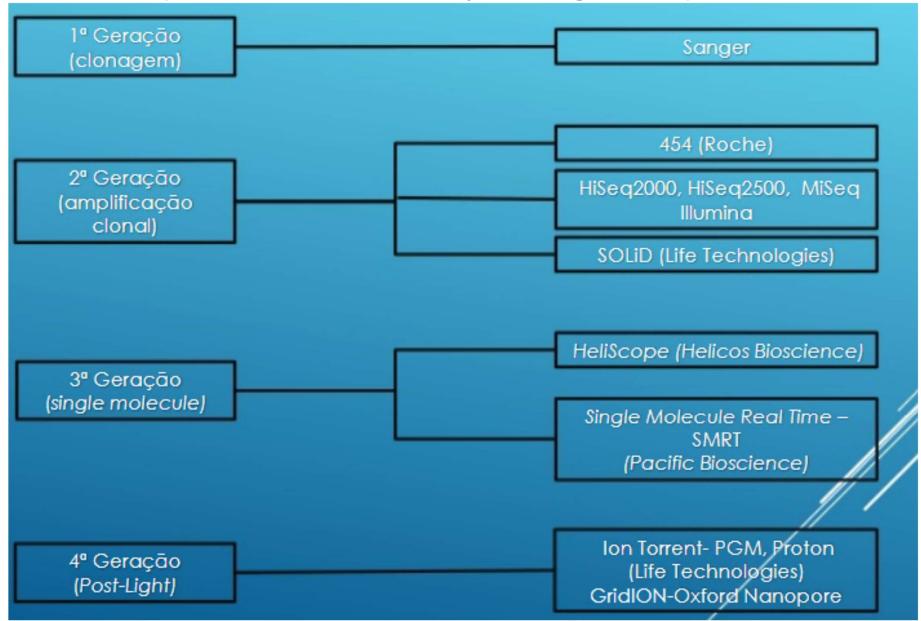


Figura 8. Modelo esquemático da técnica de Primer Walking. Um oligonucleotídeo inicial é usado no processo, determinando a sequencia a jusante após uso da técnica de sequenciamento. Como existe uma limitação da ação da polimerase e devido a baixa resolução dos géis de sequenciamento, haverá um momento em que os nucleotídeos não serão mais determinados, reduzindo a eficiência do processo. A partir desta nova sequencia determinada (1), desenha-se um novo oligonucleotído nas proximidades da posição 3'e tem-se inicio a um novo processo de sequenciamento, gerando o fragmento 2, que servirá para molde do desenho de um novo oligonucleotídeo, e assim por diante continua-se o processo. Repare que é como se fossem dados pequenos passos para se conhecer a sequencia completa, daí a definição primer walking.

O método de Sanger automatizado continua sendo usado para sequenciar DNAs pequenos.

Para o sequenciamento de Genomas usam-se outras tecnologias.

Plataformas de Sequenciamento de nova geração (Next Generation Sequencing – NGS)

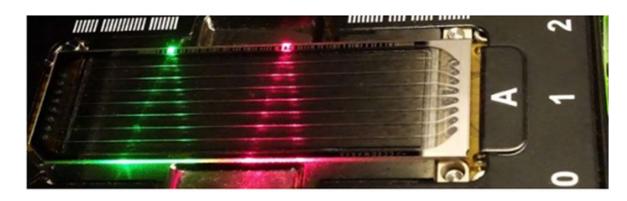


http://pos.unipar.br/simposios/x-simposio-em-biotecnologia-aplicada-a-agricultura/minicurso7.pdf

Plataformas de Sequenciamento de nova geração

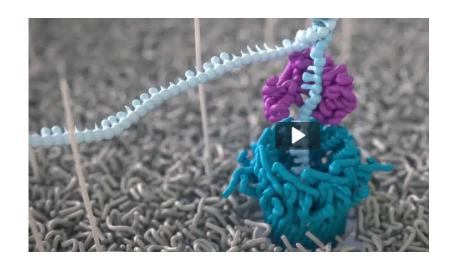


Illumina Short Read Sequencing



Nanopore technology





Plataformas de Sequenciamento de Nova Geração

Sanger ~ 6x10² nucleotídeos sequenciados/corrida, ~10⁸-10⁹ nucleotídeos sequenciados/corrida

Campo aberto para QUÍMICOS!

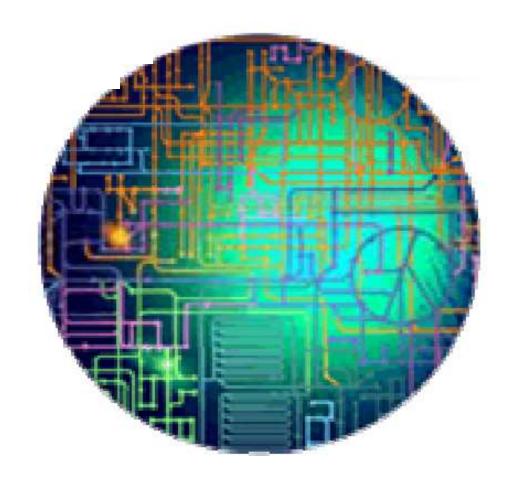
Diferentes matrizes; reagentes; métodos; equipamentos

Competição entre companhias públicas e privadas por melhor qualidade e preço

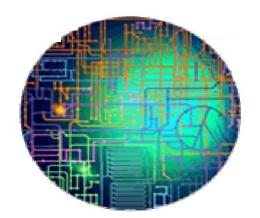
NOTA

SNGs usadas para sequenciamento **em larga escala** (Genomas, Transcriptomas)

Bioinformática



Ferramentas para a análise de dados



Bioinformática

Bioinformática é um ramo da ciência que utiliza e desenvolve ferramentas computacionais aplicadas à coleta, armazenamento e análise de dados genômicos.

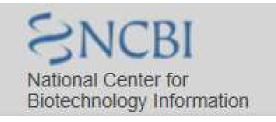
O profissional em bioinformática deve

- ter conhecimentos de biologia molecular, ciência da computação e bioinformática.
- ter conhecimentos de biologia, bioquímica, química, estatística e matemática, dentre outros.

Bancos de Dados

Repositórios de Informações





Repositório de sequências biológicas

- Sequências de Nucleotídeos
- Sequências de Proteínas
- Estruturas
- Genes
- Genomas
- Expressão Gênica

Criado em 1998

Repositório de artigos científicos

PubMed – mais de 30 milhões de citações de literatura de Biomedicina

Submissão de novas sequências

Nota: No IQUSP vários grupos têm expertise em Bioinformática

Ferramentas para busca e análise de sequências

- Entrez
- Buscas de similaridade (BLAST)
- Análise de estruturas
- Análise de Genomas, etc. etc.

Depósito da Sequência em bancos de dados

Nos bancos de dados a informação referentes a cada gene é apresentada em Nucleotídeos.

Cada sequência depositada recebe um número de acesso A sequência corresponde à extremidade 5' do gene

I - Trypanosoma cruzi strain CL Brener ABC transporter gene

ATG inicial >XM_801573.1

CTCCATCAAGGGCTCCGTCGCGCAGTTCGACGCTGTGGAACAAAATAAGAGCTCCGTGAGTGGACGCTTT TCTATTCCCGTCTCATGGCATAATTTGGCATATTCGGCAAACGGCACGAAAATTCTTTGCGGCCTCACAG GAACAGCGTTACCGTCACGATGCCTTGCTGTGATGGGATCCAGCGGTGCGGGCAAGACGACTTTTCTCAA TGCTATCTCTGACCGACTTAAAACCTCGCGTACCCTTAAGCTGACAGGGAAACGCCAGCTGGGGGACTTG GAGTACAAGCGTCATTACCGCAGGATGGTTGGTTTTGTGGCGCAAGACGACATTCTCTCACCACGGGCAA TGTTGAGGAAACTTTGGAGGAATTACGCCTTGTCCACTGCCGTGAGACCATTGTTGGCATCCCTGGCCTT GTCTCTGGTCTTTCAGGTGGTGAACGCAAACGCACAAGTATTGGAGTGGAGCTCATTTGCGATCCTAAAA TTCTCTTGCTGGATGAACCCACCTCTGGTCTGGACTCCGTGACATCTGTGAAGATTGTGCATCTTCTGAA TAACATTGCCCGAACAGGCCGCACGGTGATTTACACCATTCACCAGCCCACTGCTGAGACATTGACGTAC TTTGATGATCTCATGCTTCTCACTGGGGGTCGATGTGCTTACCATGGCACGATGGCAAAATCTGTGGAAT ACTTTGAGTCCATCGGATTCCCCTGTCCTGAACGATATACGCCAAGCGATTTCTTTATGAAGTTGCTCCA AGATCCAGAAATTTCCAAGGTACTGGTTAAAAAATGGAAGAGCTATCTAAAACACGGTGTGAGAACCCCA CATACAACCGCGGTTGAGCTAAATCCCAATCCCTCCGAGTCTCCCACCGCGAAGAATATTGAAAGCTACC TTAGTAGGTTTGGGAGCACCTCGGGTATCCAATTCCAGGAGCTTTTTCGTCGTTTTTTCCATAGATCTCAG TCGCAATCATGTATACATTTTTTCACATTTTATACAGGCTGCCTTCTTTGCAGTAATTGTGGGTCTCATA ATCGGGCTATGGGGCAGACTTTTATCATGGTCAACTCCTTTATGCAAGATAAGCCTTTGTACGTGCGGGA GCAAATGGTTGGCTCATACTCCCCTTTTATTTTCTTTTTATCAAAAACCCTGGTGGAGTTTCCAATGCGC TTTTTTACTACTTTGCGGTCATCGCGCTGCTTACTGAAGTGGCCTCGGGTCTTGGGTTTGCCATTGGTGC CACGTTTAAAAGTTTGGTCGTTGCTTCCGGTACCGCCCGTGATTTTGCTGCCGCTTGCCATGGTCGGT GGTCTTTTGGCGAACACAGATCGACTTCATCCGTATTGGTACTGGTTGGAGAAACCATCCTTTATTCGTC AGGCCTATATTCTTCTTGCCCGCAATGAATTTAAGCATATCGACCACATTCGGTGTGATGGTAGAGGCAA ACCACCGGGCTACTGTAAAGATAAGCCCCAAAACGGCGAGGATATCTTGCGCCAACTTGGGTTTCAGCAG AAGCAATATGAAAGCTGGATTTTGTGGCTAACTCTTGCCCTTTTATATATTGCTTTCCGCGGTTGGGCCG TTATTTCCCTGTACTCTGCCGCGCGTACAAGTTTTAG

Nos bancos de dados também consta a sequência de aminoácidos traduzida de muitos genes (um software faz a tradução) Cada sequência proteica tem um número de acesso

I - Trypanosoma cruzi strain CL Brener ABC transporter A sequência começa com a Metionina N-terminal

>XP_80666.1 ABC transporter [Trypanosoma cruzi strain CL Brener

MSCCRAEVNEPVTPSSASSLESDDQIASKQKGNEPLIEDYSIIAPGFSIKGSVAQFDAVEQNKSSVSGRF
SIPVSWHNLAYSANGTKILCGLTGTALPSRCLAVMGSSGAGKTTFLNAISDRLKTSRTLKLTGKRQLGDL
EYKRHYRRMVGFVAQDDILSPRATPEDSLRFSLRVRRGTSISETNKFVEETLEELRLVHCRETIVGIPGL
VSGLSGGERKRTSIGVELICDPKILLLDEPTSGLDSVTSVKIVHLLNNIARTGRTVIYTIHQPTAETLTY
FDDLMLLTGGRCAYHGTMAKSVEYFESIGFPCPERYTPSDFFMKLLQDPEISKVLVKKWKSYLKHGVRTP
HTTAVELNPNPSESPTAKNIESYLSRFGSTSGIQFQELFRRFSIDLSRNHVYIFSHFIQAAFFAVIVGLI
FLNVKDDLAGMQDREGVFFMVTMNRAMGQTFIMVNSFMQDKPLYVREQMVGSYSPFIFFLSKTLVEFPMR
VFFAFLECCILYWMVGFYRQAGAFFYYFAVIALLTEVASGLGFAIGATFKSLVVASGTAPVILLPLAMVG
GLLANTDRLHPYWYWLEKPSFIRQAYILLARNEFKHIDHIRCDGRGKPPGYCKDKPQNGEDILRQLGFQQ
KOYESWILWLTLALLYIAFRGWAVISLYSAARTKF

Não precisa mais sequenciar os aminoácidos de uma proteína!

Exemplos de buscas

Busca e caracterização de genes

Problema:

Quero desenhar primers para amplificar por PCR o gene do transportador ABC de **Trypanosoma cruzi strain CL Brener, cujo número de acesso é XM_801573.1**

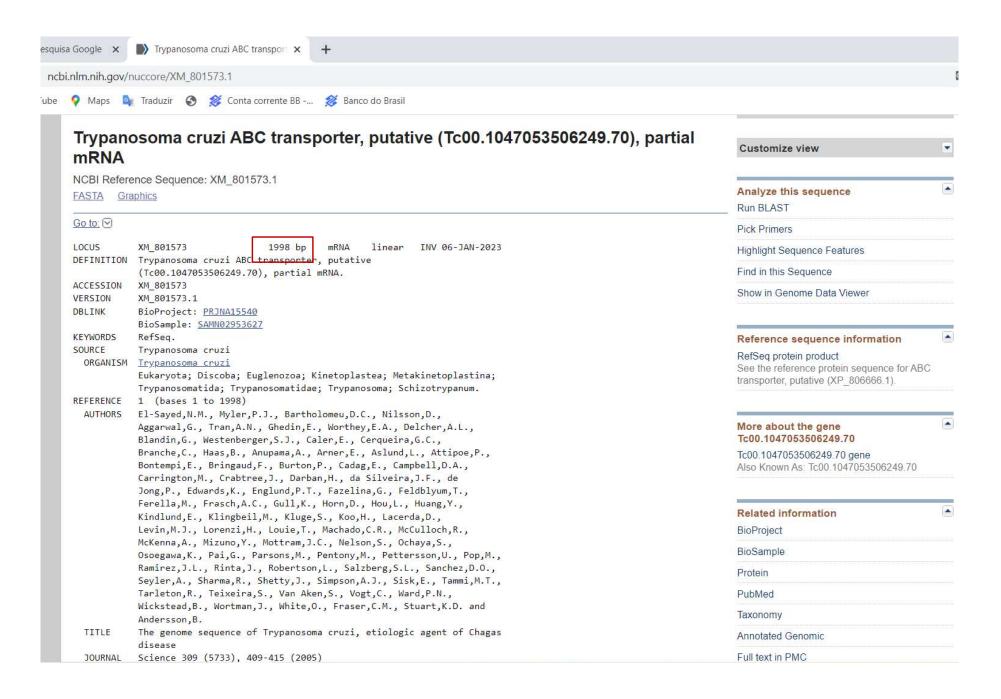
Procedimento:

Acesse: http://www.ncbi.nlm.nih.gov/genbank/

Na ferramenta Nucleotide: copie o número de acesso XM_801573.1



Resultado: Muitas informações sobre o gene



Sequência traduzida da proteína

/product="ABC transporter, putative"
/protein_id="XP_806666.1"
/db_xref="GeneID:3536733"

/translation="MSCCRAEVNEPVTPSSASSLESDDQIASKQKGNEPLIEDYSIIA
PGFSIKGSVAQFDAVEQNKSSVSGRFSIPVSWHNLAYSANGTKILCGLTGTALPSRCL
AVMGSSGAGKTTFLNAISDRLKTSRTLKLTGKRQLGDLEYKRHYRRMVGFVAQDDILS
PRATPEDSLRFSLRVRRGTSISETNKFVEETLEELRLVHCRETIVGIPGLVSGLSGGE
RKRTSIGVELICDPKILLLDEPTSGLDSVTSVKIVHLLNNIARTGRTVIYTIHQPTAE
TLTYFDDLMLLTGGRCAYHGTMAKSVEYFESIGFPCPERYTPSDFFMKLLQDPEISKV
LVKKWKSYLKHGVRTPHTTAVELNPNPSESPTAKNIESYLSRFGSTSGIQFQELFRRF
SIDLSRNHVYIFSHFIQAAFFAVIVGLIFLNVKDDLAGMQDREGVFFMVTMNRAMGQT
FIMVNSFMQDKPLYVREQMVGSYSPFIFFLSKTLVEFPMRVFFAFLECCILYWMVGFY
RQAGAFFYYFAVIALLTEVASGLGFAIGATFKSLVVASGTAPVILLPLAMVGGLLANT
DRLHPYWYWLEKPSFIRQAYILLARNEFKHIDHIRCDGRGKPPGYCKDKPQNGEDILR
QLGFQQKQYESWILWLTLALLYIAFRGWAVISLYSAARTKF"

ORIGIN

Sequência do gene 1-1998

```
1 atgtcgtgct gccgagcgga ggtgaatgag ccagtaactc ccagctccgc ttcatcattg
  61 gagtctgacg accaaatcgc gtcgaagcag aagggtaacg aacccctcat cgaagactat
 121 agcatcatcg cccccggatt ctccatcaag ggctccgtcg cgcagttcga cgctgtggaa
 181 caaaataaga gctccgtgag tggacgcttt tctattcccg tctcatggca taatttggca
 241 tattcggcaa acggcacgaa aattctttgc ggcctcacag gaacagcgtt accgtcacga
 301 tgccttgctg tgatgggatc cagcggtgcg ggcaagacga cttttctcaa tgctatctct
 361 gaccgactta aaacctcgcg tacccttaag ctgacaggga aacgccagct gggggacttg
421 gagtacaagc gtcattaccg caggatggtt ggttttgtgg cgcaagacga cattctctca
 481 ccacgggcaa cacccgaaga ttcccttcgc ttttcgctgc gcgtgaggcg tggcacaagc
 541 ataagtgaaa cgaataaatt tgttgaggaa actttggagg aattacgcct tgtccactgc
 601 cgtgagacca ttgttggcat ccctggcctt gtctctggtc tttcaggtgg tgaacgcaaa
 661 cgcacaagta ttggagtgga gctcatttgc gatcctaaaa ttctcttgct ggatgaaccc
721 acctctggtc tggactccgt gacatctgtg aagattgtgc atcttctgaa taacattgcc
 781 cgaacaggcc gcacggtgat ttacaccatt caccagccca ctgctgagac attgacgtac
 841 tttgatgatc tcatgcttct cactgggggt cgatgtgctt accatggcac gatggcaaaa
901 tctgtggaat actttgagtc catcggattc ccctgtcctg aacgatatac gccaagcgat
 961 ttctttatga agttgctcca agatccagaa atttccaagg tactggttaa aaaatggaag
1021 agctatctaa aacacggtgt gagaacccca catacaaccg cggttgagct aaatcccaat
1081 ccctccgagt ctcccaccgc gaagaatatt gaaagctacc ttagtaggtt tgggagcacc
1141 tcgggtatcc aattccagga gctttttcgt cgtttttcca tagatctcag tcgcaatcat
1201 gtatacattt tttcacattt tatacaggct gccttctttg cagtaattgt gggtctcata
1261 tttctgaatg ttaaagatga tttagctggt atgcaagatc gcgagggagt ttttttatg
1321 gtaacgatga atcgggctat ggggcagact tttatcatgg tcaactcctt tatgcaagat
1381 aagcctttgt acgtgcggga gcaaatggtt ggctcatact ccccttttat tttctttta
1441 tcaaaaaccc tggtggagtt tccaatgcgc gtgttttttg cctttcttga gtgctgtatt
1501 ttatactgga tggtgggttt ttaccgccag gcaggagctt ttttttacta ctttgcggtc
1561 atcgcgctgc ttactgaagt ggcctcgggt cttgggtttg ccattggtgc cacgtttaaa
1621 agtttggtcg ttgcttccgg taccgcgcc gtgattttgc tgccgcttgc catggtcggt
1681 ggtcttttgg cgaacacaga tcgacttcat ccgtattggt actggttgga gaaaccatcc
1741 tttattcgtc aggcctatat tcttcttgcc cgcaatgaat ttaagcatat cgaccacatt
1801 cggtgtgatg gtagaggcaa accaccgggc tactgtaaag ataagcccca aaacggcgag
1861 gatatcttgc gccaacttgg gtttcagcag aagcaatatg aaagctggat tttgtggcta
1921 actcttgccc ttttatatat tgctttccgc ggttgggccg ttatttccct gtactctgcc
1981 gcgcgtacaa agttttag
```

Desenho de primers

Muitas ferramentas disponíveis

Copie a sequência de nucleotídeos

Informe quantos primers deseja

Informe o tamanho do amplicon Ex. 1998 pb

Informe a faixa de Tm dos primers

Start

Resultado

ABC Forward 5' GTGAACGCAAACGCACAAGTA 3'

ABC Reverse 5' GAGGTGGGTTCATCCAGCAA 3'

PRONTO. JÁ POSSO FAZER A PCR!!!!!!!!!!!