

**ALISSON ESDRAS COUTINHO**

**SELEÇÃO GENÔMICA AMPLA VIA REDES NEURAS ARTIFICIAIS**

**RECIFE-PE**

**2013**

**ALISSON ESDRAS COUTINHO**

**SELEÇÃO GENÔMICA AMPLA VIA REDES NEURAIS ARTIFICIAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Agronomia “Melhoramento Genético de Plantas”, da Universidade Federal Rural de Pernambuco, como parte dos requisitos para obtenção do grau de Mestre em Agronomia.

**COMITÊ DE ORIENTAÇÃO:**

Professor Dr. José Luiz Sandes de Carvalho Filho, Orientador – UFRPE

Professor Dr. Diogo Gonçalves Neder, Co-orientador – UEPB

**RECIFE-PE**

**2013**

Ficha catalográfica

C871s      Coutinho, Alisson Esdras  
              Seleção genômica ampla via redes neurais artificiais /  
Alisson Esdras Coutinho. -- Recife, 2013.  
              63 f.: il

              Orientador(a): José Luiz Sandes de Carvalho Filho.

              Dissertação (Programa de Pós-Graduação em  
Melhoramento Genético de Plantas) – Universidade  
Federal Rural de Pernambuco, Departamento de  
Agronomia, Recife, 2013.

              Referências.

              1. Melhoramento genético 2. Correlação 3. Marcadores  
moleculares I. Carvalho Filho, José Luiz Sandes de,  
orientador II. Título

CDD 581.15

# SELEÇÃO GENÔMICA AMPLA VIA REDES NEURAIS ARTIFICIAIS

**ALISSON ESDRAS COUTINHO**

Dissertação defendida e aprovada pela Banca Examinadora em: 28/02/2013.

**ORIENTADOR:**

---

Prof. Dr. José Luiz Sandes de Carvalho Filho - UFRPE

**EXAMINADORES:**

---

Prof. Dr. Péricles de Albuquerque Melo Filho - UFRPE

---

Prof. Dr. Diogo Gonçalves Neder - UEPB

---

Dr. Roberto de Albuquerque Melo – UFRPE/PNPD

**Recife, PE  
Fevereiro, 2013**

A Deus,

Ofereço

Ao meu pai Antonio Faustino Coutinho (In memoriam), à minha mãe, Maria José de Araújo Coutinho, minha irmã, Amanda Nelma Coutinho e ao meu sobrinho Luís Henrique pelo amor e dedicação,

Dedico

## **Agradecimentos**

A Deus primeiramente por ter permitido minha chegada até aqui, pela superação, aprendizado, saúde para prosseguir sempre com força e garra.

Especialmente à minha mãe pela incrível dedicação e pela grandiosa mulher que incansavelmente lutou pelo meu melhor e por um futuro brilhante.

À minha irmã pelo caráter e determinação. À minha tia Elizabete pelo apoio e admiração. E a toda minha grande família pela alegria que transmitem. Ao meu grande amigo de velhos tempos Edson Filho pelo respeito e confiança.

Ao meu orientador Professor Dr. José Luiz Sandes de Carvalho Filho pelos ensinamentos transmitidos em sala de aula, contribuição e atenção no desenvolvimento deste trabalho.

Ao meu co-orientador Professor Dr. Diogo Gonçalves Neder que mesmo distante preocupou-se sempre com a execução deste trabalho e pelos conhecimentos inovadores transmitidos despertando o prazer pela ciência.

Aos Professores do mestrado: Clodoaldo José da Anunciação Filho; Edson Silva; Francisco Oliveira; Gerson Quirino; José Luiz Sandes; Mario Lira Junior, Péricles de Albuquerque, Vívian Loges, pelos ensinamentos e dedicação ao Programa de Melhoramento Genético de plantas da UFRPE.

Aos companheiros de formação Adriana, Alysson Jalles, Ana Luisa, Ana Rafaela, Claudia, Felipe, Guilherme, Gustavo, Hudson, Horace, Ismael Gaião, Gustavo, Jayne, José Carlos, João Filipi, Kessyana, Lucas, Lenivânia, Lindomar, Marciana, Marília, Natália, Paulo Ricardo, Rafaela, Rebeca, Ramon, Rodolfo, Renata, Samy, Tiago Batista e Thiago Prates e também aos meus amigos de graduação: Bruno Freitas, Flavio Mariano, Daniel Goulart, Arci Domingo, Helder Cruz, Rodrigo Barbosa, Carlos Fabin, Walesca Costa, Vanessa Ali, Pedro Henrique. Aos companheiros de pesquisa análises Eliane Cristina e Mairykon. Aos amigos Edvaldo Lopes, Wellington Leite, Leônia Torres e africanos: guinenses, caboverdeanos e a minha amiga angolana Vivi Domingos. E a amiga Lívia Maria pela amizade e contribuição na revisão ortográfica.

Ao apoio institucional e financeiro da Universidade Federal Rural de Pernambuco - UFRPE. À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pela concessão da bolsa.

**Muito obrigado!**

*Faça todo o bem que puder,  
Usando todos os meios que puder,  
De todas as maneiras que puder...  
Para todas as pessoas que puder,  
Durante o maior tempo que puder.*

*John Wesley*

## RESUMO

A seleção genômica ampla (GWS) consiste no uso de uma grande quantidade de marcadores moleculares cobrindo todo o genoma. Esta técnica foi proposta em 2001 por Mewissen como uma nova alternativa de seleção, em virtude das limitações da seleção assistida por marcadores (MAS). A alta densidade de marcadores é o que caracteriza a seleção genômica. O fato que teve uma contribuição para a aplicação da GWS foi o desenvolvimento dos marcadores moleculares do tipo SNP (single nucleotide polymorphism) que tornou o custo de genotipagem mais baixo. Com isso, possibilitou o uso de uma alta densidade de marcadores estimando o efeito de genes de caracteres quantitativos. Portanto, a GWS passou a ser alvo dos estudos de pesquisa, com o objetivo de investigar sua eficiência em programas de melhoramento genético de plantas e animais. Os possíveis benefícios de seu uso são a redução do ciclo de melhoramento, propiciando o ganho por unidade de tempo e diminuição de custos, a exemplo de espécies perenes, sejam elas frutíferas ou florestais. O sucesso da GWS está atrelado a escolha do método de predição na estimação dos efeitos dos marcadores. Existem vários métodos de predição dos valores genéticos genômicos (VGGs), tais como: quadrados mínimos, Bayes A, Bayes B, Aprendizado de máquina, BLUP/GWS. A escolha dentre estes preditores baseia-se na capacidade de predição e recurso computacional disponível para a análise. Alguns trabalhos demonstram comparações entre os métodos de predição, aplicando a seleção genômica via simulação. Assim, neste trabalho, visou-se aplicar outro método, qual seja: redes neurais artificiais (Artificial Neural Network - ANNs), com a finalidade de estimar os efeitos dos marcadores comparados ao RR-BLUP/GWS sob diferentes densidades de marcadores (100 e 200), níveis de desequilíbrio de ligação (0,5; 0,6; 0,7; 0,8; 0,9 e 1) e porcentagens de herdabilidade. As simulações foram realizadas no programa R, no qual foram obtidas as correlações entre os valores fenotípicos e os valores genéticos genômicos preditos (VGGs), e entre os valores genotípicos e os VGGs. Os resultados demonstraram superioridade das ANNs na predição VGGs nos tratamentos com maior densidade de marcadores, paralelo a níveis mais altos de desequilíbrio de ligação e maior herdabilidade. Por outro lado, houve correlações inferiores sob maior densidade de marcadores.

Termos para indexação: Melhoramento vegetal, correlação, marcadores moleculares.



## **ABSTRACT**

The genome wide selection (GWS) is the use of a large amount of molecular markers covering the whole genome. This technique was proposed in 2001 by Mewissen as a new alternative selection due to limitations of marker assisted selection (MAS). The high density of markers is what characterizes the genomic selection. The fact that had a contribution to the implementation of GWS was the development of molecular markers SNP (single nucleotide polymorphism) that has the lowest cost of genotyping. With this enables the use of a high density markers and then estimating the effect quantitative trait gene. Therefore, the GWS has become the subject of research studies seeking investigate its efficiency in breeding programs for plants and animals. The potential benefits of its use are reducing the breeding cycle providing the gain per unit time and cost savings such species are perennial fruit trees or forestry. GWS's success is linked to choice of prediction method for estimating the effects of the markers. Several methods for prediction of genomic breeding values (VGGs) as least squares, Bayes A, Bayes A, Machine Learning, BLUP/GWS. The choice among these predictors is based on the predictive capacity and computational resource available for analysis. Some studies have shown comparisons of prediction methods applied to genomic selection via simulation. Thus, this study aimed to apply another method, neural artificial networks (ANNs) in order to estimate the effects of markers compared to RR-BLUP/GWS under different densities of markers (100 and 200), levels of linkage disequilibrium (0.5, 0.6, 0.7, 0.8, 0.9 and 1) and percentages of heritability, which constituted the treatments. The simulations were performed in the program R where the correlations were obtained between the predicted VGGs and phenotypic values and genotypic values between and predicted VGGs. The results showed superiority of ANNs in predicting VGGs in treatments with high density markers parallel to higher levels of linkage disequilibrium and high heritability. On the other hand, there have correlations less in higher density of markers.

Index terms: Plant breeding, correlation, molecular markers

## LISTA DE TABELAS

### CAPÍTULO II - SIMULAÇÃO DOS VALORES FENOTÍPICOS E GENOTÍPICOS VIA RR- BLUP/GWS E REDES NEURAS ARTIFICIAIS

**Tabela 1.** Valores das médias das correlações entre os valores fenotípicos e preditos pelas Redes Neurais Artificiais e via BLUP/GWS para 100 genes, 20% e 40%de herdabilidade e desequilíbrios de ligação de 0,5; 0,6; 0,7; 0,8; 0,8; 0,9 e 1..60

**Tabela 2.** Valores das médias das correlações entre os valores genotípicos e preditos pelas Redes Neurais Artificiais e via BLUP/GWS para 100 genes, 20% e 40%de herdabilidade e desequilíbrios de ligação de 0,5; 0,6; 0,7; 0,8; 0,9 e 1.....61

**Tabela 3.** Valores das médias das correlações entre os valores fenotípicos e preditos pelas Redes Neurais Artificiais e via BLUP/GWS para 200 genes, 20% e 40%de herdabilidade e desequilíbrios de ligação de 0,5; 0,6; 0,7; 0,8; 0,8; 0,9 e 1..62

**Tabela 4.** Valores das médias das correlações entre os valores genotípicos e preditos pelas Redes Neurais Artificiais e via BLUP/GWS para 200 genes, 20% e 40%de herdabilidade e desequilíbrios de ligação de 0,5; 0,6; 0,7; 0,8; 0,9 e 1.....63

## SUMÁRIO

<b>CAPÍTULO I - REVISÃO BIBLIOGRÁFICA .....</b>	<b>12</b>
1. INTRODUÇÃO GERAL .....	13
2. REVISÃO BIBLIOGRÁFICA .....	17
2.1 Seleção genômica.....	17
2.2. Métodos de predição .....	22
REFERÊNCIAS.....	27
<b>CAPÍTULO II - PREDIÇÃO DOS VALORES FENOTÍPICOS E GENOTÍPICOS VIA RR- BLUP/GWS E REDES NEURAS ARTIFICIAIS.....</b>	<b>36</b>
RESUMO.....	38
ABSTRACT .....	40
1. INTRODUÇÃO .....	41
2. MATERIAL E MÉTODOS .....	42
2.1. Simulações dos dados fenotípicos e genotípicos.....	43
2.2. Estimção dos valores genéticos genômicos.....	44
2.2.1. Redes neurais artificiais (ANNs).....	44
2.2.2. RR-BLUP/GWS .....	45
3. RESULTADOS E DISCUSSÃO.....	47
4. CONCLUSÕES .....	55
REFERÊNCIAS.....	56

## **CAPÍTULO I - REVISÃO BIBLIOGRÁFICA**

## 1. INTRODUÇÃO GERAL

O melhoramento genético de plantas é uma ciência em que abrange diversas áreas como: botânica, fitopatologia, bioquímica, solos, estatística, genética, biotecnologia, bioinformática entre outras. Segundo Poelhman (1965) (citado por BESPALHOK FILHO et al., 2012) é a arte ou ciência de melhorar geneticamente plantas para o benefício da humanidade. Então, o melhoramento como arte depende da capacidade de observação do melhorista para que se possam selecionar os indivíduos com características de importância econômica (BESPALHOK FILHO et al., 2012).

A observação visual baseada nas características morfológicas, seleção fenotípica, é praticada a cerca de 10.000 anos desde no início da domesticação de plantas (BESPALHOK FILHO et al., 2012). Os primeiros agricultores já praticavam seleção por meio da observação de plantas em campo, separando as sementes dos tipos mais desejáveis. Portanto, a seleção fenotípica até hoje tem sido o ponto de partida mais importante na obtenção novas cultivares (BORÉM, 1997).

Uma importante contribuição do melhoramento foi o caso do milho híbrido nos Estados Unidos, que teve um aumento de produção devido a obtenção de cultivares híbridas. O cultivo da soja no Brasil também é outro exemplo. Sendo uma espécie de dia curto, quando plantada em locais de baixas latitudes, tem seu período vegetativo reduzido, apresentando florescimento precoce, com baixa estatura e produtividade. A introdução e expansão da soja no Brasil foram possíveis por meio do cruzamento de cultivares americanas com período juvenil longo para obtenção e seleção de genótipos com esta característica. Devido a isto, tornou-se possível o cultivo da soja no cerrado brasileiro e sua expansão para os estados do Maranhão, Piauí, Tocantins e Pará (ALMEIDA et al., 1999).

O desenvolvimento de variedades semi-anãs de trigo pelo melhorista americano Norman Erlest Borlaug constitui um destaque do melhoramento genético na revolução verde, contribuindo de maneira significativa para o aumento de produtividade de grãos, evitando uma provável fome em massa decorrente do previsto aumento da população pela teoria malthusiana, quando Thomas Robert Malthus profetizou que a população crescia em progressão geométrica, enquanto que a produção de alimentos aumentaria em progressão aritmética (ALLARD, 1971). Isto não foi possível devido à obtenção de variedades de menor porte com maior

COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

eficiência fisiológica ao absorverem o nitrogênio e que não apresentavam acamamento, com isso reduzindo perdas e possibilitando colheitas mais eficientes com a introdução de máquinas colheitadeiras (BORÉM, 1997). Em meados dos anos 60, Norman levou estas variedades e um todo pacote tecnológico para Índia e Paquistão. Entre 1964 e 2001, a produção na Índia passou de 12 para 75 milhões de toneladas, enquanto que no Paquistão aumentou de 4,5 para 22 milhões de toneladas (ORTIZ et al., 2007).

Nos últimos 50 anos, o número de habitantes do mundo quase triplicou, passando de 2 bilhões e 500 mil em 1950 e atingindo 7 bilhões em 2011. Mesmo que, na maioria dos países, as taxas de natalidade estejam diminuindo, a população mundial segue aumentando e, segundo se prevê, chegará a 8 bilhões e 900 mil até 2050 (UNFPA, 2012). Com isso, a demanda por alimentos aumentará e a biotecnologia vem como uma ferramenta útil para o melhoramento, para se ter uma maior eficiência no processo de seleção bem como reduzir o tempo de obtenção de novas cultivares (PEREIRA et al, 2012).

Como exemplo dos avanços de técnicas moleculares, ferramentas foram propostas, podendo ser exemplificada pela seleção assistida por marcadores moleculares desenvolvidos por Lande e Thompson (1990), na qual se pode integrar com os métodos convencionais de seleção nos fenótipos. Dentre as utilidades da MAS estão: o uso no retrocruzamento, como estratégia de reduzir o tempo necessário na recuperação do genoma recorrente, na realização de piramidação de alelos de resistência buscando obter cultivares resistentes a diferentes raças de um mesmo patógeno, ou a distintos patógenos (GUIMARÃES et al., 2006; ALZATE-MARIN et al., 2005). No entanto, esta metodologia, quando aplicada à características quantitativas, não obteve sucesso em função de detectar um pequeno número de QTLs de grande efeito e da grande influência ambiental nos caracteres quantitativos (DEKERS, 2004).

Com isto, em 2001 foi proposto por Mewissen uma nova metodologia denominada Seleção Genômica Ampla, (Genome Wide Selection (GWS). Esta nova metodologia consiste no uso de uma grande quantidade de marcadores saturando o genoma que então tem seus efeitos estimados. Paralela a GWS também houve o surgimento dos marcadores SNPs (*Single Nucleotide Polimorphism*) contribuindo para o aumento da genotipagem e a redução do preço por data point (JENKINS e GIBSON, 2002).

Para a estimação dos efeitos destes marcadores diversos métodos foram proposto, tais como: BLUP/GWS, Bayes A e Bayes B (MEUWISSEN et al., 2001), aprendizado de máquina (LONG et al., 2007), regressão RKHS (Reproducing Kernel Hilbert Spaces) (GIANOLA et al., 2008), LASSO Bayesiano (De LOS CAMPOS, 2009) e Regressão via Quadrados Mínimos Parciais (PLSR) (SOLBERG et al., 2009).

O método dos quadrados mínimos apresenta uma acurácia bastante baixa nas predições dos valores genéticos reais em relação aos métodos Bayes A e Bayes B. Por outro lado, Bayes A é exigente em termo computacional (tempo de conclusão de análise) o que dificulta seu uso, além de apresentar uma correlação entre os valores genéticos e os valores genéticos estimados mais baixos que o Bayes B (MEWISSEN, 2001).

Quanto ao BLUP/GWS, os efeitos alélicos podem ser ajustados simultaneamente, mesmo quando existem mais efeitos alélicos a serem preditos do que o número total de observações fenotípicas (RESENDE, 2008). Além de permitir o uso simultâneo de grande número de informações provenientes de diferentes gerações, locais e idades, gerando estimativas e predições mais precisas, além de não ser exigente em recurso computacional tão elevado quanto o Bayes A e B. (HENDERSON, 1984; RESENDE, 2002).

Outra alternativa para predizer são as redes neurais artificiais (Artificial Neural Network - ANNs) que trabalham semelhante ao neurônio biológico e têm a capacidade de aprender com ou sem professor e fazer generalizações do que foi anteriormente acumulado (LUDWING JR; MONTGOMERY, 2007). Esta técnica de inteligência artificial foi a que mais evoluiu nos últimos anos, despertando o interesse de muitos pesquisadores, apresentando um bom desempenho na solução de problemas como o reconhecimento de padrões (MARTINELLI, 1999).

As ANNs são uma ferramenta computacional que tem a capacidade de reconhecer padrões de entrada e fornecer uma saída mediante aprendizado aplicado anteriormente, possuírem a habilidade de fazer generalizações para produzirem saídas adequadas (MESSIAS, 2009; CAMPOS, 2010). De uma forma mais técnica define como sistemas paralelos distribuídos compostos por unidades de processamento simples (nodos) que calculam determinadas funções matemáticas, normalmente não-lineares (BRAGA, 2008).

Assim, as ANNs podem ser aplicadas na agricultura, como por exemplo, na estimativa de produção de álcool (OLIVEIRA et al., 2010), na programação de irrigações (RESENDE, 2010), na quantificação das classes de uso de solo (SILVA et al., 2005), no comportamento de doenças de plantas (RUGGIERO et al., 2003). No melhoramento genético de plantas o seu uso pode ser útil em estudos de divergência genética exemplificado no trabalho de Barbosa (2011), no qual classificou acessos de mamão (*Carica papaya* L.) em grupos definidos pelas ANNs como técnicas de reconhecimento de padrões.

Diante do exposto, o principal objetivo deste trabalho foi verificar a eficiência das ANNs em predizer valores genéticos genômicos sob diferentes número de marcadores, níveis de herdabilidade e desequilíbrio de ligação.



## 2. REVISÃO BIBLIOGRÁFICA

### 2.1 Seleção genômica

O melhoramento genético depende da existência de variabilidade e de métodos eficientes de seleção que permitam encontrar indivíduos superiores para determinadas características, sejam elas quantitativas ou qualitativas. Diante disto, a seleção é uma etapa muito importante dentro de um programa de melhoramento genético de plantas, contribuindo para a eficiência da escolha dos genótipos superiores.

Com os avanços da biologia molecular e a busca por uma maior eficiência na transferência de fatores genéticos, marcadores moleculares foram desenvolvidos possibilitando a seleção indireta de características desejáveis de interesse agrônomo (CAIXETA et al., 2006). Diante destas novas tecnologias foi proposta por Lande e Thompson (1990) a seleção assistida por marcadores (MAS). Este tipo de técnica baseia-se na utilização de dados fenotípicos e de dados moleculares em ligação gênica próxima a locos controladores de características quantitativas (QTLs) onde somente os marcadores significativos são utilizados para seleção (RESENDE, 2007; XU et al., 2012).

A MAS é uma ferramenta que veio como uma forma de reduzir tempo no ciclo de melhoramento, desenvolver grandes populações segregantes por várias gerações e estimar parâmetros na seleção indireta (CAIXETA et al., 2006).

As características quantitativas são afetadas por um grande número de locos de pequenos efeitos, o que torna a seleção complexa. Em virtude disto, a seleção assistida por marcadores (MAS) não tem sido eficiente no melhoramento de características como: produtividade, estresses biótico e abiótico, devido às restrições como: grande influência do ambiente sob estes caracteres, que levam a uma detecção de um pequeno número de QTLs de grandes efeitos, os quais não explicam suficientemente toda a variação genética (RESENDE, 2007). Mesmo tendo sido utilizada com sucesso em instituições públicas e companhias multinacionais, a limitação do seu uso foi devido ao baixo custo benefício, à baixa disponibilidade de sistemas de genotipagem, o menor entendimento da arquitetura genética de características quantitativas, além da falta de recursos computacionais mais avançados (XU et al., 2012).

Portanto, a seleção genômica ampla ou genome wide selection (GWS) difere-se por não utilizar somente o efeito individual de um marcador ligado a determinado loco, e sim por usar o efeito de todos os marcadores para prever o desempenho, chegando a previsões mais precisas e também levando as respostas mais rápidas e a um custo menor nos programas de melhoramento, com destaque para as características quantitativas (JANNINK et al. 2010).

A seleção genômica ampla (GWS) surgiu primeiramente como uma nova ferramenta para ser aplicada em programas de melhoramento animal, com o objetivo também de ser mais um meio de implementação em programas de melhoramento animal (MEUWISSEN et al. 2001).

A GWS pode ser uma ferramenta de grande utilidade na aplicação no melhoramento genético de plantas – levando a uma redução no tempo do ciclo do melhoramento de plantas, proporcionando aumentos nos ganhos de seleção por unidade de tempo – quando comparado com a seleção fenotípica (XU et al., 2012). Como exemplo, o melhoramento de espécies florestais em que se tem um longo tempo entre as gerações (GRATTAPAGLIA e RESENDE, 2010). Assim, em populações de *Eucalyptus* a redução no ciclo pode ser de até 7 anos, proporcionando um ganho por unidade de tempo de 724% (RESENDE JUNIOR, 2010).

Existem vários relatos que analisaram perspectivas da seleção genômica por meio de estudos de simulação, como a comparação entre GWS e MARS (seleção recorrente assistida por marcadores) para um determinado número de QTL e nível de herdabilidade, utilizando diferentes quantidades de marcadores em que se constatou superioridade da seleção genômica ampla (BERNARDO e YU, 2007). Outro exemplo de aplicação foi com Dendê (*Elaeis guineensis*) em programas de melhoramento. Comparando-se a seleção fenotípica, a seleção recorrente assistida por marcadores (MARS) e a seleção genômica ampla (GWS) para avaliar a eficiência sob pequenas populações, custo e ganho por unidade de tempo, em que a GWS se mostrou novamente superior (WONG e BERNARDO, 2008).

A frequência de fenotipagem pode ser reduzida em virtude da seleção a ser realizada sob uma maior taxa de genotipagem, na qual se utiliza uma alta densidade de marcadores cobrindo todo o genoma (XU et al., 2012). Segundo estudos empíricos e de simulações, foi constatado que a GWS pode levar a uma redução no tempo do ciclo do melhoramento, proporcionando aumentos nos ganhos de seleção

COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

por unidade de tempo, quando comparado com a seleção fenotípica (LORENZANA e BERNARDO 2009; WONG e BERNARDO 2008)

Em um estudo de GWS com dados reais em eucalipto, utilizando marcadores DarTs (Diversity Array Technology) aplicado em duas populações, chegou-se a um aumento do ganho de seleção em torno de 100% para uma população e de 45 a 50% para outra comparada com a seleção fenotípica ao reduzir o ciclo de geração de novos materiais pela metade (RESENDE JUNIOR, 2010).

Na seleção genômica são utilizados todos os dados dos alelos estimados, haplótipos ou efeitos dos marcadores para então se calcular os valores genéticos genômicos. A seleção genômica ampla baseia-se no cálculo dos indivíduos, utilizando apenas dados genotípicos através de um modelo que foi treinado para indivíduos possuindo tanto dados genotípicos como fenotípicos, isto é, primeiramente os efeitos dos marcadores são estimados através de dados de uma população de estimação a partir de dados fenotípicos. Após a estimação dos efeitos dos marcadores os mesmos são submetidos em uma população de validação e a partir daí realiza-se a seleção daqueles marcadores que explicam maior parte da variância genética de um determinado caráter (MEUWISSEN et al., 2001).

Segundo Goddard e Hayes (2007) para implementação da seleção genômica ampla são necessários três conjuntos de populações: população de descoberta ou estimação, população de validação e população de seleção.

Na população de descoberta é utilizado um grande número de marcadores SNPs e uma quantidade menor de indivíduos fenotipados e genotipados para as características de interesse. Para o caráter em estudo na seleção são utilizadas equações de predição de valores genéticos genômicos, associando a cada marcador o seu efeito no caráter de interesse (RESENDE, 2008).

Na população de validação é utilizado um menor número de indivíduos, comparado com a população de descoberta submetidos à avaliação com os marcadores SNPs para os caracteres de interesse. Para verificar a acurácia na amostra independente usam-se as equações de predição dos valores genéticos genômicos e realiza-se uma análise de correlação com os valores fenotípicos observados. Enquanto que a população de seleção engloba os marcadores SNPs avaliados nos indivíduos submetidos à seleção não sendo mais necessário fazer fenotipagem, e sim aplicar as equações de predição para predizer os valores genéticos genômicos ou fenótipos dos candidatos a seleção (RESENDE, 2008).

Em virtude de a GWS fazer o uso de um grande número de marcadores saturando todo o genoma é importante destacar a questão dos custos com o uso dos marcadores. Desde a década de 1980, os marcadores têm sido usados como uma ferramenta no auxílio do desenvolvimento de cultivares. No início dos anos 90 a cobertura ampla do genoma tornou-se um fator limitante para a MAS, pois algumas técnicas como restriction fragment length polymorphism (RFLP) ou *simple sequence repeat* (SSR) tinham um alto custo para genotipagem por data point, chegando a mais que US\$ 1 (XU et al., 2012; HEFFNER, 2009). Nos anos seguintes, com os grandes avanços na genotipagem, tem-se reduzido o custo por data point dos marcadores moleculares. Em decorrência disto, o custo vem reduzindo radicalmente, tornando-se viável a análise com grandes populações (CAETANO, 2009; XU et al., 2012). Tal redução foi devido a descoberta de grande número de polimorfismo em nível de nucleotídeo que pode ser causado por inserção e deleção de uma única base no genoma, conhecido como single nucleotide polymorphism (SNP) em muitas espécies. E para que esta variação seja considerada SNP, a mesma tem de ocorrer em pelo menos 1% da população (FARIA, 2008; CAIXETA et al., 2006).

Os SNPs apresentam características importantes, que são a baixa taxa de mutação e facilidade de genotipagem, quando comparado a outros marcadores genéticos, além de serem a forma mais abundante de variação do DNA e permitirem a detecção de polimorfismos funcionais ou polimorfismos em forte desequilíbrio de ligação com os QTNs (RESENDE, 2008). Um fator importante em relação aos marcadores na GWS é que os locos marcados se encontram em desequilíbrio de ligação (DL) com o marcador. O DL é definido como associação não aleatória entre dois genes ou entre um QTL e um marcador, ou seja, uma associação não aleatória de alelos de locos diferentes (FLINT-GARCIA et al., 2003; RESENDE, 2007). Assim, o marcador ligado ao gene está em desequilíbrio de ligação, o que representa a covariância entre o polimorfismo de dois marcadores moleculares ou genes. Portanto, o uso de uma grande quantidade de marcadores é fundamental porque mesmo uma população que já passou por várias sucessivas combinações possa existir um marcador tão próximo do QTL que esteja em desequilíbrio de ligação (GUPTA, 2005).

O grau de desequilíbrio de ligação (D) entre dois loci pode ser estimado pela máxima verossimilhança a partir de uma frequência de genótipos diploides em uma amostra de uma população de cruzamento ao acaso (HILL, 1974).

Segundo Flint-Garcia et al., (2003) dentre os métodos estatísticos utilizados para mensurar o desequilíbrio de ligação, os mais comuns são: Delvin e Risch (1995) e Jorde (2000), considerando dois loci (A e B) com dois alelos cada, sendo respectivamente A, a, B e b, juntamente com suas frequências alélicas  $f_A$ ,  $f_a$ ,  $f_B$  e  $f_b$  e frequências genotípicas ( $f_{AB}$ ,  $f_{Ab}$ ,  $f_{aB}$  e  $f_{ab}$ ). As duas medidas de DL são  $r^2$  ou  $\Delta^2$  e  $D'$ . O cálculo é baseado na diferença entre as frequências observadas e esperadas:

$$D_{ab} = (f_{AB} - f_A f_B)$$

A primeira medida de desequilíbrio de ligação é o  $r^2$  ou  $\Delta^2$ , consiste no coeficiente de determinação ou no coeficiente de correlação entre dois loci, calculado por meio da seguinte equação:

$$r^2 = \frac{(D_{ab})^2}{\pi_A \pi_a \pi_B \pi_b}$$

Já o valor de  $D'$  é baseado nas frequências alélicas observadas através da equação:

$$|D'| = \frac{(D_{ab})^2}{\min(\pi_A \pi_b, \pi_a \pi_B)} \text{ for } D_{ab} < 0;$$

$$|D'| = \frac{(D_{ab})^2}{\min(\pi_A \pi_B, \pi_a \pi_b)} \text{ for } D_{ab} > 0$$

Os valores de  $r^2$  e  $D'$  variam de 0 a 1. No caso de  $r^2 = 1$  indica associação preferencial máxima entre alelos de pares de locos, isto é, dois dos quatro haplótipos estão ausentes na amostra, caracterizando uma situação de DL perfeito (ARDLIE, 2002; KRUGLYAK, 1999). O valor de 1 não é possível, a menos que os dois loci possuam as frequências alélicas idênticas. Enquanto que  $D'$  apresentará um valor entre 0 e 1 se as frequências alélicas não diferirem entre os loci, e menor

COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

que 1 se os quatro possíveis haplótipos observados supostamente ocorrerem sob eventos de recombinação entre os loci (FLINT-GARCIA et al., 2003).

Existem diversos fatores que interferem no DL, os quais influenciam de diferentes maneiras. Dentre estes fatores estão os responsáveis pela geração do DL: deriva genética (ARDLIE, 2002), mutação, autofecundações, populações pequenas, isolamento genético entre linhagens, subdivisão populacional, baixa taxa de recombinação, mistura populacional, seleção natural e artificial (MACKAY, 2001). Outros quebram o DL como: mutação, estrutura populacional, epistasia e conversão gênica, endogamia, sistema reprodutivo, inversão e conversão de genes (entre outros que podem afetar o DL, como o tamanho de uma população) (MACKAY, 2001; HEDRICK, 2010).

## **2.2. Métodos de predição**

Os modelos mistos são aqueles que apresentam tanto efeitos fixos como aleatórios, além do erro experimental e a constante  $\mu$  (CAMARINHA FILHO, 2003; ALMEIDA, 2010). Assim, o BLUP, Best Linear Unbiased Prediction, que é uma metodologia do modelo linear misto é de grande utilidade em trabalhos de genética e melhoramento (OLIVEIRA, 2005; RESENDE et al., 2006; VIANA, 2010). Dentre suas vantagens, estão: utilizar uma grande quantidade de informações oriundas de diferentes locais, gerações, idades gerando estimativas e predições mais precisas (HENDERSON, 1984).

O método dos quadrados mínimos não possui um bom desempenho quando o número de marcadores é maior que a quantidade de genótipos, tornando-se inviável seu uso quando se aplica uma alta densidade de marcadores (RESENDE JUNIOR, 2010). Segundo Whittaker (2000), em geral não se pode incluir todos os marcadores no modelo de regressão para se selecionar um subconjunto dos marcadores. Portanto, existem os métodos de regularização citando a Ridge Regression (RR), que evita problemas na seleção, mantendo todos os marcadores no modelo (PIEPHO, 2009).

Uma característica importante que o BLUP apresenta é a matriz de parentesco completa, tornando-se superior, o comparado com outros métodos de predição, principalmente para caracteres de baixa herdabilidade (SOBREIRA, 2009).

Exemplos da aplicação do BLUP podem ser visualizados na avaliação de desempenho de genótipos de linhagens de arroz em terras altas, empregando a metodologia dos modelos mistos (BORGES et al., 2010).

Avaliando a eficiência da GWS em florescimento do milho, se comparou o RR-BLUP com o Bayes A e Bayes B na estimação dos efeitos dos marcadores nos quais a RR-BLUP teve uma melhor acurácia que os métodos bayesianos. Além de a RR-BLUP ter obtido acurácia superior com o aumento proporção da amostra de treinamento (GUO et al., 2011). Tais resultados mostram que RR-BLUP é um método preferido para estimar os efeitos do marcador em GWS, proporcionando precisão de previsão superior ou comparável à BayesA e BayesB.

HAYES et al. (2009) avaliaram a efetividade prática da seleção genômica em gado de leite nos Estados Unidos, Austrália e Nova Zelândia. Concluíram que o método BLUP mostrou-se aproximadamente igual aos métodos mais complexos, em termos de acurácia. Isso ocorre para caracteres em que o modelo infinitesimal, muitos genes de pequenos efeitos, se aplica. Adicionalmente, o método BLUP é vantajoso porque a única informação a priori necessária é uma estimativa da variância genética aditiva do caráter. Os autores relataram também a importância da inclusão do efeito poligênico no modelo de avaliação genética, como forma de capturar e selecionar QTLs de baixa frequência não capturados pelos marcadores.

Considerada como um método também capaz de realizar previsões a partir de dados, as Redes Neurais Artificiais (ANNs) podem se tornar um meio para se efetuar previsões de valores genéticos genômicos na seleção de plantas. Assim, “as ANNs são definidas como dispositivos não lineares inspirados na funcionalidade dos neurônios aplicados no reconhecimento de padrões, na otimização e na previsão de sistemas complexos” (BRUMATTI, 2005).

As ANNs trabalham baseadas no cérebro humano, mas apresentam diferenças em relação às redes neurais naturais, destacando os neurônios. São formadas por unidades de processamentos simples, denominados nodos, que calculam determinadas funções matemáticas. As unidades estão interligadas por conexões chamadas pesos, diferente dos neurônios naturais que estão interligados por meio de sinapses. Os pesos têm a finalidade de ponderar as entradas recebidas por cada neurônio da rede (BRAGA et al., 2000).

As ANNs são compostas por uma camada de entrada ou de distribuição, possuindo somente nós correspondentes ao número de sinais de entrada, que têm a

função de representar a distribuição dos dados de entrada para cada neurônio da camada seguinte. Outro componente pode ser a presença ou não de camadas ocultas ou intermediárias, constituídas por um ou mais neurônios ocultos, que por sua vez são responsáveis por capacitar a rede a extrair estatísticas e possibilitar a representação de problemas que não sejam linearmente separáveis. Por último, existe a camada de saída, que tem, necessariamente, um número de neurônios igual ao número de sinais de saída da rede (LUDWING JR; MONTGOMERY, 2007).

Desse modo, as ANNs são algoritmos capazes de aproximar os valores de saída, baseados nos valores de entrada, sem o conhecimento prévio do modelo e da complexidade de seus mecanismos, como no exemplo entre a capacidade antioxidante de um determinado óleo essencial, dados de entrada, e sua composição química (CABRERA e PRIETO, 2010).

Os neurônios biológicos são divididos em três seções: o corpo da célula, os dendritos e o axônio, cada um com funções específicas, porém complementares (Figura 1). Os dendritos têm por função receber as informações, ou impulsos nervosos, oriundos de outros neurônios e conduzi-los até o corpo celular. Aqui, a informação é processada, e novos impulsos são gerados. Estes impulsos são transmitidos a outros neurônios, passando através do axônio até os dendritos dos neurônios seguintes. O ponto de contato entre a terminação axônica de um neurônio e o dendrito de outro é chamado de sinapse. É pelas sinapses que os nodos se unem funcionalmente, formando redes neurais. As sinapses funcionam como válvulas, e são capazes de controlar a transmissão de impulso, isto é, o fluxo da informação entre os nodos na rede neural. O efeito das sinapses é variável, e é esta variação que dá ao neurônio capacidade de adaptação (BRAGA, 2000).

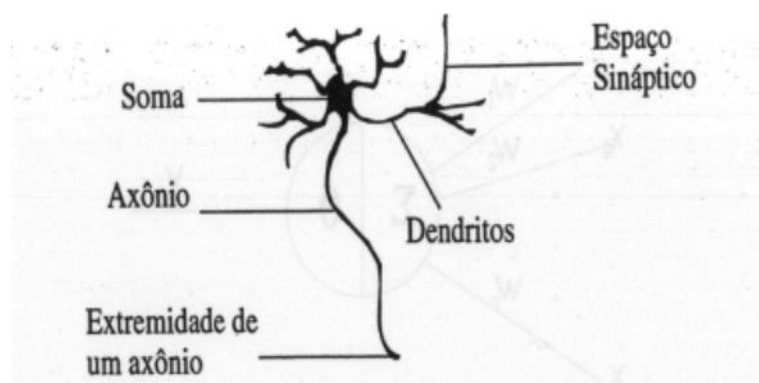


Figura 1: Neurônio biológico



Já o neurônio artificial é uma estrutura com característica lógica e matemática que procura simular a forma, o comportamento e as funções de um neurônio biológico. Desta maneira, os dendritos são substituídos por entradas, onde as ligações com o corpo celular artificial são realizadas através de elementos chamados de peso, simulando as sinapses. Os estímulos captados pelas entradas são processados pela função de soma ou de ativação, e o limiar de disparo do neurônio biológico foi substituído por uma função de transferência que determina a saída do neurônio, como mostra a figura 2 (LUDWING JR; MONTGOMERY, 2007).

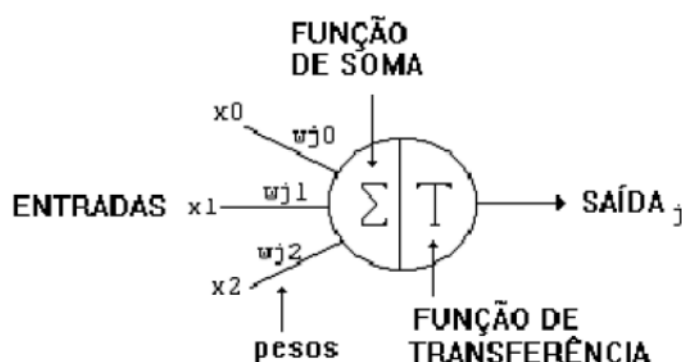


Figura 2: Neurônio artificial

Como as ANNs têm a capacidade de aprender por meio de exemplos, fazer extrapolações e interpolações do que aprenderem, são assim classificadas como: aprendizado supervisionado e aprendizado não supervisionado (BARBOSA, 2011; BRAGA et al., 2000).

No aprendizado supervisionado é apresentado para a rede um conjunto de padrões como dados de entrada, com seus correspondentes padrões de saída por um professor. Durante este processo a rede realiza um ajustamento dos pesos das conexões entre os elementos de processamento, ou seja, para cada padrão de entrada inserido na rede é comparado com a saída desejada e então realizados os ajustes necessários até que o erro quadrático seja minimizado (BARBOSA, 2011; BRAGA et al., 2000; CAMPOS et al., 2010).

Já no aprendizado não supervisionado não existe professor ou supervisor. A rede analisa os conjuntos de dados apresentados, determina algumas propriedades dos conjuntos de dados e “aprende” a refletir estas propriedades na sua saída. A rede utiliza padrões, regularidades e correlações para agrupar os conjuntos de

COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

dados em classes. As propriedades que a rede vai “aprender” sobre os dados podem variar em função do tipo de arquitetura utilizada e da lei de aprendizagem (BRAGA et al., 2000).

Por possuírem uma estrutura não linear, as ANNs conseguem captar características mais complexas dos dados, ao contrário da maioria das técnicas estatísticas (GALVÃO et al., 1999). Uma grande vantagem das ANNs em relação aos métodos convencionais é que elas não requerem informação detalhada sobre os processos físicos do sistema a ser modelado (SUDHEER et al., 2003).

A maioria das aplicações das ANNs é na engenharia, no controle de processos e reconhecimento de padrões (LUDWING JR; MONTGOMERY, 2007). Outras aplicações são previsões, simulações e otimização (BILOBROVEC, 2004). Portanto, esta ferramenta possui utilidade tanto na agricultura, como na previsão de preço de hortaliças (COSTA, 2012), no estudo de compactação de solos, na produtividade (WEBER, 2012) e no comportamento de doenças e pragas sob influência de fatores climáticos (RUGGIERO et al., 2003). E na caracterização de cultivares de arroz italiano foram classificadas 1779 amostras de acordo com suas variedades, por meio de uma RNA counterpropagation (CP-RNA) para classificação comercial. Constatou-se que a CP-RNA foi útil, predizendo 91 a 99% do conjunto teste das amostras (MARINI et al., 2004).

## REFERÊNCIAS

ARDLIE, K.G.; KRUGLYAK, L.; SEIELSTAD, M. Patterns of linkage disequilibrium in the human genome. **Nature Reviews Genetics**, v. 3, n. 4, p. 299-309, 2002.

ALMEIDA, R.V. **Seleção genética simultânea de famílias endogâmicas e topcrosses em culturas em culturas anuais via BLUP multivariado**. Viçosa-MG, 2010. Tese (Doutorado) – Universidade Federal de Viçosa, 2010.

DE ALMEIDA, L.A.; DE SOUZA KIHIL, R.A.; DE MIRANDA, M.A.C.; DE AZEVEDO CAMPELO, G.J. Melhoramento da soja para regiões de baixas altitudes. 1999. Disponível em: <[www.cptsa.embrapa.br/catalogo/livrorg/sojamelhoramento.pdf](http://www.cptsa.embrapa.br/catalogo/livrorg/sojamelhoramento.pdf)>. Acesso em: 01 dezembro. 2012.

ALZATE-MARIN, A.L.; CERVIGNI, G. D.; MOREIRA, M. A.; BARROS, E. G. Seleção assistida por marcadores moleculares visando ao desenvolvimento de plantas resistentes a doenças, com ênfase em feijoeiro e soja. **Fitopatologia Brasileira**, v. 30, n. 4, p. 333-342, 2005.

ARNHOLD, E.; MORA, F.; PACHECO, C.A.P.; CARVALHO, H.W.L.D. Prediction of genotypic values of maize for the agricultural frontier region in northeastern Maranhão, Brazil. **Crop Breeding and Applied Biotechnology**, v. 12, p. 5, 2012.

BARBOSA, C. D. Rede neural artificial aplicada à agricultura. **Vértices**, v. 13, p. 161-170, 2011.

BERNARDO, R. **Breeding for quantitative traits in plants**. Stemma Press, Minnesota. 2010. 380p.

BERNARDO, R.; YU, J. M. Prospects for genome wide selection for quantitative traits in maize. **Crop Science**, v. 47, n. 3, p. 1082-1090, 2007.

BESPALHOK F., J.C.; GUERRA, E.P.; OLIVEIRA, R. Introdução ao Melhoramento de Plantas. In: BESPALHOK F., J.C.; GUERRA, E.P.; OLIVEIRA, R. Melhoramento de Plantas. Disponível em <[www.bespa.agrarias.ufpr.br](http://www.bespa.agrarias.ufpr.br)> . p.1-9. Acesso em: 01 dez. 2012.

BILOBROVEC, M. MARÇAL, R.F.M.; PILATTI, L.A. Aplicações de redes neurais artificiais em simulações. XI SIMPEP - Bauru, SP, Brasil, 08 a 10 de novembro de 2004.

COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

BORÉM, A. **Melhoramento de plantas**. Viçosa, MG: Ed. UFV, 1997. 547 p.

BORGES, V.; SOARES, A. A.; REIS, M. S.; RESENDE, M. D.; CORNÉLIO, V. M. O.; LEITE, N. A.; VIEIRA, A. R. Desempenho genotípico de linhagens de arroz de terras altas utilizando metodologia de modelos mistos. **Bragantia**, v. 69, n. 4, p. 833-841, 2010.

BRAGA, A.P.; CARVALHO, A.P.L.; LUDEMIR, T.B. **Redes neurais artificiais: teorias e aplicações**. Rio de Janeiro: LTC, 2000. v.1. 250p.

BRUMATTI, M. Redes neurais artificiais. Disponível em: <[www.pdf-traffic-simulator.googlecode.com/...Redes%20Ne](http://www.pdf-traffic-simulator.googlecode.com/...Redes%20Ne)>. Acesso em: 18 Julho 2012.

BUENO, L.C.S. Melhoramento de plantas: princípios e procedimentos. Lavras: UFLA, 2001.

CABRERA, A. C.; PRIETO, J. M. Application of artificial neural networks to the prediction of the antioxidant activity of essential oils in two experimental in vitro models. **Food Chemistry**, v. 118, n. 1, p. 141-146, 2010.

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

CAIXETA, E. V.; OLIVEIRA, A. C. B.; BRITO, G. G.; SAKIYAMA, N. S. Tipos de marcadores moleculares. In: CAIXETA, E.; BORÉM, A. (Org) **Marcadores moleculares**. Viçosa-MG, 2006. p. 9-66.

CAMARINHA FILHO, J.A. Nota metodológica sobre modelos lineares mistos. Universidade Federal do Paraná. Curitiba-PR,2003. Disponível em: <[www.est.ufpr.br/rt/jom03a.pdf](http://www.est.ufpr.br/rt/jom03a.pdf)>. Acesso em: 20 set. 2012.

CAMPOS, J.R.; LOTUFO, ADP.; MINUSSI, C.R.; LOPES, M.L.M. Implementação de redes neurais artificiais utilizando a linguagem de programação JAVA. **Proceedings of the 9th Brazilian Conference on Dynamics Control and their Applications, 2010**. Disponível em: < [www.sbmec.org.br/dincon/trabalhos/PDF/.../67995.pdf](http://www.sbmec.org.br/dincon/trabalhos/PDF/.../67995.pdf)>. Acesso em: 20 setembro 2012.

COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

COSTA, A. L.; CARVALHO, A. C. L. F.; MENDES, R. Redes neurais para previsão de preços de hortaliças folhosas. Disponível em: <[www.sbiagro.org.br/pdf/iii\\_congresso/Artigo38.pdf](http://www.sbiagro.org.br/pdf/iii_congresso/Artigo38.pdf)>. Acesso em: 29 Novembro 2012.

DE LOS CAMPOS, G.; GIANOLA, D.; ROSA, G. J. M. Reproducing kernel Hilbert spaces regression: A general framework for genetic evaluation. **Journal of Animal Science**, v. 87, n. 6, p. 1883-1887, Jun 2009.

DEBSKA, B.; GUZOWSKA-ŚWIDER, B. Application of artificial neural network in food classification. **Analytica Chimica Acta**, v. 705, n. 33, p. 283-291, 2011.

DEKKERS, J. C. Commercial application of marker-and gene-assisted selection in livestock: strategies and lessons. **Journal of Animal Science**, v. 82, n.13 suppl, p. E313-E328, 2004.

DELVIN, B.; RISCH, N. A comparison of linkage disequilibrium measures for fine-scale mapping. **Genomics**, v.29, n.2, p.311– 22. 1995.

DUARTE, E. R. **Avaliação e desenvolvimento de algoritmos de controle: aplicação para um processo extrativo de fermentação alcoólica contínua**. Campinas, 2007 163 p. Tese (Doutorado)-Universidade Estadual de Campinas, São Paulo, 2007.

ELLIOT L. HEFFNER, M. E. S., AND JEAN-LUC JANNINK. Genomic Selection for Crop Improvement. **Crop Science**, v. 49, n.1, p.1-12, 2009.

FARIA, D.A. **Estudos genômicos em Eucalyptus: Mapeamento de genes candidatos e QTLs, estimativa de desequilíbrio de ligação e análise de estrutura de populações**. Brasília, 2008. 390p. Tese (Doutorado) – Universidade Católica de Brasília, Distrito Federal, 2008.

FLINT-GARCIA, S.A.; THORNSBERRY, J.M.; BUCKLER IV, E.S. Structure of Linkage Disequilibrium in Plants. **Annual of Plant Biology**, 54:357–74, 2003.

Fundo de população das Nações Unidas, UNFPA. Disponível em: <<http://www.unfpa.org.br/novo/index.php/populacao>>. Acesso em: 8 de Janeiro 2013.

HEDRICK, P.W. **Genetics of populations**. 4. ed. Sudbury, MA: Jones and Bartlett Publishers, 2010. 675p.

COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

GALVÃO, C.O.; VALENÇA, M.J.S.; VIEIRA, V.P.P.B.; DINIZ, L. S.; LACERDA, E.G.M.; CARVALHO, A.C.P.L.F.; LUDERMIR, T. B. **Sistemas inteligentes: Aplicações a recursos hídricos e ciências ambientais**. Porto Alegre: UFRGS/ABRH, 1999. 246p.

GIANOLA, D.; VAN KAAM, J. B. C. H. M. Reproducing kernel hilbert spaces regression methods for genomic assisted prediction of quantitative traits. **Genetics**, v. 178, n. 4, p. 2289–2303, Apr. 2008.

GODDARD, M. E.; HAYES, B. J. Genomic selection. **Journal of Animal Breeding and Genetics**, v. 124, n.6, p. 323-330, 2007.

GRATTAPAGLIA, D.; RESENDE, M. D. V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, v. 7, n.3, 241-255, 2011.

GUIMARÃES, C.T.; SCHUSTER, I.; MAGALHÃES, J.V.; JUNIOR, C.L.S. **Marcadores moleculares no melhoramento de plantas**. In: BORÉM e CAIXETA (Orgs) Marcadores moleculares. Viçosa-MG, 2006. p. 101-144.

GUPTA, P.K.; RUSTGI, S.; KULWAL, P.L. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. **Plant Molecular Biology**. V.57, n.4, p.461-485, 2005.

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.

HEFFNER, E.L.; SORRELS, JANNINK, J.L. Genomic Selection for Crop Improvement. **Crop Science**, v. 49, n.1, p. 1-12, 2009.

HENDERSON, C. R. Best linear unbiased estimation and prediction under a selection model. **Biometrics**, v. 31, n. 2 p. 423-447, 1975.

HENDERSON, C.R. **Applications of linear models in animal breeding**. Ontario, University of Guelph, 1984, 462p.

HILL, W.G. **Estimation of linkage disequilibrium in randomly mating populations**. Heredity, vol 2, n.33, p.229-239, 1974.

COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

INGVARSSON, P.K. Nucleotide polymorphism and linkage disequilibrium within and among natural populations of Europe aspen (*Populus tremula*L., Salicaceae) **Genetics**, v. 169, n.2, p.945-953, 2005.

JANNINK, J.J.; LORENZ, A.J.; IWATA, H. Genomic selection in plant breeding: from theory to practice. **Briefings in Functional Genomics**. v.9, n 2, 166-177, 2010.

JENKINS, S.; GIBSON, N. High-throughput SNP genotyping. **Comparative and functional genomics**, v. 3, n. 1, p. 57-66, 2002.

JORDE, L. Linkage disequilibrium and the search for complex disease genes. **Genome research**, v. 10, n. 10, p. 1435-1444, 2000.

KRUGLYAK, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. **Nature genetics**, v. 22, n. 2, 1999.

LANDE, R.; THOMPSON, R. Efficiency of marker-assisted selection in the improvement of quantitative traits. **Genetics**, v. 124, n. 3, p. 743-756, 1990.

LONG, N.; GIANOLA, D.; ROSA, G.; WEIGEL, K.; AVENDANO, S. Machine learning classification procedure for selecting SNPs in genomic selection: application to early mortality in broilers. **Journal of Animal Breeding and Genetics**, v. 124, n. 6, p. 377-389, 2007.

LORENZANA, R. e BERNARDO, R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. **Theoretical and applied genetics**, v.120, n.1, p.151-161, 2009.

LUDWIG, JR.; MONTGOMERY, E. **Redes Neurais: Fundamentos e Aplicações com Programas em C. Rio de Janeiro: Ciência Moderna, 2007. 125p.**

MACKAY, T. F. The genetic architecture of quantitative traits. **Annual review of genetics**, v. 35, n. 1, p. 303-339, 2001.

MARENDA, T.A.; LIMA, R.C.de.; SENNA, R.M.; ALMEIDA, M.M.; DUARTE, E.R. Desenvolvimento de um sistema para classificação do mel em floral e melato por parâmetros físico químicos combinados com redes artificiais. **Evidência Joaçaba**, v. 11, n.1, p. 61-74,2011.

COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

MARTINELLI, E. **Extração de conhecimento de redes neurais artificiais**. São Carlos-SP, 1999. 113p. Dissertação (Mestrado) – Universidade de São Paulo, São Carlos, 1999.

MENDES, F.F.; GUIMARÃES, L.J.M.; SOUZA, J.C.; GUIMARÃES, P.E.O.; PACHECO, C.A.P.; MACHADO, J.R.A.; MEIRELLES, W.F.; SILVA, A.R.; PARENTONI, S.N. Adaptability and stability of maize varieties using mixed model methodology. **Crop Breeding and Applied Biotechnology**, vol 12, p.111-117, 2012.

MESSIAS, J. A. T. **Redes neurais artificiais: Nova abordagem para identificação da qualidade da bebida do Café Arábica utilizando-se resultados de análises químicas**. Viçosa- MG, 2009. 85p. Dissertação (Mestrado) – Universidade Federal de Viçosa, Minas Gerais, 2009.

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

OLIVEIRA, A..C.S. DE; SOUZA, A. A. DE; LACERDA, W.S.; GONÇALVES. L.R. Aplicação de redes neurais artificiais na previsão da produção de álcool. **Ciência e Agrotecnologia**, Lavras, v. 34, n. 2, p. 279-284, 2010.

OLIVEIRA, C.B.; CUNHA, E.E.; CARNEIRO, P.L.S.; EUCLYDES, R.F.; MALHADO, C.H.M. Comparação de métodos de seleção em populações simuladas de frangos de corte. **Pesquisa agropecuária brasileira**, Brasília, v.40, n.10, p.969-974, 2005.

ORTIZ, R.; MOWBRAY, D.; DOWSWELL, C.; RAJARAM, S. Dedication: Norman E. Borlaug The humanitarian plant scientist who changed the world. **Plant Breeding Reviews**, v. 28, p. 1-38, 2007.

PEREIRA, A.M.; KAKITANI, F.; TRESOLDI, G.; BRAMBILA, K.; SCARPIN, L.; LOLIS, L.H.; IMAI, M.J.; TAVARES, M.S.; MATIR, R. Melhoramento Vegetal. Disponível em: <[www.dbi.uem.br/anglo.pdf](http://www.dbi.uem.br/anglo.pdf)>. Acesso em: 27 Dezembro 2012.

PETERNELLI, L.A. Program R: applications in plant breeding. **Crop Breeding and Applied Biotechnology**, v. 11, n.spe, p.91-92, 2011.

PIEPHO, H. P. Ridge Regression and Extensions for Genomewide Selection in Maize. **Crop Science**, v. 49, n. 4, p. 1165-1176, 2009.



COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

PLOTZE, R. O. **Identificação de espécies vegetais através da análise da forma interna de órgãos foliares**. São Paulo, 2004. 152p. Dissertação (Mestrado) - Universidade de São Paulo, São Carlos, 2004.

RESENDE JUNIOR, M.F.R. Seleção genômica ampla no melhoramento vegetal. Viçosa-MG, 2010. 78p. Dissertação (Mestrado) – Universidade Federal de Viçosa, Minas Gerais, 2010.

RESENDE, M. D. V. D.; BARBOSA, M. H. P. Selection via simulated individual BLUP based on family genotypic effects in sugarcane. **Pesquisa Agropecuária Brasileira**, v. 41, n. 3, p. 421-429, 2006.

RESENDE, M. D. V. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo: Embrapa Florestas, 2007. 561 p.

RESENDE, M.; OLIVEIRA, A. C. COMPARAÇÃO DE DIFERENTES ESTRATÉGIAS DE PROGRAMAÇÃO DE IRRIGAÇÃO SUPLEMENTAR EM MILHO. **Revista Brasileira de Milho e Sorgo**, v. 4, n. 02, 2010.

RESENDE, M.; OLIVEIRA, A. Comparação de diferentes estratégias de programação de irrigação em milho. **Revista Brasileira de Milho e Sorgo**, América do Norte, v. 4, 2010.

RESENDE, M.D.V. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa Informação Tecnológica, 2002, 975p.

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

RUGGIERO, C.; DURIGAN, J. F.; GOES, A. DE; NATALE, W.; BENASSI, A. C. Panorama da cultura do mamão no Brasil e no mundo: Situação atual e tendências. In: MARTINS, D. dos S. (Org.) Papaya Brasil: **Qualidade do mamão para o mercado interno**. Vitória: Incaper, p. 13-34, 2003.

SCHMID, K.J.; SORENSEN, T.R.; STRACKE, R et al. Large-scale identification and analysis of genome-wide single nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. **Genome Research**, v.13, n 6A, p.1250-1257, 2003.

SILVA, S. H. M. G. DA; MORAES, W.; MARQUE, E. S. H.; JORGE, L. A. DE C. Processamento digital de imagens para identificação da Sigatoka Negra. In:

COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

**CONGRESSO BRASILEIRO DE AGROINFORMÁTICA**, 5. 2005, Londrina. SBI-AGRO, 2005. v. 1.

SOBREIRA, F.M. **Melhor predição não viesada (BLUP) multicaracterística na seleção recorrente de plantas anuais**. Viçosa-MG, 2009. Dissertação (Mestrado) - Universidade Federal de Viçosa, 2009.

SOLBERG, T. R.; SONESSON, A. K.; WOOLLIAMS, J. A.; MEUWISSEN, T. H. Reducing dimensionality for prediction of genome-wide breeding values. **Genetics Selection Evolution**, v. 41, n. 1, p. 29, 2009.

SUDHEER, K.P.; GOSAIN, A.K.; RAMASASTRI, K.S. Estimating actual evapotranspiration from limited climatic data using neural computing technique. **Journal of Irrigation and Drainage Engineering**, v.129, n. 3, p.214-218, 2003.

VALERIO, I.P.; CARVALHO, F.I.F.; OLIVEIRA, A.C.; SOUZA, V.Q.; BENIN,G.; SCHMIDT, D.A.M.; RIBEIRO, G.; NORBERG,R.; LUCH, H. Combining ability of wheat genotypes in two models of diallel analyses. **Crop Breeding and Applied Biotechnology**, v. 9, p.100-107, 2009.

VIANA, J.; SOBREIRA, F.; DE RESENDE, M.; FARIA, V. Multi-trait BLUP in half-sib selection of annual crops. **Plant Breeding**, v. 129, n. 6, p. 599-604, 2010.

WEBER, A.R.; H. JÚNIOR, J.I.O.; PROENÇA, C.A.; GUIMARÃES, A.M.; ROCHA, J.C.F.; POZO, A.T.R. Aplicação de redes neurais artificiais para avaliação da influência da compactação do solo na produtividade de milho em dados agrícolas georreferenciados. Disponível em: <[http://www.gse.ufsc.br/sbiagro/wp-content/anais/anais/apresentacaoPoster/poster\\_sessao1/89866\\_1.pdf](http://www.gse.ufsc.br/sbiagro/wp-content/anais/anais/apresentacaoPoster/poster_sessao1/89866_1.pdf)>. Acesso em: 10 out. 2012.

WHITTAKER, J. C.; THOMPSON, R.; DENHAM, M. C. Marker-assisted selection using ridge regression. **Genetical research**, v. 75, n. 2, p. 249-252, 2000.

WONG, C.K.; BERNARDO, R. Genome wide selection in oil palm: increasing selection gain per unit time and cost with small populations. **Theoretical and Applied Genetics**, p.116:815–824, 2008.

COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

XU, Y.; LU, Y.; XIE, C.; GAO, S.; WAN, J.; PRASANNA, B. M. Whole-genome strategies for marker-assisted plant breeding. **Molecular Breeding**. v. 29, issue 4, p.833-854, 2012.

ZHU, Y.L.; SONG, Q.J.; HYTEN, D.L. Single- nucleotide polymorphisms in soybean. **Genetics**, v.163, n.3, p.1123-1114, 2003.

**CAPÍTULO II - PREDIÇÃO DOS VALORES FENOTÍPICOS E GENOTÍPICOS VIA  
RR- BLUP/GWS E REDES NEURAS ARTIFICIAIS**

**Predição dos valores fenotípicos e genotípicos via RR-BLUP/GWS e redes neurais artificiais**

Alisson Esdras Coutinho <sup>(1)</sup>, Diogo Gonçalves Neder <sup>(2)</sup>, Mairykon Coelho da Silva <sup>(3)</sup>, Eliane Cristina Arcelino <sup>(3)</sup>, Silvan Gomes de Brito <sup>(4)</sup> e José Luiz Sandes de Carvalho Filho <sup>(5)</sup>

(1) Mestrando vinculado ao Programa de Pós-Graduação em Melhoramento Genético de Plantas do Departamento de Agronomia, Universidade Federal Rural de Pernambuco, Rua Dom Manuel de Medeiros, s/n – Dois Irmãos – Recife-PE, CEP: 52171-900. Email: [alissonesdras@yahoo.com.br](mailto:alissonesdras@yahoo.com.br); (2) Professor Doutor do Departamento de Agroecologia e Agropecuária/CCAA, Universidade Estadual da Paraíba/Campus de Lagoa Seca, Sítio Imbaúba, s/n, Zona Rural, Lagoa Seca-PB, Cep: 58.117-000. Email: [dgneder@ccaa.uepb.edu.br](mailto:dgneder@ccaa.uepb.edu.br); (3) Aluno de Graduação do Departamento de Agronomia, Universidade Federal Rural de Pernambuco, Rua Dom Manuel de Medeiros, s/n – Dois Irmãos – Recife-PE, CEP: 52171- 900. Email: [mairicon@hotmail.com](mailto:mairicon@hotmail.com), [elianearcelino@gmail.com](mailto:elianearcelino@gmail.com); (4) Doutorando vinculado ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de Jaboticabal, Via de Acesso Prof. Paulo Donato Castellane, s/n – Jaboticabal-SP, CEP 14884-900. Email: [gomesilvapb@hotmail.com](mailto:gomesilvapb@hotmail.com); (5) Professor Adjunto do Departamento de Agronomia, Universidade Federal Rural de Pernambuco, Rua Dom Manuel de Medeiros, s/n – Dois Irmãos – Recife-PE, CEP: 52171- 900. Email: [joseluiz.ufrpe@yahoo.com.br](mailto:joseluiz.ufrpe@yahoo.com.br).

## RESUMO

O objetivo deste trabalho foi avaliar a eficiência das redes neurais artificiais (Artificial neural network - ANNs), comparada ao RR-BLUP/GWS (Random Regression – Best Linear Unbiased Predictor/ Genome Wide Selection). O estudo realizado foi baseado nos tratamentos constituídos por diferentes números de marcadores, desequilíbrios de ligação (DL) e porcentagem de herdabilidade em 100 repetições. Portanto, as quantidades de marcadores foram 100 e 200, os níveis de DL (0,5; 0,6; 0,7; 0,8; 0,9 e 1) e herdabilidade de 20% e 40%. As simulações foram efetuadas no programa R, fornecendo as correlações referentes às ANNs e ao RR-BLUP/GWS. Para cada método de predição foram obtidas duas correlações, uma entre o valor fenotípico e o valor genético genômico predito, e outra entre o valor genotípico e o valor genético genômico. Verificou-se para cada método em separado que as maiores médias de correlação observadas indicando melhor capacidade de predição sob níveis mais altos de DL e herdabilidade de 40%. Ao comparar os dois métodos se observou superioridade das ANNs sobre o RR-BLUP/GWS, para a maioria dos tratamentos ao nível de 5% de probabilidade pelo teste t. Houve algumas correlações inferiores quando se utilizou maior quantidade de marcadores, tanto entre os valores genotípicos e os valores genéticos genômicos, como também entre os valores fenotípicos. As menores correlações foram sob DL igual a 0,5 e as maiores com DL igual a 1 para 100 e 200 marcadores, e herdabilidade de 20% e 40%. Portanto, a RNA se revelou como bom preditor a ser utilizado em programas de melhoramento via GWS.

Termos para indexação: Redes neurais artificiais, densidade de marcadores, desequilíbrio de ligação e herdabilidade.

**Prediction of phenotypic and genotypic values by BLUP / GWS and artificial neural networks**

Alisson Esdras Coutinho <sup>(1)</sup>, Diogo Gonçalves Neder <sup>(2)</sup>, Mairykon Coelho da Silva <sup>(3)</sup>, Eliane Cristina Arcelino <sup>(3)</sup>, Silvan Gomes de Brito <sup>(4)</sup> e José Luiz Sandes de Carvalho Filho <sup>(5)</sup>

(1) Mestrando vinculado ao Programa de Pós-Graduação em Melhoramento Genético de Plantas do Departamento de Agronomia, Universidade Federal Rural de Pernambuco, Rua Dom Manuel de Medeiros, s/n – Dois Irmãos – Recife-PE, CEP: 52171-900. Email: alissonesdras@yahoo.com.br; (2) Professor Doutor do Departamento de Agroecologia e Agropecuária/CCAA, Universidade Estadual da Paraíba/Campus de Lagoa Seca, Sítio Imbaúba, s/n, Zona Rural, Lagoa Seca-PB, Cep: 58.117-000. Email: dgneder@ccaa.uepb.edu.br; (3) Aluno de Graduação do Departamento de Agronomia, Universidade Federal Rural de Pernambuco, Rua Dom Manuel de Medeiros, s/n – Dois Irmãos – Recife-PE, CEP: 52171- 900. Email: mairicon@hotmail.com, elianearcelino@gmail.com; (4) Doutorando vinculado ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de Jaboticabal, Via de Acesso Prof. Paulo Donato Castellane, s/n – Jaboticabal-SP, CEP 14884-900. Email: gomesilvapb@hotmail.com; (5) Professor Adjunto do Departamento de Agronomia, Universidade Federal Rural de Pernambuco, Rua Dom Manuel de Medeiros, s/n – Dois Irmãos – Recife-PE, CEP: 52171- 900. Email: joseluiz.ufrpe@yahoo.com.br.

### ABSTRACT

The objective of this study was to evaluate the efficiency of Artificial Neural Network (ANNs) compared to RR-BLUP/GWS (Random Regression - Best Linear Unbiased Predictor / Genome Wide Selection). The study was based on treatments consisting for different numbers of markers, levels of linkage disequilibrium (LD) and percentage of heritability in 100 replications. Then quantities of markers are 100 and 200, levels of DL (0.5, 0.6, 0.7, 0.8, 0.9 and 1) and heritability of 20% and 40%. The simulations were performed in R in the program providing correlations regarding the ANNs and RR-BLUP/GWS. For each prediction method were obtained two correlations, a phenotypic value between the genomic breeding value and another between the predicted value and the breeding value. It was found for each method separately that the highest average correlation observed indicating better predictive ability at higher density of markers, higher levels of DL and heritability of 40%. When comparing the two methods was observed superiority of ANNs on RR-BLUP/GWS for most treatments at 5% probability level by the test t. There were some correlations lower when using higher amounts of markers between genotypic values and genomic breeding value, but also between the phenotypic values and genomic breeding value. The poorer correlations were under LD equal to 0.5 and greater with LD = 1 to 100 and markers 200 and heritability of 20% and 40%. Therefore the RNA proved how good predictor to be used in breeding programs via GWS.

Index Terms: Artificial neural networks, density of markers, linkage disequilibrium and heritability.



## 1. INTRODUÇÃO

A seleção genômica ampla (GWS), como uma nova técnica de seleção proposta, tem tido diversas aplicações de estudos em programas de melhoramento animal e vegetal, utilizando de diversas ferramentas para verificar sua eficiência nos processos de seleção conforme demonstrado primeiramente por Meuwissen (2001) e em trabalhos mais recentes desenvolvidos por Oliveira (2012) e Cavalcanti et al (2012).

O potencial da GWS com estudos simulados mostra, por meio dos resultados obtidos, informações importantes acerca das perspectivas de seu uso em diversos tipos de populações, no ganho por tempo de seleção, na redução de custo, no ciclo de melhoramento, na aplicação em diferentes espécies de culturas anuais (milho e trigo) e para espécies perenes de florestais (MAYOR e BERNARDO, 2009; WONG e BERNARDO, 2008; GRATTAPAGLIA, 2011).

Na predição dos valores genéticos genômicos as informações genotípicas são usadas visando as inferências sobre os valores fenotípicos futuros dos indivíduos, isto é, na GWS é utilizado o efeito do genótipo estimado a partir de uma amostra da população para inferir sobre o fenótipo a ser expresso nos candidatos à seleção (RESENDE, 2008).

A eficiência da GWS está relacionada com a escolha do método de predição, ao qual se deve a capacidade do método de predizer os valores genéticos genômicos a partir da estimação dos efeitos dos marcadores medidos por meio das correlações entre os valores genotípicos reais e os valores preditos. Tais correlações são então avaliadas sob diferentes cenários, abordando diferentes tipos e densidades de marcadores, níveis diferentes de herdabilidade e desequilíbrio de ligação como mostram (SOLDERG et al., 2008; GUO et al 2011).

A aplicação das ANNs é de predizer o VGG dos indivíduos a partir da estimação dos efeitos dos marcadores como RR-BLUP/GWS que é um método de predição utilizado em GWS (RESENDE et al., 2010).

O potencial das ANNs no melhoramento de plantas já é visualizado em (BARBOSA et al., 2011), em estudo de diversidade genética de acessos de mamão, no qual 91,90% dos 37 acessos foram classificados de forma correta em 4 grupos estabelecidos. Outro trabalho foi com de cultivares de arroz italiano, no qual foram

COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

classificadas corretamente mais de 90% das 1779 amostras em 11 cultivares (MARINI et al., 2004). .

Na GWS é bem evidente trabalhos abordando a comparação de metodologias estatísticas (LS, BLUP, Bayes A, Bayes B) sob vários parâmetros, tais como: herdabilidade, número de marcadores, quantidade de QTLs e desequilíbrios de ligação, visando verificar a eficiência destes métodos (MEUWISSEN et al., 2001; GREDLER et al., 2009; CALUS, 2009; PUNGPAPONG, 2012). Como o sucesso da GWS depende do método de predição, torna-se importante o estudo por meio de simulações para se avaliar qual o melhor método na estimação dos efeitos dos marcadores.

Consequentemente, foram aplicadas neste trabalho as ANNs, com o objetivo de avaliar sua capacidade em prever os valores genéticos genômicos de indivíduos dentro de populações. A aplicação desta ferramenta visa ser uma nova alternativa, vislumbrando aumentar a eficiência dos programas de melhoramento de plantas.

## 2. MATERIAL E MÉTODOS

### 2.1. Simulações dos dados fenotípicos e genotípicos

As simulações dos dados fenotípicos e genotípicos foram realizadas considerando ausência de dominância. Dois níveis de herdabilidade (20 e 40%), seis níveis de desequilíbrio de ligação (0,5; 0,6; 0,7; 0,8; 0,9 e 1) e diferentes números de locos, 100 e 200, constituindo assim 24 tratamentos repetidos 100 vezes e um tamanho populacional de 1000 indivíduos. Tanto para simulação fenotípica quanto para a genotípica se atribuiu dois alelos por loco.

No procedimento da simulação genotípica, o efeito do alelo favorável foi simulado de acordo com a seguinte fórmula:

$$a_L = \frac{(L-1)}{(L+1)}$$

na qual L refere-se ao L-ésimo QTL (BERNARDO e YU 2007). Já para o alelo menos favorável ( $-a_L$ ). Os 1000 indivíduos foram simulados segundo o modelo  $f = g+e$ , no qual “g” significa o somatório dos efeitos genéticos de cada loco e “e” são os efeitos ambientais que são obtidos a partir de uma distribuição normal com média 10 e variância compatível com as herdabilidades testadas: 20% e 40%.

$$\sigma_M^2 = \frac{\sigma_g^2(1-h^2)}{h^2}$$

Em que  $\sigma_M^2$  é a variância ambiental,  $\sigma_g^2$  é a variância genética e  $h^2$  é herdabilidade.

Na simulação fenotípica, considerou-se também 1000 indivíduos, herdabilidade 20% e 40 % e desequilíbrio de ligação de 0,5; 0,6; 0,7; 0,8; 0,9 e 1 e também efeitos aditivos.

Em cada indivíduo, genótipos do tipo  $A_1A_1$ ,  $A_1A_2$ , ou  $A_2A_2$  foram sorteados aleatoriamente em cada loco e codificados como 0, 1 e 2, respectivamente. O número 1 dos genótipos do tipo  $A_1A_1$  e  $A_1A_2$  denota o alelo desfavorável e o número 2 denota o alelo favorável.

Quanto ao desequilíbrio de ligação, foram submetidos seis níveis: 0,5; 0,6; 0,7; 0,8; 0,9 e 1. Assim, em cada situação, o valor de DL irá simbolizar a porcentagem de locos marcadores em desequilíbrio de ligação com um QTL e a proporção da variância genética explicada pelos marcadores (RESENDE JUNIOR, 2010).

Logo, o marcador encontrando-se em desequilíbrio de ligação com um determinado QTL significa que há uma associação entre marcador e QTL. Desta forma, o marcador será denominado como um QTN (Quantitative Trait Nucleotide). Quando há situação de equilíbrio de ligação, o valor do genótipo do marcador é considerado como zero. Quando houver desequilíbrio de ligação o genótipo  $A_1A_1$  do QTL estará associado ao genótipo de determinado marcador. Quando for uma situação de equilíbrio, o efeito do genótipo do marcador estará associado de maneira aleatória ao QTL (RESENDE JUNIOR, 2010).

## 2.2. Estimação dos valores genéticos genômicos

### 2.2.1. Redes neurais artificiais (ANNs)

A RNA usada foi de aprendizado supervisionado e o algoritmo de aprendizagem escolhido para ajustar os pesos na fase de treinamento foi de backpropagation. O ajuste é realizado mediante a comparação dos valores de saída e os valores desejados. Caso a diferença não seja aceitável, torna-se necessário fazer o ajuste dos pesos. Esta diferença é conhecida como o erro quadrático médio. O algoritmo efetua o ajuste sempre na direção contrária à direção de maior aumento do erro médio quadrático, reduzindo desta forma o erro por duas maneiras: após a utilização de apenas um conjunto de valores de entrada, ou após todos os conjuntos de valores de entrada ter sido utilizados. Para o ajuste existe uma regra denominada de gradiente descendente, na qual os pesos são alterados da seguinte fórmula (WINANDY et al., 2007):

$$x_{i+1} = x_i - \lambda \nabla f$$

Onde  $X$  representa o vetor de pesos,  $f$  representa o gradiente do erro médio quadrático e  $\lambda$  representa um fator de aprendizagem que deve ser escolhido previamente.

Para escolha do modelo foram utilizadas redes com três camadas e testadas diferentes arquiteturas, variando o número de neurônios na camada intermediária e a função de ativação entre sigmoideal e tangente hiperbólica, ou seja, foram testados quais fatores juntos formavam a rede de melhor resultado para ser usada no desenvolvimento do sistema (MARENDA, et al., 2011).

A configuração da RNA foi composta pelos seguintes componentes: quantidade de indivíduos para o aprendizado da rede (900), quantidade de indivíduos para a validação da RNA (100), número de neurônios na camada de entrada (100), número de interações (20), taxa de aprendizado (0,01), menor peso para os neurônios (0) e maior peso para os neurônios (0,2). Portanto, na predição além dos componentes citados anteriormente, foram empregados o número de repetições (100), número de indivíduos (1000), número de genes ou QTLs (100 e 200) e número de interações (0). Assim, a configuração da RNA para 100 genes ou QTLs foi simulada (100, 1000, 100, 0, 0.4, 0.8, 900, 100, 100, 20, 0.01, 0, 0.2) e para 200 simulada (200, 1000, 200, 0, 0.4, 0.8, 900, 100, 200, 100, 0.04, 0, 0.1). Tais configurações foram estabelecidas mediante testes realizados alterando os valores via simulação, escolhendo aquelas que as ANNs superavam o RR-BLUP/GWS na maioria das correlações.

A camada de entrada das ANNs é composta por neurônios que são representados pelo valor de cada marcador referente a um dado QTL. Estes neurônios da camada de entrada são propagados por cada peso até a camada de saída (MARENDA, 2011). O resultado do somatório do produto dos dados de entrada pelos pesos sinápticos é conhecido como a função de ativação (LUDWIG JR & MONTGOMERY, 2007). Este resultado representa o valor genético genômico predito que é o sinal de saída, o qual é comparado com o valor fenotípico e genotípico.

### **2.2.2. RR-BLUP/GWS**

Outra metodologia utilizada para se estimar os valores genéticos genômicos (VGGs) foi a partir da obtenção dos haplótipos, através do BLUP tradicional. A predição dos efeitos dos marcadores é baseada no seguinte modelo linear misto segundo RESENDE (2008):

COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

$$y = Xb + Zh + e$$

$y$  é o vetor de observações fenotípicas;

$b$  é o vetor de efeitos fixos;

$h$  é o vetor dos efeitos aleatórios dos marcadores;

$e$  refere-se ao vetor de resíduos aleatórios.

$X$  e  $Z$  são as matrizes de incidência para  $b$  e  $h$ .

A matriz de incidência  $Z$  possui os valores 0, 1 e 2 referentes ao número de alelos do QTL. O  $h$  equivale ao número de intervalos multiplicado por 4, ou seja, o número de haplótipos possíveis para cada intervalo. Quanto a estrutura das médias e variâncias, elas são definidas da seguinte forma:  $h \sim N(0, G)$ ;  $E(y+Xb)$ ;  $e \sim N(0, R = I\sigma_e^2)$ ;  $\text{Var}(y) = V = ZGZ' + R$ ;  $G = I\sigma_g^2/n$  em que  $I_h$  é de ordem 4, e  $\sigma_i^2$  é a variância dos efeitos dos haplótipos no  $i$ -ésimo intervalo e  $n$  é o número total de intervalos.

Como os efeitos do vetor  $h$  são ajustados como covariáveis aleatórias associadas com base nos fenótipos, passando a ser denominada como regressão de cummeira (Ridge Regression – RR) conhecida também como RR-BLUP/GWS (MUIR, 2007). As equações do modelo misto para a predição de  $h$  via o método RR-BLUP/GWS equivalem a:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + I \frac{\sigma_e^2}{(\sigma_g^2/n)} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{h} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Em que  $\sigma_g^2$  é a variância genética total do caráter, e  $\sigma_e^2$  é a variância residual. Assim, o valor genético genômico é dado pela seguinte equação, na qual  $Z_i$  equivale a 0, 1 e 2.

Nos casos em que os marcadores não explicam toda a variação genética faz-se necessário adicionar o efeito poligênico residual, fazendo com que o modelo assuma a seguinte forma:

$$y = Xb + Zh + Wg^* + e$$

$g^*$  é o vetor dos efeitos residuais

$W$  é a matriz de incidência para  $g^*$

A capacidade preditiva das redes neurais artificiais e do RR-BLUP foi obtida comparando a correlação média entre os valores genéticos preditos e os simulados (valores fenotípicos e valores genotípicos). As análises de simulação e estimação foram realizadas por meio dos seguintes pacotes (packages), (rrBLUP) para o RR-BLUP/GWS (ENDELMAN, 2011) e RSNNS para as ANNs (BERGMEIR e BENÍTEZ, 2012) então implementados no software R (R Development Core Team, 2009), que é uma linguagem e um ambiente de domínio público, o qual possui numerosas bibliotecas, ou pacotes utilizados para análise estatística e elaboração de gráficos já instalados, além de poderem ser adicionados outros pacotes, possibilitando a criação de novas funções para serem aplicadas em avaliação e interpretação dos dados (PETERNELLI, 2011). As comparações entre as correlações do RR-BLUP e das ANNs foram realizadas ao nível de 5% de significância pelo teste t.

### 3. RESULTADOS E DISCUSSÃO

Os tratamentos analisados consistiram da combinação dos seguintes fatores: número de genes (100 e 200), herdabilidade (20% e 40%) e desequilíbrios de ligação (0,5; 0,6; 0,7; 0,8; 0,9 e 1), os quais constituíram um total de 24 tratamentos. Todos os tratamentos foram testados por meio de simulações em 100 repetições cada, comparando a capacidade preditiva via correlação dos métodos RR-BLUP/GWS e das ANNs em prever os VGGs.

Com base nos resultados das simulações se observou que as médias das correlações aumentaram na medida em que se tinha um maior nível de desequilíbrio de ligação e herdabilidade. Mas, quanto a utilização de maior quantidade de marcadores, observou-se em alguns cenários a redução das correlações, e em outros o aumento. O que foi verificado tanto para as correlações entre os valores fenotípicos e os preditos, como entre os valores genotípicos e os preditos para o RR-BLUP/GWS e as ANNs (Tabela 1, 2, 3 e 4).

Considerando a quantidade de marcadores, foram verificados resultados com maiores e menores correlações, utilizando 200 comparado com 100. Os resultados representam as médias das correlações entre os valores fenotípicos e os preditos (Tabelas 1 e 3) e entre os valores genotípicos com os preditos (Tabelas 2 e 4) para o RR-BLUP/GWS e as ANNs.

Por meio das simulações, verificou-se maior quantidade de correlações superiores para 100 marcadores, havendo médias inferiores somente para os cenários com DL de 0,9 e 1 para 20% de herdabilidade e com DL de 0,7; 0,9 e 1 para 40% de herdabilidade (Tabela 1). As médias de correlação superiores foram para os seguintes cenários: 100 marcadores e DL de 0,5; 0,6; 0,7 e 0,8 e 20% de herdabilidade e DL de 0,5; 0,6 e 0,8 para 40% de herdabilidade (Tabela 1) comparado com os mesmos DL usando 200 marcadores (Tabela 3) entre os valores fenotípicos e os preditos para as ANNs. Com os resultados obtidos, constatou-se que o uso de maior densidade de marcadores proporcionou menor capacidade preditiva. Já para o RR-BLUP/GWS, houve superioridade para 100 marcadores em todos os cenários. Neste caso o RR-BLUP também apresentou maior capacidade de predição, aplicando menor densidade de marcadores. Resultados semelhantes foram encontrados por Guo et al., (2011) quando observaram na diminuição de



COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

densidade de marcadores de 1,6 cM para 20 cM redução das acurácias de predição. Para o RR-BLUP, utilizando-se densidade de 40 cM, obteve-se acurácia de 0,29, enquanto que para 30 cM foi de 0,26. Assim, situações nas quais aplica-se menor número de marcadores podem proporcionar maior acurácia por capturar mais QTLs de maior efeito, do que usando altas densidades (GUO et al., 2011).

Avaliando-se valores de acurácia sob diferentes números de marcadores foi verificado por Lorenzana e Bernardo (2009) que em maiores densidades para o BLUP obteve-se aumento da acurácia em populações biparentais de milho (*Zea mays* L.), *Arabidopsis* e cevada (*Hordeum vulgare* L.), para as características de umidade do grão, teor de glicose em milho, rendimento de grão e teor de proteína no grão (alpha amilase). Em muitos casos o BLUP obteve melhores resultados que Bayes, além do método bayesiano ter requerido um tempo para finalizar as análises no computador 500 vezes maior que o BLUP. As maiores acurácias foram alcançadas quando se utilizou mais de 100 marcadores em *Arabidopsis*, duplo haploide de milho e cevada, além de 200 a 800 em população de milho oriunda de cruzamento aleatório.

Os valores de baixa acurácia também podem estar relacionados quando são usados grandes números de indivíduos sob alta quantidade de marcadores no BLUP.

Em Zhao et al. (2011) utilizando RR-BUP foi observado também valores mais altos de acurácia, variando de 100 a 800 SNPs, chegando a variar de 0,87 a 0,97 para umidade do grão e de 0,45 a 0,69 para rendimento de grão. A densidade de marcadores necessária para se ter predições precisas depende do nível do desequilíbrio de ligação entre os marcadores e o QTL (MEUWISSEN et al., 2001). É provável que esteja relacionado com o baixo poder de detecção do QTL, superestimação dos efeitos do QTL e ao pequeno grau de liberdade para estimar o efeito dos marcadores (MEUWISSEN et al., 2001). Outro fator importante observado é que à medida que se utilizava uma maior porcentagem (12,5 % a 100%) da população para estimação dos efeitos dos marcadores, se obtinha maiores acurácias (ZHAO et al., 2011).

Quanto ao DL as médias das correlações referentes às ANNs variaram de 0,09 a 0,39 e de 0,13 a 0,58 para o nível de DL de 0,5 a 1 em 20% e 40% de herdabilidade respectivamente (Tabela 1). Já para o RR-BLUP, obteve-se uma variação de 0,03 (DL= 0,5) a 0,35 (DL=1) e de 0,05 a 0,57 para o menor, e maior DL.

O que indica uma situação em que determinado caráter esteja havendo DL de 0,5, o método RR-BLUP/GWS não consegue ter boa predição do valor genético genômico, ou seja, seu uso não seria viável na GWS nesta condição. Já o valor DL igual a 1, mostra que o método consegue predizer melhor o valor genético genômico de um determinado caráter. Para os DL 0,9 e 1, as correlações foram inferiores aos encontrados por Resende (2010) para a mesma herdabilidade (20%), 100, 200 e 400 marcadores afetadas pela  $h^2$  de 0,2; 0,3 e 0,4. Foi observado também que as correlações aumentam na medida em que se tem um nível maior de DL, onde foram observados valores de 0,02 e 0,08 para DL igual a 0,5. E 0,41 e 0,59 referentes ao DL igual a 1 para as ANNs, mostrando que este método é capaz de fornecer uma boa predição sob estas condições. Enquanto que para o RR-BLUP /GWS foram obtidas as menores médias (0,03 e 0,04) em relação ao DL de 0,5 e as maiores (0,32 e 0,54) para o DL de 1 (Tabela 3).

Na tabela 2, constam os DLs relacionados com as correlações entre o VGG predito e o valor genotípico. As médias das correlações variaram de 0,18 (DL=0,5) a 0,89 (DL= 1) e de 0,20 (DL= 0,5) a 0,91 (DL=1) para 20% e 40% de herdabilidade para as ANNs. Da mesma forma, na tabela 4 os valores das médias das correlações também se apresentaram superiores ao empregar na simulação maiores graus de DL, demonstrando que os métodos de predição aplicados para estas condições podem ter boa capacidade preditiva.

Segundo Habier et al. (2007), quando os marcadores se encontram em desequilíbrio de ligação com o QTL é esperado que a acurácia dos VGGs preditos tenha um valor alto do que a acurácia devido ao desequilíbrio de ligação sozinho. Em um estudo de GWS em populações de cevada se verificou que a maioria das avaliações que se obteve melhor acurácia foi à presença de alto DL (ZHONG et al., 2009).

Segundo Zhong et al. (2009), independente do método utilizado, existe uma interação entre o nível de DL e densidade de marcadores. Menores DL consistem em menor colinearidade entre os marcadores. A colinearidade interfere na capacidade dos métodos de regressão, reduzindo a identificação dos QTLs (JANSEN, 2007). Ela é uma força das abordagens da regressão (ridge regression) usadas pelos métodos BLUP. Quando os efeitos do QTL são muito pequenos, podem requerer alguma colinearidade entre eles para capturar seus efeitos de todos.

As correlações maiores, de acordo com níveis mais altos de DL (Tabela 1 e 3), corroboram com os resultados encontrados por Soldberg et al. (2008). Assim como em Soldberg et al. (2008), o DL é o fator principal que proporciona o aumento das correlações entre os valores preditos e o valor fenotípico, ou seja, na presença de níveis mais altos de DL se tem correlações superiores. Nas plantas é comum que os caracteres governados por muitos genes sejam de baixa herdabilidade. Portanto, a herdabilidade foi outro fator que proporcionou, de acordo com sua maior porcentagem, também correlações maiores.

Assim como níveis mais altos de DL e porcentagens maiores de herdabilidade também proporcionaram valores superiores de correlações. Após as simulações verificou-se maiores médias de correlações entre os VGG preditos e o fenótipo sob 40% de herdabilidade para 100 e 200 alelos marcadores (Tabela 1 e 3). A mesma superioridade foi observada nas correlações entre os VGG preditos e valor genotípico com o mesmo nível de herdabilidade, para 100 e 200 marcadores (Tabela 2 e 4). Isto implica que o RR-BLUP/GWS e as ANNs mostram ter uma melhor capacidade preditiva ao estimar os efeitos dos marcadores e assim predizer o VGG para caracteres que possuam herdabilidade de 40%. Ambas as correlações para o DL igual a 0,9 (0,72) e DL= 1 (0,91) foram superiores comparado a 0,63 (DL=0,9) e 0,9 (DL=1) para os resultados encontrados por Resende (2010) com 40% de herdabilidade para o RR-BLUP/GWS. Os resultados de Guo et al. 2011 com RR-BLUP se diferenciam quando encontraram acurácias de predição 12% maiores para 60% e 70% de herdabilidade do que com 80% e 90%.

Na comparação das simulações envolvendo 100 genes com herdabilidade de 20% e DL de 0,5; 0,6; 0,7; 0,8; 0,9 e 1 entre as ANNs e do RR-BLUP/GWS baseados nas médias das correlações, houve diferença significativa ao nível de 5% pelo teste t (Tabela 1). Assim, pode-se afirmar que as ANNs para esta situação possuem uma capacidade preditiva superior de estimar o fenótipo em relação ao RR-BLUP/GWS. Já para os valores genotípicos e os preditos foi observada diferença significativa entre as ANNs e o RR-BLUP/GWS para todos os níveis de DL em que as ANNs apresentam melhor capacidade preditiva (Tabela 2).

Ainda nas simulações com 100 genes sendo com herdabilidade de 40% só não houve diferença significativa para as correlações com DL de 0,9 e 1 (Tabela 1). Para os DL de 0,5; 0,6; 0,7 e 0,8 houve diferença significativa entre as correlações das ANNs e do RR-BLUP/GWS. Com isso, há uma grande probabilidade de que as

ANNs para estas situações superem o RR-BLUP/GWS em relação a capacidade preditiva dos efeitos fenotípicos. Nas correlações entre os valores genotípicos e os VGGs preditos não ocorreu diferença significativa para o DL de 1 em 40% de herdabilidade entre as ANNs e o RR-BLUP/GWS, havendo uma chance maior do que 5% de que esta diferença seja devido ao acaso (Tabela 2). Portanto, para este nível de DL não se pode afirmar que as ANNs possuem uma melhor capacidade preditiva que o RR-BLUP/GWS. Nos outros níveis de DL (0,5; 0,6; 0,7; 0,8 e 0,9) ocorreu diferença significativa, na qual as ANNs apresentam melhor capacidade preditiva, comparado ao RR-BLUP/GWS.

As simulações envolvendo 200 genes com herdabilidade de 20%, realizadas por meio das ANNs, apresentaram médias superiores para o valor das correlações entre os valores fenotípicos e os preditos, comparado com o RR-BLUP/GWS (Tabela 3). Esta superioridade foi constatada para todos os níveis de DL submetidos na análise das simulações. A única média inferior das correlações foi para DL de 0,5. Com base no teste t ao nível de 5% de significância para este DL, as médias das correlações dos valores preditos pelas ANNs e pelo RR-BLUP/GWS não apresentaram diferença significativa. Portanto, para esta situação não se pode afirmar que as ANNs são inferiores para a capacidade preditiva dos efeitos fenotípicos. Já para os demais níveis de DL (0,6; 0,7; 0,8; 0,9 e 1) foi verificada superioridade das médias das correlações para as ANNs, superando o RR-BLUP/GWS, apresentando uma diferença significativa entre as correlações dos valores preditos e fenotípicos ao nível de 5% pelo teste t. Assim, para os DL (0,6; 0,7; 0,8; 0,9 e 1) as ANNs superam o RR-BLUP/GWS na capacidade de prever os efeitos dos fenótipos na medida em que se aumenta o DL (Tabela 3).

Já nas simulações entre os valores genotípicos e os VGGs preditos com herdabilidade de 20% e 200 genes para o nível de DL de 0,5 não houve diferença significativa entre as médias das correlações das ANNs e do RR-BLUP/GWS, (Tabela 4). Para os demais níveis de DL houve diferença significativa, na qual as ANNs apresentaram maiores valores para as correlações comparadas ao RR-BLUP/GWS, indicando que se tem menos de 5% de chance de que esta diferença seja devido ao acaso. Portanto, para os DL (0,6; 0,7; 0,8 e 0,9) as ANNs superam o RR-BLUP/GWS na capacidade de prever os efeitos dos genótipos (Tabela 4).

Nas simulações com herdabilidade de 40% para as correlações entre os VGGs preditos e valor genotípico, houve diferença significativa para todos os DL.

COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

Pelo teste t indica que se tem menos do que 5% de chance desta diferença ser devido ao acaso (Tabela 4). Pois para estes desequilíbrios de ligação pode-se afirmar que as ANNs superam o RR-BLUP/GWS na estimação dos efeitos dos marcadores. Portanto, as ANNs para este tipo de situação se comportam como bons preditores, superando o RR-BLUP/GWS na predição dos efeitos dos genótipos.

A importância das ANNs no melhoramento genético de plantas se confirma em outros trabalhos, nos quais demonstra-se sua capacidade conforme evidenciado por Marini et al., (2004), onde na caracterização de cultivares italianas de arroz, as ANNs de Kohonen, que são utilizadas para agrupamento de dados, foram capazes prever mais de 90% dos conjuntos de amostras.

As ANNs como demonstra Plotze (2004), mostra-se uma ferramenta de importância na caracterização de espécies que também podem ser estendidas para a fase pré-melhoramento no agrupamento e caracterização de acessos. Com isso, neste mesmo trabalho as ANNs foram capazes de classificar corretamente com base nos dados da complexidade das nervuras de folhas de espécies de maracujá selvagem do gênero *Passiflora* 90% e 66,7% delas. Já para os dados de complexidade dos esqueletos foliares, 80% e 70% das espécies foram classificadas corretamente quando se escolheram 75% e 50% do conjunto de dados para treinamento e 25% e 50% para testar a rede, respectivamente.

No treinamento das ANNs na predição os VGGs usou-se 90% da população para treinamento e 10% para teste. As variações de taxas de classificação encontradas por Plotze (2004) se referem às diferentes porcentagens escolhidas do conjunto de dados no treinamento e teste da rede. Isto justifica a escolha de 90% da população aplicada ao o treinamento quando se alcançou um maior número de correlações superiores para as ANNs comparado ao RR-BLUP/GWS.

Outra aplicação que demonstra o potencial das ANNs no melhoramento genético de plantas é o seu uso na classificação de indivíduos, comparado as funções discriminantes de Fischer e Anderson, nas quais foram observados um número de classificações menor que os outros dois métodos (PEREIRA, 2009). Este é mais um exemplo que fortalece as ANNs como potencial ferramenta para auxiliar as etapas de um programa de melhoramento.

O potencial da aplicação das ANNs como ferramenta na análise de divergência genética, etapa importante na seleção de indivíduos contrastantes a serem utilizados em cruzamentos em programas de melhoramento é representado

nos resultados encontrados por Barbosa et al., 2011 em que as ANNs geraram 4 grupos de acessos de mamão (*Carica papaya* L.) e 90% deles foram classificados de forma correta. Isto se demonstra como mais um exemplo para referência do potencial das ANNs a serem usadas com método de predição em estudos de GWS.

Na previsão de rendimentos de milho e soja sob condições climáticas revelou-se que as ANNs comparado a MLR (Regressão linear múltipla) apresentaram predições mais acuradas com um coeficiente de determinação ( $r^2$ ) de 0,77 e a MLR com um  $r^2$  de 0,42 para rendimento de milho. Já para a predição do rendimento de soja as ANNs obtiverem um  $r^2$  de 0,81 e a MLR resultou num  $r^2$  de 0,46 (KAUL et al., 2005 ).Em análise de discriminação de variedades de tea (*Camellia sinensis* (L) O. Kuntze) fica novamente demonstrado a potencialidade das ANNs, chegando a uma taxa de 100% na identificação das variedades (HE et al., 2007) .Isto também se confirma nos resultados encontrados por (CHO et al., 2002) num trabalho para distinguir rabanetes de ervas daninhas alcançando 100% de acurácia.

Em virtude do que foi apresentado e discutido acerca dos resultados das ANNs obtidos por meio das simulações, este método se apresenta como uma provável ferramenta com potencial nos estudos com seleção genômica ampla. Diante disto, ficam abertas ainda mais possibilidades de pesquisa, buscando melhor entendimento, bem como aplicações mais adequadas das ANNs em programas de melhoramento genético de plantas. Além disso, estudos via simulação possibilitam nestas pesquisas a busca por otimizações antes que sejam aplicadas na prática com dados reais, evitando perdas como custo e também operações mais trabalhosas no seu uso, bem como os cenários em que as ANNs tornam-se limitadas. Por fim, é importante não deixar de destacar que as simulações no presente trabalho não demandaram alto custo e proporcionaram informações úteis acerca das ANNs na predição dos VGGs nos cenários analisados, fortalecendo, portanto, sua provável potencialidade na aplicação no processo de seleção.

#### **4. CONCLUSÕES**

Maiores capacidades de predição são alcançadas sob menores densidades de marcadores juntamente com níveis altos de desequilíbrio de ligação e maior herdabilidade, sob estas condições, as redes neurais artificiais superam o RR-BLUP/GWS na predição dos valores genéticos genômicos. Nas simulações envolvendo maior número de genes foi necessário um tempo maior para concluí-las, ficando demonstrado que para a estimação de maior quantidade de marcadores é necessário computadores com maior recurso que deem maior velocidade ao processo de simulação.

**REFERÊNCIAS**

BARBOSA, C. D. Rede neural artificial aplicada à agricultura. **Vértices**, v. 13, p.10, 2011.

BERGMEIR, C.; BENITEZ, J. M. Neural Networks in R Using the Stuttgart Neural Network Simulator: RSNNS. **Journal of Statistical Software**, v. 46, n. 7, p. 1-26, 2012.

BERNARDO, R. e YU, J. M. Prospects for genomewide selection for quantitative traits in maize. **Crop Science**, v.47, n.3, p.1082-1090, 2007.

BRITO, F. V.; NETO, J. B.; SARGOLZAEI, M.; COBUCCI, J. A.; SCHENKEL, F. S. Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. **Biomed Central Genetics**, v. 12, 2011.

CHO, S. I.; LEE, D. S.; JEONG, J. Y. Weed-plant discrimination by machine vision and artificial neural network. **Biosystems Engineering**, v. 83, n. 3, p. 275-280, 2002.

CIBELLE DEGEL BARBOSA, A. P. V., SILVANA SILVA RED QUINTAL AND MESSIAS GONZAGA PEREIRA. Artificial neural network analysis of genetic diversity in *Carica papaya* L. **Crop Breeding and Applied Biotechnology**, v. 11, n. 3, p. 224-231, 2011.

DE OLIVEIRA, E. J.; DE RESENDE, M. D. V.; SANTOS, V. D.; FERREIRA, C. F.; OLIVEIRA, G. A. F.; DA SILVA, M. S.; DE OLIVEIRA, L. A.; AGUILAR-VILDOSO, C. I. Genome-wide selection in cassava. **Euphytica**, v. 187, n. 2, p. 263-276, 2012.

ENDELMAN, J. B. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome*, v. 4, n. 3, p. 250-255, Nov 2011.

MARINI, F.; ZUPAN, J.; MAGRÌ, A. On the use of counterpropagation artificial neural networks to characterize Italian rice varieties. **Analytica Chimica Acta**, v. 510, Issue 2, p. 231-240, 2004.

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



COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

GUO, Z.; TUCKER, D. M.; LU, J.; KISHORE, V.; GAY, G. Evaluation of genome-wide selection efficiency in maize nested association mapping populations. **Theoretical and Applied Genetics**, v. 124, n. 2, p. 261-275, 2011.

HABIER, D.; FERNANDO, R. L.; DEKKERS, J. C. M. The impact of genetic relationship information on genome-assisted breeding values. **Genetics**, v. 177, n. 4, p. 2389-2397, 2007.

HE, Y.; LI, X.; DENG, X. Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model. **Journal of Food Engineering**, v. 79, n. 4, p. 1238-1242, 2007.

CAVALCANTI, J. J. V.; RESENDE, M. D. V. D.; SANTOS, F. H. C. D.; PINHEIRO, C. R. Simultaneous prediction of the effects of molecular markers and genome wide selection in cashew. **Revista Brasileira de Fruticultura**, v. 34, n. 3, p. 840-846, 2012.

GREDLER, B.; NIREA, K.; SOLBERG, T.; EGGER-DANNER, C.; MEUWISSEN, T.; SÖLKNER, J. A comparison of methods for genomic selection in Austrian dual purpose Simmental cattle. *Proc Assoc Advmt Anim Breed Genet*, v. 18, p.568-571, 2009.

KAUL, M.; HILL, R. L.; WALTHALL, C. Artificial neural networks for corn and soybean yield prediction. **Agricultural Systems**, v. 85, n. 1, p. 1-18, 2005.

LANDE, R.; THOMPSON, R. Efficiency of marker-assisted selection in the improvement of quantitative traits. **Genetics**, v. 124, n. 3, p. 743-756, 1990.

LORENZANA, R. e BERNARDO, R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. **Theoretical and Applied Genetics**, v.120, n.1, p.151-161, 2009.

LUDWIG, JR.; MONTGOMERY, E. **Redes Neurais: Fundamentos e Aplicações com Programas em C. Rio de Janeiro**: Ciência Moderna, 2007. 125p.

MARENDA, T. A. L., REGINA CRISTINA A. DE; SENNA, RENATA MATTOS; ALMEIDA, MARECI MENDES; DUARTE, ELIS REGINA. Desenvolvimento de um sistema para classificação do mel em floral e melato por parâmetros físico químicos combinados com redes neurais artificiais. **Evidência, Joaçaba**, v.11, n.1, p.61-74, 2011.

COUTINHO, A. E. Seleção genômica ampla via redes neurais artificiais

MARINI, F.; ZUPAN, J.; MAGRÌ, A. L. On the use of counterpropagation artificial neural networks to characterize Italian rice varieties. **Analytica Chimica Acta**, p. 10, 2004.

MAYOR, P. J. e BERNARDO, R. Genomewide Selection and Marker-Assisted Recurrent Selection in Doubled Haploid versus F-2 Populations. **Crop Science**, v.49, n.5, p.1719-1725, 2009.

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

MUIR, W. M. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. **J. Animal Breeding Genetics**, v. 124, n. 6, p. 342-355, 2007.

PETERNELLI, L.A. Program R: applications in plant breeding. **Crop Breeding and Applied Biotechnology**, S1: p. 91-92, 2011.

PLOTZE, R. O. **Identificação de espécies vegetais através da análise da forma interna de órgãos foliares**. São Paulo, 2004. 152p. Dissertação (Mestrado) - Universidade de São Paulo, São Carlos, 2004.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Disponível em < <http://www.R-project.org>>. Acesso em: 22 Jun. 2009.

RESENDE, M.D.V. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa Informação Tecnológica, 2002, 975p.

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

RESENDE JUNIOR, M.F.R. Seleção genômica ampla no melhoramento vegetal. Viçosa-MG, 2010. 78p. Dissertação (Mestrado) – Universidade Federal de Viçosa, Minas Gerais, 2010.

SOLBERG, T. R.; SONESSON, A. K.; WOOLLIAMS, J. A.; MEUWISSEN, T. H. E. Genomic selection using different marker types and densities. **Journal of Animal Science**, v. 86, n. 10, p. 2447-2454, 2008.

WINANDY, C.E.; BORGES FILHO.; BENTO, L.V. **Algoritmos para Aprendizagem Supervisionada**. CT215 – Inteligência Artificial. Seminários, 2007. Disponível em: <[winandy.voila.net/IA\\_ApSuperv\\_NotasAula.pdf](http://winandy.voila.net/IA_ApSuperv_NotasAula.pdf)>. Acesso em: 19 Jan. 2013.

WONG, C. K. e BERNARDO, R. Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. **Theoretical and Applied Genetics**, v.116, n.6, Apr, p.815-824, 2008.

ZHAO, Y.; GOWDA, M.; LIU, W.; WÜRSCHUM, T.; MAURER, H. P.; LONGIN, F. H.; RANC, N.; REIF, J. C. Accuracy of genomic selection in European maize elite breeding populations. **Theoretical and Applied Genetics**, v. 124, n. 4, p. 769-776, 2011.

ZHONG, S.; DEKKERS, J. C.; FERNANDO, R. L.; JANNINK, J.-L. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. **Genetics**, v. 182, n. 1, p. 355-364, 2009.

**Tabela 1.** Valores das médias das correlações entre os valores fenotípicos e preditos pelas Redes Neurais Artificiais e via BLUP/GWS para 100 genes, 20% e 40% de herdabilidade e desequilíbrios de ligação de 0,5; 0,6; 0,7; 0,8; 0,8; 0,9 e 1.

Número de locos	Herdabilidade ( $h^2$ )	Desequilíbrio de ligação (LD)	Cor (valores fenotípicos e preditos)	
			ANNs	BLUP/GWS
100	20	0,5*	0,0911	0,0365
		0,6*	0,1384	0,0739
		0,7*	0,1928	0,1169
		0,8*	0,2608	0,1907
		0,9*	0,3308	0,2782
		1*	0,3904	0,3552
100	40	0,5*	0,1392	0,0526
		0,6*	0,1991	0,1253
		0,7*	0,2819	0,2244
		0,8*	0,4013	0,3429
		0,9	0,4826	0,4598
		1	0,5839	0,5763

\*Significativo ao nível de 5% pelo teste t

**Tabela 2.** Valores das médias das correlações entre os valores genotípicos e preditos pelas Redes Neurais Artificiais e via BLUP/GWS para 100 genes, 20% e 40% de herdabilidade e desequilíbrios de ligação de 0,5; 0,6; 0,7; 0,8; 0,9 e 1.

Número de locos	Herdabilidade (h <sup>2</sup> )	Desequilíbrio de ligação (DL)	Cor (valores genotípicos e preditos)	
			ANNs	BLUP/GWS
100	20	0,5*	0,1861	0,0755
		0,6*	0,3188	0,1800
		0,7*	0,4659	0,3004
		0,8*	0,6076	0,4500
		0,9*	0,7529	0,6273
		1*	0,8911	0,8236
100	40	0,5*	0,2098	0,0965
		0,6*	0,3253	0,2143
		0,7*	0,4395	0,3475
		0,8*	0,6222	0,5332
		0,9*	0,7624	0,7226
		1	0,9185	0,9175

\*Significativo ao nível de 5% pelo teste t

**Tabela 3.** Valores das médias das correlações entre os valores fenotípicos e preditos pelas Redes Neurais Artificiais e via BLUP/GWS para 200 genes, 20% e 40% de herdabilidade e desequilíbrios de ligação de 0,5; 0,6; 0,7; 0,8; 0,8; 0,9 e 1.

Número de locos	Herdabilidade ( $h^2$ )	Desequilíbrio de ligação (LD)	Cor (valores fenotípicos e preditos)	
			ANNs	BLUP/GWS
200	20	0,5	0,0202	0,0331
		0,6*	0,1096	0,0478
		0,7*	0,1870	0,0990
		0,8*	0,2472	0,1505
		0,9*	0,3442	0,2354
		1*	0,4120	0,3211
200	40	0,5*	0,0878	0,0433
		0,6*	0,1818	0,0961
		0,7*	0,2840	0,1987
		0,8*	0,3842	0,2849
		0,9*	0,4931	0,3981
		1*	0,5959	0,5413

\*Significativo ao nível de 5% pelo teste t

**Tabela 4.** Valores das médias das correlações entre os valores genotípicos e preditos pelas Redes Neurais Artificiais e via BLUP/GWS para 200 genes, 20% e 40% de herdabilidade e desequilíbrios de ligação de 0,5; 0,6; 0,7; 0,8; 0,8; 0,9 e 1.

Número de locos	Herdabilidade ( $h^2$ )	Desequilíbrio de ligação (DL)	Cor (valores fenotípicos e preditos)	
			ANNs	BLUP/GWS
200	20	0,5	0,0727	0,0464
		0,6*	0,2226	0,1057
		0,7*	0,4172	0,2121
		0,8*	0,5761	0,3606
		0,9*	0,7488	0,5100
		1*	0,9355	0,7035
200	40	0,5*	0,1315	0,0554
		0,6*	0,2912	0,1674
		0,7*	0,4551	0,3026
		0,8*	0,6244	0,4600
		0,9*	0,7850	0,6342
		1*	0,9470	0,8505

\*Significativo ao nível de 5% pelo teste t