



Capítulo 4

Bioinformática aplicada à agricultura

Poliana Fernanda Giachetto
Roberto Hiroshi Higa

1 Introdução

Os avanços nas áreas de tecnologia de informação e das novas tecnologias de sequenciamento têm provocado uma necessidade cada vez maior do uso da bioinformática na agricultura, principalmente com relação ao melhoramento genético vegetal e animal. Uma vasta quantidade de dados genômicos tem sido gerada a partir de diversas espécies de plantas, animais e micro-organismos, trazendo desafios no sentido de se desenvolver novas ferramentas de análise e de integração dos dados, além de soluções para se armazenar e tratar esse grande volume de dados.

Dada a importância da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) no cenário brasileiro da pesquisa agrícola, ações no sentido de prover à comunidade científica, condições para que informações biológicas possam ser obtidas a partir dos dados gerados pelas tecnologias genômicas e adequadamente utilizadas em programas de melhoramento genético, na caracterização de recursos genéticos e na biotecnologia, têm sido realizadas.

Este capítulo tem por objetivo colocar o leitor a par do estado da arte das principais aplicações da bioinformática na agricultura, particularmente àquelas relacionadas ao melhoramento genético animal e vegetal, executadas no âmbito da Embrapa.

2 A bioinformática e as novas tecnologias de sequenciamento

2.1 Sequenciamento e montagem de genomas

Um grande desafio veio à tona com o advento das novas tecnologias de sequenciamento, ou sequenciamento de nova geração, do inglês *Next Generation Sequencing* (NGS), em termos de capacidade de armazenamento e processamento de dados, assim como a necessidade do desenvolvimento de novas ferramentas de análise.

Caracterizada por um dramático aumento na quantidade dos dados gerados, acompanhado de uma substancial redução nos custos para produzi-los, as novas tecnologias de sequenciamento geram sequências bastante curtas, comparadas ao sequenciamento utilizando a tecnologia de Sanger, ou sequenciamento tradicional (METZKER, 2010; POP; SALZBERG, 2008; STREANGER; SALZBERG, 2012). Essa característica representa um grande desafio à bioinformática, princi-

palmente na montagem de genomas: o curto tamanho das sequências resulta em dificuldades na desambiguação de regiões repetitivas, resultando em montagens fragmentadas e demandando ferramentas otimizadas para resolver essa questão, diferentes daquelas até então utilizadas com dados gerados pelo sequenciamento Sanger.

Dificuldades à parte, sem dúvidas o surgimento do sequenciamento de nova geração tem transformado várias áreas da pesquisa biológica, incluindo a agricultura. Os estudos genéticos foram largamente beneficiados com o avanço na obtenção de informações genômicas que podem ser aplicadas no pré-melhoramento e melhoramento genético de espécies animais e vegetais de interesse econômico, na caracterização de recursos genéticos com vistas à prospecção gênica e na descoberta de ativos biotecnológicos. As ferramentas desenvolvidas para a análise dos dados gerados pelos sequenciadores de nova geração incluem aquelas que permitem o alinhamento das sequências produzidas contra um genoma referência ou a montagem *de novo* das sequências geradas, resultando em um genoma “montado”, a detecção de polimorfismos do tipo SNP (*Single Nucleotide Polymorphism*) e ainda outras variações estruturais como deleções, inserções, rearranjos e variações no número de cópias de trechos do genoma - as CNVs (*Copy Number Variation*), a análise de transcriptomas em grande profundidade e a análise de metagenomas, a partir de comunidades microbianas.

Apesar da importância econômica e biológica das plantas, e do custo reduzido do sequenciamento nos dias de hoje comparados à tecnologia tradicional, poucas espécies tiveram seu genoma sequenciado. Isso é devido, em grande parte, à natureza complexa de muitos genomas vegetais, como a existência de elementos repetitivos, transposons, duplicações gênicas e variações nos níveis de ploidia, que dificultam sua adequada montagem e anotação. Ainda, a presença de grandes famílias gênicas e um número abundante de pseudogenes, derivados de eventos recentes de duplicações do genoma e da atividade de transposons (SCHNABLE et al., 2009), fazem com que as montagens obtidas sejam bastante fragmentadas. Mesmo com todas essas dificuldades, esforços têm sido empregados na obtenção da sequência de genomas de espécies vegetais, incluindo plantas modelo e aquelas de interesse comercial. A Embrapa está participando, sozinha ou em parceria com outras instituições de pesquisa, do sequenciamento e da montagem de 2 genomas de plantas de interesse econômico: genoma da palma de óleo (*Elaeis guineensis* Jacq.) e genoma do feijão comum, *Phaseolus vulgaris*, em projetos liderados pela Embrapa Agroenergia e pela Embrapa Arroz e Feijão, respectivamente.

Na área animal, a Embrapa participa de um grande projeto em rede, a Rede Genômica Animal II (RGAI) (CAETANO et al., 2012), que prevê o sequenciamento e a montagem do genoma do Nelore (*Bos indicus*), raça bovina que compõe cerca de 80% do rebanho brasileiro de gado de corte, animais Gir leiteiro, Guzerá e Girolanda, representantes de raças de gado leiteiro, e das espécies de peixe nativas do Brasil, Cachara (*Pseudoplatystoma reticulatum*) e Tambaqui (*Colossoma macropomum*). Ao contrário dos genomas vegetais, a montagem de genomas animais não apresenta alguns desafios inerentes à constituição genética das plantas, como um grande número de famílias multigênicas e elevada frequência de poliploidia, mas a presença de sequências repetitivas torna essa tarefa não trivial também no caso dessas espécies. Os principais genomas de animais de produção publicados até o momento são: suíno (ARCHIBALD et al., 2010a), caprino (DONG et al., 2013), genoma da galinha (INTERNATIONAL CHICKEN GENOME SEQUENCING CONSORTIUM, 2004), genoma bovino, de um animal da raça Hereford (*Bos taurus*) (ELSIK et al. 2009) e ovino (ARCHIBALD et al., 2010b).

Além dos *drafts* publicados dos genomas acima mencionados (animais e vegetais), uma série de pragas agrícolas, assim como endo e ectoparasitas que causam doenças e perdas econômicas em lavouras e rebanhos, também tem tido seu genoma sequenciado. Alguns exemplos incluem o nematoide do nó da raiz (*Meloidogyne incognita*), que ataca um amplo espectro de plantas e tem alto poder destrutivo (ABAD et al., 2008), o ácaro *Tetranychus urticae*, considerado um dos ácaros fitófagos mais importantes do mundo (GOTOH et al., 1993) e um dos principais ácaros praga do Brasil, em função do número de espécies vegetais atacadas e dos danos a elas causadas (MORAES; FLECHTMANN, 2008) e o besouro *Tribolium castaneum*, que causa danos a grãos secos e às farinhas (TRIBOLIUM GENOME SEQUENCING CONSORTIUM, 2008). Genomas bacterianos de importantes patógenos de plantas como o da *Xylella fastidiosa* (THE XYLELLA CONSORTIUM OF THE ORGANIZATION FOR NUCLEOTIDE SEQUENCING AND ANALYSIS, 2000), *Xhantomonas axonopodis* pv. *citri* e *X. campestris* pv. *campestris*, bactérias causadoras do cancro cítrico e podridão negra em crucíferas (SILVA et al., 2002), respectivamente, além de outras espécies de *Xhantomonas*, como *X. oryzae* pv. *oryzae* (LEE et al., 2005; SALZBERG et al., 2008), que ataca culturas de arroz e *X. albilineans* (PIERETTI et al., 2009), patógeno da cana-de-açúcar, também foram decodificados. Outros exemplos ainda incluem algumas *Pseudomonas* (BUELL et al., 2003; JOARDAR et al., 2005) e a *Ralstonia solanacearum*, bactéria que infecta as raízes das plantas (SALANOUBAT et al., 2002). Além de genomas bacterianos, o de fungos patogênicos, como alguns do gênero *Phytophthora* (HAAS et al., 2009; TILLER et al., 2006), *Neurospora crassa* (GALAGAN et al., 2003) e o *Fusarium* (JEONG et al., 2013; MA et al., 2014), também foram publicados.

Importantes patógenos que atacam animais de produção também tiveram o seu genoma sequenciado, como a *Salmonella enterica* (CHIU et al., 2005), *Brucella abortus* (HALLING et al., 2005), *Mycobacterium avium* (LI et al., 2005), *Dichelobacter nodosus* (MYERS et al., 2007), *Clostridium perfringens* (MYERS et al., 2006), *Brucella suis* (PAULSEN et al., 2002) e a *Corynebacterium pseudotuberculosis* (SOARES et al., 2013) e uma série de outros mais.

O sequenciamento do genoma de patógenos animais e vegetais permite-nos desvendar os mecanismos responsáveis pela patogenicidade das espécies e também desenvolver testes moleculares para o seu diagnóstico. Uma vez conhecidos os mecanismos patogênicos, ações de controle podem ser implementadas. Nesse sentido, a Embrapa, por meio da RGAII (CAETANO et al., 2012), pretende sequenciar o genoma de 25 cepas de lentivírus de pequenos ruminantes (LVPRs), que causam consideráveis perdas econômicas em rebanhos caprinos e ovinos no Brasil. No caso da decodificação dos genomas das espécies animais e vegetais, além de sua importância no avanço das pesquisas biológicas, a existência de um genoma referência facilita a identificação dos transcritos sintetizados por esse organismo e a análise dos transcriptomas, como veremos a seguir.

2.2 Genômica comparativa

O crescente acúmulo de informações sobre espécies de interesse econômico depositadas nos bancos de dados, juntamente com o desenvolvimento de ferramentas de análise adequadas, tem permitido a realização de análises comparativas com dados de organismos modelo, facilitando a descoberta de genes envolvidos em características fenotípicas economicamente importantes.

De acordo com Sharma et al. (2014), o sequenciamento do genoma da *Arabidopsis*, planta modelo da família Brassicaceae, revolucionou nosso conhecimento no campo da biologia de plantas e

tornou-se um marco nos estudos envolvendo a genômica comparativa. Hoje, a existência de bancos de dados integrados acessados via web, contendo uma série de informações genéticas obtidas em estudos com *Arabidopsis*, além de ferramentas de análise, tem nos permitido obter avanços importantes em estudos envolvendo genômica comparativa. Dentre esses bancos, podemos citar o TAIR (SWARBRECK et al., 2008), considerado o repositório de dados e de ferramentas de análise mais importante relacionado à pesquisa com *Arabidopsis* (MOCHIDA; SHINOZAKI, 2010), o SIGnal, mantido pelo Salk Institute Genomic Analysis Laboratory¹ e o RARGE – RIKEN *Arabidopsis* Genome Encyclopedia (SAKURAI et al., 2005). Outros bancos de dados que incluem dados de *Arabidopsis* são o Brassica Genome Gateway², BRAD³, Phytozome⁴, PlantGDB⁵, EnsemblPlants⁶ e ChloroplastDB⁷. Além dos bancos de dados de *Arabidopsis*, informações genéticas e genômicas de plantas de interesse econômico também podem ser acessadas e utilizadas em estudos de genômica comparativa: SOL Genomics Network⁸, Gramene⁹, PLAZA¹⁰, GreenPhylDB¹¹, BarleyBase¹² e PlantTribes¹³. Os bancos acima citados compreendem dados de plantas, no entanto, dados referentes a animais e micro-organismos também existem e estão disponíveis publicamente, apesar de não citados aqui.

2.3 Análise de transcriptomas

O RNA-Seq, metodologia que utiliza o sequenciamento de nova geração na análise de transcriptomas, tem sido utilizado em estudos da expressão gênica de várias plantas de importância econômica, como o arroz (LU et al., 2010), milho (HANSEY et al., 2012), cevada (MAYER et al., 2012), laranja doce (XU et al., 2013) e cana-de-açúcar (CARDOSO-SILVA et al., 2014; FERREIRA et al., 2014), e animais de produção, como bovinos de corte (BALDWIN et al., 2012; LI et al., 2011; PATEL et al., 2013) e leite (MCCABE et al., 2012), suínos (ESTEVE-CODINA et al., 2011; RAMAYO-CALDAS et al., 2012), aves (PERUMBAKKAN et al., 2013), caprinos (GENG et al., 2013, LING et al., 2014) e ovinos (ZHANG et al., 2013).

A análise de transcriptomas é essencial para se conhecer os elementos funcionais de um genoma e os constituintes moleculares de células e tecidos específicos. Sua aplicação tem como objetivos específicos catalogar todos os transcritos identificados (RNAm, RNAs não codificadores e pequenos RNAs), determinar a estrutura transcricional dos genes (3' e 5' UTR, sítio de iniciação da transcrição, padrão de processamento dos éxons e modificações pós-transcripcionais) e quan-

¹ Disponível em: <<http://signal.salk.edu/>>.

² Disponível em: <<http://brassica.bbsrc.ac.uk/>>.

³ Disponível em: <<http://brassicadb.org/brad/>>.

⁴ Disponível em: <<http://www.phytozome.net/>>.

⁵ Disponível em: <<http://www.plantgdb.org/>>.

⁶ Disponível em: <<http://plants.ensembl.org/index.html>>.

⁷ Disponível em: <<http://chloroplast.cbio.psu.edu/>>.

⁸ Disponível em: <<http://solgenomics.net/>>.

⁹ Disponível em: <<http://www.gramene.org/>>.

¹⁰ Disponível em: <<http://bioinformatics.psb.ugent.be/plaza/>>.

¹¹ Disponível em: <<http://www.greenphylo.org/>>.

¹² Disponível em: <<http://www.plexdb.org/plex.php?database=Barley>>.

¹³ Disponível em: <<http://fgp.huck.psu.edu/tribedb/>>.

tificar as alterações nos níveis de expressão de cada transcrito ao longo do desenvolvimento e sob diferentes condições (WANG et al., 2009). Na agricultura, essa última aplicação permite, por meio do monitoramento das mudanças no nível de expressão gênica, identificar vias biológicas alteradas em situações onde uma perturbação externa é inserida no sistema, como por exemplo, a infecção por um patógeno, infestação por um parasita, alterações na dieta, restrição hídrica e outros tipos de estresse.

A análise de transcriptomas por meio da técnica de RNA-Seq revolucionou os estudos na área, por apresentar várias vantagens quando comparada a outras utilizadas até então para a análise de transcriptomas, como os microarranjos (MARTIN; WANG, 2011; WANG et al., 2009). O RNA-Seq fornece uma medida mais precisa do nível de transcritos, assim como suas isoformas, e promove uma acurada quantificação da expressão diferencial, capaz de gerar novos conhecimentos sobre mecanismos moleculares de uma dada característica de interesse, em adição à identificação de novos transcritos. O RNA-Seq permite também a identificação sistemática de SNPs em regiões transcritas, os quais podem ser utilizados na detecção de expressão alelo-específica e também como marcadores moleculares. Como exemplo dessa aplicação, citamos o trabalho de Cardoso-Silva et al. (2014) que aplicaram a tecnologia de RNA-Seq a 6 genótipos de cana-de-açúcar, contrastantes para o teor de sacarose. Os autores identificaram SNPs exclusivos de cada genótipo, os quais possuem uma alta probabilidade de associação com as características de interesse econômico particulares a cada um deles. Estratégias como essa, que têm por objetivo a busca de SNPs em regiões codificadoras, por meio do sequenciamento em larga escala do transcriptoma de tecidos alvo de cultivares ou variedades contrastantes para características de interesse, têm sido largamente utilizadas em projetos em andamento na Embrapa, envolvendo culturas como o trigo, soja, milho, café e cana-de-açúcar, e as características avaliadas incluem aumento de produtividade, resistência a doenças e pragas, tolerância à seca e melhoria da qualidade do produto final, entre outros.

Uma outra abordagem que utiliza a tecnologia de RNA-Seq, certamente a mais conhecida e empregada, foi recentemente utilizada por Nishiyama Junior, et al. (2014), que sequenciaram e compararam o transcriptoma de 3 cultivares comerciais de cana-de-açúcar e dos 2 genótipos parentais (*S. officinarum* e *S. spontaneum*), identificando diferenças na expressão gênica que possibilitarão, segundo os autores, associar transcritos espécie-específicos à produção de biomassa e outras características importantes. Os transcritos específicos identificados nesse tipo de estudo podem ser utilizados como marcadores moleculares, e serem empregados como ferramentas em programas de seleção assistida por marcadores, prestando um grande auxílio a avanços no melhoramento genético da cultura avaliada.

Importantes descobertas na área animal têm sido feitas utilizando-se a tecnologia de RNA-Seq. Genes relacionados a características de qualidade da carne em suínos (JUNG et al., 2012; RAMAYO-CALDAS et al., 2012) e gado de corte (LEE et al., 2013; SHENG et al., 2014) têm sido identificados. Essa característica tem sido alvo de estudos da Embrapa, que lidera projetos utilizando RNA-Seq com o objetivo de identificar os mecanismos que determinam a maciez da carne bovina, um atributo bastante apreciado pelos consumidores e que não é típico de animais *Bos indicus*, espécie a qual pertence o Nelore, principal raça de gado de corte criada do Brasil, como já foi citado. Esses animais apresentam tipicamente carne menos macia do que os animais da espécie *Bos taurus* e trabalhos têm sido conduzidos pela Embrapa Pecuária Sul, pela Embrapa Pecuária Sudeste e pela Embrapa Gado de Corte no sentido de comparar as duas subespécies

ou mesmo animais extremos para essa característica dentro de cada raça e identificar os genes responsáveis pela maciez.

A busca por genes e/ou mecanismos envolvidos na resistência de animais a doenças e parasitas responsáveis por perdas econômicas no setor também tem sido objeto de estudos na Embrapa. A comparação do transcriptoma de caprinos e ovinos resistentes e susceptíveis ao parasita *Haemonchus contortus*, principal helminto que ataca os rebanhos no Brasil (COSTA et al., 2000), assim como o estudo do transcriptoma do próprio parasita, estão em andamento em projetos conduzidos na Embrapa Caprinos e Ovinos. Outro parasita que causa sérios prejuízos econômicos ao Brasil e ao mundo, o carrapato bovino *Rhipicephalus microplus*, tem sido tema de muitas pesquisas na Embrapa, que buscam formas de controle por meio da compreensão dos mecanismos ativados pelo hospedeiro e/ou pelo ácaro no momento da interação e também na identificação de genes diferencialmente expressos em animais resistentes e susceptíveis. A Embrapa Pecuária Sul tem desenvolvido importantes projetos na área, juntamente com a Embrapa Pecuária Sudeste e a Embrapa Gado de Corte. Estudos acerca da identificação de genes relacionados a problemas ósseos em frangos de corte, que causam dificuldades de locomoção nas aves e consequente redução do crescimento, também têm sido investigados por meio de RNA-Seq pela Embrapa Suínos e Aves.

2.4 O estudo de comunidades microbianas

As pesquisas envolvendo o estudo de microbiomas também foram alavancadas com o surgimento das novas tecnologias de sequenciamento, as quais trouxeram inúmeras vantagens em relação aos métodos até então existentes, baseados em cultura de micro-organismos *in vitro*. A utilização de NGS tem possibilitado o conhecimento da composição da comunidade microbiana, por meio do sequenciamento de regiões hipervariáveis do gene 16S RNAr de bactérias e região Internal Transcribed Spacer (ITS) de fungos, e também de uma quantidade bastante grande de genes presentes nas comunidades microbianas, por meio do sequenciamento do DNA total delas extraído (metagenomas). A metagenômica tem se mostrado mais eficiente do que os métodos culturais principalmente pelo maior poder de identificação das cepas e por tornar possível a caracterização e o monitoramento de alterações na dinâmica de uma comunidade microbiana como um todo, e não apenas de indivíduos em particular. O conhecimento dos membros e dos genes presentes em um microbioma permite-nos inferir o papel de cada um na comunidade. No entanto, além de identificar quem está presente no microbioma, saber o que fazem é igualmente importante e nos possibilita entender a dinâmica das relações entre os micro-organismos e também a relação entre microbiota e hospedeiro. Essa informação pode ser obtida por meio do sequenciamento dos transcritos expressos pelos membros da comunidade microbiana e obtenção do metatranscriptoma. A técnica de RNA-Seq, já discutida neste capítulo, tem sido utilizada com esse objetivo.

No caso dos estudos genômicos, duas abordagens de estudo de comunidades microbianas podem ser adotadas, utilizando-se as novas tecnologias de sequenciamento: a) análise de amplicons do gene 16S RNAr, onde uma ou mais regiões hipervariáveis do gene 16S RNAr de membros individuais da comunidade é amplificada, por meio de reações em cadeia de polimerase, do inglês Polymerase Chain Reaction (PCR), utilizando iniciadores complementares a regiões conservadas do gene (que flanqueiam as regiões hipervariáveis) e sequenciada, com posterior comparação das sequências com aquelas depositadas em bancos de dados existentes, classificação e cálculo da abundância relativa e b) análise do metagenoma, por meio do sequenciamento em larga escala

do DNA extraído de toda a comunidade microbiana, o que permite a identificação dos genes que compõem o microbioma e também da diversidade da população, embora de maneira menos precisa do que a obtida com a classe anterior (OVIEDO-RONDÓN, 2009). O sequenciamento de amplicons do gene 16S RNAr é a abordagem mais acessível e utilizada em estudos que visam à identificação de membros de uma comunidade microbiana, sendo considerado um bom marcador para se caracterizar a composição filogenética de uma amostra, identificar novas espécies ou mesmo grupos filogenéticos desconhecidos (GABOR et al., 2007). O fato de vários organismos possuírem sequências do 16S RNAr idênticas, mas funções diferentes, não nos permite utilizar essa abordagem para inferir adequadamente as funções dos membros de uma comunidade microbiana. Essa limitação da técnica pode ser superada com a análise de metagenomas e/ou metatranscriptomas.

Dada a importância dos micro-organismos na agricultura, são vários os exemplos de aplicação da análise de sequências de 16S RNAr e de metagenomas nos estudos de comunidades microbianas. Em recente revisão publicada por Kao-Kniffin et al. (2013), os autores citaram estudos onde a abordagem metagenômica foi utilizada na identificação de genes e compostos com ação inseticida, e genes de resistência a herbicidas, para utilização no controle de pragas em sistemas agrícolas. De acordo com os autores, o isolamento do DNA diretamente das amostras ambientais, sem passar pelo processo de cultivo em meio de cultura, aumenta muito a diversidade de micro-organismos recuperada das amostras e, conseqüentemente, a possibilidade de identificação de compostos de interesse.

Por se tratar do ambiente que abriga a maior biodiversidade do planeta (MOCALI; BENEDETTI, 2010), o estudo de comunidades microbianas do solo apresenta um potencial imenso para a agricultura: os micro-organismos do solo desempenham um papel crítico na regulação da fertilidade do solo, saúde das plantas e na ciclagem do carbono, nitrogênio e outros nutrientes (FIERER et al., 2012a). Nesse sentido, vários estudos acerca da caracterização da microbiota de diferentes tipos de solos têm sido realizados (FIERER et al., 2007; FIERER et al., 2012b; GROS et al., 2006; SCHLOSS; HANDELSMAN, 2006). O efeito das práticas agrícolas sobre as comunidades microbianas do solo também tem sido alvo de estudos, que já identificaram o pH e a disponibilidade de água como os principais fatores abióticos que afetam as populações (KÖBERL et al., 2011), além da quantidade de N depositada no solo (FIERER et al., 2012b).

Acredita-se que as plantas recrutam micro-organismos benéficos do solo, a partir de suas rizosferas, para neutralizar o ataque de patógenos (COOK et al., 1995). Ainda, existem certos tipos de solos, conhecidos como supressivos, que abrigam comunidades microbianas que impedem que os agentes patógenos causem danos às plantas. Assim, a caracterização de microbiomas da rizosfera de plantas sob condições diversas, e aquela presente em solos supressivos, é uma outra área que tem se beneficiado dos estudos metagenômicos e com imenso potencial para gerar resultados importantes para a agricultura, como a identificação de novas bactérias fixadoras de N₂, supressoras de patógenos e promotoras do crescimento.

No mesmo contexto, na área animal estudos têm sido realizados com o objetivo de se identificar micro-organismos benéficos, que aumentem a produtividade e o bem estar animal. Nesse sentido, a caracterização de microbiomas do trato gastrointestinal tem sido foco de bastante atenção, uma vez que, além do seu papel na digestão, a microbiota do trato gastrointestinal exerce influência também na proteção contra patógenos, detoxificação e modulação do sistema imune do hospedeiro.

deiro (AMIT-ROMACH et al., 2004; OVIEDO-RONDÓN, 2009). A identificação de micro-organismos com baixa produção de metano em ruminantes também tem sido objeto de estudos, uma vez que a emissão entérica de metano tem sido apontada como responsável por cerca de 15% da emissão global de metano, de acordo com o Painel Intergovernamental sobre Mudanças Climáticas (IPCC) (ROSS et al., 2013; SHI et al., 2014).

A Embrapa tem liderado e atuado em parceria em vários projetos que utilizam a metagenômica na caracterização de microbiomas e para identificar micro-organismos com potencial biotecnológico. Dentre eles podemos citar o estudo do microbioma da rizosfera e folhas de plantas resistentes à seca e a patógenos, com participação da Embrapa Meio Ambiente, além do estudo de microbiomas do trato intestinal e sistema digestório de frangos de corte, da glândula mamária e rúmen de ovinos e sistema respiratório de suínos, como parte dos projetos em desenvolvimento na RGAII.

3 Bioinformática e a utilização de marcadores moleculares no melhoramento genético animal e vegetal

3.1 Uso de marcadores moleculares do tipo SNP em genética animal aplicada

Polimorfismos de base única (SNPs) são variações no genoma em que um único nucleotídeo - A, T, C, G, difere entre membros de uma população (FOULKES, 2009). Em geral, assume-se que SNPs são bialélicos, indicando a existência de duas possíveis bases no correspondente *locus*, e que cada base deve ter uma frequência mínima de 1% na população. Embora sua distribuição não seja homogênea ao longo do genoma (ex: eles ocorrem mais frequentemente em regiões não codificantes que em regiões codificantes), existem milhões de polimorfismos deste tipo distribuídos ao longo do genoma de espécies de interesse zootécnico, o que faz dele um marcador genômico muito atraente.

A partir da montagem dos genomas de espécies de interesse zootécnico e o subsequente mapeamento de haplótipos (ex: genoma bovino (BOVINE HAPMAP CONSORTIUM, 2009; ELSIK, 2009), foram desenvolvidas tecnologias para descoberta e genotipagem em massa de centenas de milhares de marcadores SNP a um custo que favorece sua utilização em estudos de associação e mapeamento genético, seleção genômica, detecção de doenças genéticas e/ou polimorfismos associados a características de produção, ensaios diagnósticos para confirmação de paternidade, identificação individual (rastreadabilidade) (CAETANO, 2009).

O desenvolvimento de todas essas aplicações envolve uma série de etapas experimentais e de análise de dados, exigindo infraestrutura e especialistas em diferentes áreas como biologia molecular, zootecnia, veterinária, estatística e ciência da computação. Na sequência, é apresentada uma visão geral das aplicações em seleção genômica e estudos de associação genômica ampla e as iniciativas em andamento na Embrapa.

3.2 Seleção Genômica (GS)

O objetivo do melhoramento genético é alcançar melhores níveis de produção, produtividade e/ou qualidade do produto final, em sintonia com o sistema de produção e as exigências do mercado (ROSA et al., 2013). Para isso, diversas características expressas nos animais da população em

seleção e relacionadas com o objetivo do programa de melhoramento genético, são monitoradas, como por exemplo: eficiência reprodutiva, eficiência alimentar, qualidade da carcaça e da carne para gado de corte, produção de leite, percentagem de gordura e proteínas no leite para gado de leite, tamanho da ninhada, peso ao desmame para suínos, entre outras (BOURDON, 2000).

Os valores mensurados dessas características, denominadas fenótipos (P), para um animal dependem basicamente de sua genética (G: genótipo), do ambiente no qual ele está exposto (E: ambiente) e da expressão dos genótipos quando expostos a diferentes condições ambientais. De forma sucinta, essa relação pode ser expressa como: $P = G + E + G \cdot E$ (ROSA et al., 2013).

Já as avaliações genéticas têm por objetivo identificar os animais em uma população sob seleção que sejam geneticamente superiores para que, se usados na reprodução, transmitam aos seus descendentes sua superioridade, dessa forma, alterando o desempenho médio da população (MARTINS, 2013). Tradicionalmente, o processo para determinar o valor genético dos animais (*Breeding Value* - BV) depende de três diferentes fontes de informação: seu desempenho, o desempenho de seus ancestrais (pedigree) e de seus descendentes (progênie).

Meuwissen et al. (2001) propuseram a utilização de marcadores SNP, que cobrem densamente todo o genoma, para avaliação genética de animais, um processo conhecido como seleção genômica. Neste caso, o valor genético do animal é denominado valor genético genômico, do inglês *Genomic Breeding Value* (GBV) e compreende o somatório dos efeitos de todos os marcadores SNP utilizados. A utilização dessa tecnologia em países desenvolvidos tem proporcionado ganhos genéticos aos programas de melhoramento genético, devido ao aumento das acurácias dos valores genéticos para características de baixa herdabilidade e redução do intervalo entre gerações para características de difícil mensuração ou aquelas obtidas tardiamente, como eficiência alimentar e dados de carcaça.

A Embrapa, atualmente, participa de diferentes esforços multi-institucionais para inclusão da tecnologia de seleção genômica em diferentes programas de melhoramento genético animal em que atua. Um desses esforços, desenvolvido pela Embrapa Pecuária Sul em associação com a Conexão Delta G e o Gensys Consultores Associados, foca na seleção de animais com características de resistência a carrapatos para as raças Hereford e Braford. O carrapato (*Rhipicephalus microplus*) é um dos principais problemas de saúde na pecuária de corte, implicando em custos de produção adicionais com acaricidas e no tratamento de doenças infecciosas transmitidas pelo parasita, além da queda de performance dos animais. Estima-se que a perda econômica com carrapatos no Brasil seja de dois bilhões de dólares anuais (GRISI et al., 2002). Desde 2012, a Embrapa Pecuária Sul e seus parceiros publicam o sumário de touros para resistência a carrapatos para as raças Hereford e Braford (CARDOSO et al., 2013).

Dentre os outros esforços da Embrapa, visando à inclusão da tecnologia de seleção genômica em programas de melhoramento genético animal, destacam-se: a) Genhol (NAPOLIS, 2012), que tem entre seus objetivos específicos desenvolver estudos, processos e validação de modelos de predição de valores genômicos para a raça holandesa no Brasil; b) Genomilk (SILVA, 2010), que tem entre seus objetivos específicos adequar, avaliar e/ou comparar diferentes métodos estatísticos para implementar a seleção genômica e integração de dados genômicos nos sistemas de avaliação genética e testes de progênie conduzidos pela Embrapa Gado de Leite; c) arranjo MaxiBife (SILVA, 2013), que prevê a inclusão das tecnologias para avaliação genômica no programa de melhoramento genético de gado de corte Geneplus-Embrapa; d) rede MP1 RGA II (CAETANO,

2012), que possui um projeto componente (Projeto Componente 3) focado na inclusão de seleção genômica em programas de melhoramento genético de bovinos, ovinos e caprinos.

3.3 Estudos de associação genômica amplo (GWAS)

O termo Genome-Wide Association Studies (GWAS) deve-se à ampla cobertura dos marcadores SNP ao longo do genoma, tornando-os atrativos para análises de associação genótipo:fenótipo envolvendo efeito poligênico. Estes estudos têm como objetivo identificar padrões de polimorfismos que variam sistematicamente entre indivíduos com diferentes valores de expressão de um determinado fenótipo (BALDING, 2006). Para realização de GWAS são necessários três elementos:

- 1) Uma grande quantidade de amostras;
- 2) marcadores genéticos que cubram grande parte do genoma;
- 3) métodos analíticos poderosos o suficiente para identificar sem viés a associação entre os marcadores e os fenótipos analisados (CANTOR et al., 2010).

Os procedimentos computacionais para análise de dados de experimentos de GWAS envolvem uma série de passos (ZIEGLER et al., 2008), existindo diversos métodos para identificação dos SNP associados com o fenótipo estudado. Dentre estes, os mais comumente utilizados constituem-se em testes univariados por SNP, com correção para múltiplos testes (ZIEGLER et al., 2008). Abordagens envolvendo a análise simultânea de múltiplos marcadores baseiam-se em técnicas de regressão, paramétricas ou não paramétricas, com seleção ou encolhimento (*shrinkage*) de variáveis (MOORE et al., 2010; ZIEGLER et al., 2008).

A Embrapa tem desenvolvido diversos projetos envolvendo análise de GWAS, sendo que alguns já possuem resultados publicados. Mokry et al. (2013) analisaram um conjunto de dados 400 animais da raça Canchim genotipados com o BovineHD BeadChip (ILLUMINA INC., SAN DIEGO, CA), visando identificar SNP associados com espessura de gordura. Foi encontrado um conjunto de SNPs capaz de explicar aproximadamente 50% da variância do valor genético derregredido (dEBV) para espessura de gordura e um pequeno conjunto de 5 SNPs capaz de explicar 34% da dEBV. Foram encontrados diversos Quantitative Trait Loci (QTL) relacionados com gordura na vizinhança desses SNPs, bem como genes envolvidos no metabolismo de lipídeos. Tizioto et al. (2013) conduziram um estudo de GWAS com 800 animais da raça Nelore genotipados com o BovineHD BeadChip (Illumina Inc., San Diego, CA), visando identificar SNPs associados com diferentes fenótipos relacionados com qualidade de carne: força de cisalhamento Warner-Bratzler, medida em diferentes tempos de maturação, espessura de gordura, área de olho do lombo, parâmetros de cor da carne e gordura, capacidade de retenção de água, perdas no cozimento e pH do músculo. As regiões genômicas na vizinhança dos QTL encontrados e as vias das quais esses genes fazem parte diferem daqueles identificados em raças taurinas. Espera-se que esses resultados subsidiem futuros estudos de mapeamento de QTL e o desenvolvimento de modelos para predição de mérito genético para qualidade de carne para a raça Nelore.

Diversos outros projetos encontram-se em andamento na Embrapa, fazendo uso de GWAS para identificar genes e regiões gênicas associadas com características fenotípicas de interesse zootécnico. Na RGAI (CAETANO, 2012), o projeto componente 4 está conduzindo três experimentos de GWAS: (i) estudos de associação para características de presença/ausência de chifres e criptorquidia em 250 ovinos da raça Morava Nova, genotipados com o Ovine SNP50Beadchip

(Illumina Inc., San Diego, CA); (ii) estudos de associação para 85 características avaliadas para 1.000 aves da população de Referência TT, representativa da melhor linha paterna de corte do programa de melhoramento genético da Embrapa Suínos e Aves, genotipados com o Chicken SNP60BeadChip (Illumina Inc., San Diego, CA); estudos de associação para a característica de presença de chifres e batoques em rebanhos Nelore Mocho BRGN, pertencentes à Embrapa Cerrados e de rebanhos de importantes criadores de Nelore mocho, como a Guaporé S/A, utilizando 350 animais genotipados com o Bovine SNP50Beadchip (Illumina Inc., San Diego, CA).

4 Laboratório multiusuário de bioinformática (LMB)

Em função da crescente demanda por poder computacional e competência multidisciplinar para lidar com os grandes volumes de dados, algoritmos e ferramentas de análise diversos, a Embrapa, por meio de uma decisão estratégica, inaugurou em outubro de 2011, o LMB, com sede na Embrapa Informática Agropecuária.

O LMB, que tem como missão viabilizar soluções de bioinformática para projetos de pesquisa, desenvolvimento e inovação da Embrapa, em um ambiente colaborativo¹⁴, busca incorporar e tornar disponíveis a comunidade científica novas tecnologias para armazenamento, processamento e análise de grandes volumes de dados. O LMB atua ainda na elaboração e execução de projetos e planos de ação que necessitam de ferramentas especializadas e computação de alto desempenho, na disponibilização de procedimentos computacionais para a montagem de genomas, análise de metagenomas e de transcriptomas, e na análise de dados de marcadores moleculares e de expressão gênica, desenvolvimento e implantação de recursos computacionais, criação e administração de bancos e bases de dados. A capacitação técnica, por meio da realização de cursos e treinamentos em ferramentas usadas para análise de dados, como os softwares Galaxy e Generic Genome Browse (Gbrowse), também faz parte das atuações do LMB.

5 Considerações finais

Dada a importância do melhoramento genético na agricultura sustentável e na segurança alimentar no nosso país e do mundo, o uso de dados genômicos para o melhoramento de espécies animais e vegetais de interesse econômico é de fundamental importância, e faz parte das ações da Embrapa que visam garantir a demanda da produção de alimentos frente ao crescimento da população e também aos cenários futuros de mudanças climáticas. O desenvolvimento e otimização de ferramentas de bioinformática que permitam a análise dos dados gerados pelas tecnologias genômicas em crescente avanço, assim como sua disponibilização à comunidade científica, é preocupação constante da empresa, que trabalha sempre no sentido de prover as melhores soluções para o aproveitamento das informações geradas, no sentido de contribuir para o avanço no setor e manutenção do Brasil como um país de excelência em pesquisa agrícola.

¹⁴ Disponível em: <<http://www.lmb.cnptia.embrapa.br>>.

6 Referências

ABAD, P. et al. Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. **Nature Biotechnology**, New York, v. 26, n. 8, p. 909-915, Aug. 2008.

AMIT-ROMACH, E.; SKLAN, D.; UNI, Z. Microflora ecology of the chicken intestine using 16S ribosomal DNA primers. **Poultry Science**, Champaign, v. 83, n. 7, p.1093-1098, July 2004.

ARCHIBALD, A. L.; BOLUND, L.; CHURCHER, C.; FREDHOLM, M.; GROENEN, M. A.; HARLIZIUS, B.; LEE, K. T.; MILAN, D.; ROGERS, J.; ROTHSCCHILD, M. F.; UENISHI, H.; WANG, J.; SCHOOK, L. B. Pig genome sequence - analysis and publication strategy. **BMC Genomics**, London, v. 11, p. 438, July 2010. DOI: 10.1186/1471-2164-11-438.

ARCHIBALD, A. L.; COCKETT, N. E.; DALRYMPLE, B. P.; FARAUT, T.; KIJAS, J. W.; MADDOX, J. F.; MCEWAN, J. C.; HUTTON ODDY, V.; RAADSMA, H. W.; WADE, C.; WANG, J.; WANG, W.; XUN, X. The sheep genome reference sequence: a work in progress. **Animal Genetics**, Oxford, v. 41, n. 5, p. 449-453, Oct. 2010b. DOI: 10.1111/j.13652052.2010.02100.x.

BALDING, D. J. A tutorial on statistical methods for population association studies. **Nature Review Genetics**, London, v. 7, n. 10, p. 781-791. Oct. 2006. DOI:10.1038/nrg1916.

BALDWIN, R. L.; LI, R. W.; LI, C.; THOMSON, J. M.; BEQUETTE, B. J. Characterization of the longissimus lumborum transcriptome response to adding propionate to the diet of growing Angus beef steers. **Physiological Genomics**, Bethesda, v. 44, n. 10, p. 543-550, May 2012.

BOURDON, R. M. **Understanding animal breeding**. 2nd ed. Upper Saddle River: Prentice Hall, 2000. 538 p. il.

BOVINE HAPMAP CONSORTIUM. Genome-wide survey of snp variation uncovers the genetic structure of cattle breeds. **Science**, New York, v. 342, n. 5926, p. 528-532, Apr. 2009. DOI: 10.1126/science.1167936.

BUELL, C. R. et al. The complete genome sequence of the Arabidopsis and tomato pathogen *Pseudomonas syringae* pv. Tomato DC3000. **Proceedings of the Academy of Sciences of the United States of America**, Washington, US, v. 100, n.18, p.10181-10186, Sept. 2003.

CAETANO, A. R. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **Revista Brasileira de Zootecnia**, Viçosa, MG, v. 38, p. 64-71, 2009.

CAETANO, A. R. **RGAI - Rede nacional para o desenvolvimento e adaptação de estratégias genômicas inovadoras aplicadas ao melhoramento, conservação e produção animal**. Projeto SEG: 01.11.07.002.00.00. Embrapa, 2012.

CANTOR, R. M.; LNAGE, K.; SINSHEIMER, J. S. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. **American Journal of Human Genetics**, Chicago, v. 86, n. 1, p. 6-22. Jan. 2010. DOI: 10.1016/j.ajhg.2009.11.017.

CARDOSO, F. F.; YOKOO, M. J. I.; GULIAS-GOMES, C. C.; SOLLERO, B. P.; COSTA, R. F. da; ROSO, V. M.; BRITO, F. V.; CAETANO, A. R.; AGUILAR, I. **Avaliação genômica para resistência ao carrapato de touros Hereford e Braford**. Bagé: Embrapa Pecuária Sul, 2013. 40 p. (Embrapa Pecuária Sul. Documentos, 133). Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/100508/1/DT1332013.pdf>>. Acesso em: 22 set. 2014.

CARDOSO-SILVA, C. B. et al. De novo assembly and transcriptome analysis of contrasting sugarcane varieties. **PLoS One**, San Francisco, v. 9, n. 2, e88462, Feb. 2014. DOI:10.1371/journal.pone.0088462.

CHIU, C. H. et al. The genome sequence of *Salmonella enterica* serovar choleraesuis, a highly invasive and resistant zoonotic pathogen. **Nucleic Acids Research**, v. 33, n. 5, p. 1690-1698, Mar. 21, 2005. DOI:10.1093/nar/gki297.

COOK, R. J.; COOK, R. J.; THOMASHOW, L. S.; WELLER, D. M.; FUJIMOTO, D.; MAZZOLA, M.; BANGERA, G.; KIM, D. S. Molecular mechanisms of defense root disease by rhizobacteria against root disease. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, US, v. 92, p. 4197-4201, 1995.

DONG, Y.; XIE, M.; JIANG, Y.; XIAO, N.; DU, X.; ZHANG, W.; TOSSER-KLOPP, G.; WANG, J.; YANG, S.; LIANG, J.; CHEN, W.; CHEN, J.; ZENG, P.; HOU, Y.; BIAN, C.; PAN, S.; LI, Y.; LIU, X.; WANG, W.; SERVIN, B.; SAYRE, B.; ZHU, B.; SWEENEY, D. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). **Nature Biotechnology**, New York, v. 31, p. 135-41, Dec. 2013. DOI:10.1038/nbt.2478.

ELSIK, C. G.; TELLAM, R. S.; WORLEY, K. C. The genome sequence of taurine cattle: a window to ruminant biology and evolution. **Science**, New York, v. 342, n. 5926, p. 522-528, Apr. 2009. DOI: 10.1126/science.1169588.

ESTEVE-CODINA, A.; KOFLER, R.; PALMIERI, N.; BUSSOTTI, G.; NOTREDAME, C.; PÉREZ-ENCISO, M. Exploring the gonad transcriptome of two extreme male pigs with RNA-seq. **BMC Genomics**, London, v. 12, n. 552, p. 1-14, 2011. DOI:10.1186/1471-2164-12-552.

FIERER, N.; BRADFORD, M. A.; JACKSON, R. B. Toward an ecological classification of soil bacteria. **Ecology**, Tempe, v. 88, n. 6, p. 1354-1364, June 2007.

FIERER, N.; LAUBER, C. L.; RAMIREZ, K. S.; ZANEVELD, J.; BRADFORD, M. A.; KNIGHT, R. Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. **The ISME Journal**, New York, v. 6, n. 5, p. 1007-1017, May 2012a. DOI: 10.1038/ismej.2011.159.

FIERER, N.; LEF, J. W.; ADAMS, B. J.; NIELSEN, U. N.; BATES, S. T.; LAUBER, C. L.; OWENS, S.; GILBERT, J. A.; WALL, D. H.; CAPORASO, G. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, US, v. 109, n. 52, p. 21390-21395, Nov. 2012b. DOI: 10.1073/pnas.1215210110.

FOULKES, A. S. **Applied statistical genetics with R: for population based association studies**. New York: Springer Verlag, 2009. 252 p. DOI: 10.1007/978-0-387-89554-3. Ebooks.

GABOR, E. et al. Updating the metagenomics toolbox. **Biotechnology Journal**, Weinheim, v. 2, n. 2, p. 201-206, Feb. 2007. DOI:10.1002/biot.200600250.

GALAGAN, J. E. et al. The genome sequence of the filamentous fungus *Neurospora crassa*. **Nature**, London, v. 422, n. 6934, p. 859-868, Apr. 2003.

GENG, R.; YUAN, C.; CHEN, Y. Exploring differentially expressed genes by RNA-Seq in cashmere goat (*Capra hircus*) skin during hair follicle development and cycling. **PLoS One**, San Francisco, v. 8, n. 4, e62704, 2013. DOI:10.1371/journal.pone.0062704.

GOTOH, T.; BRUIN, J.; SABELIS, M.W.; MENKEN, S.B.J. Host race formation in *Tetranychus urticae*: genetic differentiation, host plant preference, and mate choice in a tomato and a cucumber strain. **Entomologia Experimentalis et Applicata**, Dordrecht, v. 68, 171-178, 1993.

GRISI, L.; MASSARD, L. C.; BORJA, G. E. M.; PEREIRA, J. B. Impacto econômico das principais ectoparasitoses em bovinos no Brasil. **A Hora Veterinária**, Porto Alegre, v. 125, p. 8-10, Nov. 2002.

GROS, R.; MONROZIER, L. J.; FAIVRE, P. Does disturbance and restoration of alpine grassland soils affect the genetic structure and diversity of bacterial and N₂-fixing populations? **Environmental Microbiology**, Oxford, v. 8, n. 11, p. 1889-1901, Nov. 2006.

HAAS, B. J. et al. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. **Nature**, London, v. 461, p. 393-398, Sept. 2009. DOI:10.1038/nature08358.

HALLING, S. M. et al. Completion of the genome sequence of *Brucella abortus* and comparison to the highly similar genomes of *Brucella melitensis* and *Brucella suis*. **Journal of Bacteriology**, v. 187, n. 8, p. 2715-2726, Apr. 2005. DOI:10.1128/JB.187.8.2715-2726.2005.

INTERNATIONAL CHICKEN GENOME SEQUENCING CONSORTIUM. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. **Nature**, London, v. 432, n. 7018, p. 695-716, Dec. 2004. DOI: 10.1038/nature03154.

JEONG, H.; LEE, S.; CHOI, G. J.; LEE, T.; YUN, S.-H. Draft genome sequence of *Fusarium fujikuroi* B14, the causal agent of the bakanae disease of rice. **Genome Announcements**, Washington, D.C. v. 1, n. 1, p.1-2, Jan./Feb. 2013. DOI: 10.1128/genomeA.00035-13.

JOARDAR, V. et al. Whole-genome sequence analysis of *Pseudomonas syringae* pv. *Phaseolicola* 1448A reveals divergence among pathovars in genes involved in virulence and transposition. **Journal of Bacteriology**, Washington, US, v. 187, n. 18, p. 6488-6498, Sept. 2005.

JUNG, W. Y. et al. RNA-Seq approach for genetic improvement of meat quality in pig and evolutionary insight into the substrate specificity of animal carbonyl reductases. **PLoS One**, San Francisco, v. 7, n. 9, e42198, Sept. 2012. DOI:10.1371/journal.pone.0042198.

KAO-KNIFFIN, A. J.; CARVER, S. M.; DITOMMASO, A.; KAO-KNIFFIN, J. Advancing weed management strategies using metagenomic techniques advancing weed management strategies using metagenomic techniques. **Weed Science**, Champaign, v. 61, n. 2, p. 171-184, 2013. DOI: <http://dx.doi.org/10.1614/WS-D-12-00114.1>.

KÖBERL, M.; MÜLLER, H.; RAMADAN, E. M.; BERG, G. Desert farming benefits from microbial potential in arid soils and promotes diversity and plant health. **PLoS One**, San Francisco, v. 6, n. 9, p. e24452, Sept. 2011. DOI: 10.1371/journal.pone.0024452.

LEE, B.-M. et al. The genome sequence of *Xanthomonas oryzae* pathovar *oryzae* KACC10331, the bacterial blight pathogen of rice. **Nucleic Acids Research**, v. 33, n. 2, p. 577-586, 2005. DOI:10.1093/nar/gki206.

LEE, H. J. et al. Comparative transcriptome analysis of adipose tissues reveals that ECM receptor interaction is involved in the depot-specific adipogenesis in cattle. **PLoS One**, San Francisco, v. 8, n. 6, p. e66267, June, 2013. DOI: 10.1371/journal.pone.0066267.

LI, R. W.; RINALDI, M.; CAPUCO, A. V. Characterization of the abomasal transcriptome for mechanisms of resistance to gastrointestinal nematodes in cattle. **Veterinary Research**, Les Ulis, v. 30, n. 42, p. 114, Nov. 2011. DOI: 10.1186/1297-9716-42-114.

LING, Y. H.; XIANG, H.; LI, Y. S.; LIU, Y.; ZHANG, Y. H.; ZHANG, Z. J.; DING, J. P.; ZHANG, X. R. Exploring differentially expressed genes in the ovaries of uniparous and multiparous goats using the RNA-Seq (Quantification) method. **Gene**, Amsterdam, v. 550, n. 1, p. 148-153, Oct. 2014. DOI: 10.1016/j.gene.2014.08.008.

MA, L.-J.; SHEA, T.; YOUNG, S.; ZENG, Q.; CORBY, H. C. Genome sequence of *Fusarium oxysporum* f. sp. *Melonis* strain NRRL 26406, a fungus causing wilt disease on melon. **Genome Announcements**, Washington, D.C. v. 2, n. 4, e00730-14, July 2014. DOI:10.1128/genomeA.00730-14.

MARTIN, J. A.; WANG, Z. Next-generation transcriptome assembly. **Nature Review Genetics**, London, v. 12, p. 671-82, Oct. 2011. DOI:10.1038/nrg3068.

MARTINS, E. N. Avaliação Genética: dos dados às DEP's. In: ROSA, A. N.; MARTINS, E. N.; MENEZES, G. R. O.; SILVA, L. O. C. (Ed.). **Melhoramento Genético Aplicado em Gado de Corte**: Programa Geneplus-Embrapa. Brasília, DF: Embrapa; Campo Grande, MS: Embrapa Gado de Corte, 2013. 241 p. il.

MCCABE, M.; WATERS, S.; MORRIS, D.; KENNY, D.; LYNN, CREEVEY, C. RNA-seq analysis of differential gene expression in liver from lactating dairy cows divergent in negative energy balance. **BMC Genomics**, London, v. 13, n. 193, May 2012. DOI:10.1186/1471-2164-13-193.

METZKER, M. L. Sequencing technologies - the next generation. **Nature Review Genetics**, London, v. 11, n. 1, p. 31-46, Jan. 2010. DOI: 10.1038/nrg2626.

MEUWISSEN, T. H. E.; HAYES, B. J.; GOODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, London, v. 157, n. 4, p. 1819-1829, Apr. 2001.

MOKRY, F. B.; HIGA, R. H.; MUDADU, M. A.; LIMA, A. O.; MEIRELLES, S. L. C.; SILVA, M. V. G. B.; CARDOSO, F. F.; OLIVEIRA, M. M.; URBINATI, I.; NICIURA, S. C. M.; TULLIO, R. R.; ALENCAR, M. M.; REGITANO, L. C. A. Genome-wide association study for backfat thickness in Canchim beef cattle using Random Forest approach. **BMC Genetics**, London, v. 14, n. 47, 2013. DOI:10.1186/1471-2156-14-47.

MOORE, J. H.; ASSELBERGS, F. W.; WILLIAMS, S. M. Bioinformatics challenges for genome-wide association studies. **Bioinformatics**, Oxford, v. 26, n. 4, p. 445-455, Jan. 2010. DOI: 10.1093/bioinformatics/btp713.

MORAES, G. J.; FLECHTMANN, C. H.W. **Manual de acarologia**: acarologia básica e ácaros de plantas cultivadas no Brasil. Ribeirão Preto: Holos, 2008. 308 p.

MYERS, G. S. et al. Genome sequence and identification of candidate vaccine antigens from the animal pathogen *Dichelobacter nodosus*. **Nature Biotechnology**, v. 25, n. 5, p. 569-575, Apr. 29, 2007. DOI:10.1038/nbt1302.

MYERS, G. S. et al. Skewed genomic variability in strains of the toxigenic bacterial pathogen, *Clostridium perfringens*. **Genome Research**, v. 16, n. 8, p. 1031-1040, Aug. 2006. DOI: 10.1101/gr.5238106.

NAPOLIS, C. **Genhol**: desenvolvimento e integração de novos processos metodológicos no sistema de avaliação genética de bovinos leiteiros e sua filiação ao Interbull. Brasília, DF, Embrapa, 2012. Projeto SEG: 02.12.02.001.00.00.

NISHIYAMA JUNIOR, M. Y.; FERREIRA, S. S.; TANG, P.-Z.; BECKER, S.; SOUZA, G. M. PÖRTNER-TALIANA, A. Full-length enriched cDNA libraries and ORFeome analysis of sugarcane hybrid and ancestor genotypes. **PLoS One**, San Francisco, v. 9, n. 9, e107351, Sept. 2014. DOI: 10.1371/journal.pone.0107351.

OVIDO-RONDÓN, E. O. Molecular methods to evaluate effects of feed additives and nutrients in poultry gut microflora. **Revista Brasileira de Zootecnia**, Viçosa, MG, v. 38, p. 209-225, Oct. 2009. Número especial. DOI: <http://dx.doi.org/10.1590/S1516-5982009001300022>.

PATEL, A. K.; BHATT, V. D.; TRIPATHI, A. K.; SAJNANI, M. R.; JAKHESARA, S. J.; KORINGA, P. G.; JOSHI, C. G. Identification of novel splice variants in horn cancer by RNA-Seq analysis in Zebu cattle. **Genomics**, London, v. 101, n. 1, p. 57-63, Jan. 2013. DOI: 10.1016/j.ygeno.2012.10.001.

PAULSEN, I. T. et al. The *Brucella suis* genome reveals fundamental similarities between animal and plant pathogens and symbionts. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, US, v. 99, n. 20, Oct. 2002.

PERUMBAKKAM, S.; MUIR, W. M.; BLACK-PYRKOSZ, A.; OKIMOTO, R.; CHENG, H. H. Comparison and contrast of genes and biological pathways responding to Marek's disease virus infection using allele-specific expression and differential expression in broiler and layer chickens. **BMC Genomics**, London, v. 14, n. 64, Jan. 2013. DOI:10.1186/1471-2164-14-64.

PIERETTI, I. et al. The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. **BMC Genomics**, London, v. 10, n. 616, 2009. DOI: 10.1186/1471-2164-10-616.

POP, M.; SALZBERG, S. L. Bioinformatics challenges of new sequencing technology. **Trends in Genetics**, Cambridge, v. 24, n. 3, p. 142-149, Mar. 2008. DOI: 10.1016/j.tig.2007.12.006.

RAMAYO-CALDAS, Y.; MACH, N.; ESTEVE-CONDINA, A.; COROMINAS, J.; CASTELLÓ, A.; BALLESTER, M.; ESTELLÉ, J.; IBÁÑEZ-ESCRICHE, N.; FERNÁNDEZ, A. I.; PÉREZ-INCISO, M.; FOLCH, J. M. Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition. **BMC Genomics**, London, v. 13, n. 547, Oct. 2012. DOI:10.1186/1471-2164-13-547.

ROSA, A. N.; MENEZES, G. R. O.; EGITO, A. A. Recursos Genéticos e Estratégias de Melhoramento. In: ROSA, A. N.; MARTINS, E. N.; MENEZES, G. R. O.; SILVA, L. O. C. (Ed.). **Melhoramento Genético Aplicado em Gado de Corte**: Programa Geneplus-Embrapa. Brasília, DF: Embrapa; Campo Grande, MS: Embrapa Gado de Corte, 2013. 241 p. il.

- ROSS, E. M.; MOATE, P. J.; MARETT, L.; COCKS, B. G.; HAYES, B. J. Investigating the effect of two methane-mitigating diets on the rumen microbiome using massively parallel sequencing. **Journal of Dairy Science**, Champaign, v. 96, n. 9, p. 6030-6046, Sept. 2013.
- SALANOUBAT, M. et al. Genome sequence of the plant pathogen *Ralstonia solanacearum*. **Nature**, London, v. 415, n. 6871, p. 497-502, January, 2002.
- SALZBERG, S. L. et al. Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *Oryzae* PXO99A. **BMC Genomics**, London, v. 9, p. 204, May 2008. DOI: 10.1186/1471-2164-9-204.
- SCHLOSS, P. D.; HANDELSMAN, J. Toward a census of bacteria in soil. **PLoS Computational Biology**, San Francisco, v. 2, n. 7, e92, July 2006. DOI: 10.1126/science.1178534.
- SCHNABLE, P. S. et al. The B73 maize genome: complexity, diversity, and dynamics. **Science**, New York, v. 326, n. 5956, p. 1112-1115, Nov. 2009. DOI:10.1126/science.1178534.
- SHARMA, A.; LI, X.; LIM, Y. P. Comparative genomics of Brassicaceae crops. **Breeding Science**, New York, v. 64, n. 13, p. 3-13, May 2014. DOI:10.1270/jsbbs.64.3.
- SHENG, X.; NI, H.; LIU, Y.; LI, J.; ZHANG, L.; GUO, Y. RNA-seq analysis of bovine intramuscular, subcutaneous and perirenal adipose tissues. **Molecular Biology Reports**, Dordrecht, v. 41, n. 3, Mar. 2014. DOI: 10.1007/s11033-013-3010-8.
- SHI, W. et al. Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. **Genome Research**, Woodbury, v. 24, p. 1517-1525, June 2014. DOI: 0.1101/gr.168245.113.
- SILVA, A. C. R. et al. Comparison of the genomes of two *Xanthomonas pathogens* with differing host specificities. **Nature**, London, v. 417, n. 6887, p. 459-463, May 2002.
- SILVA, L. O. C. **MaxiBife** - Desenvolvimento e integração de conhecimentos de genética molecular e quantitativa visando maximizar os ganhos genéticos para produção sustentável de carne bovina: arranjo SEG. Brasília, DF: Embrapa, 2013.
- SILVA, M. V. B. S. **Genomil - Seleção Genômica em Raças Bovinas Leiteiras no Brasil**. Brasília, DF: Embrapa, 2010. Projeto SEG: 02.09.07.008.00.00.
- SOARES, S. C. et al. The pan-genome of the animal pathogen *Corynebacterium pseudotuberculosis* reveals differences in genome plasticity between the biovar ovis and equi strains. **PLoS One**, San Francisco, v. 8, n. 1, e53818, Jan. 2013. DOI:10.1371/journal.pone.0053818.
- SWARBRECK, D.; WILKS, C.; LAMESCH, P.; BERARDIN, T. Z.; GARCIA-HERNANDEZ, M.; FOERSTER, H.; LI, D.; MEYER, T.; MULLER, R.; PLOETZ, L.; RADENBAUGH, A.; SINGH, S.; SWING, V.; TISSIER, C.; ZHANG, P.; HUALLA, E. The Arabidopsis Information Resource (TAIR): gene structure and function annotation. **Nucleic Acids Research**, London, v. 36, p. 1009-1014, Jan. 2008.
- THE XYLELLA CONSORTIUM OF THE ORGANIZATION FOR NUCLEOTIDE SEQUENCING AND ANALYSIS. The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* consortium of the Organization for Nucleotide Sequencing and Analysis. **Nature**, London, v. 13, n. 406, n. 6792, p. 151-159, July 2000.
- TIZIOTO, P.C.; DECKER, J.E.; TAYLOR, J.F.; SHNABEL, R.D.; MUDADU, M.M.; SILVA, F.L.; MOURÃO, G.B.; COUTINHO, L.L.; THOLON, P.; SONSTEGARD, T.S.; ROSA, A.N.; ALENCAR, M.M.; TULLIO, R.R.; MEDEIROS, S.R.; NASSU, R.T.; FEIJÓ, G.L.D.; SILVA, L.O.C.; TORRES, R.A.; SIQUEIRA, F.; HIGA, R.H.; REGITANO, L.C.A. A genome scan for meat quality traits in Nelore beef cattle. **Physiological Genomics**, Bethesda, v. 45, n. 21, p. 1012-1020, 2013.
- TRIBOLIUM GENOME SEQUENCING CONSORTIUM The genome of the model beetle and pest *Tribolium castaneum*. **Nature**, London, v. 452, p. 949-955, Apr. 24, 2008. DOI:10.1038/nature06784.

- WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Review Genetics**, London, v. 10, n. 1, p. 57-63, Jan. 2009. DOI: 10.1038/nrg2484.
- ZIEGLER, A.; KÖNIG, I. R.; THOMPSON, J. R. Biostatistical aspects of genome-wide association studies. **Biometrical Journal**, Berlin, v. 50, n. 1, p. 8-28, Feb. 2008. DOI: 10.1002/bimj.200710398.
- XU, Q. et al. The draft genome of sweet orange (*Citrus sinensis*). **Nature Genetics**, New York, v. 45, p. 59-66, Nov. 2013. DOI:10.1038/ng.2472.