

**INSTITUTO NACIONAL DE PESQUISAS DA AMAZÔNIA – INPA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS (BOTÂNICA)**

**COMPEXO ECCLINUSA GUIANENSIS (SAPOTACEAE,  
CHRYSOPHYLLOIDEAE): UMA ABORDAGEM MORFOLÓGICA,  
ECOLÓGICA E ESPECTRAL**

TATIANA CRISTINA ROTTA DE CARVALHO

Manaus, Amazonas

Agosto, 2017

TATIANA CRISTINA ROTTA DE CARVALHO

**COMPEXO ECCLINUSA GUIANENSIS (SAPOTACEAE,  
CHRYSOPHYLLOIDEAE): UMA ABORDAGEM MORFOLÓGICA,  
ECOLÓGICA E ESPECTRAL**

ORIENTADOR: MÁRIO HENRIQUE TERRA ARAUJO

Dissertação apresentada ao Instituto Nacional de Pesquisas da Amazônia como parte dos requisitos para obtenção do título de Mestre em Ciências Biológicas, área de concentração Botânica.

Manaus, Amazonas

Agosto, 2017



ATA DEFESA PÚBLICA DE DISSERTAÇÃO DE Mestrado DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM BOTÂNICA DO INSTITUTO NACIONAL DE PESQUISAS DA AMAZÔNIA.

Aos vinte dois dias do mês de Agosto de 2017 às 14:00h, na sala de seminários da Biblioteca do INPA-Campus I, reuniu-se a Comissão Examinadora da Defesa Pública, composta pelos seguintes membros: Dr. Charles Eugene Zartman, do Instituto Nacional de Pesquisas da Amazônia (INPA), Dra. Maria de Fátima Melo, do Instituto Nacional de Pesquisas da Amazônia (INPA), e Dr. Bruno Sampaio Amorim, da Universidade Federal de Pernambuco (UFPE) tendo como suplentes: Dra. Maria Anália Duarte de Souza, da Universidade Federal do Amazonas (UFAM), e Dr. Michael John Gilbert Hopkins, do Instituto Nacional de Pesquisas da Amazônia (INPA), sob a presidência do primeiro, a fim de proceder a arguição pública da DISSERTAÇÃO DE Mestrado, intitulada: "*Ecclinusa Guianensis* Eyma (Sapotaceae, Chrysophylloideae): Uma ou mais espécies?" discente: Tatiana Cristina Rotta de Cravalho, sob orientação: Dr. Mário Henrique Terra Araújo. Após a exposição, dentro do tempo regulamentar, a discente foi arguida oralmente pelos membros da Comissão Examinadora, tendo recebido o conceito final:

EXAMINADORES	PARECER	ASSINATURA
CHARLES EUGENE ZARTMAN	<input checked="" type="checkbox"/> APROVADO	<input type="checkbox"/> REPROVADO
MARIA DE FÁTIMA FIGUEIREDO MELO	<input checked="" type="checkbox"/> APROVADO	<input type="checkbox"/> REPROVADO
BRUNO SAMPAIO AMORIM	<input checked="" type="checkbox"/> APROVADO	<input type="checkbox"/> REPROVADO
MARIA ANÁLIA DUARTE DE SOUZA	<input type="checkbox"/> APROVADO	<input type="checkbox"/> REPROVADO
MICHAEL JOHN GILBERT HOPKINS	<input type="checkbox"/> APROVADO	<input type="checkbox"/> REPROVADO

Manaus (AM), 22 de agosto de 2017.

OBS: \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_

Nada mais havendo, foi lavrada a presente ata, que, após lida e aprovada, foi assinada pelos membros da Comissão Examinadora.

Michael John Gilbert Hopkins  
 Presidente da Comissão Examinadora  
 Pro. Ciências Botânicas  
 FD. 258/2014 - 1

C331c Carvalho, Tatiana Cristina Rotta de  
Complexo *Ecclinusa guianensis* (Sapotaceae, Chrysophylloideae):  
uma abordagem morfológica, ecológica e espectral/ Tatiana Cristina  
Rotta de Carvalho. - Manaus: [s.n.], 2017, 85 f.: il.

Dissertação (Mestrado) --- INPA, Manaus, 2017.

Orientador: Terra-Araujo, Mário Henrique

Área de concentração: Botânica

1. Complexo de espécies. 2. Taxonomia. 3. Sistemática. 4.  
Morfologia. 5. NIR. I. Título.

CDD:580

**Sinopse:**

Este estudo baseia-se em um complexo de *Ecclinusa guianensis*, onde a espécie apresenta dois morfotipos distintos na região de Manaus e não reconhecidos em literatura. Com base em evidências morfológicas, espectrais e geográficas avaliou-se a existência de um ou mais táxons.

**Palavras-chave:** complexo de espécies, taxonomia, sistemática, morfologia, NIR.

## AGRADECIMENTOS

---

Ao longo destes dois anos e meio de mestrado muitas pessoas me ajudaram a passar por essa fase e superar todas as dificuldades com as quais me esbarrei.

Primeiramente agradeço à minha família, especialmente aos meus pais Elizabete e Celso e meu irmão Matheus, que sem todo o suporte que me ofereceram nada disso seria possível.

Ao meu orientador Mário Terra, que sempre esteve presente durante todo o trabalho, disponibilizando toda bibliografia e conhecimento necessários para a pesquisa, respondendo com agilidade meus e-mails, me apoiando em meus momentos de desespero com conselhos sábios e se preocupando com minha saúde mental.

Ao Alberto Vicentini, que ofereceu suporte nas análises estatísticas, discussões sobre complexos de espécie e inserção de informações nos bancos de dados.

Aos meus companheiros de república, Jana, Delano e Erika, por estarem presentes em vários momentos incríveis, datas comemorativas e dificuldades. Por todos os programas, comidas e festas aproveitados, séries e filmes vistos juntos e infinitos carinhos na Amy e Floquinho.

Aos meus companheiros de turma de mestrado, Fabi, Laura, Alex, Carol e Magno pelos inúmeros almoços, boas risadas e conversas, conforto nos momentos difíceis e apoio na realização do projeto. Levarei vocês em meu coração com muito amor e carinho, nossa amizade é muito especial e importante para mim.

Aos demais discentes do programa da botânica por toda distração, conversas enriquecedoras e cafés.

Ao Nicolas e Milena, estagiários do meu orientador que me ajudaram em parte da metodologia realizada durante o projeto.

À Carla, que mesmo distante continua sendo uma amiga incrível com uma família maravilhosa, que sempre esteve presente com conselhos válidos em momentos de desespero e sempre disposta a me dar atenção.

Ao Ronaldo, grande amigo de Alfenas, que tirou um tempo (mesmo que curto) pra me visitar em Manaus e viver os melhores e mais loucos momentos nessa selva que tanto gostaria de conhecer.

Ao Beto, que mesmo sendo de uma área de trabalho totalmente diferente sempre estava disposto a me escutar reclamando e ficando feliz a cada análise bem sucedida. Que ao longo desse tempo de amizade sincera me acolheu durante as festas de final de ano, me tornando uma pessoa melhor a cada mês. Às diversas comidinhas feitas com muito carinho, tanto quanto possuímos um com o outro. Ao abraço confortante, às conversas intermináveis e aos cafés deliciosos no começo das manhãs.

Ao Viktor, pelos momentos divertidos e risadas, muitos cinemas e filmes vistos em casa. Que sempre tem elogios guardados pra mim e que me acompanhou em um dos momentos mais importantes da minha vida, minha primeira tatuagem! Que dividiu comigo os seus problemas e me ajudou a enfrentar os meus, sendo sempre compreensivo.

Ao Dennis, muito presente nesse final de mestrado, fundamental nessa fase, bem estressante e difícil. Pelas conversas sobre a vida, por me dar suporte nas minhas agonias, ter a paciência de um monge comigo, me tirar de casa pra me exercitar (cuidando do meu bem estar) e sempre me fazer acreditar em mim mesma.

Aos meus amigos distantes, que mantenho contato apenas por internet, mas que me fazem rir, discutir sobre inúmeros assuntos e matar as saudades dos lugares os quais já visitei e morei.

Ao pessoal dos herbários do Rio de Janeiro, do INPA, e do Missouri, que ofereceram toda estrutura e material disponível para a realização de grande parte do meu mestrado.

Às secretárias da botânica, Léia e Neide, por cuidarem de toda burocracia e facilitarem sempre minha vida.

Ao coordenador do PPG BOT, Mike, pela disponibilidade e empenho em resolver os contratempos que surgiram ao longo do curso.

À CAPES, pelo financiamento durante esses dois anos.

À banca, por aceitarem o convite para minha defesa.

A todos que participaram indiretamente e contribuíram de alguma forma para a execução e finalização desse trabalho.

## RESUMO

---

Na Amazônia Central, a quantidade de espécies de Sapotaceae se iguala ou mesmo supera os índices de diversidade de outras famílias muito bem representadas na região, como Fabaceae e Lecythydaceae. Apenas na parcela de 25ha do Projeto Dinâmica Biológica de Fragmentos Florestais (PDBFF), em Manaus-AM, há 20.438 árvores pertencentes à Sapotaceae, correspondentes à 101 espécies, das quais 38 não tem nome científico formal devido à ausência de material reprodutivo e auxílio de especialistas. Sabe-se que a delimitação/identificação de espécies de Sapotaceae, principalmente em áreas ricas como a Amazônia Central, nem sempre é uma tarefa fácil, e muitas vezes se faz necessário o uso de diversas fontes de evidência que auxiliem nessa tarefa. *Ecclinusa guianensis* Eyma, popularmente conhecida como abiurana-bacuri é uma das espécies mais comuns e amostradas dentro das parcelas do PDBFF. Entretanto, apesar da sua abundância, *E. guianensis* representa um dos 23 complexos taxonômicos para Chrysophylloideae. A espécie é amplamente distribuída na Amazônia, morfológicamente variável e com muitas formas intermediárias ocorrendo em simpatria nas áreas do PDBFF e Reserva Ducke. O objetivo deste estudo foi integrar diferentes fontes de evidência (eg. análises morfométricas e dados espectrais, geografia e ecologia) para avaliar se as formas de *E. guianensis* correspondem a uma única espécie, subespécies ou mesmo, espécies diferentes. Os dados morfológicos foram submetidos a técnicas multivariadas e a análises de agrupamento para testar os limites entre os morfotipos, que permitiram o reconhecimento de dois grupos conforme as identificações que receberam *a priori*. Acrescentando os dados espectrais os mesmos agrupamentos ainda se mantêm, sem alterações, corroborando a presença dois táxons dentro da espécie. Entretanto, a ausência de flores pertencentes à forma “B”, a presença de heterofilia e as diferentes respostas espectrais nos fazem concluir que as formas reconhecidas dos morfotipos correspondem a indivíduos em diferentes estágios de desenvolvimento (i.e., jovens e adultos).

**Palavras-chave:** complexo de espécies, taxonomia, sistemática, morfologia, NIR.



## ABSTRACT

---

### ECCLINUSA GUIANENSIS COMPLEX (SAPOTACEAE, CHRYSOPHYLLOIDEAE): A MORPHOLOGICAL, ECOLOGICAL AND SPECTRAL APPROACH

In Central Amazonia, the number of species of Sapotaceae equals or even exceeds the diversity indexes of other families very well represented in the region, such as Fabaceae and Lecythidaceae. Only in the 25ha plots from the Projeto Dinâmica Biológica de Fragmentos Florestais (PDBFF), in Manaus-AM, there are ca. 10.709 Sapotaceae trees, corresponding to 101 species, of which for 38 do not have a formal scientific name. It is known that the delimitation / identification of Sapotaceae species, especially in rich areas such as Central Amazonia is not always an easy task, and it is often necessary to use several sources of evidence to assist this task. *Ecclinusa guianensis* Eyma, popularly known as Abiuruna-bacuri is one of the most common and sampled species within the plots of the PDBFF. However, in spite of its abundance, *E. guianensis* represents one of the many taxonomic complexes (n = 23) for Chrysophylloideae. This species is widely distributed in the Amazon rain forest, morphologically variable and with many intermediate forms occurring in sympathy in the areas of PDBFF and Reserva Ducke, both around Manaus. The goal of this study was to integrate different sources of evidence (eg, morphology, spectral data, geography and ecology) to evaluate if the forms of *E. guianensis* correspond to a single species, subspecies or even species. Morphological data were submitted to the multivariate techniques and sensitivity analyzes to test the boundaries between the morphotypes, which allowed the recognition of two groups according to the identifications they received *a priori*. Adding the spectral data the same groupings still remain, confirming the presence of two taxa within the species. However, the absence of flowers belonging to the "B" form, the presence of heterofilia and the different spectral responses suggest that the morphotypes correspond to individuals at different stages of development (ie., young and adult).

**Keywords:** Species complex, taxonomy, systematic, morphology, NIR.

## SUMÁRIO

---

<b>APRESENTAÇÃO</b> .....	9
<b>OBJETIVOS</b> .....	13
<b>CAPÍTULO ÚNICO</b> .....	14
Abstract.....	17
Introdução.....	18
Material e Métodos .....	19
Material amostrado.....	19
Variação morfológica.....	20
Variação espectral .....	20
Conceito de espécie.....	21
Análise dos dados.....	21
Resultados.....	22
Distribuição geográfica e habitat.....	22
Morfologia.....	22
Morfologia vs NIR .....	23
Discussão .....	24
Conclusão .....	27
Bibliografia Citada .....	28
Lista de figuras .....	35
Lista de tabelas. ....	42
Apêndice.....	43

## APRESENTAÇÃO

---

As florestas neotropicais compõem um dos principais biomas do planeta, reconhecidas pela alta diversidade e endemismo de espécies vegetais (Denslow, 1987; Burnham e Graham, 1999; Antonelli e Sanmartin, 2011; Hernández-Ruedas *et al.*, 2014) com destaque para a região amazônica, que abriga aproximadamente 16.000 espécies arbóreas (ter Steege *et al.*, 2013), sendo que destas, entre 200 e 300 podem ser encontradas em menos de um único hectare de floresta na Amazônia Central (Oliveira e Nelson, 2001).

Sapotaceae Juss., um dos representantes de Ericales (Stevens e Davis, 2001) é amplamente distribuída nos trópicos e constitui um importante componente das florestas tropicais (Swenson *et al.*, 2008a, 2008b). A origem da família ainda é incerta (Austrália ou África) estimada há cerca de 103 M.a (Swenson e Anderberg, 2005), colonizando a região neotropical por meio de um evento de dispersão a longa distância através do oceano Atlântico ou Índico há aproximadamente 59 M.a. (Bartish *et al.*, 2011). Entretanto, mesmo com alta diversidade e abundância de espécies, a taxonomia de Sapotaceae tem sido notoriamente difícil e diferentes sistemas de classificação já foram propostos (Lam, 1939; Aubréville, 1964; Baehni, 1965; Pennington, 1990, 1991), levando a grupos não naturais devido ao alto grau de homoplasia em seus caracteres morfológicos (Swenson *et al.*, 2008a, 2008b, 2014).

Na Amazônia Central, a quantidade de espécies da família se iguala ou mesmo supera os índices de diversidade de outras muito bem representadas na região, como Fabaceae Lindl. e Lecythidaceae A. Rich. (Pennington, 1991; Milliken, 1998; ter Steege *et al.*, 2000). Apenas numa parcela de 25 hectares do Projeto Dinâmica Biológica de Fragmentos Florestais (PDBFF), há 20.438 árvores pertencentes à Sapotaceae, correspondentes a 101 espécies, das quais cerca 38 não ratificaram sua identificação por especialistas. Muitas destas espécies geram incertezas quanto à sua classificação por nunca serem coletadas com flores e/ou frutos, o que impede utilização da literatura especializada. Além disso, dentro das parcelas do PDBFF também foram identificados 23 complexos de espécies pertencentes a esta família, ou seja, grupos de espécies em que os limites e/ou o status de espécie não estão bem definidos ([www.botanicaamazonica.wiki.br/plantasPDBFF](http://www.botanicaamazonica.wiki.br/plantasPDBFF), 2017).

A última revisão taxonômica para Sapotaceae foi feita por Pennington (1990). Em seu trabalho foram reconhecidos 12 gêneros e cerca de 450 espécies para os Neotrópicos, das quais 103 são consideradas endêmicas do Brasil (Carneiro *et al.*, 2015). Três clados, correspondentes a três subfamílias (i.e., Sarcospermatoideae, Sapotoideae e Chrysophylloideae) são atualmente reconhecidos e aceitos para Sapotaceae (Bartish *et al.*, 2005; Swenson *et al.*, 2007, 2008a). Sarcospermatoideae inclui um pequeno gênero, *Sarcosperma* Hook. f., e possivelmente *Eberhardtia* Lecomte (Smedmark *et al.*, 2006), apontados como grupos irmãos de Sapotaceae (Anderberg *et al.*, 2002) enquanto Sapotoideae possui a maioria dos gêneros localizada da região Indo-Pacífica e divide-se em três tribos: Sideroxyloae, Sapoteae e Isonandreae, esta última ainda pouco estudada e apresentando polifiletismo (Smedmark *et al.*, 2006). Já Chrysophylloideae é a subfamília que se diversificou mais rapidamente, contendo cerca de 28 gêneros e 600 espécies, com maior diversidade na América do Sul e Australásia (Anderberg e Swenson, 2003; Swenson *et al.*, 2008a, 2008b). Embora a subfamília seja monofilética, as relações internas ainda estão mal resolvidas, com muitos gêneros (e.g., *Chrysophyllum* Linnaeus e *Pouteria* Aublet) para ou polifiléticos (Anderberg e Swenson, 2003; Swenson e Anderberg, 2005; Swenson *et al.*, 2008b). A classificação mais atual sugere o resgate de um gênero antes extinto a fim de diminuir as politomias (Faria *et al.*, 2017).

*Ecclinusa* Martius, um dos membros de Chrysophylloideae, é um gênero neotropical com 12 espécies atualmente aceitas (Pennington, 1990; Terra-Araujo *et al.*, 2015) e monofilético (Swenson e Anderberg, 2005; Terra-Araujo *et al.*, 2015; Faria *et al.*, 2017). Morfologicamente, *Ecclinusa* é facilmente distinta dos demais gêneros de Sapotaceae pela combinação de poucos caracteres incluindo a presença de estípulas, flores sésses e ausência de estaminódios. Porém, o gênero pode ser confundido com *Chromolucuma* Ducke que também possui estípulas, mas difere quanto ao látex amarelado, branco em *Ecclinusa*, flores com pedicelo e presença de estaminódio (Pennington, 1990; 2006; Swenson *et al.*, 2008a, 2008b, Terra-Araujo *et al.*, 2015).

As flores em *Ecclinusa* podem ser uni- ou bissexuadas e seus frutos, geralmente monospermicos, possuem uma semente globosa ou lateralmente comprimida, com radícula inclusa e endosperma ausente (Pennington, 1990, 2006). Os representantes do gênero em sua maioria são árvores de médio a grande porte (8–30m alt.), e raramente arbustos, encontrados em savanas, florestas de terra firme, campinas e campinaranas. A maior parte das espécies

ocorre na região amazônica, mas *Ecclinusa ramiflora* Martius pode ser encontrada na Costa Atlântica Brasileira, desde o estado do Amapá, ao norte, até o estado do Paraná, ao sul (Pennington, 1990; <http://floradobrasil.jbrj.gov.br/reflora/listaBrasil>, 2015). Do total de espécies conhecidas, cinco são citadas pela IUCN - International Union for Conservation of Nature - (2015) como ameaçadas, sendo três destas consideradas vulneráveis (*Ecclinusa lancifolia* (Mart. & Eichler ex Miq.) Eyma, *Ecclinusa orinocoensis* Aubrév. e *Ecclinusa parviflora* T.D. Penn.).

*Ecclinusa guianensis* Eyma é uma espécie morfo e ecologicamente variável, especialmente em relação ao tamanho da folha, número de pares da venação secundária, tipo de venação, partes florais e tamanhos do fruto, sendo amplamente distribuída e encontrada em toda a região amazônica, incluindo Bolívia, Guianas, Venezuela, Brasil, Peru e Suriname (Pennington, 1990). Os indivíduos são geralmente árvores de dossel com cerca de 30 metros, mas também se encontram na forma de arvoretas de médio porte com aproximadamente 7 metros de altura, encontrados em florestas de terra firme e, menos frequentemente, em savanas. Filogeneticamente, a espécie é proximamente relacionada à *Ecclinusa atabapoensis* (Aubrév.) T.D. Penn. (Pennington, 1990; Terra-Araujo, 2013) e, de acordo com o banco de dados do PDBFF ([www.botanicaamazonica.wiki.br/plantasPDBFF](http://www.botanicaamazonica.wiki.br/plantasPDBFF), 2017), e a Flora da Reserva Ducke (Ribeiro *et al.*, 1999), no entorno de Manaus, dois morfotipos, informalmente descritos como forma "A" e "B" podem ser reconhecidos dentro de *E. guianensis*. A forma "A" consiste em indivíduos com folhas oblanceoladas, apresentando venação eucamptódroma a broquidódroma, com nervuras intersecundárias ausentes, de grande porte (> 15m alt.), com tronco rugoso pela formação de cócleas e que habitam todos os ambientes, exceto campinaranas. Já a forma "B" apresenta indivíduos com folhas lanceoladas, venação eucamptódroma e intersecundárias conspícuas, com porte médio ( $\leq$  15m alt.), tronco liso ou com poucas estrias e habitam regiões de platô e vertente (Ribeiro *et al.*, 1999). Em campo, as formas podem ser distinguidas com base na variação dos caracteres descritos anteriormente, entretanto verifica-se a presença de intermediários entre os morfotipos, dificultando o seu reconhecimento.

Na Amazônia os inventários florísticos existentes são defasados por conter, em sua grande maioria, coletas de determinados locais com maior facilidade de acesso, sendo estes reconhecidos como áreas de grande diversidade (Prance, 2001; Hopkins, 2007). Com isso, erros de identificação se tornam frequentes, pois grande parte do material coletado apresenta

apenas estruturas vegetativas, há poucos estudos taxonômicos e guias de identificação da flora local (Gomes *et al.*, 2013), e em estágio vegetativo poucos caracteres úteis para a delimitação e identificação de espécies estão disponíveis (Camargo *et al.*, 2008). Há também o fato de que as divergências morfológicas das estruturas vegetativas podem refletir diferenças entre indivíduos jovens e adultos (já que plantas jovens são incluídas nos inventários de parcelas permanentes), ou ser resultado de plasticidade fenotípica de uma mesma espécie (Schlichting, 1986; Sultan, 2000), o que torna a delimitação/identificação de espécies uma tarefa difícil. Todos os fatores citados anteriormente contribuem para o surgimento de complexos, onde os limites de uma espécie não estão bem definidos e se torna confusa sua delimitação.

Atualmente, novas tecnologias para análise e levantamento de dados, como os morfométricos (Moeller *et al.*, 2007; Esteves e Vicentini, 2013; Rellstab *et al.*, 2016), de espectroscopia do Infravermelho (FT-NIR) (Kim *et al.*, 2004; Strgulc Krajšek *et al.*, 2008; Durgante *et al.*, 2013; Jin *et al.*, 2017), e a distribuição geográfica (Medrano *et al.*, 2014; Fiala *et al.*, 2016), pretendem ajudar na resolução de alguns dos problemas relacionados à delimitação/identificação de espécies, principalmente quando combinados. Esses dois processos (identificação e delimitação) são distintos, porém relacionados. A delimitação de espécies consiste no reconhecimento de entidades biológicas baseado num conceito e no uso de algum critério, como caracteres morfológicos (De Queiroz, 1998). Já a identificação de espécies é o processo de reconhecer ou provar a identidade dos espécimes, realizada frequentemente por comparação com um espécime de herbário que se assume ter uma identificação confiável efetuada por especialistas. Nesse sentido, espécies são fundamentais para análises de estudos ecológicos, biogeográficos, evolutivos e sistemáticos e, cada vez mais se amplia a literatura acerca dos métodos empíricos para sua delimitação (Sites Jr e Marshall, 2004).

## OBJETIVOS

---

**Objetivo geral:** verificar se os morfotipos “A” e “B” de *Ecclinusa guianensis* sugeridas para a região de Manaus podem ser sustentados por diferentes fontes de evidência (e.g., morfologia, ecologia, geografia e espectral).

**Objetivos específicos:**

- Avaliar, através de informações morfológicas e ecológicas se é possível o reconhecimento de grupos entre os morfotipos diagnosticados;
- Avaliar a acurácia do NIR como ferramenta para uma identificação dos grupos dentre os morfotipos reconhecidos em *E. guianensis*;

## CAPÍTULO ÚNICO

---

CARVALHO, T.C.R. e TERRA-ARAUJO. M.H. Complexo *Ecclinusa guianensis*

(Sapotaceae, Chrysophylloideae): uma abordagem morfológica, ecológica e espectral.

Manuscrito formatado de acordo com as normas da Acta Amazônica.



**Complexo *Ecclinusa guianensis* (Sapotaceae, Chrysophylloideae): uma abordagem morfológica, ecológica e espectral**

***Ecclinusa guianensis* complex (Sapotaceae, Chrysophylloideae): a morphological, ecological and spectral approach**

Tatiana Cristina R. de CARVALHO\* e Mário H. TERRA-ARAÚJO

Instituto Nacional de Pesquisas da Amazônia, Programa de Pós-Graduação em Botânica, Av. André Araújo 2936, 69067-375, Manaus, Amazonas, Brasil; carvalhotatiana@gmail.com

## RESUMO

Sapotaceae é uma das famílias mais bem representadas no mundo, e somente na Amazônia Central supera os índices de diversidade de Fabaceae e Lecythidaceae. Sua abundância é tão significativa que em uma parcela de 25ha do Projeto Dinâmica Biológica de Fragmentos Florestais (PDBFF) são encontradas 20.438 árvores pertencentes à família. Distribuídas em 101 espécies, 38 não possuem nome científico formal pela ausência de material reprodutivo disponível e auxílio de especialistas. Assim, Sapotaceae é reconhecida por possuir uma taxonomia difícil, necessitando de diversas metodologias na resolução de complexos de espécies. Nesta região destaca-se *Ecclinusa guianensis* Eyma, uma das mais bem representadas nas parcelas do PDBFF, representando um dos 23 complexos presentes nessa área. É amplamente distribuída na Amazônia, e dois morfotipos foram reconhecidos e nomeados informalmente; as formas "A" e "B", que ocorrem em simpatria nas áreas do PDBFF e Reserva Ducke. O objetivo deste estudo foi integrar diferentes fontes de evidência (análises morfométricas, dados espectrais, geografia e ecologia) para avaliar se as formas de *E. guianensis* representam uma ou mais espécies distintas. De acordo com as análises multivariadas e de agrupamento dos dados morfológicos, foi possível o reconhecimento de dois grupos conforme as identificações iniciais. Acrescentando os dados espectrais os dois grupos se mantiveram, corroborando a presença de dois táxons dentro da espécie. Entretanto, a ausência de material reprodutivo em plantas da forma "B", a presença de heterofilia e as diferentes respostas espectrais nos levam a concluir que as formas reconhecidas dentro da espécie correspondem a indivíduos jovens e adultos.

**Palavras-chave:** complexo de espécies, taxonomia, sistemática, morfologia, NIR.

## INTRODUÇÃO

Os erros de identificação de espécie são comuns na Amazônia por diversos fatores. Os inventários contêm coletas de locais com maior acessibilidade, que são tidos como áreas de grande diversidade (Prance 2001; Hopkins 2007). Há falta de estudos taxonômicos e guias de identificação local, além de muito material somente com estruturas vegetativas (Gomes *et al.* 2013) com poucos caracteres úteis na identificação de espécies (Camargo *et al.* 2008). A variação morfológica pode heterofilia ou plasticidade fenotípica da espécie (Schlichting 1986; Sultan 2000), tornando o trabalho exigente.

Sapotaceae Juss. é altamente rica e abundante na Amazônia, as vezes até superando Fabaceae e Lecythidaceae (Pennington 1991; Milliken 1998; ter Steege *et al.* 2000), mas com taxonomia difícil, onde diversos sistemas de classificação já foram propostos (Baehni 1938, 1965; Lam 1939; Aubréville 1964; Pennington 1990, 1991). Em 25 hectares do PDBFF, cerca de 20 mil indivíduos da família equivalem a 101 espécies, 38% sem identificação retificada. A incerteza se deve a coletas sem flores e/ou frutos, o que impede utilização de guias especializados. De acordo com a base de dados do PDBFF (2017) a região contém 23 complexos, grupos em que os limites e/ou o status de espécie não estão definidos.

Apesar de Chryshophylloideae conter a maior quantidade de complexos taxonômicos em Sapotaceae, onde a maioria dos gêneros são para ou polifiléticos, como *Pouteria* e *Chrysophyllum*, *Ecclinusa* se mostra como um grupo monofilético, sustentado por dados moleculares e morfológicos (Swenson e Anderberg 2005; Terra-Araujo *et al.* 2015).

*Ecclinusa guianensis* Eyma é morfo e ecologicamente variável, presente em toda a região amazônica (Pennington 1990), abundante dentro das parcelas do PDBFF ([www.botanicaamazonica.wiki.br/plantasPDBFF](http://www.botanicaamazonica.wiki.br/plantasPDBFF) 2017), e dois morfotipos reconhecidos para a região de Manaus (Ribeiro *et al.* 1999). Na forma “A” constam indivíduos com folhas oblanceoladas, venação eucamptódroma a broquidódroma, com nervuras intersecundárias ausentes, grande porte (>20m) e tronco rugoso por cócleas, habitando todos os ambientes, exceto campinaranas. Já a forma “B” caracteriza-se por folhas lanceoladas, venação eucamptódroma e intersecundárias conspicuas, porte médio (<15m), tronco liso ou com poucas estrias e habitam regiões de platô e vertente (Ribeiro *et al.* 1999).

Das tecnologias para análise de dados, os morfométricos (Moeller *et al.* 2007; Esteves e Vicentini 2013), a espectroscopia do Infravermelho (FT-NIR) (Strgulc Krajsek *et al.* 2008, Jin

*et al.* 2017), e a distribuição geográfica (Fiala *et al.* 2015; Meldrano *et al.* 2016), pretendem ajudar na resolução de alguns dos problemas relacionados à delimitação/identificação de espécies, principalmente quando combinados. Assim, o trabalho objetiva verificar se os morfotipos de *E. guianensis* se sustentam de acordo com a separação morfológica sugerida, utilizando dados morfométricos, espectrais e geográficos.

## MATERIAL E MÉTODOS

**Material amostrado.** A fim de explorar os limites entre os morfotipos de *E. guianensis* foi realizada uma busca online em coleções botânicas através das plataformas Re flora (<http://floradobrasil.jbrj.gov.br/reflora/listaBrasil> 2015) e Species link (<http://splink.cria.org.br/> 2015) que contêm registros para a espécie para obter informações sobre o habitat e distribuição geográfica. Os dados foram compilados utilizando o programa R (R core Team 2012), com a inserção dos pacotes MASS e VEGAN, e o resultado permitiu verificar em quais herbários as amostras estão localizadas (JBRB, MO, INPA, PPBio e PDBFF) e qual a quantidade disponível (n=248). Posteriormente foi realizada a análise morfométrica do material depositado nas coleções, onde foram mensurados caracteres de 229 exsicatas (Anexo I) que procuraram abranger todas as áreas de distribuição da espécie.

Para a obtenção dos dados espectrais foram selecionadas 201 exsicatas do total de amostras disponíveis. As amostras frescas do trabalho de campo foram coletadas onde os dois morfotipos são observados, na Reserva Florestal Adolpho Ducke (RFAD) e ARIE PDBFF para a obtenção de registros fotográficos. Os espécimes foram localizados a partir das coordenadas contidas no banco de dados do PDBFF e informações nas etiquetas das exsicatas depositadas em coleções, sendo coletados cinco indivíduos de cada morfotipo e utilizados para registros fotográficos.

**Varição morfológica.** Os vinte e quatro caracteres vegetativos levantados durante o desenvolvimento do presente trabalho foram escolhidos de acordo com os caracteres descritos por Pennington (1990, 2004), e pelo livro Flora da Reserva Ducke (Ribeiro *et al.* 1999) (Tabela 1). Além dos caracteres vegetativos optou-se também pelo acréscimo de três informações ecológicas (Tabela 2). Os caracteres são, em grande parte, usualmente utilizados para distinguir as formas de *E. guianensis* em campo (Ribeiro *et al.* 1999). Um paquímetro digital foi utilizado para a tomada das medidas morfológicas.

**Variação espectral.** O NIR (Near InfraRed) é uma técnica não destrutiva, rápida e que não necessita de um tratamento prévio das amostras (Rodríguez Fernández 2008), na qual uma luz no espectro do infravermelho é incidida sobre qualquer molécula que possua em sua estrutura as ligações C=O, C-H, C-N, N-H, O-H, NO<sub>2</sub> e S-H (Pasquini 2003; Li *et al.* 2012). Este material pode ser desde partes de plantas, como raiz, casca do tronco, folhas e até animais, como insetos. A luz, que possui os comprimentos de onda do infravermelho médio medindo entre 690 a 3.000 nm, ao entrar em contato com o tecido biológico, gera uma resposta espectral de acordo com a composição química e estrutural da folha (Ponzoni 2002). Uma parte desta luz é absorvida pelo tecido biológico e outra parte refletida, onde a diferença entre os valores emitidos e recebidos demonstra a quantidade absorvida pelo material analisado, gerando diversas leituras que representam um espectro, composto por um conjunto com diversos valores de absorbância em vários comprimentos de onda (Kaye 1954; Workman Jr e Weyer 2007). Consequentemente, os materiais que possuem espectros relativamente próximos são classificados como pertencentes ao mesmo grupo. Esta ferramenta vem sendo utilizada com grande frequência em plantas amazônicas, e tem apresentado bons resultados na identificação em nível de espécie, com uma acurácia superior a 75% (Curty 2013; Durgante *et al.* 2013; Lang *et al.* 2015).

As leituras dos espectros ocorreram por meio do espectrofotômetro da marca Termo Nicollet, sistema FT-NIR Antaris II Method Development System (MDS), disponível no herbário do INPA. Para a obtenção das leituras, as amostras foram posicionadas sobre o emissor de radiação do equipamento e acima desta colocou-se um objeto opaco de cor negra para que não houvesse dispersão de luz. Para cada indivíduo foram escolhidas duas folhas e realizadas duas leituras por folha, uma da face abaxial e outra da adaxial, totalizando quatro leituras por indivíduo a fim de eliminar possíveis espectros causados por folhas com fungos ou líquens.

**Conceito de espécie.** Para testar a delimitação de *Ecclinusa guianensis* seguimos de Queiroz (1998, 2007), onde espécies são entendidas como linhagens metapopulacionais que evoluem separadamente umas de outras. Neste contexto, as linhagens podem ser morfologicamente diferentes, não necessariamente monofiléticas ou reprodutivamente isoladas para serem entendidas como espécie, dependendo de quais critérios e métodos serão utilizados para tal inferência. Na prática, qualquer informação que forneça evidências de separação entre as linhagens é relevante para inferir as fronteiras e o número de espécies de um complexo.

**Análise dos dados.** Os dados morfométricos foram inicialmente categorizados e transformados em semi quantitativos, e em seguida foi retirada a média dos caracteres que possuíam dados quantitativos. Posteriormente foi calculada a distância morfológica dos dados utilizando o índice de Gower (1971), que permite uma análise simultânea de dados qualitativos e quantitativos. Neste método os valores da matriz de distância ficam compreendidos entre 0 e 1, onde quanto mais próximo de 1 menos correlacionados os dados estão. Além disso, foi checada a correlação entre as variáveis, para que as que fossem fortemente correlacionadas e correspondessem à mesma informação pudessem ser substituídas por apenas uma para representá-las. Dessa forma, as variáveis “altura da largura máxima” e “comprimento total” foram excluídas e representadas por “comprimento da lâmina”. As análises passaram a ser rodadas com a exclusão dessas duas variáveis juntamente com as não modificadas e o novo conjunto de dados foi denominado variáveis selecionadas.

Para Técnica Estatística de Ordenação não Métrica (Non-metric Multidimensional Scale - NMDS) (Kruskal 1964) foram traçados os valores estresse, onde os índices mais baixos indicarão a quantidade de dimensões necessárias que a análise deve conter. Neste caso a NMDS foi gerada com 35 eixos (Fig. 1). Esta metodologia tem sido utilizada em trabalhos morfométricos (Mikkelsen e Seberg 2001; Moeller *et al.* 2007; Terra-Araujo *et al.* 2012) para a visualização de agrupamentos dentro de espécies com o intuito de obter uma sustentação morfológica dos grupos. Logo em seguida os dados foram submetidos à análise de agrupamento por três modelos; SVM (Support Vector Machine) (Vapnik 2013), Bayes (Carlin e Louis 2000) e LDA (Linear Discriminant Analysis) (Balakrishnama e Ganapathiraju 1998) a fim de verificar se são formados grupos morfológicos suportados pelos dados, independente da identificação contida na etiqueta das amostras das plantas. Como dentro de *E. guianensis* suspeita-se que há duas entidades, foram sugeridos até dez grupos (k=10) para o agrupamento das amostras a serem analisadas e foram considerados bons resultados os que obtiveram um suporte maior a 90% na maioria dos modelos testados.

Os dados espectrais obtidos foram filtrados, visto que para cada amostra foram feitas quatro leituras, duas da face abaxial e duas da adaxial. Dessa forma, foi necessário escolher aleatoriamente apenas um espectro para cada planta e gerar a média destes dados. Assim como feito para os dados morfológicos, foi realizada uma NMDS para que os espectros mais informativos sejam selecionados.

A combinação de caracteres morfológicos e dados espectrais seguiu o mesmo procedimento de análise de agrupamento citado anteriormente e foram selecionadas as amostras que tiveram seus caracteres vegetativos mensurados e submetidas à técnica do infravermelho próximo.

Após todas as análises de agrupamento os três grupos (k) que possuíram maiores probabilidades de serem formados foram plotados em matrizes de confusão. Em conjunto também foi gerada uma tabela com a identificação das amostras e o grupo as quais foram classificadas de acordo com cada k sugerido.

Todas as análises foram realizadas no programa R (R core Team 2012) com a adição dos pacotes MASS e VEGAN.

## RESULTADOS

**Distribuição geográfica e habitat.** *Ecclinusa guianensis* é distribuída por toda região da Amazônia, incluindo países vizinhos como Peru, Venezuela e Colômbia (Fig. 2). Na Amazônia Central a espécie possui uma boa amostragem nas áreas do PDBFF e RFAD, onde os morfotipos ocorrem. De acordo com coletas e observações realizadas, as formas “A” e “B” são encontradas nas mesmas áreas, e na maioria das vezes numa distância bem próxima. De acordo com a Flora da Reserva Ducke (Ribeiro *et al.* 1999) há uma diferença no habitat entre as duas formas, mas esta não foi observada. Ambos morfotipos foram avistados em áreas não inundáveis e solos argilosos, evidenciando que não há preferência por ambiente.

**Morfologia.** Ao checar a correlação entre cada variável (Fig. 3), comprimento da lâmina e comprimento total (0.99), altura da largura máxima e comprimento total (0.97) e altura da largura máxima e comprimento da lâmina (0.95) são as mais altas. Como todas estas variáveis expressam apenas um caráter, o tamanho da folha, estas foram substituídas por apenas uma delas, o comprimento da lâmina foliar. De acordo com o coeficiente de Gower (1971) as variáveis possuem uma alta correlação (Fig. 4).

De acordo com as análises de agrupamento, foram exploradas três hipóteses de grupos existentes; k=2, k=3 e k=4, onde k=2 é o número de grupos com melhor sustentação nos métodos testados, resultando frequentemente com uma probabilidade acima de 90% (Figs. 5 e 6). A organização das amostras em grupos gerou várias matrizes de confusão, onde as classificações dos três grupos sugeridas foram comparadas par a par. As amostras receberam a seguinte classificação: “sem classe”, quando as análises não conseguiram as encaixar em

nenhum grupo; “incertezas”, que obtiveram uma probabilidade de menor que 95% de pertencer a algum grupo; “core”, que foram classificadas com uma certeza acima de 95%.

Em relação aos agrupamentos formados apenas rodando os dados morfológicos, quando  $k=2$ , todas as plantas reconhecidas como morfotipo “B” foram colocadas no mesmo grupo (2), enquanto as demais (reconhecidas como forma “A”) formaram o outro grupo (1) (Fig. 7). Para  $k=3$ , o terceiro grupo foi formado majoritariamente por plantas contidas no grupo 1, mas não foi possível reconhecer um padrão morfológico ao formá-lo. O grupo 2 com o morfotipo B sofreu alteração de oito plantas, que foram colocadas no grupo 1. Destas oito plantas, a maioria foi classificada como incerteza, evidenciando que ao sugerir três grupos estes não apresentam nenhuma característica marcante que justifique seu agrupamento. Com  $k=4$  todas as plantas caracterizadas como forma “B” continuaram agrupadas (grupo 1), sendo este bem sustentado e quase não sofrendo alteração ao longo das análises, enquanto as amostras denominadas como forma “A” se dividiram entre os três grupos restantes. O grupo 4 possui todas as plantas classificadas anteriormente no grupo 3, o grupo 3 manteve parte de suas plantas com o acréscimo de algumas amostras e o grupo 2 foi formado com o restante das plantas. Estes três últimos grupos (2, 3 e 4) se formaram de modo que nenhuma característica morfológica exclusiva pudesse identificá-los.

Das amostras denominadas como morfotipo “A”, 11 % (26 indivíduos) foi classificada como pertencentes à forma “B” pela análise *a posteriori*. Estas amostras apresentam características referentes aos dois morfotipos e por isso foram classificadas como indivíduos intermediários. Além disso, todas as amostras pertencentes ao grupo 2, que foi sustentado mesmo com o aumento de  $k$ , são amostras provenientes do PDBFF e encontradas em uma só localidade, evidenciando que a presença de dois morfotipos ocorre apenas nesta região.

**Morfologia vs. NIR.** Na análise dos dados morfológicos e espectrais em conjunto, observa-se que para  $k=2$  todas as plantas agrupadas na análise morfológica foram sustentadas quando se acrescentaram os dados espectrais, mas 19% destas não foram atribuídas a nenhum grupo. As amostras do grupo 1 pertencem à forma “B” e as do grupo 2 à forma “A” (Fig. 8). Com  $k=3$  os dois grupos anteriores tiveram sua classificação invertida, e o terceiro grupo possui plantas antes pertencentes ao grupo 2. O número de amostras analisadas diminuiu, pois as classificadas como “sem classe” tiveram um grande aumento. Em  $k=4$  a quantidade destas amostras aumentou ainda mais, representando 37% do total. O terceiro grupo foi formado com amostras provenientes dos grupos 1 e 2, enquanto o grupo 4 foi formado



majoritariamente por plantas classificadas como grupo 1. A partir das análises com  $k=3$  os grupos 3 e 4 não apresentaram nenhuma característica morfológica em comum que pudesse os agrupar.

## DISCUSSÃO

De acordo com experiência em campo e coleta dos morfotipos, foi verificado que estes não possuem uma área de distribuição exclusiva, sendo encontrados com frequência bem próximos um ao outro e em florestas não inundáveis. Como o material herborizado raramente possuía alguma informação sobre superfície terrestre em sua etiqueta, o tipo de solo foi verificado na observação e coleta dos morfotipos e foi constatado que as formas estão presentes no mesmo tipo de ambiente, sem qualquer diferença. Desse modo, ainda há a possibilidade de plasticidade fenotípica, pois mesmo as formas de *E. guianensis* sendo encontradas lado a lado e apresentando diferenças morfológicas evidentes, a plasticidade pode ser resultado de outros fatores ambientais como competição, herbivoria e heterofilia (Schlichting 1986).

Dois grupos são reconhecidos para *Ecclinusa guianensis*, conforme sugerido inicialmente por Ribeiro *et al.* (1999) para a RFAD. As amostras classificadas estão bem delimitadas pelos caracteres morfológicos utilizados para reconhecer os dois morfotipos presentes na espécie, visto que as variáveis mais fortemente correlacionadas correspondem à forma da folha, o caráter mais evidente que os diferencia. Já ao testar se essa delimitação por morfologia ainda permanece quando sugeridos três ou quatro grupos, a ordenação das amostras se mostra completamente confusa, não sendo possível encontrar qualquer característica morfológica que sustente os novos grupos formados. Isso reforça a ideia de que *E. guianensis* é formada por duas entidades, como consta nas identificações dos bancos de dados para as plantas coletadas. Poucas amostras divergiram nas classificações *a priori* (morfotipadas sem o auxílio de sua identificação contida na etiqueta) e *a posteriori* (resultado fornecido pelas análises), mostrando que há a presença de indivíduos intermediários entre os morfotipos e que geram dúvidas quanto ao seu grupo pertencente.

Conforme observado em campo, as plantas pertencentes à forma “A” são reconhecidas como árvores de grande porte enquanto as da forma “B” não ultrapassam 15 metros de altura, sendo classificadas como arvoretas. Desse modo, pode-se inferir que as formas representam indivíduos jovens (forma “B”) e adultos (forma “A”) e que essa diferença encontrada nas

folhas dentro da espécie pode representar um caso de heterofilia, uma variação no tamanho e forma das folhas em diferentes estádios de ontogenia. Esta pode estar ligada à exposição a diferentes ambientes durante o desenvolvimento da planta (Winn 1999). Assim, a diferenciação das folhas (e até mesmo da aparência do tronco) de *Ecclinusa guianensis* pode ser justificada pela espécie apresentar uma variação ontogenética durante o desenvolvimento de plântula até a fase adulta ou pela diferença de disponibilidade de luz no ambiente, visto que quando adulta a espécie chega a atingir o dossel e obtém acesso a uma maior quantidade de luz solar. Ademais, foi encontrado material reprodutivo apenas nas coletas pertencentes ao morfotipo “A”, inferindo que os indivíduos pertencentes à forma “B” ainda não atingiram a maturidade reprodutiva. Conseqüentemente, não foi possível observar se há diferenciação nos caracteres reprodutivos, que poderiam acarretar numa distinção de polinizadores e contribuir para o isolamento reprodutivo (Armbruster 2016).

Ao combinar os dados morfológicos com os espectrais a separação da espécie em dois grupos correspondeu exatamente à delimitação resultante da análise de caracteres morfológicos. Com isso, fica evidente que o NIR também sustentou os táxons delimitados pela morfologia e, portanto, se mostra como uma ferramenta capaz de identificar variações intraespecíficas. Os resultados vão de encontro com outros estudos que também utilizaram dados espectrais, em especial a espectroscopia do FT-NIR e FT-MIR (próximo e médio, respectivamente) para discriminar espécies de plantas proximamente relacionadas e simpátricas (Xu *et al.* 2009; Maree e Viljoen 2011).

Estudos recentes desenvolvidos na Amazônia Central mostraram que pode haver distinção da resposta espectral em indivíduos jovens e adultos e estas podem ser reconhecidas com acurácia acima de 80% quando as amostras jovens representam 20% ou mais do total de amostras analisadas (Lang *et al.* 2015). No caso de *E. guianensis* as amostras do morfotipo “B” (os indivíduos que poderiam ser considerados jovens) representam cerca de 34% do total de exsicatas analisadas, o que justificaria a precisão de 100% na classificação de acordo com os agrupamentos sugeridos nas análises morfométricas. Esta clara separação entre jovens e adultos é esperada, visto que folhas são conjuntos complexos de compostos orgânicos que podem assumir diferentes respostas quando submetidas à luz infravermelha, pois sua cutícula possui uma grande variedade de estruturas e organização dos seus componentes (da Luz 2006). Abasolo *et al.* (2013) observaram em indivíduos híbridos classificações menos precisas em plântulas de quatro meses comparadas às de oito meses, podendo haver relação

com a mudança no perfil químico da planta ao longo de seu desenvolvimento ontogenético, mais evidente nos primeiro quatro meses de vida neste caso.

Durante o desenvolvimento de uma planta ocorrem várias mudanças ontogenéticas relacionadas às folhas, como diferenciação da estrutura interna (Isebrands e Larson 1973), alteração da morfologia (Thomas e Ickes 1995) e composição química (Ishida *et al.* 2005; England e Attiwill 2006), modificação na capacidade fotossintética e uso da água (Cavender-Bares e Bazzaz 2000; Mediavilla e Escudero 2004) e distribuição de biomassa e equilíbrio de taxas de carbono (Lusk 2004). Além dessas mudanças relacionadas ao crescimento das plantas, há também a infecção de fungos endofíticos que crescem dentro dos tecidos foliares, capazes até mesmo de interferir na síntese de alcaloides das plantas infectadas (Uchitel *et al.* 2011). Há relatos de insetos herbívoros que alteram a fisiologia das plantas hospedeiras para seu próprio benefício, onde sintetizam moléculas bioativas (Frago *et al.* 2012). Dessa forma, essa mudança na estrutura e composição química na folha, tanto por fatores ontogenéticos, quanto pela alteração provocada pelo parasita ou a instalação destes nos tecidos vegetais, pode ser capaz de revelar assinaturas espectrais diferentes para as duas formas, pois todo tipo de material orgânico é detectado pela técnica e talvez os seres simbiotes estejam presentes apenas na fase jovem da espécie em questão.

As evidências morfológicas e espectrais indicaram a presença de dois grupos em *Ecclinusa guianensis*. Contudo, ainda não é possível reconhecer as formas “A” e “B” como diferentes espécies. Primeiramente, grande volume do material depositado em herbário não contém informação sobre altura de indivíduo coletado e, como observado em campo, o morfotipo “B” não ultrapassa quinze metros de altura. Por ser encontrado com frequência ao lado da forma “A” deduz-se que se tratam de indivíduos jovens. Em segundo lugar, por mais que os dados espectrais também tenham corroborado a sustentação morfológica, é esperada uma diferença espectral entre indivíduos jovens e adultos, e essa pode ser detectada com precisão nas análises.

O reconhecimento de mais táxons em simpatria dentro de uma mesma espécie ocorre com frequência na região da Amazônia Central, onde ao expandir a amostragem por toda a área de distribuição da espécie em questão as diferenças morfológicas antes observadas não se mostram muito claras. Como exemplo temos os trabalhos de Barbosa (2015) e Rivera-Parada (2017), que identificaram complexos dentro de Lauraceae, uma família que também possui a taxonomia complicada. Prata (2016) identificou na região um centro de diversidade de

*Pagamea guianensis* (Rubiaceae) que apresentava um complexo de espécies crípticas, como Esteves e Vicentini (2013), que delimitaram *Pagamea coriacea* em Presidente Figueiredo. Em relação à Sapotaceae, Terra-Araujo *et al.* (2012) verificou a presença de três entidades em *Micropholis guyanensis*, um dos 23 complexos de espécies ainda existentes apenas dentro das áreas do PDBFF.

A identificação de duas entidades apenas na região da Amazônia Central e não em toda área de distribuição de *Ecclinusa guianensis* além da presença de vários complexos de espécies pode representar uma confluência de linhagens nesse local, causando essa grande variação. O fato das formas ocorrer apenas ao redor de Manaus (Reserva Ducke e áreas do PDBFF) também evidencia que ambas as áreas são bem amostradas e que por apresentarem uma grande quantidade de material foi detectada a presença desse indivíduo jovem pela alta frequência de coletas.

## CONCLUSÃO

Os dados morfológicos e espectrais permitem o reconhecimento de dois morfotipos em *Ecclinusa guianensis*, conforme sugerido previamente por Ribeiro *et al.* (1999). Entretanto, a ausência de flores/frutos e o porte reduzido (arvoretas < 15 m) em indivíduos da forma “B” sustenta a hipótese que esta forma representa jovens da forma “A”, e que não atingiram a maturidade reprodutiva. O fato de serem simpátricas nas parcelas permanentes do PDBFF e RFAD indica que as sementes não foram dispersas por longas distâncias da planta mãe.

Os resultados encontrados neste estudo indicam que a divisão de *Ecclinusa guianensis* em duas espécies é precipitado, pois o morfotipo “B” pode representar indivíduos que correspondem à fase jovem da espécie. Os dados NIR vem sendo uma eficaz ferramenta para a distinção de espécies proximamente relacionadas na região amazônica, sendo capaz de reconhecer as variações dentro de espécies e de auxiliar no entendimento da natureza de complexos taxonômicos.

## BIBLIOGRAFIA CITADA

- Abasolo, M.; Lee, D.J.; Raymond, C.; Meder, R.; Shepherd, M. 2013. Deviant near infrared spectra identifies *Corymbia* hybrids. *Forest ecology and management*, 304: 121–31.
- Anderberg, A.A.; Rydin, C.; Källersjö, M. 2002. Phylogenetic relationships in the order Ericales s.l.: analyses of molecular data from five genes from the plastid and mitochondrial genomes. *American Journal of Botany*, 89: 677–87.
- Anderberg, A.A.; Swenson, U. 2003. Evolutionary lineages in Sapotaceae (Ericales): a cladistic analysis based on *ndh F* sequence data. *International Journal of Plant Sciences*, 164: 763–73.
- Antonelli, A.; Sanmartin, I. 2011. Why are there so many plant species in the Neotropics? *Taxon*, 60: 403–14.
- Armbruster, W.S. 2016. Plant–pollinator interactions and flower diversification. In: *Encyclopedia of evolutionary biology*. Academic Press Inc.
- Aubréville, A. 1964. Les Sapotace es taxonomie et phytogéographie. *Adansonia* 1: 1–157.
- Baehni, C. 1965. Mémoires sur les Sapotacées. 3. Inventaire des genres. *Boissiera* 11: 1–262.
- Balakrishnama, S.; Ganapathiraju, A. 1998. Linear discriminant analysis-a brief tutorial. *Institute for Signal and information Processing*, 9p.
- Barbosa, A. S. da M. 2015. Testando delimitação de espécies em monografias: um estudo de caso em *Aniba Aubl.* (Lauraceae). Dissertação de Mestrado, Instituto Nacional de Pesquisas da Amazônia, Manaus, Amazonas. 90p.
- Bartish, I. V.; Antonelli, A.; Richardson, J.E.; Swenson, U. 2011. Vicariance or long-distance dispersal: historical biogeography of the pantropical subfamily Chrysophylloideae (Sapotaceae). *Journal of Biogeography*, 38: 177–90.
- Bartish, I. V.; Swenson, U.; Munzinger, J.; Anderberg, A.A. 2005. Phylogenetic relationships among New Caledonian Sapotaceae (Ericales): molecular evidence for generic polyphyly and repeated dispersal. *American Journal of Botany*, 92: 667–73.
- Burnham, R.J.; Graham, A. 1999. The history of neotropical vegetation: new developments and status. *Annals of the Missouri Botanical Garden*, 546–89.
- Camargo, J. L.C.; Ferraz, I. D. K.; Mesquita, M. R.; Santos, B. A.; Brum, H. D. 2008. Guia de propágulos e plântulas da Amazônia. Instituto Nacional de Pesquisas da Amazônia, Manaus, Amazonas, 168 p.
- Carlin, B.P.; Louis, T.A. 2000. *Bayes and empirical Bayes methods for data analysis*.

Chapman & Hall/CRC Boca Raton, FL.

Cavender-Bares, J.; Bazzaz, F.A. 2000. Changes in drought response strategies with ontogeny in *Quercus rubra*: implications for scaling from seedlings to mature trees. *Oecologia*, 124: 8–18.

Curty, C. 2013. A espectroscopia no infravermelho próximo de folhas para identificação de plantas na amazônia. Dissertação de mestrado, Instituto Nacional de Pesquisas da Amazônia, Manaus, Amazonas. 51p.

Denslow, J.S. 1987. Tropical rainforest gaps and tree species diversity. *Annual review of ecology and systematics*, 18: 431–51.

de Queiroz, K. 1998. The General Lineage Concept of Species, Species Criteria, and the Process of Speciation A Conceptual Unification and Terminological Recommendations.

de Queiroz, K. 2007. Species concepts and species delimitation. *Systematic biology*, 56: 879–86.

Durgante, F.M.; Higuchi, N.; Almeida, A.; Vicentini, A. 2013. Species spectral signature: discriminating closely related plant species in the Amazon with near-infrared leaf-spectroscopy. *Forest Ecology and Management*, 291: 240–48.

England, J.R.; Attiwill, P.M. 2006. Changes in leaf morphology and anatomy with tree age and height in the broadleaved evergreen species, *Eucalyptus regnans* F. Muell. *Trees*, 20: 79.

Esteves, S. de M.; Vicentini, A. 2013. Cryptic species in *Pagamea coriacea* sensu lato (Rubiaceae): evidence from morphology, ecology and reproductive behavior in a sympatric context. *Acta Amazonica*, 43: 415–28.

Faria, A. D., J. R. Pirani, J. E. L. S. Ribeiro, S. Nylander, M. H. Terra-Araujo, P. P. Vieira, and U. Swenson. 2017. Towards a natural classification of subfamily Chrysophylloideae (Sapotaceae) in the Neotropics. *Botanical Journal of Linnean Society*.

Fiala, B.; Slik, F.; Weising, K.; Maschwitz, U.; Mohamed, M.; Guicking, D.; et al. 2016. Phylogeography of three closely related myrmecophytic pioneer tree species in SE Asia: implications for species delimitation. *Organisms Diversity & Evolution*, 16: 39–52.

Frago, E.; Dicke, M.; Godfray, H.C.J. 2012. Insect symbionts as hidden players in insect-plant interactions. *Trends in Ecology & Evolution*, 27: 705–11.

Gomes, A.; Andrade, A.; Barreto-Silva, J.S.; Brenes-Arguedas, T.; López, D.C.; Freitas, C.C.; et al. 2013. Local plant species delimitation in a highly diverse Amazonian forest: do we all

see the same species? *Journal of Vegetation Science*, 24: 70–79.

Gower, J.C. 1971. A general coefficient of similarity and some of its properties. *Biometrics*, 857–71.

Hernández-Ruedas, M.A.; Arroyo-Rodríguez, V.; Meave, J.A.; Martínez-Ramos, M.; Ibarra-Manríquez, G.; Martínez, E.; et al. 2014. Conserving tropical tree diversity and forest structure: the value of small rainforest patches in moderately-managed landscapes. *PloS one*, 9: e98931.

Hopkins, M.J.G. 2007. Modelling the known and unknown plant biodiversity of the Amazon Basin. *Journal of Biogeography*, 34: 1400–1411.

Isebrands, J.G.; Larson, P.R. 1973. Anatomical changes during leaf ontogeny in *Populus deltoides*. *American Journal of Botany*, 199–208.

Ishida, A.; Yazaki, K.; Hoe, A.L. 2005. Ontogenetic transition of leaf physiology and anatomy from seedlings to mature trees of a rain forest pioneer tree, *Macaranga gigantea*. *Tree physiology*, 25: 513–22.

IUCN - International Union for Conservation of Nature, 2015, (<http://www.iucnredlist.org/>). Acesso em 10/07/2015.

Jin, X.; Chen, X.; Xiao, L.; Shi, C.; Chen, L.; Yu, B.; et al. 2017. Application of visible and near-infrared spectroscopy to classification of *Miscanthus* species. *PloS one*, 12: e0171360.

Kaye, W. 1954. Near-infrared spectroscopy: I. Spectral identification and analytical applications. *Spectrochimica Acta*, 6: 257E1--287E2.

Kim, S.W.; Ban, S.H.; Chung, H.; Cho, S.; Chung, H.J.; Choi, P.S.; et al. 2004. Taxonomic discrimination of flowering plants by multivariate analysis of Fourier transform infrared spectroscopy data. *Plant cell reports*, 23: 246–50.

Kruskal, J.B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29: 1–27.

Lam, H. J. 1939 On the system of the Sapotaceae, with some remarks on taxonomical methods. *Recueil Trav Bot Neerl* 36:509–525.

Lang, C.; Costa, F.R.C.; Camargo, J.L.C.; Durgante, F.M.; Vicentini, A. 2015. Near infrared spectroscopy facilitates rapid identification of both young and mature Amazonian tree species. *PloS one*, 10: e0134521.

Li, B.; Wei, Y.; Duan, H.; Xi, L.; Wu, X. 2012. Discrimination of the geographical origin of

- Codonopsis pilosula using near infrared diffuse reflection spectroscopy coupled with random forests and k-nearest neighbor methods. *Vibrational Spectroscopy*, 62: 17–22.
- Lusk, C.H. 2004. Leaf area and growth of juvenile temperate evergreens in low light: species of contrasting shade tolerance change rank during ontogeny. *Functional Ecology*, 18: 820–28.
- da Luz, B. 2006. Attenuated total reflectance spectroscopy of plant leaves: a tool for ecological and botanical studies. *New Phytologist*, 172: 305–18.
- Maree, J.E.; Viljoen, A.M. 2011. Fourier transform near-and mid-infrared spectroscopy can distinguish between the commercially important *Pelargonium sidoides* and its close taxonomic ally *P. reniforme*. *Vibrational Spectroscopy*, 55: 146–52.
- Mediavilla, S.; Escudero, A. 2004. Stomatal responses to drought of mature trees and seedlings of two co-occurring Mediterranean oaks. *Forest Ecology and Management*, 187: 281–94.
- Medrano, M.; López-Perea, E.; Herrera, C.M. 2014. Population genetics methods applied to a species delimitation problem: endemic trumpet daffodils (*Narcissus* section *Pseudonarcissi*) from the Southern Iberian Peninsula. *International Journal of Plant Sciences*, 175: 501–17.
- Mikkelsen, K.S.; Seberg, O. 2001. Morphometric analysis of the *Bersama abyssinica* Fresen. complex (Melianthaceae) in East Africa. *Plant Systematics and Evolution*, 227: 157–82.
- Milliken, W. 1998. Structure and composition of one hectare of central Amazonian terra firme forest. *Biotropica*, 30: 530–37.
- Moeller, M.; Gao, L.-M.; Mill, R.R.; Li, D.-Z.; Hollingsworth, M.L.; Gibby, M. 2007. Morphometric analysis of the *Taxus wallichiana* complex (Taxaceae) based on herbarium material. *Botanical Journal of the Linnean Society*, 155: 307–35.
- Oliveira, A.A.; Nelson, B.W. 2001. Floristic relationships of terra firme forests in the Brazilian Amazon. *Forest Ecology and Management*, 146: 169–79.
- Pasquini, C. 2003. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, 14: 198–219.
- PDBFF - Projeto Dinâmica Biológica de Fragmentos Florestais, 2017. Base de dados online - Sapotaceae. Disponível em:  
[http://www.botanicaamazonica.wiki.br/plantasPDBFF/checklistview\\_tabber.php](http://www.botanicaamazonica.wiki.br/plantasPDBFF/checklistview_tabber.php). Acesso dez/2017.
- Pennington, T.D. 1990. *Flora neotropica. Monograph 52. Sapotaceae*. New York Botanical



Garden for the Organization for Flora Neotropica.

Pennington, T.D. 1991. The genera of Sapotaceae. *Royal Botanic Gardens: Bronx, New York.: Kew & New York Botanical Garden*, 1–13.

Pennington, T.D. 2004. Sapotaceae. In: *Flowering Plants{·} Dicotyledons*. Springer.

Ponzoni, F.J. 2002. Sensoriamento remoto no estudo da vegetação: diagnosticando a mata atlântica. *São José dos Campos, Brasil: INPE*,.

Prance, G.T. 2001. Discovering the plant world. *Taxon*, 345–59.

Prata, E. M. B. 2016. Delimitação de espécies e história de diversificação do complexo *Pagamea Guianensis* (Rubiaceae) na América do Sul Tropical. Tese de Doutorado, Instituto Nacional de Pesquisas da Amazônia, Manaus, Amazonas. 143p.

Reflora, 2015. *Ecclinusa* (<http://floradobrasil.jbrj.gov.br/reflora/listaBrasil>). Acesso em 07/07/2015.

Rellstab, C.; Bühler, A.; Graf, R.; Folly, C.; Gugerli, F. 2016. Using joint multivariate analyses of leaf morphology and molecular-genetic markers for taxon identification in three hybridizing European white oak species (*Quercus* spp.). *Annals of forest science*, 73: 669–79.

Ribeiro, J.E.L.S.; Hopkins, M.J.G.; Vicentini, A.; Sothers, C.A.; Costa, M.A.S.; Brito, J.M.; et al. 1999. *Flora Da Reserva Ducke: Guia de Identificação Das Plantas Vasculares de Uma Floresta de Terra-Firme Na Amazônia Central*. Instituto Nacional de Pesquisas da Amazônia, Manaus, Amazonas. 816p.

Rivera-Parada, L. L. 2017. Delimitação das espécies do complexo *Ocotea canaliculata* (Lauraceae). Dissertação de Mestrado, Instituto Nacional de Pesquisas da Amazônia, Manaus, Amazonas. 144p.

Rodriguez Fernández, J.I. 2008. Coenosiini neotropicais (Muscidae: Diptera): taxonomia, filogenia e contribuição da espectroscopia infravermelha.

Schlichting, C.D. 1986. The evolution of phenotypic plasticity in plants. *Annual review of ecology and systematics*, 17: 667–93.

Sites Jr, J.W.; Marshall, J.C. 2004. Operational criteria for delimiting species. *Annu. Rev. Ecol. Evol. Syst.*, 35: 199–227.

Smedmark, J.E.E.; Swenson, U.; Anderberg, A.A. 2006. Accounting for variation of substitution rates through time in Bayesian phylogeny reconstruction of Sapotoideae

(Sapotaceae). *Molecular Phylogenetics and Evolution*, 39: 706–21.

SpeciesLink, 2015, (<http://splink.cria.org.br/>). Acesso em 05/07/2015.

Stevens, P.F.; Davis, H. 2001. *Angiosperm phylogeny website*. Missouri Botanical Garden St Louis, MO, USA.

Strgulc Krajšek, S.; Buh, P.; Zega, A.; Kreft, S. 2008. Identification of herbarium whole-leaf samples of *Epilobium* species by ATR-IR spectroscopy. *Chemistry & biodiversity*, 5: 310–17.

Sultan, S.E. 2000. Phenotypic plasticity for plant development, function and life history. *Trends in plant science*, 5: 537–42.

Swenson, U.; Anderberg, A.A. 2005. Phylogeny, character evolution, and classification of Sapotaceae (Ericales). *Cladistics*, 21: 101–30.

Swenson, U.; Lowry, P.P.; Munzinger, J.; Rydin, C.; Bartish, I. V. 2008a. Phylogeny and generic limits in the *Niemeyera* complex of New Caledonian Sapotaceae: evidence of multiple origins of the anisomerous flower. *Molecular Phylogenetics and Evolution*, 49: 909–29.

Swenson, U.; Munzinger, J.; Bartish, I. V. 2007. Molecular phylogeny of *Planchonella* (Sapotaceae) and eight new species from New Caledonia. *Taxon*, 56: 329–54.

Swenson, U.; Nylinder, S.; Munzinger, J. 2014. Sapotaceae biogeography supports New Caledonia being an old Darwinian island. *Journal of Biogeography*, 41: 797–809.

Swenson, U.; Richardson, J.E.; Bartish, I. V. 2008b. Multi-gene phylogeny of the pantropical subfamily Chrysophylloideae (Sapotaceae): evidence of generic polyphyly and extensive morphological homoplasy. *Cladistics*, 24: 1006–31.

Terra-Araujo, M.H., Faria, A.D., Ribeiro, J.E.L.S., Swenson, U., 2012. Flower biology and subspecies concepts in *Micropholis guyanensis* (Sapotaceae): evidence of ephemeral flowers in the family. *Aust. Syst. Bot.* 25, 295–303.

Terra-Araujo, M. H. 2013. Phylogeny, biogeography and taxonomy of *Pradosia* (Sapotaceae, Chrysophylloideae). Tese de Doutorado. Manaus, Amazonas: Instituto Nacional de Pesquisas da Amazônia, 138p.

Terra-Araujo, M. H., A. D. Faria, A. Vicentini, S. Nylinder, and U. Swenson. 2015. Species tree phylogeny and biogeography of the Neotropical genus *Pradosia* (Sapotaceae, Chrysophylloideae). *Molecular Phylogenetics and Evolution* 87: 1–13.

Terra-Araujo, M.H.; Costa, F.M.; De Carvalho, R.B.; Vicentini, A. 2015. *Ecclinusa campinae*

- (Sapotaceae), a new species from the Middle Rio Negro region, Amazonas, Brazil. *Brittonia*, 67: 180–84.
- ter Steege, H.; Pitman, N.C.A.; Sabatier, D.; Baraloto, C.; Salomão, R.P.; Guevara, J.E.; et al. 2013. Hyperdominance in the Amazonian tree flora. *Science*, 342: 1243092.
- ter Steege, H.; Sabatier, D.; Castellanos, H.; Van Andel, T.; Duivenvoorden, J.; de Oliveira, A.A.; et al. 2000. An analysis of the floristic composition and diversity of Amazonian forests including those of the Guiana Shield. *Journal of tropical ecology*, 16: 801–28.
- Team, R.C. 2012. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Thomas, S.C.; Ickes, K. 1995. Ontogenetic changes in leaf size in Malaysian rain forest trees. *Biotropica*, 427–34.
- Uchitel, A.; Omacini, M.; Chaneton, E.J. 2011. Inherited fungal symbionts enhance establishment of an invasive annual grass across successional habitats. *Oecologia*, 165: 465–75.
- Vapnik, V. 2013. *The nature of statistical learning theory*. Springer science & business media.
- Winn, A.A. 1999. The functional significance and fitness consequences of heterophylly. *International Journal of Plant Sciences*, 160: S113--S121.
- Workman Jr, J.; Weyer, L. 2007. *Practical guide to interpretive near-infrared spectroscopy*. CRC press.
- Xu, H.; Yu, P.; Fu, X.; Ying, Y. 2009. On-site variety discrimination of tomato plant using visible-near infrared reflectance spectroscopy. *Journal of Zhejiang University-Science B*, 10: 126–32.

## LISTA DE FIGURAS

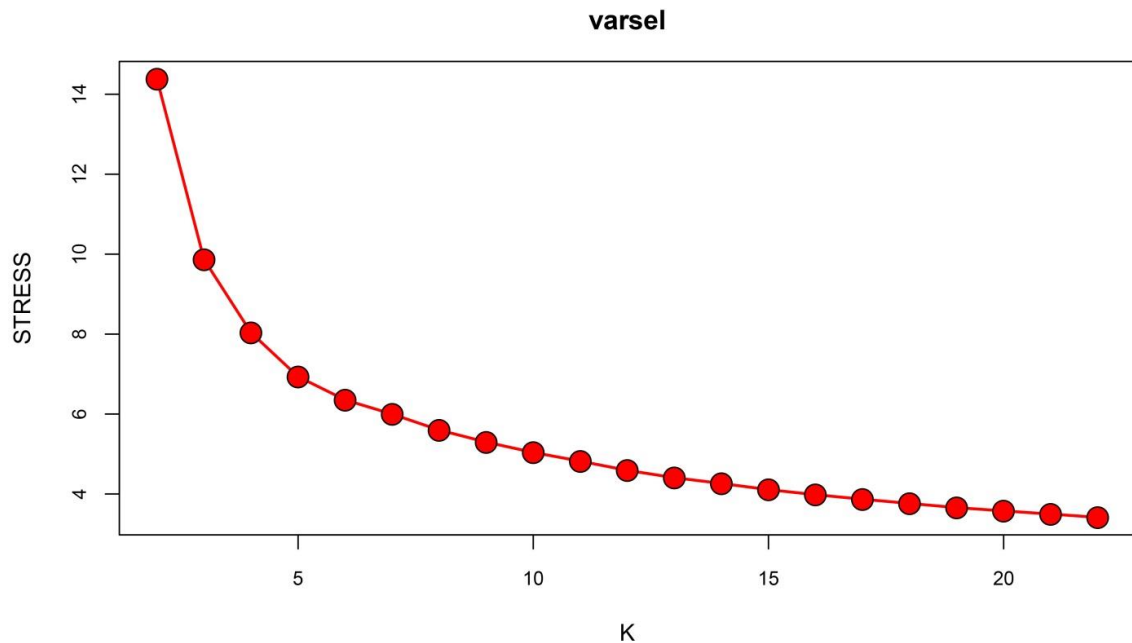


Figura 1. Comparação entre o stress causado e a quantidade de eixos (dimensões) necessários (K) para a NMDS de acordo com análises das variáveis morfológicas de *Ecclinusa guianensis* Eyma (Sapotaceae, Chrysophylloideae) selecionadas (varsel) e contidas na Tabela 1.

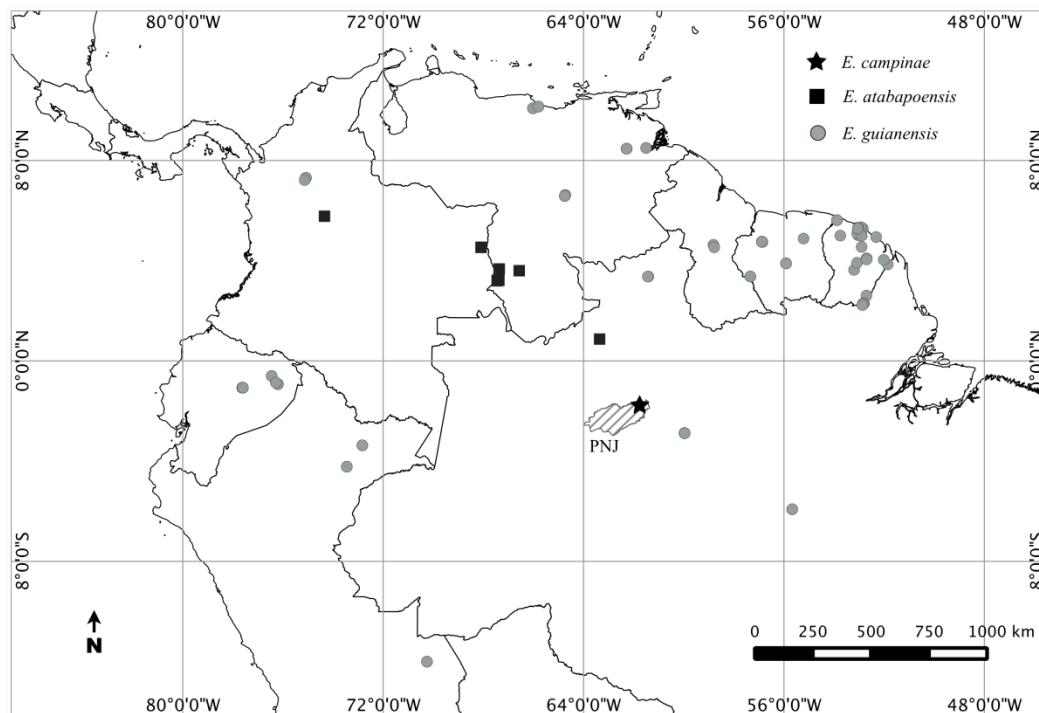


Figura 2. Distribuição geográfica conhecida para *Ecclinusa guianensis* Eyma e espécies proximoamente relacionadas filogeneticamente, *E. atabapoensis* (Aubrév.) T. D. Penn e *E. campinae* Terra-Araujo (Sapotaceae, Chrysophylloideae). O Parque Nacional do Jaú na região

do Médio Rio Negro é indicado por "PNJ".

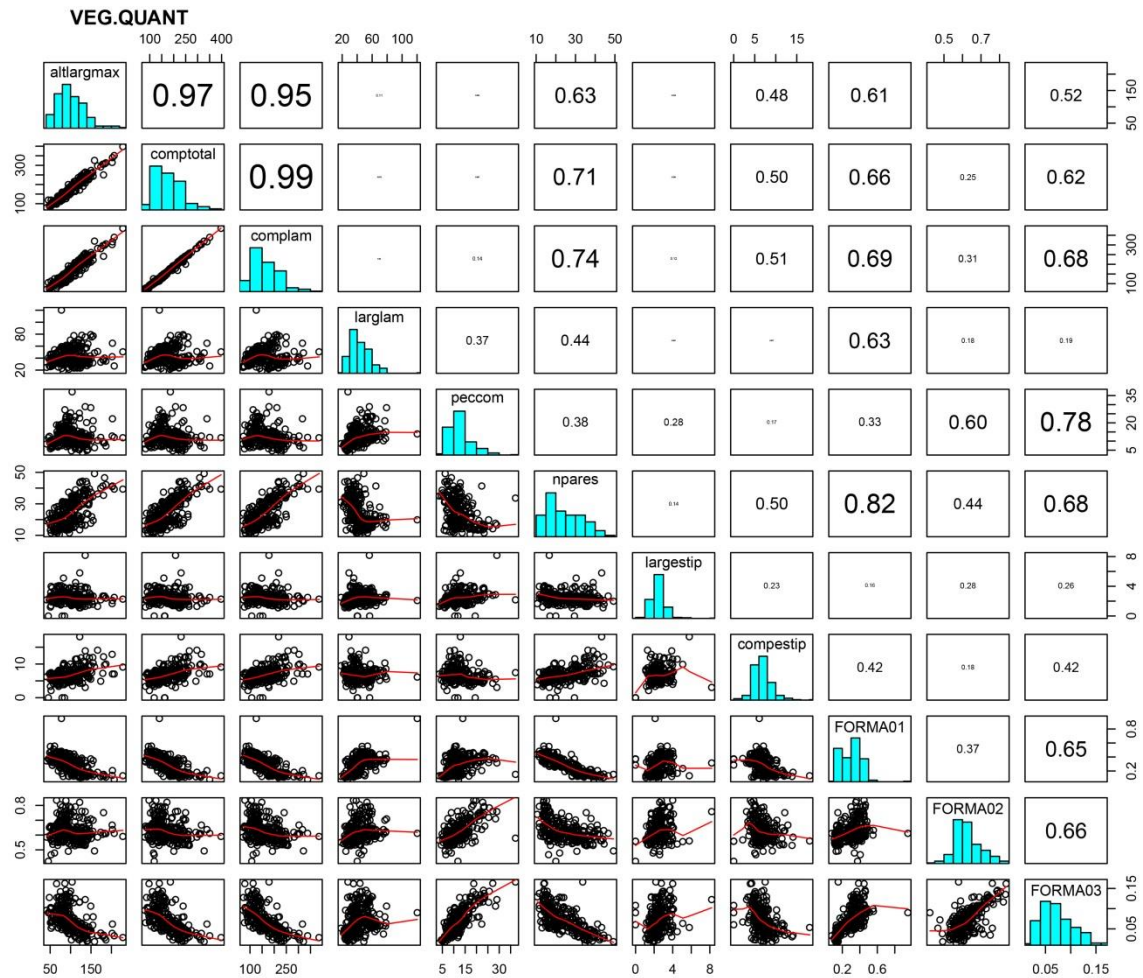


Figura 3. Correlação entre as variáveis morfológicas vegetativas de *Eclinusa guianensis* Eyma (Sapotaceae, Chrysophylloideae) mensuradas e descritas na Tabela 1. Os números contidos dentro dos quadrados indicam o grau de correlação existente entre as variáveis comparadas.

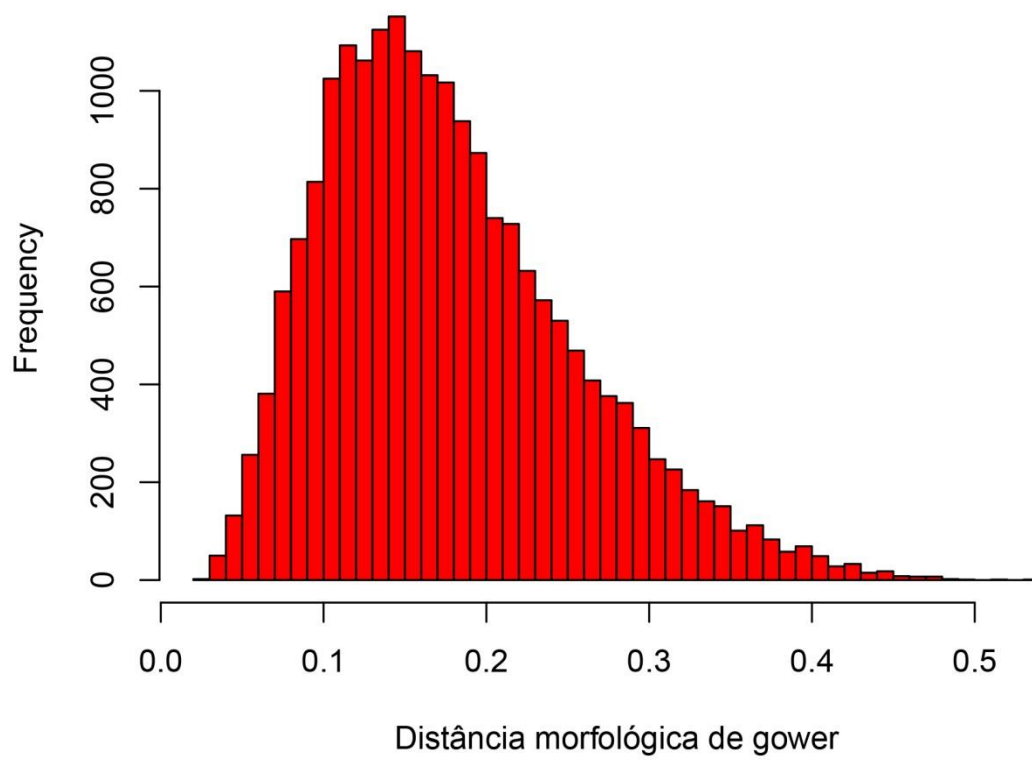


Figura 4. Distância morfológica de Gower para as variáveis morfológicas de *Ecclinusa guianensis* Eyma (Sapotaceae, Chrysophylloideae) mensuradas e contidas na Tabela 1.

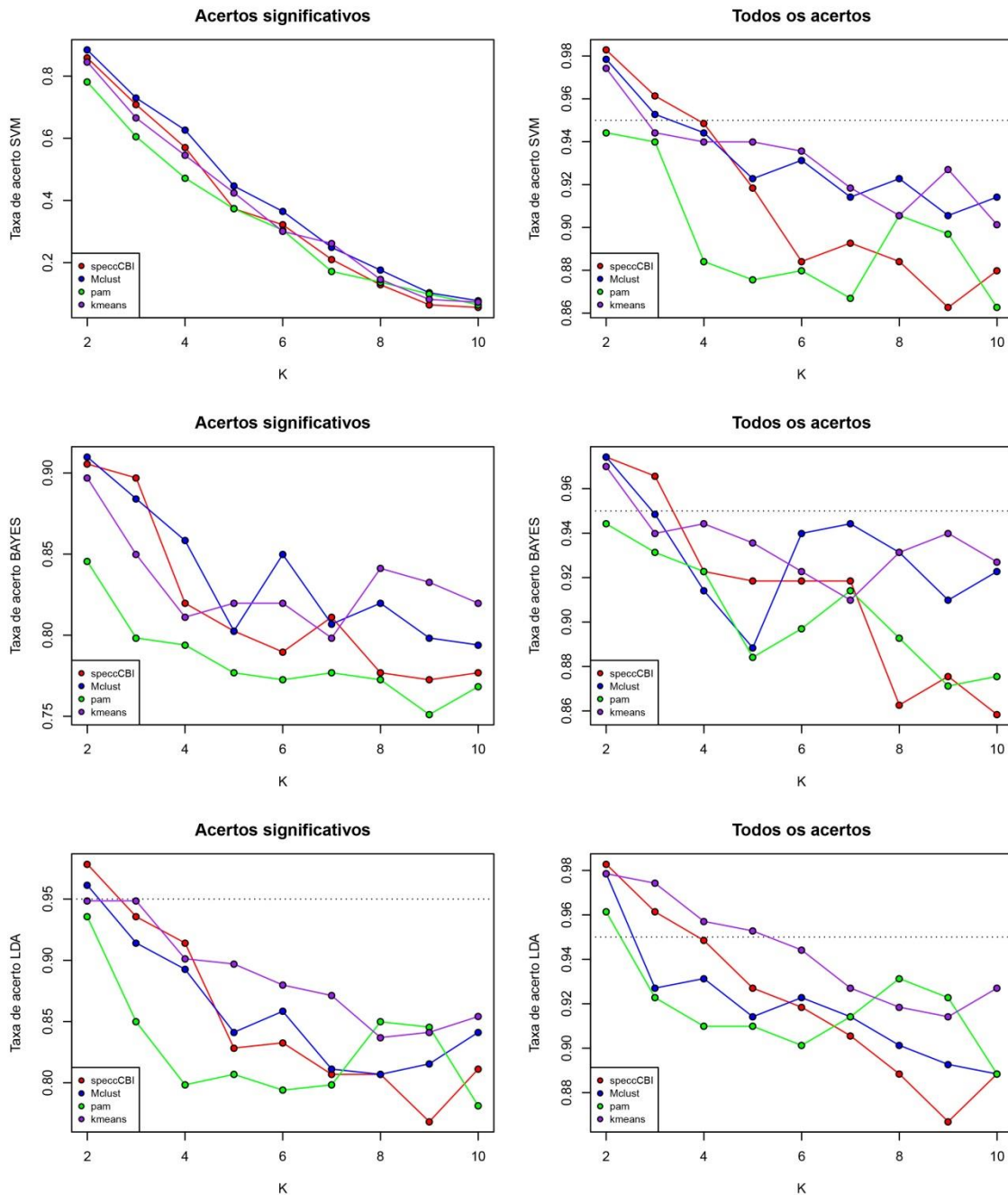


Figura 5. Análises de agrupamento para dados morfológicos de *Ecclinusa guianensis* Eyma (Sapotaceae, Chrysophylloideae) mensurados e contidos na Tabela 1. As análises foram submetidas a três modelos (SMV, Bayes e LDA), contendo todas as amostras analisadas (todos os acertos) e considerando apenas aquelas classificadas como "core" (acertos significativos).

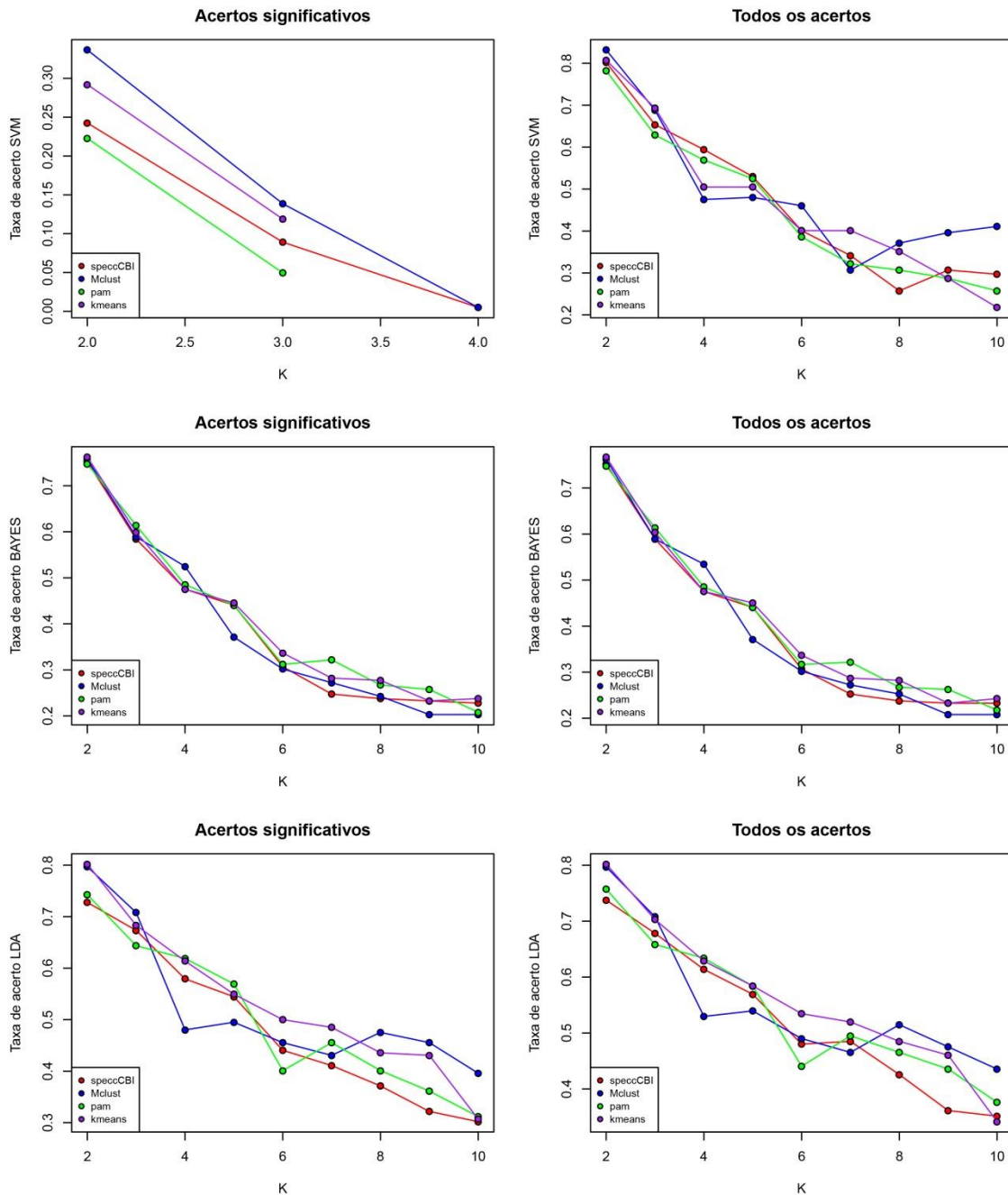


Figura 6. Análises combinadas de agrupamento para dados morfológicos (mensurados e contidos na Tabela 1) e espectrais para *Ecclinusa guianensis* Eyma (Sapotaceae, Chrysophylloideae). As análises foram submetidas a três modelos (SMV, Bayes e LDA), contendo todas as amostras analisadas (todos os acertos) e considerando apenas aquelas classificadas como "core" (acertos significativos).



K.2\_MELHOR.PREDICAO-K.3\_MELHOR.PREDICAO CORE

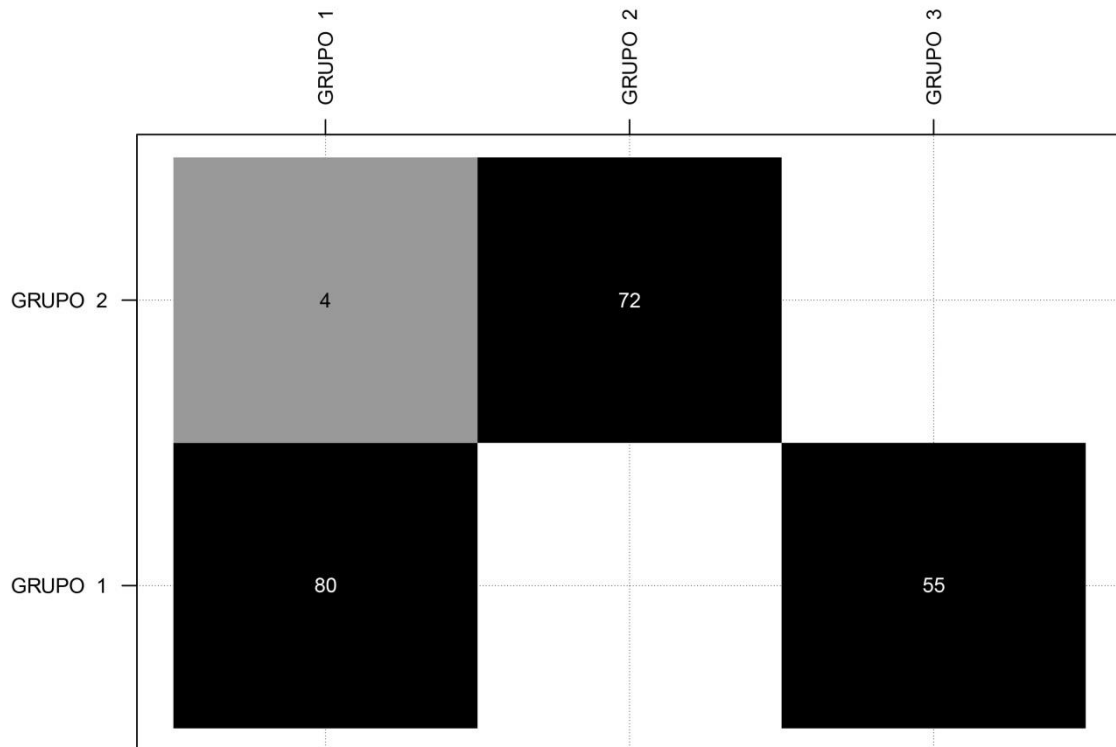


Figura 7. Matriz de confusão gerada após a análise de agrupamento com caracteres morfológicos de *Ecclinusa guianensis* Eyma (Sapotaceae, Chrysophylloideae) mensurados e contidos na Tabela 1. São comparadas as classificações das amostras “core” da análise gerada com dois e três grupos ( $k=2$  e  $k=3$ ), onde quanto mais escura a cor do quadrado maior a quantidade de amostras classificadas naquele grupo.

K.2\_MELHOR.PREDICAO-K.3\_MELHOR.PREDICAO CORE

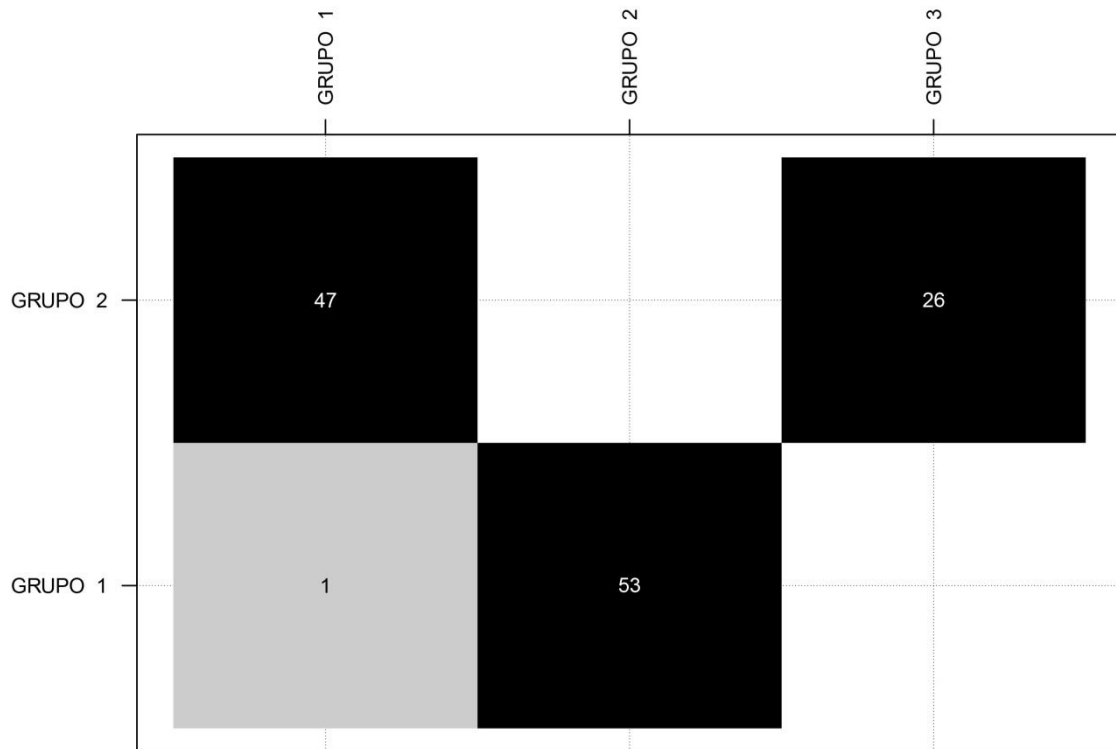


Figura 8. Matriz de confusão gerada após a análise de sensibilidade com caracteres espectrais e morfológicos mensurados (Tabela 1) de *Ecclinusa guianensis* Eyma (Sapotaceae, Chrysophylloideae). São comparadas as classificações das amostras “core” da análise gerada com dois e três grupos ( $k=2$  e  $k=3$ ), onde quanto mais escura a cor do quadrado maior a quantidade de amostras classificadas naquele grupo.

## TABELAS

Tabela 1. Variáveis morfológicas vegetativas de *Ecclinusa guianensis* Eyma (Sapotaceae, Chrysophylloideae) mensuradas.

<b>Variável morfológica</b>	<b>Descrição</b>	<b>Unidade</b>
Altura da largura máxima	Medida da base da folha até o ponto de largura máxima	mm
Comprimento total	Medida total da folha, incluindo o pecíolo	mm
Comprimento da lâmina	Medida da lâmina foliar, sem o pecíolo	mm
Largura da lâmina	Medida da região mais larga da folha	mm
Comprimento do pecíolo	Medida do pecíolo	mm
Número de pares de veias secundárias	Quantidade de veias ao longo da folha	
Largura da estípula	Medida da parte mais larga da estípula	mm
Comprimento da estípula	Média dos tamanhos das estípulas presentes	mm
Forma da folha	Obovada/elíptica/lanceolada/oblanceolada/ovada	-
Forma do ápice foliar	Acuminado/agudo/arredondado	-
Forma da base foliar	Acuminado/agudo/arredondado	-
Nervura central adaxial/abaxial	Proeminente/impressa	-
Nervura secundária adaxial/abaxial	Proeminente/impressa	-
Venação	Broquidódroma/eucamptódroma/eucampbroquidódroma	-
Veias secundárias	Oblíquas/paralelas	-
Veias intersecundárias	Presente/ausente; pequena/média/longa	-
Veias terciárias	Oblíquas/areoladas	-
Formato da estípula	Cônica/triangular	-
Indumento da estípula/pecíolo/face adaxial/face abaxial	Presente/ausente; simples ereto/simples adpresso; diminuto/pequeno/longo; glabro/glabrescente/esparso/denso/recobrimdo a superfície	-

Tabela 2. Características ecológicas de *Ecclinusa guianensis* Eyma (Sapotaceae, Chrysophylloideae) mensuradas.

<b>Variável morfológica</b>	<b>Descrição</b>	<b>Unidade</b>
Ritidoma	Liso/rugoso	-
Altura do indivíduo	Altura estimada da raiz à copa do indivíduo	m
Solo	Argiloso/arenoso/areno argiloso	-

## ANEXO I

Amostras de *Ecclinusa guianensis* Eyma (Sapotaceae, Chrysophylloideae) identificadas a priori e utilizados para a coleta de dados morfológicos, espectrais, ecológicos e geográficos. Informações disponíveis no repositório [botanicaamazonica.wiki.br/plantasPDBFF](http://botanicaamazonica.wiki.br/plantasPDBFF) (Plantas PDBFF 2015).

N.	Morfo.	wiki_id	Coleção	N.	Morfo.	wiki_id	Coleção
1	Forma A	14616	MO	116	Forma B	38324	PDBFF
2	Forma A	14191	RB	117	Forma B	155804	PDBFF
3	Forma A	14653	RB	118	Forma B	170648	PDBFF
4	Forma B	177459	PDBFF	119	Forma A	280021	PPBio
5	Forma B	200868	PDBFF	120	Forma A	280027	PPBio
6	Forma A	14528	RB	121	Forma A	280038	PPBio
7	Forma B	197846	PDBFF	122	Forma A	280046	PPBio
8	Forma B	122243	PDBFF	123	Forma A	14535	INPA
9	Forma B	5991	PDBFF	124	Forma A	14565	INPA
11	Forma A	14345	MO	125	Forma A	14269	INPA
12	Forma A	5339	PDBFF	126	Forma A	6862	PDBFF
13	Forma B	120102	PDBFF	127	Forma A	211366	PDBFF
14	Forma A	14347	MO	128	Forma B	191361	PDBFF
15	Forma A	126296	PDBFF	129	Forma B	119440	PDBFF
16	Forma B	152646	PDBFF	130	Forma B	148037	PDBFF
17	Forma B	144987	PDBFF	131	Forma A	280017	PPBio
18	Forma A	14212	INPA	132	Forma A	280025	PPBio
19	Forma B	48242	PDBFF	133	Forma A	280032	PPBio
20	Forma B	118477	PDBFF	134	Forma A	14226	INPA
21	Forma B	164287	PDBFF	135	Forma A	14445	INPA
22	Forma B	175002	PDBFF	136	Forma A	14201	INPA
23	Forma B	198318	PDBFF	137	Forma A	14200	INPA
24	Forma A	14536	INPA	138	Forma A	40439	PDBFF
25	Forma B	206646	PDBFF	139	Forma A	187502	PDBFF
26	Forma B	170536	PDBFF	140	Forma B	173177	PDBFF
27	Forma B	155317	PDBFF	141	Forma B	123370	PDBFF
28	Forma A	280012	PPBio	142	Forma B	158372	PDBFF
29	Forma A	280028	PPBio	143	Forma B	169286	PDBFF
30	Forma A	280037	PPBio	144	Forma A	280018	PPBio
31	Forma A	4434	INPA	145	Forma A	280019	PPBio
32	Forma A	137211	PDBFF	146	Forma A	280026	PPBio
33	Forma A	3197	PDBFF	147	Forma A	14187	INPA
34	Forma A	192742	PDBFF	148	Forma A	14046	INPA
35	Forma A	31826	PDBFF	149	Forma A	14264	INPA
36	Forma A	5938	PDBFF	150	Forma A	132900	PDBFF
37	Forma A	171044	PDBFF	151	Forma A	125538	PDBFF
38	Forma B	148637	PDBFF	152	Forma A	166739	PDBFF
39	Forma A	14052	INPA	153	Forma A	199690	PDBFF
40	Forma A	14255	INPA	154	Forma B	18094	PDBFF

<i>Anexo I</i>				cont.			
N	Morfo	wiki_id	Herbário	N	Morfo	wiki_id	Coleção
41	Forma A	14502	MO	155	Forma A	280023	PPBio
42	Forma A	187661	PDBFF	156	Forma A	280035	PPBio
43	Forma A	174649	PDBFF	157	Forma A	280041	PPBio
44	Forma A	124540	PDBFF	158	Forma A	14196	INPA
45	Forma A	200150	PDBFF	159	Forma A	14517	INPA
46	Forma B	2772	PDBFF	160	Forma A	14203	INPA
47	Forma B	142551	PDBFF	161	Forma A	14566	INPA
48	Forma B	157606	PDBFF	162	Forma A	14199	INPA
49	Forma A	146403	PDBFF	163	Forma A	14110	INPA
50	Forma A	170696	PDBFF	164	Forma A	14503	MO
51	Forma A	13720	PDBFF	165	Forma A	136246	PDBFF
52	Forma A	51212	PDBFF	166	Forma A	137765	PDBFF
53	Forma B	143412	PDBFF	167	Forma A	280008	PPBio
54	Forma B	171561	PDBFF	168	Forma A	280013	PPBio
55	Forma B	213117	PDBFF	169	Forma A	280029	PPBio
56	Forma B	176138	PDBFF	170	Forma A	280040	PPBio
57	Forma A	14354	MO	171	Forma A	280045	PPBio
58	Forma A	14346	MO	172	Forma A	2762	INPA
59	Forma A	146199	PDBFF	173	Forma A	53921	PDBFF
60	Forma A	171646	PDBFF	174	Forma A	32350	PDBFF
61	Forma A	167316	PDBFF	175	Forma A	164517	PDBFF
62	Forma A	52014	PDBFF	176	Forma B	280009	PPBio
63	Forma A	168666	PDBFF	177	Forma A	280010	PPBio
64	Forma A	194914	PDBFF	178	Forma A	280024	PPBio
65	Forma A	38222	PDBFF	179	Forma A	280036	PPBio
66	Forma B	159574	PDBFF	180	Forma A	14324	MO
67	Forma B	5337	PDBFF	181	Forma A	45274	PDBFF
68	Forma A	14411	RB	182	Forma A	51308	PDBFF
69	Forma A	280044	PPBio	183	Forma B	280043	PPBio
70	Forma A	14451	INPA	184	Forma A	14227	INPA
71	Forma A	14402	INPA	185	Forma A	3799	INPA
72	Forma A	14048	INPA	186	Forma A	14326	INPA
73	Forma A	14504	MO	187	Forma A	14617	MO
74	Forma A	14618	MO	188	Forma A	206613	PDBFF
75	Forma A	15398	PDBFF	189	Forma A	2761	INPA
76	Forma A	187639	PDBFF	190	Forma A	280031	PPBio
77	Forma A	113771	PDBFF	191	Forma A	280033	PPBio
78	Forma B	41933	PDBFF	192	Forma A	14204	INPA
79	Forma B	175850	PDBFF	193	Forma A	164599	PDBFF
80	Forma B	161582	PDBFF	194	Forma B	280011	PPBio
81	Forma B	162298	PDBFF	195	Forma A	205661	PDBFF
82	Forma B	173163	PDBFF	196	Forma A	152308	PDBFF
83	Forma B	176207	PDBFF	197	Forma A	175522	PDBFF
84	Forma A	280014	PPBio	198	Forma B	280015	PPBio
85	Forma A	280034	PPBio	199	Forma A	280016	PPBio

<i>Anexo I</i>				cont.			
N	Morfo	wiki_id	Herbário	N	Morfo	wiki_id	Coleção
86	Forma A	14213	INPA	200	Forma A	280042	PPBio
87	Forma A	175737	PDBFF	201	Forma A	14470	MO
88	Forma A	52940	PDBFF	202	Forma A	39639	PDBFF
89	Forma A	280022	PPBio	203	Forma A	280030	PPBio
90	Forma B	4113	INPA	204	Forma B	203867	PDBFF
91	Forma A	14207	INPA	205	Forma A	30757	PDBFF
92	Forma A	14351	MO	206	Forma A	14195	INPA
93	Forma A	172237	PDBFF	207	Forma B	14163	INPA
94	Forma A	129430	PDBFF	208	Forma A	1764	PDBFF
95	Forma A	176365	PDBFF	209	Forma A	203995	PDBFF
96	Forma A	190878	PDBFF	210	Forma A	14210	INPA
97	Forma B	34377	PDBFF	211	Forma A	14192	INPA
98	Forma B	198720	PDBFF	212	Forma A	242744	PDBFF
99	Forma B	14209	RB	213	Forma A	14189	INPA
100	Forma B	280047	PPBio	214	Forma A	154036	PDBFF
101	Forma A	280048	PPBio	215	Forma A	1654	PDBFF
102	Forma A	14537	INPA	216	Forma A	32571	PDBFF
103	Forma A	14205	INPA	217	Forma A	280039	PPBio
104	Forma A	14265	INPA	218	Forma B	8294	PDBFF
105	Forma A	14267	INPA	219	Forma A	14142	MO
106	Forma A	14333	MO	220	Forma A	564	PDBFF
107	Forma A	212922	PDBFF	221	Forma A	118019	PDBFF
108	Forma A	24970	PDBFF	222	Forma A	148862	PDBFF
109	Forma A	119813	PDBFF	223	Forma A	14624	MO
110	Forma A	151153	PDBFF	224	Forma A	37762	PDBFF
111	Forma A	146169	PDBFF	225	Forma A	187576	PDBFF
112	Forma A	31841	PDBFF	226	Forma A	5120	PDBFF
113	Forma A	213068	PDBFF	227	Forma A	14440	MO
114	Forma B	14469	MO	228	Forma A	149818	PDBFF
115	Forma B	200313	PDBFF	229	Forma A	6757	PDBFF

## ANEXO II

Scripts utilizados no programa R para organização e obtenção dos resultados apresentados.

Script 1: lista de funções necessárias utilizadas como base para as demais análises

```
library(e1071)
```

```
library(MASS)
```

```
library(cluster)
```

```
library(gridExtra)
```

```
library(mclust)
```

```
library(fpc)
```

```
library(multcompView)
```

```
library(multcomp)
```

```
testa.classificacao <- function(dad,colgrp="GROUPING",  
analises=c("svm","bayes","lda"),method=c("loo","perc"),perc=30) {
```

```
#methods: garante que coluna e fator
```

```
  dad[,colgrp] = as.factor(dad[,colgrp])
```

```
#dad[160,colgrp], loo = leave one out
```

```
#perc = train.set perc% of the samples of each category in colgrp. In this case, 100  
#permutations are conducted and the best score is retrieved.
```

```
#CRIA OBJETOS PARA SALVAR OS RESULTADOS
```

```
#RESUMO PARA MATRIX DE CONFUSÃO:
```

```
spp = as.vector(unique(dad[,colgrp]))
```

```
resumo.svm = matrix(0,nrow=length(spp),ncol=length(spp),dimnames = list(spp,spp))
```

```
resumo.lda = resumo.svm
```

```
resumo.bayes = resumo.lda
```

```
resumos = NULL
```



```

#funções para pegar ids do trainset

pegatrain <- function(x=namednames,aperc=perc) {

  n.por.grupo = table(as.vector(x))

  idstrain = NULL

  for(i in 1:length(n.por.grupo)) {

#print(i)

    n = n.por.grupo[i]

    gp = names(n.por.grupo)[i]

    vl = as.vector(x)[gp] & !is.na(as.vector(x))

#quantas amostras são necessárias selecionar para o grupo (100-aperc%)

    quantas = ceiling(n*((100-aperc)/100))

#pega os ids para essas amostras

    if (n<=3 | quantas==n) {

      if (n>2) {

        print(paste(gp,"tem apenas",n,"amostras! O modelo irá incluir sempre duas!"))

        ids = sample(names(x[vl]),2)

      } else {

#inclui todas no modelo, esta espécie não será testada pois tem menos do que 3 amostras, mas
#ela integra o modelo

        print(paste(gp,"tem apenas",n,"amostras! O modelo irá incluir todas! O grupo não será
testado"))

        ids = names(x[vl])

      }

    } else {

      osids = names(x[vl])

      ids = sample(osids,quantas,replace=F)

```

```

    }

    idstrain = c(idstrain,ids)

  }

  return(idstrain)

}

pegamelhorprob <- function(x,predicao) {

  pt = as.vector(predicao[x,"MELHOR.PREDICAO"])

  val = predicao[x,pt]

  return(val)

}

pegapreditocomo <- function(x,tb) {

  gp = which(tb[x,]==max(tb[x,]))

  return(colnames(tb)[gp])

}

#RESULTADO BRUTO - que contém a probabilidade da amostra pertencer à classe predita

oslv = levels(as.factor(dad[,colgrp]))

resultados.brutos = matrix(NA,nrow=1,ncol=length(oslv)+6,dimnames =
list(NULL,c("ID","ANALISE",oslv,"ORIGINAL","MELHOR.PREDICAO","ACERTO","M
ELHOR.PREDICAO.PROB")))

#FAZ OS LEAVE-ONE-OUTs

if ("loo"%in%tolower(method)) {

#para cada amostra (linha) em dad

  for (i in 1:nrow(dad)){

    cat(".")

#índice da coluna com nome dos grupos

    idxcolgrp = which(colnames(dad)==colgrp)

```

```

#seleciona os dados para o modelo (- a amostra i)
  train = dad[-i,-idxcolgrp]

#seleciona os dados da amostra (apenas variáveis)
  test=dad[i,-idxcolgrp]

#seleciona a classe original da amostra
  original = as.vector(dad[i,idxcolgrp])

  if ("svm"%in%analises) {

#validação via support vector machine svm()

#define o modelo
  svmm <- svm(dad[-i,colgrp] ~ ., train, probability=TRUE)

#prediz a amostra i
  predsvm = predict(svmm,test,probability=T)

#qual foi a melhor predição
  preditocomo = as.vector(predsvm)

#quais são as probabilidades de predição
  predicoes = as.data.frame(attr(predsvm,"probabilities"))

#adiciona colunas
  predicoes = cbind(ID=rownames(dad)[i],ANALISE='SVM',predicoes,ORIGINAL =
original,MELHOR.PREDICAO = preditocomo)

#resume o acerto
  if (preditocomo==original) {
    acerto = "acerto"
  } else {
    acerto= "ERRO"
  }

  predicoes$ACERTO = acerto

```

```

predicoes$MELHOR.PREDICAO.PROB = predicoes[1,preditocomo]

#junta os dados

cln = colnames(resultados.brutos)

cln2 = cln[!cln%in%colnames(predicoes)]

if(length(cln2)>0) {
  predicoes[,cln2] = NA
}

resultados.brutos = rbind(resultados.brutos,predicoes[,cln])

#print(paste(i,original,nrow(test),"vs",nrow(train)))

#colnames(resultados.brutos)==colnames(predicoes)

#print(paste("SVM [",paste(i,"de",nrow(dad)),"]:",original," predito",preditocomo,":", acerto))

#guarda matriz de confusão svm

resumo.svm[original,preditocomo] = resumo.svm[original,preditocomo]+1
}

if ("bayes"%in%analises) {

#agora faz naiveBayes

bayescl <- naiveBayes(dad[-i,colgrp] ~ ., train,probability=TRUE)

predbayes = predict(bayescl,test,type='raw')

predbayes = as.data.frame(predbayes)

rownames(predbayes) = rownames(dad)[i]

predv = predbayes[1,]

preditocomo = unique(colnames(predbayes)[predv==max(predv)])

predicoes = cbind(ID=rownames(dad)[i],ANALISE='BAYES',predv,ORIGINAL =
original,MELHOR.PREDICAO = preditocomo)

#resume o acerto

if (!is.na(preditocomo) & preditocomo==original) {

```

```

    acerto= "acerto"

  } else {

    acerto= "ERRO"

  }

  predicoes$ACERTO = acerto

  if (!is.na(preditocomo)) {

    oprob= predicoes[1,preditocomo]

#guarda matriz de confusão bayes

    resumo.bayes[original,preditocomo] = resumo.bayes[original,preditocomo]+1

#print(paste("BAYES [",paste(i,"de",nrow(dad)),"]:",original," predito",preditocomo,":",
acerto))

    } else {

#print(paste("BAYES [",paste(i,"de",nrow(dad)),"]:",original," não foi predita como nada:",
acerto))

    oprob=0

    }

#junta os dados

  predicoes$MELHOR.PREDICAO.PROB = oprob

#junta os dados

  cln = colnames(resultados.brutos)

  cln2 = cln[!cln%in%colnames(predicoes)]

  if(length(cln2)>0) {

    predicoes[,cln2] = NA

  }

  resultados.brutos = rbind(resultados.brutos,predicoes[,cln])

}

```

```

if ("lda"%in%analises)
#faz uma simples LDA

mylda <- lda(dad[-i,colgrp] ~ ., train)

predlda = predict(mylda,test)

predv = as.data.frame(predlda$posterior)

preditocomo = as.vector(predlda$class)

predicoes = cbind(ID=rownames(dad)[i],ANALISE='LDA',predv,ORIGINAL =
original,MELHOR.PREDICAO = preditocomo)

#resume o acerto

if (preditocomo==original) {

  acerto= "acerto"

} else {

  acerto= "ERRO"

}

predicoes$ACERTO = acerto

predicoes$MELHOR.PREDICAO.PROB = predicoes[1,preditocomo]

#junta os dados

c1n = colnames(resultados.brutos)

c1n2 = c1n[!c1n%in%colnames(predicoes)]

if(length(c1n2)>0) {

  predicoes[,c1n2] = NA

}

resultados.brutos = rbind(resultados.brutos,predicoes[,c1n])

#print(paste("LDA [" ,paste(i,"de",nrow(dad)),"]:",original," predito",preditocomo,":", acerto))

#guarda matriz de confusão LDA

resumo.lda[original,preditocomo] = resumo.lda[original,preditocomo]+1

```

```
}  
}  
if (!is.null(resumo.svm)) {  
  resumos[["SVM"]] = resumo.svm  
}  
if (!is.null(resumo.bayes)) {  
  resumos[["BAYES"]] = resumo.bayes  
}  
if (!is.null(resumo.lda)) {  
  resumos[["LDA"]] = resumo.lda  
}  
}  
else {  
  res.per = NULL  
  namednames = as.vector(dad[,colgrp])  
  names(namednames) = rownames(dad)  
  melhorfits = vector("list", 3)  
  todaspreds = vector("list", 3)  
  for(p in 1:100) {  
#cat("\014")  
#cat(paste(p,"de",100))  
  cat(".")  
#define o trainset  
  perc=30  
  train.set = pegatrain(x=namednames,aperc=perc)
```

```

test.set = names(namednames)

test.set = test.set[!test.set%in%train.set]

#índice da coluna com nome dos grupos

idxcolgrp = which(colnames(dad)==colgrp)

#seleciona os dados para o modelo (- a amostra i)

train = dad[train.set,-idxcolgrp]

#seleciona os dados da amostra (apenas variáveis)

test=dad[test.set,-idxcolgrp]

#seleciona as classes originais das amostras

original = as.vector(dad[test.set,idxcolgrp])

if ("svm"%in%analises) {

  svmm <- svm(as.factor(dad[train.set,colgrp]) ~ ., train, probability=TRUE)

#prediz a amostra i

  predsvm = predict(svmm,test,probability=T)

#qual foi a melhor predição

  preditocomo = as.vector(predsvm)

#quais são as probabilidades de predição

  predicoes = as.data.frame(attr(predsvm,"probabilities"))

#adiciona colunas

  predicoes = cbind(ID=rownames(dad)[i],ANALISE='SVM',predicoes,ORIGINAL =
original,MELHOR.PREDICAO = preditocomo)

  vl = original==preditocomo

  predicoes$ACERTO = "ERRO"

  predicoes$ACERTO[vl] = "acerto"

  mpb = sapply(1:nrow(predicoes),pegamelhorprob,predicoes)

  predicoes$MELHOR.PREDICAO.PROB = mpb

```



```

#porcentagem de acerto significativo

vll = vl & mpb>0.95

pasg = round((sum(vll)/nrow(predicoes))*100,1)

if (!is.null(melhorfits)) {

  melhorfits[['svm']] = c(melhorfits[['svm']],sum(vll))

} else {

  melhorfits[['svm']] = c(sum(vll))

}

todaspreds[['svm']][[p]] = predicoes

}

if ("bayes"%in%analises) {

  bayescl <- naiveBayes(as.factor(dad[train.set,colgrp]) ~ ., train,probability=TRUE)

  predbayes = predict(bayescl,test,type='raw')

  predbayes = as.data.frame(predbayes)

  rownames(predbayes) = rownames(test)

  preditocomo = sapply(1:nrow(predbayes),pegapreditocomo,tb=predbayes)

  predicoes = cbind(ID=rownames(test),ANALISE='BAYES',predbayes,ORIGINAL =
original,MELHOR.PREDICAO = preditocomo)

  vl = original==preditocomo

  predicoes$ACERTO = "ERRO"

  predicoes$ACERTO[vl] = "acerto"

  mpb = sapply(1:nrow(predicoes),pegamelhorprob,predicoes)

  predicoes$MELHOR.PREDICAO.PROB = mpb

#porcentagem de acerto significativo

vll = vl & mpb>0.95

pasg = round((sum(vll)/nrow(predicoes))*100,1)

```

```

if (!is.null(melhorfits)) {
  melhorfits[['bayes']] = c(melhorfits[['bayes']],sum(vll))
} else {
  melhorfits[['bayes']] = c(sum(vll))
}
todaspreds[['bayes']][[p]] = predicoes
}

if ("lda"%in%analises) {
  mylda <- lda(as.factor(dad[train.set,colgrp]) ~ ., train)
  predlda = predict(mylda,test)
  predv = as.data.frame(predlda$posterior)
  preditocomo = as.vector(predlda$class)

  predicoes = cbind(ID=rownames(test),ANALISE='LDA',predv,ORIGINAL =
original,MELHOR.PREDICAO = preditocomo)

  vl = original==preditocomo

  predicoes$ACERTO = "ERRO"
  predicoes$ACERTO[vl] = "acerto"

  mpb = sapply(1:nrow(predicoes),pegamelhorprob,predicoes)

  predicoes$MELHOR.PREDICAO.PROB = mpb

#porcentagem de acerto significativo

vll = vl & mpb>0.95

pasg = round((sum(vll)/nrow(predicoes))*100,1)

if (!is.null(melhorfits)) {
  melhorfits[['lda']] = c(melhorfits[['lda']],sum(vll))
} else {
  melhorfits[['lda']] = c(sum(vll))
}

```

```

    }

    todaspreds[['lda']][[p]] = predicoes
  }

#final da permutação

nn = sapply(melhorfits,length)

gp = which(nn>0)

for(f in 1:length(gp)) {
  on = names(gp)[f]

  ofit = which(melhorfits[[on]]==max(melhorfits[[on]]))

  if (length(ofit)>1) {
    ofit=sample(ofit,1)
  }

#pega a predição com melhor fit

  apred = todaspreds[[on]][[ofit]]

#salva o resultado

  resultados.brutos = rbind(resultados.brutos,apred)

#gera tabela confusão

#guarda matriz de confusão LDA

  tbb= table(apred$ORIGINAL,apred$MELHOR.PREDICAO)

  resumos[[toupper(on)]] = tbb
}

}

resultados.brutos = as.data.frame(resultados.brutos,stringsAsFactors=F)

res1 = NULL

for(r in 1:length(resumos)) {

```

```

res = resumos[[r]]

tp = names(resumos)[r]

r1 = diag(res)/apply(res,1,sum)

ns = sum(res)

r1 = c(r1,TOTAL=sum(diag(res))/ns)

res1 = rbind(res1,r1)

}

rownames(res1) = names(resumos)

res1 = as.data.frame(res1,stringsAsFactors = F)

res1 = round(res1*100,1)

final = list(BRUTOS = resultados.brutos, RESUMOS = resumos, ANALISES.RES =res1 )

return(final)

}

plotmatriz <- function(matriz,bg.cols=NULL,txt.cols=NULL, valcex=1, cexaxis=1) {

  tb = matriz

#define cores do backgroun e texto

  if (is.null(bg.cols)) {

    bg.cols= tb

    bg.cols[tb==0] = gray(level=1)

    bg.cols[tb==1] = gray(level=0.8)

    bg.cols[tb>1 & tb<=5] = gray(level=0.6)

    bg.cols[tb>5 & tb<=10] = gray(level=0.4)

    bg.cols[tb>10] = gray(level=0)

  }

  if (is.null(txt.cols)) {

```

```

txt.cols = tb

#se for 0 é branco

txt.cols[tb==0] = gray(level=1)

#se for 1 é cinza claro

txt.cols[tb==1] = gray(level=0)

#se for entre 1 e 5 é um pouco mais escuro

txt.cols[tb>1 & tb<=5] = gray(level=0)

txt.cols[tb>5 & tb<=10] = gray(level=1)

#se for > 10 o fundo é preto e o texto branco

txt.cols[tb>10] = gray(level=1)

#1 para 0 coloca NA, isso não é plotado

}

#plota uma matriz vazia

xx = 1:ncol(tb)

yy = seq(1,nrow(tb),length.out=length(xx))

plot(xx,yy,type='n',xaxt='n',yaxt='n',xlab="",ylab="",pty='m',
xlim=c(0.5,max(xx)+.5),ylim=c(0.5,max(yy)+0.5))

##ax = yy[1:(length(yy)-1)]+(diff(yy)/2)

##ay = xx[1:(length(xx)-1)]+(diff(xx)/2)

ay = 1:ncol(tb)

ax = 1:nrow(tb)

abline(v=ay,h=ax,lty='dotted',lwd=0.5)

nn = rownames(tb)

nncl = colnames(tb)

for(l in 1:nrow(tb)) {

  for(cc in 1:ncol(tb)) {

```

```

cl = bg.cols[l,cc]

txtc = txt.cols[l,cc]

val = tb[l,cc]

if (!is.na(val)) {
  rect(xleft=cc-0.5,ybottom=l-0.5,xright=cc+0.5,ytop=l+0.5,density=-1,border=NA,col=cl)
  text(cc,l,labels=val,cex=valcex,col=txtc)
}
}
}

axis(side=3,at=ay,labels=colnames(tb),cex.axis=cexaxis,las=2)
axis(side=2,at=ax,labels=rownames(tb),cex.axis=cexaxis,las=2)
}

#função para calcular a média por célula
m.media<-function(x,nomecoluna="") {
  res = NA
  if (!is.na(x)) {
    if (is.character((x))) {
      xx = strsplit(x,";")[[1]]
      xx = xx[xx!=""]
      xx1 = as.numeric(xx)
      xx1 = xx1[!is.na(xx1)]
      if (length(xx1)!=length(xx)) {
        print(paste("Coluna",nomecoluna,"tem valor não numérico:",paste(xx,collapse =
";"),""))
      }
      xx = xx1
    }
  }
}

```

```

    } else { xx = x }

    xx = xx[!is.na(xx)]

    xx = mean(xx,na.rm = T)

    if (length(xx)==1) { res = xx }

  }

  return(res)

}

#função que aplica m.media para todas as colunas em dad

aplicamedia <- function(dad) {

  for(cl in 1:ncol(dad)) {

    v1 = dad[,cl]

    vm = sapply(v1,m.media,nomecoluna=colnames(dad)[cl])

    dad[,cl] = as.numeric(as.vector(vm))

  }

  return(dad)

}

qualnaoena <- function(x) {

  return(which(!is.na(x)))

}

#funções para plotar as matrizes de correlação entre variáveis com pares()

panel.hist <- function(x, ...)

{

  usr <- par("usr"); on.exit(par(usr))

  par(usr = c(usr[1:2], 0, 1.5) )

  h <- hist(x, plot = FALSE)

```

```

breaks <- h$breaks; nB <- length(breaks)

y <- h$counts; y <- y/max(y)

rect(breaks[-nB], 0, breaks[-1], y, col = "cyan", ...)

}

panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  dx = cbind(x,y)
  dx = na.omit(dx)
  x = dx[,1]
  y = dx[,2]
  r <- abs(cor(x, y))
  txt <- format(c(r, 0.123456789), digits = digits)[1]
  txt <- paste0(prefix, txt)
  if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)
  text(0.5, 0.5, txt, cex = cex.cor * r)
}

#faz o teste de correlações par a par entre
#variáveis - normalizando os dados antes de correlacionar
fazascor <- function(dv)
{
  mt = NULL
  for(j in 1:ncol(dv)) {
#print(j)

```



```

toc = 1:ncol(dv)

toc = toc[toc!=j]

for(n in toc) {
#print(n)

  dtoc = dv[,c(j,n)]

  dtoc = na.omit(dtoc)

#print(nrow(dtoc))

  dtoc = scale(dtoc,center=T,scale=T)

  r <- try(cor.test(dtoc[,1],dtoc[,2]),silent = T)

  acor= NA

  opv = NA

  if (!class(r)=='try-error'){

    if (!is.na(r$estimate)) {

      acor = round(r$estimate,3)

      opv = round(r$p.value,4)

    }    }

  vars = sort(colnames(dv)[c(j,n)])

  rr =
data.frame(VAR1=vars[1],VAR2=vars[2],CORRELACAO=acor,PV=opv,stringsAsFactors =
F)

  mt = rbind(mt,rr)

}

}

mt = as.data.frame(mt)

mt = unique(mt)

mt = mt[order(abs(mt$CORRELACAO),decreasing=T),]

```

```

return(mt)
}

plotatable <- function(x) {

nl = nrow(x)

nc = ncol(x)

xx = seq(1.5,nc+0.5)

yy = (nl+1):1

plot(1:nc,seq(1,nl+1,length.out=nc),xlim=c(1,nc+1),type='n',xaxt='n',yaxt='n',xlab="",ylab="",bt
y='n')

text(xx,yy[1],labels=colnames(x),cex=0.8,adj=0.5,font=2)

for(l in 1:nrow(x)) {

for(j in 1:ncol(x)) {

if (j<=2) {ad = 0} else {ad=0.5}

text(xx[j],yy[l+1],labels=x[l,j],cex=0.6,adj=ad)

}

}

}

removeacentos <- function(x) {

xx = iconv(x, to="ASCII//TRANSLIT")

xx = gsub("[~|^|~|\"|'|..Â°.|.|. ]", "", xx)

#retorna o valor da posicao indicada

return(xx)

}

mytrim <- function (x) {gsub("^\\s+|\\s+$", "", x)}

colunavazia <- function(x) {

x = gsub(" ", "", x)

```

```

x[x=="" & !is.na(x)] = NA

xx = x[!is.na(x)]

if (length(xx)==0) {

  return(FALSE)

} else {

  return(TRUE)

}

}

x = "10/18/2016"

pegadata <- function(x,qual="dd") {

  xx = strsplit(x,split="/")[[1]]

  xx= as.numeric(xx)

  xx= xx[xx!=10 & xx!=2016 & xx!=11 & xx!=2017]

  if (length(xx)==1) {

    return(xx)

  } else { return(NA)}

}

faltam <- function(dd) {

#COLUNAS QUE FALTAM

v1 = grep("_corr",colnames(dd))

c11 = colnames(dd)[v1]

c11 = gsub("_corr","",c11)

v2 = which(colnames(dd)%in%c11)

gp = unique(c(v1,v2))

fl = colnames(dd)[-gp]

```

```

return(fl)
}
#qual linha tem o maior valor?
qualmax <- function(x) {
  which(x==max(x,na.rm=T) & !is.na(x))
}
#refina grupos
#grupos são vetores com nomes de linha
#dados.morfo são os dados morfológicos para refinar a classificação
#refina a classificação do grupo2, em geral mais segmentada que grupo1 (k1<k2)
refinagrupos <- function(grupo1,grupo2,dados.morfo) {
  tbb = as.matrix(table(grupo1,grupo2))
  qlm = apply(tbb,2,max,na.rm=T)
  qlm0 = qlm
  qlm = qlm+1
  #para cada grupo enquanto não estabilizar
  while(sum(!qlm0==qlm)>0) {
    print(sum(!qlm0==qlm))
    novaclasse = data.frame(NOVACLASSE=grupo2,stringsAsFactors = F)
    rownames(novaclasse)= names(grupo2)
  }
  #para cada coluna ou grupo2
  for(q in 1:ncol(tbb)) {
    #qual o grupo
    gp = colnames(tbb)[q]
    #quais os dados do grupo coluna (k=8) são do grupo

```

```

    vl = grupo2==gp & !is.na(grupo2)
#quais linhas tem o maior valor?

    ql = qualmax(tbb[,gp])
#qual é o grupo da linha

    gp2 = rownames(tbb)[ql]
#quais os nomes dessa linha

    vl2 = grupo1==gp2 & !is.na(grupo1)
#combinando isso quais amostras do grupo2 não estão nessa linha?

    vll = vl & !vl2
#desclassifica essas amostras

    novaclasse$NOVACLASSE[vll] = NA
}
#pega os dados train e dados test
#train

    vl = !is.na(novaclasse$NOVACLASSE)

    rn = rownames(novaclasse)[vl]

    train = dados.morfo[rn,]
#test são as demais amostras

    rnn = rownames(dados.morfo)

    rnn = rnn[!rnn%in%rn]

    test = dados.morfo[rnn,]
#faz uma análise discriminante com o train (modelo) pega o agrupamento core

    vl = !is.na(novaclasse$NOVACLASSE)

    grouping = novaclasse$NOVACLASSE[vl]

    mylda <- lda(grouping ~ ., train)

```

```

#prediz as demais

predlda = predict(mylda,test)

predv = as.data.frame(predlda$posterior)

preditocomo = as.vector(predlda$class)

#pega a maior predição

mpred = round(apply(predv,1,max),2)

#salva a classe predita e a probabilidade da predição

rn = rownames(predv)

novaclasse[rn,"NOVACLASSE"] = preditocomo

novaclasse[rn,"PROBABILIDADE"] = mpred

novaclasse[rownames(train),"PROBABILIDADE"] = 1

#compara novamente

tbb2 = as.matrix(table(grupo1,novaclasse$NOVACLASSE))

#refaz a operação

tbb= tbb2

qlm0=qlm

qlm = apply(tbb,2,max,na.rm=T)

}

#agora classifica as amostras, GRUPO CORE = todas aquelas que continuam juntas em
ambas as classificações

novaclasse2 = data.frame(NOVACLASSE=grupo2,stringsAsFactors = F)

rownames(novaclasse2)= names(grupo2)

#para cada coluna ou grupo2

for(q in 1:ncol(tbb)) {

#qual o grupo

gp = colnames(tbb)[q]

```

```

#quais os dados do grupo coluna (k=8) são do grupo
  v1 = grupo2==gp & !is.na(grupo2)

#quais linhas tem o maior valor?
  q1 = qualmax(tbb[,gp])

#qual é o grupo da linha
  gp2 = rownames(tbb)[q1]

#quais os nomes dessa linha
  v12 = grupo1==gp2 & !is.na(grupo1)

#combinando isso quais amostras do grupo2 não estão nessa linha?
  v11 = v1 & !v12

#desclassifica essas amostras
  novaclasse2$NOVACLASSE[v11] = NA
}

v1 = !is.na(novaclasse2$NOVACLASSE)

novaclasse$CATEGORIA = NA

novaclasse$CATEGORIA[v1] = "core"

#todas as demais são incertezas, ainda que com alta probabilidade
novaclasse$CATEGORIA[!v1] = "incerteza"

novaclasse = data.frame(GRUPO1.ORG=grupo1,GRUPO2.ORG=grupo2,novaclasse)

return(novaclasse)
}

```

Script 2: utilizado para a organização de dados e verificação de variáveis mais influenciáveis

```
#carrega pacotes e funcoes necessarias
```

```
source("funcoesNecessarias.R")
```

```
#define caminho e arquivos para ler
```

```

dados.brutos = 'morfogui_quanti.csv'

#caminhos para salvar as figuras, tabelas e dados gerados pelo script, irá repor arquivos com o
mesmo nome se já gerados

#criar pasta se não houver no diretório atual de trabalho (getwd())

pathparafiguras= "figuras"

pathparatabelas= "tabelas"

pathparadados = "dadosParaAnalises"

#lê os dados brutos, com variáveis ordinais e nominais codificadas como números ou 0 e 1

dad.morfo.bruto = read.table(file=dados.brutos,sep="\t",na.strings=c("NA",""),row.names =
NULL,as.is=T,header=T)

#obtendo os valores médios de cada amostra

dados.media = aplicamedia(dad.morfo.bruto)

#qual o número de variáveis para cada amostra? Para cada amostra qual o número mínimo de
amostras com quem pode comparar? Classificar MANUALMENTE as variáveis

#com nome das colunas vegetativas quantitativas

veg.quant =
c('altlargmax','comptotal','complam','larglam','peccom','npares','largestip','compestip')

#nome das colunas vegetativas categóricas, ou seja, todas as variáveis que originalmente eram
categóricas

veg.quali =
c('foleliptica','folobovada','foloblan','folance','apiceacum','apiceagu','baseacum','baseagu','nec
enadpro','necenadimp','necenabpro','necenabimp','nersecadpro','nersecadimp','nervsecaba','v
eneucampt','venbroq','vencras','veneucampbr','veiassecobl','intersec','intercur','intermed','interl
on','tercobli','estipcon','indestip','indestipsimpad','indestipdim','indestipdenso','indpec','indpessi
mpere','indpecpeq','indpecdenso','indfada','indfaba','indfabasimpadp','indfabadim','indfabaesp')

#analise somente de caracteres vegetativos

#para cada conjunto de variáveis, conta para cada variável o número de amostras que ela
permite comparar

contaramostraporvar <- function(y,dd) {

  yy = as.numeric(as.vector(dd[,y]))

  sum(!is.na(yy))

```



```

}

pd.veg.quant = sapply(veg.quant,contamostraporvar,dd=dados.media)

pd.veg.ind = sapply(veg.quali,contamostraporvar,dd=dados.media)

#junta tudo

mm1 = data.frame(CATEG="Vegetativos Quantitativos",NAMOSTRAS=pd.veg.quant)

mm2 = data.frame(CATEG="Vegetativos Qualitativos",NAMOSTRAS=pd.veg.ind)

mm = rbind(mm1,mm2)

tbb = table(mm$NAMOSTRAS)

tbb = as.data.frame(tbb)

row.names(tbb)= tbb[,1]

categs = levels(as.factor(mm$CATEG))

for(l in 1:length(categs)){

  vl = as.factor(mm$CATEG)==categs[l]

  tbb1 = table(mm$NAMOSTRAS[vl])

  tbb[,categs[l]] = NA

  tbb[names(tbb1),categs[l]] = tbb1

}

tbb = tbb[,-1]

tbb = as.matrix(tbb[,-1])

tbb = t(tbb)

dir.create(pathparafiguras)

dir.create(pathparatabelas)

dir.create(pathparadados)

pdf(paste(pathparafiguras,"/01_amostrasPorvariaveis1.pdf",sep=""),width=8.5,height=10)

par(las=1,mar=c(4,12,3,2))

```

```
barplot(tbb,beside=T,col=c("green","red"),horiz=T,ylab="Número de amostras",xlab="Número de variáveis")
```

```
legend(x=2,y=30,legend=row.names(tbb),pch=22,pt.bg=c("green","red"),ncol=1,cex=0.8)
```

```
dev.off()
```

#há variáveis que permitem comparar poucas amostras? Qual o mínimo aceitável para inclusão na análise? Ou seja, quais variáveis devem ser eliminadas porque têm muitos valores ausentes na coluna? Melhor eliminar essas variáveis dos dados, pois terão pouco efeito de qualquer forma nas comparações

#Qual o numero mínimo de amostras para incluir a variável?

```
n.minimo = 5
```

```
vl = mm$NAMOSTRAS<=n.minimo
```

#então quais eliminar?

```
quais.vars = rownames(mm[vl,])
```

#elimina essas colunas dos dados.media e das listas de categorias

```
rn = colnames(dados.media)
```

```
rn = rn[!rn%in%quais.vars]
```

```
dados.media = dados.media[,rn]
```

```
veg.quant = veg.quant[!veg.quant%in%quais.vars]
```

```
veg.quali= veg.quali[!veg.quali%in%quais.vars]
```

#definindo similaridades:

#1) qual importância das variáveis medidas, 2) elas expressam de fato o que eu considero informativo no meu grupo?

#FORMA DE FOLHAS

#definindo variáveis de acordo com o que eu percebo que é importante. A forma da folha expressa pelas seguintes variáveis melhor descreve os padrões que eu percebo nas minhas amostras:

#a) As folhas variam na proporção entre largura e comprimento da lamina (há folhas mais obovadas, outras mais elípticas e outras mais redondas). Isso pode ser codificado em duas razões numéricas:

#a.1) máxima largura dividida pelo comprimento da lamina (forma da lâmina) - FORMA01

#salva figuras das variáveis criadas num arquivo

```
pdf(paste(pathparafiguras, "/02_formaFolhas1.pdf", sep=""), width=8.5, height=10)
```

```
FORMA01 = dados.media$larglam/dados.media$complam
```

```
par(mfrow=c(3,1), las=1, mar=c(5,4,3,3), mgp=c(2.5,0.5,0), cex.axis=0.8, tck=-0.01)
```

```
h = hist(FORMA01, breaks = 30, col='blue', xlim=c(0, max(FORMA01, na.rm =
T))+min(FORMA01, na.rm=T)), ylab = "Número amostras", xlab="Forma01 = (máxima
largura/comprimento) da Lâmina Foliar", main="")
```

```
text(h$mids[1], h$count[1]+0.5, labels="Folhas + estreitas", srt=90, adj=c(0,0.5), cex=0.8)
```

```
text(h$mids[length(h$mids)], h$count[length(h$count)]+0.5, labels="Folhas +
arredondadas", srt=90, adj=c(0,0.5), cex=0.8)
```

#b) outra medida que mostra se lâmina é mais obovada, elíptica ou ovada, é a razão entre a altura da máxima largura e comprimento total (FORMA 02)

```
FORMA02 = dados.media$altlargmax/dados.media$complam
```

```
par(las=1, mar=c(5,4,3,2), mgp=c(2.5,0.5,0), cex.axis=0.8, tck=-0.01)
```

```
h <- hist(FORMA02, breaks = 30, col='red', xlim=range(FORMA02, na.rm=T)+c(-0.1,+0.1), ylab
="Número amostras", xlab="Forma02 = (altura da máxima largura/comprimento) da Lâmina
Foliar", main="")
```

```
text(h$mids[1], h$count[1]+0.5, labels="Folhas + elípticas", srt=90, adj=c(0,0.5), cex=0.8)
```

```
text(h$mids[length(h$mids)], h$count[length(h$count)]+0.5, labels="Folhas +
obovadas", srt=90, adj=c(0,0.5), cex=0.8)
```

#c) uma última variável de forma, é a proporção da folha que é pecíolo, ou seja do comprimento total da folha, quanto é pecíolo. Isso mostra proporcionalmente se o pecíolo é muito pequeno ou se a folha é peciolada (FORMA 03)

```
FORMA03 = dados.media$peccom/(dados.media$peccom+dados.media$comptotal)
```

```
par(las=1, mar=c(5,4,3,2), mgp=c(2.5,0.5,0), cex.axis=0.8, tck=-0.01)
```

```
h <- hist(FORMA03, breaks = 30, col='green', xlim=range(FORMA03, na.rm=T)+c(-
0.04,+0.04), ylab = "Número amostras", xlab="Forma03 = (comprimento pecíolo/comprimento
da folha)", main="")
```

```
text(h$mids[1], h$count[1]+0.5, labels="Folhas sésseis", srt=90, adj=c(0,0.5), cex=0.8)
```

```
text(h$mids[length(h$mids)], h$count[length(h$count)]+0.5, labels="Folhas +
pecioladas", srt=90, adj=c(0,0.5), cex=0.8)
```

```

dev.off()

#adiciona essas variaveis aos vetores e dados

veg.quant= c(veg.quant,c("FORMA01","FORMA02","FORMA03"))

dados.media = data.frame(dados.media,FORMA01,FORMA02,FORMA03,stringsAsFactors
= F)

#CARACTERES FLORAIS

#no caso deste exemplo nada faz sentido e significa algo (mas será testado apenas para checar
correlações das variáveis)

pdf(paste(pathparafiguras,"/05_formaFlores.pdf",sep=""),width=8.5,height=10)

FORMA.ESTILETE = dados.media$altpisto/dados.media$compestilete

par(mfrow=c(3,1),las=1,mar=c(5,4,3,3),mgp=c(2.5,0.5,0),cex.axis=0.8,tck=-0.01)

h = hist(FORMA.ESTILETE,breaks
=30,col='blue',xlim=range(FORMA.ESTILETE,na.rm=T)+c(-0.04,+0.04),ylab ="Número
amostras",xlab="FormaEstilete = (comprimento pistilo/comprimento estilete",main="")

text(h$mids[1],h$count[1]+0.5,labels="Estilete I + curtos",srt=90,adj=c(0,0.5),cex=0.8)

text(h$mids[length(h$mids)],h$count[length(h$count)]+0.5,labels="Estilete I +
longos",srt=90,adj=c(0,0.5),cex=0.8)

#agora vamos juntar os dados apenas com as variáveis

FORMA.OVARIO = dados.media$altpisto/dados.media$compovario

par(las=1,mar=c(5,4,3,2),mgp=c(2.5,0.5,0),cex.axis=0.8,tck=-0.01)

h = hist(FORMA.OVARIO,breaks
=30,col='green',xlim=range(FORMA.OVARIO,na.rm=T)+c(-0.04,+0.04),ylab ="Número
amostras",xlab="FormaOvario = (comprimento pistilo/comprimento ovario",main="")

text(h$mids[1],h$count[1]+0.5,labels="Ovarios menores",srt=90,adj=c(0,0.5),cex=0.8)

text(h$mids[length(h$mids)],h$count[length(h$count)]+0.5,labels="Ovarios
maiores",srt=90,adj=c(0,0.5),cex=0.8)

#adiciona essas variáveis aos vetores e dados

flor.quant = c(flour.quant,c("FORMA.ESTILETE","FORMA.OVARIO"))

dados.media =
data.frame(dados.media,FORMA.ESTILETE,FORMA.OVARIO,stringsAsFactors = F)

```

## #CHECA AS CORRELAÇOES ENTRE VARIÁVEIS. MOSTRA AS DISTRIBUIÇÕES UNIVARIADAS

```
osdados = list(veg.quant=veg.quant,flor.quant=flor.quant,qualitativos=c(veg.quali,flor.quali))
```

```
#note que juntei os dois grupos de dados qualitativos para ver a correlação entre eles, pois se referem neste caso ao indumento
```

```
#para cada conjunto de variáveis checa as correlações entre variáveis e gera uma figura para cada dado (num pdf unico) e uma tabela mostrando as correlações (significativas) entre todos os pares de variáveis
```

```
arqn = paste(pathparafiguras,"/06_CorrelacoesEntreVariaveis.pdf",sep="")
```

```
pdf(arqn,width=10,height=8.5)
```

```
for(d in 1:length(osdados)) {
```

```
  colunas = osdados[[d]]
```

```
  nomedado = names(osdados)[d]
```

```
  #pega esses dados
```

```
  dv = dados.media[,colunas]
```

```
  #salva num arquivo
```

```
  pairs(dv, lower.panel = panel.smooth, upper.panel = panel.cor, diag.panel = panel.hist)
```

```
  mtext(side=3,line=2.5,text = toupper(nomedado),cex=1,adj=0, font=2)
```

```
  #testa correlacoes
```

```
  ascors = fazascor(dv)
```

```
  #mostra apenas as significativas
```

```
  v1 = ascors$PV<0.05
```

```
  par(mar=c(5,5,3,3))
```

```
  plotatable(ascors[v1,])
```

```
  mtext(side=2,line=2,text = paste("Correlações significativas para",toupper(nomedado)),cex=1,adj=1, font=2)
```

```
  fn = paste(pathparatabelas,"/",nomedado,".csv",sep="")
```

```
  write.table(ascors,fn,sep="\t",na="",row.names = T,quote=T)
```

```
}
```

```
dev.off()
```

```
#ELIMINANDO VARIÁVEIS CORRELACIONADAS, EXPLICITANDO O QUE VOCE  
ACHA IMPORTANTE
```

```
#na figura gerada acima algumas variáveis estão fortemente correlacionadas. Elas expressam  
apenas 1 coisa, tamanho das parte vegetativas. Isso pode estar representado na análise por  
apenas 1 dessas variáveis, indicando tamanho. Note que as variáveis de forma não estão  
correlacionadas com as variáveis brutas medidas, expressando outro tipo de informação
```

```
#a variável escolhida para representar isso foi comprimento da lâmina
```

```
veg.quant.excluidas = c("atlargmax", "comptotal")
```

```
gp = which(veg.quant%in%veg.quant.excluidas)
```

```
veg.quant = veg.quant[-gp]
```

```
#Em relação à flor, os caracteres comprimento da flor, comprimento da sépala, largura da flor,  
comprimento da corola, largura da corola e largura da sépala estão fortemente correlacionados  
( $r > 0.8$ ). Todos eles expressam tamanho da flor em geral. Portanto, vamos ficar apenas com  
comprimento da flor expressando isso
```

```
#Em relação aos estames, serão representados por altura dos estames, excluindo comprimento  
do filete e comprimento da antera
```

```
#Em relação ao pistilo, este será representado em partes por altura do pistilo, sendo excluído  
comprimento do estigma
```

```
flor.quant.excluidas = c("compcorola", "largcorola", "largflor", "compsep", "largsep",  
"compfilete", "compantera", "compestigma")
```

```
flor.quant = flor.quant[!flor.quant%in%flor.quant.excluidas]
```

```
#junta todas as variáveis num objeto medias
```

```
cls = c(veg.quant, flor.quant)
```

```
dados.media2 = dados.media[,cls]
```

```
#todo registro precisa pelo menos ter dados vegetativos, elimina linhas que não tem
```

```
vl = is.na(dados.media2$complam)
```

```
dados.media2 = dados.media2[!vl,]
```

```

vl = is.na(dados.media$complam)

dados.media = dados.media[!vl,]

#ultima checada

contanas <- function(x) { sum(is.na(x))/length(x) }

nas1 = apply(dados.media,1,contanas)

nas2 = apply(dados.media2,1,contanas)

hist(nas1,main='NAs em dados.media',breaks=20,col="red")

hist(nas2,main='NAs em dados.media2',breaks=20,col="red")

#salva esse arquivo

fn = paste(pathparadados,"/dados_medias_varsel.csv",sep="")

write.table(dados.media2,file=fn,sep="\t",row.names = T,na="",quote=T)

#salva o arquivo com todas as variáveis

fn = paste(pathparadados,"/dados_medias_todasvar.csv",sep="")

write.table(dados.media,file=fn,sep="\t",row.names = T,na="",quote=T)

#gera um arquivo com histogramas para cada variável

odado = dados.media

fn = paste(pathparafiguras,"/07_dadoMediasHistogramas_todas.pdf",sep="")

pdf(fn,width=8.5,height=10)

par(mfrow=c(3,2),mar=c(5,4,3,3),mgp=c(2.5,0.5,0),tck=-0.01,cex.axis=0.8)

for(cl in 1:ncol(odado)) {

  cln = colnames(odado)[cl]

  var = odado[,cl]

  hist(var,xlab=cln,col='red',breaks=30 ,main="")

}

dev.off()

```

Script 3: com os dados organizados é gerada a NMDS

```
#MATRIZ DE DISTANCIA E NMDS
```

```
rm(list=ls())
```

```
source("funcoesNecessarias.R")
```

```
#calcula a distância morfológica e o dado NMDS resultante
```

```
#CALCULA A DISTANCIA MORFOLÓGICA USANDO GOWER
```

```
#lê os dados
```

```
arq1 = "dadosParaAnalises/dados_medias_todasvar.csv"
```

```
dados.media = read.table(file=arq1,sep="\t",row.names = 1,na.strings =c("NA",""))
```

```
#elimina umas colunas desses dados
```

```
eliminadas = c("FOL_LARDALAMNAMET", "COL_ALT", "COL_DAP")
```

```
cls = colnames(dados.media)
```

```
cls = cls[!cls%in%eliminadas]
```

```
dados.media = dados.media[,cls]
```

```
#dados com variáveis selecionadas
```

```
arq2 = "dadosParaAnalises/dados_medias_varsel.csv"
```

```
dados.media2 = read.table(file=arq2,sep="\t",row.names = 1,na.strings =c("NA",""))
```

```
#qual desses dados vai rodar?
```

```
qualdados = list(varsel=dados.media2,todasvar=dados.media)
```

```
#define caminho para salvar figuras e objetos
```

```
pathparadados = "dadosParaAnalises"
```

```
pathparafiguras = "figuras"
```

```
pathparatabelas = "tabelas"
```

```
dir.create(pathparadados)
```

```
dir.create(pathparafiguras)
```



```

dir.create(pathparatabelas)

#calcula a distância morfológica de gower

#veja detalhes no help da função daisy para entender o que ele faz com NAs (pode ter dados
faltantes)

#para cada dado

parasalvar = NULL

for(d in 1:length(qualdados)) {
  odado = qualdados[[d]]

  nomedado = names(qualdados)[d]

  md = daisy(odado, metric = "gower", stand = F, warnBin = F)

#pega parte inferior da matriz, excluindo a diagonal

  tt = md[lower.tri(md,diag = F)]

#plota essas distâncias numa figura

  fn = paste(pathparafiguras,"/Histograma_morfoGowerDist_",nomedado,".pdf",sep="")

  pdf(fn,width=6,height=5)

  hist(tt,breaks=40,col='red',xlab="Distância morfológica de gower",main="")

  dev.off()

#se for NA, então a dissimilaridade é 0 (No help: "If all weights w_k delta(ij;k) are zero, the
dissimilarity is set to NA")

#vamos tratar tudo como 0 de dissimilaridade para esses pares de amostra

  md[is.na(md)] = 0

#NMDS não funciona se tem amostras com distancia 0. Então, coloca uma dissimilaridade
muito pequena para esses casos

  md[md==0] = table(round(md,3))[2]/100000

#guarda objetos de distancia numa lista para um Rdata

  parasalvar[[nomedado]] = list(morfodist=md,dado=odado)
}

```

```

#salva distâncias de gower num Rdata para uso

fn = paste(pathparadados,"/morfoDistGower.Rdata",sep="")

save(parasalvar,file=fn)

#QUAL O NÚMERO DE EIXOS MINIMOS (K) QUE DEVO REDUZIR MEUS DADOS
SEM PERDER INFORMAÇÃO

#qual o número de eixos que minimiza o valor de stress no NMDS?

#testa de k=2 a k=número de variáveis originais

fn = paste(pathparafiguras,"/EixosMinimosParaNMDSbruto_",nomedado,".pdf",sep="")

pdf(fn,width=8.5,height=10)

par(mfrow=c(2,1),mar=c(5,4,3,2),cex.axis=0.8)

for(d in 1:length(parasalvar)) {

  odado = parasalvar[[d]][['dado']]

  nomedado = names(parasalvar)[d]

  md = parasalvar[[d]][['morfodist']]

  qualk = NULL

#para tantas variáveis quanto em odado

  for(k in 2:ncol(odado)) {

    print(paste(k,"de",ncol(odado)))

    m.nmds = isoMDS(md,k=k)

    rr = data.frame(K=k,STRESS=m.nmds$stress)

    qualk = rbind(qualk,rr)

  }

  K = qualk$K

  STRESS = qualk$STRESS

  plot(K,STRESS,type='l',lwd=2,lty='solid',main=nomedado,col='red')

  points(K,STRESS,pch=21,cex=2,bg='red')

```

```

}

dev.off()

#TRANSFORMANDO OS DADOS

#qual o K mínimo após o qual o valor de stress estabiliza?

#vendo a correlação entre matrizes

#md1 =parasalvar[[1]][['morfodist']]

#md2 =parasalvar[[2]][['morfodist']]

#plot(md1,md2)

#um vetor com um valor de K para cada dado em parasalvar

um.bom.k = c(30,30)

#gerar o NMDS e salvar os dados

for(d in 1:length(parasalvar)) {

  odado = parasalvar[[d]][['dado']]

  nomedado = names(parasalvar)[d]

  md = parasalvar[[d]][['morfodist']]

  m.nmnds = isoMDS(md,k=um.bom.k[d])

  dado.nmnds = m.nmnds$points

  colnames(dado.nmnds) = paste("NMDS",1:ncol(dado.nmnds),sep="")

  fn = paste(pathparadados,"/dadosMORFOnmnds",nomedado,".csv",sep="")

  write.table(dado.nmnds,file=fn,sep="\t",na="",row.names=T,quote=T)

}

#pronto dadosMorfoNMDS gerados para análises

Script 4: dados submetidos às análises de agrupamento. Neste script são descritos os passos apenas para dados morfológicos. Ao inserir os dados NIR os scrips 4 e 6 são iguais, havendo apenas a inserção da planilha de dados espectrais

#TESTANDO CLASSIFICAÇÕES A POSTERIORI

```

```

rm(list=ls())

library(mclust)

#funções e pacotes necessários

source("funcoesNecessarias.r")

#para salvar resultados

pathparadados = "dadosParaAnalises"

pathparafiguras = "figuras"

pathparatabelas = "tabelas"

#dir.create(pathparadados)

#dir.create(pathparafiguras)

#dir.create(pathparatabelas)

#pega os dados. O exemplo testado é uma matriz de NMDS gerada a partir de distâncias
morfológicas

arq = "dadosParaAnalises/dadosMORFOnmdsvarel.csv"

dados.morfo = read.table(file=arq,sep='\t',as.is=F,row.names = 1,header=T)

#PASSO 01 - QUAL O NÚMERO DE GRUPOS?

#qual a porcentagem de acerto usando testa.classificacao para cada k e metodo de particao
(mclust,kmeans, pam ou specc)? O melhor k (e método de partição) que indique
agrupamentos das amostras é aquele em que há maior taxa de acerto numa validação cruzada.
A validação pode ser feita com os mesmos dados que geram os grupos ou com dados
diferentes

dado.grupos = dados.morfo

dado.teste = dados.morfo #pode usar NIR aqui por exempo, se tiver

#ambos esses dados precisam ter o mesmo número de linhas e as amostras na mesma ordem
garante isso

dado.teste = dado.teste[row.names(dado.grupos),]

#qual a sequencia de valores de K a serem testados (k=número de grupos) K=10 GRUPOS

kseq = seq(10,2,by=-1)

```

```

#para cada valor de K, gera diferentes agrupamentos para diferentes métodos

ltotest = NULL

for(oka in kseq) {

  cat("0\14")

  cat(paste("Gerando K=",oka,"grupos"))

  tkn = speccCBI(dado.grupos,k=oka)

  ltotest[["speccCBI"]][[as.character(oka)]] = tkn$partition

  clb1 = Mclust(dado.grupos,G=oka)

  ltotest[["Mclust"]][[as.character(oka)]] = clb1$classification

  clb1 = pam(dado.grupos,k=oka,cluster.only=T)

  ltotest[["pam"]][[as.character(oka)]] = clb1

  clb1 = kmeans(dado.grupos, center=oka,iter.max=10000,nstart=1000)

  ltotest[["kmeans"]][[as.character(oka)]] = clb1$cluster

}

str(ltotest[[2]])

#agora testa a estabilidade dos agrupamentos. Isso pode demorar bastante - para cada grupo
acima ira testar a classificação dos dados

bestks.tb =
data.frame(FUN=NA,K=NA,MAX.TAXA.ACERTO=NA,SVM=NA,BAYES=NA,LDA=NA
,SVM.ALL=NA,BAYES.ALL=NA,LDA.ALL=NA)

bestks = NULL

for(l in 1:length(ltotest)) {

  lt = ltotest[[l]]

  nlt = names(ltotest)[l]

  for(n in 1:length(lt)) {

    print(paste(names(ltotest)[l],"k=",names(ltotest[[l]])[n]))

  }

}

#pega o agrupamento

```

```

cl = ltotest[[1]][[n]]

oka = names(ltotest[[1]])[n]

#copia o dado

dd = dado.teste

#adiciona o agrupamento gerado

rn = row.names(dd)

dd$GROUPING = as.factor(paste("GRUPO",cl))

#testa o agrupamento com LDA loo

oteste =
testa.classificacao(dad=dd,colgrp="GROUPING",analises=c("svm","bayes","lda"),method=c(
"loo","perc")[1],perc=15)

#pega o fit do teste (maior taxa de acerto total)

br =oteste$BRUTOS

#pega melhor predição

vl = br$MELHOR.PREDICAO.PROB>0.95 & br$ACERTO=='acerto' &
!is.na(br$ACERTO)

ovv = table(as.vector(br[vl,"ANALISE"]))

ovv1 = ovv/length(unique(br$ID))

c1l = c("BAYES","LDA","SVM")

mcll = c1l[!c1l%in%names(ovv1)]

rz = c(FUN=nlt,K=oka,MAX.TAXA.ACERTO=max(ovv1),ovv1)

if (length(mcll)>0) {rz[mcll] = NA}

#pega acertos totais

vl = br$ACERTO=='acerto'

ovv = table(as.vector(br[vl,"ANALISE"]))

ovv = ovv/length(unique(br$ID))

c1l = c("BAYES","LDA","SVM")

```

```

mcll = cll[!ccl%in%names(ovv)]
if (length(mcll)>0) {ovv[mcll] = NA}

ovv = ovv[ccl]

names(ovv) = c("BAYES.ALL","LDA.ALL","SVM.ALL")

rz = c(rz,ovv)

rz = rz[colnames(bestks.tb)]

print(paste(round(max(ovv,na.rm=T),4)," taxa de acerto máxima para K=",oka))

#salva o resultado do teste para k

bestks.tb = rbind(bestks.tb,rz)

bestks[[nlt]][[as.character(oka)]] = oteste

}

}

#faz uma figura com os resultados

bestks.tb = as.data.frame(bestks.tb)

rownames(bestks.tb) = paste(bestks.tb$FUN,bestks.tb$K)

bestks.tb = bestks.tb[order(as.numeric(bestks.tb$K)),]

#bestks.tb = na.omit(bestks.tb)

fn = paste(pathparafiguras,"/otimoNgruposKmeansTest.pdf",sep="")

v1 = is.na(bestks.tb$K)

sum(v1)

bestks.tb = bestks.tb[!v1,]

pdf(fn,width=8.5,height=10)

par(mfrow=c(3,2),mar=c(5,4,3,2))

for(acol in 4:6) {

  clnam = colnames(bestks.tb)[acol]

```

```

plot(bestks.tb[!is.na(bestks.tb$FUN),c(2,acol)],lwd=1,type='n',main='Acertos
significativos',ylab=paste('Taxa de acerto',clnam))

ff = unique(bestks.tb$FUN)

ff = ff[!is.na(ff)]

lty = c("solid","dotted","dashed","dotdashed")

pchscol = c("red","blue","green","purple")

for(f in 1:length(ff)) {

  vll = bestks.tb$FUN==ff[f] & !is.na(bestks.tb[,acol])

  points(bestks.tb[vll,c(2,acol)],lwd=1,col=pchscol[f],type='l')

  points(bestks.tb[vll,c(2,acol)],pch=21,bg=pchscol[f])

}

legend("bottomleft",legend=ff,pch=21,pt.bg=pchscol,cex=0.8,bg='white')

abline(h=0.95,lty='dotted')

plot(bestks.tb[!is.na(bestks.tb$FUN),c(2,acol+3)],lwd=1,type='n',main='Todos os
acertos',ylab=paste('Taxa de acerto',clnam))

for(f in 1:length(ff)) {

  vll = bestks.tb$FUN==ff[f] & !is.na(bestks.tb[,acol+3])

  points(bestks.tb[vll,c(2,acol+3)],lwd=1,col=pchscol[f],type='l')

  points(bestks.tb[vll,c(2,acol+3)],pch=21,bg=pchscol[f])

}

legend("bottomleft",legend=ff,pch=21,pt.bg=pchscol,cex=0.8,bg='white')

abline(h=0.95,lty='dotted')

}

dev.off()

Script 5: filtrando os dados espectrais para submeter às análises

#Calculando a media dos espectros por individuo

```



```

dados = read.table('dadoscompletosNIR.csv',sep="\t", as.is=T, header=T)

colunaind = "FileName" #coluna contendo o identificador dos individuos

colunacat = "NOME" #coluna contendo a categoria que seja testar

#pega colunas NIR

cls = grep("X",colnames(dados),ignore.case=F)

#filtra os dados que pode calcular a média

dd = dados[,cls]

#calcula a média por individuo

dt = aggregate(dd,by=list(dados[, colunaind]),FUN=mean)

colnames(dt)[1] = colunaind

rownames(dt) = dt[,colunaind]

#pega os dados dos individuos

db = dados[,c(colunaind,colunacat)]

db = unique(db)

head(db)

rownames(db) = db[,colunaind]

#junta os dados agregados

rn = rownames(dt)

dados.media = data.frame(colunacat=db[rn,2],dt[rn,],stringsAsFactors=F)

#ver o começo dos dados

dados.media[1:5,1:5]

#salva como arquivo no computador

write.table(dados.media, file='dadosNIRmedia.csv',sep="\t", row.names=F)

```

Script 6: após separar quais os melhores agrupamentos, estes são convertidos em matrizes de confusão para a interpretação dos dados

## #INTERPRETANDO RESULTADOS E CLASSIFICANDO AMOSTRAS: APENAS MORFOLOGIA

#a) eu quero K grupos que tem alta taxa de predição em modelos de validação cruzada (loo) por modelos (SVM, Bayes e/ou LDA)

#a.1) no exemplo, quando K=2, em todos os modelos de gerar grupos a taxa de acerto total é > 95% para SVM, BAYES, LDA e kmeans

#a.2) K=3 e k=4 são outros agrupamentos que tem relativo alto suporte em vários métodos > 90%

#explorando essas três hipóteses (k=2, k=3, k=4). Os diferentes ks explorados aqui permitirão visualizar a hierarquia das similaridades entre os grupos

#primeiro vamos obter as classificações para cada K, depois vamos plotar essas classificações em espaço discriminante

#reclassifica e identifica incertezas para esses Ks

```
quaisks=c(2,3,4)
```

#cria um objeto para salvar resultados

```
resd = dado.grupos[,1:2]
```

#refaz o agrupamento para esse valor de K (número de grupos)

```
for(q in 1:length(quaisks)) {
```

#com Kmeans

```
ttk = kmeans(dado.grupos, center=quaisks[q],iter.max=10000,nstart=1000)
```

```
clk = ttk$cluster
```

#com Mclust

```
ttmc = Mclust(dado.grupos,G=quaisks[q])
```

```
clmc = ttmc$classification
```

```
tbb = table(KMEANS=clk,MCLUST=clmc)
```

#testa essa classificação

```
dd=dado.teste
```

```
grps = clk
```

```

dd$GROUPING=paste("GRUPO",sprintf("%02s",grps))

#com LDA e loo

oteste =
testa.classificacao(dad=dd,"GROUPING",analises=c("svm","bayes","lda")[3],method=c("loo",
"perc")[1])

#pega o resultado

br = oteste$BRUTOS

#classifica as amostras

br$CLASSIFICACAO = NA

#quais amostras tem predição correta nos grupos

vl = br$MELHOR.PREDICAO.PROB>=0.95 & br$ACERTO=='acerto' &
!is.na(br$ACERTO)

sum(vl)

br$CLASSIFICACAO[vl] = "core"

#quais são acertos não significativos

vl = br$MELHOR.PREDICAO.PROB<=0.95 & br$ACERTO=='acerto' &
!is.na(br$ACERTO)

sum(vl)

br$CLASSIFICACAO[vl] = "incerteza"

#quais amostras são erros

vl = br$ACERTO=='ERRO' & !is.na(br$ACERTO)

sum(vl)

#desclassifica os erros

br$MELHOR.PREDICAO[vl] = NA

br$MELHOR.PREDICAO.PROB[vl] = NA

br$CLASSIFICACAO[vl] = "sem classe"

cls = c("MELHOR.PREDICAO","MELHOR.PREDICAO.PROB","CLASSIFICACAO")

```

```

brr = br[,cls]

colnames(brr) = paste("K=",quaisks[q],"_",cls,sep="")

brr = brr[-1,]

rnn = rownames(brr)

resd = cbind(resd[rnn,],brr[rnn,])

}

resd = resd[,-c(1:2)]

head(resd)

#salva essas classificações

fn = paste(pathparadados,"/classificacaoAposterioriK.csv",sep="")

write.table(resd,file=fn,sep="\t",na="",quote=T,row.names=T)

#REFINANDO E OBTENDO HIPÓTESE FINAL A POSTERIORI: COMPARANDO
CLASSIFICAÇÕES A POSTERIORI

fn = "dadosParaAnalises/classificacaoaposteriorik.csv"

resd = read.table(file=fn,sep="\t",na.strings=c("NA",""),row.names=1,header=T,as.is=T)

colunasgrupos = seq(1,ncol(resd)-2,by=3)

comparacoes = NULL

tbbs = NULL

for(cl in colunasgrupos) {

  resto = colunasgrupos[!colunasgrupos==cl]

  for(cl2 in resto) {

    comparacao =paste(sort(colnames(resd)[c(cl,cl2)]),collapse="-")

    if (!comparacao%in%comparacoes) {

      comparacoes = c(comparacoes,comparacao)

      tb = table(resd[,cl],resd[,cl2])

      tbbs[[comparacao]] = tb

```

```

#apenas core

v1 = resd[,c1+2]== 'core' & !is.na(resd[,c1+2])

v12 = resd[,c12+2]== 'core' & !is.na(resd[,c12+2])

v11 = v1 & v12

tb = table(resd[v11,c1],resd[v11,c12])

tbbs[[paste(comparacao,"CORE")]] = tb

}

}

}

fn = paste(pathparafiguras, "/matrizConfusaoEntreKsAposteriori.pdf", sep="")

pdf(file=fn,width=8.5,height=6)

for(t in 1:length(tbbs)) {

  print(t)

  par(mar=c(2,7,7,2),mfrow=c(1,1),cex.axis=0.8)

  tb = tbbs[[t]]

  comparacao = names(tbbs)[t]

#define cores do backgroun e texto

  bg.cols= tb

  txt.cols = tb

#se for 0 é branco

  bg.cols[tb==0] = gray(level=1)

  txt.cols[tb==0] = gray(level=1)

#se for 1 é cinza claro

  bg.cols[tb==1] = gray(level=0.8)

  txt.cols[tb==1] = gray(level=0)

```

```
#se for entre 1 e 5 é um pouco mais escuro
bg.cols[tb>1 & tb<=5] = gray(level=0.6)
txt.cols[tb>1 & tb<=5] = gray(level=0)
bg.cols[tb>5 & tb<=10] = gray(level=0.4)
txt.cols[tb>5 & tb<=10] = gray(level=1)
#se for > 10 o fundo é preto e o texto branco
bg.cols[tb>10] = gray(level=0)
txt.cols[tb>10] = gray(level=1)
#1 para 0 coloca NA, isso não é plotado
tb[tb==0] = NA
plotmatriz(tb,bg.cols=bg.cols,txt.cols=txt.cols,valcex=0.8, cexaxis = 0.8)
mtext(side=3,text=comparacao,adj=0,line=6,cex=0.8)
}
dev.off()
```