

BIOLOGIA MOLECULAR PARA O BACHARELADO – BIO 0307
Sequenciamento de genoma - 2017

Etapas consecutivas, em comum para o sequenciamento massivo, em paralelo de DNA total de um genoma (métodos diferentes):

- 1) Fragmentação do DNA genômico;
- 2) Ligação a adaptadores para formação da biblioteca de DNA;
- 3) Amplificação *in vitro* dos fragmentos de DNA genômico;
- 4) Sequenciamento, em matriz sólida, de todas as amostras (massivo) ao mesmo tempo (em paralelo);
- 5) Leituras mais curtas do que o sequenciamento de Sanger;
- 6) Captura e leitura digitais da sequência;
- 7) Formação de *contigs* e análise bioinformática.

Tabela comparativa entre três plataformas

454™, Roche	SOLiD™*, ABI (Applied Biosyst)	Illumina
<u>Amplificação <i>in vitro</i> dos fragmentos de DNA genômico</u>		
PCR em emulsão (emPCR, Fig. 1)	PCR em emulsão (Fig 1)	Amplificação em ponte (em vidro, Fig 2)
<u>Base molecular para o sequenciamento</u>		
Sequenciamento por síntese	Sequenciamento por ligação	Sequenciamento por síntese
Detecção do pirofosfato liberado após formação da ligação fosfodiéster (Fig. 3)	Ligação de oligonucleotídeo a iniciador universal. As duas bases a 3' no oligonucleotídeo são sequenciadas (Fig. 4)	Nucleotídeos com fluoróforo em 3' que permite detecção e bloqueia temporariamente a progressão da reação (Fig 5)
<u>Método para detecção da sequência</u>		
Sulfurilase converte pirofosfato em ATP que é usado pela luciferase para emissão de luz (Fig 3)	LASER detecta fluorocromo a 5' do oligonucleotídeo utilizado na ligação	Quatro filtros para detecção de fluorescência específica para cada um dos quatro nucleotídeos
<u>Método complementar para iniciar novo ciclo de sequenciamento com o mesmo molde</u>		
Apirase degrada nucleotídeos (trifosfatados)	Sais de prata degradam três bases de cada sonda restituindo a extremidade 5'. Após dez ciclos, o polímero é desnaturado e todo o processo se repete por quatro vezes.	Clivagem enzimática do fluoróforo após emissão de fluorescência restitui hidroxila em 3'
<u>Média de leitura (bases)</u>		
700 bases (para GS FLX)	75 bases	125 bases (modelo HiSeq 2500)
<u>Rendimento (Mb)</u>		
700 Mb (para GS FLX)	120.000 Mb (para modelo 5500)	10.000–400.000 Mb (HiSeq2500)
<u>Tempo de corrida</u>		
24 horas a dias (GS FLX)	24 horas a 6 dias (para 5500)	Até 6 dias (para HiSeq 2500)
<u>Número de sequências por corrida</u>		
1 X 10 ⁶ (para GS FLX)	1 X 10 ⁸ (para 5500)	3 X 10 ⁶ a 4 X 10 ⁹ (HiSeq2500)

*SOLiD: Support Oligonucleotide Ligation Detection

Métodos para amplificação genômica

Após fragmentação do DNA genômico, sequências de dois tipos de adaptadores A ou B, são ligadas às extremidades de cada fragmento. Os fragmentos são distribuídos aleatoriamente seja a uma suspensão de água e óleo (PCR em emulsão, Fig. 1) ou sobre uma superfície de vidro que contém sequências complementares às dos adaptadores (para amplificação em ponte, Fig. 2)

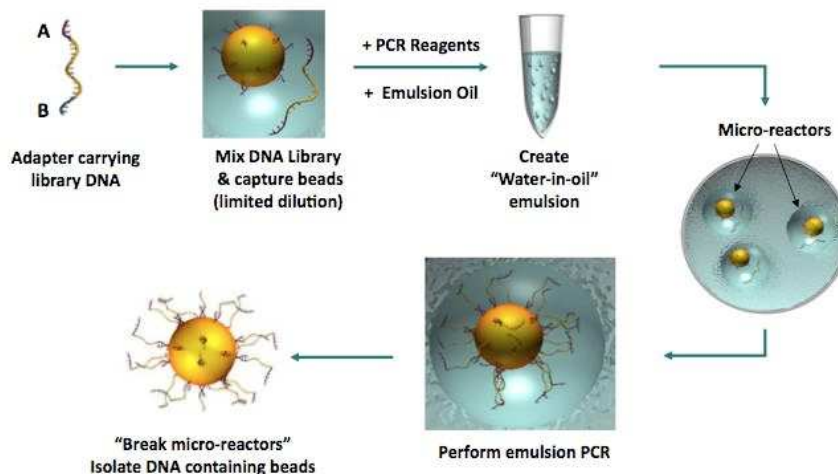


Figura 1: Para PCR em emulsão, o DNA genômico, fragmentado, ligado a adaptadores é misturado a *beads* hidrofílicas. Isso permite a ligação de uma fita simples a cada *bead* por complementariedade entre o adaptador do fragmento e iniciadores para PCR ligados à *bead*. Esta solução é misturada a uma emulsão lipídica e a reagentes para PCR. Forte agitação faz com que gotículas de água contendo *bead* fiquem emulsificadas em óleo. A emulsão é distribuída a microrreatores, de modo que cada um conterá uma *bead*. A PCR será realizada nessas condições. A cada ciclo, a desnaturação do produto permitirá a hibridação com novo iniciador contido na própria *bead*. A *bead* contendo milhões de cópias do mesmo fragmento é separada da emulsão e terá o DNA submetido ao sequenciamento.

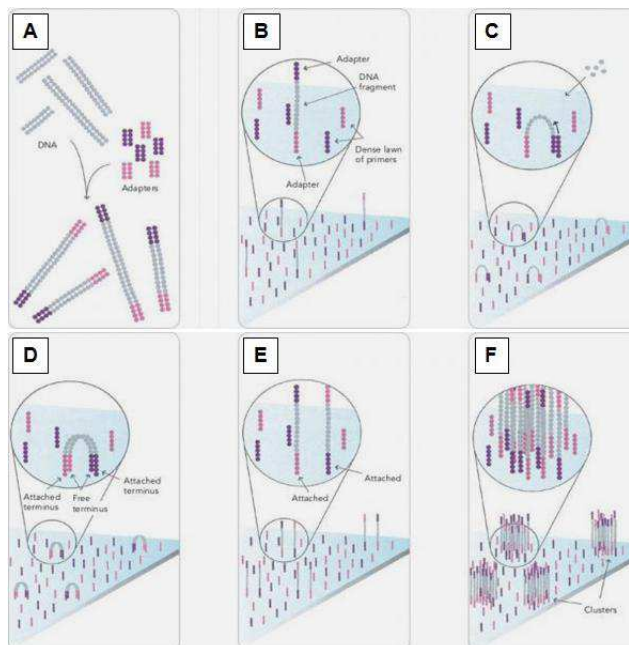


Figura 2: Após a ligação de fragmentos de DNA a adaptadores (A), eles são distribuídos a uma matriz sólida, onde se ligam aleatoriamente (B). A extremidade livre do fragmento de DNA genômico, em fita simples, liga-se por complementariedade a sequências da placa que servirão como iniciadores para síntese do DNA (C). Essa curvatura do DNA com ambas as extremidades ancoradas na placa permite a amplificação em ponte do DNA (D). Uma vez concluída, o molde estará disponível para novas cópias (E), formando, ao final, um *cluster* com cerca de um milhão de cópias (F).

Métodos para sequenciamento genômico

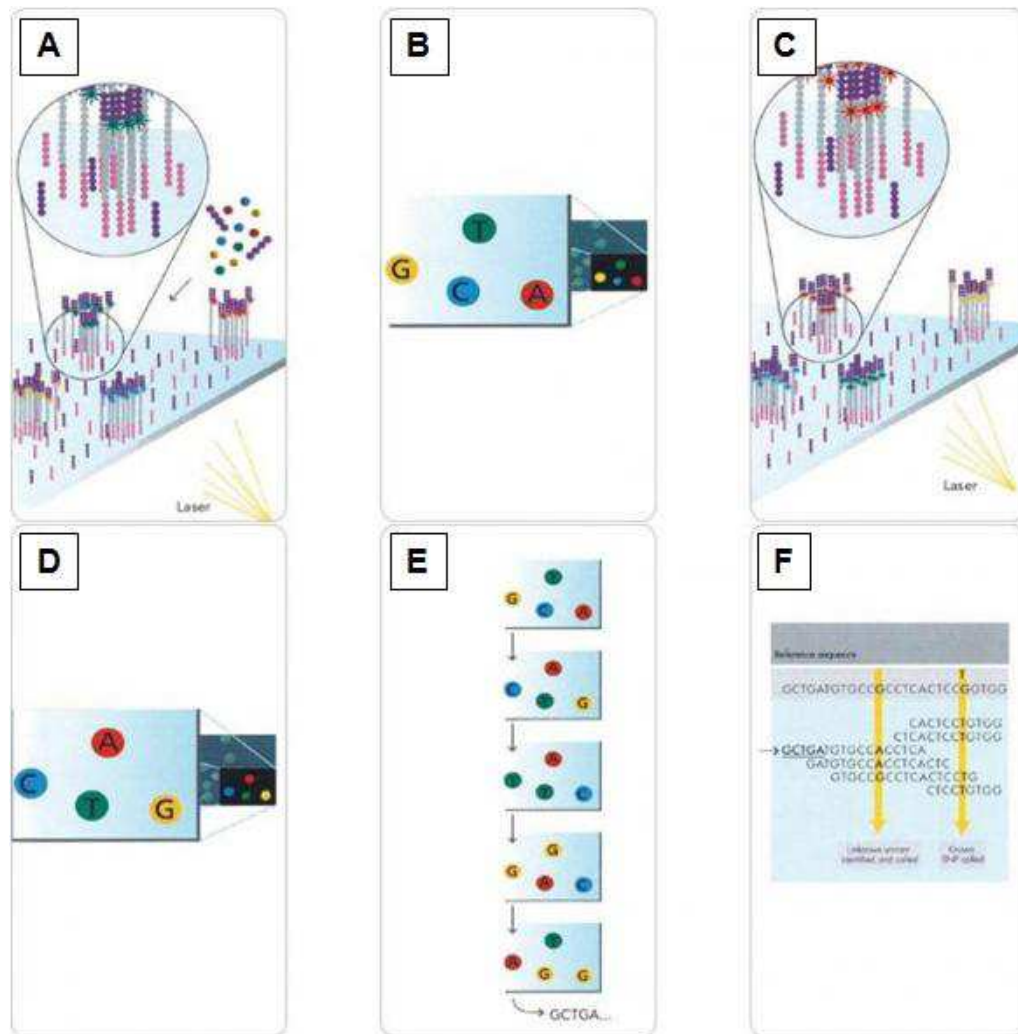


Figura 3: Para o processo de sequenciamento por síntese na plataforma Illumina, uma mistura de nucleotídeos e polimerase do DNA é acrescentada à placa contendo os *clusters* de fragmentos amplificados de fita simples de DNA (A). Os nucleotídeos são bloqueados a 3' pelo fluoróforo, implicando na parada da reação quando o nucleotídeo é incorporado ao DNA. Para cada um dos quatro tipos de nucleotídeos, há uma espécie de fluoróforo. Por isso, os quatro tipos são sempre adicionados juntos. Após incorporação do nucleotídeo, há lavagem da placa e o LASER escaneia a superfície e captura a fluorescência emitida (B). Cada ciclo – base sequenciada – resulta em um número de imagens, uma para cada cor. Após a emissão de fluorescência, há clivagem química do fluoróforo, com recuperação da hidroxila em 3' para novo ciclo. O processo é repetido (C) e uma nova base é registrada para cada posição (D). Novos ciclos permitem a leitura da sequência (E), a qual será alinhada a um genoma de referência (F).

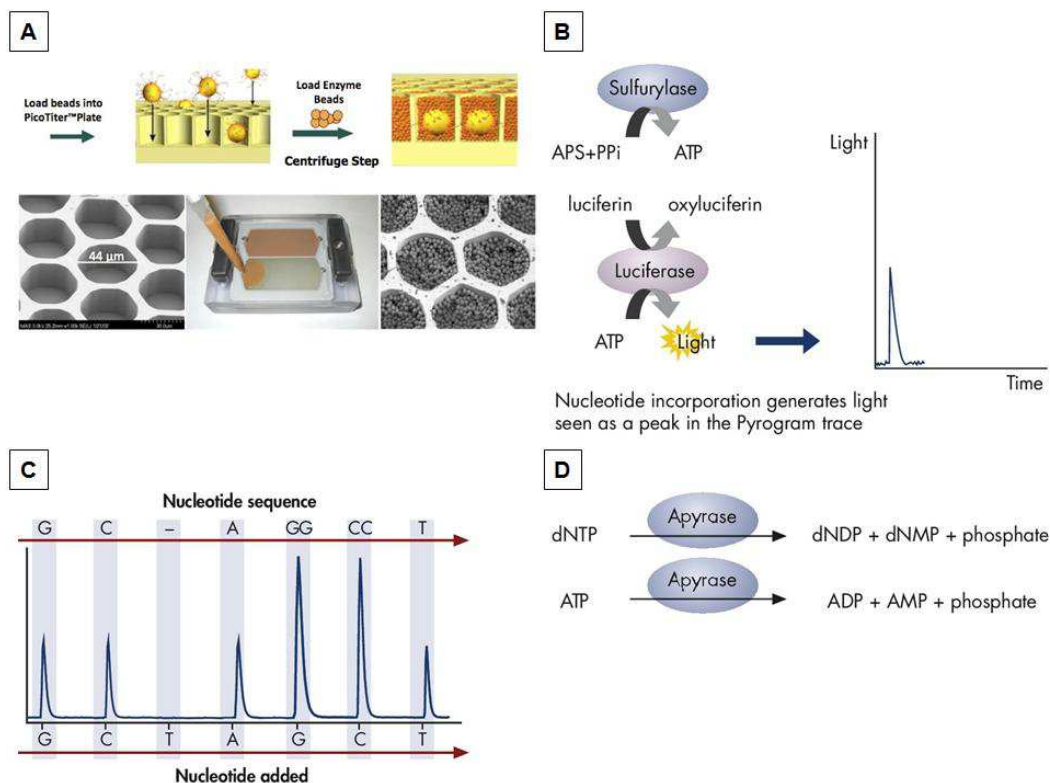


Figura 4: Para pirosequenciamento na plataforma 454, as *beads* são adicionadas a uma placa com poros de 44 µm de diâmetro, onde também são acrescentadas as enzimas (polimerase do DNA, sulfurilase de ATP, luciferase), além de adenosina 5'-fosfossulfato (APS) e luciferina (**A**). Nucleotídeos são adicionados sequencialmente, um tipo de cada vez. Se o primeiro nucleotídeo for complementar à base do molde, a 3' do iniciador, uma ligação fosfodiéster é formada e há a liberação de pirofosfato (PPi) em quantidade equimolar (não mostrado). A sulfurilase de ATP converte PPi a ATP na presença de APS. O ATP favorece a catálise da conversão de luciferina a oxiluciferina pela luciferase, produzindo luz visível de modo proporcional à quantidade de ATP (**B**). A luz gerada é detectada pelo sensor da câmera CCD e registrada como um pico do pirograma (**C**). A altura do pico é proporcional ao número de nucleotídeos incorporados ao DNA. (**C**) Apyrase é uma enzima que degrada nucleotídeos transformando-os em mono e difosfatos. O acréscimo de apirase à reação degrada os nucleotídeos não incorporados ao DNA, incluindo o ATP. Quando a degradação é concluída, um novo nucleotídeo é acrescentado, iniciando nova reação de sequenciamento. Deve-se notar que o dATP usado na reação de sequenciamento não é reconhecido pela luciferase.

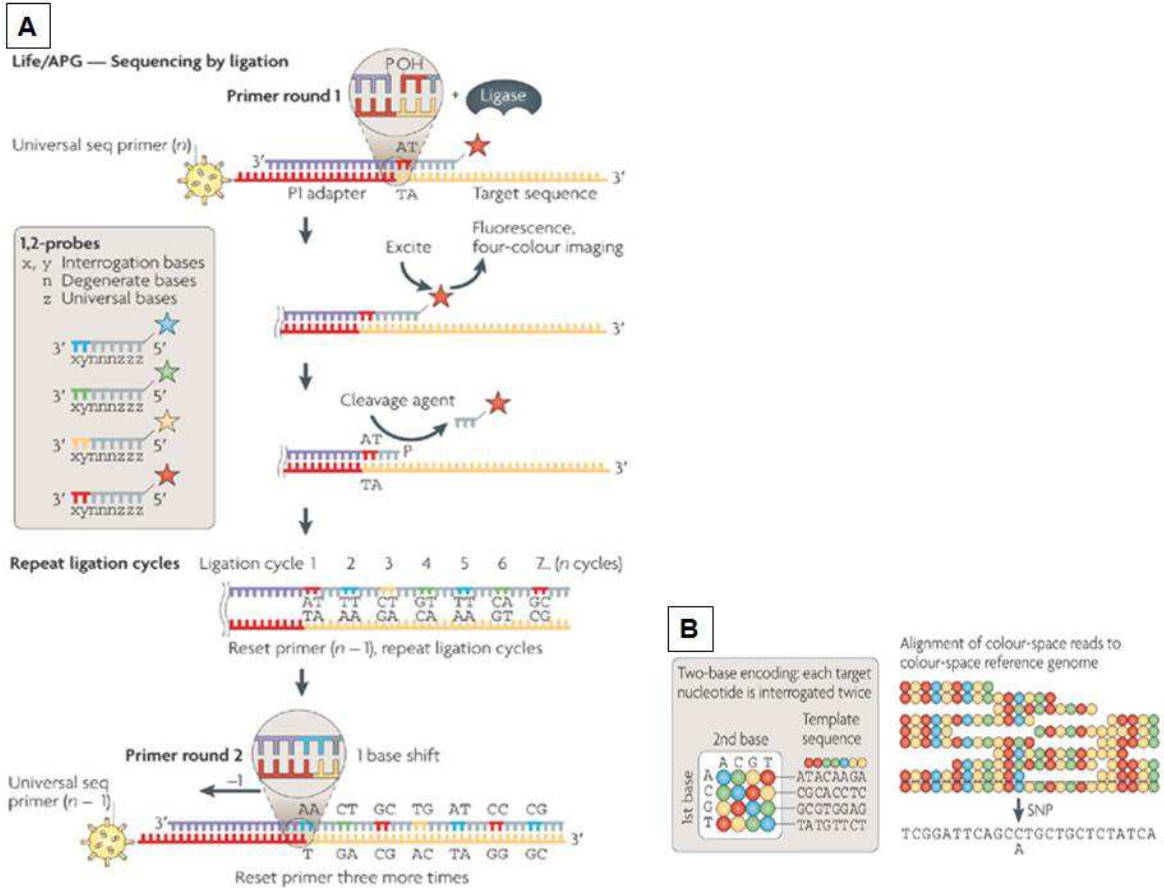


Figura 5: O sequenciamento de DNA por ligação na plataforma SOLiD requer a hibridação de um oligonucleotídeo universal à sequência do adaptador e uma biblioteca de sondas, marcadas a 5' com um de quatro fluorocromos. Não há polimerização, mas ligação da sonda ao oligonucleotídeo universal. Após a emissão de fluorescência pela sonda, íons de prata clivam a ligação não usual – fosforotiolato – que há entre as quinta e sexta bases da sonda, gerando a extremidade 5' fosforilada da sonda, ligada ao oligonucleotídeo universal. O ciclo SOLiD é repetido mais nove vezes, permitindo a ligação de um total de dez sondas de cinco bases cada (50 pb). Este oligonucleotídeo estendido é desnaturado do molde e os dez ciclos repetidos quatro vezes. Cada sonda só interroga duas bases de cada vez: a base x (3') e a anterior (y). À 5' da base (y) há três bases degeneradas, que serão mantidas no polímero, seguidas de três bases universais (ex. inosina), as quais serão removidas pelo tratamento com prata. A cada repetição dos dez ciclos, o oligonucleotídeo universal, utilizado tem uma base a menos em 5'. Assim, após a desnaturação do polímero formado pelo primeiro grupo de dez sondas, as bases a serem sequenciadas serão deslocadas em uma posição. A segunda base que foi sequenciada (y) não o será na segunda vez e a primeira base (x) será resequenciada. **(B)** Dezesesseis tipos de sondas são possíveis, de acordo com a combinação de bases, em duas posições na extremidade 3' (4^2). Quatro tipos de oligonucleotídeos associam-se ao mesmo fluoróforo (por exemplo, sondas com AA, CC, GG ou TT têm o fluoróforo azul). Cada molde é interrogado duas vezes e compilado como tiras de *bits* de cores e espaços (bases degeneradas), sendo alinhadas com sequências do genoma de referência.