



INTRODUÇÃO

O estudo de diversidade genética de pimentas é fundamental para atender às demandas do mercado. O uso de descritores morfo-agronômicos é uma das formas utilizadas. Técnicas de aprendizado de máquina podem auxiliar na seleção de caracteres que permitem a escolha de genótipos divergentes e contribuem para ganhos em programas de melhoramento genético de *Capsicum* spp.

O objetivo desse trabalho foi desenvolver um algoritmo de classificação de descritores de *Capsicum frutescens* e *C. chinense* por meio de técnicas de aprendizado de máquina..

METODOLOGIA

A metodologia de Data Mining é o principal suporte para a elaboração do modelo de previsão de diferenciação dos caracteres de *Capsicum* spp. Avaliaram-se 39 genótipos, através de 13 descritores morfológicos multicategóricos.

Para planta foram: Altura = AP (cm), diâmetro da copa = DCO (cm), diâmetro do caule = DCA (cm), tamanho da folha = TF (cm), longitude do caule = LC (cm), diâmetro da corola = DCOR (cm).

Já para o fruto foram: peso = PF (g), comprimento = CF (cm) e diâmetro = DF (cm), espessura do pericarpo = EP (cm), matéria fresca (apenas polpa sem presença de sementes) = MF (g) e seca = MS (g), número de sementes = SF.

A análise incluiu a estimativa da distância de Mahalanobis, da Distância Média Euclidiana, da similaridade e dissimilaridade, bem como o agrupamento de Tocher e a rede de correções.

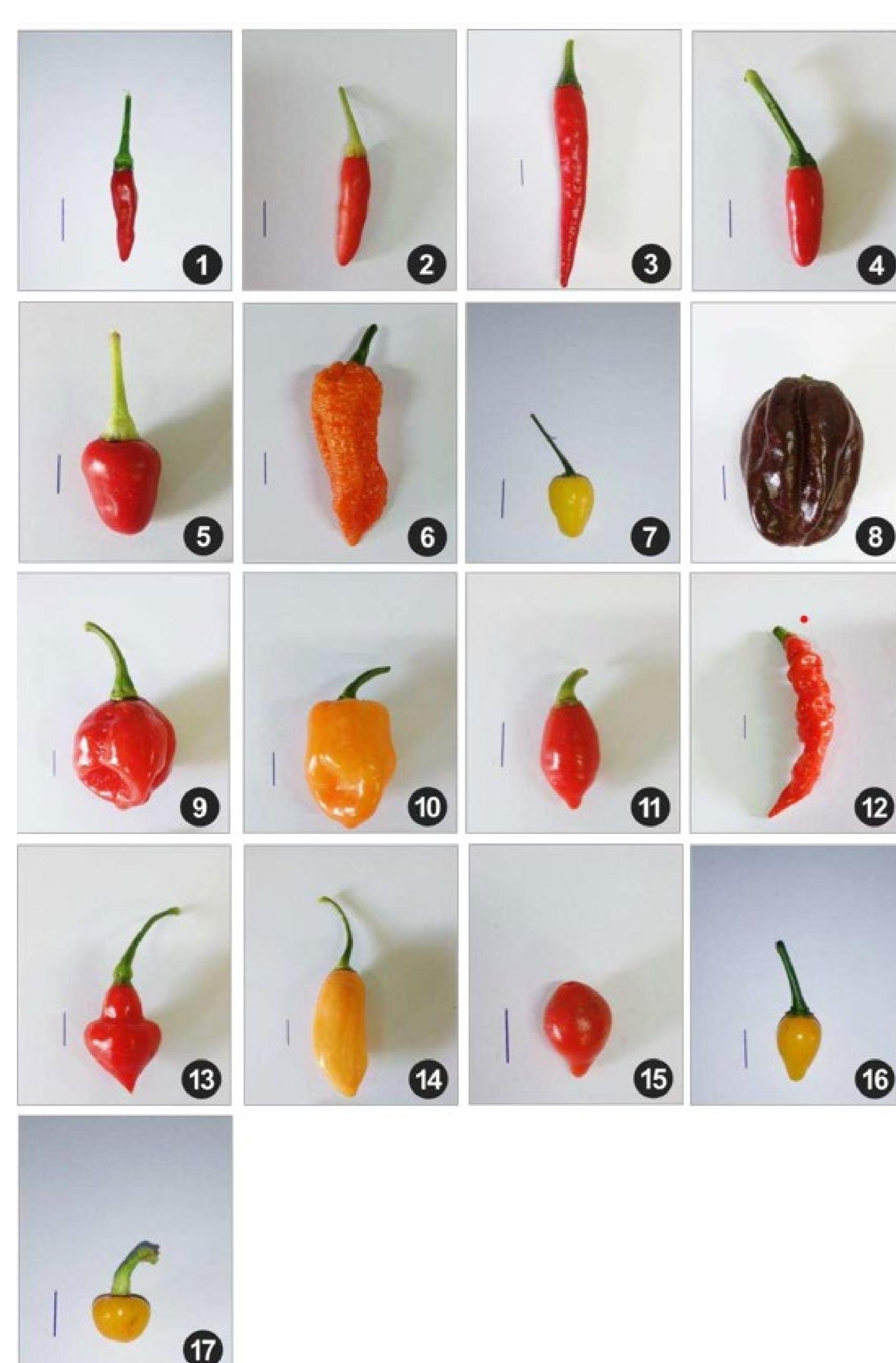


Figura 1 - Imagens dos 17 frutos das duas espécies estudadas (*C. chinense* e *C. frutescens*), em que: 1 = Malagueta, 2 = Tabasco, 3 = Cayene, 4 = Etna Ornamental, 5 = Pirâmide Ornamental, 6 = Bhut Jolokia, 7 = Cumari, 8 = Habanero chocolate, 9 = Habanero vermelha, 10 = Habanero amarela, 11 = Peito de Moça, 12 = Murupi, 13 = Piãozinho, 14 = Cheiro do Norte, 15 = Biquinho vermelha, 16 = Biquinho amarela e 17 = Arari Bode amarela.

RESULTADOS E CONCLUSÕES

No caso das variedades de *C. frutescens*, os descritores PF, MF, CF e EP contribuíram com 89,95% da divergência genética, enquanto para *C. chinense*, essa contribuição foi de 68,72%. As correlações mais relevantes para os caracteres de plantas foram AP e LC. Já para os caracteres de fruto, destacaram-se PF, MF, MS, DF e CF. Por meio do método de otimização de Tocher, os acessos foram subdivididos em grupos.

Grupos	Acessos								
I	15	16	11	13	17	7	10	14	
II	8	9	6						
III	3	12							
IV	1	2							
V	4								
VI	5								

Tabela 1 -Análise de agrupamento de *C. Chinense* e *C. Frutescens* 1 = Malagueta, 2 = Tabasco, 3 = Cayene, 4 = Etna Ornamental, 5 = Pirâmide Ornamental, 6 = Bhut Jolokia, 7 = Cumari, 8 = Habanero chocolate, 9 = Habanero vermelha, 10 = Habanero amarela, 11 = Peito de Moça, 12 = Murupi, 13 = Piãozinho, 14 = Cheiro do Norte, 15 = Biquinho vermelha, 16 = Biquinho amarela e 17 = Arari Bode amarela. Acessos indicados pelos números 1, 2, 3, 4 e 5 são classificados como *C. Frutescens* e os indicados por 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 e 17, como *C. chinense*

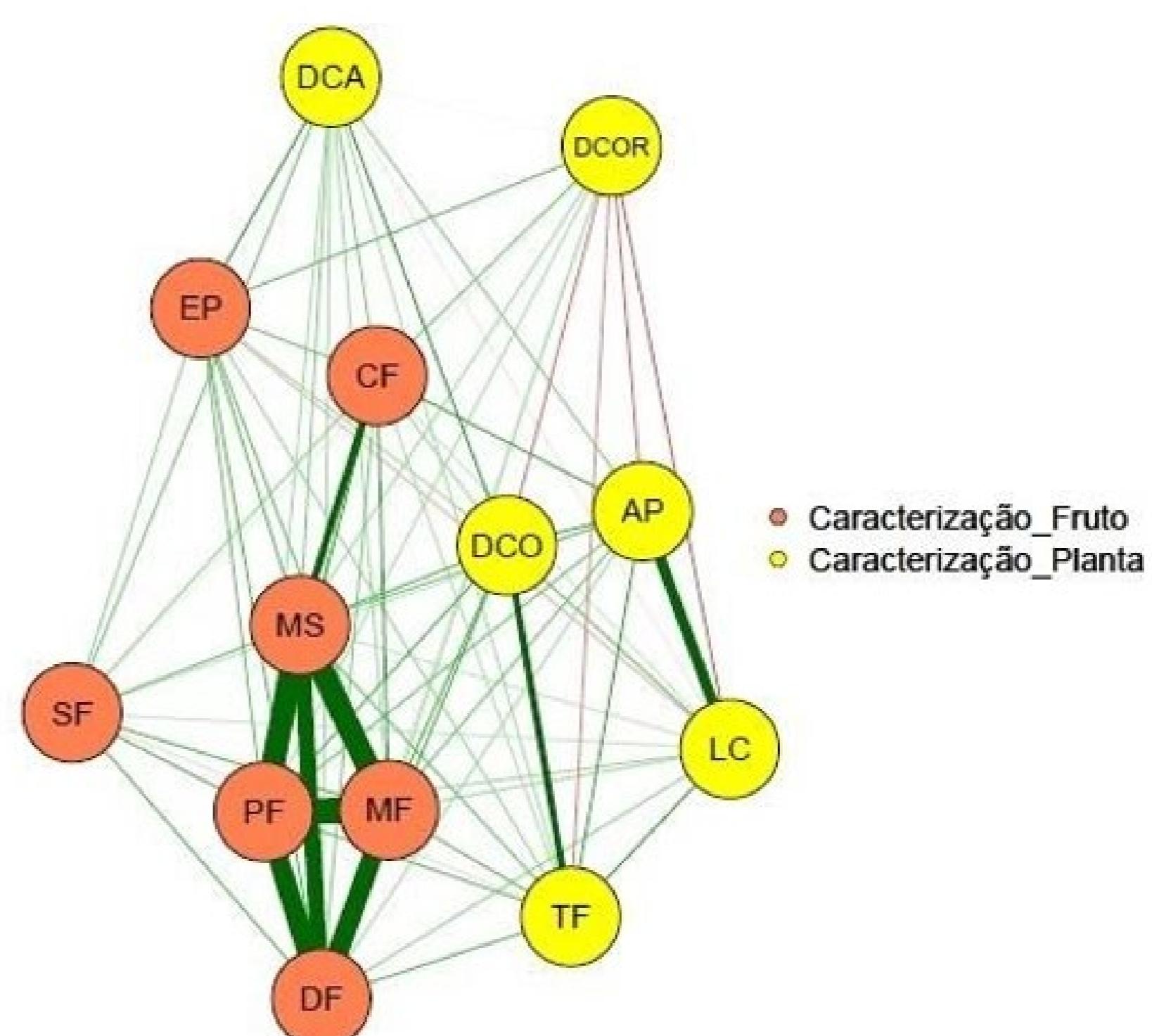


Figura 2 - Rede de correlação fenotípica entre caracteres de *C. chinense* e *C. frutescens*. Linhas vermelhas e verdes representam correlações negativas e positivas, respectivamente. PF: Peso do fruto maduro (g), CF: Comprimento do fruto maduro (cm), EP: Espessura do pericarpo (cm), SF: Número de sementes por fruto, MF: Massa total do fruto fresco (g), MS: Massa total do fruto seco (g), DF: Diâmetro do fruto (cm), AP: Altura da planta (cm), DCO: Diâmetro da copa (cm), TF: Tamanho da folha (cm), LC: Longitude do caule (cm), DCA: Diâmetro do caule (cm) e Diâmetro da corola (cm).

Os caracteres selecionados contribuem para a discriminação da diversidade genética em *Capsicum* spp.

O uso de algoritmos de classificação é eficiente na avaliação de caracteres morfo-agronômicos e na seleção de genótipos divergentes em programas de melhoramento.

AGRADECIMENTOS

Os autores agradecem à FAPEMIG (Processo: BPD-01045-22).