

4

Genômica aplicada à genética e melhoramento de *Eucalyptus* na Embrapa: 25 anos de avanços e as perspectivas para o futuro

Dario Grattapaglia

Introdução

Eucalyptus L'Hér. (Myrtaceae) representa atualmente o gênero de espécies florestais folhosas (*hardwoods*) mais amplamente plantado no mundo. As assim chamadas *big nine*, ou seja, as nove principais espécies plantadas e seus híbridos, todas elas pertencentes ao subgênero *Symphyomyrtus* (*E. grandis*, *E. urophylla*, *E. globulus*, *E. camaldulensis*, *E. nitens*, *E. pellita*, *E. saligna*, *E. tereticornis*, *E. dunnii*) constituem mais de 95% das florestas plantadas de eucalipto do mundo (Harwood, 2011). Crescimento rápido e adaptabilidade a uma ampla diversidade de regiões tropicais e subtropicais, combinados com propriedades versáteis de madeira para energia, produtos sólidos de madeira, celulose e papel têm contribuído para a posição de destaque do eucalipto na atual silvicultura mundial. *Eucalyptus grandis* (Maiden) Hill, *E. urophylla* S.T. Blake, *E. camaldulensis* Dehn. e seus híbridos são as principais espécies plantadas em regiões tropicais, enquanto *E. globulus* Labill e *E. nitens* H.Deane & Maiden são as espécies mais importantes em regiões temperadas (Grattapaglia et al., 2012). A enorme diversidade genética encontrada nas procedências dentro das espécies e as oportunidades de explorar a complementaridade e heterose em híbridos, a partir de pools gênicos contrastantes que evoluíram separadamente no centro de origem do gênero, tem sido uma grande vantagem para o desenvolvimento de estoques genéticos de alta qualidade por seleção direcional no melhoramento genético (Grattapaglia; Kirst, 2008).

É consenso hoje que o melhoramento genético tornou-se um elemento competitivo chave das operações industriais de base florestal intensiva com espécies de eucalipto em todo o mundo, fornecendo clones e sementes geneticamente superiores, que aumentam consideravelmente o valor econômico das florestas plantadas. Independentemente da espécie alvo e objetivos industriais, os programas de melhoramento genético enfrentam, entretanto, um desafio comum: a longa duração de um ciclo de melhoramento, que pode levar vários anos ou mesmo décadas. Espécies de eucalipto têm ciclos de vida relativamente longos e as árvores tornam-se reprodutivamente ativas somente após alguns anos. O progresso e o sucesso dos programas de melhoramento são, portanto, fortemente dependentes do tempo necessário para completar um ciclo de melhoramento. Isso pode durar vários anos a décadas, dependendo da região geográfica, da biologia da espécie, da idade na qual os fenótipos podem ser medidos com precisão e da forma de implantação de material melhorado, seja sementes ou clones. Além disso, as incertezas associadas à execução de um programa de melhoramento ao longo de décadas podem ser altas. O intervalo de tempo prolongado entre o investimento em melhoramento e a efetiva utilização do material genético selecionado na floresta e, em seguida, na fábrica, torna a prática do melhoramento um empreendimento suscetível a mudanças nas demandas de mercado, nos objetivos de negócios e nas políticas de gestão, além de alterações climáticas e de acesso a áreas para a expansão do plantio florestal. O desafio posto pelos longos ciclos de um programa de melhoramento

resultou, ao longo dos anos, em esforços substanciais para entender as correlações juvenil-adulto, para características de expressão tardia (Borrvalho et al., 1992; Eldridge et al., 1993), desenvolver maneiras de acelerar a recombinação por indução artificial de floração (Griffin et al., 1993; Hasan; Reid, 1995) e praticar a seleção precoce de características juvenis (Blake; Bevilacqua, 1990; Bouvet et al., 2004).

Com a expansão da biologia molecular de plantas a partir do final dos anos 80 e das tecnologias genômicas a partir de meados dos anos 90, novas perspectivas se abriram para atacar os desafios existentes em duas grandes áreas das ciências florestais: (i) acelerar e aumentar a efetividade do melhoramento genético florestal, visando o aumento da produtividade, qualidade e proteção das florestas plantadas e (ii) o manejo de populações naturais de espécies florestais, pelo estudo detalhado da distribuição e evolução da diversidade neutra e adaptativa em um contexto ecológico e evolutivo. A biologia molecular aplicada às espécies florestais tem lançado mão de um amplo arsenal de tecnologias moleculares, que podem ser agrupadas em dois grupos principais: (i) o estudo, caracterização e manipulação de genes individuais por técnicas de biologia molecular, o que representaria a biotecnologia florestal no sentido mais restrito do termo, e (ii) a descrição, caracterização e análise da estrutura, variação e expressão de sequências de DNA e RNA com técnicas de alto desempenho o que constituiria, de forma geral, a genômica.

Enquanto a biotecnologia florestal tem predominantemente o objetivo de realizar intervenção direta no genoma, via técnicas de transgenia e edição gênica, o foco da genômica florestal tem sido, por um lado, a estimativa dos níveis e distribuição de variabilidade genética global do genoma em indivíduos e populações utilizando marcadores moleculares e, por outro, o entendimento dos mecanismos genéticos, bioquímicos, moleculares e celulares envolvidos principalmente no crescimento, formação da madeira e resposta de resistência ou tolerância da árvore a estresses bióticos e abióticos. Nesta segunda vertente, as tecnologias genômicas buscam, em suas aplicações mais avançadas, a detecção e compreensão das relações complexas entre variabilidade genética, em nível de sequência de DNA, e a variabilidade fenotípica observada para características complexas de interesse florestal. A genômica contrasta com a abordagem de biologia molecular tradicional baseada no estudo de um ou poucos genes de cada vez, ao analisar, em paralelo, milhares de sequências, genes ou mesmo genomas inteiros de um ou mais indivíduos ou espécies. Para um maior aprofundamento no tema geral da genômica florestal, foco técnico principal deste capítulo, aqui voltado para a genômica do eucalipto, vale destacar algumas revisões que abordam o tópico de uma forma ampla contemplando todas as principais espécies florestais plantadas (Grattapaglia et al., 2009; Neale; Kremer, 2011; Harfouche et al., 2012; Plomion et al., 2016; Tuskan et al., 2018) e mais especificamente as espécies de *Eucalyptus* (Poke et al., 2005; Grattapaglia et al., 2012).

Considerando que espécies florestais como o eucalipto não se qualificam como um sistema modelo para o estudo da biologia molecular de plantas, o grande impulso na evolução da genômica do eucalipto nos últimos 25 anos foi motivado predominantemente pelo potencial de utilização de marcadores moleculares, na otimização de programas de melhoramento genético, e pela investigação da relação complexa entre variação no genoma e variação nos fenótipos relacionados com a qualidade da madeira, crescimento e resistência a doenças. Foi na perspectiva de facilitar a prática do melhoramento que as tecnologias de análise genômica com marcadores moleculares despertaram a atenção dos geneticistas de eucalipto no início da década de 1990. Além dos marcadores moleculares permitirem a quantificação e o gerenciamento da variabilidade genética de bancos de germoplasma e populações de melhoramento, identificar indivíduos, determinar parentesco e reconstruir pedigrees, a possibilidade de praticar seleção assistida por marcadores em idade precoce das árvores foi proposta como um meio de acelerar os ciclos de melhoramento, aumentar a intensidade de seleção, reduzir o esforço de testes de campo e melhorar a precisão de seleção, principalmente para características de baixa herdabilidade e expressão tardia (Grattapaglia et al., 1992; Williams; Neale, 1992). No entanto, o potencial da seleção assistida por marcadores para espécies florestais foi logo questionado com base no estado de equilíbrio de ligação de populações de melhoramento, considerando o tamanho efetivo elevado e histórico recente de domesticação, além da variabilidade da expressão dos locos controladores de características quantitativas (*QTLs* – *quantitative trait loci*) entre populações e ambientes distintos (Strauss et al., 1992), críticas estas que ao longo dos anos se provaram procedentes porém contornáveis ao se utilizar a abordagem de seleção genômica (Grattapaglia; Resende, 2011).

O papel da pesquisa pública e privada brasileira na geração de conhecimento e desenvolvimento de produtos e processos nas áreas de silvicultura, manejo, conservação, melhoramento, proteção e tecnologia da madeira do eucalipto é amplamente reconhecido internacionalmente. A Embrapa, em colaboração com diversas universidades e empresas privadas do País e exterior, no desenvolvimento da genômica florestal de espécies de *Eucalyptus* a partir de 1994 e vem mantendo uma posição de destaque até hoje, seja na geração de recursos e conhecimento fundamentais para a pesquisa genômica de *Eucalyptus*, bem como na implementação inovadora de tecnologias genômicas na prática operacional do melhoramento genético do eucalipto. Neste histórico de contribuições, deve ser destacado o papel que teve o Projeto Genolyptus, Rede Brasileira de Pesquisa do Genoma do Eucalipto, concebido, formatado e liderado pela Embrapa entre 2001 e 2008. Como descrito a seguir, o Genolyptus foi o principal “motor” que deu início à evolução da pesquisa e aplicação da genômica no melhoramento e produção do eucalipto no país.

Este capítulo apresenta um breve histórico das contribuições da Embrapa na área das ciências genômicas aplicadas ao eucalipto, realizadas ao longo dos últimos

25 anos (Figura 1). Após uma breve descrição do histórico e impactos do Projeto Genolyptus, a discussão é apresentada de forma organizada em três grandes temas. O primeiro envolve o desenvolvimento de recursos genômicos desde as plataformas de marcadores moleculares, passando pela geração de bancos de dados de sequências, até o trabalho de sequenciamento do genoma de referência de *Eucalyptus grandis*. O segundo apresenta aplicações de ferramentas de análise genômica que não incluem a tentativa de correlacionar variação na sequência do DNA com características fenotípicas, mas que tem por objetivo o estudo da variabilidade e estrutura genética de populações, a identificação individual e determinação e quantificação de parentesco. Essas aplicações, hoje rotina em programas de melhoramento e propagação clonal de eucalipto em diversas empresas no Brasil e no mundo, têm tido um impacto direto e quantificável na otimização dos vários processos que dependem da correta certificação de identidade e parentesco, reconstrução de pedigrees e gerenciamento de variabilidade genética em populações de melhoramento. O terceiro tema aborda aplicações tecnicamente mais desafiadoras que vêm buscando conectar a variação de sequência no DNA com a variação de características quantitativas complexas, por meio de mapeamento de QTLs, mapeamento de associação e, mais recentemente, predição e seleção genômica, com o intuito de auxiliar diretamente a seleção mais rápida e precisa de árvores superiores em programas de melhoramento. Em vista do foco da obra da qual este capítulo faz parte, o tratamento da biotecnologia e genômica florestal não pretende ser exaustivo. Para aqueles que desejam um maior aprofundamento em cada tema abordado, são indicadas referências úteis ao longo da discussão.

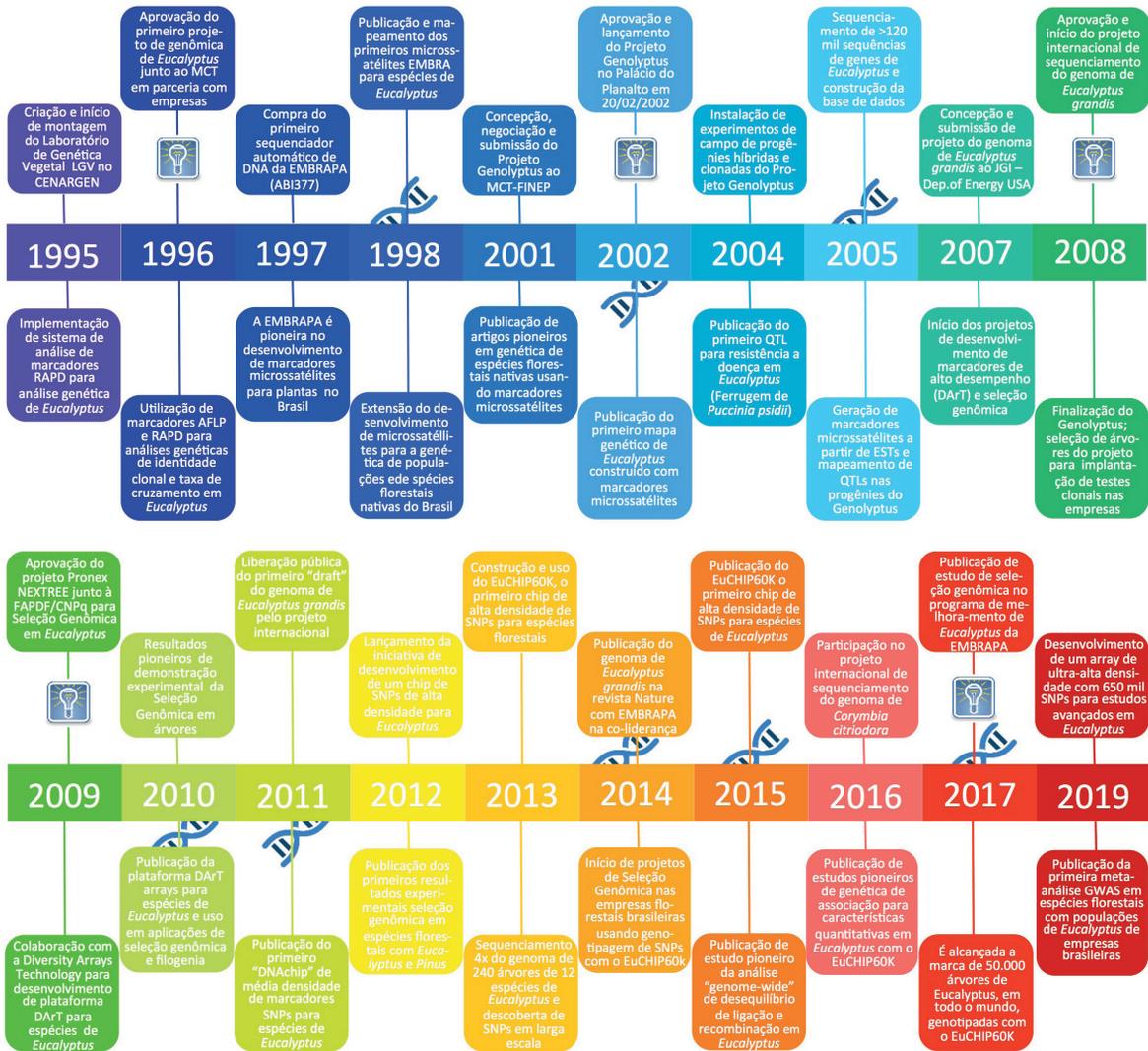


Figura 1. Linha do tempo das contribuições da Embrapa na área de genômica aplicada a espécies de *Eucalyptus* ao longo dos últimos 25 anos, através do trabalho do Laboratório de Genética Vegetal e todos os seus colaboradores nacionais e internacionais. O ícone representando uma fita de DNA indica contribuições mais significativas (milestones) de geração de recursos genômicos de impacto internacional. O ícone representando uma lâmpada indica os momentos nos quais projetos captados externamente viabilizaram em grande parte a realização dos trabalhos descritos. Detalhes sobre cada uma das contribuições são fornecidas ao longo do texto.

O Projeto Genolyptus: a Embrapa lidera uma ampla rede de pesquisa genômica público-privada na área florestal

Dois projetos de pesquisa de genômica do eucalipto foram iniciados paralelamente no Brasil entre 2001 e 2002: o Projeto Forests em São Paulo, e o Projeto Nacional Genolyptus - Rede Brasileira de Pesquisa do Genoma do Eucalipto (Figura 2). O Projeto Forests, financiado pela Fapesp e por quatro empresas paulistas de base florestal, teve como objetivo específico a geração de um banco de sequências de genes expressos de *Eucalyptus grandis* segundo a lógica dos projetos genoma que a Fapesp apoiava na época. Houve um avanço relevante com os trabalhos de anotação *in silico* dos dados de sequenciamento gerados, tema de um fascículo especial da revista brasileira *Genetics and Molecular Biology* (Carrer, 2005). Entretanto, do ponto de vista da efetiva internalização da genômica no setor florestal, o *Forests* teve uma contribuição limitada, seja pelo fato dos dados de sequenciamento terem permanecido inacessíveis à comunidade científica até hoje, por questões contratuais do projeto, bem como, e principalmente, por não associar a geração de dados genômicos a experimentos de campo que permitissem a busca de um impacto mais direto no melhoramento genético do eucalipto.

Indo além de um típico projeto de sequenciamento de porções do genoma, o Projeto Genolyptus, por outro lado, se fundamentou em uma estratégia de interconexão entre as diferentes tecnologias genômicas disponíveis na época com uma intensa experimentação de campo, otimização de métodos de avaliação de qualidade da madeira e a solução de desafios na área de patologia florestal (Grattapaglia, 2004; Grattapaglia et al., 2004a). Como primeiro passo do projeto, entre 2002 e 2003 foi realizado um importante trabalho de interação entre as empresas participantes na geração e instalação de uma ampla rede experimental de campo. Este trabalho, fundamental para o sucesso do projeto, demandou a mobilização de dezenas de pessoas e recursos materiais e logísticos nas várias empresas. A rede experimental instalada cobriu cinco regiões do País, desde o Rio Grande do Sul até o Pará, amostrando uma ampla variabilidade ambiental onde o eucalipto é plantado. Com o ingresso de uma empresa portuguesa no projeto, novas populações e material genético de *E. globulus* foram agregados ao esforço experimental, consolidando assim uma base experimental florestal inédita em nível mundial especificamente voltada para a investigação genômica do eucalipto. Estes experimentos de campo não serviram apenas como base fundamental para os estudos de mapeamento genético e expressão gênica, mas também forneceram produtos na forma de estoque genético elite para programas de produção e melhoramento florestal. Ao final do projeto, passados cerca de 6 a 7 anos de experimentação de campo, mais de 100 árvores foram selecionadas pelo seu

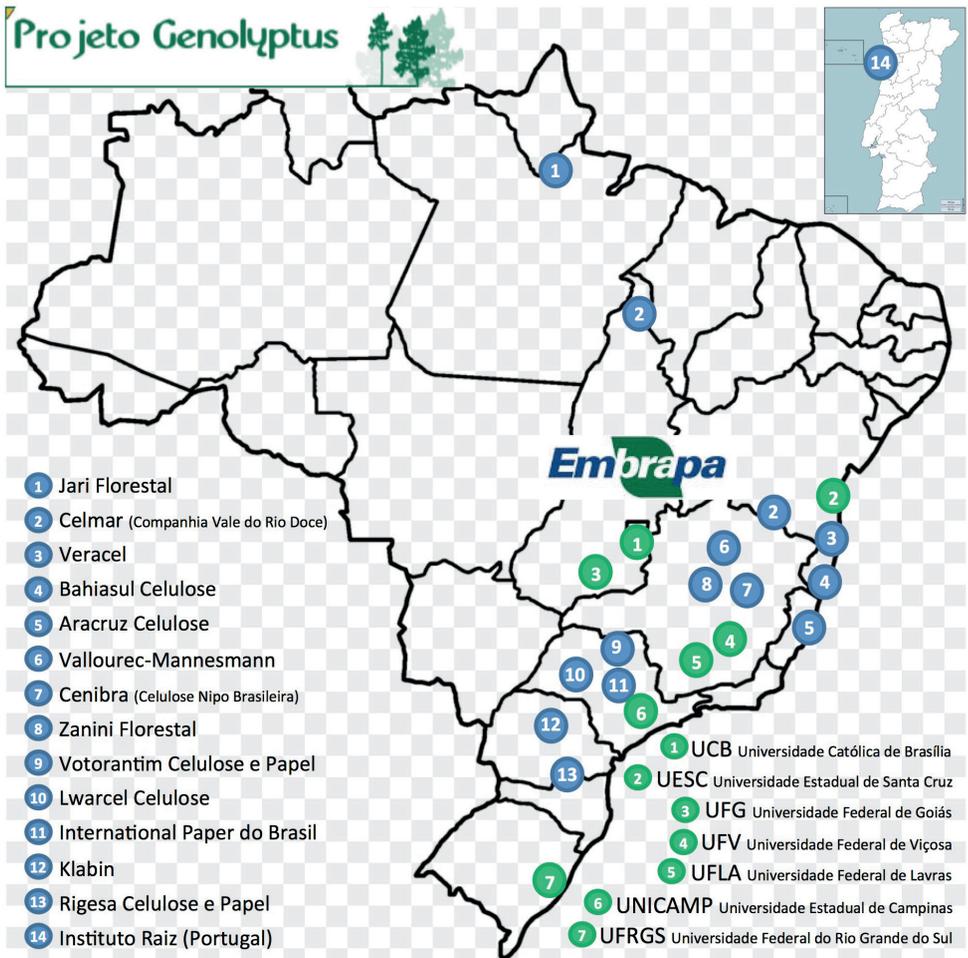


Figura 2. Configuração do projeto de parceria pública-privada da Rede Brasileira de Pesquisa do Genoma do Eucalipto – Projeto Genolyptus (2002-2008) liderada pela Embrapa, envolvendo 14 empresas de base florestal e sete universidades brasileiras localizadas no mapa. Várias das empresas de base florestal na época hoje ou não existem mais pois foram compradas por outras empresas, ou mudaram de nome no forte movimento de consolidação do setor que ocorreu nos últimos dez anos.

excelente crescimento, forma e propriedades da madeira nos diferentes ambientes. Estas árvores, várias delas de constituição híbrida de interesse de várias empresas, foram propagadas e distribuídas entre os participantes e instaladas em diversos testes clonais ou utilizadas como genitores em programas de hibridação interespecífica.

O Projeto Genolyptus se baseou em uma parceria entre o Governo Federal por meio da Financiadora de Estudos e Projetos (Finep) agência do Ministério da Ciência e Tecnologia, o setor acadêmico representado por sete universidades, dois centros da Embrapa (Embrapa Recursos Genéticos e Biotecnologia e Embrapa Florestas) e o setor privado com, na época, 14 empresas florestais, 13 brasileiras e uma portuguesa (Figura 2). O projeto consumiu cerca de R\$ 8 milhões, correspondendo, na época, a 2 milhões de dólares americanos, isso sem contar os custos de salários dos pesquisadores envolvidos das várias instituições. A Finep participou com R\$ 4,5 milhões do Fundo Setorial Verde Amarelo, o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) com cerca de R\$ 500 mil em bolsas e as empresas com R\$ 2,5 milhões em recursos financeiros diretos e outros R\$ 500 mil em recursos materiais e logísticos.

A formação desta rede representou um posicionamento estratégico importante do País e das empresas brasileiras do setor na área de biotecnologia florestal em uma efetiva parceria público-privada. Dificilmente cada uma das empresas individualmente teria fôlego tecnológico para sustentar um esforço concentrado na área de ciências genômicas na época, que se caracterizava pela novidade constante, complexidade técnica, necessidade de equipes especializadas e uma imprevisibilidade do retorno efetivo no curto e médio prazo do investimento realizado. Concebido dentro de um inovador espírito pré-competitivo, ao longo dos seus oito anos de duração, o Projeto Genolyptus, envolveu mais de 40 pesquisadores das várias universidades, centros da Embrapa e empresas privadas, incluindo a importante participação de dezenas de estudantes de graduação e pós-graduação como veículo essencial de formação de massa crítica, disseminação e continuidade da área de desenvolvimento e uso da genômica aplicada ao eucalipto pelos anos subsequentes. De fato, diversos profissionais que atuaram como estudantes no Projeto Genolyptus, foram, mais tarde contratados pelas várias empresas e atuam até hoje na interface entre genômica e melhoramento.

O foco central do Projeto Genolyptus foi a geração de uma plataforma integrada de recursos experimentais de campo e bases de dados genômicos para descobrir, mapear e buscar o entendimento da variação nos genes e regiões genômicas subjacentes às várias características de importância econômica em *Eucalyptus*, com ênfase no crescimento, formação da madeira e resistência a doenças. O Projeto Genolyptus gerou uma coleção de mais de 130 mil sequências expressas de quatro espécies de *Eucalyptus* incluindo *E. grandis*, *E. globulus*, *E. pellita* e *E. urophylla*, buscando assim capitalizar na variação interespecífica do gênero. Esta coleção cobriu os estimados 22 mil genes

do eucalipto na época e permitiu a identificação de vários genes de lignificação e síntese de celulose, entre outros, e o estudo das diferenças entre as espécies. Esta base de dados foi utilizada para a realização de um experimento de análise de expressão gênica em microarranjo com 390.000 sondas representando os 22 mil genes identificados. Os resultados deste experimento forneceram informações inéditas sobre o desempenho deste primeiro microarranjo de todo o transcriptoma de *Eucalyptus* em nível mundial, bem como uma primeira visão da variação nos níveis de expressão gênica entre e dentro de espécies de *Eucalyptus* que possuem características contrastantes em termos de propriedades físicas e químicas da madeira. A partir da mineração desta base de dados foram também desenvolvidos e geneticamente mapeados 540 marcadores microssatélites, os quais permitiram a construção de múltiplos mapas genéticos e a execução de experimentos de mapeamento genético de QTLs.

Com base na avaliação dos experimentos de campo, avanços foram realizados na construção de diversos mapas genéticos e localização de QTLs controlando características de crescimento volumétrico, densidade básica da madeira, teores de hemiceluloses e lignina, florescimento precoce e resistência a doenças fúngicas e bacterianas. A interação estreita entre os cientistas das instituições públicas e privadas permitiu grandes avanços na área de tecnologia da madeira e de fitopatologia. Foram internalizadas novas metodologias de avaliação precisa e rápida das propriedades químicas da madeira usando espectrometria do infravermelho próximo. Árvores que apresentavam resistência às quatro doenças bacterianas ou fúngicas alvo do projeto foram identificadas e o mapeamento genético em progênies segregando para estas resistências permitiu a localização de QTLs de maior efeito controlando estas resistências. Os experimentos de mapeamento genético de QTL desenvolvidos no âmbito do Genolyptus foram fundamentais para que se fizesse uma avaliação crítica desta abordagem, até então canônica, para a tentativa de seleção assistida, abrindo espaço, mais tarde para a mudança de paradigma neste tema com a pesquisa e adoção da seleção genômica no melhoramento do eucalipto (Grattapaglia; Resende, 2011; Resende et al., 2012).

Além do banco de sequências, mapas genéticos e marcadores moleculares, o Genolyptus gerou um importante recurso genômico na forma de duas bibliotecas de fragmentos longos de DNA do eucalipto, clonados em Bacterial Artificial Chromosomes (BAC), sendo uma delas a partir do DNA da árvore cujo genoma foi, em seguida, sequenciado integralmente. Estes fragmentos longos de DNA foram sequenciados visando a clonagem de genes envolvidos na síntese de lignina e celulose (Paiva et al., 2011), mas se tornaram efetivamente essenciais para a pesquisa internacional ao dar suporte ao sequenciamento e montagem completa do genoma do eucalipto. Durante um congresso internacional em 2005, a comunidade científica trabalhando com genética e genômica de *Eucalyptus* formatou o que viria a ser um projeto de sequenciamento do genoma do eucalipto, submetido em 2007 ao centro de sequenciamento (Joint Genome Institute) do Departamento de Energia dos EUA.

Após concorrer com mais de 120 projetos do mundo todo, a proposta foi aprovada em Junho de 2007 e o projeto começou em janeiro de 2008 com a entrega da amostra de DNA da árvore brasileira BRASUZ1 de *E. grandis*, eleita como a mais adequada para o projeto para facilitar o trabalho de montagem do genoma (Myburg et al., 2011). Este projeto internacional, de acesso totalmente público, representou um salto quantitativo e qualitativo gigantesco para a genômica e genética do eucalipto, comparável àquele que ocorreu com o sequenciamento do genoma humano. A participação da Embrapa e do Brasil como um todo, por meio dos seus pesquisadores organizados em grupo, juntamente com o acesso aos recursos genômicos, material vegetal e bases de dados geradas no âmbito do Projeto Genolyptus, foi decisiva para alavancar a aprovação e concretização deste projeto internacional que, ao publicar em 2013 a sequência completa e interpretada do genoma do eucalipto na prestigiosa revista *Nature* (Myburg et al., 2014), colocou a espécie entre as poucas árvores cujo genoma foi sequenciado até hoje.

Um aspecto interessante de destacar como uma contribuição da Embrapa por meio do Projeto Genolyptus é o fato da rede ter sido citada reiteradamente como modelo de sucesso na área de inovação via redes de cooperação tecnológica, para a criação de conhecimento em publicações da Fundação Getúlio Vargas (FGV) (Figueiredo, 2014, 2016) e se tornou um *case study* na Fundação Dom Cabral (FDC). O estudo detalhado feito pela FDC se baseou em diversos documentos gerados ao longo do projeto e entrevistas semiestruturadas com os principais participantes seja das universidades bem como das empresas. Foram analisados vários aspectos, desde as particularidades do setor florestal, o histórico da formatação inicial do Projeto Genolyptus com seus termos de compromisso, confidencialidade e regimento interno de funcionamento da rede, até a rotina de troca de informações entre os participantes e a importância do capital relacional do seu líder para que o projeto se concretizasse. Entre as várias conclusões deste estudo, que resultou em uma tese de doutorado e algumas publicações internacionais (Milagres; Silveira, 2008; Foss; Milagres, 2014), os autores destacaram que a rede Genolyptus teve sucesso pois, aproveitando o histórico cooperativo do setor florestal, conseguiu gerar forte identidade entre seus membros, realizou uma repartição clara do trabalho e responsabilidades entre os membros e teve um caráter democrático dos processos de decisão criados. Dizem os autores:

O que a Rede Genolyptus pode ensinar a outras redes? Muito. Sugere-se a observação da estrutura de coordenação implementada e as relações de proximidade entre empresas, institutos de pesquisa e universidades, que são os elementos diferenciadores dessa rede. (Milagres, 2009, p. 35).

A Embrapa protagoniza o desenvolvimento de recursos genômicos para espécies de *Eucalyptus*

Marcadores moleculares vêm sendo utilizados em espécies de *Eucalyptus* desde os anos 80, quando marcadores isoenzimáticos permitiram realizar os primeiros estudos de sistemas de cruzamento em pomares de sementes e analisar a diversidade genética de populações naturais (Moran et al., 1983; Moran, 1992). Desde então, a análise genética de espécies de *Eucalyptus* progrediu essencialmente em razão do desenvolvimento de técnicas moleculares começando com marcadores Restriction Fragment Length Polymorphism (RFLP) no início dos anos 90 (Byrne et al., 1993) e, principalmente, com as técnicas baseadas na reação de polimerase e cadeia (PCR) dos marcadores Random Amplified Polymorphic DNA (RAPD) (Grattapaglia et al., 1992) e Amplified Fragment Length Polymorphisms (AFLP) (Gaiotto et al., 1997). Marcadores RFLP logo foram abandonados como técnica viável para análises na escala desejada, em vista de sua baixa eficiência e alto custo. Marcadores RAPD e, em menor escala, AFLP tiveram assim um papel fundamental ao facilitar o acesso rápido e econômico do eucalipto ao mundo da genética molecular, apesar da absoluta carência de recursos moleculares, em comparação a espécies anuais que já contavam com elevado financiamento e amplas redes de pesquisa na época. Ao permitir a análise de várias dezenas ou centenas de marcadores no genoma, marcadores RAPD permitiram realizar análises de variabilidade, distância e identidade genética entre indivíduos de *Eucalyptus* (Grattapaglia et al., 1992). Ao reunir a flexibilidade de marcadores RAPD com a abordagem simples, porém criativa e inédita até então de mapeamento genético por pseudo-cruzamento teste, foram gerados os primeiros mapas genéticos para espécies de *Eucalyptus* (Grattapaglia; Sederoff, 1994; Verhaegen, 1996) e, logo em seguida, o mapeamento dos primeiros QTLs no gênero (Grattapaglia et al., 1995, 1996; Verhaegen et al., 1997), caminho seguido também com marcadores AFLP (Marques et al., 1998, 1999; Myburg et al., 2003).

Apesar da importância histórica de marcadores RAPD e AFLP ao inaugurar a entrada do eucalipto e, na verdade, de inúmeras outras espécies de plantas e animais na época, na era da genômica, estas tecnologias, por suas características metodológicas, não se tornaram recursos genômicos efetivos, ou seja recursos de ampla e duradoura utilização pela comunidade técnica e científica. Além da limitação da herança dominante e baixo conteúdo informativo, marcadores RAPD e AFLP se baseavam na amplificação simultânea de múltiplos segmentos de DNA por meio de iniciadores de PCR com sequência arbitrária, gerando marcadores sem endereço específico conhecido no genoma de um indivíduo e que apenas em parte aderiam aos princípios básicos de segregação mendeliana. Estas metodologias apresentavam, assim, limitações de reprodutibilidade entre laboratórios ou mesmo entre análises distintas em um mesmo laboratório a depender dos reagentes utilizados e, principalmente do rigor e critérios

do experimentador na coleta e análise de dados. Informações genéticas geradas com marcadores RAPD e AFLP tinham, portanto, vida curta e os marcadores propriamente ditos não eram facilmente portáveis entre estudos, o que impedia a comparação e reprodução de resultados.

Marcadores microssatélites: contribuição pioneira da Embrapa para a análise genômica de *Eucalyptus*

Na primeira metade dos anos 90 com a primeira onda de avanço tecnológico e redução de custos dos métodos e equipamentos de sequenciamento de DNA, a capacidade de geração de sequências de DNA aumentou significativamente. Este avanço, por sua vez, permitiu o desenvolvimento de outros tipos de marcadores moleculares baseados na PCR, desta vez amplificados em um endereço preciso e único no genoma, com herança mendeliana codominante. Além disso, estes marcadores se baseavam em uma classe de sequências de DNA repetitivo, altamente variável entre indivíduos, chamadas de sequências simples repetitivas (SSR simple sequence repeats), repetições curtas em tandem (STR short tandem repeats) ou microssatélites. Amplas baterias de microssatélites foram logo desenvolvidas para seres humanos, rapidamente revolucionando a velocidade de mapeamento genético e identificação de genes envolvidos no controle de doenças mendelianas (Weissenbach et al., 1992). O desenvolvimento de marcadores microssatélites logo foi seguido para algumas das principais culturas agrícolas como soja (Akkaya et al., 1992) e milho (Senior et al., 1996). Marcadores microssatélites se tornaram, assim, o padrão ouro em termos de metodologia de análise genética com marcadores moleculares em plantas, animais e microorganismos eucariotos (Figura 3).

Em 1995 foi criado o Laboratório de Genética Vegetal (LGV) dentro da área de recursos genéticos da Embrapa Recursos Genéticos e Biotecnologia (Embrapa Cenargen), com o mandato de apoiar projetos de caracterização e conservação de recursos genéticos e programas de melhoramento de plantas da empresa, com base no desenvolvimento, otimização e aplicação de marcadores moleculares e demais tecnologias genômicas que estavam começando a surgir. Em vista da revolução que marcadores microssatélites representavam para o avanço da aplicação da genética molecular nestas áreas, o laboratório estabeleceu, mediante interações pessoais anteriores dos líderes do LGV-Embrapa, uma colaboração informal com a empresa Dupont nos EUA, cujo time já estava muito avançado no desenvolvimento e uso de microssatélites para o melhoramento de milho (Taramino; Tingey, 1996). Dois estudantes de doutorado na época sendo treinados no LGV-Embrapa, passaram um período de treinamento sanduiche na Dupont e retornaram com centenas de marcadores microssatélites descobertos para *Eucalyptus* e arroz e, o mais importante, todo o conhecimento tecnológico para o desenvolvimento de microssatélites. Em paralelo a este treinamento, em

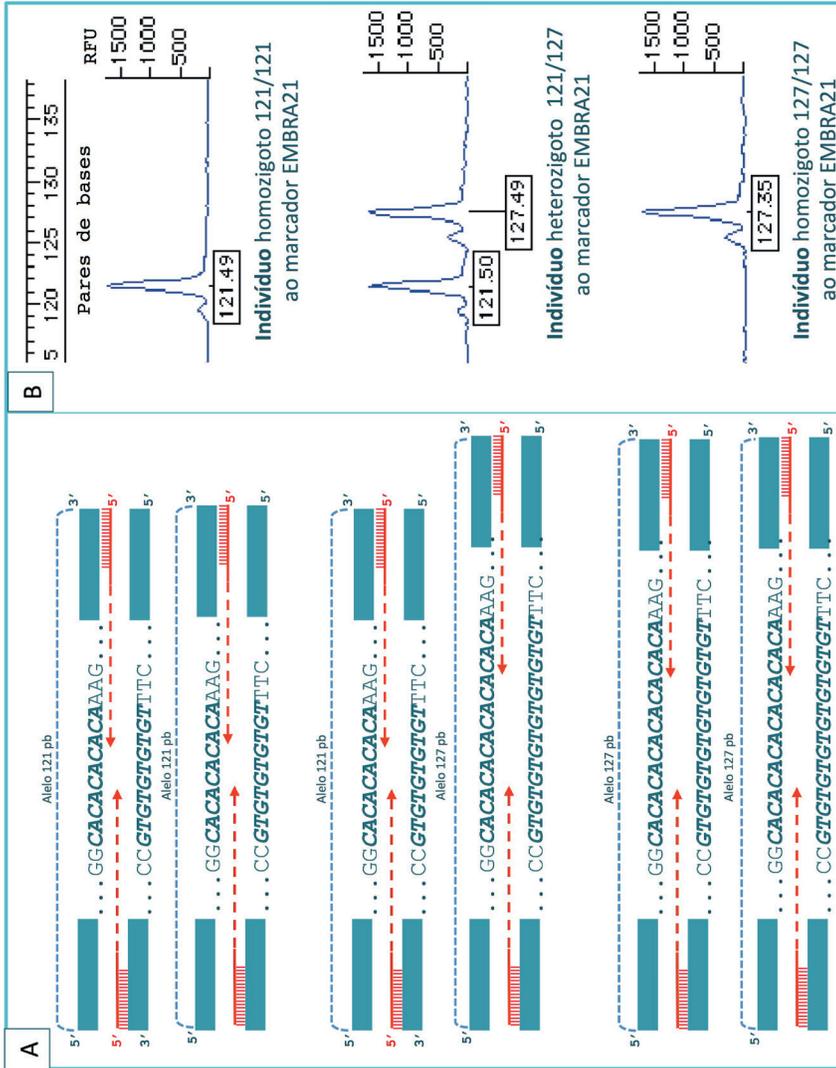


Figura 3. (A) Base genética do marcador molecular microsatélite, ilustrando um diagrama das sequências de DNA amplificadas via PCR de um dos mais de 500 marcadores microsatélites EMBRA desenvolvidos pela Embrapa. São destacados três indivíduos de *Eucalyptus*, apresentando genótipos distintos pela variabilidade no número de unidades repetitivas do dinucleotídeo (CA)_n nos diferentes alelos em configurações de homocigose ou heterocigose. (B) Eletroferogramas os três indivíduos, ilustrando a detecção dos alelos ao marcador microsatélite com os respectivos tamanhos estimados em pares de bases, considerando que frações decimais são ajustadas para a unidade via software na exportação final dos genótipos.

1996 o LGV-Embrapa adquiriu o primeiro sequenciador automático de DNA ABI-377 no Brasil e demais equipamentos necessários para a implementação da tecnologia de desenvolvimento e análise automatizada de microssatélites por meio de recursos competitivos obtidos com um projeto do Programa de Apoio ao Desenvolvimento Científico e Tecnológico (PADCT) do CNPq. Com esta iniciativa, a Embrapa se tornou em 1996 a instituição pioneira no país no desenvolvimento de microssatélites para espécies vegetais, logo estendendo o trabalho para fungos fitopatogênicos. Nos anos seguintes o LGV-Embrapa se tornou uma verdadeira “fábrica” de descoberta e desenvolvimento de microssatélites para mais de 30 espécies de plantas e fungos, divulgando a tecnologia (Kirst et al., 1997; Brondani et al., 2007), treinando profissionais por meio de cursos, e viabilizando inúmeros projetos de pesquisa da Embrapa em colaboração com várias universidades e instituições de pesquisa do Brasil e algumas do exterior.

Especificamente para espécies de *Eucalyptus*, a Embrapa apresentou os primeiros resultados deste trabalho no congresso internacional da International Union of Forest Research Organization (Iufro) de silvicultura de Eucalipto em Salvador, BA, em 1997 (Brondani et al., 1997) e, em seguida, publicou os primeiros microssatélites para espécies de *Eucalyptus* com informação de mapeamento (Brondani et al., 1998), dando início a uma das principais contribuições para a comunidade científica internacional na área de genética e melhoramento do eucalipto, além de iniciar a internalização da genética molecular aplicada no setor industrial de base florestal do Brasil e do mundo. Com os recursos de sequenciamento do genoma do eucalipto gerados no Projeto Genolyptus (Rede Brasileira de Pesquisa do Genoma do Eucalipto), discutido anteriormente, a Embrapa seguiu neste trabalho de uma forma mais intensa desenvolvendo coleções de várias centenas de microssatélites geneticamente mapeados (Brondani et al., 2002, 2006; Marques et al., 2002; Faria et al., 2011b; Grattapaglia et al., 2015) e sistemas otimizados de marcadores para “fingerprinting” de clones e populações (Kirst et al., 2005; Faria et al., 2010, 2011a).

Estes marcadores com a denominação Embra, que remete à Embrapa, bem como a um acrônimo para Eucalyptus Microsatellites from Brazil, são usados hoje por dezenas de laboratórios públicos e privados em todo o mundo. Apesar de terem sido publicadas algumas dezenas de microssatélites por outros grupos na Austrália, os microssatélites Embra representam um dos mais importantes recursos para pesquisa genética e genômica de espécies de *Eucalyptus* e, seguramente, o recurso mais amplamente usado pelo fácil acesso, grande número de marcadores, ampla aplicabilidade em diversas áreas de pesquisa e baixo custo de utilização (Grattapaglia et al., 2012). Vários trabalhos aplicando microssatélites na solução de problemas práticos do melhoramento e descoberta de associações entre variação genômica e fenotípica, discutidos na seção de aplicações a seguir, foram realizados pelo time do LGV-Embrapa em colaboração com empresas florestais, universidades brasileiras e internacionais por meio da

atuação de dezenas de estudantes, trabalhos estes relatados em revisões bibliográficas publicadas pela equipe do LGV-Embrapa ou em colaboração com outros grupos ao longo dos anos (Grattapaglia, 2000, 2006; Myburg et al., 2007; Grattapaglia; Kirst, 2008; Grattapaglia et al., 2012).

Marcadores SFPs, SNPs e GbS: a Embrapa avalia novas tecnologias de genotipagem de *Eucalyptus*

A partir de 2005, com o avanço extraordinário das novas tecnologias de sequenciamento de DNA (Next Generation Sequencing (NGS)), em paralelo ao desenvolvimento de plataformas de genotipagem de alto desempenho, a área de análise genômica passou por uma verdadeira revolução na possibilidade de descobrir e analisar a variação existente diretamente na sequência de DNA, decorrente de inserções e deleções de sequência, bem como dos chamados polimorfismos de base individual (Single Nucleotide Polymorphisms (SNP)), a maior fonte de variação de sequência em genomas (Figura 4). Ao longo do Projeto Genolyptus, atento a estes desenvolvimentos tecnológicos na área da genômica, e em vista da necessidade de tecnologias acessíveis, econômicas e de maior cobertura genômica para implementar ações genômicas mais efetivas no melhoramento de *Eucalyptus*, a equipe do LGV-Embrapa, a partir de 2007, enxergou oportunidades de evoluir de microssatélites para metodologias mais avançadas que permitissem genotipar polimorfismos de sequência em larga escala. Embora microssatélites oferecessem vantagens analíticas do ponto de vista de conteúdo informativo e transferibilidade, o número de marcadores e a velocidade com a qual podiam ser genotipados limitavam seu uso para aplicações em escala genômica (*genome-wide*). Tecnologias de sequenciamento massal (Barbazuk et al., 2007) e metodologias de genotipagem de SNPs, desenvolvidas principalmente para seres humanos, sistemas modelo e algumas plantas cultivadas, já permitiam, respectivamente, a descoberta e a análise de milhares de polimorfismos de sequência de forma muito mais rápida e a custos inferiores aos custos dos marcadores microssatélites, quando avaliados na base do custo por genótipo e por indivíduo (*data point*) (Murray et al., 2004; Gunderson et al., 2005; Ganai et al., 2009).

Entretanto, embora o custo por genótipo de SNPs já tivesse alcançado naquele momento valores da ordem de alguns centavos de dólar, os custos por amostra ainda eram da ordem de 150 a 200 US\$ para a genotipagem de alguns milhares de SNPs por meio da tecnologia “padrão ouro” usada para seres humanos e baseada em microarranjos fixos de SNPs. Na área florestal, somente para as coníferas *Pinus taeda* e *Picea abies*, ambas com amplo financiamento de grandes projetos de genômica florestal no Canadá e EUA, respectivamente, microarranjos fixos com centenas a poucos milhares de SNPs haviam sido desenvolvidos (Pavy et al., 2008; Eckert et al., 2009). No caso de *Eucalyptus*, apesar da disponibilidade de bancos de sequências do Projeto Genolyptus

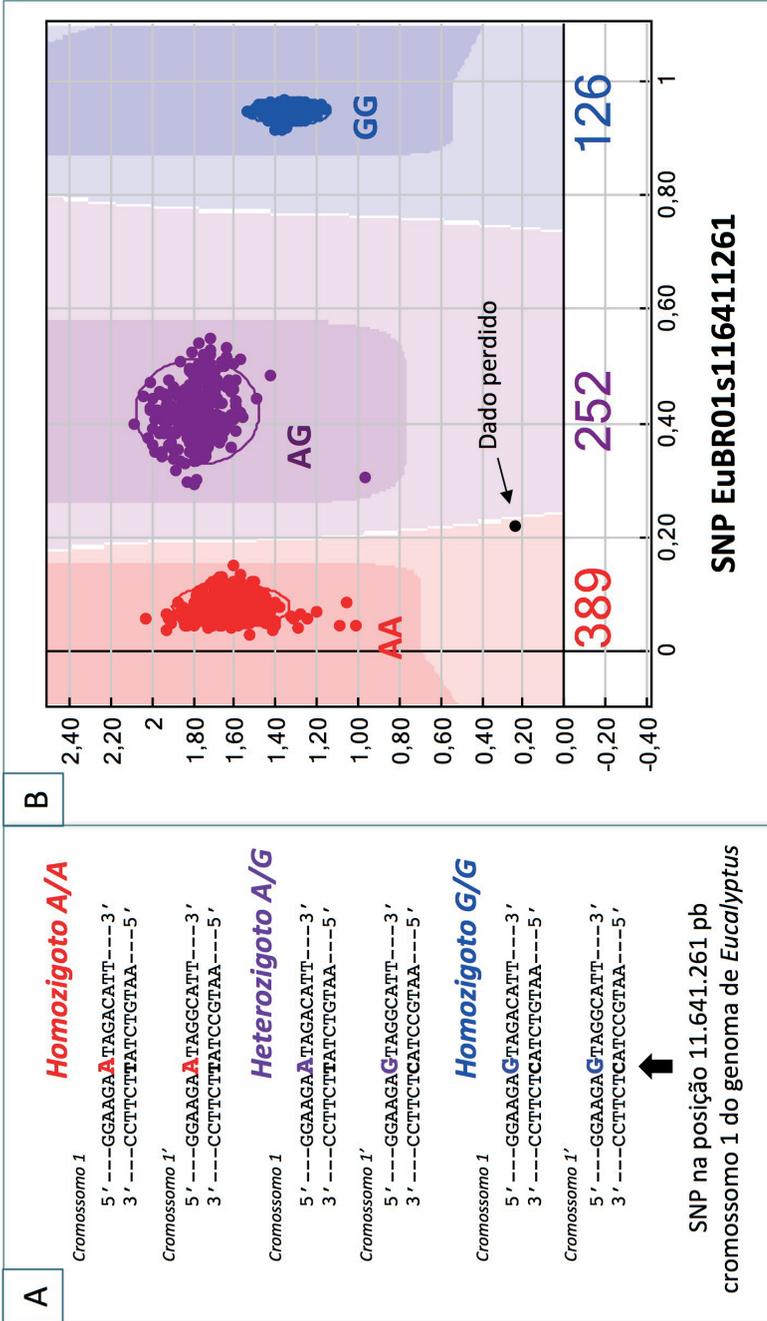


Figura 4. (A) Base genética do marcador SNP (*Single Nucleotide Polymorphism*). Para cada um de três indivíduos, o diagrama ilustra os dois homólogos do cromossomo 1 do genoma de *Eucalyptus grandis* e a posição de coordenada 11.641.261 pares de bases na qual ocorre um SNP cuja variação de base é A ou G correspondendo aos dois alelos. (B) Detecção e genotipagem automatizada via software do marcador SNP realizada com o “chip” de 60.000 SNPs (EuCHIP60K) desenvolvido pela Embrapa. O SNP EuBR01s11641261 (*Eucalyptus* Brasil; cromossomo 01; posição 11.641.261 pb) foi genotipado em 768 indivíduos de diferentes espécies de *Eucalyptus* resultando em 389 indivíduos homozigotos AA, 252 indivíduos heterozigotos AG, 126 indivíduos homozigotos GG e um indivíduo com dado faltante.

e bancos maiores ainda de sequências gerados com as novas tecnologias NGS (Novaes et al., 2008) e sequências disponíveis de forma crescente com o projeto do genoma completo, ainda não existia esforço em nível mundial até aquele momento para o desenvolvimento de metodologias de genotipagem de SNPs para espécies do gênero e os recursos eram escassos. Quatro projetos paralelos foram então desenvolvidos no LGV-Embrapa entre 2007 e 2010 visando avaliar diferentes abordagens paralelizadas de genotipagem de centenas a milhares de polimorfismos de sequência de DNA que fossem acessíveis a custos relativamente baixos.

O primeiro deles envolveu o desenvolvimento de um microarranjo de Single Feature Polymorphisms (SFPs) em parceria com a University of Florida envolvendo o treinamento de um mestrando do LGV-Embrapa, naquela instituição. SFPs são, traduzindo o termo, polimorfismos de elementos únicos, ou seja, um tipo de marcador baseado na presença ou ausência de hibridização de DNA complementar a sondas desenhadas a partir de sequências de RNA, detectando dessa forma qualquer tipo de polimorfismo de sequência em regiões transcritas do genoma que resultasse em hibridização ou não da sonda. Um microarranjo de triagem com 103 mil sondas desenhadas a partir dos 20.276 genes únicos descritos na base de dados do Projeto Genolyptus foi construído. Um microarranjo operacional com um total de 43.777 sondas selecionadas representando 15.698 genes foi então utilizado para a construção de um mapa genético sobre o qual foram posicionados 1845 genes, 87% dos quais mapearam na localização esperada na sequência do genoma de *Eucalyptus grandis* (Neves et al., 2011). Esta metodologia embora interessante, não se mostrava suficientemente automatizável e econômica para uso em larga escala para aplicações no melhoramento além de herança dominante, embora tentativas tenham sido feitas para analisar a intensidade do sinal de hibridização dos SFPs e com isso derivar marcadores codominantes.

O segundo projeto representou o primeiro trabalho em nível internacional de desenvolvimento de um chip de SNPs para espécies de *Eucalyptus*. Para o genoma do eucalipto, a abundância de SNPs como resultado da elevada heterozigidade e diversidade genética facilitavam a sua identificação. Por outro lado, por apresentar uma alta diversidade de nucleotídica, um diagnóstico mais refinado do contexto de sequência de localização do SNP era necessário para desenhar sondas que permitissem a genotipagem acurada com a tecnologia disponível na época para o chip (Golden Gate Genotyping Technology, GGGT). Esta questão se tornava ainda mais exacerbada para o plano do projeto que era permitir a transferibilidade de SNPs entre as diferentes espécies de *Eucalyptus*, um tema que ainda era praticamente inexplorado, uma vez que chips de SNPs somente eram desenvolvidos para uma mesma espécie de planta ou animal. A equipe do LGV-Embrapa iniciou nesta época uma parceria mais intensa ainda com o laboratório de bioinformática da Embrapa Recursos Genéticos e Biotecnologia que já participava ativamente do Projeto Genolyptus e do projeto internacional de sequenciamento do genoma de *Eucalyptus grandis*. Neste momento, uma

primeira versão do genoma de *Eucalyptus grandis* com 4,5X de cobertura de sequenciamento já era acessível para os participantes do projeto internacional juntamente com uma ampla base de dados de sequências de genes expressos. O acesso a estes recursos genômicos possibilitou realização do primeiro trabalho sistemático de descoberta de SNPs em larga escala para espécies do gênero. Foi desenvolvido o primeiro chip de baixa densidade com 768 SNPs para *Eucalyptus* a partir da análise de um banco de dados com 1.164.695 de sequências de genes expressos de *Eucalyptus* envolvendo as cerca de 130 mil sequências geradas pela tecnologia Sanger no Projeto Genolyptus e de trabalhos pioneiros de sequenciamento em larga escala com a tecnologia 454 realizado pela University of Florida.

O trabalho descrevendo este avanço (Grattapaglia et al., 2011), além de resultar no primeiro sistema de genotipagem de SNPs para *Eucalyptus* no mundo, fez uma importante contribuição para a área no direcionamento da necessidade de avaliação sistemática de SNPs, em genomas altamente heterozigotos, no processo de desenvolvimento de chips de SNPs. O trabalho quantificou o impacto da aplicação de restrições de ausência de SNPs adicionais nas sequências flanqueantes ao SNP, no desempenho qualitativo da genotipagem. A partir de um total de 162.141 SNPs identificados na base de sequências gênicas, 768 SNPs foram selecionados para o chip, e destes, os 288 que haviam sido selecionados com mais restrições de contexto de sequência forneceram genotipagem de mais alta qualidade. Obteve-se sucesso da genotipagem acima de 93% em um painel diversificado de indivíduos de nove espécies de *Eucalyptus* pertencentes a três seções dentro do subgênero *Symphomyrtus*, e ainda satisfatórias para as espécies de dois subgêneros adicionais, embora o polimorfismo fosse cada vez menor à medida que a distância filogenética aumentava em relação às espécies do subgênero *Symphomyrtus*. Um avanço importante deste trabalho foi a demonstração que seria viável o desenvolvimento de um conjunto muito maior de SNPs informativos para *Eucalyptus* com grande possibilidade de transferibilidade entre as principais espécies plantadas. Para um desenvolvimento de um chip de mais alta densidade de SNPs, entretanto, concluiu-se que seria vital partir de uma coleção extensa de sequências de DNA com ampla representatividade de espécies e indivíduos, para permitir uma seleção precisa de SNPs com as devidas restrições de contexto de sequência ao longo de todo o genoma.

A terceira metodologia avaliada, denominada Diversity Array Technology (DArT), foi fruto de uma parceria com a empresa australiana Diversity Array Technologies e o grupo de pesquisa em eucalipto da University of Tasmania. A metodologia DArT foi descrita há quase 20 anos nos primórdios do surgimento da tecnologia de microarranjos de ácidos nucleicos (Jaccoud, Peng et al., 2001), e apresentava várias das características desejadas em termos de paralelização da análise de centenas a milhares de marcadores, a possibilidade de genotipar diferentes espécies do gênero, e o baixo custo, o que atenderia com sucesso o que era buscado para *Eucalyptus*. O mesmo modelo adotado

pelo LGV-Embrapa mais uma vez se mostrou vencedor: dois doutorandos que atuavam no LGV-Embrapa trabalharam na empresa DArT na Austrália por mais de um ano no desenvolvimento e posteriormente utilização de um microarranjo com 7.680 marcadores para *Eucalyptus*. Estes marcadores correspondem a uma coleção representativa de sequências genômicas de cópia única com 300 a 500 pares de bases produzidas via redução de complexidade com enzimas de restrição e arranjadas sobre um slide de vidro. As amostras a serem analisadas são preparadas com o mesmo procedimento de redução de complexidade genômica, amplificadas via PCR e, em seguida, hibridizadas no microarranjo, fornecendo marcadores dominantes do tipo presença ou ausência como resultado da captura do sinal de hibridização em um scanner.

O microarranjo DArT para *Eucalyptus* (Sansaloni et al., 2010) pelas suas características técnicas, foi particularmente valioso por permitir genotipar milhares de marcadores para centenas ou milhares de amostras a baixo custo para as várias espécies do gênero. Logo alcançou amplo uso por diversos grupos trabalhando com espécies de *Eucalyptus* no mundo, constituindo mais uma importante contribuição da equipe da Embrapa, em parceria com os colegas australianos, para a comunidade mundial do eucalipto. Em um trabalho subsequente, os fragmentos de DNA que compõem o arranjo foram sequenciados, caracterizados e geneticamente mapeados (Petroli et al., 2011), fornecendo marcadores âncoras que foram fundamentais, mais tarde, juntamente com outros mapas genéticos (Hudson et al., 2012) para auxiliar o processo de montagem e ordenamento dos pseudo-cromossomos do genoma de referência do eucalipto. O microarranjo DArT para *Eucalyptus* foi também chave na investigação de questões filogenéticas até aquele momento pendentes no gênero, envolvendo diferenciação de espécies, identificação de híbridos interespecíficos e resolução de disjunções biogeográficas entre espécies (Steane et al., 2011). Finalmente, na área da genômica aplicada ao melhoramento, o microarranjo DArT foi a ferramenta genômica essencial para a genotipagem em larga escala de diversas populações, o que permitiu ao grupo da Embrapa entrar pioneiramente na área experimental de seleção genômica em espécies florestais, de maneira geral e para o gênero *Eucalyptus* em particular (Resende et al., 2012), abrindo uma longa história de avanços neste tema, discutido em detalhe mais à frente.

Com o avanço das tecnologias e redução de custos de sequenciamento no período de 2008 a 2009, metodologias de genotipagem de SNPs por sequenciamento Genotyping by Sequencing (GbS) foram surgindo, explorando a possibilidade de genotipar polimorfismos em sequências curtas geradas via sequenciamento massal Next Generation Sequencing (NGS) de pools de indivíduos indexados com adaptadores específicos que permitem recuperar a identidade de cada indivíduo no pool (Davey et al., 2011). A dificuldade de montar, acessar e manter plataformas de sequenciamento NGS no Brasil e a dificuldade crônica do país de importar reagentes, a equipe do LGV-Embrapa passou a utilizar cada vez mais o conceito de contratação de serviços

genômicos no exterior. Nesta época, empresas de prestação de serviços de sequenciamento e genotipagem começaram a se estabelecer nos EUA, China e outros países, criando uma saudável concorrência internacional, o que causou, juntamente com o avanço das tecnologias, uma redução rápida dos custos. Tornou-se assim muito mais econômico, rápido e com maior qualidade de dados, contratar serviços no exterior do que tentar comprar e manter equipamentos de rápida obsolescência, importar reagentes e treinar pessoas para atividades de rotina. Atualmente, o *modus operandi* estabelecido para trabalhos na área de análise genômica envolvendo sequenciamento e genotipagem é mediante a contratação de serviços genômicos no exterior, permitindo às equipes dedicar mais tempo ao planejamento experimental, análise, interpretação e utilização dos dados gerados.

Três abordagens de genotipagem baseada em sequenciamento existiam e foram avaliadas como uma terceira alternativa para a análise genética de *Eucalyptus*, em larga escala. A técnica de GbS (Elshire et al., 2011) foi avaliada tanto no LGV-Embrapa como por meio da contratação de serviço na Cornell University; a técnica RAD-sequencing (Restriction Amplified DNA Sequencing) (Baird et al., 2008) foi contratada com a empresa Floragenex também nos EUA e a versão da metodologia DArT baseada em sequenciamento, neste caso denominada DArT-seq (Sansaloni et al., 2011), com a empresa DArT na Austrália. As três metodologias são essencialmente equivalentes com apenas algumas variações na forma de preparo inicial do DNA da amostra. As metodologias se baseiam, resumidamente, na redução de complexidade genômica da amostra de DNA total, utilizando corte com uma ou duas enzimas de restrição e, no caso de RAD-sequencing, realizando uma quebra aleatória (*shearing*) do DNA após o corte com enzima. Após a redução de complexidade e antes do sequenciamento, cada amostra de DNA a ser genotipada recebe adaptadores com sequências indexadoras (*barcodes*) que permitem mais tarde rastrear as sequências geradas para cada amostra. Fragmentos obtidos dentro de uma faixa de tamanho adequado são, então, selecionados e sequenciados gerando dezenas de milhões de sequências curtas (70 a 150 bases) em plataformas de sequenciamento NGS. Com a indexação de cada amostra com um *barcode* específico, dezenas de amostras podem ser sequenciadas conjuntamente em cada canaleta de sequenciamento, otimizando significativamente os custos de geração dos dados. Milhares ou dezenas de milhares de marcadores potenciais, na forma de polimorfismos de presença ou ausência, derivados da variabilidade na distribuição dos sítios de restrição, e SNPs entre sequências em comum entre as amostras são detectados por meio do mapeamento das sequências geradas sobre um genoma de referência. Os dados são filtrados intensamente para métricas de qualidade, resultando em alguns até vários milhares de marcadores efetivamente úteis, dependendo da diversidade da espécie. Ao longo dos projetos, a equipe do LGV-Embrapa testou o protocolo GbS (Elshire et al., 2011) em *Eucalyptus* obtendo cerca de 17.000 SNPs segregando em uma população de mapeamento, e mais de 130.000 SNPs potenciais

em uma amostra de 24 indivíduos geneticamente não relacionados de *E. grandis* (Faria et al., 2012). A técnica DArT-seq foi utilizada para a construção de mapas genéticos de alta densidade (Sansaloni et al., 2011) e a técnica RAD-sequencing para descoberta de SNPs (Grattapaglia et al., 2011), os quais, foram utilizados mais tarde para calibrar a descoberta e seleção de SNPs realizada visando a construção de um micro arranjo fixo de SNPs para *Eucalyptus*, discutido a seguir.

Estes métodos de genotipagem baseados em sequenciamento eram atraentes na época, pois supostamente forneceriam um grande número de SNPs de qualidade a um custo menor por amostra, quando comparados a microarranjos de conteúdo fixo de SNP. Esta promessa se concretizou relativamente bem para espécies autógamas, para as quais a homozigose de plantas é elevada ou praticamente total em linhagens comerciais, o que facilita a análise de SNPs. Entretanto, os resultados obtidos com as três metodologias em plantas de *Eucalyptus* revelaram que os desafios eram expressivos para se considerar a genotipagem direta de SNPs com alta acurácia em plantas alógamas com elevada heteroziguidade. A baixa cobertura de sequenciamento utilizada para manter os custos acessíveis resultava em uma grande quantidade de dados faltantes derivados de uma amostragem altamente variável dos locos, afetando seriamente a reprodutibilidade entre os experimentos, seja em termos da capacidade de amostrar os mesmos SNPs, bem como de reproduzir o mesmo genótipo ao SNP amostrado. Nos experimentos de GbS realizados no LGV-Embrapa as estimativas de reprodutibilidade somente alcançaram cerca de 65%, uma taxa claramente inaceitável. Serviços de GbS contratados junto ao centro de prestação de serviços genômicos da Cornell University forneceram resultados equivalentes. Entre as três técnicas avaliadas, a DArT-seq foi a que forneceu os melhores resultados, por utilizar uma maior cobertura de sequenciamento e processar os dados com uma filtragem intensa, procedimentos estes que, por outro lado, resultavam em maiores custos e em uma quantidade final de SNPs abaixo do necessário. Diversos trabalhos publicados nos anos subsequentes destacaram problemas com estas técnicas, mostrando que a divergência entre o genoma das amostras genotipadas e o genoma de referência utilizado na análise também contribuía para uma reprodutibilidade genotípica variável que se mostrava mais exacerbada em genomas heterozigotos (Myles, 2013), impactando as medidas de diversidade genética e, por conseguinte, as conclusões de estudos realizados com estas técnicas (Arnold et al., 2013; Gautier et al., 2013; Lowry et al., 2016).

A avaliação das tecnologias de genotipagem de polimorfismos de seqüências de DNA permitiu à equipe do LGV-Embrapa chegar a algumas conclusões importantes que direcionaram os passos subsequentes. A tecnologia SFP não era eficiente em termos de tempo, custo e limitações técnicas, pois envolvia a redução de complexidade genômica com base em RNA. A tecnologia DArT de microarranjos de sondas tinha uma limitação ao número de marcadores que poderiam ser analisados, apenas 7.680, além de ser fornecida por apenas uma empresa no mundo que detinha a proteção da

tecnologia, o que representava um risco de falta de fornecimento de serviço caso a empresa saísse do mercado. Todas as metodologias de genotipagem por sequenciamento não haviam apresentado resultados satisfatórios de robustez analítica, mesmo contratando serviços em centros especializados. A única metodologia com desempenho adequado para as aplicações previstas no melhoramento foi o chip de 768 SNPs que, além de fornecer dados robustos com elevada reprodutibilidade, também demonstrou a possibilidade de genotipar os mesmos SNPs em diferentes espécies. A tecnologia GGGT utilizada para este chip de SNPs era o “padrão ouro” até então, mas novas nanotecnologias já haviam surgido e vinham revolucionando a genotipagem em larga escala em seres humanos, permitindo a análise de um número muito maior de SNPs. Além disso, nesta altura dos acontecimentos, com a concorrência crescente na área de genotipagem em nível mundial, micro arranjos fixos de SNPs estavam se tornando significativamente mais acessíveis em termos de custo, desde que existisse a perspectiva concreta de genotipar uma quantidade suficientemente grande de alguns milhares de amostras. Viu-se, então, a oportunidade de propor um projeto de desenvolvimento de um sistema de genotipagem de alto desempenho para espécies de *Eucalyptus* que viesse a efetivamente revolucionar a capacidade de integração da genômica no melhoramento do eucalipto.

EUCHIP60K de SNPs para *Eucalyptus*: a Embrapa inova a genotipagem mundial em larga escala

Os projetos descritos anteriormente que visaram a avaliação de tecnologias de genotipagem para avançar na área de seleção genômica foram financiados por um projeto competitivo denominado Núcleo de Excelência de Genômica Florestal Aplicada (Nextree), aprovado pela equipe do LGV-Embrapa em 2009 junto à Fundação de Apoio à Pesquisa do Distrito Federal (FAP-DF), em um edital conjunto com o CNPq, voltado para a consolidação de núcleos de excelência (Pronex). Com base neste projeto e com os trabalhos de simulação (Grattapaglia; Resende, 2011) e resultados experimentais promissores sobre o potencial de seleção genômica para *Eucalyptus* (Resende et al., 2012), em meados de 2012 a equipe do LGV-Embrapa lançou uma iniciativa de desenvolvimento de um chip de genotipagem em larga escala para poder efetivamente tornar a seleção genômica uma realidade para o setor florestal brasileiro. A experiência e reputação do LGV-Embrapa adquiridas ao longo do Projeto Genolyptus em trabalhar com as empresas de base florestal no País foi fundamental para viabilizar mais esta iniciativa.

Um convite informal foi feito a todas as empresas de base florestal no país para participarem desta iniciativa com base em algumas premissas. A primeira delas era o fato que a operacionalização da seleção genômica na rotina de uma empresa florestal demandava uma tecnologia de genotipagem avançada, rápida, robusta, de uso simples

e flexível, e a mais econômica possível. A segunda premissa era que a tecnologia deveria permitir genotipar milhares de amostras para milhares ou dezenas de milhares de marcadores, com custos reduzidos, com elevada reprodutibilidade, sem demandar infraestrutura ou pessoal altamente especializado em bioinformática ou genômica. A seleção genômica muito mais do que uma biotecnologia, deveria ser vista como um novo método de melhoramento que se caracteriza por utilizar dados genéticos de marcadores moleculares. A terceira premissa era que o custo inicial de descoberta e seleção dos SNPs para a construção de chip de alta densidade tinha baixado consideravelmente e representava um investimento viável. Por outro lado a efetiva construção inicial do chip tinha um custo inversamente proporcional à escala de produção de chips para uso posterior. Ou seja, quanto maior o número de chips encomendados ao fabricante, menor se tornava o preço individual. Isso fez com que o desenvolvimento de chips de SNPs para plantas cultivadas como o milho, soja e animais domésticos como bovinos e suínos fosse realizado via projetos colaborativos entre diversas instituições públicas e privadas interessadas no melhoramento genético e genômica da espécie alvo.

O projeto de desenvolvimento do chip de SNPs para *Eucalyptus*, batizado de EUChip60K foi idealizado e executado em um modelo inovador, semelhante ao sistema de financiamento de interesse coletivo (*crowdfunding*) que consiste na obtenção de capital para iniciativas de interesse coletivo mediante agregação de múltiplas fontes de financiamento. Neste tipo de iniciativa é geralmente estipulada uma meta de arrecadação que deve ser atingida para que o projeto seja viabilizado. Para fabricar o chip e disponibilizá-lo a um custo acessível, foi negociado com a empresa americana Neogen (na época Geneseek), a maior e mais competitiva empresa de genotipagem na área agropecuária mundial, o compromisso de genotipar pelo menos 15 mil amostras de DNA. Nem a Embrapa ou nenhuma empresa florestal individualmente poderia ou estaria disposta a pagar por 15 mil amostras ao custo de 60 dólares, o que somava um compromisso de 900 mil dólares. A Embrapa, também interessada em utilizar o EUChip60K para seus programas de melhoramento e caracterização de germoplasma de *Eucalyptus*, entraria com alguns recursos do Projeto Nextree, para o trabalho de descoberta e seleção de SNPs e o trabalho da equipe técnica do seu LGV. Para as empresas a proposta colocada foi que cada empresa participasse com o compromisso de genotipar pelo menos 1.000 amostras do seu próprio material genético. Para estimular as empresas a participarem foi concedido a empresa participante um desconto e a oportunidade de pagar apenas 51 dólares nas amostras subsequentes. Foi obtida uma ampla adesão de mais de dez empresas que permitiu não apenas tornar o EUChip60K uma realidade, mas também iniciou a integração do uso da informação genômica de SNPs, em larga escala nos programas de melhoramento genético de várias empresas brasileiras e da Embrapa, tornando o Brasil, mais uma vez, pioneiro em uma nova vertente de inovação tecnológica na área florestal.

Para o desenvolvimento do chip, o primeiro passo do projeto explorou as novas tecnologias de sequenciamento NGS no sentido de gerar uma extensa base de dados para a descoberta de SNPs, aprendizado obtido com projetos anteriores. Foi, então, realizado um trabalho inédito de resequenciamento do genoma inteiro (*whole genome resequencing*) de 240 árvores de 12 espécies de eucalipto, incluindo as chamadas *big nine* mais plantadas no mundo - *E. grandis*, *E. urophylla*, *E. globulus*, *E. pellita*, *E. camaldulensis*, *E. tereticornis*, *E. saligna*, *E. nitens*, *E. dunnii* - além de *E. pilularis*, *E. benthamii*, *E. cloeziana* e *Corymbia citriodora* como espécies que vêm sendo utilizadas de forma crescente em algumas regiões. Com base nesta ampla base de dados de sequências genômicas, quase 47 milhões de SNPs foram identificados dentro e entre espécies, em um trabalho inédito no Brasil de análise de dados e descoberta de variantes realizado pela equipe do laboratório de bioinformática da Embrapa Recursos Genéticos e Biotecnologia. A partir dessa ampla base de dados, após diversos passos de seleção com base em métricas de análise de sequências, foram selecionados 208 mil SNPs com características adequadas para genotipagem, com a tecnologia Infinium da empresa Illumina. Destes 208 mil, 60.904 SNPs foram então utilizados para compor o EUChip60K, fornecendo uma cobertura genômica de 96% do genoma com uma distribuição regular de um SNP a cada 12 kb a 20 kb ao longo dos cromossomos, e 47 mil SNPs próximos a 30 mil dos 34 mil genes anotados no genoma do eucalipto (Figura 5). O genoma de referência de *E. grandis*, a esta altura, já havia sido finalizado e disponibilizado por meio de um projeto internacional do qual o LGV-Embrapa participava ativamente na execução e co liderança.

O EUChip60K publicado em 2015 (Silva-Junior et al., 2015) foi o primeiro chip desenvolvido deliberadamente para atender a múltiplas espécies de um mesmo gênero de plantas. Além dos aspectos operacionais do chip, os dados levantados de milhares de SNPs compartilhados entre as diferentes espécies revelaram a existência de SNPs ancestrais surgidos antes da separação das espécies de *Eucalyptus*, sugerindo uma radiação recente do gênero. O EUChip60K representou uma contribuição significativa da Embrapa para a comunidade internacional, mais uma vez em parceria com o setor florestal privado nacional. O EUChip60K foi o primeiro sistema de genotipagem de alto desempenho para uma espécie florestal, colocando o eucalipto em condições equivalentes às grandes culturas agrícolas tais como milho e a soja, no que se refere ao potencial de avançar no estudo da genômica populacional, seleção genômica e estudos de associação. A possibilidade de genotipar espécies de eucalipto plantadas nos vários países de clima tropical e temperado em uma plataforma comum, e o fato do EUChip60K ter sido desenvolvido no conceito *open access*, ou seja, não patentado e disponível para todos os interessados, tem promovido o intercâmbio de dados e ampliado o conhecimento científico. Vários estudos vêm sendo publicados por diversos grupos no mundo utilizando o EUChip60K, revisados recentemente (Grattapaglia et al., 2018).

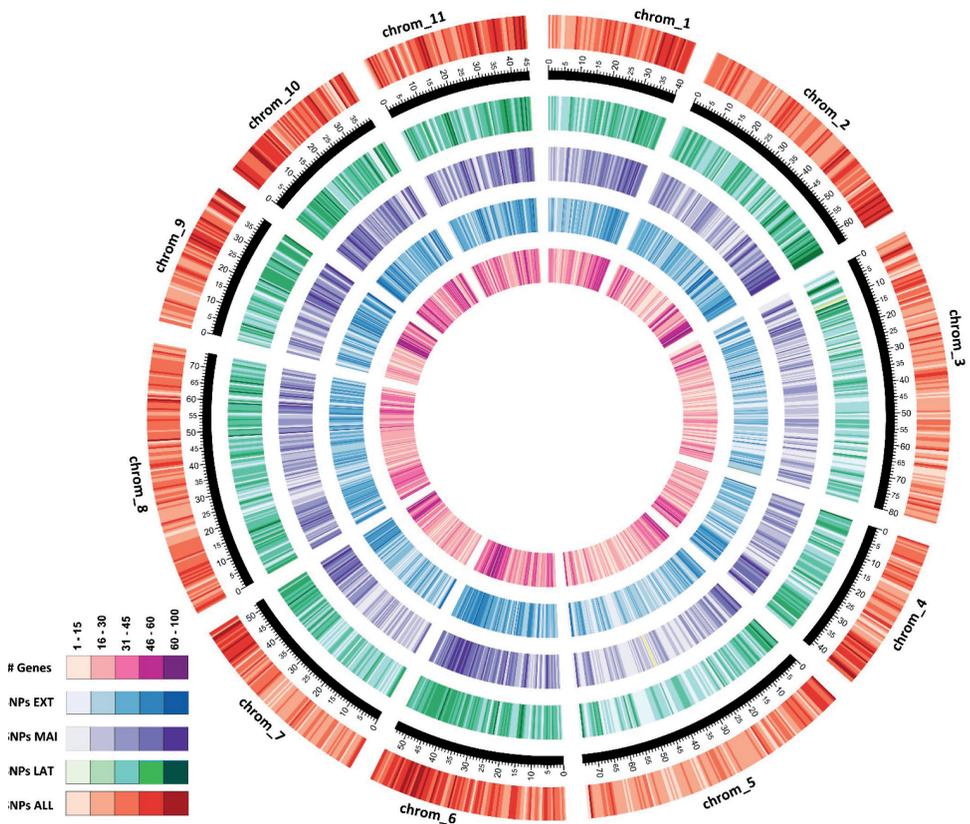


Figura 5. Mapa de calor da distribuição de densidade dos 51.204 SNPs polimórficos posicionados ao longo dos 11 cromossomos do genoma de *Eucalyptus* em intervalos de uma megabase (10^6 pares de bases), nas três principais seções do subgênero *Symphyomyrtus* (LAT Latoangulatae; MAI Maidenaria; EXT Exsertaria) e no consolidado de indivíduos das 14 espécies contempladas, com a densidade correspondente de modelos gênicos no genoma de referência de *Eucalyptus grandis*. Figura reproduzida com permissão (Silva-Junior et al., 2015).

O genoma do eucalipto: a Embrapa colidera a iniciativa mundial de sequenciamento

A iniciativa de produzir uma sequência do genoma completo do eucalipto foi discutida tecnicamente pela primeira vez em um encontro na cidade de Hobart na Austrália em 2004 da qual o LGV-Embrapa participou. Na época os grupos de pesquisa naquele país convidaram representantes dos principais grupos no mundo trabalhando com a genômica do eucalipto, para discutir os aspectos técnicos de como empreender este projeto ambicioso, considerando que o eucalipto é uma planta icônica e uma importante contribuição da Austrália para o mundo. Existia interesse de uma

ala do governo da Austrália em apoiar este projeto. Infelizmente o partido que havia sinalizado o apoio acabou saindo do poder nas eleições parlamentares fazendo com que o projeto não acontecesse. A ideia foi retomada em 2005 durante a conferência *IUFRO Tree Biotechnology* em Pretoria, África do Sul, do qual participaram ativamente vários pesquisadores na época trabalhando no Projeto Genolyptus. Um plano foi formatado para a formação da Rede Internacional Eucagen - Eucalyptus Genome Network, visando a submissão de uma proposta para competir na chamada pública anual que o centro de sequenciamento Joint Genome Institute (JGI) do Departamento de Energia do EUA lançava, visando apoiar projetos de interesse científico mundial que demandassem grande capacidade de sequenciamento de DNA. Após debate sobre qual deveria ser a espécie de eucalipto a ter seu genoma sequenciado, se *E. globulus*, de interesse de países como Austrália, Chile e Portugal ou *E. grandis*, de interesse do Brasil, África do Sul e demais países tropicais, foi decidido que a árvore a ser sequenciada seria da espécie *Eucalyptus grandis*. Isto ocorreu não só em função da forte participação de pesquisadores daqueles dois países na coordenação da proposta, mas principalmente em vista das várias contribuições técnicas que o Projeto Genolyptus faria em termos de bases de dados de genes sequenciados, mapas genéticos e bibliotecas de fragmentos longos de DNA para sequenciamento Bacterial Artificial Chromosomes (BAC). Além destas importantes contribuições feitas através do Genolyptus, a escolha de *E. grandis* foi, por fim, alavancada pela disponibilização, por parte da empresa brasileira Suzano Celulose, de uma árvore adulta, denominada Brasuz1 (Brasil Suzano S₁), derivada de uma geração de autofecundação (S₁) que, pela sua maior homozigose, facilitaria consideravelmente o trabalho de bioinformática de montagem do genoma.

O interesse do JGI em apoiar projetos de sequenciamento era movido pela importância do organismo alvo para questões de sustentabilidade energética mundial, e principalmente pela dimensão, interesse e contribuições ao projeto que a comunidade científica internacional, que viria a se beneficiar do acesso à sequência do genoma completo do organismo, poderia fazer. Ao longo de 2006, o trabalho da rede Eucagen foi no sentido de listar as contribuições efetivas que cada participante poderia fazer e buscar apoio formal de vários países, organizações públicas e privadas a esta iniciativa. O Brasil, e certamente a Embrapa, pela coordenação que já existia com o Projeto Genolyptus, teve papel central neste processo. Após passar com sucesso por uma etapa de pré-propostas em Janeiro de 2007, uma proposta final foi submetida em Março de 2007, liderada pelos pesquisadores Alexander Myburg da University of Pretoria, África do Sul, Dario Grattapaglia da Embrapa e Jerry Tuskan do Oak Ridge National Laboratory do Departamento de Energia dos EUA. Cartas formais de apoio e colaboração de mais de 100 membros da rede Eucagen de cerca de 20 países foram anexadas à proposta aprovada em primeiro lugar ao competir com mais de 120 outras propostas submetidas ao JGI.

Em Janeiro de 2008 foi oficialmente iniciado o projeto de sequenciamento com a entrega das amostras de DNA e RNA realizada pessoalmente pela equipe da Embrapa à coordenação do projeto no JGI. Sendo o JGI uma instituição financiada por recursos públicos, por obrigação de mandato, todos os dados de sequenciamento são públicos e foram, portanto, sendo depositados diariamente no banco de dados GenBank pelo JGI, à medida que iam sendo produzidos. O sequenciamento do genoma do eucalipto foi realizado exclusivamente com a tecnologia Sanger. Ao mesmo tempo que o sequenciamento genômico pela abordagem “shotgun” era conduzido nos laboratórios do JGI, o sequenciamento das extremidades de cerca de 120 mil clones BAC da biblioteca fornecida pelo Genolyptus foi realizado pelo centro de sequenciamento da Stanford University, também associada ao JGI. Em 2011, uma primeira versão da montagem e anotação do genoma de *Eucalyptus grandis*, ainda com cobertura preliminar de 4,5 vezes, foi disponibilizada para o mundo pelo JGI na plataforma de dados genômicos Phytozome.com (Goodstein et al., 2012). Nesta plataforma todos os genomas de plantas sequenciados pelo JGI ou com participação direta do JGI, são depositados para acesso público. Após um intenso trabalho de análise realizado por diversos grupos internacionais participantes da rede EUCAGEN, a montagem e anotação final da sequência de genoma de *Eucalyptus grandis*, disponível para a comunidade internacional na plataforma Phytozome com tamanho total de 605,9 Mbp, foi finalmente submetida para publicação em Setembro de 2013 e publicada na prestigiosa revista *Nature* (Myburg et al., 2014) em um artigo envolvendo 55 autores de 18 países. Este foi o primeiro genoma de uma planta sequenciado com a participação do Brasil na liderança, formatação e execução do projeto.

A disponibilização pública da sequência completa do genoma de *Eucalyptus grandis* representou um importante avanço não apenas para a comunidade científica internacional, mas também para os programas operacionais de melhoramento de *Eucalyptus*. A partir de uma referência genômica acessível aos vários grupos de pesquisadores no mundo, na qual o endereço genômico específico de marcadores SNPs e genes anotados é conhecido e facilmente compartilhado, tem sido possível a geração de informações e conhecimento sobre a evolução e genômica comparativa, redes de expressão e regulação gênica e posicionamento de associações entre marcadores e fenótipos. Uma coletânea de 14 artigos científicos contribuídos pelos diversos grupos que participaram diretamente ou indiretamente da produção do genoma, explorando seus diversos aspectos foi publicada em 2015, na revista científica de impacto *New Phytologist*. Duas foram as contribuições do LGV-Embrapa nesta coleção. A primeira foi o desenvolvimento da plataforma de genotipagem de SNPs EUChip60K descrita anteriormente (Silva-Junior et al., 2015), que efetivamente revolucionou a capacidade de análise genética no gênero, permitindo a implementação de estudos de associação e seleção genômica por diversas universidades e empresas florestais em todo o mundo. Em um segundo estudo, de cunho mais fundamental, foram examinados os

padrões genômicos de recombinação, desequilíbrio de ligação e diversidade nucleotídica, por meio da genotipagem de SNPs e resequenciamento do genoma de 30 árvores de *E. grandis*. Foi construído o mapa de ligação mais denso até hoje, para espécies de *Eucalyptus* com 4.396 marcadores SNPs e revelado, pela primeira vez, que o padrão real de desequilíbrio de ligação, quando analisado no genoma todo, era bem mais extenso do que se achava, fato este com importante impacto na capacidade de realizar predição genômica à espécie. Além disso, verificou-se que a mutação, como processo evolutivo, teve papel mais importante do que a recombinação na moldagem da variabilidade genética do eucalipto. Gráficos de recombinação ancestral cromossômica permitiram estimar o tamanho efetivo e, com isso, propor uma datação molecular para o tempo até o ancestral comum mais recente de *E. grandis* estimado entre 1,7 a 4,8 milhões de anos atrás. Esta datação por sua vez forneceu uma importante contribuição molecular em nível de genoma completo para o esclarecimento da controvérsia que persistia quanto à datação da origem dos eucaliptos baseada, até então, na análise de apenas pequenos trechos de sequência de ITS e DNA de cloroplasto. A análise ampla do genoma (*genome-wide*) deu suporte para a origem mais recente do gênero, de 5 a 10 milhões de anos, consistente com registros fósseis que colocam a radiação de espécies do subgênero *Symphomyrtus* em 2 a 5 milhões de anos atrás, questionando a origem mais antiga proposta por alguns grupos de 26 a 38 milhões de anos. A capacidade de compartilhamento de SNPs entre espécies de *Eucalyptus* e a colinearidade genômica observada entre espécies também contribuem com evidências adicionais em suporte à hipótese de uma origem mais recente do gênero *Eucalyptus* (Silva-Junior; Grattapaglia, 2015).

Aplicações operacionais da análise genômica: a Embrapa implementa e transfere metodologias para o setor florestal

Ao longo dos últimos 25 anos a Embrapa fez importantes contribuições não apenas em nível nacional, mas também internacional, demonstrando eficiência do uso de marcadores moleculares na resolução de questões de quantificação de variabilidade de populações de melhoramento, determinação de identidade clonal e investigação de parentesco. Vários tipos de marcadores foram utilizados para isso, ao longo dos anos. Embora marcadores RAPD e AFLP forneceram inicialmente um poder razoável de discriminação, o seu modo de herança dominante não era adequado para a determinação inequívoca de identidade entre indivíduos e para investigações mais refinadas de relacionamento genético em populações. Com o desenvolvimento pioneiro de marcadores microssatélites codominantes, com maior conteúdo informativo e elevada

reprodutibilidade realizado pela Embrapa a partir de 1996, marcadores dominantes foram logo abandonados. Conjuntos de dezenas de marcadores microssatélites são hoje utilizados em rotina para estas aplicações por serem acessíveis, em termos de necessidade de equipamento, apresentarem baixo custo, ter elevada capacidade de discriminação e fornecerem resultados em poucos dias.

A partir de 2013, com o desenvolvimento da plataforma de genotipagem de SNPs, uma ferramenta adicional se tornou disponível, fornecendo um número consideravelmente maior de marcadores para aplicações que demandam uma análise mais ampla do genoma, como, por exemplo estudos de associação e seleção genômica. Embora marcadores SNPs possam vir a substituir microssatélites, SNPs ainda demandam equipamentos mais caros e sofisticados, a análise demanda um mínimo de amostras analisadas de cada vez que, em geral, é de um mínimo de 24 a 384. Além disso, como a geração de dados é realizada no exterior, a entrega de resultados leva algumas semanas. A perspectiva, portanto, é que ambos os tipos de marcadores, microssatélites e SNPs continuem convivendo, ambos muito úteis, dependendo da aplicação desejada e do prazo com o qual se pretende receber os resultados. A seguir são brevemente detalhadas algumas das principais aplicações operacionais de marcadores moleculares em programas de análise populacional e melhoramento de eucalipto, especificamente destacando as contribuições do LGV-Embrapa e suas colaborações internacionais. Para um maior aprofundamento neste tema algumas referências são sugeridas (Grattapaglia, 2000; Grattapaglia; Kirst, 2008; Grattapaglia et al., 2012).

Identificação de clones de eucalipto

A identificação correta de árvores, seja clones usados em plantios operacionais ou genitores em pomares de sementes, é, no momento, a aplicação de marcadores moleculares mais comum. Marcadores são rotineiramente utilizados por diversas empresas em todo o mundo, como parte de seus programas de controle de qualidade de processos de propagação, principalmente quando estes são terceirizados. A certificação genética do material propagado é um aspecto crucial das grandes operações de plantios clonais, principalmente em sistemas verticalmente integrados de produção, nos quais é feito um planejamento detalhado de utilização de matéria-prima por parte das fábricas, com base na disponibilização de determinados clones ou famílias com determinadas propriedades químicas e físicas da madeira. Dada à escala destas operações que, frequentemente, têm por desafio alimentar os programas anuais de plantio com milhões de mudas de diferentes clones, erros de identificação podem afetar seriamente a produtividade prevista. A identidade clonal correta também tem implicações importantes em programas de produção de sementes em pomares ou via polinizações controladas. Erros podem vir a impactar negativamente os ganhos esperados em cada ciclo de melhoramento.

Da mesma forma que na análise forense de seres humanos e animais domésticos, hoje o método-padrão para a identificação de clones de *Eucalyptus* se baseia na utilização de microssatélites em sistemas multiplex de coamplificação de 10 a 20 marcadores analisados em sequenciador automático, que permite alta resolução e reprodutibilidade dos dados genéticos (Faria et al., 2011a). A identidade de duas amostras geneticamente indistinguíveis é declarada com base em uma verossimilhança na qual a probabilidade de se observar os dados genéticos, condicional à hipótese das duas amostras serem oriundas do mesmo clone, é comparada à hipótese alternativa, ou seja, de que as duas amostras são oriundas de clones diferentes. Além disso, a reprodutibilidade e precisão da declaração dos genótipos permite a comparação precisa de resultados entre laboratórios e épocas diferentes de análise. O elevado polimorfismo de microssatélites em *Eucalyptus* foi logo demonstrado ao permitir a discriminação de 192 indivíduos de uma população de melhoramento de *E. grandis*, utilizando apenas três marcadores. A probabilidade de dois indivíduos geneticamente distintos terem o mesmo genótipo multiloco em seis microssatélites, devido simplesmente ao acaso, foi menor que 1 em 2 bilhões (Kirst et al., 2005). Estudos posteriores otimizaram sistemas de análise mais potentes com 10 a 20 microssatélites incluindo marcadores baseados em tri e tetranucleotídeos utilizados atualmente em rotina de análise (Faria et al., 2011a). Marcadores SNPs também podem ser utilizados com sucesso para a geração de perfis genéticos individuais visando análises de identidade e parentesco com a vantagem adicional de maior precisão na declaração de genótipos, uma vez que este procedimento é totalmente automatizado e menos sujeito a variações de condições de eletroforese que podem impactar a análise com microssatélites (Telfer et al., 2015). Entretanto, tendo em vista o bialelismo do SNP e conseqüente menor conteúdo informativo, estudos mostraram que são necessários cerca de sete vezes mais marcadores SNPs do que microssatélites, para alcançar o mesmo poder de discriminação em *Eucalyptus* (Correia et al., 2011). Apesar do menor conteúdo de informação de cada SNP individualmente, a utilização de painéis de algumas dezenas de SNPs selecionados para elevado polimorfismo tornar-se-á uma alternativa cada vez mais atraente para aplicação operacional em programas de melhoramento.

Proteção varietal de clones de eucalipto

Material genético na forma de clones elite representa hoje um ativo essencial para a manutenção e o crescimento da competitividade de uma indústria de base florestal. Parece evidente, portanto, que a proteção legal contra a potencial utilização e/ou comercialização indevida por terceiros é um aspecto preventivo importante que não pode ser ignorado por um setor que obteve e continuará a obter ganhos extraordinários com base na genética e no melhoramento. A identificação correta de clones de eucalipto tem implicações importantes no intercâmbio e licenciamento de

clones e execução de contratos de terceirização de produção de mudas. No Brasil, a lei de proteção de cultivares nº 9.456, sancionada em 1997, representou um importante avanço no sentido de proteger e estimular o contínuo desenvolvimento de materiais genéticos superiores de culturas agrícolas, frutíferas e florestais. A Embrapa liderou ativamente a compilação e validação das instruções para a execução dos ensaios de distinguibilidade, homogeneidade e estabilidade (DHE) de cultivares de eucalipto, por meio da instrução normativa publicada em fevereiro de 2002 pelo Serviço Nacional de Proteção de Cultivares (SNPC), órgão do Ministério da Agricultura, Pecuária e Abastecimento. Espécies do gênero *Eucalyptus* foram as primeiras espécies florestais incluídas na lista daquelas cujos cultivares passaram a ser protegidos.

A proteção de clones superiores de eucalipto envolve hoje a utilização de um conjunto selecionado e validado de 36 marcadores morfológicos, principalmente em folhas, flores e casca. Esses descritores foram selecionados por atender aos requisitos básicos de estabilidade e pouca influência ambiental. Entretanto, além das dificuldades de caracterização morfológica, às vezes encontradas em razão da mudança de fase fisiológica entre o estado juvenil e o adulto, clones elite de eucalipto muitas vezes apresentam coancestralidade. É comum, por exemplo, que clones elite sejam selecionados dentro da mesma família de irmãos completos ou meios-irmãos. A baixa variabilidade observada em características morfológicas dificulta, assim, a clara distinção entre os indivíduos, tornando-se necessária a utilização de outros descritores de maior poder de resolução.

Em vista das potenciais limitações inerentes ao processo de descrição morfológica, já na instrução normativa que regulamentou a proteção de cultivares de eucalipto, marcadores moleculares microssatélites foram incluídos como descritores complementares facultativos (Grattapaglia et al., 2003). Isso representou um avanço importante no panorama mundial da proteção varietal de espécies florestais perenes. Interessante observar que, apesar de serem apenas facultativos, perfis genéticos de 10 a 15 marcadores microssatélites têm sido rotineiramente incluídos nos pedidos de proteção e registro de clones de eucalipto no Brasil por diversas empresas. Isso se deve, por um lado, à percepção que os melhoristas têm do valor de um clone-elite para o negócio florestal e, por outro, à posição inovadora que o Brasil ocupa internacionalmente na pesquisa e desenvolvimento em genética, melhoramento e análise molecular do eucalipto. A perspectiva para os próximos anos deverá ser, portanto, de um aumento significativo na solicitação de proteção de clones superiores por parte das várias empresas florestais. Isso, sem dúvida, poderá contribuir para a melhor organização e documentação do material genético proprietário existente no país e facilitar contratos de intercâmbio de clones entre os programas de melhoramento e produção das indústrias. Embora refinamentos adicionais nos procedimentos de seleção e utilização de marcadores sejam feitos continuamente, o Brasil, por meio do trabalho da Embrapa já está hoje na vanguarda da utilização desta tecnologia genômica à proteção de variedades florestais.

Caracterização de coleções de germoplasma e populações de melhoramento

Espécies de eucalipto, em geral, encontram-se em estágio de melhoramento em comparação à maioria das culturas agrícolas. Existe ainda ampla flexibilidade sobre quais materiais genéticos incluir nas populações de melhoramento que, em regra geral, são constituídas por dezenas ou centenas de árvores geneticamente não-relacionadas, selecionadas diretamente de populações naturais, testes de procedência/progênie ou mesmo clones elite utilizados comercialmente. Marcadores moleculares têm sido utilizados nessas etapas iniciais dos programas para caracterizar e quantificar os níveis e a organização da variabilidade genética existente como recentemente demonstrado em estudos com *E. urophylla* e *E. pilularis* (Silva et al., 2015, 2018). Esses dados podem ser utilizados para melhorar a estrutura das populações de melhoramento, para infundir material novo em populações e decidir sobre a seleção, o enriquecimento ou a eliminação de indivíduos.

A redução de variabilidade genética em populações de melhoramento é essencialmente o resultado da seleção direcional, aliada ao incremento da endogamia, devido a cruzamentos aparentados e deriva genética. A restrição da contribuição genética de uma mesma família para a população selecionada, i.e., número de indivíduos selecionados por família, promove a manutenção da variabilidade, mas, por outro lado, tipicamente reduz o ganho genético potencial. A restrição do número de famílias pode aumentar o ganho, porém reduz a variabilidade genética. Em sistemas de *pedigree* incompleto, frequentemente usados na forma de famílias de meios-irmãos para avançar gerações de melhoramento de espécies florestais, um sistema de monitoramento genético baseado em marcadores para acompanhar os níveis de variabilidade genética ao longo dos diferentes ciclos de um programa de melhoramento é particularmente útil, uma vez que permite flexibilidade maior durante o processo de seleção e auxilia no controle da taxa de incremento da endogamia. É importante lembrar, entretanto, que a informação fornecida pela análise de um conjunto de marcadores moleculares anônimos, independentemente do tipo utilizado, se restringe à caracterização dos aspectos de ancestralidade, heterozigosidade observada e divergência ou distância genômica entre as árvores. Não existe, a princípio, qualquer correlação entre a informação fornecida e o potencial desempenho dos indivíduos ou seus descendentes para características de importância silvicultural ou econômica, como muitas vezes equivocadamente se espera, a menos que um delineamento experimental específico em termos de estrutura populacional e uma densidade de marcadores elevada sejam utilizados conforme descrito mais à frente na seção sobre seleção genômica. Isso ocorre, pois, dados experimentais que vêm sendo levantados têm mostrado que as regiões genômicas envolvidas no controle de características quantitativas, além de serem da ordem de centenas ou mesmo milhares, cada uma delas contribuindo com

efeito muito pequeno, são distribuídas por todo o genoma e potencialmente com forte interação com o ambiente.

Determinação precisa de parentesco e estudos de fluxo gênico

O conhecimento das taxas de fecundação cruzada *versus* autofecundação é essencial para a manutenção de níveis adequados de variabilidade genética nas populações para ganhos contínuos, bem como para fundamentar corretamente as estimativas de parâmetros genéticos em testes de progênie de polinização aberta. Marcadores isoenzimáticos foram originalmente utilizados para estimar o sistema preferencial de cruzamento dentro de populações ou pomares de semente, logo substituídos por marcadores de DNA dominantes na década de noventa (Gaiotto et al., 1997; Gaiotto; Grattapaglia, 1997) e hoje por marcadores microssatélites que permitem análises mais refinadas e precisas de parentesco e fluxo gênico. Microssatélites vêm sendo utilizados rotineiramente por diversas empresas florestais para controle de qualidade do parentesco de descendentes em programas de polinização controlada (Telfer et al., 2015), confirmar eventos de autofecundação em projetos de desenvolvimento de linhagens, estimar o fluxo gênico de pólen externo contaminante em um pomar, bem como detectar eventos de fluxo gênico a distância. Marcadores microssatélites foram utilizados com sucesso para estudar o fluxo de pólen e sucesso reprodutivo em povoamentos clonais fragmentados de eucalipto, mostrando que a polinização ocorre predominantemente até distâncias de menos de 200 m., compatível com o vôo de abelhas polinizadoras, embora níveis baixos de sucesso de polinização sejam esperados em distâncias maiores (Silva et al., 2015). Em um estudo subsequente marcadores microssatélites permitiram verificar que a maior taxa de fluxo de pólen a partir de árvores geneticamente modificadas (GM) de eucalipto ocorre a distâncias curtas de até 15 metros, diminuindo rapidamente até a maior distância avaliada de 650 m., semelhante ao observado em estudos de fluxo de pólen de eucalipto não-GM, fornecendo uma informação importante na definição de estratégias de biossegurança para contenção de fluxo transgênico mediado por pólen (Silva et al., 2017).

A determinação precisa de paternidade de árvores utilizando marcadores moleculares pode também ser utilizada como uma estratégia de seleção retrospectiva de genitores de maior capacidade geral ou específica de combinação, visando melhorar o desempenho médio dos plantios derivados de um pomar de sementes. Marcadores microssatélites permitem identificar exatamente quais foram os genitores de um pomar que geraram maior número de descendentes geneticamente superiores e, com base nisso eliminar genitores menos interessantes de um pomar, ou realizar cruzamentos controlados entre os genitores que apresentaram melhor capacidade específica de combinação. Esta estratégia de utilizar análise de parentesco no melhoramento foi originalmente proposta e utilizada em *Eucalyptus* (Grattapaglia et al., 2004) e

mais tarde ampliada para uma abordagem denominada *breeding without breeding* que incluiu a estimativa de parâmetros genéticos com base no relacionamento genético entre indivíduos determinado pelos marcadores moleculares (El-Kassaby; Lstiburek, 2009). O sucesso dessa abordagem depende, entretanto, do número total de genitores envolvidos na análise e do grau de relacionamento genético entre eles. Quanto maior o número de genitores, e quanto maior o relacionamento entre eles, maior o número de marcadores necessários para se declarar categoricamente o parentesco correto. Com o advento da possibilidade de genotipar milhares de marcadores SNPs em eucalipto, hoje é possível reconstruir parentesco com alta precisão mesmo envolvendo um grande número de genitores potenciais. Esta abordagem de reconstrução de parentesco por meio da geração de uma matriz de relacionamento genômico Genomic Relationship Matrix (GRM) com base em marcadores SNPs efetivamente estimando o parentesco realizado entre indivíduos tem sido cada vez mais utilizada em programas avançados de melhoramento de eucalipto por algumas empresas no Brasil (Resende et al., 2017; Tan et al., 2018; Lima et al., 2019). O uso de GRM em substituição à matriz de relacionamento esperado também chamada de matriz do numerador do coeficiente de parentesco, permite desemaranhar as variâncias aditivas das não-aditivas com precisão e assim obter estimativas mais fidedignas de parâmetros genéticos tais como herdabilidade e contribuições relativas das duas classes de variâncias genéticas (Grattapaglia et al., 2018).

A Embrapa transita do mapeamento genético para a Seleção Genômica de eucalipto

Neste tópico são apresentadas as contribuições do LGV-Embrapa visando aplicações tecnicamente mais complexas, ao tentar conectar a variação de sequência no DNA com a variação de características quantitativas de interesse florestal, por meio de mapeamento de Quantitative Trait Loci (QTLs), mapeamento de associação e, nos últimos dez anos, com a predição e seleção genômica. O intuito deste esforço é o de auxiliar diretamente a seleção mais rápida e precisa de árvores superiores em programas de melhoramento. O melhoramento genético de eucalipto envolve ciclos recorrentes de seleção, cruzamento e testes para obter sementes geneticamente melhoradas ou material clonal-elite, maximizando o ganho genético por unidade de tempo ao menor custo possível. Tendo em vista os ciclos relativamente longos de vida e florescimento tardio do eucalipto, o progresso e o sucesso de um programa de melhoramento são fortemente dependentes do tempo necessário para completar uma geração. No início dos anos 90, quando as tecnologias de análise genômica se tornaram mais acessíveis, a Seleção Assistida por Marcadores (SAM) foi imediatamente vista como uma

ferramenta potencialmente poderosa para superar alguns destes desafios (Grattapaglia et al., 1992). No entanto, o potencial da SAM no melhoramento florestal foi questionado com base no estado de equilíbrio de ligação, na qual se encontraria uma população de melhoramento grande com histórico recente de domesticação. Além disso, questões relativas à estabilidade da expressão dos locos controladores de características quantitativas (QTLs) ao longo do tempo, em diferentes locais e em diferentes backgrounds genéticos, seria um forte impedimento (Strauss et al., 1992).

Mapeamento de QTLs e mapeamento de associação

Apesar destes argumentos iniciais bem fundamentados, que mais tarde se provaram corretos em grande parte, vários trabalhos de construção de mapas genéticos e mapeamento de QTL foram realizados em espécies de *Eucalyptus*, revisados em (Grattapaglia et al., 2012). A premissa nesta fase inicial da pesquisa genômica, também adotada no Projeto Genolyptus, foi a que, pela abordagem de dissecação genética, seria possível mapear e estimar os efeitos de todos os genes ou QTLs relevantes para características como crescimento e qualidade da madeira durante a vida da árvore, em todas as populações e ambientes (Grattapaglia, 2004; Grattapaglia; Kirst, 2008). Diversos estudos descrevendo mapeamento de QTLs em espécies de *Eucalyptus* foram relatados utilizando uma ou mais populações biparentais envolvendo o cruzamento entre genitores heterozigotos e análise na configuração pseudo-cruzamento teste (Grattapaglia; Sederoff, 1994). No âmbito do Projeto Genolyptus, QTLs para características de crescimento, qualidade da madeira e tolerância a doenças foram mapeados em diferentes famílias biparentais. Um importante foco deste trabalho foi desenvolvido em forte colaboração com a equipe da Universidade Federal de Viçosa visando a resistência a doenças que apresentavam uma herança aparentemente menos complexa e, portanto, mais propensas à identificação de locos de maior efeito, além da relevância crescente de doenças do ponto de vista silvicultural. O primeiro QTL de grande efeito mapeado em *Eucalyptus* foi na espécie *E. grandis* para resistência à ferrugem do eucalipto causada por *Puccinia psidii* (Junghans et al., 2003), mais tarde complementado por estudos de validação do QTL em diferentes pedigrees (Mamani et al., 2010) e identificação de interação entre QTLs (Alves et al., 2012). A estes trabalhos seguiram-se outros semelhantes localizando QTLs em *Eucalyptus* para resistência à desfolha causada por *Cylindrocladium pteridis* (Zarpelon et al., 2011), mancha-de-Colonectria causada por *Calonectria pteridis* (Zarpelon et al., 2015) e murcha de *Ceratocystis* causada por *Ceratocystis fimbriata* (Rosado et al., 2016).

Diversos experimentos de mapeamento de QTLs para características de crescimento e qualidade da madeira também foram realizados ao longo do Projeto Genolyptus, os quais somente foram publicadas no âmbito de teses de mestrado com exceção de um estudo que localizou um loco envolvido em florescimento precoce

(Missiaggia et al., 2005). Tentativas de transferência da informação de posicionamento de QTLs gerada em uma família biparental para uma outra família, para a prática da seleção assistida não mostraram, entretanto, o sucesso esperado.. Além disso, verificou-se que a magnitude do efeito estimado de um QTL, além de superestimado, variava de acordo com o desenho experimental, background genético e ambiente. Estes resultados, em linha com resultados que vinham sendo publicados por outros grupos no mundo, mostraram, portanto, que as abordagens de mapeamento de QTL em famílias biparentais não seriam capazes de capturar variação genética suficiente que pudesse resultar na implementação efetiva da SAM no melhoramento genético operacional do eucalipto (Grattapaglia et al., 2009).

Como alternativa ao mapeamento genético em famílias, estudos posteriores também foram desenvolvidos para avaliar a abordagem de genética de associação em populações não estruturadas como uma forma de superar as limitações intrínsecas da SAM baseada em QTLs mapeados em famílias biparentais. Estes estudos foram inicialmente desenvolvidos com base na análise de SNPs localizados em genes candidatos, mostrando-se totalmente ineficientes do ponto de vista de utilização posterior da informação gerada, na medida que partiam de uma premissa que aqueles genes selecionados para estudo seriam de fato genes com efeito relevante na característica, premissa esta que se mostrou falsa. Com o advento de marcadores DArT e em seguida SNP que permitiram a análise em escala genômica ampla, uma série de estudos de GWAS (*Genome-Wide Association Studies*) foram realizados pelo LGV-Embrapa em colaborações com empresas florestais. Diferentes abordagens analíticas e meta-análises amplas foram utilizadas no intuito de avaliar, em detalhe, a possibilidade de detectar associações discretas entre marcadores SNPs e características complexas de crescimento e qualidade da madeira para posterior uso no melhoramento (Cappa et al., 2013; Muller et al., 2017, 2019; Resende et al., 2017). Os resultados, entretanto, mostraram que, além do número muito pequeno de associações significativas detectadas, a magnitude do efeito de cada uma delas era muito pequena, raramente excedendo 1% a 5% da variância fenotípica de forma que pouco ou nenhum impacto teria para a SAM. Estes resultados corroboraram a natureza complexa das características quantitativas de importância econômica em eucalipto, e deram suporte experimental ao modelo infinitesimal de Ronald Fisher, como sendo a melhor aproximação teórica para a arquitetura genética de características quantitativas também no eucalipto (Fisher, 1918). Com efeito, com exceção de algumas poucas características qualitativas ou monogênicas simples em culturas anuais, raros exemplos de mutações de grande efeito em espécies frutíferas e defeitos genéticos recessivos em animais domésticos, esta tem sido a conclusão geral não só para espécies florestais, mas também para a maioria, se não a totalidade, de espécies de plantas e animais para as quais se pretende utilizar a genômica no melhoramento. A menos que seja possível capturar uma grande proporção da variação genética para as características alvo, as

abordagens convencionais da genética quantitativa provavelmente continuariam a ser mais eficientes para o avanço do melhoramento. Discussões mais detalhadas deste tema são apresentadas em alguns artigos tais como os de Bernardo (2008), Goddard e Hayes (2009), Grattapaglia e Resende (2011).

Esta constatação causou uma mudança de paradigma na perspectiva de utilização da análise genômica no melhoramento de animais e plantas. Ao invés de tentar descobrir, validar e somente então utilizar as associações marcador-característica, passou-se a lidar com o agregado do efeito do genoma inteiro, da mesma forma que a genética quantitativa sempre fez, com a diferença que o efeito do genoma inteiro sobre a característica alvo é estimado com o uso de grande quantidade de marcadores moleculares distribuídos em todo o genoma. Esta revolução só foi possível com o avanço na capacidade de desenvolver e analisar simultaneamente, de forma automatizada e econômica, um grande número de marcadores SNPs. Esta abordagem foi denominada Seleção Genômica (GS) ou Seleção Genômica Ampla (SGA).

Seleção Genômica: princípios básicos

Inspirado no modelo infinitesimal de Fisher e na prática convencional da genética quantitativa, o conceito de se utilizar o efeito agregado de todo o genoma sobre a característica alvo para a prática de seleção direcional vinha sendo proposto ao final dos anos 90, quando tecnologias de genotipagem de marcadores SNPs se tornavam cada vez mais acessíveis técnica e financeiramente. Com base na genotipagem em escala genômica, viu-se que seria possível praticar a “seleção genômica total”, assegurando que marcadores estariam em completa associação (ou seja, em desequilíbrio de ligação) com grande parte dos locos controladores da característica e, portanto, capturariam a maioria dos efeitos genômicos subjacentes à característica (Haley; Visscher, 1998). No entanto, foi o trabalho seminal de Meuwissen e colaboradores (Meuwissen et al., 2001) que antecipou e demonstrou por simulação que a seleção baseada em valores genéticos preditos a partir de marcadores poderia aumentar consideravelmente o ganho genético em programas de melhoramento.

Do ponto de vista conceitual, seja a SAM bem como a Seleção Genômica (SG), também chamada de Seleção Genômica Ampla (SGA) (do inglês GWS Genome-wide Selection), começam por estabelecer associações entre genótipos de marcadores moleculares discretos e fenótipos contínuos em populações relevantes. No entanto, as duas abordagens são fundamentalmente diferentes após essa etapa. A SAM normalmente tem como alvo a descoberta e caracterização de QTLs individuais em uma ou mais populações biparentais ou painéis de mapeamento de associação usando testes rigorosos de significância, com o objetivo posterior de usar tais associações para a seleção. A SG, diferentemente, usa um painel extenso de marcadores cujos efeitos no fenótipo são estimados simultaneamente numa “população de treinamento”, também

chamada de “população de estimação”, sem a aplicação de testes de significância, mas mantendo todos ou grande parte dos marcadores como preditores do fenótipo. Modelos matemáticos preditivos baseados no uso de todos os marcadores, capturando assim todo o efeito do genoma, são então gerados para serem aplicados posteriormente nos candidatos à seleção, para os quais apenas genótipos são coletados e os fenótipos preditos pelo modelo (Figura 6). Assim, na SG, o efeito de um marcador não precisa exceder um limiar de significância rigoroso para ser usado na fase subsequente de seleção. Além disso, os efeitos dos marcadores são estimados em uma população muito maior e mais representativa do que apenas em uma ou algumas famílias biparentais, ou em um painel de indivíduos geneticamente distintos da população de melhoramento.

A população de treinamento de um modelo preditivo envolve pelo menos várias centenas a milhares de indivíduos da população alvo do melhoramento, os quais são genotipados para o painel de marcadores e fenotipados para todas as características de interesse. Os modelos preditivos desenvolvidos para cada característica são validados em uma população de «validação», ou seja, um subconjunto de indivíduos amostrados aleatoriamente da mesma população alvo do melhoramento, mas que não participaram da estimativa dos efeitos dos marcadores. Uma vez que um modelo preditivo fornece uma capacidade preditiva satisfatória, ou seja, uma boa correlação entre os valores genéticos observados e os valores preditos pelo modelo, ele pode ser utilizado na fase de seleção para estimar os valores genéticos genômicos dos candidatos à seleção.

A SG explora tanto o desequilíbrio de ligação entre marcadores e QTLs, bem como o relacionamento genético entre os indivíduos da população de treinamento e os candidatos à seleção. Por evitar a seleção a priori de QTLs, ou seja, dos marcadores que estão associados com a característica, e estimando os efeitos de todos os marcadores em uma população grande e representativa, a SG tende a capturar grande parte da variância genética resultante do efeito combinado do grande número de locos em todo o genoma, consistente com a filosofia do modelo infinitesimal. A SG se tornou, assim, o paradigma para a seleção assistida em plantas e animais e é hoje a tecnologia padrão utilizada amplamente para o melhoramento avançado de bovinos leiteiros e outras espécies animais, como suínos e frangos (Van Eenennaam et al., 2014). Enquanto a SG vinha sendo rapidamente adotada no melhoramento animal a partir de 2006, ela também se tornou um tópico de interesse no melhoramento de culturas anuais (Bernardo; Yu, 2007; Heffner et al., 2009; Lin et al., 2014; Crossa et al., 2017) e espécies florestais para as quais as contribuições teóricas e experimentais da Embrapa foram reconhecidamente as pioneiras em nível mundial (Resende et al., 2008; Grattapaglia et al., 2009; Grattapaglia; Resende, 2011; Resende et al., 2012; Grattapaglia, 2017).

O sucesso da SG depende de vários fatores, incluindo os aspectos teóricos fundamentais previstos pela genética de populações e genética quantitativa, e os aspectos

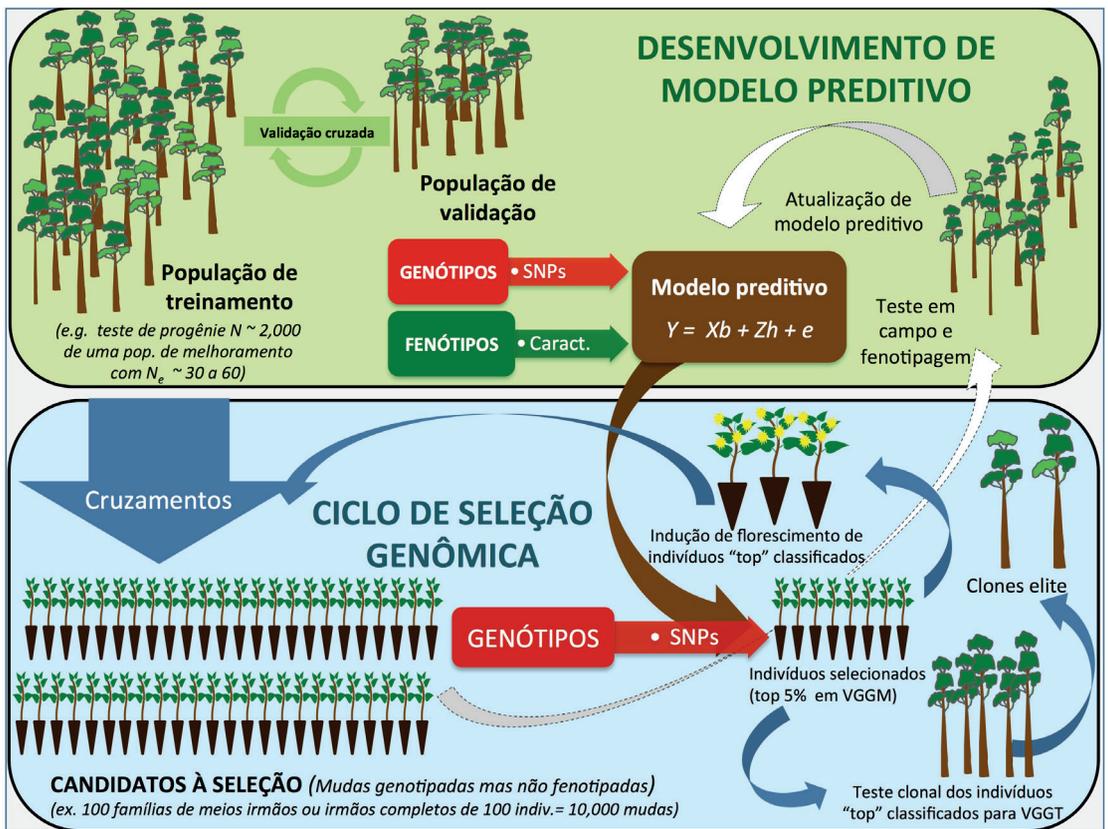


Figura 6. Etapas envolvidas na implementação da Seleção Genômica em um programa de melhoramento de *Eucalyptus*. O painel superior ilustra o desenvolvimento e validação do modelo de predição genômica a partir de uma população de treinamento de ~2.000 indivíduos, genotipada e fenotipada a partir de um teste de progênie com tamanho efetivo N_e entre 30 e 60. O painel inferior ilustra o ciclo da efetiva aplicação da Seleção Genômica começando com a geração do candidatos à seleção a partir da recombinação da própria população de treinamento ou de genitores selecionados dela. Os candidatos à seleção, no estágio de muda jovem, correspondendo a descendentes de famílias de meios-irmãos ou irmãos completos, são genotipados aos SNPs utilizados no modelo e têm seus valores genéticos genômicos de melhoramento (VGGM) e valores genéticos genômicos totais (VGGT) estimados com base nos seus genótipos usando os modelos preditivos desenvolvidos na etapa anterior. Os indivíduos melhor classificados para VGGM são submetidas à indução de florescimento e cruzados para produzir a próxima geração de melhoramento. Os indivíduos melhor classificados para VGGT são clonados e levados a testes clonais de verificação em campo, resultando, mais à frente, em clones recomendados para plantio operacional. As setas brancas pontilhadas indicam a ação de atualização do modelo preditivo, no qual um subgrupo de candidatos é plantado em delineamento experimental e fenotipado até a idade de rotação para fornecer dados adicionais para atualização contínua do modelo preditivo de seleção genômica à medida que gerações avançam. Figura adaptada reproduzida com permissão (Grattapaglia, 2017).

mais práticos e logísticos da alocação de recursos e análise de custo-benefício. Foge à dimensão deste capítulo uma discussão mais ampla sobre os principais fatores nestas duas áreas, lembrando que eles são interligados e interdependentes. A acurácia de um modelo de predição genômica, ou seja, a correlação entre o valor genômico predito e o valor fenotípico é, sem dúvida, o fator chave que terá o maior impacto no sucesso da SG. Quatro são os fatores fundamentais do ponto de vista de genética de populações e quantitativa que afetam a acurácia da predição genômica (Hayes et al., 2009):

- O tamanho efetivo da população (N_e), (isto é, o tamanho de uma população teórica que geraria a mesma quantidade de deriva genética observada na população real sob consideração) e a densidade de genotipagem que, por sua vez, determina a extensão do DL entre marcadores e QTLs. Quanto menor o tamanho efetivo e maior a densidade de marcadores, maior será a extensão do DL e maior será a acurácia preditiva.
- O tamanho e a representatividade da população de treinamento, isto é, o número de indivíduos com fenótipos e genótipos a partir dos quais os efeitos dos marcadores são estimados e o relacionamento desta com os candidatos à seleção. Quanto maior e mais representativa for a população de treinamento e quanto maior o relacionamento desta com os candidatos à seleção, maior será a acurácia.
- A herdabilidade da característica em questão. Quanto maior a herdabilidade, maior será a acurácia. Cabe destacar aqui que, por outro lado, a SG se torna cada vez mais interessante e útil quanto menor a herdabilidade da característica, posto que para características de alta herdabilidade, a vantagem relativa do uso de marcadores moleculares se torna menor, com exceção de características que, mesmo altamente herdáveis, possuem expressão tardia ou mensuração complexa ou cara.
- A arquitetura genética da característica alvo, ou seja, a distribuição dos efeitos do QTL (número de locos envolvidos e o tamanho do seus efeitos). Quanto menos poligênica for a característica ou quanto maior for o número de locos envolvidos com grande efeito, maior será a acurácia.

Do ponto de vista dos aspectos práticos e da relação de custo/benefício no melhoramento, alguns dos fatores mais relevantes que afetam as perspectivas da SG incluem: (1) o tamanho, composição e esforço de genotipagem dedicado à população de treinamento; (2) a plataforma de genotipagem empregada, incluindo a qualidade dos dados resultantes, o custo envolvido e a velocidade com a qual os dados são gerados; (3) a extensão da interação genótipo por ambiente ($G \times A$); e (4) o desempenho no longo prazo dos modelos de predição genômica, incluindo a necessidade de atualização do modelo e o efeito potencial da SG na redução da diversidade, com a aceleração dos ciclos de seleção. Um tratamento mais detalhado foge do escopo deste capítulo, mas pode ser encontrado em revisões recentes sobre a seleção genômica em espécies florestais e de plantas em geral (Isik, 2014; Lin et al., 2014; Grattapaglia, 2017).

Impacto da seleção genômica no melhoramento de *Eucalyptus*

O objetivo do melhoramento genético é maximizar a resposta à seleção por unidade de tempo. O fator tempo é extremamente relevante para o melhoramento florestal, em vista da duração longa de um ciclo de seleção e recombinação. Na expressão de ganho genético, também chamada de equação do melhorista (*breeder's equation*) ($\Delta G = i \cdot r \cdot s_a / L$), a intensidade de seleção (i), a acurácia de seleção (r) e o desvio padrão aditivo (s_a) impactam diretamente no ganho, enquanto o denominador é dado pelo tempo necessário para completar um ciclo de melhoramento (L) com impacto inverso. A SG pode, portanto, aumentar o ganho genético por unidade de tempo em um programa de melhoramento de duas formas. A SG permite aumentar a intensidade de seleção, pois um número maior de plantas podem ser genotipadas e seus fenótipos preditos pela SG do que o número de plantas tipicamente plantadas em testes de progênie. Isto é particularmente relevante para características caras para serem mensuradas (ex. conteúdo de lignina) ou características expressas em idade adulta (ex. densidade da madeira). No entanto, o maior impacto da SG resultará da redução radical do tempo necessário para completar um ciclo. Os fenótipos das plantas candidatas à seleção podem ser preditos em idades ultra precoces, por exemplo, quando as mudas estão com poucas semanas ainda no viveiro, em vez de esperar vários anos até as árvores expressarem os fenótipos. A acurácia da seleção (r) evidentemente também é um dos principais propulsores do ganho genético e se a SG fornecer uma acurácia maior do que a seleção fenotípica convencional, melhor. No melhoramento genético, esta acurácia é fornecida pela raiz quadrada da herdabilidade, isto é, a proporção da variância fenotípica explicada pelos componentes genéticos. Na SG, a acurácia de seleção ou capacidade preditiva, é estimada pela correlação entre o valor genético genômico predito e o valor genético observado. Quando o objetivo de melhoramento é selecionar indivíduos a serem utilizados como clones, o valor genético predito vai envolver não apenas os efeitos aditivos, mas também os componentes não-aditivos.

A implementação de um programa de SG no melhoramento florestal envolve essencialmente duas etapas (Figura 6). A primeira refere-se à construção do modelo preditivo. Nesta etapa é definida uma «população de treinamento» composta por indivíduos que são genotipados com milhares de marcadores de DNA e mensurados para todas as características de interesse, visando assim desenvolver e validar modelos preditivos para serem usados posteriormente, na segunda etapa, na qual a SG é efetivamente colocada em prática. Uma população de treinamento é geralmente amostrada a partir de um teste de progênie existente, derivado do cruzamento de um grupo de genitores elite que foram selecionados para compor a população alvo a ser submetida a processos recorrentes de seleção e recombinação no programa de melhoramento. Normalmente, esse grupo de genitores elite tem um tamanho efetivo populacional (N_e) entre 30 e 100 e um número censitário que variará nesse mesmo intervalo levando

em consideração eventuais parentescos existentes entre os indivíduos. A população de treinamento deverá ter pelo menos 1.000 e preferencialmente 2.000 indivíduos ou mais. Entretanto, vale lembrar que quanto mais indivíduos forem genotipados e fenotipados para treinar os modelos, melhor estimados serão os efeitos dos marcadores e mais robusto se tornará o modelo preditivo.

Na segunda etapa, a SG é efetivamente aplicada aos candidatos à seleção. Estes candidatos à seleção tipicamente correspondem a um grande número de indivíduos de famílias de meios-irmãos ou irmãos completos derivados do cruzamento dos genitores elite originais ou do cruzamento de indivíduos selecionados na própria população de treinamento ou do teste de progênie utilizado para compor a população de treinamento. Esses candidatos à seleção são genotipados e têm seus valores genéticos genômicos estimados, usando o modelo preditivo desenvolvido anteriormente. As plantas melhor classificadas podem ser submetidas à indução de florescimento e cruzadas para gerar a próxima geração e/ou propagadas por clonagem e levadas a testes clonais de verificação, para posterior seleção para a recomendação operacional. Além disso, todos ou apenas um subconjunto aleatório de candidatos à seleção já genotipados pode ser plantado em delineamento experimental e fenotipado até a idade de rotação, para fornecer dados adicionais para a atualização do modelo preditivo a ser utilizado nas gerações subsequentes de SG, mitigando assim a erosão da associação (desequilíbrio de ligação) entre marcadores e QTL, e o decaimento do relacionamento genético entre a população de treinamento e os candidatos à seleção, procurando manter, portanto, a acurácia preditiva da SG ao longo das gerações.

Estudos experimentais e de simulação detalharam as perspectivas promissoras da SG em aumentar a eficiência dos programas de melhoramento florestal. No eucalipto e demais espécies plantadas na forma de clones, a SG não só poderia eliminar o teste de progênie, mas também reduziria o tempo e os custos envolvidos na fase de testes clonais, reduzindo o número de árvores testadas (Figura 7). Em coníferas, a SG combinada à embriogênese somática poderia aumentar consideravelmente a eficiência dos protocolos de propagação clonal atuais, permitindo a pré-seleção de embriões zigóticos com base na genotipagem e estimativa de seu valor genético genômico e a expansão imediata de linhagens embriogênicas elite para o estabelecimento de ensaios clonais ou diretamente em plantações comerciais.

Além do ganho de tempo e intensidade de seleção, uma vantagem pouco citada da SG tem a ver com a possibilidade de realizar com eficiência a seleção simultânea para várias características, em um grande número de indivíduos. É praticamente impossível para qualquer programa de melhoramento concluir uma avaliação rigorosa de todas as características de interesse em todas as árvores de um teste de progênie. Dependendo dos objetivos da produção florestal, as características alvo geralmente incluem crescimento volumétrico, conicidade e retidão do caule, propriedades físicas e químicas da madeira, capacidade de enraizamento, tolerância a pragas, doenças, seca e geada.

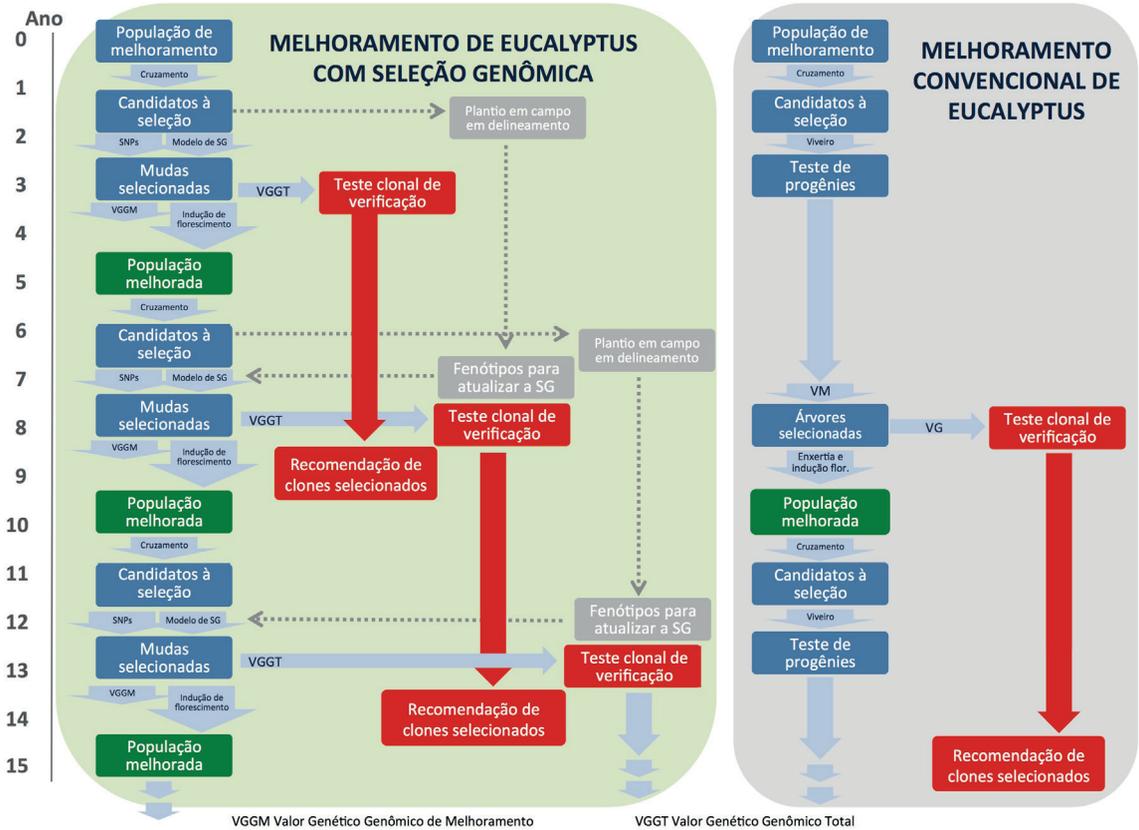


Figura 7. Comparação dos fluxogramas temporais e das etapas de execução de programas de seleção recorrente intra-populacional (caixinhas azuis e verdes) e seleção e recomendação de clones (caixinhas vermelhas) utilizando a seleção genômica (SG) ou a metodologia convencional de seleção fenotípica para o melhoramento de *Eucalyptus*. Ambos os métodos iniciam no ano zero com a mesma população base de melhoramento. No caso da seleção genômica assume-se que os modelos preditivos já foram previamente desenvolvidos. Caixinhas cinza correspondem às ações necessárias para atualização contínua e progressiva dos modelos preditivos de seleção genômica. Figura adaptada reproduzida com permissão (Grattapaglia, 2017).

Em eucaliptos tropicais, por exemplo, mesmo em testes clonais, a avaliação de todas as características de interesse é tipicamente realizada somente nos estágios finais dos testes clonais ampliados para um número muito limitado de clones potenciais que foram pré-selecionados apenas para crescimento e densidade de madeira (Rezende et al., 2014). Devido a razões práticas e de custos, uma abordagem sequencial em tandem é normalmente usada, combinando níveis independentes de seleção para se chegar à classificação (*ranking*) do valor final dos candidatos. Na SG, uma vez que os valores genéticos são preditos para cada característica separadamente, índices de seleção

podem ser usados combinando as predições de todas as características sob análise em um único valor, para cada candidato. A validade desta abordagem de múltiplos passos repousa sobre uma propriedade que o Best Linear Unbiased Prediction (Blup) de qualquer combinação linear de características é igual à combinação linear dos valores preditos por Blup das características individuais (White et al., 2007). Portanto, o efeito líquido da SG seria um aumento notável na intensidade de seleção na fase de viveiro para todas as características, simultaneamente, melhorando consideravelmente a eficiência geral do programa de melhoramento. Além disso, a SG para múltiplas características poderia aumentar significativamente a acurácia da predição para uma característica de baixa herdabilidade ou para características com um número limitado de registros fenotípicos, quando uma característica correlacionada e de alta herdabilidade é disponível.

Seleção genômica: resultados experimentais em espécies de *Eucalyptus*

Os experimentos de SG em espécies de *Eucalyptus* têm procurado espelhar a estrutura das populações reais de melhoramento e adotar delineamentos que respondem, com satisfação às expectativas teóricas de diversidade e relacionamento entre os indivíduos utilizados no treinamento e na validação. Os trabalhos desenvolvidos pela Embrapa, em estreita colaboração com empresas florestais brasileiras, têm buscado avaliar o impacto das diferentes variáveis, particularmente relevantes para o melhoramento florestal, na acurácia da predição, incluindo o nível de relacionamento entre a população de treinamento e de validação, o efeito da interação genótipo*ambiente, e o desempenho de diferentes abordagens analíticas que se baseiam em diferentes premissas a respeito da arquitetura genética das características quantitativas alvo da seleção. De forma geral, as capacidades preditivas relatadas nos diferentes estudos com *Eucalyptus*, bem como com outras espécies florestais, principalmente de coníferas, têm sido muito satisfatórias (Grattapaglia, 2017; Grattapaglia et al., 2018), consistentes com as expectativas de trabalhos iniciais de simulação em espécies florestais (Grattapaglia; Resende, 2011; Iwata et al., 2011).

Dois estudos experimentais de SG foram os pioneiros na área florestal com participação destacada da Embrapa. Um em *Eucalyptus* envolvendo duas populações de melhoramento não relacionadas, com tamanhos efetivos populacionais contrastantes, avaliadas em ambientes diferentes (Resende et al., 2012), e um segundo envolvendo um conjunto de famílias de irmãos completos de *Pinus taeda* clonados e plantados em quatro ambientes diferentes, e mensurados em duas idades diferentes (Resende et al., 2012). Em ambos os estudos, capacidades preditivas equivalentes àquelas obtidas por seleção fenotípica foram estimadas por validação cruzada para características de crescimento e qualidade da madeira. Os resultados sugeriram que ganhos potenciais

de 50% a 200% na eficiência de seleção poderiam ser alcançados. Entretanto, estes estudos também mostraram que as capacidades preditivas dependem fortemente da existência de relacionamento genético entre a população de treinamento e de validação, e são impactadas pelas interações G*A e juvenil-adulto, de modo que as predições serão efetivas quando realizadas no mesmo ambiente onde, e na mesma idade quando, os dados de treinamento foram coletados.

Nos últimos anos diversos estudos experimentais de SG vêm sendo publicados em espécies de eucalipto pela equipe da Embrapa, sempre em colaboração com o setor produtivo no intuito de fomentar a internalização desta nova abordagem técnica nos programas de melhoramento das empresas florestais brasileiras (Lima, 2014; Torres-Dini et al., 2016; Muller et al., 2017; Resende et al., 2017; Tan et al., 2017, 2018; Cappa et al., 2019). Todos estes estudos, sem exceção, corroboraram as perspectivas promissoras da SG para acelerar o melhoramento de eucalipto, ressaltando que todos eles utilizaram a plataforma comum de genotipagem EUChip60K desenvolvida pela Embrapa para espécies de *Eucalyptus* (Silva-Junior et al., 2015). O desenvolvimento e disponibilização deste sistema de genotipagem no domínio público realizada pela Embrapa não apenas acelerou significativamente o acesso à SG por empresas e grupos de pesquisa interessadas em todo o mundo, mas também vem permitindo a execução de estudos comparativos e meta-análises de dados, uma vez que os mesmos SNPs foram utilizados, seguindo a mesma filosofia adotada para plataformas de genotipagem de seres humanos. Por fim, cabe destacar que em função do tempo necessário para obter dados experimentais de duas gerações em espécies florestais, até hoje os estudos ainda não puderam avaliar o desempenho da SG entre gerações, ou seja, usando dados de treinamento de uma geração ancestral para prever e validar fenótipos de descendentes na geração seguinte. Validação cruzada e externa, bem como estimativas de acurácia preditiva foram realizadas exclusivamente dentro da mesma geração, embora um estudo recente tenha tentado observar, com sucesso, a predição entre gerações ao consolidar dados de alguns poucos genitores e descendentes em mesmo conjunto de treinamento (Bartholome et al., 2016). Um detalhamento dos principais resultados e contribuições de todos esses estudos na avaliação das perspectivas da SG para o melhoramento florestal foge ao objetivo deste capítulo, mas foram revisados recentemente (Grattapaglia, 2017; Grattapaglia et al., 2018).

Aspectos operacionais e perspectivas para a implementação da seleção genômica

Conforme descrito anteriormente, a operacionalização de um programa de SG demanda inicialmente o desenvolvimento dos modelos preditivos, utilizando para isso uma população de treinamento devidamente genotipada com alta densidade de marcadores e fenotipada com alta qualidade para o maior número possível de características

de interesse ao programa. Em seguida, já na etapa de utilização efetiva da SG, uma vez realizada a genotipagem de SNPs e a predição genômica de todas as características-alvo para os candidatos à seleção (em geral mudas jovens ainda no viveiro) três possíveis rotas não exclusivas podem ser seguidas com estas plantas:

- As plantas melhor classificadas para valor genético genômico de melhoramento (VGGM) (efeitos aditivos) (genomic breeding value) são encaminhadas imediatamente para a indução de florescimento com os vários métodos disponíveis tais como a indução com fitoreguladores, ou enxertia de topo, e logo recombinadas para rapidamente gerar uma nova população melhorada, completando assim o ciclo recorrente de seleção.
- As plantas melhor classificadas para valor genético genômico total (VGGT) (efeitos aditivos e não aditivos) são selecionadas com intensidade de seleção desejada e clonadas ainda no viveiro e testes clonais de verificação implantados, visando uma seleção final de clones operacionais para plantio comercial.
- Uma vez que o investimento da genotipagem já foi realizado, todas ou apenas um subconjunto de algumas centenas de plantas pode ser instalado em ensaios de campo para fornecer, no devido tempo, dados fenotípicos a serem agregados ao conjunto de dados iniciais da população de treinamento, permitindo assim uma atualização contínua do modelo preditivo.

Com este esquema tentativo sugerido, espera-se que a SG possa eliminar a etapa de testes de progênie, acelerando a conclusão de um ciclo de melhoramento e permitindo a seleção de clones elite mais rapidamente. Com a SG, um ciclo de seleção recorrente de eucalipto tropical poderá durar de 5 a 6 anos, dependendo da capacidade de induzir florescimento, enquanto no melhoramento tradicional ele dura entre 12 e 18 anos dependendo da estratégia e agilidade na condução. Duas gerações de clones elite podem ser recomendadas via SG em 14 a 15 anos, enquanto a seleção fenotípica padrão forneceria apenas uma recomendação no mesmo período (Figura 7). É importante destacar que, mesmo com o uso da SG, embora predições precisas de propriedades de madeira possam ser obtidas nas plantas jovens, um teste clonal de verificação de desempenho no campo, com duração até a idade de rotação (6 anos) deve ser mantido da mesma forma que no melhoramento convencional, para avaliar aspectos de adaptabilidade que a SG pode ter dificuldade em prever. Dependendo do desempenho da SG, entretanto, e à medida que o programa avança gerando mais dados sobre todo o processo pode ser possível até mesmo eliminar ou encurtar o teste clonal de verificação final, acelerando ainda mais a recomendação de novos clones.

A aplicação efetiva da predição genômica em um programa de melhoramento florestal, entretanto, variará de caso a caso, com base em uma análise detalhada da

relação custo/benefício. A SG pode não ser uma opção para programas de melhoramento de eucalipto em pequena escala ou para espécies de menor expressão, mercado limitado ou de nicho, pouca informação genética anterior e orçamentos modestos. Por outro lado, para programas de melhoramento agressivos como aqueles conduzidos por empresas florestais que sustentam grandes operações industriais, parece claro que os ganhos de tempo e intensidade de seleção, juntamente com a racionalização dos testes clonais, representa um avanço importante. A adoção da SG pode, portanto, tornar-se uma vantagem competitiva ao acelerar a recomendação de clones.

Ao finalizar a discussão da SG para *Eucalyptus*, algumas conclusões derivadas dos relatos experimentais de predição genômica podem ser, por fim, mencionadas. Espera-se que essas conclusões auxiliem a tomada de decisão quanto aos investimentos em projetos de pesquisa e implementação operacional da SG nos programas de melhoramento das organizações.

Para a implementação bem-sucedida da SG, todos os estudos experimentais demonstraram que é crucial que os candidatos à seleção sejam geneticamente relacionados à população de treinamento. Vários estudos ratificaram que a remoção do relacionamento genético entre os indivíduos utilizados no treinamento e na validação resultou na redução abrupta da capacidade preditiva. A avaliação do impacto do relacionamento genético sobre a capacidade preditiva, à medida que as gerações de seleção avançam, se afastando, assim, da população original de treinamento é uma área de estudo importante. Neste sentido, estratégias de atualização dos modelos preditivos como as descritas por (Iwata et al., 2011) serão muito importantes para contrabalançar a deterioração esperada do relacionamento genético e do desequilíbrio de ligação resultante da recombinação genética, e com isso manter capacidades preditivas elevadas ao longo do programa.

Praticamente todos os estudos em espécies florestais estimaram capacidades preditivas satisfatórias, usando densidades de genotipagem da ordem de 5.000 a 10.000 marcadores SNPs, muito provavelmente devido ao papel central do relacionamento genético como fator propulsor da capacidade preditiva. Densidades mais elevadas de marcadores, entretanto, são recomendadas para capturar desequilíbrio de ligação efetivo e, com isso, em teoria, sustentar a capacidade preditiva de modelos no longo prazo. Uma plataforma de genotipagem eficiente já é disponível para espécies de *Eucalyptus*. Embora métodos de genotipagem por seqüenciamento possam vir a ser uma alternativa de mais baixo custo, eles ainda precisam melhorar significativamente a qualidade dos dados gerados. As tecnologias de genotipagem em microarranjos de conteúdo fixo de SNPs constituem, hoje, o padrão ouro em termos de qualidade, reprodutibilidade e velocidade de geração de dados e facilidade de uso por parte do melhorista. Os custos de genotipagem caíram significativamente nos últimos anos, e já alcançaram hoje custos essencialmente equivalentes aos custos da genotipagem por sequenciamento. Com o uso crescente de microarranjos comuns por grande número

de organizações, estes custos podem cair mais ainda. O custo de genotipagem, antes considerado um fator limitante para a adoção da SG, não parece ser mais um entrave relevante.

Estudos de SG mostraram que as capacidades preditivas diminuíram quando modelos treinados em um ambiente foram validados em um ambiente diferente, embora a magnitude de tal redução variasse conforme a característica considerada. Basicamente, dados existentes de estudos convencionais da interação G*A ou juvenil-adulto são um indicativo preciso do que esperar da predição genômica entre diferentes ambientes e idades. Como regra geral, boas predições genômicas exigirão modelos de treinamento construídos com fenótipos mensurados no mesmo ambiente e idade onde e quando se pretende utilizá-los nos candidatos à seleção.

Apesar das estimativas encorajadoras de capacidade preditiva até agora, deve-se ressaltar mais uma vez que todos os estudos até o momento avaliaram apenas o potencial da SG dentro de uma mesma geração. Em outras palavras, os conjuntos de treinamento e validação eram contemporâneos. Nenhum resultado de SG em eucalipto em gerações verdadeiramente independentes de genitores e descendentes foi gerado e muito menos o desempenho da SG em gerações mais distantes da população de treinamento. Além disso, o impacto da recombinação e seleção ao longo das gerações sobre a capacidade preditiva ainda não pôde ser avaliado, uma questão que pode se tornar mais relevante à medida que as gerações de SG avançam. Existe um esforço global de vários grupos de pesquisa que trabalham na área para gerar dados experimentais sobre capacidade preditiva efetivamente realizada pela SG entre gerações nos próximos anos. Vários experimentos estão em andamento, principalmente com eucalipto e pinus, visando a avaliação comparativa entre as classificações de indivíduos obtidas com dados genômicos, e aquelas observadas na idade de rotação para características como crescimento e qualidade da madeira.

O melhoramento genético florestal convencional já enfrenta hoje um desafio crescente para a recomendação de clones em vista das flutuações ambientais cada vez mais frequentes e imprevisíveis. Em outras palavras, o ambiente que resultou no desempenho de um indivíduo selecionado em testes de progênie ou teste clonal pode não vir a ser exatamente o ambiente ao qual ele será exposto quando ele vier a ser plantado em florestas comerciais. No caso da SG este desafio pode ser mais crítico ainda. Modelos preditivos construídos para a predição de fenótipos em animais domésticos tendem a manter boa acurácia uma vez que o manejo (nutrição, temperatura, controle de pragas e doenças) pode ser controlado com bastante precisão, o que ajuda a explicar o grande sucesso da SG nesta área. No caso florestal, entretanto, com as flutuações ambientais, é praticamente inviável assegurar que o ambiente alvo no qual os fenótipos utilizados para treinar um modelo preditivo venha a ser o mesmo para os indivíduos selecionados ao longo das gerações. Isto torna o aspecto do re-treinamento dos modelos preditivos uma questão chave para o sucesso da SG

no melhoramento florestal. A implementação eficiente da SG vai demandar a atualização contínua dos modelos pela inclusão de novos dados fenotípicos e genotípicos coletados a cada geração, de modo a capturar os efeitos das flutuações ambientais nos modelos e com isso assegurar a sustentabilidade da capacidade preditiva dos mesmos.

Um método de predição genômica deve fornecer alta precisão, limitar o sobre-ajuste (*overfitting*) do modelo sobre a população de treinamento e capturar LD além de parentesco, para maior estabilidade no longo prazo. Um bom método deve ser fácil de implementar, confiável para uma ampla gama de características e conjuntos de dados e computacionalmente eficiente. Em inúmeros trabalhos realizados com espécies anuais, florestais e de animais domésticos, o método Ridge Regression-BLUP (RR-BLUP) e o seu equivalente G-BLUP foram eficazes em fornecer o melhor compromisso entre o tempo computacional e a eficiência de predição. O RR-BLUP assume que a característica é controlada por muitos locos de pequeno efeito, sugerindo, portanto, que as características quantitativas aderem adequadamente à premissa do modelo infinitesimal. Vários softwares de acesso aberto estão disponíveis para implementar este método e cursos de treinamento sobre seu uso são oferecidos regularmente por várias instituições em todo o mundo. Ainda assim, pesquisas sobre o assunto vêm sendo desenvolvidas para oferecer abordagens mais eficientes de aprendizado de máquina (*machine learning*) e inteligência artificial, para incorporar componentes de variação não aditiva e otimizar métodos de classificação individual de árvores, quando a seleção clonal é o objetivo.

O sucesso da implementação operacional da SG vai depender também, em boa parte, de aspectos logísticos como infraestrutura de viveiro, coleta e rastreamento de amostras, extração e qualificação de DNA em larga escala, contratação de provedores de serviços de genotipagem e análise de dados. Evidentemente a correta identificação de amostras ao longo de todo o processo passa a ser uma questão central para o sucesso da SG. No entanto, vários desses componentes já são usados rotineiramente, embora em menor escala, em operações de viveiro de grandes empresas de base florestal ou podem ser facilmente estabelecidos internamente ou acessados por meio de prestadores de serviços especializados em sistemas de informatizados de gerenciamento utilizados em laboratórios de análise clínica.

Uma análise detalhada da relação de custo e benefício em adotar a GS utilizando metodologias de valor presente líquido é um passo absolutamente necessário antes de considerar a sua implementação. O avanço inovador que a GS causou no melhoramento de bovinos de leite é frequentemente usado como um exemplo do uso economicamente bem-sucedido dessa tecnologia. Entretanto, enquanto bovinos e árvores compartilham o mesmo desafio dos longos tempos de geração, a logística e o custo de testar um touro são substancialmente mais altos do que a avaliação de uma árvore no campo, de modo que o custo da genotipagem em bovinos frente ao ganho na intensidade de seleção é facilmente justificado. Entretanto, a consolidação de um

número grande de amostras, em programas de melhoramento de várias organizações para genotipagem de SNP em contratos de longo prazo, já fornece hoje a economia de escala necessária para viabilizar os custos. Trabalhos de bioinformática e validação de SNPs em microarranjos estão bem avançados no LGV-Embrapa visando o desenvolvimento de disponibilização de uma plataforma de genotipagem de SNPs de segunda geração em parceria com a empresa detentora da tecnologia de microarranjos. Este novo microarranjo de genotipagem de SNPs com um maior número de SNPs, estimado não faixa de 70 a 75 mil, consideravelmente mais otimizado deverá ser oferecido no mercado em breve com custos extremamente competitivos, de ordem de 40% do custo do EuCHIP60K, de forma a tornar a SG uma opção cada vez mais acessível a todo o setor florestal baseado em florestas de eucalipto.

Conclusões

O Brasil é reconhecido mundialmente como um dos líderes no desenvolvimento e aplicação de inovações na área de genética, melhoramento e propagação de eucalipto. Este sucesso ocorreu principalmente em função do espírito empreendedor do setor produtivo de base florestal sempre estimulado pela pressão do mercado a desenvolver soluções locais e buscar inovações tecnológicas para manter e aumentar a competitividade. Outro aspecto fundamental que tem contribuído de forma decisiva para o sucesso do setor florestal produtivo brasileiro tem sido a forte predisposição, seja entre as empresas bem como delas com a Embrapa, as diversas universidades e instituições de pesquisa, em compartilhar conhecimentos e encarar desafios de pesquisa e desenvolvimento de forma colaborativa, numa visão público-privada que, espera-se, possa se tornar cada vez mais frequente em outros setores industriais do País. A evolução que o Brasil teve na área de utilização efetiva de ferramentas genômicas no melhoramento florestal iniciada principalmente com o Projeto Genolyptus e continuada com o Projeto Nexttree e outras iniciativas que, de uma forma ou de outra, surgiram a partir do Genolyptus ou do projeto internacional de sequenciamento do genoma do eucalipto, é um dos exemplos nesta linha, que segue ativa até hoje com diversos projetos em várias organizações. As contribuições da Embrapa, retratadas neste capítulo, notadamente tiveram, e continuam tendo, papel importante para que o país alcançasse este estágio atual.

Para a consolidação das novas ferramentas genômicas como um instrumento adicional efetivo do melhoramento genético, é fundamental, entretanto, que os investimentos brasileiros em pesquisa sejam contínuos, seja mediante de parcerias público-privadas bem como com crescentes investimentos em projetos competitivos dentro das empresas, profundamente interconectados com as estratégias e tecnologias da genética biométrica e melhoramento florestal. A Seleção Genômica no melhoramento

é hoje um tópico quente no melhoramento de plantas e animais em todo o mundo, e representa talvez o melhor exemplo atual de convergência entre as tecnologias genômicas de alto desempenho e os conceitos fundamentais da genética quantitativa (Grattapaglia et al., 2018). Embora os princípios da genética quantitativa e melhoramento que fundamentam a SG devam permanecer essencialmente como são, aqueles relacionados com a geração e análise de dados certamente irão evoluir à medida que novas tecnologias de genotipagem e sequenciamento de DNA se materializarem e abordagens mais potentes de ciência de dados e inteligência artificial forem desenvolvidas. À medida que a SG for sendo adotada nas diversas organizações e, com isso, uma quantidade crescente de dados experimentais for sendo coletada e disponibilizada, a predição genômica poderá se tornar uma plataforma experimental poderosa para auxiliar na investigação da base molecular das características quantitativas que, por sua vez, vai auxiliar no direcionamento de estratégias efetivas de edição gênica. A elucidação completa das conexões entre o genoma e a expressão final das características multifatoriais no eucalipto, no entanto, continuará a ser um empreendimento desafiador devido à dinâmica de tempo e espaço à qual árvores são sujeitas, e aos processos estocásticos de difícil predição que impactam as relações complexas entre genótipos e fenótipos.

Agradecimentos

Os trabalhos descritos neste capítulo foram apoiados, ao longo dos anos, por recursos de diversos projetos competitivos de agências de fomento, incluindo a Financiadora de Estudos e Projetos-Finep, Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Capes e Fundação de Apoio a Pesquisa do Distrito Federal-FAP-DF, bem como por várias empresas privadas do setor florestal, a quem o autor estende seus sinceros agradecimentos. Um agradecimento especial a todos os estudantes e colaboradores das várias universidades, organizações de pesquisa e empresas privadas que participaram ativamente da execução destes projetos, sem os quais nada do que foi descrito teria sido possível. Também agradeço, em particular à Embrapa por prover um ambiente estimulante e uma infraestrutura de pesquisa que permitiram estabelecer e manter, com sucesso, as redes de colaboração nacionais e internacionais necessárias para a execução continuada dos projetos que resultaram nos avanços descritos. Por fim, um agradecimento especial ao meu melhor amigo, pesquisador da Embrapa e grande parceiro de iniciativas e realizações ao longo destes 25 anos no LGV, Dr. Marcio Elias Ferreira, pela leitura crítica, correções e sugestões neste capítulo.

Referências

AKKAYA, M. S.; BHAGWAT, A. A.; CREGAN, P. B. Length polymorphisms of simple sequence repeat DNA in soybean. **Genetics**, v. 132, n. 4, p. 1131-1139, 1992.

ALVES, A. A.; ROSADO, C. C. G.; FARIA, D. A.; GUIMARÃES, L. M. da S.; LAU, D.; BROMMONSCHENKEL, S. H.; GRATTAPAGLIA, D.; ALFENAS, A. C. Genetic mapping provides evidence for the role of additive and non-additive QTLs in the response of inter-specific hybrids of *Eucalyptus* to *Puccinia psidii* rust infection. **Euphytica**, v. 183, n. 1, p. 27-38, 2012. DOI: <https://doi.org/10.1007/s10681-011-0455-5>.

ARNOLD, B.; CORBETT-DETIG, R. B.; HARTL, D.; BOMBLIES, K. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. **Molecular Ecology**, v. 22, n. 11, p. 3179-3190, 2013. DOI: <https://doi.org/10.1111/mec.12276>.

BAIRD, N. A.; ETTER, P. D.; ATWOOD, T. S.; CURREY, M. C.; SHIVER, A. L.; LEWIS, Z. A.; SELKER, E. U.; CRESKO, W. A.; JOHNSON, E. A. Rapid SNP Discovery and genetic mapping using sequenced RAD markers. **Plos One**, v. 3, n. 10, 2008. DOI: <https://doi.org/10.1371/journal.pone.0003376>.

BARBAZUK, W. B.; EMRICH, S. J.; CHEN, H. D.; LI, L.; SCHNABLE, P. S. SNP discovery via 454 transcriptome sequencing. **Plant Journal**, v. 51, n. 5, p. 910-918, 2007. DOI: <https://doi.org/10.1111/j.1365-313X.2007.03193.x>.

BARTHOLOME, J.; VAN HEERWAARDEN, J.; ISIK, F.; BOURY, C.; VIDAL, M.; PLOMION, C.; BOUFFIER, L. Performance of genomic prediction within and across generations in maritime pine. **BMC Genomics**, v. 17, n. 1, p. 604, 2016. DOI: <https://doi.org/10.1186/s12864-016-2879-8>.

BERNARDO, R. Molecular markers and selection for complex traits in plants: learning from the last 20 years. **Crop Science**, v. 48, n. 5, p. 1649-1664, 2008. DOI: <https://doi.org/10.2135/cropsci2008.03.0131>.

BERNARDO, R.; YU, J. M. Prospects for genomewide selection for quantitative traits in maize. **Crop Science**, v. 47, n. 3, p. 1082-1090, 2007. DOI: <https://doi.org/10.2135/cropsci2006.11.0690>.

BLAKE, T.; BEVILACQUA, E. Early selection of fast-growing *Eucalyptus* clones and species. **IPEF International**, v. 1, p. 26-34, 1990.

BORRALHO, N. M. G.; COTTERILL, P. P.; KANOWSKI, P. J. Genetic-parameters and gains expected from selection for dry-weight in *Eucalyptus globulus* ssp. *globulus* in Portugal. **Forest Science**, v. 38, n. 1, p. 80-94, 1992. DOI: <https://doi.org/10.1093/forestscience/38.1.80>.

BOUVET, J.-M.; VIGNERON, P.; SAYA, R.-A.; GOUMA, R. Early selection of *Eucalyptus* clones in retrospective nursery test using growth, morphological and dry matter criteria, in Republic of Congo. **Southern African Forestry Journal**, v. 200, n. 1, p. 5-17, 2004. DOI: <https://doi.org/10.1080/20702620.2004.10431756>.

BRONDANI, R. P.; BRONDANI, C.; GRATTAPAGLIA, D. Towards a genus-wide reference linkage map for *Eucalyptus* based exclusively on highly informative microsatellite markers. **Molecular Genetics and Genomics**, v. 267, n. 3, p. 338-47, 2002. DOI: <https://doi.org/10.1007/s00438-002-0665-6>.

BRONDANI, R. P. V.; BRONDANI, C.; GRATTAPAGLIA, D. **Manual prático para o desenvolvimento de marcadores microssatélites em plantas**. Brasília, DF: Embrapa, 2007. 192 p.

- BRONDANI, R. P. V.; BRONDANI, R.; TARCHINI, R.; GRATTAPAGLIA, D. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E.urophylla*. **Theoretical and Applied Genetics**, v. 97, p. 816-827, 1998. DOI: <https://doi.org/10.1007/s001220050961>.
- BRONDANI, R. P. V.; BRONDANI, C.; TARCHINI, R.; TINGEY, D.; GRATTAPAGLIA, D. Discovery and development of microsatellite based markers in *Eucalyptus*. In: IUFRO CONFERENCE ON SILVICULTURE AND IMPROVEMENT EUCALYPTUS=CONFERÊNCIA IUFRO SOBRE SILVICULTURA E MELHORAMENTO DE EUCALIPTOS, 1997, Salvador. **Proceedings...= Anais...** Colombo: EMBRAPA-CNPQ, 1997. v. 2. p. 30-35.
- BRONDANI, R. P.; WILLIAMS, E. R.; BRONDANI, C.; GRATTAPAGLIA, D. A microsatellite-based consensus linkage map for *Eucalyptus* species and a novel set of 230 microsatellite markers for the genus. **BMC Plant Biology**, article number 20, 2006. DOI: <https://doi.org/10.1186/1471-2229-6-20>.
- BYRNE, M.; MORAN, G. F.; TIBBITS, W. N. Restriction map and maternal inheritance of chloroplast DNA in *Eucalyptus nitens*. **Journal of Heredity**, v. 84, p. 218-220, 1993.
- CAPPA, E. P.; EL-KASSABY, Y. A.; GARCIA, M. N.; ACUÑA, C.; BORRALHO, N. M. G.; GRATTAPAGLIA, D.; POLTRI, S. N. M. Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: a case study in *Eucalyptus globulus*. **Plos One**, v. 8, n. 11, 2013. DOI: <https://doi.org/10.1371/journal.pone.0081267>.
- CAPPA, E. P.; LIMA, B. M.; SILVA-JUNIOR, O. B.; GARCIA, C. C.; MANSFIELD, S. D.; GRATTAPAGLIA, D. Improving genomic prediction of growth and wood traits in *Eucalyptus* using phenotypes from non-genotyped trees by single-step GBLUP. **Plant Science**, v. 284, p. 9-15, 2019. DOI: <https://doi.org/10.1016/j.plantsci.2019.03.017>.
- CARRER, H. Seeing the FORESTs for the trees. **Genetics and Molecular Biology**, v. 28, n. 3, p. I-II, 2005.
- CORREIA, L.; FARIA, D.; GRATTAPAGLIA, D. Comparative assessment of SNPs and microsatellites for fingerprinting, parentage and assignment testing in species of *Eucalyptus*. **BMC Proceedings**, v. 5, suppl. 7, p. 41, 2011. DOI: <https://doi.org/10.1186/1753-6561-5-S7-P41>.
- CROSSA, J.; PÉREZ-RODRÍGUEZ, P.; CUEVAS, J.; MONTESINOS-LÓPEZ, O.; JARQUÍN, D.; CAMPOS G. DE LOS; BURGUEÑO, J.; GONZÁLEZ-CAMACHO, J. M.; PÉREZ-ELIZALDE, S.; BEYENE, Y.; DREISIGACKER, S.; SINGH, R.; ZHANG, X.; GOWDA, M.; ROORKIWA, M.; RUTKOSKI, J.; VARSHNEY, R. K. Genomic selection in plant breeding: methods, models, and perspectives. **Trends in Plant Science**, v. 22, n. 11, p. 961-975, 2017. DOI: <https://doi.org/10.1016/j.tplants.2017.08.011>.
- DAVEY, J. W.; HOHENLOHE, P. A.; ETTER, P. D.; BOONE, J. Q.; CATCHEN, J. M.; BLAXTER, M. L. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. **Nature Reviews Genetics**, v. 12, n. 7, p. 499-510, 2011. DOI: <https://doi.org/10.1038/nrg3012>.
- ECKERT, A. J.; PANDE, B.; ERSOZ, E. S.; WRIGHT, M. H.; RASHBROOK, V. K.; NICOLET, C. M.; NEALE, D. B. High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (*Pinus taeda* L.). **Tree Genetics & Genomes**, v. 5, n. 1, p. 225-234, 2009. DOI: <https://doi.org/10.1007/s11295-008-0183-8>.
- EL-KASSABY, Y. A.; LSTIBUREK, M. Breeding without breeding. **Genetics Research**, v. 91, n. 2, p. 111-120, 2009. DOI: <https://doi.org/10.1017/S001667230900007X>.

- ELDRIDGE, K.; DAVIDSON, J.; HARDWIID, C.; WYK, G. van. **Eucalypt domestication and breeding**. Oxford: Clarendon Press, 1993. 288 p.
- ELSHIRE, R. J.; GLAUBITZ, J. C.; SUN, Q. I.; POLAND, J. A.; KAWAMOTO, K.; BUCKLER, E.; MITCHELL, S. E. A robust, simple genotyping-by-sequencing (Gbs) approach for high diversity species. **PLoS One**, v. 6, n. 5, 2011. DOI: <https://doi.org/10.1371/journal.pone.0019379>.
- FARIA, D. A.; MAMANI, E. M. C.; PAPPAS, G. J.; GRATTAPAGLIA, D. Genotyping systems for *Eucalyptus* based on tetra-, penta-, and hexanucleotide repeat EST microsatellites and their use for individual fingerprinting and assignment tests. **Tree Genetics & Genomes**, v. 7, n. 1, p. 63-77, 2011a. DOI: <https://doi.org/10.1007/s11295-010-0315-9>.
- FARIA, D. A.; MAMANI, E. M. C.; PAPPAS, M. R.; PAPPAS, G. J.; GRATTAPAGLIA, D. A selected set of EST-derived microsatellites, polymorphic and transferable across 6 species of *Eucalyptus*. **Journal of Heredity**, v. 101, n. 4, p. 512-520, 2010. DOI: <https://doi.org/10.1093/jhered/esq024>.
- FARIA, D. A.; MAMANI, V.; SENA, J.; ALVES, A.; FALCAO, C.; LOURENÇO, R.; PAPPAS, J.; GRATTAPAGLIA, D. A new set of 182 microsatellites for *Eucalyptus*: characterization and mapping in a four-species consensus linkage map. **BMC Proceedings**, v. 5, n. suppl. 7, p. P34, 2011b. DOI: 10.1186/1753-6561-5-S7-P34.
- FIGUEIREDO, P. N. Beyond technological catch-up: an empirical investigation of further innovative capability accumulation outcomes in latecomer firms with evidence from Brazil. **Journal of Engineering and Technology Management**, v. 31, p. 73-102, 2014. DOI: <https://doi.org/10.1016/j.jengtecman.2013.10.008>.
- FIGUEIREDO, P. N. Evolution of the short-fiber technological trajectory in Brazil's pulp and paper industry: the role of firm-level innovative capability-building and indigenous institutions. **Forest Policy and Economics**, v. 64, p. 1-14, 2016. DOI: <https://doi.org/10.1016/j.forpol.2015.12.008>.
- FISHER, R. A. The correlation between relatives on the supposition of Mendelian inheritance. **Transactions of the Royal Society of Edinburgh**, v. 52, p. 399-433, 1918.
- FOSS, N.; MILAGRES, R. Pro-social motivation beyond firm boundaries: the case of the Genolyptus Network. **BAR. Brazilian Administration Review**, v. 11, p. 364-384, 2014.
- GAIOTTO, F. A.; BRAMUCCI, M.; GRATTAPAGLIA, D. Estimation of outcrossing rate in a breeding population of *Eucalyptus urophylla* with dominant RAPD and AFLP markers. **Theoretical and Applied Genetics**, v. 95, p. 842-849, 1997. DOI: <https://doi.org/10.1007/s001220050634>.
- GAIOTTO, F. A.; GRATTAPAGLIA, D. Estimation of genetic variability in a breeding population of *Eucalyptus urophylla* using AFLP (Amplified Fragment Length Polymorphism) markers. In: IUFRO CONFERENCE ON SILVICULTURE AND IMPROVEMENT EUCALYPTUS=CONFERÊNCIA IUFRO SOBRE SILVICULTURA E MELHORAMENTO DE EUCALIPTOS, 1997, Salvador. **Proceedings... = Anais...** Colombo: EMBRAPA-CNPq, 1997. v. 2. p. 46-52.
- GANAL, M. W.; ALTMANN, T.; RODER, M. S. SNP identification in crop plants. **Current Opinion in Plant Biology**, v. 12, n. 2, p. 211-217, 2009. DOI: <https://doi.org/10.1016/j.pbi.2008.12.009>.
- GAUTIER, M.; GHARBI, K.; CEZARD, T.; FOUCAUD, J.; KERDELHUÉ, C.; PUDLO, P.; CORNUET, J. M.; ESTOUP, A. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. **Molecular Ecology**, v. 22, n. 11, p. 3165-3178, 2013. DOI: <https://doi.org/10.1111/mec.12089>.

- GODDARD, M. E.; HAYES, B. J. Mapping genes for complex traits in domestic animals and their use in breeding programmes. **Nature Reviews Genetics**, v. 10, n. 6, p. 381-391, 2009. DOI: <https://doi.org/10.1038/nrg2575>.
- GOODSTEIN, D.; RUSSELL, H.; NEUPANE, R.; HAYES, R.D.; FAZO, J.; MITROS, T.; DIRKS, W.; HELLSTEN, U.; PUTNAM, N.; ROKHSAR, D. D. Phytozome: a comparative platform for green plant genomics. **Nucleic Acids Research**, v. 40, n. D1, p. D1178-D1186, 2012. DOI: <http://dx.doi.org/10.1093/nar/gkr944>.
- GRATTAPAGLIA, D.; ALENCAR, S.; PAPPAS, G. J. Genome-wide genotyping and SNP discovery by ultra-deep Restriction-Associated DNA (RAD) tag sequencing of pooled samples of *E. grandis* and *E. globulus*. **BMC Proceedings**, v. 5, suppl. 7, p. P45, 2011. DOI: <https://doi.org/10.1186/1753-6561-5-S7-P45>.
- GRATTAPAGLIA, D.; ALFENAS, A.; COELHO, A.; PAPPAS, G.; PASQUALI, G.; PEREIRA, G.; COLODETTE, J.; GLOMIDE, J. L.; BUENO, J.; CASCARDO, J.; BRONDANI, R.; BROMMONSCHENKEL, S. Building resources for molecular breeding of *Eucalyptus*. In: EUCALYPTUS IN A CHANGING WORLD: international IUFRO conference of the WP2.08.03 on silviculture and improvement of eucalypts, 2004, Aveiro. **Proceedings**. Aveiro: Instituto de Investigação da Floresta e Papel, 2004a. p. 20-32.
- GRATTAPAGLIA, D.; BERTOLUCCI, F. L.; SEDEROFF, R. R. Genetic mapping of QTLs controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross strategy and RAPD markers. **Theoretical and Applied Genetics**, v. 90, n. 7-8, p. 933-947, 1995. DOI: <https://doi.org/10.1007/BF00222906>.
- GRATTAPAGLIA, D.; CHAPARRO, J.; WILCOX, P.; MCCORD, S.; WERNER, D.; AMERSON, H.; MCKEAND, S.; BRIDGWATER, F.; WHETTEN, R.; O'MALLEY, D.; SEDEROFF, R. R. Mapping in woody plants with RAPD markers: applications to breeding in forestry and horticulture. In: SYMPOSIUM OF THE APPLICATIONS OF RAPD TECHNOLOGY TO PLANT BREEDING: Analysis of genetic relationships using RAPD marker data, 1992, Minneapolis. **Proceedings** [...]. Minneapolis: Crop Science Society of America, American Society of Horticultural Science, American Genetic Association, 1992. p. 37-40.
- GRATTAPAGLIA, D. Integrating genomics into *Eucalyptus* breeding. **Genetics and Molecular Research**, v. 3, n. 3, p. 369-379, 2004.
- GRATTAPAGLIA, D.; KIRST, M. *Eucalyptus* applied genomics: from gene sequences to breeding tools. **New Phytologist**, v. 179, n. 4, p. 911-929, 2008.
- GRATTAPAGLIA, D.; MAMANI, E. M. C.; SILVA-JUNIOR, O. B.; FARIA, D. A novel genome-wide microsatellite resource for species of *Eucalyptus* with linkage-to-physical correspondence on the reference genome sequence. **Molecular Ecology Resources**, v. 15, n. 2, p. 437-448, 2015. DOI: <https://doi.org/10.1111/1755-0998.12317>.
- GRATTAPAGLIA, D. Marker assisted selection in *Eucalyptus*. In: GUIMARÃES, E. (Ed.). **Marker-assisted selection (MAS) in crops, livestock, forestry and fish: current status and the way forward**. Rome: FAO, 2006.
- GRATTAPAGLIA, D. Molecular breeding of *Eucalyptus*, state of the art, operational applications and technical challenges. In: JAIN, S. M.; MINOCHA, S. C. (ed.). **Molecular biology of woody plants**. The Netherlands: Kluwer Academic Publishers, 2000. p. 451-474. (Forestry sciences, 64).

GRATTAPAGLIA, D.; PENCHEL, R.; SEDEROFF, R. R. Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. **Genetics**, v. 144, n. 3, p. 1205-14, 1996.

GRATTAPAGLIA, D.; PIMENTA, D.; CAMPINHOS, E.; RESENDE, G. D.; ASSIS, T. Marcadores moleculares na proteção varietal de *Eucalyptus*. In: CONGRESSO FLORESTAL BRASILEIRO, 8., 2003, São Paulo. **Benefícios, produtos e serviços da floresta: oportunidades e desafios do século XXI**. São Paulo: Sociedade Brasileira de Silvicultura: Sociedade Brasileira de Engenheiros Florestais, 2003. p. 1-13.

GRATTAPAGLIA, D.; PLOMION, C.; KIRST, M.; SEDEROFF, R. Genomics of growth traits in forest trees. **Current Opinion in Plant Biology**, v. 12, n. 2, p. 148-156, 2009. DOI: <https://doi.org/10.1016/j.pbi.2008.12.008>.

GRATTAPAGLIA, D.; RESENDE, M. D. V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, v. 7, n. 2, p. 241-255, 2011.

GRATTAPAGLIA, D.; RIBEIRO, V. J.; REZENDE, G. D. Retrospective selection of elite parent trees using paternity testing with microsatellite markers: an alternative short term breeding tactic for *Eucalyptus*. **Theoretical and Applied Genetics**, v. 109, n. 1, p. 192-9, 2004b. DOI: <https://doi.org/10.1007/s00122-004-1617-9>.

GRATTAPAGLIA, D.; SEDEROFF, R. R. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. **Genetics**, v. 137, n. 4, p. 1121-37, 1994.

GRATTAPAGLIA, D.; SILVA-JUNIOR, O. B.; KIRST, M.; LIMA, B. M. de; FARIA, D. A.; PAPPAS JUNIOR, G. J. High-throughput SNP genotyping in the highly heterozygous genome of *Eucalyptus*: assay success, polymorphism and transferability across species. **BMC Plant Biology**, v. 11, n. 65, 2011. DOI: <https://doi.org/10.1186/1471-2229-11-65>.

GRATTAPAGLIA, D.; SILVA-JUNIOR, O. B.; RESENDE, R. T.; CAPPA, E. P.; MÜLLER, B. S. F.; TAN, B.; ISIK, F.; RATCLIFFE, B.; EL-KASSABY, Y. E. Quantitative genetics and genomics converge to accelerate forest tree breeding. **Frontiers in Plant Science**, v. 9, p. 1693, 2018. DOI: <https://doi.org/10.3389/fpls.2018.01693>.

GRATTAPAGLIA, D. Status and perspectives of genomic selection in forest tree breeding. In: VARSHNEY, R. K.; ROORKIWAL, M.; SORRELLS, M. E. (ed.). **Genomic selection for crop improvement: new molecular breeding strategies for crop improvement**. Cham: Springer International Publishing, 2017. p. 199-249.

GRATTAPAGLIA, D.; VAILLANCOURT, R. E.; SHEPHERD, M.; THUMMA, B. R.; FOLEY, W.; KÜLHEIM, C.; POTTS, B. M.; MYBURG, A. A. Progress in Myrtaceae genetics and genomics: *Eucalyptus* as the pivotal genus. **Tree Genetics & Genomes**, v. 8, n. 3, p. 463-508, 2012. DOI: <https://doi.org/10.1007/s11295-012-0491-x>.

GRIFFIN, A. R.; WHITEMAN, P.; RUDGE, T.; BURGESS, I. P.; MONCUR, M. Effect of paclobutrazol on flower-bud production and vegetative growth in two species of *Eucalyptus*. **Canadian Journal of Forest Research**, v. 23, n. 640-647, 1993. DOI: <https://doi.org/10.1139/x93-084>.

GUNDERSON, K. L.; STEEMERS, F. J.; LEE, G.; MENDOZA, L. G.; CHEE, M. S. A genome-wide scalable SNP genotyping assay using microarray technology. **Nature Genetics**, v. 37, n. 5, p. 549-54, 2005. DOI: <https://doi.org/10.1038/ng1547>.

- HALEY, C. S.; VISSCHER, P. M. Strategies to utilize marker-quantitative trait loci associations. **Journal of Dairy Science**, v. 81, p. 85-97,
- HARFOUCHE, A.; MEILAN, R.; KIRST, M.; MORGANTE, M.; BOERJAN, W.; SABATTI, M.; MUGNOZZA, G. S. Accelerating the domestication of forest trees in a changing world. **Trends in Plant Science**, v. 17, n. 2, p. 64-72, 2012. DOI: <https://doi.org/10.1016/j.tplants.2011.11.005>.
- HARWOOD, C. New introductions: doing it right. In: WALKER, J. (ed.). **Proceedings of the workshop of developing a eucalypt resource: learning from Australia and elsewhere**. Christchurch: University of Canterbury, 2011. p. 125-136.
- HASAN, O.; REID, J. B. Reduction of generation time in *Eucalyptus globulus*. **Plant Growth Regulation**, v. 17, n. 1, p. 53-60, 1995. DOI: <https://doi.org/10.1007/BF00024495>.
- HAYES, B. J.; BOWMAN, P. J.; CHAMBERLAIN, A. J.; GODDARD, M. E. Invited review: genomic selection in dairy cattle: progress and challenges. **Journal of Dairy Science**, v. 92, n. 2, p. 433-443, 2009. DOI: <https://doi.org/10.3168/jds.2008-1646>.
- HEFFNER, E. L.; SORRELLS, M. E.; JANNINK, J. L. Genomic Selection for crop improvement. **Crop Science**, v. 49, n. 1, p. 1-12, 2009.
- HUDSON, C. J.; FREEMAN, J. S.; KULLAN, A. R. K.; PETROLI, C. D.; SANSALONI, C. P.; KILIAN, A.; DETERING, F.; GRATTAPAGLIA, D.; POTTS, B. M.; MYBURG, A. A.; VAILLANCOURT, R. E. A reference linkage map for *Eucalyptus*. **BMC Genomics**, v. 13, article 240, 2012. DOI: <https://doi.org/10.1186/1471-2164-13-240>.
- ISIK, F. Genomic selection in forest tree breeding: the concept and an outlook to the future. **New Forests**, v. 45, n. 3, p. 379-401, 2014.
- IWATA, H.; HAYASHI, T.; TSUMURA, Y. Prospects for genomic selection in conifer breeding: a simulation study of *Cryptomeria japonica*. **Tree Genetics & Genomes**, v. 7, n. 4, p. 747-758-758, 2011. DOI: <http://dx.doi.org/10.1007/s11295-011-0371-9>.
- JUNGHANS, D. T.; ALFENAS, A. C.; BROMMONSCHENKEL, S. H.; ODA, S.; MELLO, E. J. GRATTAPAGLIA, D. Resistance to rust (*Puccinia psidii* Winter) in *Eucalyptus*: mode of inheritance and mapping of a major gene with RAPD markers. **Theoretical and Applied Genetics**, v. 108, n. 1, p. 175-80, 2003. DOI: <https://doi.org/10.1007/s00122-003-1415-9>.
- KIRST, M.; BRONDANI, R. P. V.; BRONDANI, C.; GRATTAPAGLIA, D. Screening of designed primer pairs of recovery of microsatellite markers and their transferability among species of *Eucalyptus*. In: IUFRO CONFERENCE ON SILVICULTURE AND IMPROVEMENT EUCALYPTUS=CONFERÊNCIA IUFRO SOBRE SILVICULTURA E MELHORAMENTO DE EUCALIPTOS, 1997, Salvador. **Proceedings...=Anais...** Colombo: EMBRAPA-CNPQ, 1997. v. 2. p. 167-171.
- KIRST, M.; CORDEIRO, C. M.; REZENDE, G. D.; GRATTAPAGLIA, D. Power of microsatellite markers for fingerprinting and parentage analysis in *Eucalyptus grandis* breeding populations. **Journal of Heredity**, v. 96, n. 2, p. 161-166, 2005. DOI: <https://doi.org/10.1093/jhered/esi023>.
- LIMA, B. M. **Bridging genomics and quantitative genetics of *Eucalyptus***: genome-wide predic. 2014. 93 f. Thesis (Doctor in Science) - University of São Paulo "Luiz de Queiroz" College of Agriculture, Piracicaba.

- LIMA, B. M.; CAPPA, E. P.; SILVA-JUNIOR, O. B.; GARCIA, C.; MANSFIELD, S. D.; GRATTAPAGLIA, D. Quantitative genetic parameters for growth and wood properties in *Eucalyptus* “urograndis” hybrid using near-infrared phenotyping and genome-wide SNP-based relationships. **Plos One**, v. 14, n. 6, p. e0218747, 2019. DOI: <https://doi.org/10.1371/journal.pone.0218747>.
- LIN, Z.; HAYES, B. J.; DAETWYLER, H. D. Genomic selection in crops, trees and forages: a review. **Crop & Pasture Science**, v. 65, n. 11, p. 1177-1191, 2014.
- LOWRY, D.; HOBAN, S.; KELLEY, J.; LOTTERHOS, K.; REED, L.; ANTOLIN, M. F.; STORFER, A. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. **Molecular Ecology Resources**, v. 17, n. 2, p. 142-152, 2016. DOI: <https://doi.org/10.1111/1755-0998.12635>.
- MAMANI, E. M. C.; BUENO, N. W.; FARIA, D. A.; GUIMARÃES, L. M. S.; LAU, D.; ALFENAS, A. C.; GRATTAPAGLIA, D. Positioning of the major locus for *Puccinia psidii* rust resistance (Ppr1) on the *Eucalyptus* reference map and its validation across unrelated pedigrees. **Tree Genetics & Genomes**, v. 6, n. 6, p. 953-962, 2010. DOI: <https://doi.org/10.1007/s11295-010-0304-z>.
- MARQUES, C. M.; ARAÚJO, J. A.; FERREIRA, J. G.; WHETTEN, R.; O’MALLEY, D. M.; LIU, B.-H.; SEDEROFF, R. AFLP genetic maps of *Eucalyptus globulus* and *E. tereticornis*. **Theoretical and Applied Genetics**, v. 96, p. 727-737, 1998. DOI: <https://doi.org/10.1007/s001220050795>.
- MARQUES, C. M.; VASQUEZ-KOOL, J.; CAROCHA, V. J.; FERREIRA, J. G.; O’MALLEY, D. M.; LIU, B.-H.; SEDEROFF, R. Genetic dissection of vegetative propagation traits in *Eucalyptus tereticornis* and *E. globulus*. **Theoretical and Applied Genetics**, v. 99, n. 6, p. 936-946, 1999. DOI: <https://doi.org/10.1007/s001220051400>.
- MARQUES, M.; BRONDANI, V.; GRATTAPAGLIA, D.; SEDEROFF, R. Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species. **Theoretical and Applied Genetics**, v. 105, n. 2-3, p. 474-478, 2002. DOI: <https://doi.org/10.1007/s00122-002-0899-z>.
- MEUWISSEN, T. H.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, n. 4, p. 1819-29.
- MILAGRES, R. Redes de empresas, a chave para inovar. **HSM Management**, v. 72, p. 1-6, 2009.
- MILAGRES, R. Rotinas e redes: o caso Genolyptus. **Reuna**, v. 19, p. 105-122, 2014.
- MILAGRES, R.; SILVEIRA, F. Evaluating routines for network knowledge generation and transfer: the Genolyptus case. In: DRUIDs 25th Celebration Conference. Copenhagen: DRUID, 2008. 28 p. (Danish Research Unit for Industrial Dynamics).
- MISSIAGGIA, A.; PIACEZZI, A.; GRATTAPAGLIA, D. Genetic mapping of Eef1, a major effect QTL for early flowering in *Eucalyptus grandis*. **Tree Genetics & Genomes**, v. 1, p. 79-84, 2005. DOI: <https://doi.org/10.1007/s11295-005-0011-3>.
- MORAN, G. F.; BELL, J. C. *Eucalyptus*. In: TANKSLEY, S. D.; ORTON, T. J. (ed.). **Isozymes in plant genetics and breeding**. Amsterdam: Elsevier, 1983. p. 423-441.
- MORAN, G. F. Patterns of genetic diversity in Australian tree species. **New Forests**, v. 6, p. 49-66, 1992.
- MÜLLER, B. S. F.; ALMEIDA FILHO, J. E. de; LIMA, B. M.; GARCIA, C. C.; MISSIAGGIA, A.; AGUIAR, A. M.; TAKAHASHI, E.; KIRST, M.; GEZAN, S. A.; SILVA-JUNIOR, O. B.; NEVES, L. G.; GRATTAPAGLIA, D. Independent and Joint-GWAS for growth traits in *Eucalyptus* by assembling

genome-wide data for 3373 individuals across four breeding populations. **New Phytologist**, v. 221, n. 2, p. 818-833, 2019. DOI: <https://doi.org/10.1111/nph.15449>.

MÜLLER, B. S. F.; NEVES, L. G.; ALMEIDA FILHO, J. E. de; RESENDE JUNIOR, M. F. R.; MUÑOZ, P. R.; SANTOS, P. E. T. dos; PALUDZYSZYN FILHO, E.; KIRST, M.; GRATTAPAGLIA, D. Genomic prediction in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of *Eucalyptus*. **BMC Genomics**, v. 18, article 524, 2017. 17 p. DOI: <https://doi.org/10.1186/s12864-017-3920-2>.

MURRAY, S. S.; OLIPHANT, A.; SHEN, R.; MCBRIDE, C.; STEEKE, R. J.; SHANNON, S. G.; RUBANO, T.; KERMANI, B. G.; FAN, J.-B.; CHEE, M. S.; HANSEN, M. S. T. A highly informative SNP linkage panel for human genetic studies. **Nature Methods**, v. 1, n. 2, p. 113-7, 2004. DOI: <https://doi.org/10.1038/nmeth712>.

MYBURG, A.; GRATTAPAGLIA, D.; TUSKAN, G.; JENKINS, J.; SCHMUTZ, J.; MIZRACHI, E.; HEFER, C.; PAPPAS, G.; STERCK, L.; DE PEER, Y. V.; HAYES, R.; ROKHSAR, D. The *Eucalyptus grandis* Genome Project: genome and transcriptome resources for comparative analysis of woody plant biology. **BMC Proceedings**, v. 5, suppl. 7, p. I20, 2011. DOI: <https://doi.org/10.1186/1753-6561-5-S7-I20>.

MYBURG, A. A.; GRATTAPAGLIA, D.; TUSKAN, G. A.; HELSTEN, U.; HAYES, R. D.; GRIMWOOD, J.; JENKINS, J.; LINDQUIST, E.; BAUER, D.; GOODSTEIN, D. M.; DUBCHAK, I.; POLIAKOV, A.; MIZRACHI, E.; KULLAN, A. R. K.; HUSSEY, S. G.; PINARD, D.; MERWE, K. van der; SINGH, P.; JAARVELD, I. van; SILVA JUNIOR, O. B.; TOGAWA, R. C.; PAPPAS, M. R.; FARIA, D. A.; SANSALONI, C. P.; PETROLI, C. D.; YANG, X.; RANJAN, P.; TSCHAPLINSKI, T. J.; YE, C.-Y.; LI, T.; STERCK, L.; VANNESTE, K.; MURAT, F.; SOLER, M.; SAN CLEMENTE, H.; SAIDI, N.; CASSAN-WANG, H.; DUNAND, C.; HEFER, C. A.; BORNBERG-BAUER, E.; KERSTING, A. R.; VINING, K.; AMARASINGHE, V.; RANIK, M.; NAITHANI, S.; ELSER, J.; BOYD, A. E.; LISTON, A.; SPATAFORA, J. W.; DHARMWARDHANA, P.; RAJA, R.; SULLIVAN, C.; ROMANEL, E.; ALVES-FERREIRA, M.; KULHEIM, C.; FOLEY, W.; CAROCHA, V.; PAIVA, J.; KUDRNA, D.; BROMMONSCHENKEL, S. H.; PASQUALI, G.; BYRNE, M.; RIGAUULT, P.; SPOKEVICIUS, A.; JONES, R. C.; STEANE, D. A.; VAILLANCOURT, R. E.; POTTS, B. M.; JOUBERT, F.; BARRY, K.; PAPPAS JUNIOR, G. J.; STRAUSS, S. H.; JAISWAL, P.; GRIMA-PETTENATI, J.; SALSE, J.; PEER, Y. van de; ROKHSAR, D. S.; SCHMUTZ, J. The genome of *Eucalyptus grandis*. **Nature**, v. 510, p. 356-362, 2014. DOI: <https://doi.org/10.1038/nature13308>.

MYBURG, A. A.; GRIFFIN, A. R.; SEDEROFF, R. R.; WHETTEN, R. W. Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F1 hybrid based on a double pseudo-backcross mapping approach. **Theoretical and Applied Genetics**, v. 107, n. 6, p. 1028-42, 2003. DOI: <https://doi.org/10.1007/s00122-003-1347-4>.

MYBURG, A. A.; POTTS, B. M.; MARQUES, C. M.; KIRST, M.; GION, J.-M.; GRATTAPAGLIA, D.; GRIMA-PETTENATTI, J. *Eucalyptus*. In: KOLE, C. (ed.) **Genome mapping and molecular breeding in plants**. New York: Springer, 2007. p. 115-160. (Forest trees, 7).

MYLES, S. Improving fruit and wine: what does genomics have to offer? **Trends in Genetics**, v. 29, n. 4, p. 190-196, 2013. DOI: <http://dx.doi.org/10.1016/j.tig.2013.01.006>.

NEALE, D. B.; KREMER, A. Forest tree genomics: growing resources and applications. **Nature Reviews Genetics**, v. 12, n. 2, p. 111-122, 2011. DOI: <https://doi.org/10.1038/nrg2931>.

NEVES, L. G.; MAMANI, E. M.; ALFENAS, A. C.; KIRST, M.; GRATTAPAGLIA, D. A high-density transcript linkage map with 1,845 expressed genes positioned by microarray-based Single Feature Polymorphisms (SFP) in *Eucalyptus*. **BMC Genomics**, v. 12, 2011. DOI: <https://doi.org/10.1186/1471-2164-12-189>.

NOVAES, E.; DROST, D. R.; FARMERIE, W. G.; PAPPAS JUNIOR, G. J.; GRATTAPAGLIA, D.; SEDEROFF, R. R.; KIRST, M. High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. **BMC Genomics**, v. 9, p. 312, 2008. DOI: <https://doi.org/10.1186/1471-2164-9-312>.

PAIVA, J. A. P.; PRAT, E.; VAUTRIN, S.; SANTOS, M. D.; SAN CLEMENTE, H.; BROMMONSCHENKEL, S.; FONSECA, P. G. S.; GRATTAPAGLIA, D.; SONG, X.; AMMIRAJU, J. S. S.; KUDRNA, D.; WING, R. A.; FREITAS, A. T.; BERGÈS, H.; GRIMA PETTENATI, J. Advancing *Eucalyptus* genomics: identification and sequencing of lignin biosynthesis genes from deep-coverage BAC libraries. **BMC Genomics**, v. 12, article 137, 2011. DOI: <https://doi.org/10.1186/1471-2164-12-137>.

PAVY, N.; PELGAS, B.; BEAUSEIGLE, S.; BLAIS, S.; GAGNON, F.; GOSSELIN, I.; LAMOTHE, M.; ISABEL, N.; BOUSQUET, J. Enhancing genetic mapping of complex genomes through the design of highly multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. **BMC Genomics**, v. 9, article 21, 2008. DOI: <https://doi.org/10.1186/1471-2164-9-21>.

PETROLI, C. D.; SANSALONI, C. P.; CARLING, J.; MAMANI, E. M. C.; STEANE, D. A.; MYBURG, A. M.; VAILLANCOURT, R. E.; KILIAN, A.; PAPPAS JUNIOR, G. J.; SILVA, O. B. da; GRATTAPAGLIA, D. Genomic characterization, high-density mapping and anchoring of DArT markers to the reference genome of *Eucalyptus*. **BMC Proceedings**, v. 5, suppl. 7, p. P35, 2011. DOI: <https://doi.org/10.1186/1753-6561-5-S7-P35>.

PLOMION, C.; BASTIEN, C.; BOEAT-TRIBOULOT, M.-B.; BOUFFIER, L.; DÉJARDIN, A.; DUPLESSIS, S.; FADY, B.; HEUERTZ, M.; LE GAC, A.-L.; LE PROVOST, G.; LEGUÉ, V.; LELU-WALTER, M.-A.; LEPLÉ, J.-C.; MAURY, S.; MOREL, A.; ODDOU-MURATORIO, S.; PILATE, G.; SANCHEZ, L.; SCOTTI, I.; SCOTTI-SAINTAGNE, C.; SEGURA, V.; TRONTIN, J.-F.; VACHER, C. Forest tree genomics: 10 achievements from the past 10 years and future prospects. **Annals of Forest Science**, v. 73, n. 1, p. 77-103, 2016. DOI: <https://doi.org/10.1007/s13595-015-0488-3>.

POKE, F. S.; VAILLANCOURT, R. E.; POTTS, B. M.; REID, J. B. Genomic research in *Eucalyptus*. **Genetica**, v. 125, n. 1, p. 79-101, 2005. DOI: <https://doi.org/10.1007/s10709-005-5082-4>.

RESENDE JUNIOR, M. F. R.; MUÑOZ, P.; ACOSTA, J. J.; PETER, G. F.; DAVIS, J. M.; GRATTAPAGLIA, D.; RESENDE, M. D. V. de; KIRST, M. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. **New Phytologist**, v. 193, n. 3, p. 617-624, 2012. DOI: <http://dx.doi.org/10.1111/j.1469-8137.2011.03895.x>.

RESENDE, M. D. V. de; LOPES, P. S.; SILVA, R. L. da; PIRES, I. E. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, v. 56, p. 63-77, 2008.

RESENDE, M. D. V. de; RESENDE JUNIOR, M. F. R.; SANSALONI, C. P.; PETROLI, C. D.; MISSIAGGIA, A. A.; AGUIAR, A. M.; ABAD, J. M.; TAKAHASHI, E. K.; ROSADO, A. M.; FARIA, D. A.; PAPPAS JUNIOR, G. J.; KILIAN, A.; GRATTAPAGLIA, D. Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex

traits in forest trees. **New Phytologist**, v. 194, p. 116-128, 2012. DOI: <http://dx.doi.org/10.1111/j.1469-8137.2011.04038.x>.

RESENDE, R. T.; RESENDE, M. D. V. de; SILVA, F. F. S.; AZEVEDO, C. F. A.; TAKAHASHI, E. K. T.; SILVA JUNIOR, O. B. da; GRATTAPAGLIA, D. Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in *Eucalyptus*. **New Phytologist**, v. 213, p. 1287-1300, 2017. DOI: <https://doi.org/10.1111/nph.14266>.

RESENDE, R. T.; RESENDE, M. D. V.; SILVA, F. F.; AZEVEDO, C. F.; TAKAHASHI, E. K.; SILVA JUNIOR, O. B.; GRATTAPAGLIA, D. Assessing the expected response to genomic selection of individuals and families in *Eucalyptus* breeding with an additive-dominant model. **Heredity**, v. 119, p. 245-255, 2017. DOI: <http://dx.doi.org/10.1038/hdy.2017.37>.

REZENDE, G. D. S. P.; RESENDE, M. D. V.; ASSIS, T. F. *Eucalyptus* breeding for clonal forestry. In: FENNING, T. (ed.). **Challenges and opportunities for the world's forests in the 21st Century**. Dordrecht: Springer Netherlands, 2014. p. 393-424.

ROSADO, C. C. G.; GUIMARÃES, L. M. da S.; FARIA, D. A.; RESENDE, M. D. V. de; CRUZ, C. D.; GRATTAPAGLIA, D.; ALFENAS, A. C. QTL mapping for resistance to *Ceratocystis* wilt in *Eucalyptus*. **Tree Genetics & Genomes**, v. 12, n. 4, article 72, 2016. 10 p.

SANSALONI, C. P.; PETROLI, C. D.; CARLING, J.; HUDSON, C. J.; STEANE, D. A.; MYBURG, A. A.; GRATTAPAGLIA, D.; VAILLANCOURT, R. E.; KILIAN, A. A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. **Plant Methods**, v. 6, 30 2010. DOI: <https://doi.org/10.1186/1746-4811-6-16>.

SANSALONI, C.; PETROLI, C.; JACCOUD, D.; CARLING, J.; DETERING, F.; GRATTAPAGLIA, D.; KILIAN, A. Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. **BMC Proceedings**, v. 5, suppl. 7, p. P54, 2011. DOI: <https://doi.org/10.1186/1753-6561-5-S7-P54>.

SENIOR, M. L.; CHIN, E. C. L.; LEE, M.; SMITH, J. S. C.; STUBER, C. W. Simple sequence repeat markers developed from maize sequences found in the GENBANK database: map construction. **Crop Science**, v. 36, n. 6, p. 1676-1683, 1996. DOI: <https://doi.org/10.2135/cropsci1996.0011183X003600060043x>.

SILVA-JUNIOR, O. B.; FARIA, D. A.; GRATTAPAGLIA, D. A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing 240 *Eucalyptus* tree genomes across 12 species. **New Phytologist**, v. 206, n. 4, p. 1527-1540, 2015.

SILVA-JUNIOR, O. B.; GRATTAPAGLIA, D. Genome-wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of *Eucalyptus grandis*. **New Phytologist**, v. 208, n. 3, p. 830-845, 2015. DOI: <https://doi.org/10.1111/nph.13505>.

SILVA, P. H. M. D.; BRUNE, A.; ALVARES, C.; AMARAL, W.; TEIXEIRA, M.; GRATTAPAGLIA, D.; PAULA, R. C. Selecting for stable and productive families of *Eucalyptus urophylla* S.T. Blake across a country wide range of climates in Brazil. **Canadian Journal of Forest Research**, v. 49, p. 87-95, 2018. DOI: <https://doi.org/10.1139/cjfr-2018-0052>.

SILVA, P. H. M.; SEBBENN, A. M.; GRATTAPAGLIA, D.; CONTI JUNIOR, J. L. Realized pollen flow and wildling establishment from a genetically modified eucalypt field trial in Southeastern Brazil. **Forest Ecology and Management**, v. 385, p. 161-166, 2017. DOI: <https://doi.org/10.1016/j.foreco.2016.11.043>.

SILVA, P. H. M.; SEBBENN, A. M.; GRATTAPAGLIA, D. Pollen-mediated gene flow across fragmented clonal stands of hybrid eucalypts in an exotic environment. **Forest Ecology and Management**, v. 356, p. 293-298, 2015a. DOI: <https://doi.org/10.1016/j.foreco.2014.12.005>.

SILVA, P. H. M.; SEBBENN, A. M.; GRATTAPAGLIA, D.; SEBBENN, A. M. Use of genetic markers to build a new generation of *Eucalyptus pilularis* breeding population. **Silvae Genetica**, v. 64, n. 4, p. 170-181, 2015. 2015. DOI: <https://doi.org/10.1515/sg-2015-0016>.

STEANE, D. A.; NICOLLE, D.; SANSALONI, C. P.; PETROLI, C. D.; CARLING, J.; KILIAN, A.; MYBURG, A. A.; GRATTAPAGLIA, D.; VAILLANCOURT, R. E. Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (Myrtaceae) using high-throughput, genome-wide genotyping. **Molecular Phylogenetics and Evolution**, v. 59, n. 1, p. 206-224, 2011. DOI: <https://doi.org/10.1016/j.ympcv.2011.02.003>.

STRAUSS, S. H.; LANDE, R.; NAMKOONG, G. Limitations of molecular-marker-aided selection in forest tree breeding. **Canadian Journal of Forest Research**, v. 22, n. 7, p. 1050-1061, 1992. DOI: <https://doi.org/10.1139/x92-140>.

TAN, B.; GRATTAPAGLIA, D.; MARTINS, G. S.; FERREIRA, K. Z.; SUNDBERG, B.; INGVARSSON, P. K. Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F1 hybrids. **BMC Genomics**, v. 17, p. 110, 2017. DOI: <https://doi.org/10.1186/s12870-017-1059-6>.

TAN, B.; GRATTAPAGLIA, D.; WU, H. X.; INGVARSSON, P. K. Genomic relationships reveal significant dominance effects for growth in hybrid *Eucalyptus*. **Plant Science**, v. 267, p. 84-93, 2018. DOI: <https://doi.org/10.1016/j.plantsci.2017.11.011>.

TARAMINO, G.; TINGEY, S. Simple sequence repeats for germplasm analysis and mapping in maize. **Genome**, v. 39, n. 2, p. 277-87, 1996. DOI: <https://doi.org/10.1139/g96-038>.

TELFER, E. J.; STOVOLD, G. T.; SILVA-JUNIOR, O. B.; GRATTAPAGLIA, D.; DUNGEY, H. S. Parentage reconstruction in *Eucalyptus nitens* using SNPs and microsatellite markers: a comparative analysis of marker data power and robustness. **PLoS One**, v. 10, n. 7, p. e0130601, 2015. DOI: <https://doi.org/10.1371/journal.pone.0130601>.

TORRES-DINI, D.; NUNES, A. C. P.; AGUIAR, A. V.; NIKICHUCK, A. V.; CENTURIÓN, C.; CABRERA, M.; MORAES, M. L. T.; RESENDE, M. D. V.; SEBBENN, A. M. Clonal selection of *Eucalyptus grandis* x *Eucalyptus globulus* for productivity, adaptability, and stability, using SNP markers. **Silvae Genetica**, v. 65, n. 2, p. 30, 2016. DOI: <https://doi.org/10.1515/sg-2016-0014>.

TUSKAN, G. A.; GROOVER, A. T.; SCHMUTZ, J.; DIFAZIO, S. P.; MYBURG, A.; GRATTAPAGLIA, D.; SMART, L. B.; YIN, T.; AURY, J.-M.; KREMER, A.; LEROY, T.; LE PROVOST, G.; PLOMION, C.; CARLSON, J. E.; RANDALL, J.; WESTBROOK, J.; GRIMWOOD, J.; MUCHERO, W.; JACOBSON, D.; MICHENER, J. K. Hardwood tree genomics: unlocking woody plant biology. **Frontiers in Plant Science**, v. 9, 2018. DOI: <https://doi.org/10.3389/fpls.2018.01799>.

VAN EENENNAAM, A. L.; WEIGEL, K. A.; YOUNG, A. E.; CLEVELAND, M. A.; DEKKERS, J. C. M. Applied animal genomics: results from the field. **Annual Review of Animal Biosciences**, v. 2, p. 105-139, 2014. DOI: <https://doi.org/10.1146/annurev-animal-022513-114119>.

VERHAEGEN, D.; PLOMION, C. Genetic mapping in *Eucalyptus urophylla* and *Eucalyptus grandis* using RAPD markers. **Genome**, v. 39, p. 1051-1061, 1996.

VERHAEGEN, D.; PLOMION, C.; GION, J.-M.; POITEL, M.; COSTA, P.; KREMER, A. Quantitative trait dissection analysis in *Eucalyptus* using RAPD markers, 1. Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. **Theoretical and Applied Genetics**, v. 95, p. 597-608, 1997. DOI: <https://doi.org/10.1007/s001220050601>.

WEISSENBACH, J.; GYAPAY, G.; DIB, C.; VIGNAL, A.; MORISSETTE, J.; MILLASSEAU, P.; VAYSSEIX, G.; LATHROP, M. A 2nd-generation linkage map of the human genome. **Nature**, v. 359, n. 6398, p. 794-801, 1992. DOI: <https://doi.org/10.1038/359794a0>.

WHITE, T. L.; ADAMS, W. T.; NEALE, D. B. **Forest Genetics**. Cambridg: CABI Publishing, 2007. 682 p.

WILLIAMS, C. G.; NEALE, D. B. Conifer wood quality and marker-aided selection: a case-study. **Canadian Journal of Forest Research**, v. 22, n. 7, p. 1009-1017, 1992. DOI: <https://doi.org/10.1139/x92-135>.

ZARPELON, T. G.; GUIMARÃES, L. M. S.; COUTINHO, M. M.; CÁPUA NETO, B.; FARIA, D. A.; GRATTAPAGLIA, D.; ALFENAS, A. C. QTL associated with resistance to defoliation (*Cylindrocladium pteridis*) in *Eucalyptus* spp. **BMC Proceedings**, v. 5, n. 7, p. P27, 2011. DOI: <https://doi.org/10.1186/1753-6561-5-S7-P27>.

ZARPELON, T. G.; GUIMARÃES, L. M. da S.; FARIA, D. A.; COUTINHO, M. M.; CÁPUA NETO, B.; TEIXEIRA, R. U.; GRATTAPAGLIA, D.; ALFENAS, A. C. Genetic mapping and validation of QTLs associated with resistance to Calonectria leaf blight caused by *Calonectria pteridis* in *Eucalyptus*. **Tree Genetics & Genomes**, v. 11, n. 1, p. 803, 2015. DOI: <https://doi.org/10.1007/s11295-014-0803-4>.

