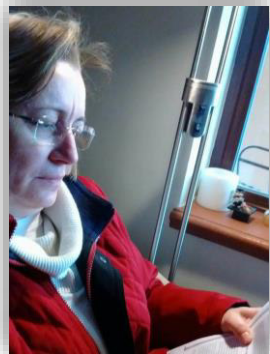


2. Genômica Populacional: técnicas, aplicações e desafios



Marínes Marli Gniech Karasawa

Eng. Agr. pela Universidade Federal de Pelotas – RS (UFPeL, 1993), Me. na Universidade Federal de Lavras – MG (UFLA, 2001), PhD. pela Universidade de São Paulo (ESALQ, 2005), Pós-Doutorado na Universidade Estadual do Sudoeste da Bahia - BA (UESB), na Universidade Estadual de Feira de Santana - BA (UEFS) e na Universidade de Palermo – Itália (UNIPA, 2014). Atuou como Professora Convidada na UESB (2007), ministrando biologia geral para graduação e, como Professora Adjunta na Universidade Federal de Alfenas – MG (UNIFAL, 2009-2012), com as disciplinas: genética clássica, genética humana, genética de populações, biologia molecular, e engenharia genética.

Introdução

A genômica populacional hoje é tema recorrente em recursos genéticos, se referindo ao estudo de variações das frequências gênicas ou genotípicas que ocorrem em populações em consequência de processos evolutivos como seleção natural, mutação, deriva genética e fluxo gênico, levando em consideração os fatores, tais como: recombinação, estrutura e subdivisão populacional que podem ou não estar associados a fenômenos adaptativos ou de especiação.

Ao longo dos anos os geneticistas de populações vinham estudando os padrões de variação na estrutura genética (causadas por seleção, mutação, migração, fluxo gênico e distribuição espacial dentro e entre populações) com base em estudos de um número limitado de marcadores anônimos de genes ou locos gênicos (BRAGG *et al.*, 2015). Carente de ferramentas genômicas, a genética de populações se concentrava na análise de amostras de genes, relativamente pequenas, obtidas de populações naturais. No entanto, a partir dos anos 90's, ferramentas de análise genômica passaram a ser aplicadas para estudos de polimorfismo genético em seres humanos e, em consequência, o termo “genômica de populações” começou a ser usado para descrever a análise de polimorfismos observados em larga-escala (ELLEGRÉN, 2014). Desde então, cada vez mais, passou-se a utilizar o termo genômica de populações para designar os estudos de genética de populações que avaliam polimorfismos de DNA em escala genômica.

Do que se trata?

O termo “genômica populacional” se refere ao estudo abrangente de todos os genes do genoma de um organismo (HARTL & CLARK, 2010). Em seu sentido mais amplo, esta é uma área de estudo que combina os objetivos da genética de populações naturais com a amostragem genômica ampla (GWAS – *genome wide association*), para compreender a adaptação específica, a evolução, e conhecer a força evolutiva que está afetando a população (GONZÁLEZ-MARTÍNEZ *et al.*, 2006).

Como citado anteriormente, somente recentemente as abordagens da genômica de populações incluíram o sequenciamento de DNA em larga-escala, ensaios sobre abundância de

transcritos e estudos funcionais por meio de eliminação ou inibição sistemática da atividade de cada gene. Ainda, métodos computacionais foram adotados para manejo, comparação e interpretação, com o objetivo de estudar as frequências alélicas, identificar genes polimórficos e definir a sua posição no genoma, verificar a presença de desequilíbrio de ligação gênica, estabelecer a estrutura e a distribuição da diversidade genética, realizar inferências sobre a história demográfica de uma população e definir qual das forças evolutivas (mutação, migração, seleção e deriva genética) está atuando sobre os indivíduos da população em estudo (HARTL & CLARK, 2010).

Estudos genômicos de populações também tem permitido acessar questões relativas a conservação das espécies, tais como: efeitos da mudança de *habitat* sobre a adaptação local, sobre a estrutura e a diversidade genética (ALLENDORF *et al.*, 2010). Além disto, os estudos espaciais e temporais estão ajudando a entender os efeitos da fragmentação de *habitat* e das atividades humanas sobre a estrutura e a diversidade genética das populações naturais. Estas informações ajudam estabelecer planos de manejo em *habitats* perturbados e heterogêneos (VEGA *et al.*, 2016), responder questões sobre a relação filogenética entre táxons não resolvidos e determinar os locos gênicos responsáveis pela especiação, adaptação local, interação entre as espécies e a depressão por endogamia (FUENTES-PARDO & RUZANTE, 2017).

Em tempos de mudança climática crescente, torna-se muito relevante o estudo da genômica populacional. Compreender a base genética da adaptação e da divergência entre populações é um dos principais objetivos da genética de populações e serve de caminho para entender a evolução (KRUTOVSKY, 2006). Mudanças extremas em variáveis ambientais (tais como: temperatura, humidade, luminosidade) geralmente tem efeito sobre o genoma das espécies que podem ajustar a distribuição da variabilidade genética, caso exista variação ao nível de genoma, ou podem gerar novas variações por meio de mutações que produzam adaptação ou, ainda, serem extintas por não serem capazes de se adaptar. De acordo com Bragg *et al.* (2015) o estudo da distribuição da variação genômica ao longo de paisagens pode indicar interações complexas entre genes e ambiente, os quais podem determinar e/ou influenciar a distribuição da espécie, podendo ainda indicar a associação/adaptação de determinados fenótipos a condições locais. Deste modo, determinar a direção da evolução e o status adaptativo das espécies pode ajudar a entender melhor como os mecanismos evolutivos afetam genes e causam adaptação (YANG *et al.*, 2017).

O desenvolvimento do sequenciamento de última geração (*Next Generation Sequence* - NGS) utilizando plataformas Illumina (HiSeq ou MiSeq), Roche (454) e *Life technologies* (Ion Proton, Torrent ou SOLiD) e o desenvolvimento de ferramentas mais robustas na bioinformática, que são capazes de avaliar grande quantidade de polimorfismos de DNA, tem permitido acessar uma vasta gama de questões genéticas e evolutivas em organismos não-modelo, que é o caso das populações nativas, contudo a quantidade de DNA repetido segundo Ellegren (2014) ainda constitui um obstáculo para acessar todo o genoma e afeta a continuidade do sequenciamento.

Técnicas de estudo de genomas

Com o advento do sequenciamento de última geração (NGS) foram desenvolvidas técnicas de genotipagem reduzida do genoma em larga escala através da análise de polimorfismos de nucleotídeos isolados (*Single Nucleotide Polymorphisms* – SNPs). Dentre as técnicas que diminuem a complexidade são descritos os GBS (*Genotyping by Sequencing*), RAD e ddRAD (*Restriction Site Analysis*) e a RNAseq (Transcriptômica) que consiste no sequenciamento de cDNA a partir do mRNA. Utilizando estas técnicas, é possível realizar a análise de polimorfismo de genoma completo baseado em grande quantidade de amostras obtidas de populações não-modelo, embora este tipo de estudo possa ser um pouco complicado pela inexistência de um genoma de referência e pelo elevado custo gerado na análise de múltiplos indivíduos (FUENTES-PARDO & RUZANTE, 2017).

RAD, ddRAD

A técnica RAD (*Restriction-Site Associated DNA-Sequencing*) proposta por Baird *et al.* (2008) envolve o uso de enzimas de restrição para fragmentar o DNA, o que promove considerável redução nos custos de genotipagem, se tornou bastante popular por poder ser aplicada a qualquer espécie sem a necessidade de caracterização prévia dos SNPs (*Single Nucleotide Polymorphisms*). Esta metodologia tem sido utilizada para realizar estudos de estrutura genética, hibridização entre espécies, história demográfica, migração e filogeografia, bem como para detectar locos envolvidos na adaptação e divergência de populações naturais.

Modificações na técnica de RAD tem sido feitas para melhorar a eficiência no sequenciamento utilizando uma enzima de corte frequente e uma enzima de corte raro e corte com enzima seguido de sonicação (para inserção de adaptador randômico). Mas, de acordo com Sheben *et al.* (2017) para o sequenciamento de múltiplas amostras em uma única *lane* (linha de sequenciamento) adaptadores com código de barras devem ser ligados às extremidades digeridas. E, a biblioteca gerada deve passar por um screening para seleção de fragmentos de tamanho adequado para o sistema de sequenciamento adotado, os quais devem ser amplificados com ambos os adaptadores. Entretanto, para Lowry *et al.* (2017) embora a técnica RAD seja eficiente e tenha custo baixo, em função dos problemas detectados, ao realizar estudos adaptativos antes de utilizar o sequenciamento baseado na técnica RAD deveriam ser considerados outros métodos, pelo baixo polimorfismo apresentado e pela dificuldade em capturar variações existentes em regiões gênicas envolvidas na adaptação. Um exemplo seria utilizar RNAseq (transcriptômica) ou exoma quando o DNA de referência não estiver disponível.

GBS

Genotyping by Sequencing é uma variação da técnica RADseq que foi simplificada por Elshire *et al.* (2011) e se baseia no uso de enzimas de restrição que são sensíveis a metilação, tais como a *ApeKI*. Estas, são capazes de eliminar regiões repetitivas do genoma o que simplifica muito os problemas computacionais durante o alinhamento em espécies com elevado nível de diversidade genética. Nesta metodologia o DNA dos indivíduos em estudo deve ser cortado com a enzima

ApeKI (Figura 1) que é sensível a regiões metiladas ou qualquer outra enzima que deixa saliências (extremidades coesivas) usadas para ancoragem de *barcodes* e ligação de adaptadores de 4 a 8 pares de base (pb). Após amplificar o DNA, procede-se a purificação dos fragmentos, verificação de tamanho dos fragmentos, sendo ideal 170 a 350 pares de base (pb), e o sequenciamento.

Estudos posteriores, otimizaram a técnica de GBS para genomas altamente repetitivos para espécies sem genoma de referência utilizando um sistema com duas enzimas (*PstI/MspI*) permitindo à amplificação das sequências *forward* e *reverse*. Neste sistema, dois grupos de adaptadores são construídos, um possuindo *barcodes* e sequências complementares ao *primer forward* e outro complementar ao *primer* de sequenciamento *reverse* (POLAND *et al.*, 2012). Estudos posteriores realizados por Sonah *et al.* (2013) avaliaram a eficiência das enzimas *ApeKI*, *PstI* e *MspI* verificaram que a enzima *ApeKI* gerou a maioria dos fragmentos no tamanho desejado, enquanto *MspI* gerou inúmeros fragmentos muito pequenos que devem dificultar o bom performance durante o sequenciamento na plataforma Illumina, e a enzima *PstI*, ainda que associada a *MspI* (estratégia de duas enzimas), foi capaz de acessar poucas regiões genômicas de interesse. Outras adaptações na técnica RAD-GBS têm permitido o sequenciamento pelo sistema Ion torrente em espécies com pouco DNA (LEBORDOUS *et al.*, 2014).

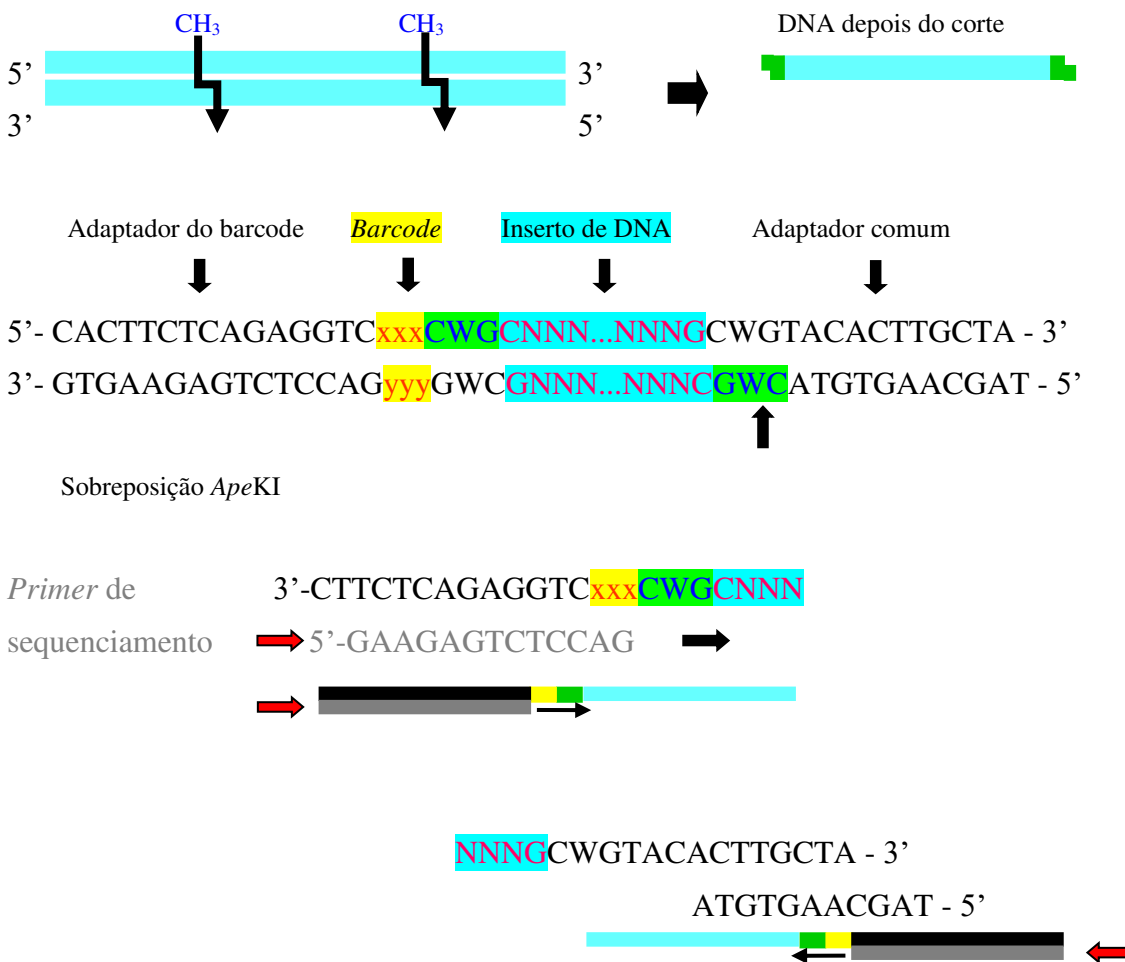


Figura 1: Etapas no preparo do DNA pela técnica de GBS utilizando a enzima *ApeKI* (ELSHIRE *et al.* 2011).

O uso desta técnica tem reduzido drasticamente o custo e o tempo de preparo das bibliotecas. Além disto, o uso de *barcodes* tem permitido que amostras individuais fossem agrupadas para o conjunto comum de *primers* permitindo amplificar as bibliotecas por PCR, realizar estudos de espécies não-modelo e, ainda, facilitando a identificação das amostras tornando mais precisa a associação entre genótipo e o fenótipo (BATHIA *et al.*, 2013). Assim, a técnica de GBS vem desempenhando papel importante na transição da área de genética de populações para genômica de populações que, associada com caracteres fenotípicos, permite o mapeamento e a rápida identificação de genes para a introgressão no germoplasma de plantas cultivadas (SCHEBEN *et al.*, 2017).

Transcriptômica e proteômica

A análise de transcriptoma lida com questões que envolvem partes transcritas do genoma e o nível de atividade dos genes transcritos. No passado, estas questões eram avaliadas com os microarranjos (que se baseia na hibridação de sondas de DNA com amostras de RNA). A utilização da abordagem NGS nos estudos de RNA-seq (transcriptômica) não depende da anotação do genoma, pois o sequenciamento do RNA mensageiro (mRNA - genes transcritos) da população permite saber qual parte do genoma está sendo transcrita (WANG, 2016). Para estudos de RNAseq trabalhos clássicos envolvem a extração do RNA total e a separação mRNA por purificação utilizando kits de purificação ou colunas contendo esferas de oligonucleotídeos de timina (T) que separam o mRNA (Figura 2). Em seguida, procede-se a síntese do cDNA por transcriptase reversa. O cDNA obtido passa pela etapa de fragmentação, reparo das extremidades, adição de adaptadores únicos e amplificação por PCR para a formação das bibliotecas que, posteriormente, serão sequenciadas.

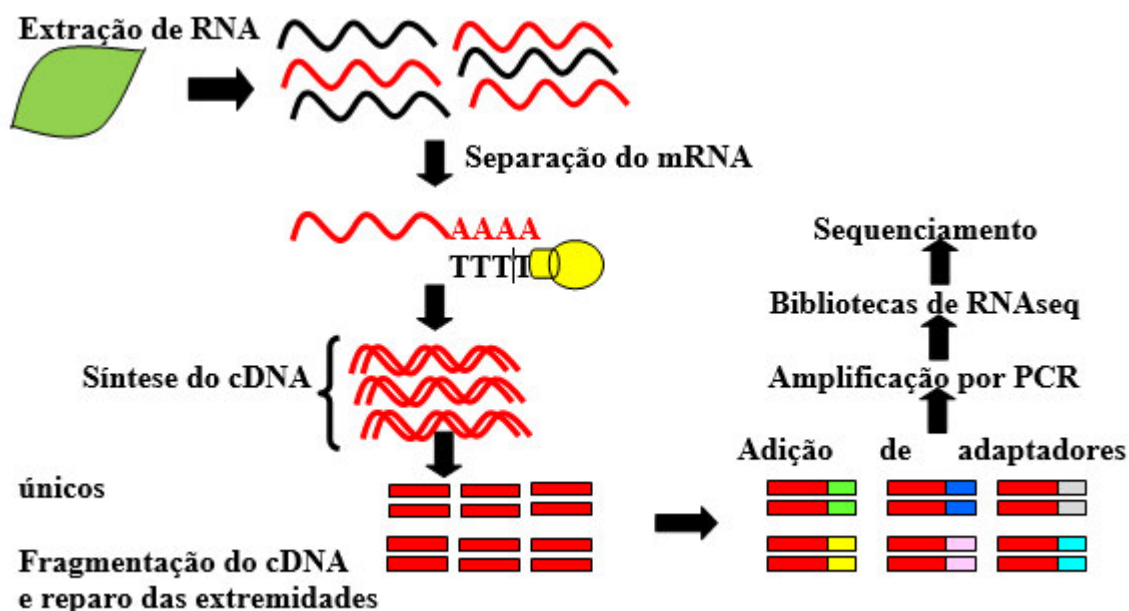


Figura 2. Esquema das etapas envolvidas no RNAseq (esquema baseado em Kumar *et al.*, 2012)

A qualidade superior e a rápida obtenção dos resultados tornaram viável a execução de projetos pequenos e análise da expressão gênica em larga escala, apesar da dificuldade e do custo um pouco elevado na construção das bibliotecas (KUMAR *et al.*, 2012). Uma vantagem do RNA-seq é a capacidade de coletar expressão gênica e polimorfismo gênico (SNP) de um único conjunto de dados. Pela comparação dos resultados do controle com os tratamentos é possível identificar genes ativos e reprimidos em resposta a um tratamento particular (expressão gênica diferencial) em espécies não-modelo (TOOD *et al.*, 2016) e agrupar estes genes em categorias funcionais (GO – Gene Ontology). Assim, é possível descobrir quais processos moleculares estão sendo afetados no organismo pela mudança ambiental (DE WIT *et al.*, 2012).

Apesar da técnica ainda ser subutilizada em estudos adaptativos, estas metodologias têm criado ferramentas com potencial de promover significativo avanço na compreensão da diversificação adaptativa, além de estar proporcionando *insights* biológicos sobre o desenvolvimento, função gênica e respostas fisiológicas ao estresse sendo recomendada em estudos que visam melhorar a compreensão sobre evolução e a adaptação (VOECKEL *et al.*, 2017; LOWRY *et al.*, 2017).

Estudos de genoma completo

Sequências de genoma completo, em contraste às sequências de locos individuais, revelam a biologia do genoma e como o material genético é organizado permitindo verificar a abundância de elementos transponíveis, densidade dos genes no genoma, além de retratar muitas outras características, tais como: composição de bases, RNAs codificadores e, marcas de modificações nos nucleotídeos e na cromatina (ex: taxa de recombinação) que são parâmetros críticos em estudos de genética de populações e evolução. Uma das percepções obtidas dos estudos de genoma completo é que a taxa de recombinação ao longo do genoma é bem mais heterogênea do que se pensava e que existem regiões, tais como as extremidades dos braços do cromossomo, que apresentam um elevado índice de recombinação. Uma das implicações é que a recombinação afeta a eficiência da seleção de sítios ligados e a seleção de sítios locais. Este fenômeno, conhecido como efeito *Hill-Robertson*, ocorre porque diferentes mutações favoráveis devem surgir em contextos genéticos diferentes e, conforme os alelos favorecidos aumentam de frequência eles produzem desequilíbrio de ligação negativa com alelos de regiões vizinhas fazendo com que o produto das frequências dos gametas que contém um alelo favorável e outro desfavorável seja maior do que o dos gametas que contém ambos os alelos favoráveis ou ambos desfavoráveis, o que dificulta a seleção do alelo favorável (HARTL & CLARK, 2010). Neste caso, a seleção deve favorecer a taxa de recombinação nestas regiões. Assim, a possibilidade de obter estimativas da taxa de recombinação (combinando análises de ligação e sequências genômicas) tem produzido *insights* sobre a evolução da composição de nucleotídeos e a sua relação com a história evolutiva do organismo (ELLEGREN, 2014).

De acordo com Weigel & Nordborg (2015) o problema começa quando algumas características são altamente herdáveis sem que qualquer efeito marginal seja detectado nos polimorfismos, mostrando que a “seleção” mais do que a “deriva” determina o destino do alelo

desde que o seu efeito sobre a adaptação seja maior que o inverso do tamanho efetivo da população (N_e). Organismos grandes que tenham tamanho populacional pequeno também podem sofrer forte efeito da seleção a favor (ou contra) o aumento (ou redução) da aptidão do alelo através de efeitos que jamais possam ser vistos no desenvolvimento, na fisiologia e na morfologia. Além disto, fatores ambientais bióticos e abióticos que jamais são constantes afetam os efeitos da seleção. Entretanto, quando regiões codantes ou genes expressos são afetados pela deleção, mudança na função gênica ou tipo de proteína, é possível caracterizar polimorfismos moleculares com importância adaptativa.

Bioinformática aplicada ao estudo de sequências genômicas

Para responder qualquer pergunta a nível biológico, os arquivos (no formato FASTA) contendo as informações geradas no sistema NGS devem ser submetidos a *verificação de qualidade das leituras* que consiste na remoção de bases com escore inferior a 20, remoção dos adaptadores contendo *barcodes*, elaboração de gráficos contendo informações a respeito da qualidade dos escores e dos nucleotídeos e, cálculo das frações de leituras em duplicata e *singleton* presente nos dados. Verificada a qualidade das leituras é realizado o *alinhamento das sequências* para a formação dos *contigs* (sequências de leitura mais longas) que irão representar cada mRNA transcrito (para estudos de RNAseq).

Estabelecer um *link* entre a variação observada a nível de sequência e a função gênica é um grande desafio para dados de RNAseq oriundos de organismos não-modelo. Neste sentido, utiliza-se a ferramenta de busca para alinhamento básico local (BLAST – *Basic Local Alinhament Search Tool*) para proceder anotações com base na semelhança de genes/proteínas/funções conhecidas. Consultando as três maiores bases de dados (*GenBank*, *Uniprot Swiss-Protein* e TrEMBL) é possível identificar as sequências mais semelhantes para cada um dos *contigs* e a informação disponível para as sequências identificadas pode ser utilizada para anotar as propriedades mais prováveis dos *contigs* obtidos no estudo. No caso de coincidências significantes entre os *contigs* e a base de dados será possível, também, identificar o nome dos genes, obter sua descrição geral e a classificação de agrupamento para categorias específicas de genes (GO - *Gene Ontology*) baseados na função celular e molecular.

Depois disto, é realizado o *mapeamento* que consiste no alinhamento das leituras a uma sequência de referência. Para realizar o mapeamento das leituras em uma sequência de referência existem vários programas que diferem nos algoritmos e na velocidade. O objetivo é criar um arquivo conhecido como mapa de alinhamento de sequência (SAM – *Sequence Alignment Map*) para cada uma das amostras. Este arquivo será composto por uma linha para as leituras de cada amostra chamada de sequência de referência (gene, *contig* ou região gênica) para a qual ela foi mapeada, a posição na sequência de referência e o score do mapeamento na escala Phred, entre outros. As informações SAM serão utilizadas para extrair informações sobre a expressão gênica e para identificar o polimorfismo presente nos dados. Vários parâmetros podem ser definidos nesta etapa. Para maiores detalhes acessar <http://bio-bwa.sourceforge.net/bwa.shtml>

A etapa seguinte envolverá a contagem do número de leituras, normalização e teste para diferenças e, análise de enriquecimento funcional. A contagem do número de leituras no arquivo de cada amostra (mapeados uma única vez para cada *contig* ou gene) permite ter uma noção do nível de expressão gênica de cada transcrito, pois transcritos mais abundantes apresentam um maior nível de expressão.

Diferentes populações em geral são geneticamente diversas, o que pode apresentar implicações importantes no potencial de adaptação e na resiliência das espécies. Assim, os padrões de diversidade genética detectada no polimorfismo de nucleotídeos (SNP) podem fornecer uma idéia da história demográfica e evolutiva da espécie. Os níveis de polimorfismo observados entre as espécies são bastante variáveis, mas leituras curtas obtidas de uma única lane facilmente identificam aproximadamente 1000 SNPs. Para detectar a variação real é preciso separar os polimorfismos do sequenciamento e alinhamento por meio de uma filtragem dos dados baseada em modelos estatísticos, tal como o modelo mendeliano e o de equilíbrio de Hardy-Weinberg.

Espera-se que nível de expressão gênica varie entre amostras e/ou indivíduos devido a fatores externos. Uma vez conhecido o fator que está produzindo a variação (ex.: diferenças na temperatura) este poder ser utilizado para verificar a sua relação com questões relativas a estrutura e diversidade genética das populações, adaptação e fatores evolutivos que afetam a espécie.

Estatísticas para estudo das populações naturais

Sabe-se que as forças estudadas pela genética de populações clássica, tais como a mutação, migração, deriva e seleção apresentam impacto direto no padrão de variação que pode ser observado ao nível de sequência de nucleotídeos na molécula de DNA. Assim, o conhecimento da sequência completa dos genes fornece um retrato muito mais rico a respeito da variação genética existente nas populações do que as informações obtidas pelo modelo clássico “A” e “a” (onde genes eram entidades quimicamente indefinidas). Em função disto, a genética de populações molecular passou a requerer uma parametrização muito mais detalhada de processos envolvidos na evolução das espécies, tais como a mutação. Assim, sempre que padrões de variação inesperados são observados nas sequências de DNA surgem novas demandas para o desenvolvimento teórico da genética de populações (HARTL & CLARK, 2010).

Atualmente, os estudos de genética de populações passaram a ter abordagens em escala de paisagem, onde a frequência de genes e alelos passa a ser olhada dentro da paisagem onde a população se encontra, incorporando aspectos relativos ao ambiente ecológico em que vivem (tais como: altitude, luminosidade, humidade, temperatura, etc.). Para viabilizar inferências biológicas mais amplas foram necessários avanços na estatística de populações afim de permitir o refinamento e ampliação de abordagens, tais como: regressão múltipla de matrizes de distância e o teste de restrições de Mantel, e o desenvolvimento métodos completamente novos (tais como: a teoria dos circuitos, gráficos populacionais, gráficos de componentes principais e matrizes vizinhas).

A vasta gama de questões potenciais geradas no estudo das espécies, da história de vida e o *habitat* onde vivem fez com que a abordagem robusta utilizada na genética de paisagem tenha resultado no desenvolvimento de estruturas de software que exigem do pesquisador de genética de populações o conhecimento e experiência no uso do programa R e das linguagens de programação Perl e Python (DYER, 2016).

Como dados genômicos podem refletir a seleção do passado?

De acordo com Clark *et al.* (2003) os atributos do DNA que refletem a seleção do passado são:

- Padrão de variação sinônima (d_S) e não-sinônima (d_N) em genes codificadores de proteínas ($\omega = d_N / d_S$)
- Espectro de frequências locais (quantos alelos/nucleotídeos polimórficos existem no gene). Excesso de alelos raros indicam seleção purificadora (alelo mutante deletério), excesso de alelos comuns indicam seleção positiva e a vantagem do heterozigoto indica seleção balanceadora
- Padrão de desequilíbrio aberrante
- Haplótipos muito longos
- Padrão de polimorfismo entre espécies fora dos níveis d
- e divergência
- Divergência excepcional entre populações

Problemas no uso de SNPs em estudos de seleção

- Viés na identificação do SNP
- Mistura populacional devido a cruzamentos ou estratificação populacional (quando consideramos uma única unidade panmítica populações subdivididas)
- História demográfica confunde o teste (ex.: crescimento demográfico muito acelerado gera mais variação nos nucleotídeos)

Conclusões

A genômica populacional vem permitindo identificar a estrutura e a diversidade genética de populações, indicar os genes responsáveis pela adaptação das espécies e solucionar problemas referentes à mudança climática, conservação e preservação de recursos genéticos, entretanto a mudança de abordagem e a grande quantidade de informações obtidas a partir do estudo direto dos genes existentes no genoma tem tornando urgente a necessidade de reciclagem de pesquisadores desta área oferecendo vários desafios no conhecimento das novas abordagens e metodologias utilizadas no sequenciamento de última geração, na bioinformática e na programação para ser capaz de ajustar programas estatísticos utilizando álgebras mais complexas de modo a atender às necessidades das investigações.

Referências

- ALLENDORF, FW; HOHENLOHE, PA; LUIKART, G. 2010. Genomics and the future of conservation genetics. **Nature Reviews Genetics**, 11: 697-709.
- BAIRD, NA; ETTER, PD; ATWOOD, TS; CURREY, MC; SHIVER, AL; LEWIS, ZA; SELKER, EU; CRESKO, WA; JOHNSON, EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. **PLoS ONE**, 3(10): e3376. www.plosone.org
- BATHIA, D; WING, R.A.; SING, K. 2013. Genotyping by sequencing, its implications and benefits. **Crop Improvement**, 40(2): 101-111.
- BRAGG, JG; SUPPLE, MA; ANDREW, RL; BOREVITZ, JO. 2015. **Genomic variation across landscapes: insights and applications**. **New Phytologist** (2015) 207: 953–967. doi: 10.1111/nph.13410
- CLARK, AG; GLANOWSKI, S; NIELSEN, R; THOMAS, PD; KEJARIWAL, A; TOOD, MA; TANENBAUM, DM; CIVELLO, D; LU, F; MURPHY, B; FERRIERA, S; WANG, G; ZHENG, X; WHITE, TJ; SNISKY, JJ; ADAMS, MD; CARGIL, M. 2003. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trio. **Science**, 302:1960-1963
- De Wit, P; Pespeni, MH; Ladner, JT; Barshis, DH; Seneca, F; Jaris, H; Overgaard Therkildsen, N; Morikawa, M; Palumbi, SR. 2012. The simple fool's guide to population genomics via RNA-seq: an introduction to high-throughput sequencing data analysis. **Molecular Ecology Resources**, 12: 1058-1067.
- DYER, RJ. Landscape and plant population genetics. In: BALKENHOL, N; CUSHMAN, SA; STORFER, AT; WAITS, LP. **Landscape genetics: concepts, methods and applications**. 2016, chapter 11.
- ELSHIRE, RJ; GLAUBITZ, JC; SUN, Q; POLAND, JA; KAWAMOTO, K; BUCKLER, ES; MITCHELL, ES. 2011. A robust, simple genotype-by-sequencing (GBS) approach for high diversity species. **PLoS ONE**, 6(5):e19379. www.plosone.org
- ELLEGREN, H. 2014. Genome sequencing and population genomics in non-model organisms. **Trends in Ecology & Evolution**, 29(1): 51-63.
- FUENTES-PARDO, AP; RUZANTE, DE. 2017. Whole-genome sequencing approaches for conservation biology: advantages, limitations and practical recommendations. **Molecular Ecology**, 26:5369-5406.
- GONZÁLEZ-MARTÍNEZ, SC; KRUTOVSKY, KV; NEALE, DB. 2006. Forest-tree population genomics and adaptive evolution. **New Phytologist**, 170: 227-238.
- HARTL, D.L.; CLARK, A.G. **Princípios de genética de populações**. 4ªed. Porto Alegre: Artmed, 2010. 659p.
- KRUTOVSKY, KV. 2006. From population genetics to population genomics of forest-tree: integrated population genomics approach. **Russian Journal of Genetics**, 42(10): 1088-1100.
- KUMAR, R; YASUNORY, I; KIMURA, S; CHITWOOD, DH; HEADLAND, LR; PENG, J; MALOOF, JN; SINHA, NR. 2012. A high-throughput method for Illumina RNA-seq library preparation. **Frontiers in Plant Science**, 3. DOI:10.3389/fpls.2012.00202
- LEBOLDOS, JM; KINZER, K; RICHARDS, J; YA, Z; YAN, C; FRIESEN, TL; BRUEGGEMAN, R. 2014. Genotyping-by-sequencing of plant-pathogenic fungi *Pyrenophora teres* and *Sphaerulina musiva* utilizing Ion Torrent sequencing technology. **Molecular plant pathology**, DOI: 10.1111/mpp.12214
- LOWRY, DB; HOBAN, S; KELLEY, JS; LOTHEROS, KE; REED, LK; ANTOLIN, MF; STORFER, A. 2017. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. **Molecular Ecology Resources**, 17: 142-152.
- POLAND, JA; Brown, PJ, Sorrells, ME, Jannink, J-L. 2012. Development of a high-density genetic map for barley and wheat using a novel two-enzyme genotype-by-sequencing approach. **PLoS ONE**, 7: e32253. www.plosone.org
- SHEBEN, A; BATLEY, J; EDWARDS, D. 2017. Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. **Plant Biotechnology Journal**, 15: 149-161.
- SONAH, H; BASTIEN, M; IQUIRA, E, TARDIVEL, A; LEGARÈ, G; BOYLE, B; NORMANDEU, E; LAROCHE, J; LAROSE, S; JEAN, M; BELZILE, F. 2013. An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. **PLoS ONE**, 8(1): e-54603. www.plosone.org

- TOOD, EV; BLACK, MA; GEMMEL, MJ. 2016. The power and promise of RNAseq in ecology and evolution. **Molecular Ecology**, 25: 1224-1241.
- VEGA, R; VÁZQUEZ-DOMÍNGUEZ, E; WHITE, T; VALENZUELA-GALVÁN, D; SEARLE, J. 2016. Population genomics applications for conservation: the case of the tropical dry forest dweller *Peromyscus melanophrys*. **Conservation Genetics**. ISSN: 1566-0621.
- VOECKEL, C; GRUENHEIT, N; LUCKHART, P. 2017. Evolutionary transcriptomics and proteomics: insights into plant adaptation. **Trends in Plant Science**. DOI: 10.1016/j.tplants.2017.03.001
- YANG, J; MIAO, C-Y; MAO, R-L, LI; Y. 2017. Landscape population genomics of Forsythia (*Forsythia suspensa*) reveal that ecological habitats determine adaptive evolution of species. **Frontiers in Plant Science**, 8: 481. Doi: 10.3389/fpls.2017.00481. www.frontiersin.org
- WEIGEL, D; NORDBORG, M. 2015. Population genomics for understanding adaptation in wild plant species. **Annual Review of Genetics**, 49:11.1-11.24
- WANG, X. **Next-generation sequencing data analysis**, 2016. Chapter 7.