

Universidade de São Paulo (USP) – Escola Superior de Agricultura Luiz de Queiroz (ESALQ)

André Sader Garcia

Gustavo Medeiros Nyssen

Júlia Niero Costa

Marcelo Aguiar Carrieri

Pedro Coutinho

Sequenciamento de Nova Geração

Piracicaba, 2016

SUMÁRIO

1	INTRODUÇÃO	3
2	REVISÃO DE LITERATURA	4
2.1	SEQUENCIAMENTO MANUAL (DE SANGER).....	4
2.2	SEQUENCIAMENTO AUTOMÁTICO	4
2.3	SEQUENCIAMENTO DE NOVA GERAÇÃO	5
2.3.1	<i>Minlon</i>	5
2.3.2	<i>Illumina</i>	5
2.3.3	<i>PacBio RS II</i>	6
3	OBJETIVOS	6
4	METODOLOGIA	8

1. INTRODUÇÃO

Sequenciamento consiste em determinar a ordem das bases nitrogenadas (A,C,G,T) presentes em determinada molécula de DNA.

Em 1977, Fred Sanger e Alan Coulson publicaram um trabalho, mostrando que era possível sequenciar o DNA dos organismos. Sua metodologia consistia basicamente em utilizar nucleotídeos modificados, didesoxiribonucleotídeos, para parar a síntese do DNA e fazer o sequenciamento. Sanger, devido a tal feito, é, por muitos, considerado o pai da genômica.

Com o desenvolvimento da tecnologia, é possível utilizar novos meios para realizar esse sequenciamento, como o uso de algumas plataformas, como a da Applied Biosystems, denominada SOLiD System e o HeliscopeTrue Single Molecule Sequencing (tSMS), da Helicos. Essas tecnologias conseguem sequenciar e transmitir milhões de informações em muito menos tempo e com menor custo, fazendo isso através da utilização da clonagem *in vitro* e de sistemas de suporte sólido para as unidades de sequenciamento, não sendo mais necessários gastar muito tempo no laboratório para produzir clones bacterianos, montar as placas para sequenciamento ou separar os fragmentos do DNA em géis.

Esse sequenciamento é extremamente útil, tanto no campo da saúde como no agrônômico e em muitos outros.

Nas plantas, vem sendo utilizado para o sequenciamento dos genomas plastidiais, sequências expressas, clones de interesse, ressequenciamento e detecção de variantes genotípicas.

Pode ser usado também para caracterização de doenças (descobrir o agente causador, por exemplo) e, com estudos, trazer o tratamento para tal.

2. REVISÃO DE LITERATURA

2.1 SEQUENCIAMENTO MANUAL (DE SANGER)

- Esse método foi desenvolvido por Frederick Sanger em meados dos anos 70, trata-se do método mais difícil e caro de sequenciamento, sendo ele feito grande parte de forma manual, e com a adição de muitos reagentes que tem um alto valor comercial, tendo ele uma baixa eficiência perante os métodos modernos de sequenciamento
- Consiste em utilizar produtos químicos para quebrar a molécula de DNA; ddNTP (didesoxinucleotídeo trifosfato) para interromper a incorporação de outros nucleotídeos a partir dele, acabando com a síntese de DNA; primers para definir o DNA molde e produzir o filamento, e usar da radioatividade para marcar e visualizar os fragmentos, os quais são gerados pela eletroforese
- "Produz-se" longos reads, porém pouquíssimas bases
- Técnica muito demorada, difícil e tem o preço muito elevado por base sequenciada
- Sequenciamento teria de ser feito manualmente
- Quando o método utilizado é o manual, os dideoxinucleotídeos são marcados com radiação e não com substâncias fluorescentes, sendo então necessária uma radiografia para a visualização da seqüência de DNA. Já no seqüenciamento automático, o seqüenciador identifica os fragmentos por meio da incidência de laser sobre os dideoxinucleotídeos fluorescentes

2.2 Sequenciamento automático

- Essa forma de sequenciamento apresenta o mesmo princípio do sequenciamento manual

- Nesse caso, os ddNTPs possuem própria fluorescência (não radioatividade, como no manual)
- Após a realização de todos os processos de purificação e demarcação dos fragmentos, os produtos são submetidos ao sequenciador automático e a medida que são bombardeados por feixes de laser vão migrando pelos capilares do sequenciador e emitem luz, a qual é detectada por um fotomultiplicador. Em seguida a informação é passada para o computador e processada, dados são recuperados em formato de eletroferograma ou de texto, com a emissão de gráficos para melhor interpretação dos fragmentos.
- "Produz-se" reads médios e poucas bases

2.3 Sequenciamento de Nova Geração:

"Produz-se" reads curtos e muitas bases, em geral

- 2.3.1 **Minlon:** Trata-se de um mini sequenciador "móvel", desenvolvido pela Oxford Nanopore, sua forma de obtenção de energia é pelo cabo USB, sendo assim bem funcional em áreas isoladas do planeta, sendo possível a realização de sequenciamento em locais inóspitos, ele utiliza nanoporos para sequenciar o DNA; Ao contrário de outras máquinas de sequenciamento comercial, que podem ser do tamanho de um refrigerador e requerem enormes quantidades de produtos químicos caros, este analisa o DNA diretamente à medida que a molécula atravessa um pequeno poro suspenso em uma membrana. Alterações na corrente elétrica são usadas para ler a cadeia de letras A, G, C e T. O Minlon trata-se de uma tecnologia muito avançada, tendo ele somente 4 cm, é ideal para uso de campo, mas apresenta erros, tendo uma precisão de 60-80%, não sendo muito bem desenvolvido, faltando algumas correções que com o avanço da tecnologia serão corrigidos rapidamente, permitindo um uso perfeito do aparelho.
- 2.3.2 **Illumina:** Esse sequenciador foi desenvolvido em 2006 baseado no conceito de amplificação clonal e seqüenciamento por síntese (SBS) química para permitir a rápida e precisa (aproximadamente 99% de precisão) sequencia; o processo identifica simultaneamente bases de DNA, incorporando-as em uma cadeia de ácidos nucleicos. Cada base

emite um sinal fluorescente único, uma vez que é adicionada à cadeia em crescimento, o qual é utilizado para determinar a ordem da sequência de DNA, os quatro nucleotídeos tem as quatro bases que se completam umas com as outras para se ligar ao alvo, esta competição natural garante a alta precisão. Depois de cada síntese os fluorocromos são excitados por um laser, a cor obtida identifica a base que foi adicionada. Pode ser aplicada a pequenas regiões, orientada ou no genoma inteiro através de uma variedade de métodos, permitindo aos investigadores investigar e entender tudo sobre o DNA.

2.3.3 PacBio RS II: A Pacific Biosciences desenvolveu a tecnologia de sequenciamento de molécula única em tempo real para sequenciamento em larga escala de amplicons, target sequencing, análise de populações, transcriptoma, epigenoma e montagem de genomas complexos. O robusto equipamento RSII possui um sistema ótico de alta performance, uma workstation para manejo automatizado das SMRT® Cells e reagente, além de um software de controle intuitivo para monitoramento e análise do sequenciamento, para que seja feita a análise não é necessário com que seja realizado a ampliação do material, preservando assim informações essenciais da amostra. Ele apresenta capacidade de registro em vídeo que permite o sequenciamento de moléculas varias vezes, com uma duração de 30 minutos a 6 horas, ele ainda apresenta interface touchscreen, facilitando assim o manuseio do aparelho, a grande tecnologia empregada nesse aparelho permite com que ele faça o sequenciamento o mais rápido possível.

3. Objetivos:

As tecnologias de sequenciamento de nova geração começaram a ser comercializadas no ano de 2005 e desde então vem sendo aprimoradas muito rapidamente, evoluindo cada vez mais.

Essas tecnologias geram em uma única corrida o sequenciamento de DNA, trazendo informações sobre milhares de pares de bases.

As plataformas com maior utilização são a 454 FLX da Roche e a Solexa da Illumina. Há também dois sequenciamentos que estão sendo aderidos, a plataforma da Applied Biosystems, denominada SOLiD System, e o HeliscopeTrue Single Molecule Sequencing (tSMS), da Helicos. Essas novas plataformas podem gerar informações em menor tempo e com um menor custo por base para o sequenciamento, sendo mais produtivas que o sequenciamento de Sanger. Essa maior eficiência se deve ao uso da clonagem in vitro e de sistemas de suporte sólido para as unidades de sequenciamento, não necessitando mais da produção de clones bacterianos, da montagem de placas de sequenciamento e da separação dos fragmentos em géis. A clonagem in vitro em suporte sólido possibilita que possam ser produzidas milhares de leituras de uma só vez com a plataforma Solexa, SOLiD ou 454.

A primeira plataforma de sequenciamento de nova geração a ser comercializada foi o sistema 454, realizando o sequenciamento baseado em síntese, chamado pirosequenciamento (RONAGHI et al., 1998). A leitura da sequência de bases nessa plataforma é formada pela combinação de reações enzimáticas que se inicia com a liberação de um pirofosfato, provindo da adição de um desoxinucleotídeo na cadeia. O pirofosfato é convertido para ATP, sendo utilizado pela luciferase a fim de oxidar a luciferina. Essa reação é capaz de produzir um sinal de luz que é capturado por uma câmera CCD (charge-coupled device) acoplada ao sistema. O DNA é mecanicamente fragmentado em sequências de 300 a 800 pb, transformando-os em fragmentos fosforilados e ligados a adaptadores de sequência específica (Figura 1). Após isso, o DNA é ligado a adaptadores A e B nas extremidades 3' e 5', utilizados posteriormente no isolamento dos fragmentos (A-B) e amplificação nas reações de sequenciamento. Apenas os fragmentos AB são eluídos na reação de purificação e são especificamente ligados às microesferas que carregam várias cópias da sequência complementar exata ao adaptador B de um único fragmento (MARGULIES et al., 2005). O outro adaptador é utilizado no anelamento do primer que inicia a reação de sequenciamento. As microesferas ligadas aos fragmentos são emulsionadas em uma mistura de água e óleo com reagentes de PCR para a clonagem do fragmento em cerca de 1 milhão de cópias. Com a PCR em emulsão, o óleo em solução aquosa forma micelas, capturando as microesferas. Cada micela produzirá diversas cópias de um mesmo fragmento isoladamente em um

microsuporte. Cada micela produzirá cópias idênticas de um mesmo fragmento isoladamente (DRESSMAN ET AL., 2003). Após a PCR de emulsão, as microesferas juntamente aos fragmentos de DNA são depositadas em poços distintos em uma placa de sílica onde os reagentes para o sequenciamento são distribuídos. Em cada poço ocorre as reações de sequenciamento, para aquele fragmento, não havendo competição para reagentes. A placa de sequenciamento é inserida junto ao sistema óptico de leitura no equipamento. Os reagentes e as soluções de sequenciamento são então distribuídos por toda a placa a cada ciclo para obtenção do sequenciamento paralelo dos 1,6 milhões de poços. O sequenciamento é realizado em ciclos, e a cada ciclo um tipo determinado de nucleotídeo é adicionado à reação. Se o nucleotídeo adicionado for incorporado à sequência em síntese, é emitido um sinal de luz, sendo a intensidade desse sinal um reflexo do número de nucleotídeos desse tipo que foram sucessivamente incorporados na molécula. As leituras produzidas possuem geralmente cerca de 250pb, o que representa um comprimento de leitura muito menor que o produzido pelo sistema de Sanger (~700pb). A Roche divulgou recentemente o lançamento da série Titanium de pirosequenciamento, em que leituras maiores que 400pb são conseguidas. O custo do sequenciamento com essa plataforma é superior ao custo das plataformas Solexa e SOLiD, mas, nos casos em que a produção de leituras maiores é necessária, a plataforma 454 deve ser a melhor opção.

Plataforma Solexa

O sequenciamento na plataforma Solexa, assim como o sequenciamento de Sanger, é realizado por síntese usando o DNA polimerase e os nucleotídeos terminadores marcados com diferentes fluoróforos. A inovação dessa plataforma é a clonagem in vitro dos fragmentos em uma plataforma de vidro, processo conhecido como PCR de fase sólida (FEDURCO et al., 2006; TURCATTI et al., 2008). A superfície de clonagem é dividida em oito linhas, onde pode ser utilizado para o sequenciamento de oito bibliotecas. Em cada linha, adaptadores são fixados à superfície pela extremidade 5', deixando a extremidade 3' livre para servir na iniciação da reação de sequenciamento dos fragmentos imobilizados no suporte por hibridização. No

primeiro ciclo de amplificação, nucleotídeos não marcados são fornecidos para que haja a síntese da segunda fita do fragmento imobilizado no suporte. A alta densidade de adaptadores no suporte viabiliza a hibridização do adaptador livre dos fragmentos imobilizados a sua sequência complementar fixa perto do clone inicial durante o ciclo de anelamento. Após o anelamento, esse fragmento forma uma estrutura em “ponte” na superfície de sequenciamento, correndo a extensão, formando a fita complementar também em “ponte”. Ocorre um ciclo de desnaturação onde as fitas são separadas e linearizadas. Os ciclos são repetidos 35 vezes e assim cerca de mil cópias geradas de cada fragmento nessa PCR de fase sólida permanecem próximas umas das outras, formando um cluster de sequenciamento. . A alta densidade dos clusters de sequenciamento permite que o sinal de fluorescência formado com a incorporação de cada um dos nucleotídeos terminadores tenha a intensidade necessária para garantir sua detecção exata. Até 50 milhões de clusters podem ser produzidos por linha. . Após a incorporação de cada nucleotídeo no fragmento em síntese, a leitura do sinal de fluorescência é realizada. Logo após, ocorre uma etapa de lavagem para retirada dos reagentes excedentes e remoção do terminal 3' bloqueado e do fluoróforo do nucleotídeo incorporado no ciclo anterior para que a reação de sequenciamento prossiga. A leitura das bases é feita pela análise sequencial das imagens capturadas em cada ciclo de sequenciamento. Em geral, leituras de 25-35 bases são obtidas de cada cluster (SHENDURE & JI, 2008).

4. METODOLOGIA

Foi realizado o experimento sequenciando 94% do genoma de *Eucalyptus Grandis*, 640 megabases em 36.376 genes.

Foram utilizadas três bibliotecas de tamanhos diferentes para o processo de sequenciamento de subclones plasmídicos, sequenciando em dois pares. As leituras de sequência foram montadas utilizando uma versão modificada de Arachne v.20071016. Os andaimes foram rastreados contra proteínas bacterianas, sequências de organelas, e foram removidos se contaminados. Para a construção de pseudomólulos cromossômicos, os marcadores do mapa genético foram colocados usando dois métodos baseado em SSR; resumidamente, as ESTs <300

pb foram removidas, juntamente com cloroplasto, mitocondrial ou rDNA ESTs. Todas as ESTs duplicadas foram colocadas com o genoma usando BLAT em comparação.

5. Referencia Bibliografica

RONAGHI, M. et al. A sequencing method based on real-time pyrophosphate. *Science*, v.281, p.363–365, 1998. Disponível em: . Acesso em: 5 jun. 2009. doi: 10.1126/ science.281.5375.363.

MARGULIES, M. et al. Genome sequencing in open microfabricated high density picoliter reactors. *Nature*, v.437, n. 7057, p. 376-380, 2005. Disponível em: . Acesso em: 5 jun. 2009. doi: 10.1038/nature03959.

TURCATTI, G. et al. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic acids research*, v.36, e25, 2008. Disponível em: . Acesso em: 5 jun. 2009. doi: 10.1093/nar/gkn021.

SHENDURE, J.; JI, H. Next-generation DNA sequencing. *Nature biotechnology*, v.26, n.10, p.1135-1145, 2008. Disponível em: . Acesso em: 5 jun. 2009. doi:10.1038/nbt1486.