

Sequenciamento de nova geração e entomologia: Novas perspectivas para antigos questionamentos

Modern sequencing technologies and entomology: New perspectives for old questions

Patrícia Regina Ströher¹

¹ Programa de Pós-Graduação em Entomologia, Universidade Federal do Paraná, Curitiba, Paraná, Brasil.

*Contato: patricia.stroher@gmail.com

Resumo. Este artigo é uma revisão das novas tecnologias disponíveis para sequenciamento de DNA e sua aplicabilidade nos estudos entomológicos. O texto traz uma perspectiva histórica do assunto, as principais plataformas disponíveis na atualidade e também informações de como estes instrumentos funcionam, discutindo sua utilidade para resolver questões como biogeografia, taxonomia e evolução. No passado, importantes questões evolutivas ficavam com respostas limitadas em várias áreas (incluindo a entomologia), quando o cenário envolvia organismos não-modelo. Os métodos anteriores muitas vezes não possuíam protocolos aplicáveis às peculiaridades de cada indivíduo ou espécie que permitissem recuperar a informação contida em seus genomas a fim de resolver indagações filogenéticas e biogeográficas. Outro ponto interessante do sequenciamento de nova geração e insetos diz respeito à possibilidade de recuperação de DNA antigo, como o contido em exemplares de coleções científicas e museus. Estas novas tecnologias tornam possível o acesso a um novo universo de informações que até então estava inacessível aos pesquisadores.

Palavras-chave. *Interdisciplinaridade; organismos não-modelo; DNA fóssil.*

Abstract. *This manuscript is a revision of the available new technologies to perform DNA sequencing and its application in entomology studies. A historical view is presented as the currently accessible tools to study biogeography, taxonomy and evolution. In the past, important evolutive questions were limited when non-model organisms were involved. Previous methodologies commonly did not have protocols adapted to single individuals or species in order to be able to recover the information behind their genomes to resolver phylogenic and biogeographic topics. Other point related with these new sequencing technologies and insects is the possibility of ancient DNA recovery, as the information in scientific collections and museums. These technologies made possible to address a new informational universe that was previously inaccessible.*

Keywords. *Interdisciplinary approach; non-model organisms; Fossil DNA .*

Recebido: 09dez16
Publicado: 27ago18

Editado por Laila
Asth e revisado por
Anônimo.

Introdução

Um campo de estudo ainda jovem e em plena revolução: é assim que podemos falar da Biologia Molecular. Quando analisamos esta área aplicada aos estudos entomológicos (particularmente com a finalidade de resolver questões de sistemática, taxonomia e biogeografia), percebemos que este campo é ainda mais recente, mas que também está em franco crescimento. E justamente neste momento em que a maioria dos pesquisadores começa a quebrar a barreira entre as duas áreas (entomologia e biologia molecular), seguindo uma tendência científica mundial, a interdisciplinaridade, surgem novas tecnologias que os fazem reaprender o conteúdo há tão pouco tempo assimilado (a autora deste texto inclusa). Faz parte da ciência e de ser um pesquisador, aliás, são as constantes novidades e curiosidade com o desconhecido que atraíram muitos destes pesquisa-

dores para a carreira científica, a necessidade de contínua adaptação não é novidade. Os que mais recentemente aliam dados moleculares nas suas pesquisas provavelmente já foram inseridos nas técnicas mais atuais e, portanto, não estão passando por esta fase de transição necessária para quem já trabalhava com tais dados. O “Next-Generation Sequencing” ou “Massively Parallel Sequencing” conhecidos pelas siglas NGS e MPS, respectivamente consistem no sequenciamento de DNA através de milhares de leituras em paralelo gerando milhões de sequências em uma única corrida de um equipamento automatizado. Assim, a intenção deste artigo é fazer uma revisão do sequenciamento de nova geração e sua aplicação, particularmente nos estudos com insetos a fim de colaborar com entomólogos que queiram incorporar esta nova tecnologia em seus projetos.

Sequenciamento de DNA

Atualmente o sequenciamento de DNA (principalmente em humanos e em outros organismos modelo), permite-nos vislumbrar possibilidades como a medicina personalizada, sequenciamento completo de genomas e reconstrução de genomas fósseis como plenamente plausíveis, mesmo que estes termos em um passado recente fossem considerados enredo de ficção científica.

O sequenciamento de DNA é resultado de esforços de biologia molecular que começaram a se tornar proeminentes na década de 1950. No entanto, as pesquisas com o intuito direto de sequenciamento do material genético só começaram no início da década de 1970 comandadas por Frederick Sanger. A equipe do Professor Sanger em 1977 publicou dois importantes trabalhos (Sanger et al., 1977a; Sanger et al., 1977b) descrevendo todos os procedimentos necessários para se chegar à leitura dos hoje populares quatro nucleotídeos A, C, G e T que compõem o material genético de todos os organismos, bem como, publicaram o primeiro genoma a ser desvendado, o fago phi X17. O princípio desta técnica perdurou por mais de três décadas como a única forma pela qual cientistas sequenciavam DNA em seus laboratórios, mesmo quando o processo manual passou a ser automatizado no fim da década de 1980 e início da década de 1990, o método era essencialmente o mesmo. Trata-se de um sequenciamento enzimático baseado na interrupção da cadeia pela adição de um dideoxinucleotídeo, que nada mais é do que um nucleotídeo com uma modificação química: não possui o grupo hidroxila (OH) 3', não permitindo a ligação do próximo nucleotídeo. Estes nucleotídeos modificados são marcados com P32 que podem ser visualizados através de raios-X. Assim, após a geração dos fragmentos estes eram submetidos a um gel de eletroforese e observados em filmes de raios-X onde tinham suas sequências determinadas (Sanger et al., 1977a). Era um processo manual, com múltiplas etapas, com as leituras realizadas a olho nu e que deixava os cientistas constantemente expostos aos raios-X (Mardis, 2013), mas de tamanha importância para ciência que rendeu o segundo prêmio Nobel em química para Sanger no ano de 1980. O método Sanger começou a ganhar maior escala e popularização nos laboratórios quando em 1986 a empresa Applied Biosystems, Inc. (ABI) iniciou a comercialização de um sequenciador de DNA baseado em fluorescência que foi desenvolvido anteriormente por Smith et al. (1986). Neste aperfeiçoamento do método, a marcação com P32 foi eliminada, gerando uma grande economia de tempo extinguindo várias etapas manuais e com isso aumentando a acurácia ao eliminar uma das fontes de erro. Outro avanço da mesma época que não pode deixar de ser citado foi o desenvolvimento da Reação em cadeia da Polimerase (PCR) (Mullis et al., 1986). Esta técnica permite a amplificação de sequências de DNA em milhões de cópias através de reações em ciclos catalisadas por uma enzima polimerase termoestável, assim a quantidade inicial de DNA usada pôde ser bem menor (Mardis, 2013).

Após aproximadamente uma década sem grandes revoluções no sequenciamento de DNA, o lançamento de sequenciadores de capilar teve um grande impacto na área. Estes instrumentos resolveram o problema com relação ao

gel eletroforético ao injetar as amostras diretamente em uma matriz de capilares, proporcionando uma resolução individual dos nucleotídeos. O equipamento de maior sucesso foi o ABI PRISM® 3700, que começou a ser comercializado em 1999. Este modelo, logo seguido de sua versão posterior ABI 3730, foram os principais instrumentos utilizados no sequenciamento do genoma humano e do rato (Mardis, 2013). O projeto genoma humano foi oficialmente declarado finalizado em 2003, dois anos antes do previsto (Chial, 2008), mas teve seu início ainda no ano de 1990 (Watson, 1990), portanto, levou mais de uma década para que as três bilhões de bases fossem decifradas com um orçamento inicial de também três bilhões de dólares (Watson, 2001). Com este grande salto em conhecimento e entusiasmo dos pesquisadores ao redor do mundo em utilizar dados genômicos, logo empresas de biotecnologia perceberam que uma nova forma de sequenciamento mais rápida e menos laboriosa seria necessária para tornar a genômica trivial a todos os organismos. Era o início de uma nova era no sequenciamento de DNA, onde novas tecnologias passaram a ser desenvolvidas não mais por biólogos, mas também por engenheiros, não mais em universidades, mas em empresas. Isto individualmente, resultou em uma revolução que produziu equipamentos capazes de realizarem o mesmo trabalho de trinta mil sequenciadores dos modelos usados no início dos anos 2000, juntos (Schnable, 2010).

Sequenciamento de nova geração

Em meados de 2005, após quase trinta anos de sequenciamento pelo mesmo método, surgiram os primeiros equipamentos com uma abordagem alternativa à Sanger (Margulies et al., 2009). Uma tecnologia onde o DNA a ser sequenciado é usado para construir uma biblioteca de fragmentos que possuem DNAs sintéticos (adaptadores) ligados covalentemente a eles. Estes adaptadores são sequências universais, específicos para cada plataforma, usados pela polimerase para amplificar os fragmentos das bibliotecas. Este processo ocorre geralmente em uma superfície sólida, seja em vidro plano ou em volta de "beads" (microesferas magnéticas). Em essência, os instrumentos de sequenciamento de nova geração conduzem a amplificação dos nucleotídeos e a detecção (o sequenciamento em si) em uma única etapa, enquanto o método Sanger realiza os dois processos separados (Mardis, 2013). Além do baixo rendimento, o sequenciamento via método Sanger, para obter bons resultados, requer que o DNA amplificado esteja em alta concentração (Polz and Cavanaugh, 1998) e fornece um único padrão de sinal, ou eletroferograma, para cada sequência gerada. Ambas as características diferem da tecnologia de nova geração que pode ser conduzida com baixas concentrações de DNA e produz milhões de sinais em paralelo para a mesma sequência (Shokralla et al., 2014).

Outro contraste entre os dois métodos que vale ser mencionado é o tamanho da sequência produzida. Enquanto o método Sanger (após décadas de refinamento) produz fragmentos com ≈1000 pares de base (pb) (Shendure and Hanlee, 2008), os primeiros sequenciadores de nova geração produziam fragmentos com apenas 150-200 pb (Schuster, 2007). Atualmente, o sequenciamento de nova geração já

capaz de produzir fragmentos com mais de 500 pb (Glenn, 2016). A taxa de erro no sequenciamento de nova geração que no início era mais elevada do que a do método Sanger, foi sendo diminuída juntamente com o aprimoramento das plataformas e atualmente não pode mais ser apontada como uma crítica às novas tecnologias (Shokralla et al., 2014). Mas, uma das principais vantagens do sequenciamento de nova geração ainda não foi mencionada: o custo. Após o investimento inicial com um novo equipamento o custo por megabase é inferior a \$ 0.10 com a possibilidade de sequenciamentos completos de genomas em um período inferior a uma semana (Glenn, 2016). Assim, os custos e prazos são apenas uma fração dos necessários para o sequenciamento via método Sanger.

O pequeno tamanho das sequências produzidas inicialmente aliada à necessidade de compra de novos equipamentos podem ser explicações para a maioria dos cientistas, a princípio, não adotarem o novo método. Mesmo as versões iniciais da plataforma 454 (considerado o primeiro sequenciador de nova geração) que facilmente geravam o equivalente a mesma informação de 50 sequenciadores modelo 3730XL da Applied Biosystems (método Sanger), com apenas um sexto do custo, tiveram reações surpreendentemente reservadas (Schuster, 2007). Na verdade, houve certa resistência da comunidade científica em alterar a forma que há tanto tempo vinha sendo utilizada no sequenciamento de material genético. Este fato pode ser observado pelo reduzido número de artigos que utilizaram a nova tecnologia nos primeiros anos após o lançamento das primeiras plataformas. Ao invés de abraçar a nova tecnologia e rapidamente adaptar-se para aproveitar seu enorme potencial, muitos cientistas acostumados ao método Sanger, levantaram questionamentos com relação à acurácia, tamanho dos fragmentos gerados, custo de infraestrutura ou apenas a objeção de lidar com o grande volume de dados gerados pela nova tecnologia. Este ceticismo, que se iniciou nas agências de financiamento, pode ter sido impulsionado pelo medo de que investimentos substanciais no sequenciamento Sanger poderiam se tornar obsoletos (Schuster, 2007).

Somente agora, mais de uma década após o lançamento do sequenciamento de nova geração, é que estas tecnologias ganham maior popularidade no meio científico. O estranhamento inicial e as preocupações com relação à acurácia e eficiência já foram superados. Cada dia mais cientistas percebem que esta é uma tecnologia que está transformando “o que nós pesquisamos, como nós pesquisamos e quanto irá custar” (Glenn, 2011). Conseqüentemente, muitos centros de genômica têm reduzido gradualmente os seus sequenciadores Sanger e passaram a incorporar a tecnologia NGS (Shokralla et al., 2014). O desafio agora é treinar e formar profissionais para que o sequenciamento de nova geração que já é viável se torne realidade como ferramenta de estudos para outras áreas além da genética e biologia molecular.

Plataformas

Nesta revisão será usada a mesma nomenclatura proposta por Glenn 2011, em que são chamadas plataformas de 2a geração as que necessitam da amplificação prévia de DNA

para o sequenciamento, 3a geração para as plataformas que sequenciam diretamente e individualmente as moléculas de DNA e simplesmente plataformas de sequenciamento de nova geração, indicadas com a sigla “NGS” para de forma genérica fazer referência aos instrumentos de 2a e 3a geração. As plataformas aqui discutidas serão as que estão disponíveis atualmente e/ou as mais largamente utilizadas, tendo em vista que algumas plataformas que surgiram já estão praticamente extintas ou simplesmente não possuem instrumentos disponíveis para venda, como o sequenciador Helicos (Nikolaki and Tsiamis, 2013). A maioria das plataformas necessita que o DNA seja fragmentado em sequências curtas, entre 200-100 pb, e que a solução contenha primers ou adaptadores, ou seja, a construção de uma biblioteca se faz necessária (Glenn, 2011). Abaixo segue uma breve descrição destes equipamentos:

- 454- A primeira plataforma NGS comercialmente disponível foi lançada no ano de 2005, atualmente faz parte dos equipamentos comercializados pela empresa Roche e seu princípio usa o pirosequenciamento. Esta técnica utiliza beads e inicia o processo com uma única molécula de DNA que é amplificada em uma PCR de emulsão, após, as milhões de beads são transferidas para uma placa. Nesta placa cada bead só encaixa em apenas uma nanofenda onde serão sequenciadas em paralelo. O processo utiliza uma enzima luciferase que produz luz detectável a cada vez que um nucleotídeo é incorporado e assim permite o sequenciamento (Margulies et al., 2005). Em outubro de 2013 foi realizado um anúncio pela empresa Roche comunicando que até meados de 2016 a empresa iria gradualmente descontinuar a venda de equipamentos baseados em pirosequenciamento. Atualmente o sequenciador continua anunciado no website da empresa, mas começam a surgir os primeiros relatos de pesquisadores comunicando que não há mais a possibilidade de adquirir os kits de reagentes para o correto funcionamento do sequenciador.

- Illumina- Inicialmente lançada com o nome de Solexa foi subsequentemente adquirida pela companhia Illumina de onde vem o nome atual. O princípio deste método usa bridge PCR, onde pontes de DNA são criadas e amplificadas em uma superfície sólida, criando agrupamentos chamados de clusters. Os milhões de clusters produzidos possuem, cada um, mais de mil cópias do DNA inicial que serão sequenciados com uma estratégia similar ao método Sanger. A diferença é que apenas terminadores marcados por fluorescência são utilizados e há a garantia de que a cada etapa somente um nucleotídeo é incorporado. A cada incorporação o nucleotídeo é detectado por um dispositivo de leitura que interpreta qual a identidade do mesmo, o sequenciamento ocorre em paralelo em todos os clusters ao mesmo tempo (Morozova and Marra, 2008). No início de 2014 a companhia anunciou um novo sistema, o HiSeq X™ Ten Sequencing System. Este produto quebrou a “barreira do som” da genômica ao se tornar capaz de sequenciar um genoma humano completo por apenas mil dólares gerando seis bilhões de bases por dia, podendo sequenciar mais de 18 mil genomas por ano. Atualmente o portfólio da plataforma Illumina é composto por seis equipamentos que se diferenciam entre si na escala e tamanho das sequências geradas: Illumina MiniSeq,

Illumina MiSeq, Illumina NextSeq 500/550, Illumina HiSeq 1500/2500, Illumina HiSeq 3000/4000 e Illumina HiSeq X Ten.

- SOLiD- No presente esta plataforma é administrada pela Thermo Fisher Scientific, empresa resultante primeiramente pela fusão das companhias Applied Biosystems e Invitrogen, formando a Life Technologies Corporation em 2009. Em 2014 a empresa foi adquirida pela Thermo Fisher Scientific, por este motivo ainda é comum encontrarmos nomes de diferentes empresas associadas a esta plataforma. A tecnologia SOLiD também faz uso das beads, mas aqui elas são inseridas diretamente em chip onde ocorre a leitura das seqüências. Estes chips podem ser divididos em oito áreas, permitindo a análise de oito bibliotecas simultaneamente. Outra diferença é que as bases são determinadas por meio de ligação, ou seja, a enzima utilizada é uma DNA ligase ao invés da conhecida DNA polimerase (Glenn, 2011). Atualmente o alto custo para aquisição e a limitação de aplicações tornam esta plataforma praticamente inviável, sendo considerada um “potencial zumbi” (Glenn, 2016).

- Ion Torrent- É uma plataforma também administrada pela empresa Thermo Fisher Scientific. Sua estratégia de sequenciamento é similar a aplicada pela 454, com o contraste de que esta tecnologia detecta íons de hidrogênio (H⁺) ao invés de sinal luminoso. A utilização de H⁺ implica na não necessidade de uso de lasers, câmeras ou marcadores fluorescentes. O lançamento dos primeiros equipamentos ocorreu no final no ano de 2010. No segundo semestre de 2012 foi lançado o sucessor deste equipamento, o Ion Proton que utiliza chips maiores sendo capaz de gerar 10 gigas de informação por corrida (Nikolaki and Tsiamis, 2013). Em setembro de 2015 foi lançado o Ion Torrent – S5 com a possibilidade de uso de três novos chips, permitindo ao pesquisador escolher qual o mais conveniente com a sua necessidade de escala de sequenciamento. Infelizmente os novos chips não são operáveis entre os equipamentos da plataforma “Ion”. Esta tecnologia, em especial o S5, pode ser vista como uma alternativa aos equipamentos Illumina MiniSeq, Illumina MiSeq e Illumina NextSeq 500/550, no entanto, sua baixa representatividade nos laboratórios e a baixa disponibilidade de softwares ainda não o tornaram competitivo (Glenn, 2016).

- PacBio- Foi a plataforma de 3a geração que alcançou maior notoriedade, a empresa Pacific Biosciences desenvolveu um equipamento capaz de sequenciar moléculas únicas de DNA em tempo real. Isto é possível pela observação imediata da enzima DNA polimerase onde as bases são detectadas no exato momento em que ocorre a síntese, isto é praticável por uma marcação fluorescente (Glenn, 2011). Uma desvantagem dos sequenciadores de 3a geração é a taxa de erro, superior aos índices das outras plataformas (Quail et al., 2012). No início de 2016 a empresa lançou um novo equipamento chamado Sequel System, um equipamento menor, mais leve que custa metade do valor e com capacidade de sequenciamento sete vezes maior do que o seu antecessor. No entanto, a tecnologia continua muito mais cara do que o Illumina e com maior taxa de erro, mas a promessa de aperfeiçoamento na plataforma a torna muito promissora (Glenn, 2016).

- MinION- Plataforma apresentada pela empresa britânica Oxford Nanopore em 2012, mas sua tecnologia vem sendo desenvolvida desde a segunda metade da década de 1990 (Kasianowicz et al., 1996). Aqui o sequenciamento se faz por meio de sinais elétricos detectados no momento em que os nucleotídeos passam através de nanoporos (Ninomiya et al., 2013). O equipamento todo possui o tamanho similar a um HD externo onde as amostras são inseridas e o sequenciador é acoplado diretamente aos computadores dos pesquisadores. Causou grande entusiasmo na comunidade científica, ao permitir, por exemplo, imaginar o sequenciamento de DNA tão acessível que este poderia ser realizado em campo. Infelizmente, após vários adiamentos, ainda não apresenta data para ser lançado no mercado. Em abril de 2014, alguns cientistas ao redor do globo foram selecionados para testar o equipamento em seus laboratórios e imediatamente surgiram fotos nas mídias sociais da plataforma em uso no dia-a-dia (Nextgenseek, 2014). Após dois anos de uso a plataforma apresenta grandes melhorias, embora sua alta taxa de erro ainda seja considerada como um fator limitante e que possivelmente está impedito sua ampla comercialização (Glenn, 2016). É um produto que gera muita expectativa, mas após quatro anos de espera continua sendo considerado apenas como “uma promessa”.

- GenapSys- Faz parte do portfólio da empresa Sigma-Aldrich, ainda não foi lançado oficialmente e não possui detalhes da sua tecnologia divulgados, mas promete ser uma nova alternativa de sequenciador portátil e atualmente está recrutando pesquisadores para testar a tecnologia (Glenn, 2016).

- Genia- Finalizando os sequenciadores de moléculas únicas, este equipamento é considerado uma combinação das tecnologias da Oxford Nanopore, Illumina e Pacbio (Robison, 2016). Detalhes do funcionamento da tecnologia foram surpreendentemente publicados em acesso aberto recentemente (Fuller et al., 2016). A empresa foi adquirida pela Roche em 2014 e afirma que em breve será possível sequenciar um genoma humano por apenas 100 dólares (GENIA TECHNOLOGIES, 2016).

As plataformas 454, Illumina e SOLiD são as mais populares em laboratórios acadêmicos (Glenn, 2016), estes três equipamentos possuem estratégias distintas quanto ao tamanho de fragmentos produzidos. Enquanto o 454 produz um número de leituras menor em cada corrida com fragmentos maiores, o Illumina e SOLiD produzem um número maior de leituras e fragmentos com tamanhos menores. Aqui cabe uma consideração com relação ao número de pares de base necessários para se considerar uma leitura curta ou longa. Segundo Glenn (2011), são considerados fragmentos curtos aqueles ≤ 50 bases consecutivas, fragmentos médios aqueles que possuem ≥ 51 , mas < 400 bases consecutivas, fragmentos longos aqueles com ≥ 400 , mas < 1000 bases consecutivas e, finalmente fragmentos estendidos aqueles > 1000 bases.

Leituras longas são mais indicadas em casos onde o objetivo é a caracterização de todo o genoma ou transcriptoma, uma vez que fragmentos longos são mais fáceis de serem montados do que fragmentos curtos. No entanto, o baixo custo e o incremento no número de leituras associado os sequenciadores de fragmentos curtos os tornam mais ade-

quados em casos de resequenciamento (quando um genoma completo já está disponível) ou em casos em que a frequência de leituras é importante como estudos de expressão gênica (Glenn, 2011). Mas, é importante lembrar que praticamente todas as plataformas estão aumentando progressivamente o tamanho dos fragmentos gerados, assim muitas das plataformas podem ser usadas para as mesmas aplicações. Então, entre tantas possibilidades, qual escolher? Levando em consideração o tempo, custos e tamanhos de fragmentos, um veredito atual sobre qual é a plataforma mais apropriada para estudos em organismos não modelo, o título fica com a plataforma Illumina. Enquanto a tecnologia da Oxford Nanopore e similares não chegam a todos e ao mercado, Illumina aparece como a plataforma de escolha na maioria dos laboratórios e mesmo quando as mais recentes tecnologias se consolidarem e se tornarem confiáveis, o equipamento Illumina provavelmente continuará dominante, competitivo e estes deverão coexistir por um período (Yoder, 2014).

Se o sequenciamento de nova geração é uma realidade que vem aumentando o tamanho dos fragmentos e baixando os custos criando a sensação de que só existem benefícios, precisamos mencionar que as novas tecnologias também possuem suas desvantagens. O Calcanhar de Aquiles do NGS é sem dúvida o recurso computacional necessário para as análises (El-metwally et al., 2013). A montagem dos genomas surge como um problema, pois é impossível sequenciar diretamente um genoma inteiro em apenas uma leitura usando as tecnologias atuais. Os métodos de sequenciamento quebram as leituras randomicamente e as leem de forma individual, o processo de reconstrução total do genoma requer grande poder computacional para fazer a junção correta dos fragmentos até o nível cromossômico.

Devido a estas limitações em recursos computacionais, a construção dos genomas é a tarefa mais difícil enfrentada pelas novas tecnologias de sequenciamento. Alguns equipamentos geraram um volume de dados tão grande que foram considerados impossíveis de serem totalmente examinados (Egan et al., 2012). A necessidade de computadores tão potentes deve-se também aos milhões de fragmentos curtos produzidos, tendo em vista que estas sequências requerem análises mais intensas para serem agrupadas.

Não bastasse o desafio computacional dos extensos dados gerados, outra dificuldade é o fato destes não serem comunicáveis entre si. Cada plataforma gera seus arquivos em formatos únicos e possuem seus próprios protocolos para montagem e análise dos dados (Metzker, 2009). E quando o sequenciamento é feito de forma terceirizada, muitas vezes, é mais rápido enviar as sequências produzidas, gravadas em dispositivos físicos, via serviço postal tradicional ao invés de fazer o download das sequências via web (Jones et al., 2012). Em língua inglesa o termo Sneakernet é informalmente usado para definir esta transferência de arquivos usando um componente físico, estratégia que também é usada no caso da entrega dos resultados do NGS.

Ferramentas e conhecimento em bioinformática são essenciais para quem deseja lidar com os dados de forma crua, ou seja, assim que eles saem das plataformas de NGS. As ferramentas atuais ainda não possuem interfaces interativas aos usuários, de fácil configuração e independência

do sistema operacional, tornando-se um desafio para os pesquisadores sem profundos conhecimentos em informática (El-metwally et al., 2013). Embora promessas de novos programas que irão permitir avanços técnicos para que as tecnologias de NGS continuem evoluindo e ao mesmo tempo se tornem mais acessíveis e populares tenham sido feitas (Egan et al., 2012), atualmente a grande maioria das análises continuam sendo executadas através de “linha de comando”.

Como é possível usar a maioria das plataformas para a maior parte das aplicações, a economia de dinheiro, o tempo para aquisição dos dados e o tempo necessário para analisar a quantidade de informação gerada são itens importantes para a escolha da plataforma (Glenn, 2011). Com o número de equipamentos disponíveis aumentando e os custos diminuindo, dependerá apenas do nosso conhecimento em relação aos sistemas e da nossa criatividade desenvolver e adaptar estas tecnologias para obtermos dados de nosso interesse de forma eficiente. Ao aplicar estas novas tecnologias aos nossos organismos de estudo, que até pouco tempo estavam fora do alcance deste campo, expandimos também toda uma área de conhecimento. Esta afirmação tem base na ideia de que plataformas de NGS estão ajudando a abrir áreas de investigação biológica inteiramente novas, incluindo a investigação de genomas antigos, a caracterização de diversidade ecológica e a identificação de agentes etiológicos desconhecidos (Mardis, 2008).

Sequenciamento de nova geração e organismos não modelo: Entomologia

As plataformas de nova geração tornaram o sequenciamento mais acessível aos organismos não modelo, como boa parte dos insetos. É uma era em que praticamente qualquer organismo pode ser estudado de forma genômica (Ekblom and Galindo, 2010) e que faz surgir inclusive novas áreas de estudo, como a “ecologia genômica” (Gilad et al., 2009). Com a mudança do reino da genômica de endereço, saindo de estudos baseados em grandes laboratórios com espécies modelo para estudos de populações naturais, as pesquisas podem agora abordar importantes questões ecológicas e evolutivas em uma escala de precisão que soaria irrealista há apenas alguns anos atrás.

O método anterior requer o desenvolvimento de protocolos individualizados para a maioria das espécies, com primers específicos para a PCR (necessária na amplificação preliminar do DNA), além de ajustes finos em tempos e temperaturas. Pela imensa diversidade encontrada em insetos, na grande maioria das vezes não era possível aplicar o mesmo protocolo em diferentes organismos, em alguns casos até mesmo entre espécies muito próximas filogeneticamente. Isto limitava o número de marcadores moleculares que poderiam ser usados em insetos, ficando possível em boa parte dos estudos o uso apenas do DNA mitocondrial, pois este possui um genoma infinitamente menor e mais conhecido do que o nuclear. Isto facilitou o desenvolvimento de alguns primers genéricos que permitem seu uso em diferentes espécies, além das mitocôndrias possuírem múltiplas cópias em cada célula, aumentando a taxa de sucesso na amplificação final. No entanto, o material genético mitocondrial muitas vezes

não contém informação necessária para responder muitas perguntas filogenéticas. Assim, parece até “injusto” que após o sequenciamento de nova geração superar a inércia de praticamente três décadas na área (Schuster, 2007), e que vivendo em plena era genômica alguns pesquisadores continuem a usar a mesma tecnologia que traz resultados semelhantes há tantos anos.

É preciso saber tirar proveito desta nova era de informação a fim de responder os principais questionamentos que envolvem estudos em organismos não modelo. Inicialmente a aplicação mais comum para o NGS nestes casos foi a caracterização de transcriptomas (Ekblom and Galindo, 2010), ou seja, descrever como os genes são expressados em determinados tecidos, estágios de vida ou sob pressão de condições específicas. Um exemplo deste tipo de abordagem é o trabalho de Feldmeyer et al., (2013) utilizando insetos sociais. Este estudo com a espécie *Temnothorax longispinosus* utilizou sequenciamento de nova geração para tentar responder um questionamento que remota a época de Darwin: explicar a evolução das castas sociais, em especial das operárias (Darwin, 1859). A tecnologia NGS permitiu a comparação dos genomas de uma rainha e de três operárias com funções distintas na colônia, com os resultados apontando para cerca de 2500 genes que são diferencialmente expressos entre os indivíduos estudados. A técnica empregada neste caso chama-se RNA-Seq e consiste na detecção e quantificação de RNA em uma amostra em um determinado momento usando uma plataforma NGS (Wang et al., 2009). Mas, um dado não relacionado com o transcriptoma (expressão dos genes) e sim com a bioinformática chamou a atenção neste estudo: menos de 50% das sequências produzidas foram anotadas, um resultado que não era esperado uma vez que já existem diversos genomas de formigas sequenciados e anotados.

Aqui chegamos a um ponto que pode ser frágil quanto ao uso de NGS e sequenciamento completo de genomas em organismos não modelo. Como apenas fragmentos curtos são gerados, para que o genoma possa ser montado e anotado é preciso que já exista um genoma previamente sequenciado de outro organismo filogeneticamente próximo. Este genoma parente serve como referência para que as sequências curtas produzidas pelos equipamentos NGS sejam encaixadas, como em um quebra-cabeça. Assim, o sequenciamento completo de genomas, bem como o desenvolvimento de marcadores é facilitado para os organismos que já possuem parentes filogenéticos com genomas sequenciados. Em insetos, este é o caso principalmente de organismos pertencentes à Diptera (Celniker et al., 2002), Coleoptera (Richards et al., 2008), Lepidoptera (The International Silkworm Genome consortium, 2008) e Hymenoptera (Gadua et al., 2012). Isto não significa que organismos sem espécies proximamente relacionadas com genomas sequenciados não possam se beneficiar da tecnologia NGS, a estes ainda estão disponíveis outras abordagens capazes de produzir milhares de sequências, com baixo custo em um curto espaço de tempo como o RADSeq (Davey and Blaxter, 2011), cujo detalhes não são o foco desta revisão.

É importante ressaltar que já há alguns anos estas modernas tecnologias de sequenciamento estão sendo consideradas como o novo padrão para alguns pesquisadores na área filogenética (McCormack and Faircloth, 2012), inclusive sendo aplicadas para elucidar questões de filogeografia (O’neil et al., 2013). O uso de NGS permite uma coleta de dados em escalas maiores viabilizando estudos mais robustos sobre a história demográfica de populações e suas variações de aptidão associadas aos ambientes (Steiner et al., 2013).

Outra aplicação do NGS aos estudos entomológicos em especial é a implementação do “DNA barcoding” ou código de barras. Nesta técnica uma pequena parte variável do genoma (usualmente mitocondrial) é sequenciada a partir de amostras não especificadas. A informação contida na sequência é então utilizada para identificar as espécies presentes na amostra (Valentini et al., 2009). O DNA código de barras pode ser usado para detectar espécies crípticas (Janzen et al., 2012; Chacon et al., 2013), para investigar interações ecológicas complexas (Smith et al., 2011) e até para determinar o conteúdo de produtos comerciais (Wallace et al., 2012). Esta abordagem também está sendo empregada com insetos em pesquisas de monitoramento ambiental, como em BRODIN et al., (2012) que mostrou um incremento de 10% para mais de 90% no sucesso em identificação de espécies de Chironomidae (Diptera) em amostras bentônicas do Mar Báltico através do uso de NGS.

Uma vez que as tecnologias de NGS estão se tornando acessíveis para vários cientistas, tornam-se possíveis também trabalhos que possibilitem inferências puramente ecológicas e/ou evolutivas através destes tipos de dados (Fukatsu, 2012). Como referência a estas aplicações pode-se citar a pesquisa com uma espécie de Pyrrhocoridae (Hemiptera) e a caracterização de sua microbiota intestinal. A conclusão principal foi que a maioria das bactérias está localizada na região anóxica do sistema digestivo, levando a considerações importantes com relação ao mutualismo e à coevolução destes organismos (Sudakaran et al., 2012). O estudo conseguiu avaliar diferentes regiões do trato digestivo dos insetos, em diferentes fases de desenvolvimento, de diferentes regiões geográficas e que se alimentaram de distintas dietas. Esta enorme quantidade de sequências que revelou em detalhes a microbiota destes insetos seria praticamente impossível de ser produzida utilizando apenas técnicas tradicionais.

É importante também ressaltar a importância da tecnologia em estudos médico-veterinários relacionados aos insetos, como sequenciamento do genoma da mosca tsé-tsé (Attardo et al., 2014). A *Glossina morsitans* é o vetor para a famosa doença do sono e que constitui um flagelo para as populações da África Subsaariana, o tratamento é longo e complicado e sem ele a doença é geralmente fatal (Welburn et al., 2009). A decodificação deste material genético permitirá o reconhecimento de genes responsáveis pelo desenvolvimento inseto, contribuindo para elaborar novas estratégias de controle. O genoma levou uma década para ser concluído e custou \$10 milhões, o projeto teve início em um período pré-NGS, mas mesmo assim chegou

a usar as novas tecnologias em fases subsequentes (Attardo et al., 2014). Como o sequenciamento do genoma ficou mais rápido e barato o Instituto Nacional de Saúde dos Estados Unidos fomentou o sequenciamento de outras cinco espécies de *Glossina* que foi realizado em menos de dois anos, permitindo inclusive estudos de caracterização de potenciais genes alvo no controle da mosca (Macharia et al., 2016).

Em termos de conservação, análises genômicas que antes eram limitadas aos organismos modelo agora podem ser aplicadas a espécies ameaçadas a fim de estimar recentes acontecimentos demográficos, variação genética e a estrutura das populações (Steiner et al., 2013). Tendo em vista que o fluxo gênico entre unidades de conservação é um fator muito importante em termos de manutenção destas populações, uma técnica como NGS que traz uma grande quantidade de informações, em um curto período de tempo e com baixo custo é uma ferramenta importante para mensurar estas conexões.

DNA Histórico e Coleções Científicas

O campo de estudos com DNA antigo começou a ganhar ânimo na metade da década de 1980 quando ocorreu a extração e sequenciamento do DNA do Quagga, um equídeo da África do Sul que foi extinto no século XIX (Higuchi et al., 1984), seguido do anúncio do sequenciamento do material genético de uma múmia egípcia (Pääbo, 1985). Estes trabalhos utilizaram clonagem bacteriana para amplificar pequenas sequências de DNA (as amostras foram retiradas da pele destes espécimes) e mostraram que a origem da maioria do material genético extraído era de micróbios ou fungos. Apenas uma pequena porcentagem do DNA pertencia aos organismos em estudo e eram compostos principalmente de DNA mitocondrial (Rizzi et al., 2012). Alguns anos mais tarde o desenvolvimento da PCR deixou mais acessível à amplificação de DNA antigo, resultando em uma rápida diversificação e crescimento da área (Pääbo et al., 1989) e (Thomas et al., 1989). No entanto, devido à enorme capacidade da PCR em amplificar mesmo poucas cópias de material genético que estejam na amostra, a contaminação com DNA moderno acabou se tornando um problema importante (Rizzi et al., 2012).

Até recentemente, análises com DNA antigo eram na maioria das vezes restritas a fragmentos curtos, geralmente provenientes do genoma mitocondrial. Alguns avanços foram realizados em termos técnicos para aumentar o tamanho dos fragmentos, mas nenhuma inovação havia revolucionado a estudo com material genético antigo como o surgimento do NGS fez (Knapp and Hofreiter, 2010). Assim que a primeira plataforma foi lançada em 2005, já foi quase que imediatamente implementada na pesquisa com DNA fóssil. Apenas alguns meses após a introdução do NGS, Poinar et al., (2006) publicaram um artigo com 13 milhões de pares de base de DNA provenientes do genoma nuclear do extinto mamute. Isto representou um incremento de 480X no número de bases produzidas a partir de um exemplar fóssil em relação ao estudo que havia conseguido maior

quantidade de informação genética até então (Knapp and Hofreiter, 2010).

O acesso a grande diversidade genômica armazenada em coleções científicas criaria oportunidades sem precedentes para estudos evolutivos, filogenéticos e de genética de populações (Staats et al., 2013). Porém, várias complicações surgem ao se manipular material antigo, e existem problemas técnicos que são específicos deste campo de estudo. Estes empecilhos muitas vezes são acentuados em espécimes de museus, onde a qualidade do DNA varia muito entre as amostras e os níveis de contaminação são geralmente altos (Millar et al., 2008). Mas, a principal dificuldade consiste na produção de quantidades suficientes de sequências de DNA autênticas para que se possa chegar a resultados conclusivos. Este obstáculo é consequência de processos de degradação do DNA post mortem, que podem levar a erros na leitura ou na destruição física da molécula, aumentando as chances de sequências contaminantes exógenas serem amplificadas (Rizzi et al., 2012). Porém, a qualidade do DNA recuperado parece ser mais influenciada pelo modo como os espécimes são tratados e armazenados nas coleções do que simplesmente pelo tempo que se passou deste a coleta (Mason et al., 2011). Especificamente com relação às coleções entomológicas é sabido que métodos tradicionais usados para sacrificar os organismos, como o acetado de etila e substâncias similares, são responsáveis por danos ao material genético (Dillon et al., 1996). Estas adversidades já estavam presentes nas tentativas de sequenciamento de material antigo via método Sanger, mas agora podem ser mais facilmente contornadas via NGS. Trabalhos que utilizaram sequenciamento de nova geração em exemplares provenientes de museus mostraram que a quantidade e a qualidade das sequências geradas de espécimes antigos não foram inferiores às sequências produzidas de tecidos frescos das mesmas espécies (Staats et al., 2013).

Uma característica comum a todas as plataformas de NGS faz com que organismos coletados no passado, inclusive os insetos pertencentes às coleções entomológicas, se tornem propícios a este método de sequenciamento: a fragmentação do material genético. Com relação aos espécimes de museus a variação no tamanho dos fragmentos é gigantesca, mas em material genético fóssil geralmente são menores do que 150 pares de base, com alguns chegando a apenas 50 pares de base (Knapp and Hofreiter, 2010). A conexão entre este tipo de amostra e o NGS acontece porque as principais plataformas de NGS requerem como etapa inicial a fragmentação do DNA após a extração, seja de forma mecânica ou química (Mardis, 2013). Justamente a degradação do passar do tempo, que fragmentava o DNA de espécimes fósseis ou dos depositados em coleções, e que na maioria das vezes impossibilitava o sequenciamento via o método tradicional, agora é uma característica que não os inviabilizam mais. Aliás, este passo de fragmentação é inclusive descartado no NGS quando as amostras são de espécimes antigos (Knapp and Hofreiter, 2010). Assim, o NGS é considerado ideal para este tipo de pesquisa, uma

vez que fornecem sequências para todas as moléculas de DNA no extrato, independentemente do seu tamanho, e a probabilidade de preferência por moléculas contaminantes atuais é menor (Knapp and Hofreiter, 2010).

O advento das plataformas de NGS tornou possível pela primeira vez amostrar o DNA nuclear antigo do mamute (Poinar et al., 2006) e do Homem de Neandertal (Green et al., 2006), ambos com milhares de anos. Se estes trabalhos com resultados favoráveis fazem muitos imaginar um mundo cinematográfico de possibilidades, um trabalho mais recente, que tentou acessar o DNA fóssil de insetos impregnados em resina, nos mostrou que em alguns casos, nem mesmo a tecnologia NGS pode ser suficiente para trazer o passado à tona (Penney et al., 2013). O trabalho da equipe de pesquisadores ingleses comprovou que a existência de DNA em amostras fossilizadas em resina é altamente improvável. Os resultados do início da década de 1990 que haviam conseguido extrair DNA de âmbar de milhões de anos e que não puderam ser replicados (Austin et al., 1997), hoje são tratados como exemplos clássicos de contaminação das amostras por DNA moderno (Hebsgaard et al., 2005). Como citado anteriormente, nos protocolos tradicionais o DNA era amplificado via PCR, que preferencialmente irá amplificar qualquer molécula de DNA moderna, não danificada e que esteja como contaminante no extrato antigo gerando falsos positivos ao serem confundidos com DNA fóssil genuíno (Brown and Brown, 2011). A degradação do DNA é influenciada por vários fatores como o conteúdo de água e oxigênio, temperatura ambiente e tempo decorrido desde a morte do organismo (Lindahl, 1993). Assim, intuitivamente, poderíamos imaginar que o rápido envolvimento em resina, resultando em uma morte quase instantânea e gerando um ambiente anóxico, promoveria uma preservação do material genético do inseto, mas por razões físico-químicas ainda desconhecidas isto não ocorre (Penney, et al., 2013). Certamente os processos de preservação e fossilização não são uniformes (nem mesmo na fossilização em resina) e o resultado negativo em uma fonte de material fóssil não pode ser vista como uma negativa para todas as outras formas fósseis ou desencorajar futuras pesquisas. Dados teóricos e empíricos indicam que fragmentos de DNA podem estar presentes em material geológico bem preservado com até 100 mil anos de idade e sugerem que alguns materiais com até um milhão de anos poderiam fornecer sequências (Hebsgaard et al., 2005), nestes cenários novas tecnologias de NGS são essenciais para acessarmos tais dados.

Se no presente as novidades para o DNA fossilizado em âmbar não são tão animadoras, o mesmo não se pode dizer para amostras de coleções científicas e em especial para coleções entomológicas que possuem resultados estimulantes, ainda mais por possuírem uma riqueza incomparável de biodiversidade e de dados de distribuição geográfica protegidos em seu interior. A informação contida nestas coleções está se tornando cada vez mais acessível à medida que mais museus dispõem seus catálogos na web e os curadores destas coleções

permitem estudos moleculares com amostras mais raras. No entanto, como a maioria dos exemplares em coleções científicas são insubstituíveis, tentativas de extração de DNA destrutivas geralmente são desencorajadas. Neste sentido existem grandes progressos já realizados com o desenvolvimento de métodos não destrutivos de recuperação de DNA proveniente de insetos em museus (Gilbert et al., 2007) e (Thomsen et al., 2009), incluindo amostras do século XIX que foram fixadas de forma tradicional (Chapco and Litzenger, 2004). Quanto à contaminação existem também algumas recomendações técnicas que são importantes e que podem ser aplicadas de forma simples. São lavagens prévias dos exemplares a fim de eliminar fungos e outros microrganismos que possam ter se desenvolvido na superfície dos espécimes, além do uso de pipetas com filtro exclusivas e laboratórios distintos para extração e o processamento do material genético (Staats et al., 2013).

Resultados promissores com DNA antigo de outros grupos ou com material genético fóssil nos permitem presumir toda sua aplicabilidade latente às coleções entomológicas, como a possibilidade de estudar as relações genéticas entre organismos extintos e seus familiares contemporâneos. A preservação de muitos indivíduos, todos provenientes de uma única localidade, coletados anteriormente por naturalistas, oferece uma oportunidade única para acompanhar as mudanças em uma população ao longo do tempo (Rizzi et al., 2012). Além disso, o NGS de DNA histórico permite recuperar informações genéticas essenciais de antigos espécimes tipo e, portanto, abre também uma nova fronteira para a pesquisa taxonômica (Staats et al., 2013). Com relação às espécies extintas, espécimes de museus podem representar “fósseis em alfinetes” (Larsen, 2005), e sequências recuperadas destes exemplares podem ajudar no posicionamento destes táxons nas filogenias (Penney et al., 2013). Esta abordagem também seria útil para o estudo de faunas extintas e endêmicas de ilhas, não apenas para entender suas relações evolutivas, mas também para elucidar sua história biogeográfica.

Um ponto adicional a favor da utilização de insetos depositados em coleções científicas é a dificuldade na coleta de insetos em determinadas regiões do planeta, seja por limitações orçamentárias, conflitos armados ou até por leis restritivas de conservação da vida selvagem em determinados ambientes (Kothamasi, 2009). Algumas espécies são tão raras que não há garantia alguma de que serão coletadas novamente durante uma nova expedição de campo. Outra vantagem potencial de dados provenientes de material genético de insetos em museus seria a geração de um conjunto de dados para examinar os efeitos moleculares de eventos de poluição que ocorreram no passado, tal como testes com bombas atômicas e de outras formas de radiação ou exposição a substâncias químicas (Penney et al., 2013), como foi explorado mais recentemente em um estudo com borboletas após o acidente nuclear de Fukushima (Hiyama et al., 2012).

Até recentemente, quase todos os estudos genéticos realizados em espécimes antigos tinham como alvo

regiões do genoma mitocondrial. As expectativas futuras são de que por meio da tecnologia NGS o sequenciamento completo de genomas antigos, tanto mitocondriais como nucleares, sejam possíveis, inclusive com exemplos de sucesso em insetos (Staats et al., 2013). A identificação, estimativa e eliminação de sequências contaminantes em amostras antigas são questões que ainda precisam ser discutidas e melhor resolvidas (Rizzi et al., 2012), mas é inegável que as novas tecnologias abriram uma janela de acesso a faunas antigas de formas que não podiam ser consideradas possíveis em um passado não tão distante. O sequenciamento de nova geração pode trazer a pesquisa com DNA antigo para o centro da biologia evolutiva e torna-la uma parte crucial da genética moderna. Afinal de contas, o que melhor poderia ajudar a compreender a evolução do que visualizar seu andamento em tempo real? (Knapp and Hofreiter, 2010).

Perspectivas futuras

Com os contínuos avanços tecnológicos desta área, é arriscado especular muito adiante no futuro. No entanto, pode ser assegurado que uma das grandes vantagens das tecnologias de NGS para entomologia e campos afins, é a pequena quantidade de material genético necessário para se iniciar o sequenciamento. Isto torna estas tecnologias adequadas para estudos com espécies ameaçadas de extinção, onde é necessária uma amostragem não invasiva, bem como para exemplares de coleções científicas que requerem métodos não destrutivos de extração do DNA. Há inclusive esforços que visam ir ainda mais longe e abrem a possibilidade de análises baseadas em sequências de células individuais (Yan et al. 2013). Além disso, podemos esperar que tecnologias disponíveis no momento continuem melhorando sua capacidade de produção de sequências juntamente com a diminuição da taxa de erro, dando maior qualidade a montagem dos genomas.

O incremento da velocidade e decréscimo dos custos na geração de dados moleculares está transformando vários campos da biologia. As tecnologias de NGS estão revolucionando a ecologia ao derrubar as barreiras entre ecologia, biologia molecular, genética e genômica (Tautz et al., 2010). O futuro parece muito promissor e é muito interessante que cada vez mais pesquisas adicionem dados moleculares aos seus resultados. Certamente estas novas tecnologias de sequenciamento irão contribuir para que a informação contida no material genético dos organismos se difunda como ferramenta aos cientistas.

Cientistas que se entusiasmarem com as possibilidades apresentadas neste texto e que não se intimidarem com os desafios ainda existentes, podem decidir por utilizar estas novas tecnologias de sequenciamento. No entanto, precisarão de cautela para não desperdiçar recursos. Raramente laboratórios isolados de zoologia ou ecologia terão uma demanda tão grande de sequências genéticas conforme a capacidade destes novos equipamentos. Assim, o ideal é realizar parcerias com laboratórios que já disponham destes sequenciadores, a fim de otimizar o uso e não obter um patrimônio que em pou-

cos anos estará ultrapassado em termos de tecnologia. Uma alternativa é a terceirização do sequenciamento, atualmente existem muitas empresas no mercado que realizam o sequenciamento de forma comercial. Aos que pretendem adquirir os novos sequenciadores vale lembrar que é preciso considerar os custos com mão de obra altamente especializada para a operação e manuseio que são específicas de cada plataforma. É preciso também considerar os custos com a manutenção dos instrumentos, gasto que muitas vezes não é levado em consideração quando da aquisição deste tipo de equipamento e que geralmente possui um suporte técnico oneroso.

Agradecimentos

Agradeço ao Professor Dr. Marcio R. Pie pelas positivas sugestões neste manuscrito e fundamental incentivo durante todo o período de desenvolvimento deste trabalho. Sou grata ao Programa de Pós-Graduação em Ciências Biológicas (Área de concentração em Entomologia) da Universidade Federal do Paraná (UFPR) e aos seus professores que estiveram presentes durante toda a minha formação científica. Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de doutorado número 140262/2013-0, e agradeço a bolsa doutorado-sanduíche no exterior concedida pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) número 99999.002880/2014-08. O apoio financeiro de ambas as agências foi essencial para o desenvolvimento deste artigo e pesquisa desenvolvida durante a revisão.

Referências

- Attardo G, Abila, PP, Auma JE, Baumann AA, Benoit JB, Brelsfoard CL, et al., 2014. Genome sequence of the Tsetse fly (*Glossina morsitans*): Vector of African Trypanosomiasis. *Science* 344(6182): 380–386.
- Austin JJ, Ross AJ, Smith AB, Fortey RA, Thomas RH. 1997. Problems of reproducibility - does geologically ancient DNA survive in amber-preserved insects? *Proceedings of the Royal Society B: Biological Sciences* 264(1381): 467–474.
- Brodin Y, Ejdung G, Strandberg J, Lyrholm T. 2012. Improving environmental and biodiversity monitoring in the Baltic sea using DNA barcoding of Chironomidae (Diptera). *Molecular Ecology Resources* 13(06): 996–1004.
- Brown KA, Brown TA. 2011. *Biomolecular archaeology: An introduction*. Malden, MA: Wiley-Blackwell (an imprint of John Wiley & Sons Ltd).
- Celniker S E, Wheeler DA, Kronmiller B, Carlson JW, Halpern A, Patel S. et al., 2002. Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Genome Biology* 3(12):research0079.1–0079.14
- Chacón I, Janzen D, Hallwachs W, Sullivan J, Hajibabaei M. 2013. Cryptic species within cryptic moths: New species of *Dunama schaus* (Notodontidae, Nystaleinae) in Costa Rica. *ZooKeys* 264: 11–45.
- Chapco W, Litzemberger G. 2004. A DNA investigation into the mysterious disappearance of the Rocky Mountain grasshopper, mega-pest of the 1800s. *Molecular Phylogenetics and Evolution* 30(3): 810–814.
- Chial H. 2008. Human genome project: Sequencing the human

- genome. *Nature Education* 1(1):219.
- Darwin C. 1859. *The origin of species (by means of natural selection)*. London: John Murray.
- Davey J W, Blaxter ML. 2011. RADSeq: Next-generation population genetics. *Briefings in Functional Genomics* 9(5-6):416–423.
- Dillon, N, Austin AD, Bartowsky E. 1996. Comparison of preservation techniques for DNA extraction from hymenopterous insects. *Insect Molecular Biology* 5(1):21–24.
- Egan A N, Schlueter J, Spooner DM. 2012. Applications of next-generation sequencing in plant biology. *American Journal of Botany* 99(2):175–185.
- Eklom R, Galindo J. 2010. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* 107(1): 1–15.
- El-Metwally S, Hamza T, Zakaria M, Helmy M. 2013. Next-generation sequence assembly: Four stages of data processing and computational challenges. *PLoS Computational Biology* 9(12): e1003345.
- Feldmeyer B, Elsner D, Foitzik S. 2013. Gene expression patterns associated with caste and reproductive status in ants: Worker-specific genes are more derived than queen-specific ones. *Molecular Ecology* 23(1):151–161.
- Fukatsu T. 2012. Next-generation sequencing sheds light on intricate regulation of insect gut microbiota. *Molecular Ecology* 21(24):5908–5910.
- Fuller C W, Kumar S, Porel M, Chien M, Bibillo A, Stranges PB, et al. 2016. Real-time single-molecule electronic DNA sequencing by synthesis using polymer-tagged nucleotides on a nanopore array. *Proceedings of the National Academy of Sciences* 113(19):5233–5238.
- Gadau J, Helmkamp M, Nygaard S, Roux J, Simola DF, Smith CR, et al. 2012. The genomic impact of 100 million years of social evolution in seven ant species. *Trends in Genetics* 28(1):14–21.
- Genia Technologies. 2016. <http://www.geniachip.com/technology/>
- Gilad Y, Pritchard JK, Thornton K. 2009. Characterizing natural variation using next-generation sequencing technologies. *Trends in Genetics* 25(10):463–471.
- Gilbert MTP, Moore W, Melchior L, Worobey M. 2007. DNA extraction from dry museum beetles without conferring external morphological damage. *PLoS ONE* 2(3):e272.
- Glenn TC. 2011. Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* 11(5):759–769.
- Glenn T. 2016 NGS field guide: Overview. <http://www.molecularecologist.com/next-gen-fieldguide-2016>
- Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons J F, et al. 2006. Analysis of one million base pairs of Neanderthal DNA. *Nature* 444(7117):330–336.
- Hebsgaard MB, Phillips MJ, Willerslev E. 2005. Geologically ancient DNA: Fact or artefact? *Trends in Microbiology* 13(5):212–220.
- Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312(5991): 282–284.
- Hiyama A, Nohara C, Kinjo S, Taira W, Gima S, Tanahara A, Otaki JM. 2012. The biological impacts of the Fukushima nuclear accident on the pale grass blue butterfly. *Scientific Reports* 2:570.
- Janzen DH, Hallwachs W, Harvey DJ, Darrow K, Rougerie R, Hajibabaei M, et al. 2012. What happens to the traditional taxonomy when a well-known tropical saturniid moth fauna is DNA barcoded? *Invertebrate Systematics* 26(6):478.
- Jones DC, Ruzzo WL, Peng X, Katze MG. 2012. Compression of next-generation sequencing reads aided by highly efficient de novo assembly. *Nucleic Acids Research* 40(22): e171.
- Kasianowicz JJ, Brandin E, Branton D, Deamer DW. 1996. Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences* 93(24): 13770–13773.
- Knapp M, Hofreiter M. 2010. Next generation Sequencing of ancient DNA: Requirements, strategies and perspectives. *Genes* 1(2):227–243.
- Kothamasi D, KIERS ET. 2009. Emerging conflicts between Biodiversity conservation laws and scientific research: The case of the Czech Entomologists in India. *Conservation Biology* 23(5):1328–1330.
- Larsen TB. 2005. Hazards of butterfly collecting – fossil on a pin. *The entomologist's record and journal of variation* 117:109–111.
- Lindahl T. 1993. Instability and decay of the primary structure of DNA. *Nature* 362(6422):709–715.
- Macharia R, Mireji P, Murungi E, Murilla G, Christoffels A, Aksoy S, Masiga D. 2016. Genome-wide comparative analysis of Chemosensory gene families in Five Tsetse fly species. *PLOS Neglected Tropical Diseases* 10(2):e0004421.
- Mardis ER. 2008. Next-generation DNA Sequencing methods. *Annual Review of Genomics and Human Genetics* 9(1):387–402.
- Mardis ER. 2013. Next-generation Sequencing platforms. *Annual Review of Analytical Chemistry* 6(1):287–303.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376–380.
- Mason VC, Li G, Helgen KM, Murphy WJ. 2011. Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome Research* 21(10):1695–1704.
- McCormack JE, Faircloth BC. 2012. Next-generation phylogenetics takes root. *Molecular Ecology* 22(1):19–21.
- Metzker ML. 2009. Sequencing technologies — the next generation. *Nature Reviews Genetics* 11(1):31–46.
- Millar CD, Huynen L, Subramanian S, Mohandesan E, Lambert DM. 2008. New developments in ancient genomics. *Trends in Ecology & Evolution* 23(7):386–393.
- Morozova O, Marra MA. 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92(5):255–264.
- Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. 1986. Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology* 51:263–273.
- Nikolaki S, Tsiamis G. 2013. Microbial diversity in the era of Omic technologies. *BioMed Research International* 2013: 1–15.
- Ninomiya M, Ueno Y, Shimosegawa T. 2013. Application of deep sequence technology in hepatology. *Hepatology Research* 44(2):141–148.
- Penney D, Wadsworth C, Fox G, Kennedy SL, Preziosi RF, Brown TA. 2013. Absence of ancient DNA in sub-fossil insect inclusions preserved in “Anthropocene” Colombian Copal. *PLoS ONE* 8(9): e73150.
- Poinar, HN. 2006. Metagenomics to Paleogenomics: Large-scale Sequencing of mammoth DNA. *Science* 311(5759):92–394.
- Pääbo, S. 1985. Molecular cloning of ancient Egyptian mummy DNA. *Nature* 314(6012):644–645.
- Pääbo S, Higuchi R, Wilson A. 1989. Ancient DNA and the poly-

- merase chain reaction. The emerging field of molecular archaeology. *The Journal of biological chemistry* 264(17):9709–12.
- Quail M, Smith ME, Coupland P, Otto TD, Harris SR, Connor TR, et al. 2012. A tale of three next generation sequencing platforms: Comparison of ion torrent, pacific biosciences and illumina MiSeq sequencers. *BMC Genomics* 13(1):341.
- Richards S, Gibbs R, Weinstock G, Brown S, Denell R, Beeman R, et al. 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature* 452(7190):949–55
- Rizzi E, Lari M, Gigli E, De Bellis G, Caramelli, D. 2012. Ancient DNA studies: New perspectives on old samples. *Genetics Selection Evolution*, 44(1):21.
- Robison K. 2016. Omics! Omics! <http://omicsomics.blogspot.com.br/2016/04/genia-publishes-platform-progress.html>
- Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes J C, et al. 1977a. Nucleotide sequence of bacteriophage ϕ X174 DNA. *Nature* 265(5596):687–695.
- Sanger F, Nicklen S, Coulson AR. 1977b. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12):5463–5467.
- Schnable J. 2010. Even faster Sequencing. <http://www.jamesandthegiantcorn.com/2010/01/13/even-fastersequencing/>
- Schuster SC. 2007. Next-generation sequencing transforms today's biology. *Nature Methods* 5(1):16–18.
- Shendure J, Ji H. 2008. Next-generation DNA sequencing. *Nature Biotechnology* 26(10):1135–1145.
- Shokralla S, Gibson JF, Nikbakht H, Janzen DH, Hallwachs W, Hajibabaei M. 2014. Next-generation DNA barcoding: Using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Molecular Ecology Resources* 14(5): 892–901.
- Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell C R, et al. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* 321(6071): 674–679.
- Smith MA, Eveleigh ES, McCann KS, Merilo MT, McCarthy PC, Van Rooyen K I. 2011. Barcoding a quantified food web: Crypsis, concepts, ecology and hypotheses. *PLoS ONE* 6(7): e14424.
- Staats M, Erkens RHJ, van de Vossen B, Wieringa JJ, Kraaijeveld K, Stielow B, et al. 2013. Genomic treasure Troves: Complete genome Sequencing of Herbarium and insect museum specimens. *PLoS ONE*, 8(7):e69189.
- Sudakaran S, Salem H, Kost C, Kaltenpoth M. 2012. Geographical and ecological stability of the symbiotic mid-gut microbiota in European firebugs, *Pyrrhocoris apterus* (Hemiptera, Pyrrhocoridae). *Molecular Ecology* 21(24):6134–6151.
- Tautz D, Ellegren H, Weigel D. 2010. Next generation molecular ecology. *Molecular Ecology* 19: 1–3.
- The International Silkworm Genome Consortium. 2009. The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect biochemistry and molecular biology* 38(12):1036–45.
- Thomas R H, Schaffner W, Wilson AC, Pääbo S. 1989. DNA phylogeny of the extinct marsupial wolf. *Nature* 340(6233):465–467.
- Thomsen PF, Elias S, Gilbert MTP, Haile J, Munch K, Kuzmina S, et al. 2009. Non-destructive sampling of ancient insect DNA. *PLoS ONE*, 4(4):e5048.
- Valentini A, Pompanon F, Taberlet P. 2009. DNA barcoding for ecologists. *Trends in Ecology & Evolution* 24(2):110–117.
- Wallace, LJ, Boilard SMAL, Eagle SHC, Spall JL, Shokralla S, Hajibabaei M. 2012. DNA barcodes for everyday life: Routine authentication of natural health products. *Food Research International* 49(1):446–452.
- Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: A revolutionary tool for transcriptomics. *Nat Rev Genet* 10(1):57–63.
- Watson JD. 1990. The human genome project: Past, present, and future. *Science* 248(4951):44–49.
- Watson JD. 2001. The human genome revealed. *Genome Research* 11(11):1803–1804.
- Welburn SC, Maudlin I, Simarro PP. 2009. Controlling sleeping sickness – a review. *Parasitology* 136(14):1943.
- Yan L, Yang M, Guo H, Yang L, Wu J, Li R, et al. 2013. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nature Structural & Molecular Biology* 20(9):1131–1139.
- Yoder J. 2014. 2014 NGS field guide: Resistance is futile (mostly, at least for a while). <http://www.molecularecologist.com/2014/03/2014-ngs-field-guide-resistance-is-futilemostly-at-least-for-a-while/>