

Metodologias de sequenciamento de genomas e transcriptomas

Desirrê Petters-Vandresen

Módulo I – Genômica no Estudo de Microrganismos

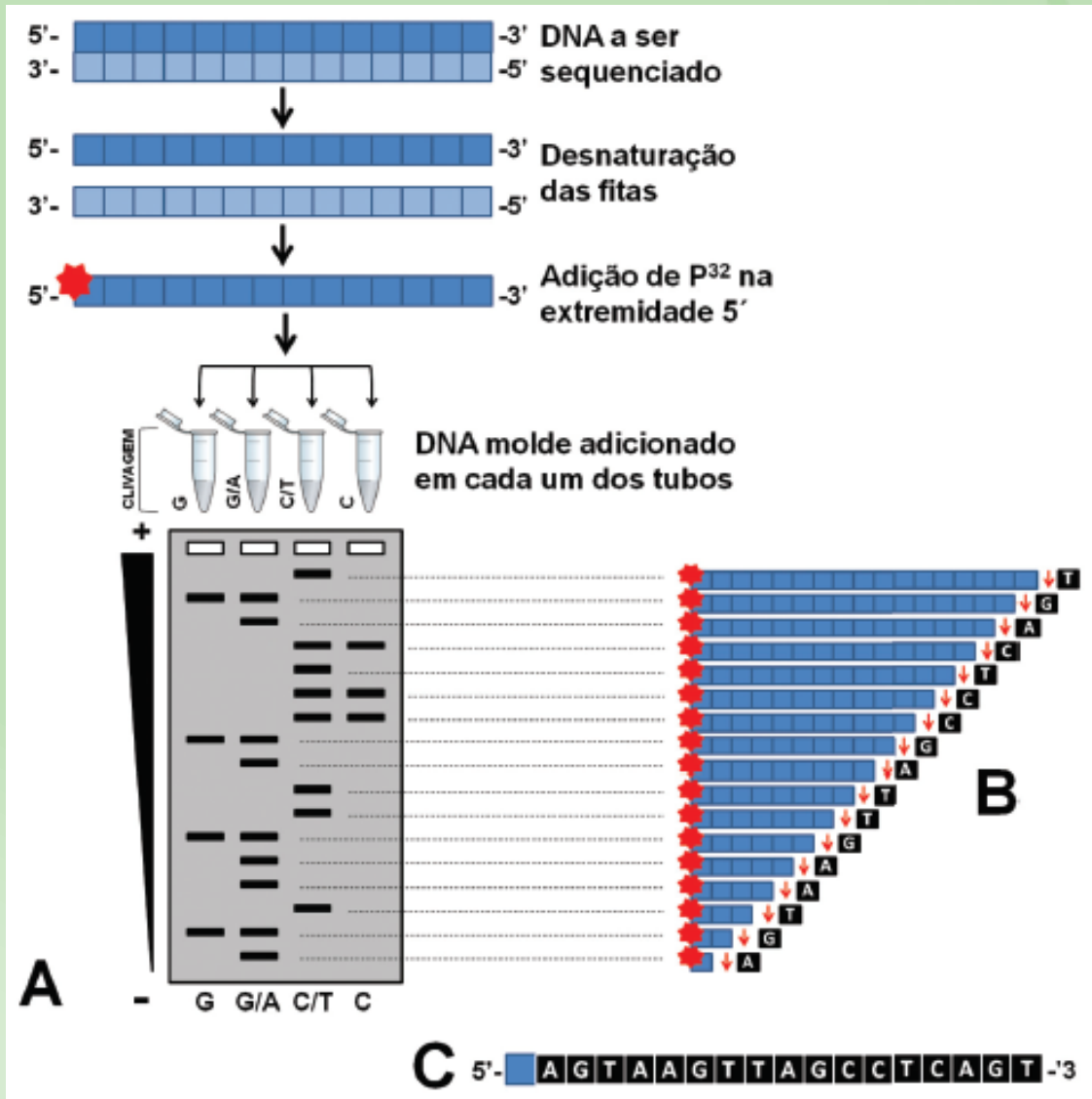
Sequenciamento de genoma ou transcriptoma

- Identificação da sequência de nucleotídeos de uma molécula de DNA ou RNA na sua ordem correta, para conhecer a informação genética presente nesta estrutura
- Além da identidade de cada base, o sequenciamento também fornece informações sobre a confiabilidade de cada uma das bases identificadas
- Avanços na escala de sequenciamento nos últimos 50 anos: do sequenciamento manual ao sequenciamento maciço e paralelo de genomas inteiros em um curto período de tempo

Sequenciamento em pequena escala

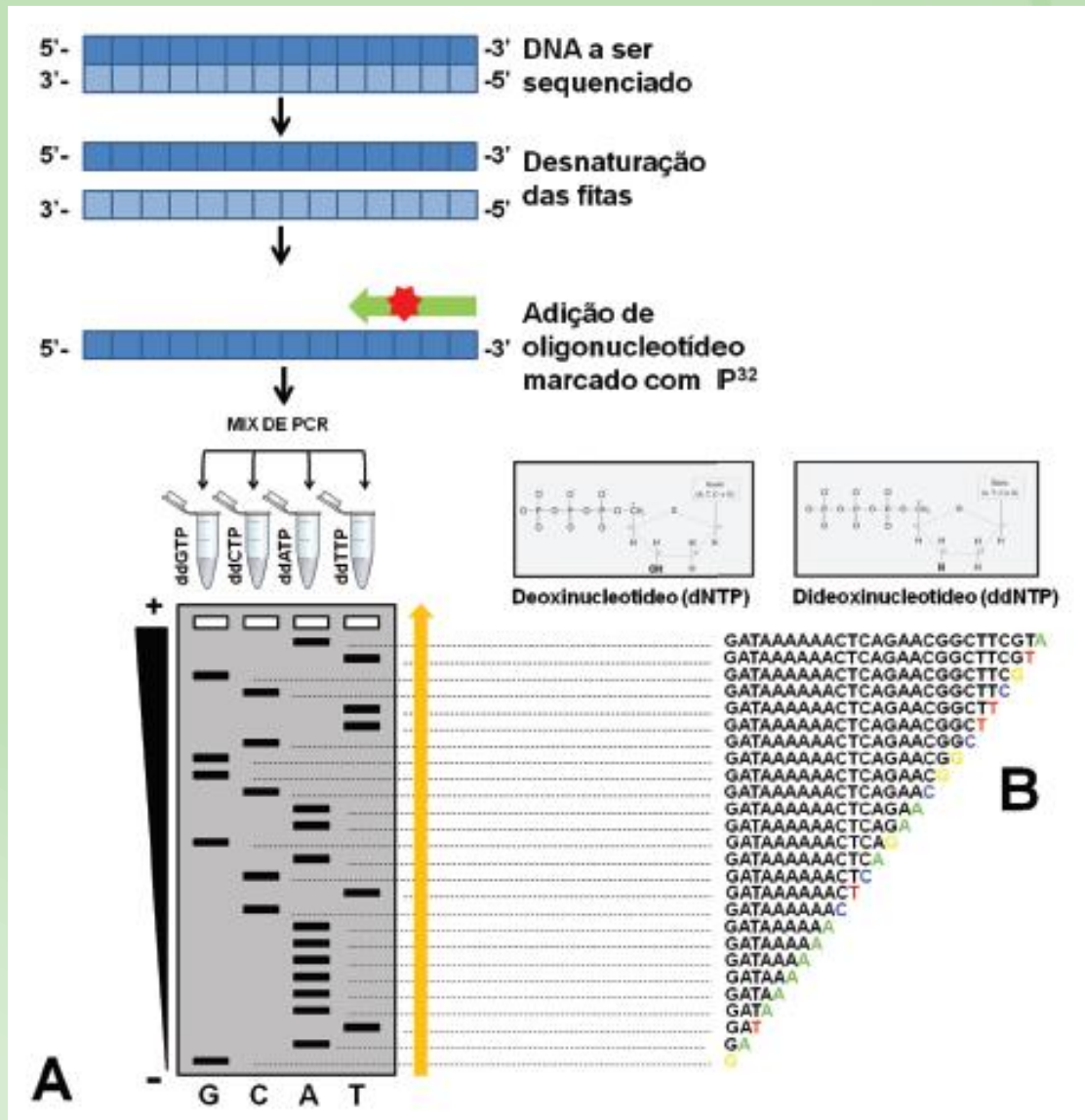
- Até a década de 70, obter a sequência de nucleotídeos de um fragmento de DNA era um processo complexo e difícil
- Duas tecnologias resolveram este problema e revolucionaram as pesquisas científicas:
 - Sequenciamento químico de Maxam-Gilbert (hidrólise química)
 - Sequenciamento de Sanger (reações enzimáticas)

Sequenciamento químico de Maxam-Gilbert



- Marcação do DNA alvo com fósforo radioativo (P³²) após desnaturação e separação das fitas
- Clivagem do DNA em posições específicas (antes de "G", antes de "G" ou "A", antes de "C" ou "T", antes de "C")
- Produtos da clivagem submetidos à eletroforese para separação por tamanho, e perfil de bandas é lido de baixo para cima para determinar a sequência

Sequenciamento de Sanger

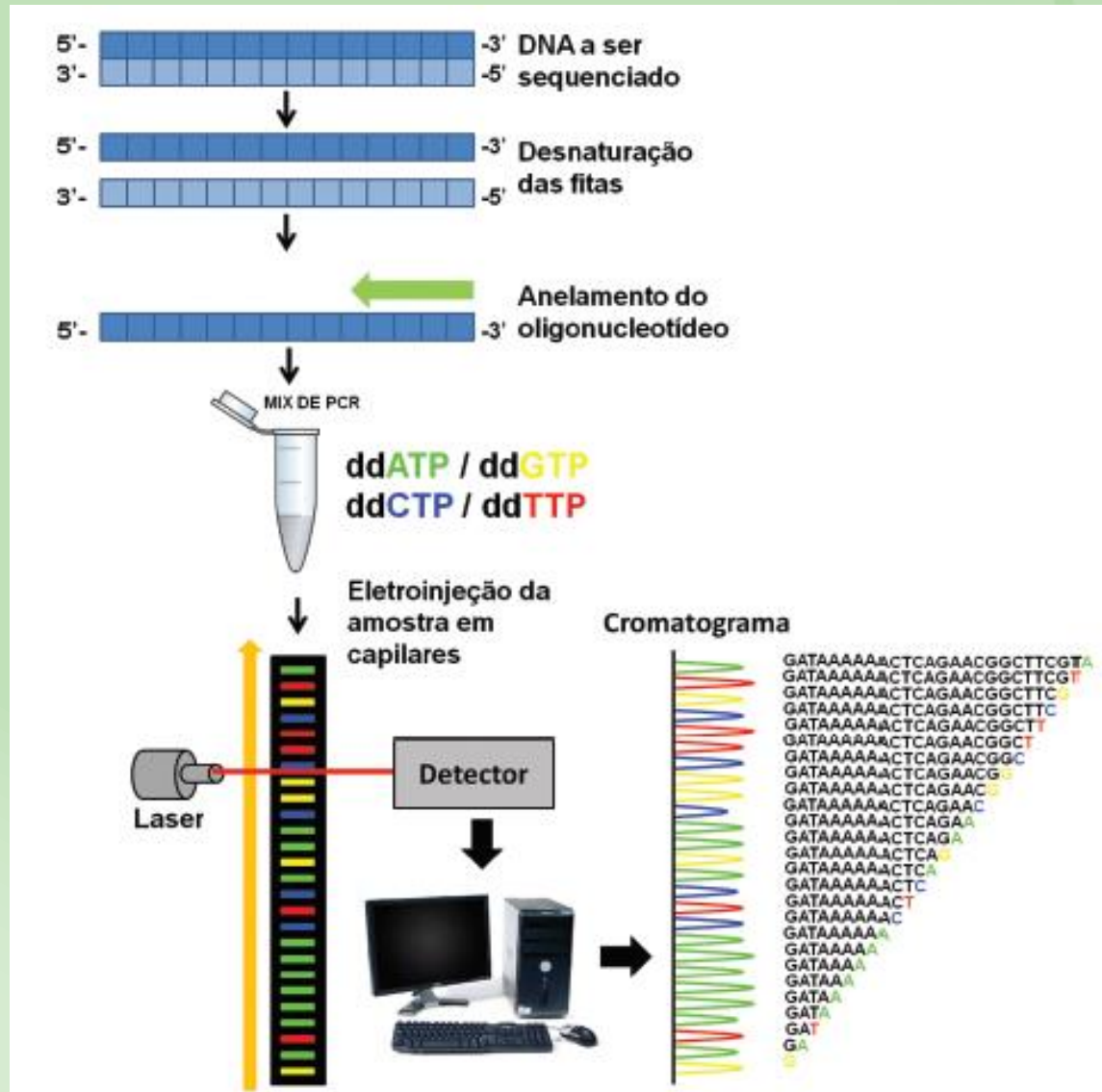


- Reação de PCR com deoxinucleotídeos modificados (dideoxinucleotídeos), marcados com fósforo (P^{32}) ou enxofre (S^{35}) radioativos
- Incorporação de dideoxinucleotídeos interrompe a síntese da nova molécula de DNA
- Produtos das reações de PCR com dideoxinucleotídeos são submetidos à eletroforese para separação por tamanho, e perfil de bandas é lido de baixo para cima para determinar a sequência

Automatização do sequenciamento de Sanger

- Substituição dos dideoxinucleotídeos marcados com radiação por dideoxinucleotídeos marcados com fluoróforos:
 - Menor risco à saúde
 - Uso de fluoróforos diferentes para cada uma das bases: emissão de fluorescência em comprimentos de onda distintos e possibilidade de realizar a reação num tubo único
- Geis de eletroforese substituídos por capilares preenchidos com gel
 - Maior quantidade de amostras analisadas no mesmo período de tempo
 - Maior automação e diminuição do trabalho manual do analista

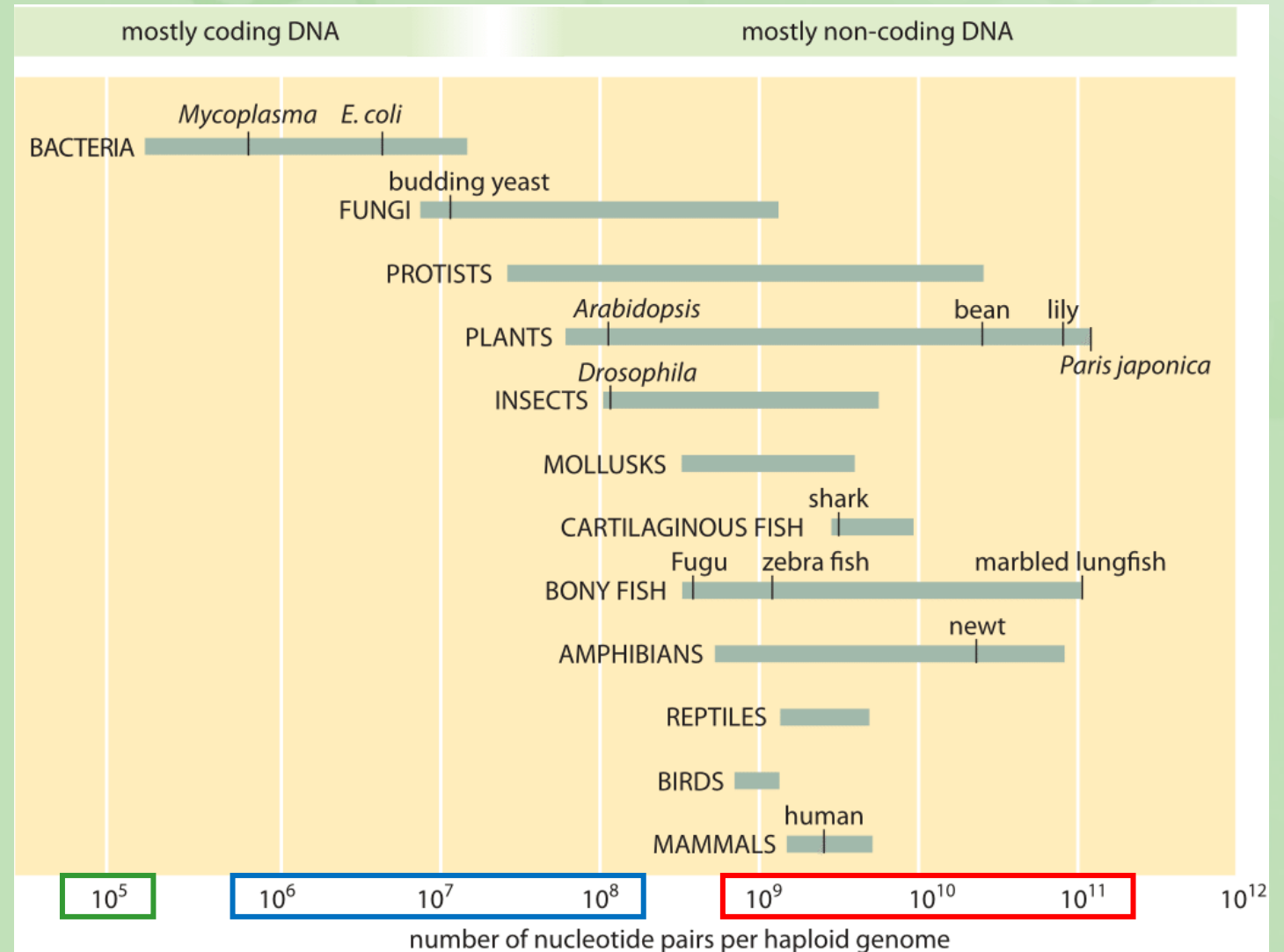
Automatização do sequenciamento de Sanger



- Reação de PCR com deoxinucleotídeos marcados com fluoróforos
- Produtos de PCR migram ao longo dos capilares e passam por um feixe de raios laser que excita os fluoróforos, fazendo com que emitam fluorescência
- A intensidade e comprimento de onda da fluorescência é registrada pelo detector e interpretada pelo computador para gerar o cromatograma
- O cromatograma é decodificado na sequência de nucleotídeos do fragmento

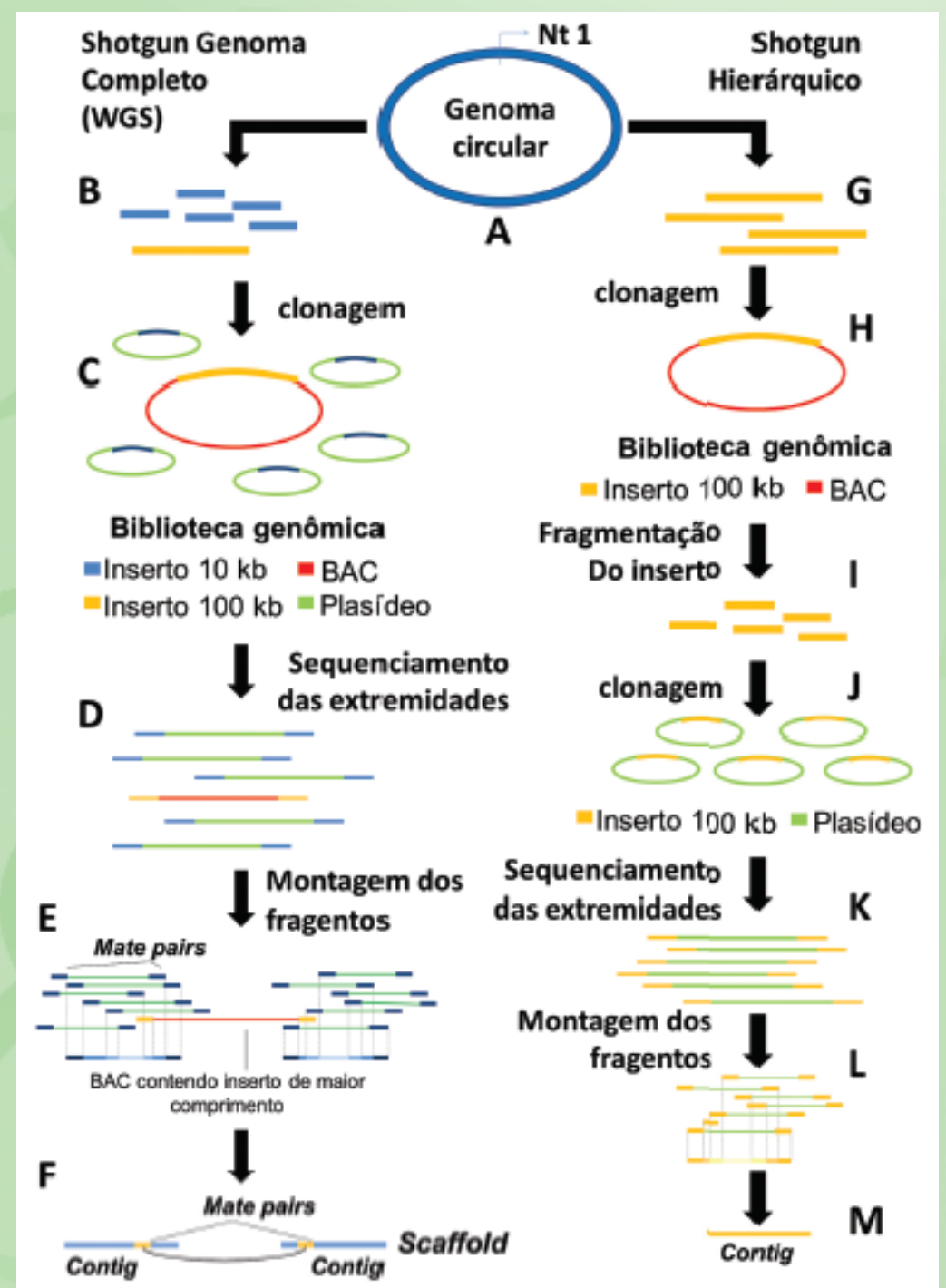
Estratégias de sequenciamento de genoma usando sequenciamento de pequena escala

- Sequenciamento automatizado de pequena escala: fragmentos de ~700 nucleotídeos
- Genomas completos: **milhares**, **milhões** ou até **bilhões** de pares de bases
- Necessidade de fragmentação e posterior montagem para obtenção do genoma completo



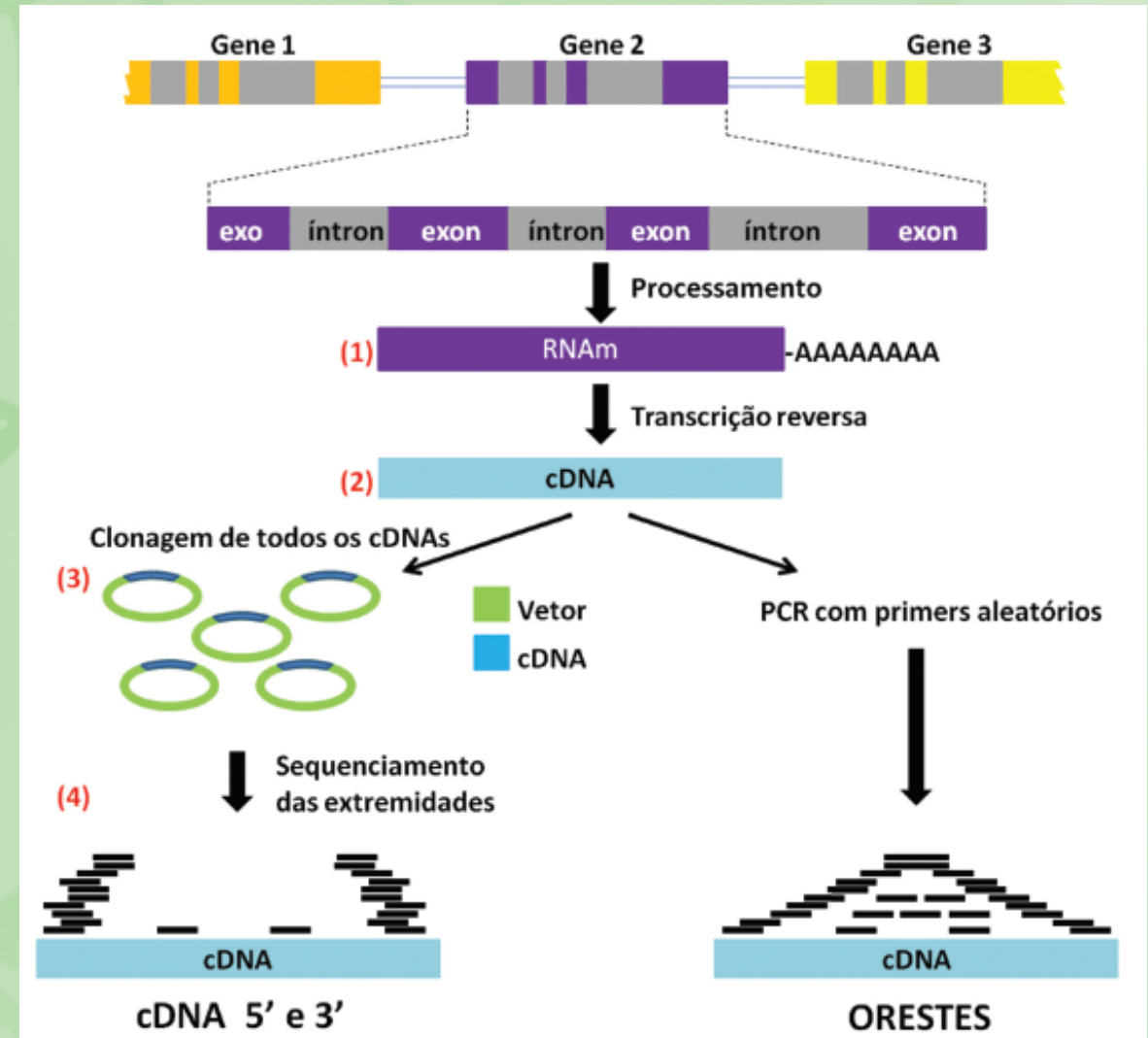
Shotgun

- Fragmentação mecânica do DNA
- Inserção dos fragmentos em vetores e clonagem para construção de bibliotecas genômicas
- Sequenciamento das bibliotecas e montagem dos fragmentos para obtenção de regiões maiores



Estratégias de sequenciamento de transcriptoma usando sequenciamento de pequena escala

- Conversão de RNA mensageiro para DNA utilizando transcriptase reversa
- **EST (Expressed Sequence Tags):** inserção do cDNA em vetores e clonagem, e sequenciamento das extremidades dos insertos
- **ORESTES (Open Reading Frames ESTs):** PCR com primers degenerados aleatórios antes da clonagem e sequenciamento para facilitar o sequenciamento de regiões internas dos genes



Limitações do sequenciamento de pequena escala nos projetos de genomas e transcriptomas

- Erros ou não detecção das bases iniciais
- Preparação das amostras demorada e laboriosa
- Fragmentos pequenos
- **Necessidade de estratégias mais rápidas, baratas, precisas e com maior capacidade de leitura**

Sequenciamento de nova geração (larga escala, 2ª geração)

- Reações não são baseadas em eletroforese
- Alta capacidade de geração de uma grande quantidade dados: genomas sequenciados em uma única corrida
- Preparação de bibliotecas independentes de clonagem

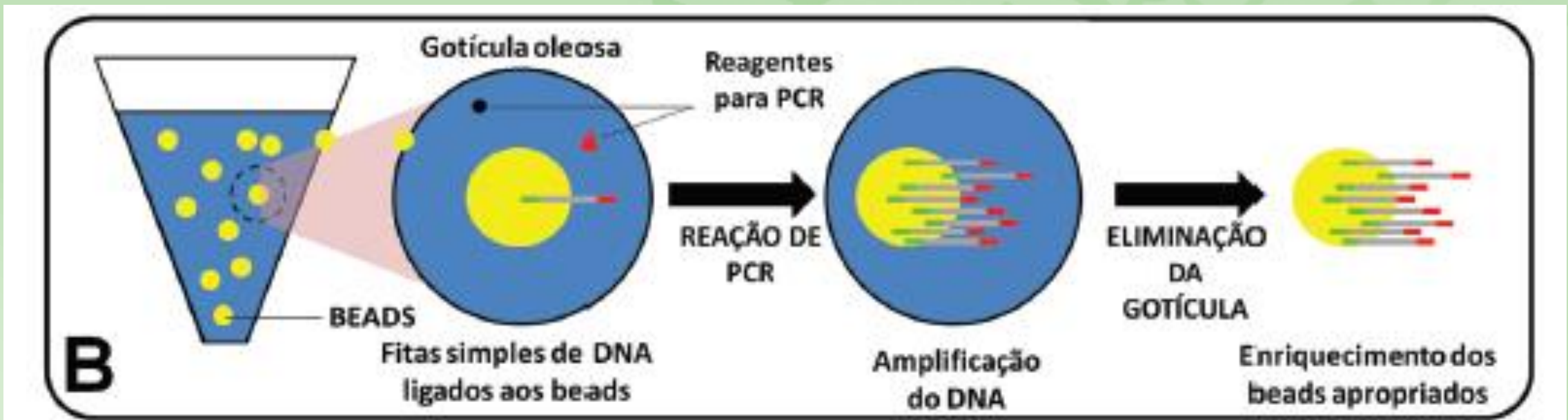
Pirosequenciamento (Plataforma 454 Roche)

- **Preparo da amostra:** fragmentação aleatória do DNA por nebulização, seleção por tamanho e ligação dos adaptadores "A" e "B" às extremidades dos fragmentos



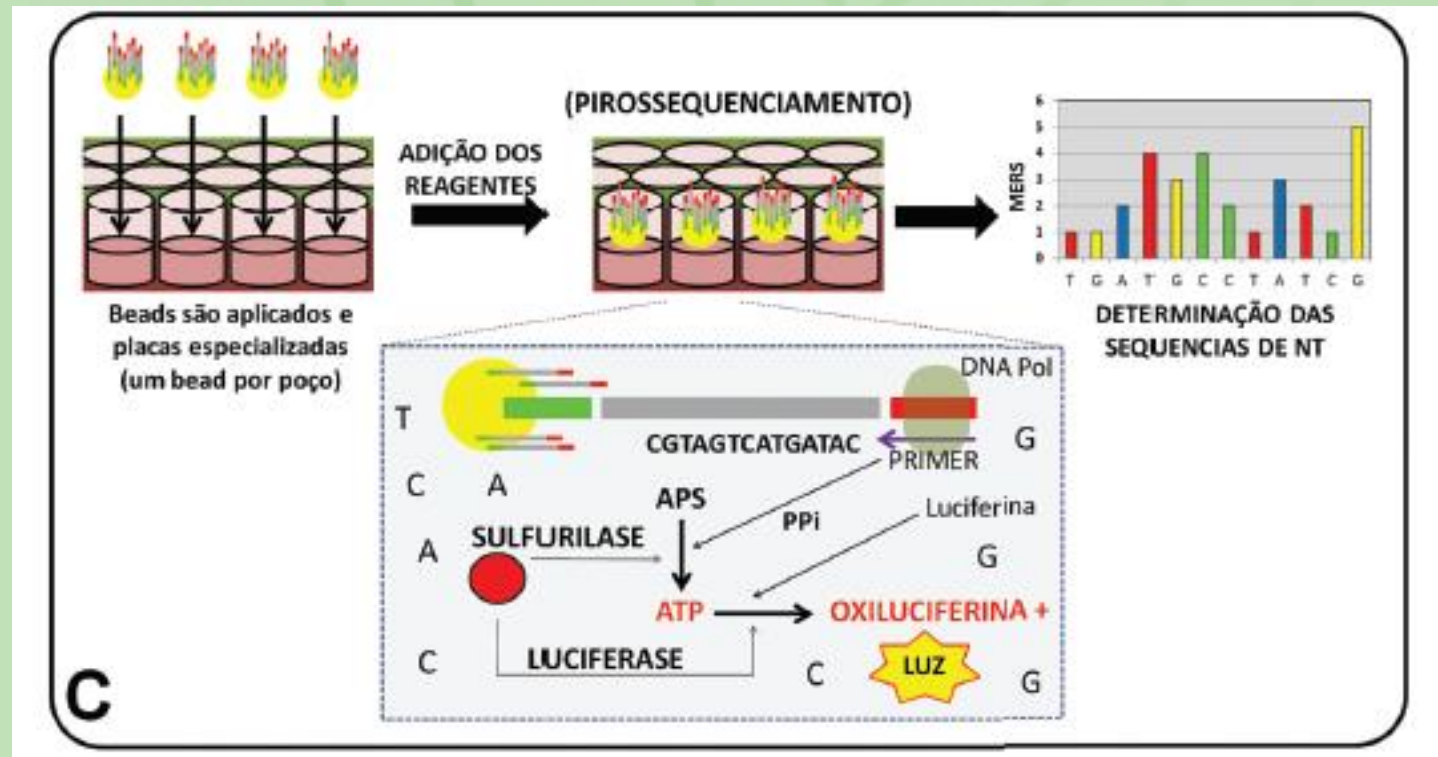
Pirosequenciamento (Plataforma 454 Roche)

- **PCR em emulsão:** ligação dos fragmentos com adaptadores à microesferas magnéticas por meio do adaptador "B" e o adaptador "A" serve como molde para amplificação. Milhares de cópias são produzidas a partir de um único molde.



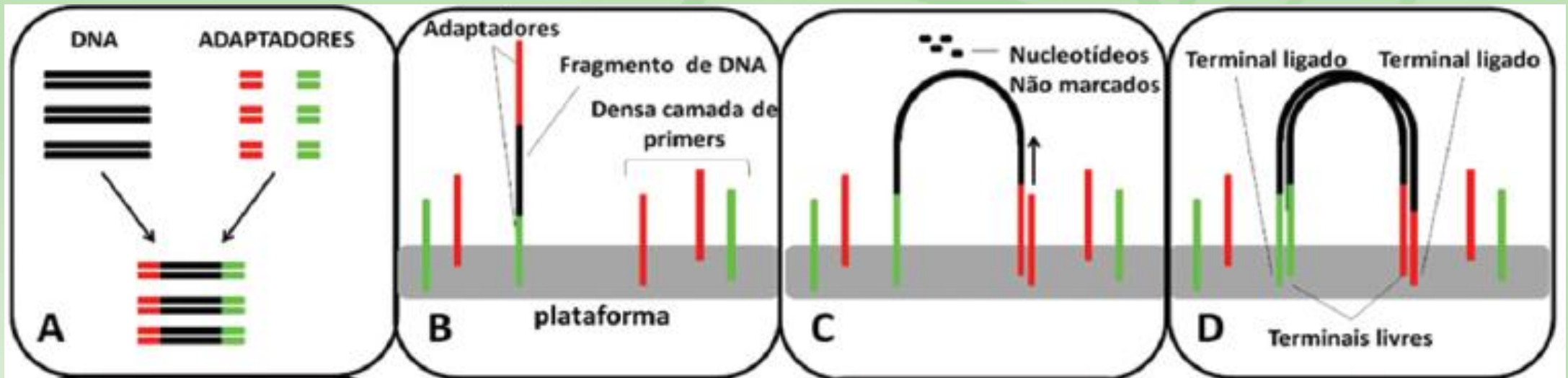
Pirosequenciamento (Plataforma 454 Roche)

- **Sequenciamento:** as microesferas contendo fragmentos amplificados são aplicadas em placas especializadas, e os reagentes para sequenciamento são adicionados aos poços
- Pirofosfato inorgânico (PPi) é liberado a cada incorporação de um nucleotídeo durante a reação de sequenciamento, e convertido em ATP pela ATP sulfurilase
- ATP serve como energia para oxidação de luciferina em oxiluciferina pela enzima luciferase, emitindo luz



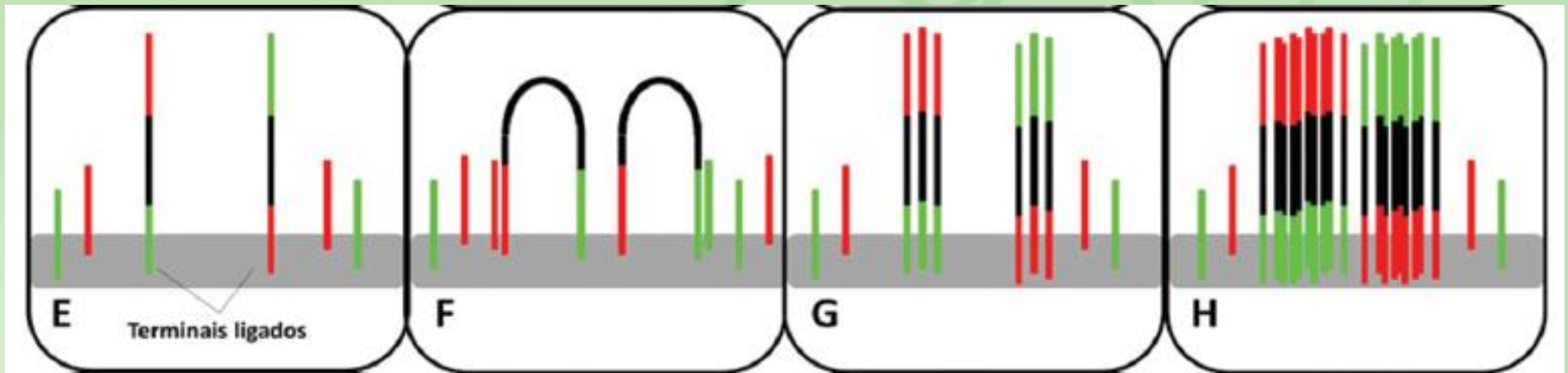
Illumina

- Preparação: fragmentação do DNA por nebulização, seleção por tamanho e ligação dos adaptadores às extremidades dos fragmentos
- Ligação dos fragmentos à plataforma, que contém uma densa camada de primers. Ocorre a incorporação de nucleotídeos não marcados com fluorescência até que ocorra a amplificação de todo o fragmento
- Há formação da estrutura em ponte (amplificação em ponte), com dois adaptadores presos à placa e dois livres



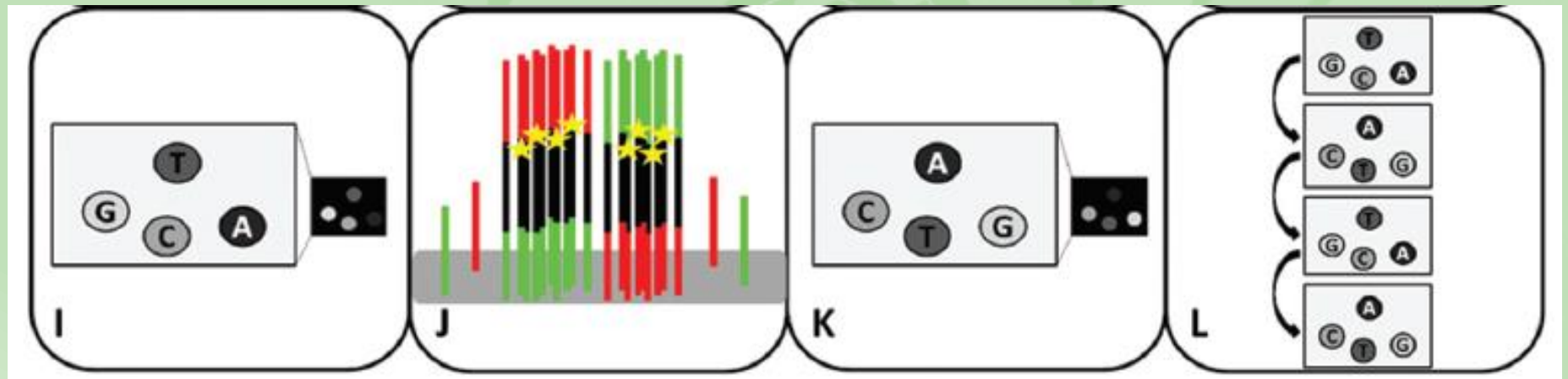
Illumina

- Novos ciclos de amplificação em ponte ocorrem, até a formação de clusters com mais de um milhão de cópias do mesmo fragmento



Illumina

- Após a formação dos clusters, dideoxinucleotídeos são adicionados antes da próxima amplificação
- Após a incorporação dos dideoxinucleotídeos durante a amplificação, um feixe de raios laser excita os fluoróforos e a luz emitida é registrada pelo detector, fazendo a leitura de bases naquela posição
- Ocorre uma lavagem para remoção dos grupos bloqueadores presentes nas extremidades 3' dos dideoxinucleotídeos para que a reação possa continuar
- Esse processo é repetido sucessivamente até que toda a extensão do fragmento de DNA seja polimerizada e o fragmento seja sequenciado
- Por fim, as leituras são decodificadas para determinar a sequência de bases dos fragmentos

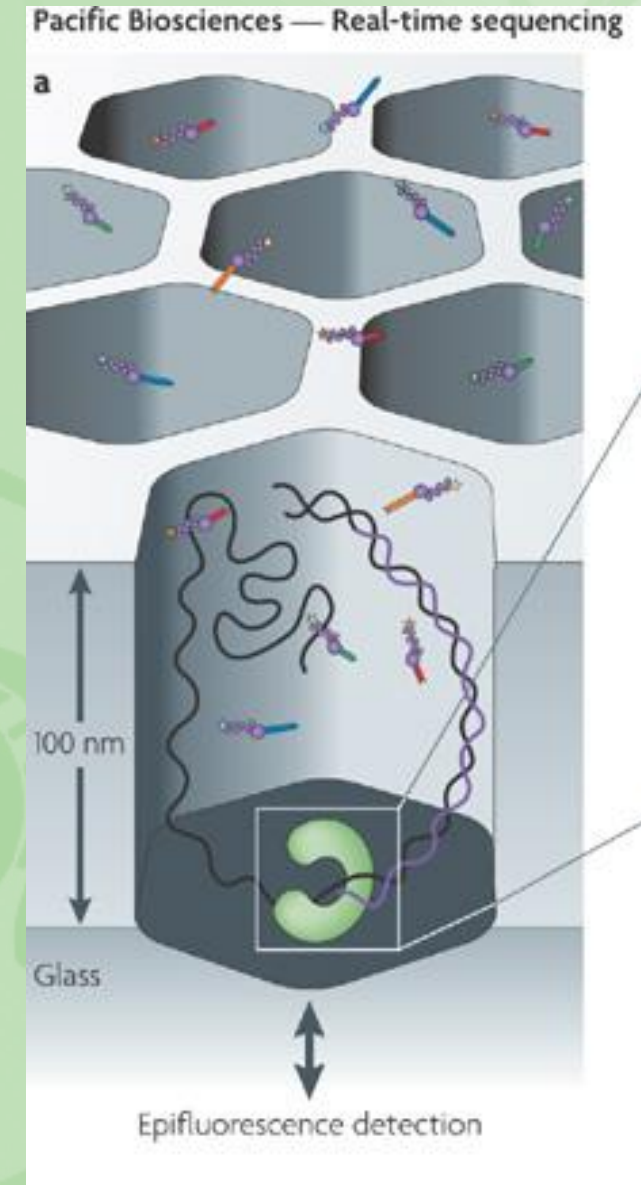


Sequenciamento de nova geração (larga escala, 3ª geração)

- Capacidade de sequenciamento de uma única molécula de DNA
- Sem necessidade de amplificação
- Altíssima capacidade de geração de um grande volume de dados em curto período de tempo: um genoma humano pode ser sequenciado em uma única corrida

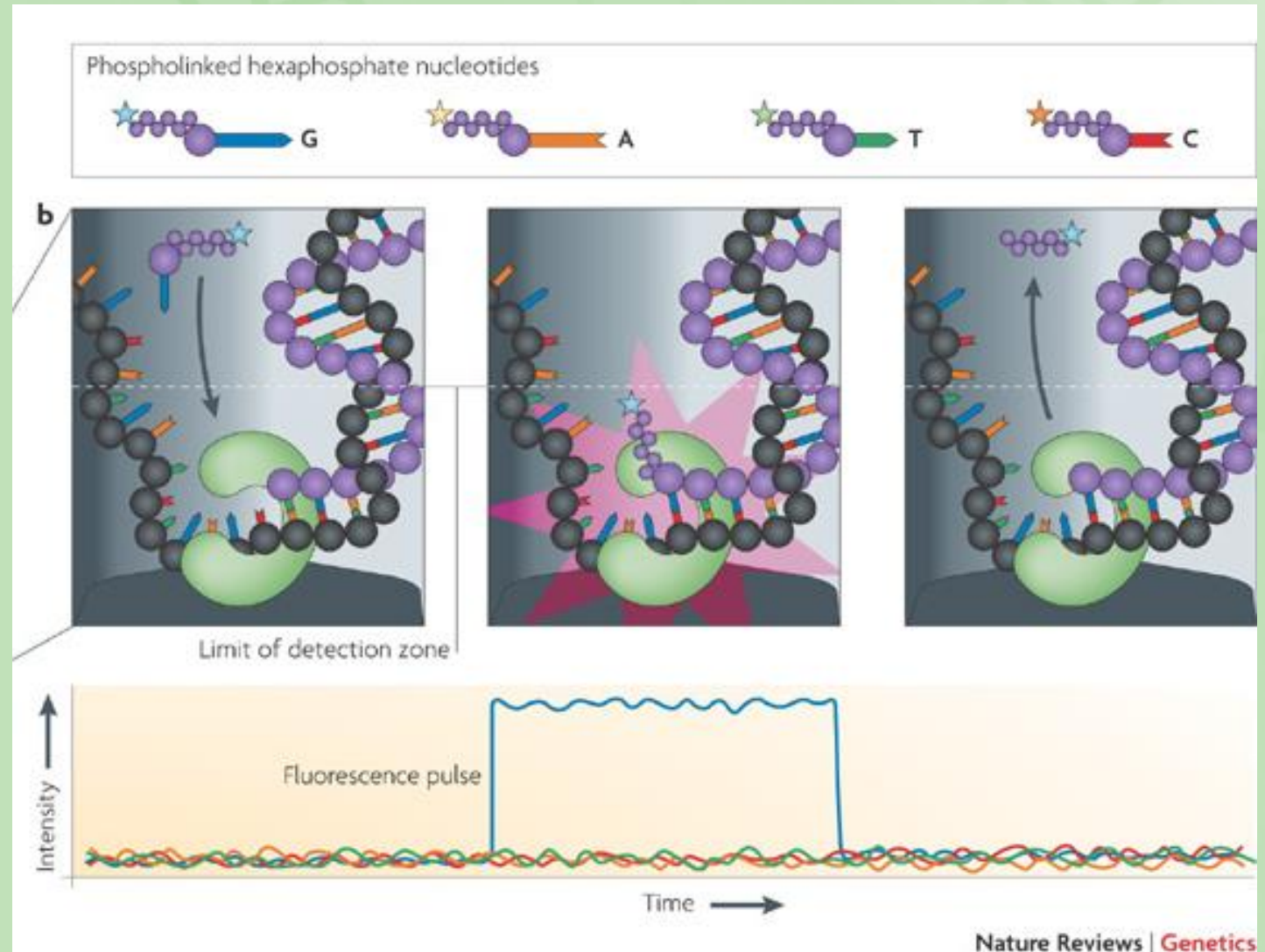
Pacific Biosciences SMRT Sequencing (*Single Molecule Real Time Sequencing*)

- Janela de observação em nano-escala (ZMW, *zero-mode waveguide*), com um volume extremamente reduzido, suficiente para visualizar a incorporação de um único nucleotídeo pela DNA polimerase
- Uma única DNA polimerase contendo uma única molécula de DNA molde é fixada no fundo da ZMW



Pacific Biosciences SMRT Sequencing (Single Molecule Real Time Sequencing)

- Nucleotídeos com fluoróforos são utilizados, e quando um nucleotídeo é incorporado a marcação fluorescente é clivada, emitindo luz, que é detectada e transformada em dado de sequência

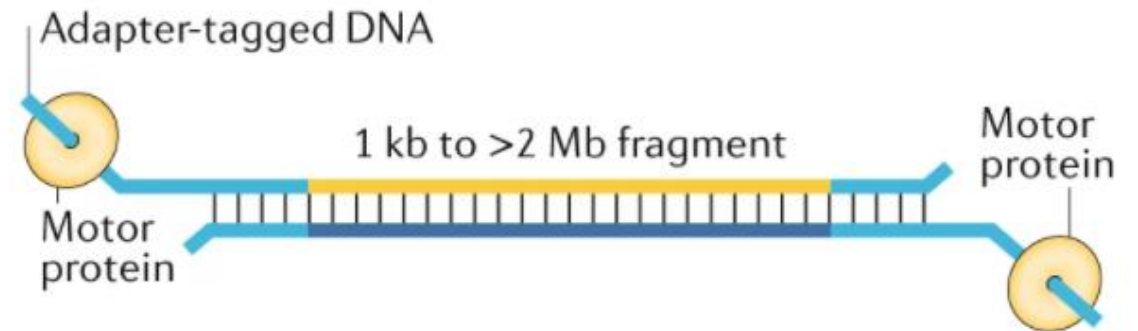


Oxford Nanopore Technologies

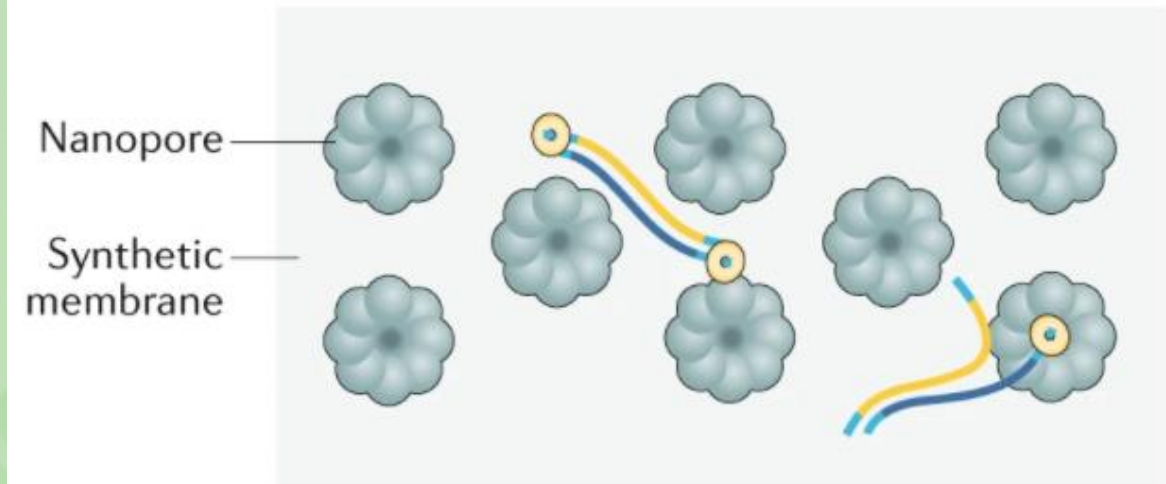
- DNA é marcado com adaptadores com proteínas motoras em uma ou ambas as extremidades e é combinado à proteínas carregadoras, que o direcionarão aos nanoporos
- A plataforma de sequenciamento contém milhares de nanoporos de proteicos associados à uma membrana sintética

b ONT sequencing

Template topology

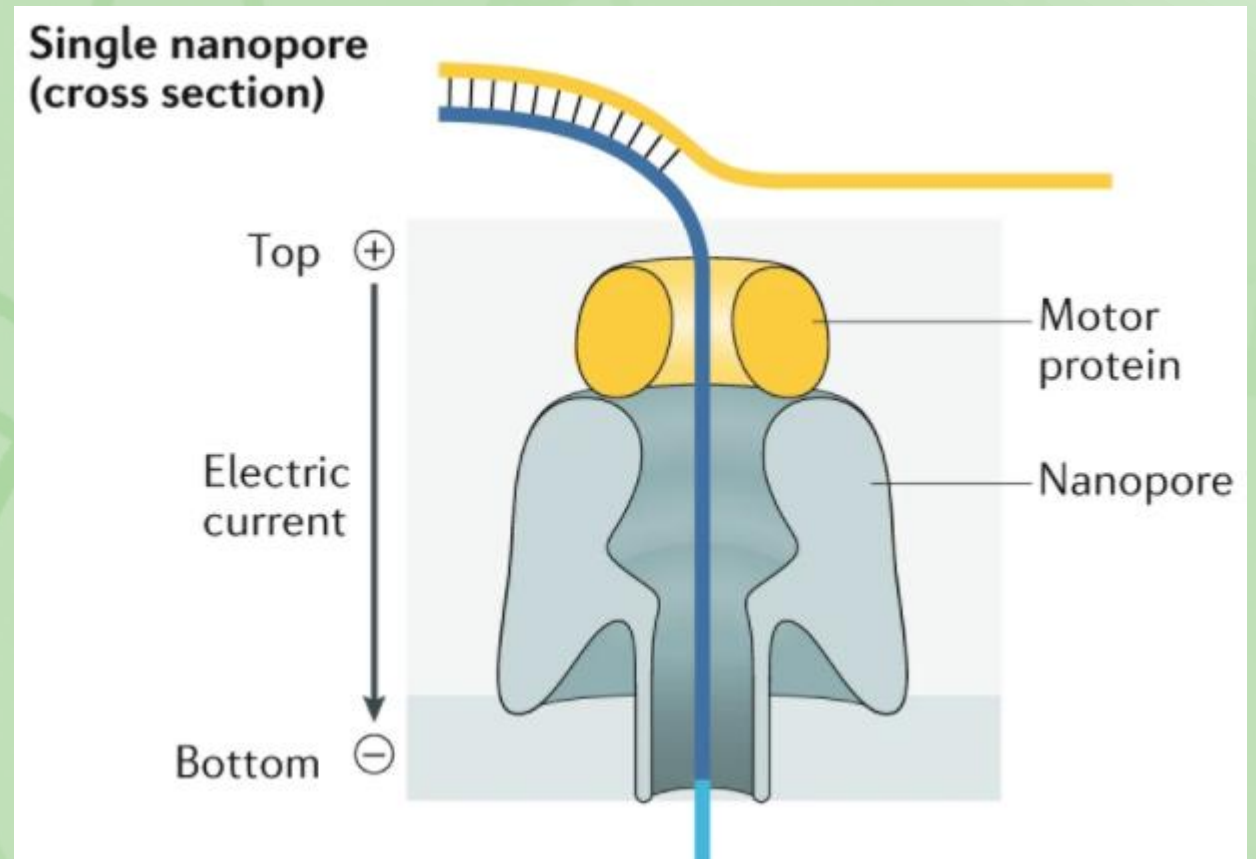


Flow cell (top view)



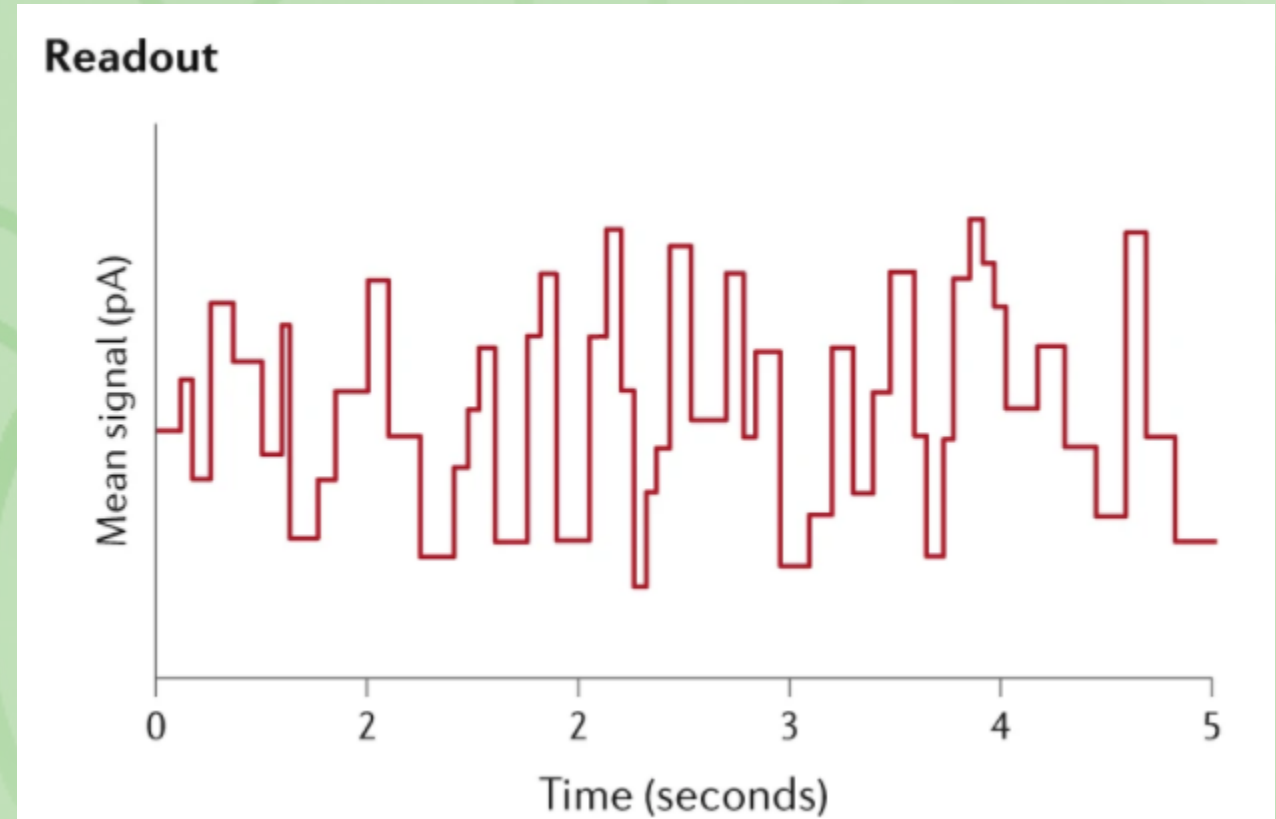
Oxford Nanopore Technologies

- O adaptador se insere na abertura do nanoporo, e a proteína motora começa a separar as fitas do DNA
- Uma corrente elétrica é aplicada e, em conjunto com a proteína motora, conduz o DNA carregado negativamente através do poro numa velocidade de ~450 bases por segundo



Oxford Nanopore Technologies

- À medida que o DNA se move pelo poro, causa perturbações à corrente elétrica, as quais são específicas para cada um dos nucleotídeos
- O perfil de mudanças na corrente elétrica pode ser utilizado para identificar a sequência de bases da molécula



Comparativo entre metodologias

Método	Sanger	454	Illumina	PacBio	Nanopore
Comprimento dos reads	400 - 900 pb	700 bp	100 – 300 pb	10 – 100 kb	Variável (até 1000 kb)
Taxa de erro	0.01 %	0.1 %	0.1%	5 – 15%*	5 – 20%*
Eficiência (bases por corrida)	1.9 - 84 Kb	1 Mb	200 – 600 Gb	10 – 20 Gb	5 – 10 Gb
Tempo de corrida	20 min – 3 horas	24 horas	1 – 3 dias	~ 30 horas	1 minuto até 72 horas
Prós	Alta confiabilidade	Velocidade	Alta confiabilidade e custo baixo	Reads longos, velocidade e alta eficiência	Reads longos, velocidade e alta eficiência
Contras	Baixa eficiência	Baixa eficiência e alto custo	Reads curtos, velocidade	Taxa de erro elevada e alto custo	Taxa de erro elevada e alto custo

Adaptado de:

Basal & Boucher et al. 2019 **iScience**. DOI: [10.1016/j.isci.2019.06.035](https://doi.org/10.1016/j.isci.2019.06.035)

Liu et al. 2012 **BioMed Research International**. DOI: [10.1155/2012/251364](https://doi.org/10.1155/2012/251364)

<https://www.pacb.com/products-and-services/sequel-system/>

* Em baixa cobertura