UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA
Programa de Pós-Graduação em Química

MAURO VICENTINI CORREIA

**Redes Neurais e Algoritmos Genéticos no estudo
Quimiossistemático da Família Asteraceae.**

São Paulo

Data do Depósito na SPG:
01/02/2010

MAURO VICENTINI CORREIA

**Redes Neurais e Algoritmos Genéticos no estudo
Quimiossistemático da Família Asteraceae.**

*Dissertação apresentada ao Instituto de Química
da Universidade de São Paulo para obtenção do
Título de Mestre em Química (Química Orgânica)*

Orientador: Prof. Dr. Vicente de Paulo Emerenciano.

São Paulo
2010

Mauro Vicentini Correia
Redes Neurais e Algoritmos Genéticos no estudo Quimiossistemático da Família
Asteraceae.

*Dissertação apresentada ao Instituto de Química
da Universidade de São Paulo para obtenção do
Título de Mestre em Química (Química Orgânica)*

Aprovado em: _____

Banca Examinadora

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

DEDICATÓRIA

À minha mãe, Silmara Vicentini pelo suporte e apoio em todos os momentos da minha vida, sem os quais esse trabalho não seria possível.

Ao Grande Amigo Braz Gerônimo de Queiroz (in memoriam) pelos preciosos conselhos na minha primeira juventude.

AGRADECIMENTOS

Ao Ciclo Pentano, como a muito prometido, pela amizade e por compartilhar comigo ainda nos tempos da graduação a idéia desse mestrado. Em especial aos amigos Mário Henrique Tomassi e Rafael Vianna Croffi.

Ao Professor Dr. Vicente de Paulo Emerenciano, por me aceitar em seu grupo, pela orientação, principalmente pelas idéias que acrescentaram muito em qualidade ao trabalho, e pelo conhecimento compartilhado.

Aos amigos Harold Hilarion Fokoue, Marcus Tullius Scotti, Alejandro Speck Planche, Sarah Ramalho Rodrigues e Michelle Rossini pelos bons momentos no laboratório e pela grande contribuição nesta dissertação.

Aos professores da USP, principalmente aqueles com que tive contato na Faculdade de Ciências Farmacêuticas, Instituto de Química e Instituto de Ciências Biomédicas, por compartilhar comigo sua ciência.

Aos alunos do curso de Farmácia e Bioquímica 2008 (noturno), com os quais apreendi a compartilhar o conhecimento.

A CAPES e à FAPESP pelas bolsas de Mestrado Fornecidas.

“Erros são, no final das contas, fundamentos da verdade. Se um homem não sabe o que uma coisa é, já é um avanço do conhecimento saber o que ela não é.”

Carl Gustav Jung

RESUMO

Correia, M.V. **Redes Neurais e Algoritmos Genéticos no estudo Quimiossistemático da Família Asteraceae**. 2010. 189p. Dissertação (Mestrado) – Programa de Pós-Graduação em Química. Instituto de Química, Universidade de São Paulo, São Paulo.

No presente trabalho duas metodologias da área de inteligência artificial (Redes Neurais e Algoritmos Genéticos) foram utilizadas para realizar um estudo Quimiossistemático da família Asteraceae.

A família Asteraceae é uma das maiores famílias entre as Angiospermas, conta com aproximadamente 24.000 espécies. As espécies da família produzem grande diversidade de metabólitos secundários, entre os quais merecem destaque os terpenóides, poliacetilenos, flavonóides e cumarinas.

Para um melhor entendimento da diversidade química da família construiu-se um Banco de Dados com as ocorrências de doze classes de metabólitos (monoterpenos, sesquiterpenos, sesquiterpenos lactonizados, diterpenos, triterpenos, cumarinas, flavonóides, poliacetilenos, benzofuranos, benzopiranos, acetofenonas e fenilpropanóides) produzidos pelas espécies da família. A partir desse banco três diferentes estudos foram realizados.

No primeiro estudo, utilizando os mapas auto-organizáveis de Kohonen e o banco de dados químico classificado segundo duas das mais recentes filogenias da família foi possível realizar com sucesso separações de tribos e gêneros da família Asteraceae. Também foi possível indicar que a informação química concorda mais com a filogenia de Funk (Funk et al. 2009) do que com a filogenia de Bremer (Bremer 1994, 1996).

No estudo seguinte, onde se objetivou a criação de modelos de previsão dos números de ocorrências das doze classes de metabólitos, utilizando o perceptron de múltiplas camadas com algoritmo de retropropagação de erro, o resultado foi insatisfatório. Apesar de em algumas classes de metabólitos a fase de treino da rede apresentar resultados satisfatórios, a fase de teste mostrou que os modelos criados não são capazes de realizar previsão para dados aos quais eles não foram submetidos na fase de treino, e portanto não são modelos adequados para realizar previsões.

Finalmente, o terceiro estudo consistiu na criação de modelos de regressão linear utilizando como método de seleção de variáveis os algoritmos genéticos. Nesse estudo foi possível indicar que os monoterpenos e os sesquiterpenos são bastante relacionados biossinteticamente, também foi possível indicar que existem relações biossintéticas entre monoterpenos e diterpenos e entre sesquiterpenos e triterpenos.

Palavras-chave: Asteraceae, Redes Neurais, Quimiossistemática, Mapas Auto-Organizáveis, Perceptron de Múltiplas Camadas, Algoritmos Genéticos.

ABSTRACT

Correia, M.V. **Neural Network and Genetic Algorithms in the Chemosystematic study of Asteraceae Family.** 2010. 189p. Masters Thesis - Graduate Program in Chemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

In this study two methods of artificial intelligence (neural network and genetic algorithms) were used to work out a Chemosystematic study of the Asteraceae family.

The family Asteraceae is one of the largest families among the Angiosperms, having about 24,000 species. The species of the family produce a large diversity of secondary metabolites, and some worth mentioning are the terpenoids, polyacetylenes, flavonoids and coumarins.

For a better understanding of the chemical diversity of the family a database was built up with the occurrences of twelve classes of metabolites (monoterpenes, sesquiterpenes, lactonized sesquiterpenes, diterpenes, triterpenes, coumarins, flavonoids, polyacetylenes, Benzofurans, benzopyrans, acetophenones and phenylpropanoids) produced by species of the family. From this database three different studies were conducted.

In the first study, using the Kohonen self-organized map and the chemical data classified according to two of the most recent phylogenies of the family, it was possible to successfully separate the tribes and genera of the Asteraceae family. It was also possible to indicate that the chemical information agrees with the phylogeny of Funk (Funk et al. 2009) than with the phylogeny of Bremer (Bremer 1994, 1996).

In the next study, which aims at creating models to predict the number of occurrences of the twelve classes of metabolites using multi-layer perceptron with backpropagation algorithm error, the result was found unsatisfactory. Although in some classes of metabolites the training phase of the network has satisfactory results, the test phase showed that the models created are not able to make prevision for data to which they were submitted in the training phase and thus are not suitable models for predictions.

Finally, the third study was the creation of linear regression models using a genetic algorithm method of variable selection. This study could indicate that the monoterpenes and sesquiterpenes are closely related biosynthetically, and was also possible to indicate that there are biosynthetic relations between monoterpenes and diterpenes and between sesquiterpenes and triterpenes.

Keywords: Asteraceae, Neural Network, Chemosystematic, Self-Organizing Maps, Multi-layer Perceptron, Genetic Algorithms.

ÍNDICE DE FIGURAS

1. INTRODUÇÃO

Figura 1.2.1. Caracteres da família Asteraceae. A) Capítulo, B) Anteras fundidas, C) Tipos de cipselas (funk et al. 2005).....	26
Figura 1.3.1. Interrelações de ordens e algumas famílias do clado das angiospermas (APG 2003).....	27
Figura 1.3.2. Relações entre as famílias do clado euasteride II (campanulid). Destaque para a família Asteraceae no clado Asterales.....	28
Figura 1.3.3. Diagrama de inter-relações entre as 19 tribos de Asteraceae segundo Cassini 1816. Synanthérés = Asteraceae, Boopidéas = Calyceraceae.....	30
Figura 1.3.4. Classificações da família Asteraceae de Bentham e Hoffmann. A tribo Cynaroideae/Cynareae atualmente é chamada de Cardueae e Cichoriaceae/Cichorieae é agora a tribo Lactuceae.....	31
Figura 1.3.5. Diagrama de inter-relações filogenéticas das tribos da família Asteraceae, segundo Bremer [1].....	35
Figura 1.3.6. Diagrama resumido da análise filogenética das tribos da família Asteraceae segundo Meta-tree (adaptado de Funk et al. 2009).....	36
Figura 1.4.1.1. Funções Booleanas de duas variáveis representadas no plano binário (Kovács 2002).....	40
Figura 1.4.2.1. Ilustração do neurônio biológico (adaptado de Junqueira e Carneiro 2004).....	41
Figura 1.4.2.2. Classificação de algumas arquiteturas de redes neurais (adaptado de Lawrence 1994).....	42
Figura 1.4.3.1. Representação da rede do tipo Perceptrons de Múltiplas Camadas (adaptado de Zupan e Gasteiger 1993).....	43
Figura 1.4.3.2. Modelo não linear de neurônio (Haykin 2001).....	44
Figura 1.4.3.3. Tipos de função de ativação. a) função de limiar, b) função rampa, c) função sigmóide para parâmetro de inclinação a variável (Haykin, 2001).....	45

Figura 1.4.3.4. Resumo do algoritmo de retropropagação de erros (error backpropagation) (Corne 2006).....	47

Figura 1.4.4.1. Exemplo de rede de kohonen (SOM) unidimensional.....	48

Figura 1.5.1. Diagrama que posiciona os algoritmos genéticos como técnica de busca (Linden 2006).....	52

3. MATERIAIS E MÉTODOS

Figura 3.2.1. Rotas biossintéticas de produtos naturais em plantas (Mann et al.)....	57

Figura 3.2.2. Biossíntese dos Monoterpenos, Sesquiterpenos, Diterpenos, Triterpenos e Esteróides a partir do pirofosfato de isopentenila (IPP).....	58

Figura 3.2.3 – Biossíntese do IPP: Rota do ácido mevalônico.....	59

Figura 3.2.4 - Biossíntese do IPP: Rota do 1-desoxi-D-xilose-5-fosfato.....	60

Figura 3.2.5. Rota de formação do ácido chiquimico.....	60

Figura 3.2.6. Formação do ácido cinâmico a partir do ácido chiquimico.....	61

igura 3.2.7. Esquema biossintéticos de um composto aromático através da via dos policetídeos.....	62

Figura 3.2.8. Esquema de biossíntese de alguns flavonóides a partir do precursor 4-coumaroil-CoA (Dewick 2002).....	63

Figura 3.2.9. Esquema biossintético dos poliacetilenos.....	63

Figura 3.5.1.1. Exemplos do cálculo de NOX.....	67

Figura 3.5.2.1 Exemplos de cálculo de PO.....	68

Figura 3.5.5.1. Exemplos de tipos de esqueletos (ESQUEL) e moléculas que compartilham esses esqueletos.....	71

4. RESULTADOS

Figura 4.1.1. Distribuição das doze classes de metabólitos secundários nas 17 tribos (propostas por Bremer) da família Asteraceae.....	84
Figura 4.1.2. Distribuição das doze classes de metabólitos secundários nas 41 tribos presentes no banco (propostas na Meta-tree) da família Asteraceae.....	85
Figura 4.1.3. Esqueletos com maior número ocorrência na classe dos monoterpenos.....	87
Figura 4.1.4. Esqueletos com maior número ocorrência na classe dos sesquiterpenos.....	88
Figura 4.1.5. Esqueletos com maior número ocorrência na classe dos sesquiterpenos lactonizados.....	88
Figura 4.1.6. Esqueletos com maior número ocorrência na classe dos diterpenos.....	89
Figura 4.1.7. Esqueletos com maior número ocorrência na classe dos triterpenos.....	90
Figura 4.1.8. Esqueletos com maior número ocorrência na classe das cumarinas.....	90
Figura 4.1.9. Esqueletos com maior número ocorrência na classe dos flavonóides.....	91
Figura 4.1.10. Esqueletos com maior número ocorrência na classe dos poliacetilenos.....	92
Figura 4.1.11. Esqueletos com maior número ocorrência na classe dos benzofuranos.....	92
Figura 4.1.12. Esqueletos com maior número ocorrência na classe dos benzopiranos.....	93
Figura 4.1.13. Esqueletos com maior número ocorrência na classe das acetofenonas.....	93
Figura 4.1.14. Esqueletos com maior número ocorrência na classe dos fenilpropanóides.....	94

Figura 4.2.1.1. Mapas auto-organizáveis. a) Mapa obtido da classificação de cinco tribos da família Asteraceae (segundo Bremer): ANT (Azul), AST (Amarelo), HLT (Vermelho), SEN (Rosa), VER (Verde). b) Mapa obtido da classificação de sete tribos da família Asteraceae (segundo Meta-tree): ANT (Azul), AST (Amarelo), COR (Vermelho), MAD (Marrom), NAU (Azul Claro), SEN (Rosa), VER (Verde). c) Mapa obtido da classificação de nove gêneros da família Asteraceae: Acritopappus (Azul), Bellis (Verde), Geigeria (Azul claro), Mutisia (Vermelho), Osteospermum (Amarelo), Petasites (Rosa), Relhania (Cinza), Rudbeckia (Preto), Ursinia (Marrom).....**98**

Figura 4.2.1.2. Continuação Mapas. a) Mapa obtido da classificação de quatro Tribos da família Asteraceae segundo Bremer: CAR (Azul), LAC (Amarelo), MUT (Vermelho), SEN (Rosa). b) Mapa obtido da classificação quatro Tribos da família Asteraceae segundo Meta-tree: CAR (Azul), CIC (Amarelo), MUT (Vermelho), SEN (Rosa). c) Mapa obtido da classificação de três tribos da subfamília Asteroideae (segundo Bremer): GNA (Azul), INU (Vermelho), SEN (Amarelo). d) Mapa obtido da classificação de três tribos da subfamília Asteroideae (segundo Meta-tree): GNA (Azul), INU (Vermelho), SEN (Amarelo).....**99**

Figura 4.2.1.3. Continuação Mapas. a) Mapa obtido da classificação de quatro gêneros da família Asteraceae: Saussurea (Azul), Lactuca (Vermelho), Trixis (Amarelo) e Doronicum (Verde). b) Mapa obtido da classificação de quatro gêneros da família Asteraceae: Carthamus (Azul), Picris (Vermelho), Mutisia (Amarelo), Petasites (Verde). c) Mapa obtido da classificação de três gêneros da Subfamília Asteroideae: Geigeria (Azul), Petasites (Vermelho), Relhania (Amarelo). d) Mapa obtido da classificação de quatro Gêneros da Tribo Asterae: Aster (Azul), Erigeron (Vermelho), Heterotheca (Amarelo), Psidia (Verde).....**100**

Figura 4.2.2.1. Gráficos dos valores reais versus previsto obtidos no treino da rede neural com algoritmo de retro propagação. a) Gráfico dos monoterpenos, b) Gráfico dos sesquiterpenos, c) Gráfico dos sesquiterpenos lactonizados.....**106**

Figura 4.2.2.2. Gráficos dos valores reais versus previsto obtidos no teste da rede neural com algoritmo de retro propagação. a) Gráfico dos monoterpenos, b) Gráfico dos sesquiterpenos, c) Gráfico dos sesquiterpenos lactonizados.....**106**

Figura 4.2.2.3. Gráficos dos valores reais versus previsto obtidos no treino da rede neural com algoritmo de retro propagação. a) Gráfico dos diterpenos, b) Gráfico dos triterpenos, c) Gráfico das cumarinas.....**107**

Figura 4.2.2.4. Gráficos dos valores reais versus previsto obtidos no teste da rede neural com algoritmo de retro propagação. a) Gráfico dos diterpenos, b) Gráfico dos triterpenos, c) Gráfico das cumarinas.....**107**

Figura 4.2.2.5. Gráficos dos valores reais versus previsto obtidos no treino da rede neural com algoritmo de retro propagação. a) Gráfico dos flavonóides, b) Gráfico dos poliacetilenos, c) Gráfico dos benzofuranos.....**108**

Figura 4.2.2.6. Gráficos dos valores reais versus previsto obtidos no teste da rede neural com algoritmo de retro propagação. a) Gráfico dos flavonóides, b) Gráfico dos poliacetilenos, c) Gráfico dos benzofuranos.....**108**

Figura 4.2.2.7. Gráficos dos valores reais versus previsto obtidos no treino da rede neural com algoritmo de retro propagação. a) Gráfico dos benzopiranos, b) Gráfico das acetofenona, c) Gráfico dos fenilpropanóides.....**109**

Figura 4.2.2.8. Gráficos dos valores reais versus previsto obtidos no teste da rede neural com algoritmo de retro propagação. a) Gráfico dos benzopiranos, b) Gráfico das acetofenonas, c) Gráfico dos fenilpropanóides.....**109**

Figura 4.3.1. Esqueletos de Monoterpenos que aparecem nos modelos de regressão das equações 4.3.1 a 4.3.40.....**111**

Figura 4.3.2. Esqueletos de Sesquiterpenos que aparecem nos modelos de regressão das equações 4.3.1 a 4.3.40.....**112**

Figura 4.3.3. Esqueletos de Sesquiterpenos lactonizados que aparecem nos modelos de regressão das equações 4.3.12-4.3.13. 4.3.2.15, 4.3.2.20, 4.3.2.32 e, 4.3.37 – 4.3.39.....**113**

Figura 4.3.4. Esqueletos de diterpenos e triterpeno que aparecem nos modelos de regressão das equações 4.3.16, 4.3.19 e 4.3.40.....**113**

ÍNDICE DE TABELAS

1. INTRODUÇÃO

Tabela 1.2.1. Exemplos de espécies de Asteraceae e principais usos.....	25

Tabela 1.3.1. Principais classificações e filogenias recentes da família Asteraceae (Modificado de Bremer 1992).....	32

Tabela 1.3.2. Tribos da família Asteraceae segundo as filogenias de Bremer e a da Meta-tree, e os respectivos códigos de três letras utilizados para representar cada uma das tribos.....	37

3. MATERIAIS E MÉTODOS

Tabela 3.3.1. Ilustração da planilha criada com os dados químicos da família Asteraceae.....	65

Tabela 3.5.4.1. Reações características dos esqueletos presentes no banco de dados.....	69

Tabela 3.7.1. Índices estatísticos utilizados nos modelos de MLR utilizando algoritmos genéticos.....	76

4. RESULTADOS

Tabela 4.1.1. Fragmento do Banco de dados químicos de Asteraceae.....	78

Tabela 4.1.2 Quantidade de espécies presentes na família Asteraceae segundo Bremer [1], quantidade de espécies presentes no banco de dados químicos de Asteraceae do grupo, e porcentagem de espécies presentes no banco em relação à filogenia de Bremer.....	78

Tabela 4.1.3. Quantidade de gêneros presentes na família Asteraceae segundo Bremer, quantidade de gêneros presentes no banco de dados químicos de Asteraceae do grupo, e porcentagem de gêneros presentes no banco em relação à filogenia de Bremer.....	79

Tabela 4.1.4. Quantidade de gêneros presentes na família Asteraceae segundo Meta-tree, quantidade de gêneros presentes no banco de dados químicos de Asteraceae do grupo, e porcentagem de gêneros presentes no banco em relação à filogenia da Meta-tree.....	79

Tabela 4.1.5. Ocorrência e porcentagem das doze classes de metabólitos secundários nas 17 tribos (propostas por Bremer) da família Asteraceae. Monoterpenos (MONO), Sesquiterpenos (SESQ), Sesquiterpenos lactonizados (LACT), Diterpenos (DITE), Triterpenos (TRITE), Cumarinas (CUMA), Flavonóides (FLAV), Benzofuranos (BENZF), Benzopiranos (BENZP), Acetofenonas (ACET), Fenilpropanóides (FENIL).....**81**

Tabela 4.1.6. Ocorrência das doze classes de metabólitos secundários nas 41 tribos presentes no banco (propostas na Meta-tree) da família Asteraceae. Monoterpenos (MONO), Sesquiterpenos (SESQ), Sesquiterpenos lactonizados (LACT), Diterpenos (DITE), Triterpenos (TRITE), Cumarinas (CUMA), Flavonóides (FLAV), Benzofuranos (BENZF), Benzopiranos (BENZP), Acetofenonas (ACET), Fenilpropanóides (FENIL).....**82**

Tabela 4.1.7. Quantidade e porcentagem de esqueletos em classe de metabólito secundário na família Asteraceae.....**86**

Tabela 4.1.8. Descrição dos esqueletos com maior número de ocorrência na classe dos monoterpenos.....**87**

Tabela 4.1.9. Descrição dos esqueletos com maior número de ocorrência na classe dos sesquiterpenos.....**87**

Tabela 4.1.10. Descrição dos esqueletos com maior número de ocorrência na classe dos sesquiterpenos lactonizados.....**88**

Tabela 4.1.11. Descrição dos esqueletos com maior número de ocorrência na classe dos diterpenos.....**89**

Tabela 4.1.12. Descrição dos esqueletos com maior número de ocorrência na classe dos triterpenos.....**89**

Tabela 4.1.13. Descrição dos esqueletos com maior número de ocorrência na classe das cumarinas.....**90**

Tabela 4.1.14. Descrição dos esqueletos com maior número de ocorrência na classe dos flavonóides.....**91**

Tabela 4.1.15. Descrição dos esqueletos com maior número de ocorrência na classe dos poliacetilenos.....**91**

Tabela 4.1.16. Descrição dos esqueletos com maior número de ocorrência na classe dos benzofuranos.....**92**

Tabela 4.1.17. Descrição dos esqueletos com maior número de ocorrência na classe dos benzopiranos.....**92**

Tabela 4.1.18. Descrição dos esqueletos com maior número de ocorrência na classe das Acetofenonas.....**93**

Tabela 4.1.19. Descrição dos esqueletos com maior número de ocorrência na classe dos fenilpropanóides.....**94**

Tabela 4.2.1.1. Resultados dos Mapas Auto-Organizáveis, e suas respectivas dimensões, com os valores das ocorrências, os números de acertos absolutos e relativos para cinco Tribos segundo a filogenia de Bremer (figura 4.2.1.1 a) e sete tribos segundo a filogenia da Meta-tree (figura 4.2.1.1 b).....**95**

Tabela 4.2.1.2. Resultados do Mapa Auto-Organizável, e sua dimensão, com os valores das ocorrências, os números de acertos absolutos e relativos para nove gêneros da família Asteraceae (figura 4.2.1.1 c).....**95**

Tabela 4.2.1.3. Resultados dos Mapas Auto-Organizáveis, e suas respectivas dimensões, com os valores das ocorrências, os números de acertos absolutos e relativos para quatro tribos da família Asteraceae segundo Bremer (figura 4.2.1.2 a) e quatro tribos da família Asteraceae segundo a Meta-tree (figura 4.2.1.2 b).....**96**

Tabela 4.2.1.4. Resultados dos Mapas Auto-Organizáveis, e suas respectivas dimensões, com os valores das ocorrências, os números de acertos absolutos e relativos para quatro gêneros da família Asteraceae. Gêneros: Saussurea, Lactuca, Trixis e Doronicum. Mapa na figura 4.2.1.3 a. Gêneros: Carthamus, Picris, Mutisia e Petasites, mapa na figura 4.2.1.3 b.....**96**

Tabela 4.2.1.5. Resultados dos Mapas Auto-Organizáveis, e suas respectivas dimensões, com os valores das ocorrências, os números de acertos absolutos e relativos para três tribos da subfamília Asteroideae (Asteraceae) segundo Bremer (figura 4.2.1.2 c) e três tribos da subfamília Asteroideae (Asteraceae) segundo a Meta-tree(figura 4.2.1.2 d).....**97**

Tabela 4.2.1.6. Resultados dos Mapas Auto-Organizáveis, e suas respectivas dimensões, com os valores das ocorrências, os números de acertos absolutos e relativos para três gêneros da subfamília Asteroideae (Asteraceae) (figura 4.2.1.3 c) e quatro gêneros da tribo Astereae (figura 4.2.1.3 d).....**97**

Tabela 4.2.2.1. Fragmento da planilha de contagens do número de ocorrência das classes de metabólitos secundários por gêneros, gerada pelo programa Gertaxo (Gertaxo 2004).....**101**

Tabela 4.2.2.2. Média e Desvio padrão das ocorrências das doze classes de metabólitos nos gêneros da família Asteraceae.....**102**

Tabela 4.2.2.3. Classificação heurística da probabilidade relativa de ocorrência das doze classes metabólitos nos gêneros da família Asteraceae.....**102**

Tabela 4.2.2.4. Fragmento da planilha gerada com a substituição dos valores obtidos na contagem do Gertaxo (Gertaxo 2004 pelas probabilidades relativas para cada classe de composto.....**103**

Tabela 4.2.2.5. Médias dos acertos da série de treino e série de teste na previsão da variável utilizando como entrada para a rede os dados das outras 11 variáveis.....**104**

Tabela 4.2.2.6. Distribuição das cinco faixas probabilidades relativas das ocorrências nas doze classes de compostos nos 674 gêneros da família Asteraceae.....**105**

Tabela 4.3.1. Resumo dos modelos de regressão da previsão dos esqueletos 1 e 101.....**110**

Apêndice C

Tabela 5.2.1. Acertos e erros da rede supervisionada por gêneros para os sesquiterpenos lactonizados, com heurística de 0 e 1, na fase de treino.....**179**

Tabela 5.2.2. Acertos e erros da rede supervisionada por gêneros para os sesquiterpenos lactonizados, com heurística de 0 e 1, na fase de teste.....**188**

LISTA DE ABREVIATURAS E SIGLAS

ACET	acetofenonas
ANN	Rede Neural Artificial (Artificial Neural Network)
BENZF	Benzofuranos
BENZP	Benzopiranos
BMU	Best Match Unit
CUMA	Cumarinas
DITE	Diterpenos
DMPA	Pirofosfato de 3,3-dimetilalila
DNA	Ácido desoxiribonucléico
DOXP	1-desoxi-d-xilose-5-fosfato
“E”	Função Booleana E (AND)
ESQUEL	Tipo De Esqueleto
FENIL	Fenilpropanóides
FLAV	Flavonóides
GA-3P	Gliceraldeido-3-fosfato
IPP	Pirofosfato De Isopentenila
LACT	Sesquiterpenos Lactonizados
MLP	Perceptron de Múltiplas Camadas (Multi-Layer Perceptron)
MLR	Regressão Linear Múltipla (Multiple Linear Regression)
MONO	Monoterpenos
MVA	Ácido Mevalônico
NOX	Grau De Oxidação
“OU”	Função Booleana OU (OR)
PO	Passo Oxidativo

POLI	Poliacetilenos
Qcv^2	Coeficiente de predição pelo método full cross-validation
r^2	Coeficiente de correlação
$r^2(\text{adj})$	Coeficiente de correlação ajustado
REA	Reação Característica
RNA	Ácido ribonucleico
S	Índice De Especialização De Esqueletos
SESQ	Sesquiterpenos
SOM	Mapas Auto-Organizáveis (Self-Organizing Maps)
TRITE	Triterpenos
TPP-E	Pirofosfato de tiamina-enzima
VSS-GA	Variable Selection – Genetic Algorithm
“XOR”	Função Booleana XOR

SUMÁRIO

1. INTRODUÇÃO.....	22
1.1. Sistemática.....	22
1.1.1. Quimiosistemática.....	23
1.2. Família Asteraceae: Aspectos gerais.....	24
1.3. Classificação e Filogenia da Família Asteraceae.....	26
1.4. Redes Neurais.....	38
1.4.1. Breve histórico.....	38
1.4.2. Fundamentos das redes neurais.....	41
1.4.3 Perceptrons de Múltiplas Camada (MLP).....	43
1.4.4. Mapas auto-organizáveis (SOM).....	48
1.4.5 Aplicações das Redes Neurais em Química.....	51
1.5. Algoritmos Genéticos.....	52
2. Objetivos.....	55
3. MATERIAIS E MÉTODOS.....	56
3.1. Marcadores químicos.....	56
3.2. Biossíntese dos diferentes marcadores.....	56
3.3. Coleta e arquivamento dos dados.....	64
3.4. Número de ocorrência.....	65
3.5. Parâmetros evolutivos.....	66
3.5.1. Grau de oxidação (NOX).....	66
3.5.2. Passo Oxidativo (PO).....	67
3.5.3. Especialização dos esqueletos (S).....	68
3.5.4. Reação Característica (REA).....	69
3.5.5. Tipo de Esqueleto (ESQUEL).....	71
3.6. Redes Neurais.....	72
3.6.1. Rede Neural Não Supervisionada (Mapas auto-organizáveis).....	72
3.6.2. Rede Neural Supervisionada(MLP).....	73
3.7. Algoritmos Genéticos.....	74

4. RESULTADOS.....	77
4.1. Banco de Dados Químico.....	77
4.2. Redes Neurais.....	95
4.2.1. Redes Neurais não Supervisionadas.....	95
4.2.2. Redes Neurais Supervisionadas.....	101
4.3. Algoritmos Genéticos.....	110
5. DISCUSSÃO.....	114
5.1. Banco de Dados Químico.....	114
5.2. Redes Neurais.....	118
5.2.1. Redes Neurais não Supervisionadas.....	118
5.2.2. Redes Neurais Supervisionadas.....	120
5.3. Algoritmos Genéticos.....	122
6. CONCLUSÕES.....	125
7. REFERÊNCIAS.....	129
APÊNDICES.....	137
Apêndice A.....	138
Apêndice B.....	174
Apêndice C.....	179

1. INTRODUÇÃO

1.1. Sistemática

O estudo científico da diversidade biológica e sua história evolutiva são chamados de sistemática. Para Carl Linnaeus, naturalista que iniciou a moderna classificação biológica, e para os seus sucessores imediatos, o objetivo no estudo sistemático era nomear, descrever e relacionar todas as espécies de plantas, animais e minerais. Entretanto, após a publicação de origem das espécies (Darwin 1859), as diferenças e semelhanças entre indivíduos passaram a ser vistas como produto da sua história evolutiva, ou filogenia. Daí em diante os pesquisadores passaram a ambicionar classificações que fossem não apenas informativas e úteis, mas que também refletissem as relações evolutivas entre os organismos (Raven 2001).

Atualmente a Sistemática inclui a taxonomia (ciência da descoberta, descrição e classificação das espécies e grupos de espécies, seguindo normas e princípios) e também a filogenia (relações evolutivas entre os organismos) (Raven 2001).

No estudo sistemático utilizam-se caracteres dos mais diversos tipos: Fósseis, morfológicos, anatômicos, palinológicos, embriológicos, químicos (metabólitos secundários) e macromoleculares (DNA).

A importância da sistemática para a ciência e sociedade está ligada ao fato de esse campo do conhecimento fornecer respostas cada vez mais bem elaboradas sobre perguntas como: Como a vida foi originada? Como a vida mudou pelo tempo? Qual foi a origem da espécie humana e sua evolução? Questões essas que influenciam muito outros campos do conhecimento (Stuessy 1990).

1.1.1. Quimiosistemática

Quando se utiliza dados químicos isoladamente para classificar e analisar a história evolutiva de um grupo de organismos, chama-se esse estudo de Quimiosistemática (Grayer et al. 1999). A Quimiosistemática se restringe a análise de organismos onde se empregam caracteres químicos; micromoléculas (metabolitos secundários) e macromoléculas (proteínas) como caracteres.

O estudo Quimiosistemático tem mostrado grandes contribuições para a sistemática de plantas (Grayer 1999; Alvarenga et al. 2001, Calabria et al. 2006; Hegnauer 1986; Gershenson e Mabry 1983; Gotlieb et al. 1996;).

As principais limitações da quimiosistemática é que um número elevado de espécies não foi estudado quimicamente, e das espécies estudadas o principal problema é a falta de padronização dos dados. Abaixo são descritos os principais problemas:

1. Em muitos casos não são mencionadas as partes da planta que foram estudadas,
2. Os estudos são dirigidos à obtenção de novas substâncias, então muitas vezes não é relatada a presença de algumas substâncias usuais nos organismos estudados.
3. Existem falsos positivos, por identificação errônea de um dado composto, e falsos negativos, pois por muitas vezes um dado composto isolado não é mencionado nas publicações.

Na década de 90, a sistemática micromolecular (Quimiosistemática) foi eclipsada pelas filogenias moleculares (Waterman 2007). Nessa época as técnicas

de extração e identificação de DNA e RNA tornaram-se bastante eficientes e rápidas, gerando grande quantidade de dados moleculares para análise, dados esses, padronizados e menos sujeitos a erro do que os dados de micromoléculas. Assim, houve uma grande explosão no número de publicações utilizando dados moleculares.

Embora a Quimiosistemática tenha perdido espaço nos últimos anos e as filogenias moleculares tenham trazido avanço para a sistemática, cada vez mais os pesquisadores enxergam que os dados moleculares isoladamente são insuficientes para resolver todos os problemas na área (Soltis 2005). Como já foi dito, filogenia não é o mesmo que classificação.

Atualmente as filogenias moleculares produzem grupos monofiléticos, mas não estabelecem grupos taxonômicos, os quais necessitam de caracteres fenotípicos: as sinapomorfias (similaridades compartilhadas com um ancestral comum). Para a maioria dos ramos das angiospermas (APG 2003), não se dispõem ainda de sinapomorfias fenotípicas, nesse caso caracteres químicos são bem vindos. Também deve ser enfatizado que muitas vezes, caracteres moleculares isoladamente não são suficientes para a obtenção de ramos bem consolidados, e mais uma vez caracteres fenotípicos, inclusive químicos, podem ser combinados aos moleculares, a fim de aumentar a consistência de ramos, ou melhorar sua resolução.

1.2. Família Asteraceae: Aspectos gerais

A família Asteraceae ou Compositae é uma das maiores famílias de plantas florescentes no mundo. Foram descritas, aceitadas botanicamente, nessa família, aproximadamente 24.000 espécies (Heywood e Harbone 1977; Harbone 1988,

Ferreira et al. 1982, Gotlieb 1989, 1990, Emerenciano et al. 1985, 1987, 1998a, 1998b, Bremer 1994; Funk et al. 2009). Sua distribuição é vasta e envolve os mais variados ecossistemas, é uma família de plantas cosmopolita (Brant 2003). Existem 1600-1700 gêneros distribuídos por todo o mundo, exceto na Antártica (funk et al. 2009). Asteraceae é chamada mais comumente de “a família do girassol” ou “a família da margarida” (Herman et al. 2000), As plantas da família têm grande importância econômica, incluindo alimentos, plantas ornamentais, ervas medicinais e plantas com propriedades inseticidas (funk et al. 2009). Algumas espécies de Asteraceae e os seus usos estão mostrados na tabela 1.2.1.

Tabela 1.2.1. Exemplos de espécies de Asteraceae e principais usos.

Nome científico	Nome comum	Uso
<i>Lactuca sativa</i>	Alface	Alimentar
<i>Helianthus annuus</i>	Girassol	Alimentar (óleo)
<i>Tanacetum cinerariifolium</i>	Piretro	Inseticida
<i>Artemisia annua</i>	Artemisia	Medicinal (Febre, dores de cabeça)
<i>Grindelia robusta</i>	Mal-me-quer do campo	Medicinal (Inflamação do trato respiratório)
<i>Chrysanthemum morifolium</i>	Crisântemo	Ornamental
<i>Chrysanthemum leucanthemum</i>	Margarida	Ornamental

A família é monofilética, ou seja, todos os representantes da família derivam de um ancestral comum e incluem todos os seus descendentes. Muito cedo nos trabalhos de classificação as espécies da família já foram consideradas como um grupo e em todos os tipos de análises realizadas com a família o resultado encontrado é de que a família Asteraceae é monofilética (funk et al. 2009).

Asteraceae é caracterizada por flores arranjadas em uma estrutura chamada capítulo, que consiste de muitas flores individuais pequenas, chamadas flósculos; por anteras fundidas em um anel com o pólen sendo empurrado ou forçado pelo

estilo e pela presença de cipselas (frutos secos). Essas estruturas estão mostradas na figura 1.2.1.

Figura 1.2.1. Caracteres da família Asteraceae. A) Capítulo, B) Anteras fundidas, C) Tipos de cipselas (funk et al. 2005).

Embora a família seja bem definida há grande variabilidade entre os seus membros. O hábito varia de ervas anuais e perenes para subarbustos e arbustos, ocorrendo também ervas anuais, lianas e árvores (Bremer 1994; Funk et al. 2009).

1.3. Classificação e Filogenia da Família Asteraceae

A família Asteraceae está localizada em posição avançada entre as angiospermas. Na filogenia mais recente apresentada pelo grupo de filogenia das Angiospermas (APG 2003) a família aparece no ramo Asterales que se localiza no ramo Euasterides I, que por sua vez é parte do ramo Asterids que faz parte do ramo das Eudicotiledôneas (figuras 1.3.1 – 1.3.2).

Figura 1.3.1. Interrelações de ordens e algumas famílias do ramo das angiospermas (APG 2003).

Figura 1.3.2. Relações entre as famílias do ramo euasteride II (campanulid). Destaque para a família Asteraceae no ramo Asterales.

A família tem sido alvo de estudos já há alguns séculos. Na primeira metade do século XVI Jean Ruel apresentou a descrição de muitas espécies de Compositae, identificando claramente o capítulo como sendo composto de muitos tipos de pequenas flores (Funk et al. 2009).

Entretanto, uma compreensão profunda da verdadeira natureza das Compostas não chegou antes do século XVII (na era chamada pré-Cassini) com os trabalhos de Joseph Pitton de Tournefort, Sébastien Vaillant e Johann Le Francq van Berkhey (Funk et al. 2009).

Tournefort entendeu a família como um grupo e a dividiu em três famílias; Vaillant dividiu as Compostas também em três grupos, em adição à Cichorieae reconhecida por Tournefort, ele identificou Cardueae e o restante das espécies formou o grupo conhecido como “Corymbifères”. Por último, Berkhey é importante por ser o primeiro a escrever um livro (*Expositio Characteristica Structurae florum qui Dicuntur Compositi* (1760)) sobre as características dos gêneros de Asteraceae (Funk et al. 2009).

A classificação das Asteraceae mais próxima das classificações modernas foi criada pelo botânico Alexandre-Henri-Gabriel de Cassini (1781-1832) que é considerado o fundador da classificação da família Asteraceae. Em 1816 Cassini publicou um diagrama onde são mostradas as inter-relações das 19 tribos em que a família foi dividida (Cassini 1816) (figura 1.3.3), e em 1821 ele reduziu a tribo Xeranthemeae para uma sinonímia de Carlineae e propôs duas novas tribos, Tageteae e Nassauvieae (Bremer 1994, Funk et al. 2009).

Christian Friedrich Lessing em 1832 (Lessing 1832) publicou um livro onde apresenta sua classificação da família, que inclui sete tribos. Em 1873 George Bentham apresentou uma revisão da classificação tribal da família em “Genera

Plantarum” (Bentham 1873), em grande parte foi um retorno a classificação de Cassini, na qual a família é dividida em 13 tribos.

Em “Die Natürlichen Pflanzenfamilien”, Karl August Hoffmann (Hoffmann 1890) apresenta sua classificação das Asteraceae, em essência ele repete o trabalho de Bentham com pequenas alterações. As classificações de Bentham e Hoffmann encontram-se na figura 1.3.4.

Figura 1.3.3. Diagrama de inter-relações entre as 19 tribos de Asteraceae segundo Cassini 1816. Synanthérés = Asteraceae, Boopidéas = Calyceraceae.

Bentham (1873)	Hoffmann (1890)
Vernoniaceae	Vernonieae
Eupatoriaceae	Eupatorieae
Asteroideae	Astereae
Inuloideae	Inuleae
Helianthoideae	Heliantheae
Helenioideae	Helenieae
Anthemideae	Antemideae
Senecionideae	Senecioneae
Calendulaceae	Calenduleae
Arctotideae	Arctotideae
Cynaroideae	Cynareae
Mutisieaceae	Mutisieae
Cichoriaceae	Cichorieae

Figura 1.3.4. Classificações da família Asteraceae de Bentham e Hoffmann. A tribo Cynaroideae/Cynareae atualmente é chamada de Cardueae e Cichoriaceae/Cichorieae é agora a tribo Lactuceae

No século XX o trabalho de classificação continuou avançando. A tabela 1.3.1 mostra as principais classificações e filogenias propostas para a família da segunda metade do século XX até meados do século XXI.

A partir de Wagenitz e Carlquist passou-se a classificar a família em duas subfamílias (Wagenitz 1976; Carlquist 1976). Com Bremer (Bremer 1994) e Kin e Jansen (Kin e Jansen 1995,1996) a família passou a ser dividida em três subfamílias, em Bremer (Bremer 1996) a família foi dividida em cinco grupos, quatro subfamílias e a tribo Mutisieae e finalmente na filogenia mais atual (Funk et al. 2009) a família apresenta-se dividida em doze subfamílias.

Para a construção das classificações e filogenias mencionadas na tabela 1.3.1 os caracteres utilizados foram quase que exclusivamente caracteres morfológicos, exceto nas filogenias de Kin e Jansen (1995, 1996) e Funk (Funk et al.2009), em que utilizaram-se exclusivamente caracteres moleculares para construir as filogenias.

Tabela 1.3.1. Principais classificações e filogenias recentes da família Asteraceae (Modificado de Bremer 1992).

[illegible]

Tabela 1.3.1. Continuação.

Bremer et al. (1992)	Bremer (1994)	Bremer (1996)	Kin e Jansen (1995, 1996)	Funk et al. (2009)	Cont. Funk et al. (2009)
Barnadesioideae	Barnadesioideae	Barnadesioideae	Barnadesioideae	Barnadesioideae	Cichorioideae
Barnadesieae	Barnadesieae	Barnadesieae	Barnadesieae	Barnadesieae	Cichorieae
					Platycarpeae
Cichorioideae	Cichorioideae	Mutisieae	Cichorioideae	Stifftioideae	Liabeae
Mutisieae	Mutisieae		Mutisieae	Stifftieae	Moquinieae
Cardueae	Cardueae	Carduoideae	Tarchonantheae		Vernonieae
Lactuceae	Lactuceae	Cardueae	Cardueae	Wunderlichioideae	Eremothamneae
Vernonieae	Vernonieae		Liabeae	Hyalideae	Arctotideae
Liabeae	Liabeae	Cichorioideae	Vernonieae	Wunderlichieae	
Arctoteae	Arctoteae	Lactuceae	Arctoteae		Corymbioideae
		Vernonieae	Lactuceae	Mutisoideae	Corymbieae
Asteroideae	Asteroideae	Liabeae		Onoserideae	
Inuleae	Inuleae	Arctoteae	Asteroideae	Naussauvieae	Asteroideae
Astereae	Plucheeae		Senecioneae	Mutisieae	Senecioneae
Anthemideae	Gnaphalieae	Asteroideae	Anthemideae		Calenduleae
Senecioneae	Calenduleae	Inuleae	Astereae	Gochnatioideae	Gnaphalieae
Calenduleae	Astereae	Astereae	Calenduleae	Gochnatieae	Astereae
Helenieae	Anthemideae	Anthemideae	Gnaphalieae		Anthemideae
Coreopsideae	Senecioneae	Senecioneae	Inuleae	Hecastocleidoideae	Inuleae
Tageteae	Helenieae	Calenduleae	Athroisma group	Hecastocleideae	Athroismeae
Heliantheae	Heliantheae	Helenieae	Heliantheae		Feddeae
Eupatorieae	Eupatorieae	Coreopsideae		Carduoideae	Helenieae
		Tageteae		Tarchonantheae	Coreopsideae
		Heliantheae		Oldenburgieae	Neurolaeneae
		Eupatorieae		Dicomeae	Tageteae
				Cardueae	Chaenactideae
					Bahieae
				Pertyoideae	Polymnieae
				Pertyeae	Heliantheae
					Millerieae
				Gymnarrhenoideae	Peritytileae
				Gymnarrheneae	Euparotieae
					Madieae

• Cont. = Continuação

No presente trabalho foram utilizadas as filogenias de Bremer (Bremer 1994a, 1996) e de Funk (Funk et al. 2009). Assim sendo, as duas filogenias serão discutidas com mais detalhe abaixo.

A filogenia da família segundo Bremer, descrita na figura 1.3.5, divide-a em quatro subfamílias mais a tribo Mutisieae, as subfamílias são: Barnadesoideae (1 tribo), Carduoideae (1 tribo), Cichorioideae (7) e Asteroideae (8). Assim o total de tribos na família é de 17 (tabela 1.3.1).

A árvore de inter-relações de Bremer foi quase que exclusivamente construída com base em dados morfológicos, em menor escala foram utilizadas informações moleculares (inversão de DNA do cloroplasto e duplicação no gene *rbcL*), e dados químicos. A explicação dada pelo autor para a preferência de dados morfológicos (Bremer 1994) foi a de que na época que o trabalho estava sendo realizado menos que 1% das espécies tinha dados moleculares disponíveis para realizar as análises, quanto aos dados químicos, já que existe abundante dados dessa classe, a explicação foi de que é difícil interpretar esses dados no contexto filogenético, uma vez que a presença do composto é reportada, mas raramente sua ausência.

Na filogenia de Funk, a primeira Super-árvore (=Meta-tree) para Compositae foi publicada em 2005 (Funk et al. 2005), e desde de então houve um considerável progresso na reconstrução das relações evolucionárias em muitos ramos. Na filogenia publicada em 2009 (funk et al. 2009), a qual a partir desse ponto será chamada somente por Meta-tree, há filogenias mais robustas para a maioria dos ramos da família, essa filogenia foi desenvolvida usando uma compilação de árvores. A árvore de base da Meta-tree foi formada da filogenia da família publicada em 2008 (Panero e Funk 2008).

Como mencionado anteriormente, segundo a filogenia da Meta-tree a família foi dividida em doze subfamílias (tabela 1.3.1). As subfamílias são: Barnadesioideae (1 tribo), Stifftioideae (1 tribo), Wunderlichioideae (2 tribos), Mutisioideae (3 tribos), Gochnatioideae (1 tribo), Hecastocleideae (1 tribo) Carduoideae (4 tribos), Pertyoideae (1 tribo), Gymnarrhenoideae (1 tribo) Cichorioideae (10 tribos), Corymbioideae (1 tribo) e Asteroideae (20 tribos). A figura 1.3.6 apresenta um diagrama resumido da filogenia da família Asteraceae segundo a Meta-tree.

As correspondências entre as filogenias de Bremer e da Meta-tree são mostradas na tabela 1.3.2, assim como o código de três letras utilizado para representar cada uma das tribos.

Figura 1.3.5. Diagrama de inter-relações filogenéticas das tribos da família Asteraceae, segundo Bremer [1].

Figura 1.3.6. Diagrama resumido da análise filogenética das tribos da família Asteraceae segundo Meta-tree (adaptado de Funk et al. 2009).

Tabela 1.3.2. Tribos da família Asteracea segundo as filogenias de Bremer e a da Meta-tree, e os respectivos códigos de três letras utilizados para representar cada uma das tribos.

Tribos segundo Bremer	Código	Tribos segundo Metatree	Código
ANTHEMIDEAE	ANT	ANTHEMIDEAE	ANT
ARCTOTEAE	ARC	ARCTOTIDEAE	ARC
ARCTOTEAE	ARC	PLATYCARPHEAE	PLA
ASTEREAE	AST	ASTEREAE	AST
BARNADESIEAE	BAR	BARNADESIEAE	BAR
CALENDULEAE	CAL	CALENDULEAE	CAL
CARDUEAE	CAR	CARDUEAE	CAR
EUPATORIEAE	EUP	EUPATORIEAE	EUP
GNAPHALIEAE	GNA	GNAPHALIEAE	GNA
HELENIEAE	HLN	ATHROISMEAE	ATH
HELENIEAE	HLN	BAHIEAE	BAH
HELENIEAE	HLN	CHAENACTIDEAE	CHA
HELENIEAE	HLN	HELENIEAE	HLN
HELENIEAE	HLN	MADIEAE	MAD
HELENIEAE	HLN	PERITYTILEAE	PER
HELENIEAE	HLN	TAGETAE	TAG
HELIANTHEAE	HLT	COREOPSIDEAE	COR
HELIANTHEAE	HLT	HELIANTHEAE	HLT
HELIANTHEAE	HLT	MILLERIEAE	MIL
HELIANTHEAE	HLT	NEUROLAENEAE	NEU
HELIANTHEAE	HLT	POLYMNIEAE	POL
INULEAE	INU	INULEAE	INU
LACTUCEAE	LAC	CICHORIEAE	CIC
LIABEAE	LIA	LIABEAE	LIA
MUTISIEAE	MUT	DICOMEA	DIC
MUTISIEAE	MUT	GOCHNATIEAE	GOC
MUTISIEAE	MUT	HYALIDEAE	HYA
MUTISIEAE	MUT	MUTISIEAE	MUT
MUTISIEAE	MUT	NAUSSAUVIEAE	NAU
MUTISIEAE	MUT	OLDENBURGIEAE	OLD
MUTISIEAE	MUT	ONOSERIDEAE	ONO
MUTISIEAE	MUT	PERTYAE	PET
MUTISIEAE	MUT	STIFFTIEAE	STI
MUTISIEAE	MUT	TARCHONANTHEAE	TAR
MUTISIEAE	MUT	WUNDERLICHIEAE	WUN
NÃO CLASSIFICADO	UNA	CORYMBIEAE	COY
NÃO CLASSIFICADO	UNA	EREMOTHAMNEAE	ERE
NÃO CLASSIFICADO	UNA	FEDDEAE	FED
NÃO CLASSIFICADO	UNA	GYMNARRHENEAE	GYM
NÃO CLASSIFICADO	UNA	HECASTOCLEIDEAE	HEC
SENECIONEAE	SEN	SENECIONEAE	SEN
VERNONIEAE	VER	MOQUINIEAE	MOQ
VERNONIEAE	VER	VERNONIEAE	VER

Como exposto acima, a família foi extensamente estudada, não só com a finalidade de entender as relações entre as espécies, gêneros e tribos que a constituem, mas também com relação a sua importância econômica (alimentos, ervas medicinais, inseticidas, etc.). Como consequência, uma grande quantidade de informação sobre a família foi gerada, e estas estão disponíveis na literatura (Bremer et al. 1992; Bremer 1994; Ito et al. 2000; Zdero e Bohlmann 1993; Funk et al. 2009).

Dentre os dados obtidos estão as informações sobre a produção de metabólitos secundários. Para o presente estudo foi utilizado a abundância de dados sobre metabólitos secundários para estudos Quimiosistemáticos na família.

1.4. Redes Neurais

1.4.1. Breve histórico

O advento da era moderna das redes neurais teve início com o trabalho pioneiro de McCulloch e Pitts (McCulloch et al. 1943). Pitts era um matemático prodigioso e McCulloch era psiquiatra, ambos se associaram em 1942. No artigo publicado em 1943 os autores descrevem um cálculo lógico das redes neurais que unificava os estudos de neurofisiologia e da lógica matemática.

O neurônio de McCulloch e Pitts era um dispositivo binário: a sua saída poderia ser pulso ou não pulso, e as suas entradas tinham ganhos arbitrários e poderiam ser excitatórias ou inibitórias. Para determinar a saída do neurônio, calculava-se a soma ponderada das entradas com os respectivos ganhos como fatores de ponderação, positivo nos casos excitatórios e negativo nos casos

inibitórios. Caso o valor fosse maior ou igual a um determinado limiar, a saída do neurônio seria pulso, e caso contrário a saída seria não pulso (Kovács 2002).

O próximo significativo desenvolvimento no campo veio com a publicação do livro de Hebb “The Organization of Behavior” (Hebb 1949), no qual foi apresentado pela primeira vez uma formulação explícita de uma regra de aprendizagem fisiológica para a modificação sináptica.

Cerca de 15 anos após a publicação do artigo de McCulloch e Pittis, uma nova abordagem para o problema de reconhecimento de padrões foi introduzida por Rosenblat (Rosenblat 1958) em seu trabalho sobre o perceptron de camada simples, um método inovador de aprendizagem supervisionada (perceptron de camada simples é a forma mais simples de uma rede neural usada para classificar padrões linearmente separáveis).

Em 1960 Widrow e Hoff introduziram o algoritmo do mínimo quadrado médio (LMS. Least Mean-Square) e o usaram para formular o Adaline (adaptive linear element, em português: elemento linear adaptativo) (Haykin 2001). A diferença entre esse último sistema e o perceptron de Rosenblat está no fato de que no Adaline o aprendizado é construído em torno de um neurônio linear, enquanto no perceptron é construído em torno de um neurônio não linear.

Em 1969 foi escrito um livro (Minsk e Papert 1969) sobre as limitações do perceptron de camada simples, ressaltando que essas limitações não deveriam ser resolvidas nos sistemas de camadas múltiplas. As limitações eram que os perceptrons de camada simples só podiam classificar funções linearmente separáveis, do tipo “E” e “OU”, não sendo capazes de separar a função do tipo “XOR” ou exclusivo, que não é uma função linearmente separável (exemplos dessas

funções estão na figura 1.4.1.1). A obra publicada desestimulou a pesquisa na área de redes neurais até meados da década de 80.

Figura 1.4.1.1. Funções Booleanas de duas variáveis representadas no plano binário. O ponto escuro significa que ali a saída deve ser 1, enquanto no ponto claro a saída é nula (Kovács 2002).

Nos anos 80, importantes contribuições foram feitas em várias frentes à teoria e ao projeto de redes neurais, e com isso houve o ressurgimento do interesse nessa área do conhecimento.

Em 1982 Hopfield (Hopfield 1982) utilizou uma abordagem do campo da física estatística que despertou o interesse para a modelagem de redes neurais por parte dos cientistas. Também em 1982 foi publicado o artigo de Teuvo Kohonen sobre os mapas auto-organizáveis (Kohonen, 1982).

Em 1986 houve um grande aumento nas pesquisas em redes neurais por influência do livro "Parallel Distributed Processing" (Processamento Distribuído Paralelamente) editado por David Rumelhart e James McClelland (Rumelhart e McClelland 1986). Nesse livro foi relatado o algoritmo de retropropagação (back-propagation), que emergiu como o algoritmo de aprendizagem mais popular para o treinamento de perceptrons de múltiplas camadas (Haykin 2001).

As redes neurais certamente trilharam um longo caminho desde McCulloch e Pitts, e atualmente estão estabelecidas em diversas áreas.

1.4.2. Fundamentos das redes neurais

O cérebro humano é um sistema de processamento paralelo altamente complexo e não-linear, capaz de realizar tarefas como reconhecimento de padrões, percepções e coordenação motora, muito mais agilmente que o mais rápido dos computadores digitais existentes hoje. (Haykin 2001).

A neurocomputação estuda a simulação computacional do cérebro humano, fugindo das características da computação convencional, simulando a estrutura composta por neurônios, suas interconexões e os estímulos que, através da sinapse, passam de um neurônio para outro. O sistema resultante dessa simulação é denominado Rede Neural Artificial (ANN).

Um neurônio biológico, como mostra a figura 1.4.2.1, é basicamente composto por três partes: dendritos, corpo celular (ou soma) e axônio. Em uma rede neural biológica, impulsos elétricos são transferidos de uma célula para outra através de um processo químico de conexão denominado sinapse, que ocorre entre o axônio de um neurônio e dendritos de outros neurônios vizinhos.

Figura 1.4.2.1. Ilustração do neurônio biológico (adaptado de Junqueira e Carneiro 2004).

As ANNs são sistemas computacionais que têm como unidades de processamento neurônios artificiais, os quais são interconectados através de pesos sinápticos.

Segundo Lawrence (Lawrence 1994), existem dois tipos principais de idéias nas arquiteturas das redes neurais, uma delas apresenta alimentação recorrente (feedback) e a outra apresenta alimentação direta (feedforward) (figura 1.4.2.2).

No escopo desse trabalho serão utilizadas somente duas arquiteturas das mostradas na figura 1.4.2.2, perceptrons de múltiplas camadas e redes de kohonen, que serão detalhas abaixo.

Figura 1.4.2.2. Classificação de algumas arquiteturas de redes neurais (adaptado de Lawrence 1994).

As ANNs são treinadas com o intuito de ajustar os pesos sinápticos dos neurônios, de forma a definir o melhor conjunto de pesos que minimize o erro. O treinamento pode ser realizado de forma supervisionada ou não supervisionada, sendo que no primeiro tipo, além do conjunto de parâmetros de entrada, é apresentado também o respectivo conjunto de parâmetros desejados na saída.

A configuração de uma ANN, no que diz respeito às quantidades de neurônios nas camadas de entrada e saída, é definida de acordo com o número de parâmetros que se deseja apresentar na entrada e o número de respostas que se deseja obter na saída, respectivamente. O número de neurônios na(s) camada(s) oculta(s) está associado à capacidade de generalização da rede.

1.4.3 Perceptrons de Múltiplas Camadas (MLP)

As MLP são redes de treinamento supervisionado que utilizam comumente aprendizagem por retropropagação, possuem duas ou mais camadas, sendo uma de saída e uma ou mais intermediárias (ou ocultas), a camada de entrada não é computada, pois não realiza nenhum processamento.

A figura 1.4.3.1 ilustra uma MLP com três camadas: duas ocultas, uma com três neurônios e outra com dois neurônios, e a de saída com quatro neurônios. Alguns autores consideram a camada de entrada para computar o número de camadas da rede, segundo Kovács 2002 se for mantida a referência explícita às dimensões mencionadas, passa a ser irrelevante qual é a convenção adotada quanto ao número de camadas da rede. Assim, a rede na figura 1.4.3.1 pode ser descrita como uma rede de quatro camadas, uma camada de entrada, duas ocultas e uma camada de saída.

Figura 1.4.3.1. Representação da rede do tipo Perceptrons de Múltiplas Camadas (adaptado de Zupan e Gasteiger 1993).

A figura 1.4.3.2 ilustra os elementos envolvidos na propagação de um sinal em um neurônio k de uma MLP. Este processo consiste, resumidamente, na soma ponderada das entradas x_n pelos pesos w_{kn} , com o acréscimo do bias b_k , gerando v_k , que é ativada pela função de ativação φ , gerando a saída y_k .

Dependendo da saída deste processo, o neurônio k decide se dispara um sinal y_k ou não. O sinal disparado pode ser 1, 0, ou pode ser um valor real entre 1 e 0; dependendo se estamos trabalhando com valores binários ou reais (Zupan et al. 1993).

FIGURA 1.4.3.2. Modelo não linear de neurônio (Haykin 2001).

O bias, ou viés, têm o efeito de aumentar ou diminuir a entrada líquida da função de ativação, o que depende se o seu valor é positivo ou negativo, respectivamente.

A função de ativação, ou função de transferência, é aplicada sobre a ativação interna de um neurônio, produzindo o sinal de saída do neurônio em questão. Abaixo e na figura 1.4.3.3 são apresentados três tipos básicos de função de ativação:

1. Função limiar:
$$\varphi(v) = \begin{cases} 1 & \text{se } v \geq 0 \\ 0 & \text{se } v < 0 \end{cases}$$

2. Função rampa: $\varphi(v) = \begin{cases} 1, & v \geq +\frac{1}{2} \\ v, & -\frac{1}{2} < v < +\frac{1}{2} \\ 0, & v \leq -\frac{1}{2} \end{cases}$

3. Função sigmoidal: $\varphi(v) = \frac{1}{1 + \exp(-av)}$

Figura 1.4.3.3. Tipos de função de ativação. a) função de limiar, b) função rampa, c) função sigmóide para parâmetro de inclinação a variável (Haykin, 2001).

As funções sigmoidais são as mais utilizadas em ANNs (Haykin, 2001). A função sigmoidal apresentada acima é uma função do tipo logística.

O processamento de um neurônio do tipo perceptron (figura 1.4.3.2), consiste em quatro etapas:

1. Apresentação das entradas: as entradas x_i podem ser os parâmetros de entrada da rede (se o neurônio estiver na primeira camada oculta) ou as saídas dos neurônios da camada anterior (se o neurônio estiver em uma camada oculta mais interna ou na camada de saída).

2. Ativação interna: é realizada uma soma ponderada das m entradas apresentadas ao neurônio k , de acordo com o conjunto de pesos w_{ki} definido para aquele neurônio, da forma:

$$v_k = \sum_{i=1}^m w_{ki} X_i \quad \text{Equação 1.4.3.1}$$

3. Função de ativação: a ativação interna v_k , somada ao bias b_k , passa pela função ϕ , que gera o valor de saída y_k do neurônio k :

$$y_k = f(v_k + b_k) \quad \text{Equação 1.4.3.2}$$

4. Propagação da saída: a saída y_k é propagada para os neurônios da próxima camada (no caso deste neurônio pertencer a uma camada oculta), ou a saída y_k é também a saída da rede (no caso do neurônio pertencer à camada de saída).

O treinamento supervisionado é caracterizado pela apresentação do conjunto de saídas desejadas à rede. O aprendizado é feito em ciclos ou épocas chamados “*epochs*”, ou seja, define-se uma época como a apresentação completa do conjunto de padrões à rede. Cada ciclo corresponde a um período mínimo no qual todos os pares de entradas e saídas são apresentados uma vez para a rede.

No algoritmo de retropropagação de erros, os valores desejados são comparados aos calculados pela rede. Esse algoritmo está resumido abaixo e as equações e mais detalhes do algoritmo podem ser encontrados em Haykin 2001 e Zupan e Gasteiger 1993.

1. Inicialização do conjunto de pesos: os conjuntos de pesos sinápticos e bias, que envolvem os sub-conjuntos relativos a cada neurônio da rede, são inicializados com valores aleatórios entre -1 e 1;

2. Apresentação do conjunto de treinamento: os padrões de entrada são apresentados, podendo cada padrão ser um valor único ou um vetor de valores. Um padrão é apresentado por vez e este propaga para frente (forward) e em seguida retro-propaga.
3. Propagação do sinal de entrada: o primeiro padrão é apresentado aos neurônios da primeira camada, é ativado e apresentado à camada seguinte até que seja recebido e processado pelos neurônios da camada de saída e seja calculado o erro.
4. Retropropagação do erro: o erro calculado na camada de saída serve de base para o cálculo dos gradientes locais dos neurônios de saída e dos ocultos.
5. Ajuste dos pesos: os gradientes locais possibilitam a correção dos pesos dos neurônios de cada camada;
6. Iterações subseqüentes: os passos anteriores são realizados para todos os padrões de entrada e são repetidos por mais iterações até que o critério de parada - número máximo de iterações ou erro mínimo tolerável - seja atingido.
7. Conjunto ideal de pesos: o conjunto ideal de pesos, gerados na iteração que retornou o menor erro durante o treinamento, é armazenado para utilização na fase de testes.

A figura 1.4.3.4 mostra um resumo do algoritmo de retropragação de erros.

```

do while error > acceptable-error
  read-input-vector();           forward pass
  read-target-output();

  calculate-actual-output();
  calculate-errors();           reverse pass
  adjust-weights();
end-do

```

Figura 1.4.3.4. Resumo do algoritmo de retropropagação de erros (error backpropagation) (Corne 2006).

1.4.4. Mapas auto-organizáveis (SOM)

Um mapa auto-organizável (Self-organizing map, SOM) tem como principal objetivo o de transformar um sinal de dimensão elevada em um mapa discreto uni- ou bidimensional e realizar esta transformação de maneira topologicamente ordenada (Haykin 2001).

Esse tipo de rede é também chamado de rede de Kohonen em homenagem ao seu criador Teuvo Kohonen (Kohonen, 2001). Na rede de Kohonen o tipo de aprendizado é não supervisionado, ou seja, nenhum “professor” é envolvido, a rede é exposta a um número de entradas e se organiza de modo a fazer suas próprias classificações unicamente com base nesses dados. A aprendizagem não supervisionada pode ser usada como módulo de “descoberta de características” que precede a aprendizagem supervisionada.

Em analogia ao que ocorre nas MLP, a cada uma das ligações entre as camadas de entrada e saída é atribuído um valor de peso sináptico, como pode ser observado na figura 1.4.4.1. Nota-se que cada padrão de entrada é formado por um vetor de m parâmetros e que a camada de saída possui n neurônios, ou *clusters*. Pode-se denotar o peso de cada ligação como w_{ji} , em que i e j identificam, respectivamente, o número do parâmetro da entrada e o neurônio de saída.

Figura 1.4.4.1. Exemplo de rede de kohonen (SOM) unidimensional.

O treinamento da rede de kohonen pode ser feito pelo método padrão ou pelo método em lote (batch training). No treinamento pelo método padrão, também chamado de método seqüencial o processo de treinamento da rede começa com a inicialização dos pesos sinápticos, atribuindo-lhes valores aleatórios. Em seguida, três etapas são realizadas: competição, cooperação e adaptação sináptica (Haykin 2001).

1. Competição: consiste em determinar um neurônio vencedor K , que melhor represente um dado padrão de entrada x_i . O critério utilizado é o cálculo da distância euclidiana entre o valor da entrada e os pesos de cada neurônio de saída. O neurônio do qual a entrada for menos distante, é considerado o vencedor. A distância euclidiana d_j entre a entrada x_i e o vetor de pesos do neurônio j é dada pela equação 1.4.4.1 em que i é o número do parâmetro de entrada.

$$d_j = \sqrt{\sum_{i=1}^m (x_i - w_{ji})^2} \quad \text{Equação 1.4.4.1}$$

2. Cooperação: fase realizada quando se é pré-definida uma vizinhança topológica, ou seja, neurônios que estão a um raio R do neurônio vencedor e cujos pesos serão atualizados juntamente aos do neurônio vencedor. A vizinhança é definida pela equação 1.4.4.2, abaixo:

$$h_{ci}(t) = \exp(-\|r_c - r_i\|^2 / 2\sigma^2(t)) \quad \text{Equação 1.4.4.2}$$

$$\sigma(t) = \sigma(0) \cdot \exp(-t/\tau_i) \quad \text{Equação 1.4.4.3}$$

Onde: τ_i é uma constante.

3. *Adaptação sináptica*: última etapa, atualiza os pesos dos neurônios selecionados, caso seja considerada vizinhança $R=0$, da forma:

$$W_{ji}^{(old)} = W_{ji}^{(new)} + a(x_i - W_{ji}^{(old)}) \quad \text{Equação 1.4.4.4}$$

Com isso diminui-se a distância dessas unidades com o padrão de entrada.

As três etapas são repetidas até serem apresentados todos os padrões do vetor de entrada e durante um número máximo de interações ou outro critério de parada.

No treinamento em lote, por sua vez, é apresentada toda série de dados ao mapa antes que qualquer ajuste seja feito. Em cada treinamento, dividi-se a série de dados em concordância com as regiões de Voronoi dos vetores de peso do mapa, que significa que cada vetor de dados pertence a uma série de dados do mapa da unidade ao qual está mais próximo. A equação 1.4.4.5 mostra como os pesos são calculados (Vesanto et al. 1999; Scotti 2008).

$$m(t+1) = \frac{\sum_{j=1}^n h_{ic(j)}(t)x_j}{\sum_{j=1}^n h_{ic(j)}(t)} \quad \text{Equação 1.4.4.5}$$

Onde: $c(j)$ é o neurônio do vetor da amostra x_j , $h_{ic(j)}$ é a função de vizinhança aqui utilizado como um fator de ponderação, e n é o número de vetores das amostras;

Em geral a determinação dos parâmetros de aprendizagem é empírica. A dimensionalidade do mapa auto-Organizável e seu tamanho (m) dependerão do tipo de problema e propósito a serem tratados. A literatura mostra que a determinação do tamanho do SOM é um processo empírico (Kohonen, 2001). Em geral, o SOM bidimensional $N \times M$ é utilizado devido sua capacidade de projeção dos dados de dimensão p num Mapa bidimensional que pode ser facilmente interpretado. Para

grandes volumes de dados, em geral mapas maiores são mais adequados. Todavia, grandes mapas comprometem o desempenho do algoritmo e por sua vez mapas muito pequenos comprometem a integridade da formação topológica do SOM (Kohonen 2001).

1.4.5. Aplicações das Redes Neurais em Química

As redes neurais encontram ampla área de atuação nos mais diversos campos do conhecimento. Na área de química estão colocados alguns exemplos abaixo da utilização de MLP com algoritmo de retropropagação e das redes de Kohonen, que ilustram a importância das redes neurais como ferramenta para extrair informação de grande quantidade de dados.

- MLP: Previsão de ocorrência de metabólitos (Ferreira et al. 2004), identificação de esqueletos (Emerenciano et al. 2006a), eletroquímica (Lvova et al. 2008), análise de limonóides em Meliaceae (Fraser et al. 1997), análise de espectros no infravermelho (Cleva et al. 1999), em RMN ^{13}C (Doucet et al. 1993) e em identificação quantitativa de mistura de gases (Gulbag e Temurtas 2006).
- SOM: Quimiotaxonomia da família Asteraceae (Da Costa et al. 2005; Hristov et al. 2007; Correia et al. 2008), em relações entre estrutura química, atividade biológica (Gasteiger et al. 2003; Wagner et al. 2006; Fernandes et al. 2008), classificação de metabólitos (Gupta & Aires-de-Souza 2007) e previsão de esqueletos diterpênicos (Emerenciano et al. 2006b).

1.5. Algoritmos Genéticos

Algoritmos evolucionários são técnicas de busca de um ramo de busca chamado de “Técnicas Aleatórias-Guiadas”, que utilizam componentes aleatórios com informações do estado corrente para guiar a pesquisa. Esses algoritmos usam modelos computacionais dos processos naturais de evolução como uma ferramenta para resolver problemas.

Todos os modelos computacionais propostos usam o conceito de simulação da evolução das espécies através de seleção, mutação e reprodução, processos que dependem do desempenho dos indivíduos desta espécie dentro do “ambiente” (Linden 2006). Como se pode ver na figura 1.5.1 os algoritmos evolucionários se dividem em vários tipos distintos. Um dos tipos são os algoritmos genéticos.

Figura 1.5.1. Diagrama que posiciona os algoritmos genéticos como técnica de busca (Linden 2006).

A história dos algoritmos genéticos tem início na década de 40 na criação do ramo de inteligência artificial. Na década de 60 Rechenberg desenvolveu as estratégias evolucionárias na tentativa de usar processos evolutivos para resolver problemas (Rechenberg, 1965).

Entretanto, apesar de não ser o primeiro a trabalhar na área, o inventor dos algoritmos genéticos foi John Holland que estudou formalmente a evolução das espécies e propôs um modelo heurístico computacional que quando implementado poderia oferecer boas soluções para problemas insolúveis quando tratados com outros métodos (Linden 2006).

Os algoritmos genéticos funcionam mantendo uma população de estruturas, denominadas indivíduos ou cromossomos, que se comportam de forma semelhante à evolução das espécies. A estas estruturas são aplicados os chamados operadores genéticos, como crossing-over e mutação, entre outros. Cada indivíduo recebe uma “avaliação” que é uma quantificação da sua qualidade para a solução do problema em questão. Com base nesta avaliação serão aplicados os operadores genéticos de forma a simular a sobrevivência do mais apto. Tais operadores são aproximações computacionais de fenômenos naturais, como mutação, crossing-over e reprodução sexuada (Linden 2006).

Algoritmos genéticos trabalham com uma grande população de pontos, sendo uma heurística de busca no espaço de soluções, que se diferencia dos métodos de busca enumerativa por não procurar em todos os pontos possíveis, mas sim subconjuntos desses pontos. Além disso, os algoritmos genéticos diferenciam-se dos métodos de busca puramente aleatórios por fazerem uma busca que utiliza informação pertinente ao problema e não trabalhar com caminhadas aleatórias pelo espaço de soluções, mas sim direcionando sua busca através do mecanismo de seleção (carvalho 2003).

Tais algoritmos são usados com sucesso em problemas como o do “caixeiro viajante”, em que o número de soluções é proporcional à fatorial do número de cidades que o caixeiro teria de visitar. Em problemas desse tipo, dependendo do

número de cidades a solução do problema não pode ser encontrada por métodos enumerativos, assim sendo, os algoritmos genéticos são uma boa opção, pois procuram o número de soluções que os seus parâmetros definirem, e nunca uma fração significativa das soluções possíveis, processo que poderia levar a um tempo computacional muito alto para a obtenção de uma solução satisfatória (Leardi 1992).

2. OBJETIVOS

- Atualizar e acrescentar novas classes ao banco de dados químico existente, criando um banco mais completo com informações químicas das doze classes de metabólitos secundários isolados de espécies da família Asteraceae.
- Utilizar o banco criado para estudos quimiotaxonômicos da família Asteraceae. Verificar a possibilidades de separar tribos e gêneros da família com a utilização de parâmetros evolutivos, usando como ferramenta as redes neurais de Kohonen.
- Verificar se é possível realizar previsões dos números de ocorrências das doze classes de metabólitos da família, utilizando uma classificação heurística de cinco faixas dos valores de ocorrência através de uma rede neural de perceptrons de camadas múltiplas com algoritmo de retropropagação de erros.
- Construir modelos de regressão linear (MLR), utilizando como método de seleção de variáveis os algoritmos genéticos, com o objetivo de verificar se com os dados presentes no banco de dados químico é possível encontrar algum equilíbrio na produção dos esqueletos das classes de substâncias do grupo dos terpenóides, e caso seja possível, verificar quais classes estão mais relacionadas.

3. MATERIAIS E MÉTODOS

3.1. Marcadores químicos

Em nosso estudo quimiotaxonômico utilizamos como marcadores as seguintes classes de metabólitos secundários: Monoterpenos (MONO), sesquiterpenos (SESQ), sesquiterpenos lactonizados (LACT), diterpenos (DITE), triterpenos (TRITE), flavonóides (FLAV), cumarinas (CUMA), poliacetilenos (POLI), benzofuranos (BENZF), benzopiranos (BENZP), fenilpropanóides (FENIL) e acetofenonas (ACET).

3.2. Biossíntese dos diferentes marcadores

A figura 3.2.1 (Mann et al. 1994) apresenta de modo geral as principais rotas biossintéticas de metabolitos secundários em plantas vasculares. As demais rotas passam a ser mais específicas para cada tipo de classe estudada.

Todos os terpenóides (monoterpenos, sesquiterpenos, diterpenos, triterpenos) são formados de unidades básicas chamadas de isopreno, os quais são substâncias com cinco átomos de carbono. A biossíntese dos terpenóides consiste na condensação dos precursores básicos o pirofosfato de isopentenila (IPP) e seu isômero o pirofosfato de 3,3-dimetilalila (DMPA). Uma reação nucleofílica entre o IPP e o DMPA (cauda-cabeça), mediadas por enzimas denominadas preniltransferases, gera o pirofosfato de geranila, unidade precursora dos terpenos, como ilustra a figura 3.2.2 (Dewick 2002).

Figura 3.2.1. Rotas biossintéticas de produtos naturais em plantas (Mann et al. 1994).

Figura 3.2.2. Biossíntese dos Monoterpenos, Sesquiterpenos, Diterpenos, Triterpenos e Esteróides a partir do pirofosfato de isopentenila (IPP).

São conhecidas duas rotas para a biossíntese do IPP, a primeira rota o é desde o fim da década de 1950 (Disch et al. 1998), esta via envolve a biossíntese do ácido mevalônico (MVA). A rota do ácido mevalônico está esquematizada na figura 3.2.3.

Figura 3.2.3 – Biossíntese do IPP: Rota do ácido mevalônico.

A segunda rota, independente da via do ácido mevalônico para a síntese do IPP, é a via do 1-desoxi-D-xilose-5-fosfato (DOXP). Essa última via foi descoberta em uma eubactéria (Rohmer et al. 1993; Lichtenthaler 1999), a seguir verificou-se que esta rota está presente em algas verdes (*Chlorophyta*), plantas superiores e outros grupos de algas (Lichtenthaler 1999). Nesta rota (figura 3.2.4), o piruvato após reagir com pirofosfato de tiamina-enzima (TPP-E), reage com gliceraldeído-3-fosfato (GA-3-P) formando 1-desoxi-D-xilose-5-fosfato, que a seguir produz o pirofosfato de isopentenila (IPP).

Em plantas a síntese dos terpenos é compartimentalizada, monoterpenos e diterpenos são produzidos pela via do 1-desoxi-D-xilose que tem suas enzimas localizadas nos plastídios. Os sesquiterpenos e triterpenos são produzidos no citoplasma, onde se encontram as enzimas da via do ácido mevalônico (Kirby e Keasling 2009; McGarvey e Croteau 1995). Existem evidências da passagem de precursores entre o citoplasma e os plastídios, entretanto a extensão da passagem é

limitada e a perda de uma via não pode ser compensada pela outra (Kirby e Keasling 2009).

Figura 3.2.4 - Biossíntese do IPP: Rota do 1-desoxi-D-xilose-5-fosfato.

As cumarinas, assim como os fenilpropanóides têm como precursor o ácido cinâmico, sendo, portanto, biossintetizadas através da rota do ácido chiquimico (figuras 3.2.5 e 3.2.6) (Dewick 2002).

Figura 3.2.5. Rota de formação do ácido chiquimico.

Figura 3.2.6. Formação do ácido cinâmico a partir do ácido chiquimico.

Os benzofuranos, benzopiranos e acetofenonas são substâncias aromáticas de grande diversidade que podem ser biossintetizadas através da rota dos policetideos (figura 3.2.7) (Dewick 2004).

Figura 3.2.7. Esquema biossintéticos de um composto aromático através da via dos policetídeos.

Os flavonóides possuem origem mista, são formados a partir de duas rotas distintas. A rota do ácido chiquimico (porção C3-C6) e a dos policetídeos (Acetato/malonato) (porção C6). Na figura 3.2.8 temos o esquema da biossíntese de alguns flavonóides a partir dos precursores 4-coumaroil-CoA (rota do ácido chiquimico, figuras 3.2.5 e 3.2.6) e malonil-CoA (rota dos policetídeos, figura 3.2.7) (Dewick 2002).

Figura 3.2.8. Esquema de biossíntese de alguns flavonóides a partir do precursor 4-coumaroil-CoA (Dewick 2002).

Os poliacetilenos provêm do ácido oléico através de sucessivas etapas de desidrogenação e perdas de carbono, como pode ser visto na figura 3.2.9 (Dewick 2002).

Figura 3.2.9. Esquema biossintético dos poliacetilenos.

3.3. Coleta e arquivamento dos dados

Como ponto de partida para a construção do banco atual de dados químicos da família Asteraceae foi utilizado um banco prévio já existente no grupo (Ferreira et al. 2001).

A partir do banco existente o trabalho resumiu-se ao levantamento das ocorrências de algumas classes inexistentes ou bastante deficientes no banco inicial (sesquiterpenos, benzopiranos, acetofenonas e fenilpropanóides) e atualização das ocorrências naturais das demais classes de metabólitos já presentes no banco. O levantamento foi realizado a partir da literatura especializada englobando revisões anteriores a 2008. Na primeira etapa os dados levantados foram os seguintes:

- 1- Ocorrências apenas naturais dos metabólitos em questão nas espécies estudadas.
- 2- Parte da planta estudada, que podem ser toda ou específicas: partes áreas, folhas, flores, caules, raízes, frutos, etc.
- 3- Local de coleta.
- 4- Registro do gênero e espécie da qual o metabólito de interesse foi extraído.
- 5- Referência, na qual as informações foram obtidas. Dado importante para quaisquer conferências.

Posteriormente, foram calculados os seguintes parâmetros:

- 1- Grau de Oxidação (NOX).
- 2- Passo Oxidativo (PO).
- 3- Grau de especialização de esqueletos (S).
- 4- Reação Característica (REA).

O passo seguinte consistiu na organização dos dados em uma planilha eletrônica no formato Excel. Os dados da parte da planta estudada, local de coleta e referências não foram adicionados à planilha, sendo reservados para análises futuras. A tabela 3.3.1 ilustra como os dados foram organizados.

Tabela 3.3.1. Ilustração da planilha criada com os dados químicos da família Asteraceae.

TRIBO	SUBTRIBO	GÊNERO	ESPECIE	PO	POB	S	ESQUEL	CLASSE	REA
Tribo1	Subtribo1	Gênero1	Especie1	Po1	Pob1	S1	Esquel1	Classe1	Rea1
Tribo2	Subtribo2	Gênero2	Especie2	Po2	Pob2	S2	Esquel2	Classe2	Rea2
Tribo3	Subtribo3	Gênero3	Especie3	Po3	Pob3	S3	Esquel3	Classe3	Rea3

Na tabela acima, PO = Passo Oxidativo, POB = PO da porção C3-C6 de flavonóides, CLASSE = Código da classe de metabólito secundário, REA = reação característica, ESQUEL = Tipo de esqueleto e S = Índice de Especialização do Esqueleto.

3.4. Número de ocorrência

O “Número de ocorrência” (NO) é um parâmetro químico que fornece o grau de relevância de uma determinada categoria metabólica (X) para um táxon escolhido (A). Na prática, se para duas espécies de um táxon foram registradas cinco substâncias diferentes da mesma categoria para uma e sete para a outra, considerando-se para o táxon referente à categoria micromolecular em questão NO = 12, mesmo que haja coincidência de substâncias nas duas espécies. “Ao contar n vezes uma mesma substância encontrada em n espécies diferentes, consegue-se caracterizar a tendência de sua produção no táxon” (Gottlieb 1996).

3.5. Parâmetros evolutivos

A evolução da família Asteraceae pode ser estudada através da utilização de índices de avanço evolutivos calculados usando metabólitos secundários (Gotlieb 1982). No presente estudo foram calculados alguns índices que, espera-se, possam representar o estágio evolutivo de alguns táxons. Esses índices são: Grau de oxidação (NOX), Passo Oxidativo (PO), grau de especialização (S) e Reações Características (REA). Também se classificou as classes de metabólitos entre os diferentes tipos de esqueletos (ESQUEL) e esse parâmetro foi também utilizado como parâmetro evolutivo da família Asteraceae.

3.5.1. Grau de oxidação (NOX)

O trabalho envolvendo oxidação de micro moléculas baseia-se no cálculo do estado relativo de oxidação de cada átomo de carbono, que somados, fornecem o grau de oxidação de cada substância (NOX). Para cada substância o número de oxidação foi calculado através das regras de Hendrickson *et al.* (Hendrickson et al. 1970) . Essas regras podem ser resumidas em uma equação: $NOX = \sum n_i C-B$, onde n_i é o número de ligações entre o átomo C e o átomo B. Nessa equação, B corresponde a H, C ou heteroátomo X, isto é, as ligações resultantes são respectivamente C-H, C-C e C-X. Os valores de oxidação para cada carbono são obtidos adicionando cada um dos seguintes valores para suas quatro ligações: -1 para C-H, zero para C-C e +1 para C-X (X= heteroátomo). Na figura 3.5.1.1 temos dois exemplos de cálculos de NOX.

Figura 3.5.1.1. Exemplos do cálculo de NOX.

3.5.2. Passo Oxidativo (PO)

Embora o método de cálculo de NOX se aplique bem a cada classe, fica difícil a comparação de dois valores de oxidação vindos de diferentes classes. Pretendemos ampliar esta metodologia introduzindo o conceito de passos oxidativos (“oxidative step” – OS) (Emerenciano et al. 1998b). Como passo oxidativo entende-se as mudanças oxidativas em relação a um precursor considerado como marco zero na escala. O passo oxidativo de uma substância é calculado através do absoluto da diferença entre os valores de NOX do precursor e do composto, e finalmente dividindo o resultado por 2 como demonstrado a seguir.

$$PO = \frac{|(NOX_{\text{precursor}} - NOX_{\text{substância}})|}{2} \quad \text{Equação 3.5.2.1}$$

Cada passo oxidativo representa uma insaturação ou uma ligação carbono-heteroátomo presente em dada substância quando comparada ao seu precursor. Cada diferença de dois pontos no NOX é considerada um passo oxidativo (PO), vide figura 3.5.2.1.

Figura 3.5.2.1 Exemplos de cálculo de PO.

3.5.3. Especialização dos esqueletos (S)

Índice de especialização do esqueleto é definido como o número de reações necessárias para se chegar a um esqueleto a partir de um precursor (Gottlieb 1982).

A especialização de esqueletos (S) é calculada da seguinte maneira:

$$S = \frac{q + f + c + u}{n} \quad \text{Equação 3.5.3.1}$$

- n: número de átomos de carbono do esqueleto molecular
- q: número de ligações C-C quebradas com respeito a um precursor
- f: número de ligações C-C formadas com relação ao mesmo precursor
- c: número de sistemas cíclicos formados envolvendo heteroátomos com relação ao precursor
- u: número de unidades carbônicas adicionais com relação ao mesmo precursor.

3.5.4. Reação Característica (REA)

Reação característica foi um parâmetro criado no âmbito desse trabalho a partir da diversidade de esqueletos obtidos no levantamento bibliográfico. A idéia na criação desse parâmetro foi estimar quais possíveis reações químicas deram origem a cada esqueleto a partir do esqueleto precursor. Na tabela 3.5.4.1 estão descritas as reações características presentes no banco de dados.

Tabela 3.5.4.1. Reações características dos esqueletos presentes no banco de dados.

Código	REA
0	Desconhecida
3	Formação Carbono, Carbono (C-C) e migração
4	Formação C-C
5	Dimerização
6	Formação C-C + abertura
7	Mudança de configuração
8	Formação C-C + expansão do anel
9	Formação C-C + migração + perda de C
10	Formação C-C + perda de C
11	Formação C-C + migração + abertura
12	Formação C-C + contração + perda de C
14	Formação C-C + migração + perda de C + abertura
16	Formação C-C + perda de C + abertura
18	Adição de C extra
19	Adição de unidade de outro caminho biossintético
20	Formação C-C + migração + perda de C + ciclo de três
21	Formação C-C + migração + contração de anel
22	Formação C-C + perda de C + ciclo de três
23	Formação C-C + migração + ciclo de três
24	Formação C-C + migração + contração de anel + ciclo de quatro
25	Formação C-C + expansão do anel + migração
26	Formação C-C + migração + contração
27	Formação C-C + ciclo de quatro
28	Formação C-C + abertura + contração do anel
29	Formação C-C + anel de três
30	Formação C-C + anel de quatro + migração
31	Formação C-C + abertura + anel de três
33	Formação C-C + abertura de anel + C extra
36	Formação C-C + perda de C + ciclo de quatro
37	Formação C-C + abertura de anel + anel de quatro
39	Formação C-C irregular + migração
41	Formação C-C irregular + ciclo de três
44	Formação C-C irregular
45	Cumarina + R
46	Cumarina + R + Prenil
47	Cumarina + prenil irregular
48	REA 56 - Metil + R
49	Formação C-C + migração + abertura + perda de C + contração
50	Cumarina + R + Mono - C
51	REA 56 -Metil
52	Cumarina + R + Sesqu - C

Tabela 3.5.4.1. Continuação.

Código	REA

53	Cumarina + R + Mono + C
55	Acetofenona Simples
56	4 Acetatos simples
57	Acetofenona (ACTF) + R
58	ACTF + prenil
59	ACTF + 2 prenil
60	Dímero de Cumarina
61	ACTF + R + prenil
62	THF simples
63	THF + R
64	Cumarina + Sesq
65	THF + THP
66	THP + R + Prenil
67	Dímero de THP
68	THP + R + R-Aromático
69	REA 70 - C
70	THP simples
71	THP + prenil
72	THF + R - C
73	THP + R
74	Dímero de THF + R
75	Dímero de THF
76	Furano Cumarina
77	ACTF + R-Aromático
78	Dímero de ACTF
79	Perda de C + R em THP
80	THF + perda de C
81	Cumarina simples
82	Cumarina + prenil
83	Formação C-C em Poliacetilenos
84	Fenilpropanóide simples (FPRO)
85	FPRO + prenil
86	FPRO + 2 prenil
87	FPRO + R
88	Dímero de THP + R + prenil
89	THP + 2 prenil
91	Formação C-C + abertura + expansão + Migração de R
92	Formação C-C + abertura + expansão + Migração de Metila
93	Cumarina + Mono + metila
94	Formação C-C + abertura + expansão + Perda de C
96	REA 90 + C extra
97	Formação C-C + abertura + contração do anel + C extra + ciclo de três
98	Cumarina + Sesq + metila
99	REA 98 + abertura
100	Flavonóides

3.5.5. Tipo de Esqueleto (ESQUEL)

Ao remover todos os grupos funcionais de metabólitos secundários (exemplo: terpenóides), ou parcialmente, mantendo um núcleo comum (exemplo: flavonóides), chega-se a estruturas simplificadas que são chamadas esqueletos. A figura 3.5.5.1 mostra exemplos de flavonóides e sesquiterpenos, com os seus respectivos esqueletos.

As moléculas das doze classes de metabólitos secundários foram organizadas em tipos de Esqueleto. Cada tipo de esqueleto diferente recebeu um número de acordo com a sua classe. Todos os tipos de esqueletos encontram-se no apêndice A.

Figura 3.5.5.1. Exemplos de tipos de esqueletos (ESQUEL) e moléculas que compartilham esses esqueletos.

3.6. Redes Neurais

3.6.1. Rede Neural Não Supervisionada (Mapas auto-organizáveis)

O Banco de dados químico, utilizando os dados que eram comuns as filogenias proposta por Bremer e da Meta-tree (24.687 ocorrências), foi utilizado como dados de entrada na rede neural Kohonen. Nesta rede, os parâmetros evolutivos (PO, S, REA, ESQUEL) e o código da classe de metabólito secundário foram as variáveis de entrada, ou seja, cada amostra (tribo ou gênero) corresponde a um vetor, o qual é constituído pelos valores dos parâmetros evolutivos e o código da classe do metabólito deste composto.

Como está descrito na seção 1.4.4 a rede neural Kohonen utiliza a aprendizagem não supervisionada. Os dados de ocorrência botânica, as tribos ou gêneros, são utilizados no SOM apenas para “rotular” áreas do mapa, não participando do treinamento. Nesta fase apenas os parâmetros evolutivos e os códigos das classes de metabólitos foram utilizados como dados de entrada.

Para gerar os mapas auto-organizáveis (SOM) foi utilizado o aplicativo SOM toolbox 2.0 (Vesanto *et al.* 2005) para Matlab 6.5 (Matlab 2004). Todas as estruturas dos mapas foram geradas em duas dimensões, e os neurônios foram organizados de forma retangular, no qual cada neurônio tem quatro vizinhos. Através da utilização da equação 1.4.4.1 foi medida a semelhança entre o vetor de entrada (composto) e o vetor de ponderação do neurônio, possibilitando encontrar o BMU (“Best Match Unit”). Na determinação da vizinhança utilizou-se a função gaussiana (equação 1.4.4.2) e o treinamento foi realizado em lote, como descrito no item 1.4.4.

Ao final da fase de treinamento da rede auto-organizável de Kohonen, os neurônios do mapa são rotulados pelo maior número de casos, ou seja, caso o táxon com o maior número de amostras seja do gênero *Geigeria*, este neurônio será

rotulado como uma “região” de Geigeria. Assim, os casos neste determinado neurônio são considerados acertos se forem do gênero Geigeria, e consideradas erro em caso contrário.

Em relação aos mapas, as dimensões dos mesmos foram determinadas empiricamente, minimizando o erro.

Onze análises foram realizadas utilizando a metodologia descrita acima. Nas primeiras três análises objetivou-se separar o maior número de tribos da família Asteraceae, utilizando como entrada na rede as informações do banco de dados químico por tribo e por gênero. As Quatro análises seguintes tiveram como alvo a separação de tribos e gêneros dos quatro maiores grupos em que a família Asteraceae é dividida na filogenia de Bremer. Outras três análises foram concentradas na separação de tribos e gêneros da subfamília Asteroideae. E a última das análises realizada consistiu na tentativa de separar gêneros da tribo Astereae (AST).

3.6.1. Rede Neural Supervisionada (MLP)

Para a previsão das ocorrências das doze classes de metabólitos secundários foi utilizado a rede supervisionada de perceptrons de múltiplas camadas, na qual o sinal de entrada se propaga para frente através da rede, camada por camada, utilizando o algoritmo de retropropagação de erro (error back-propagation) (Haykin 2001). As explicações de como as análises foram realizadas encontra-se no item 4.2.2.

3.7. Algoritmos Genéticos

Para esse trabalho utilizou-se o programa MOBYDIGS v. 1.0 (Talet 2004) que possibilita selecionar as variáveis por algoritmo genético, usando o comando “Variable Subset Selection - Genetic Algorithm (VSS-GA)”. O parâmetro de avaliação para esse estudo foi o Q_{cv}^2 . No algoritmo genético do programa MobyDigs v 1.0, cada indivíduo é determinado por um cromossomo que é um vetor binário, onde cada posição (um gene) corresponde a uma variável, ou seja, esqueleto (1 se for incluído no modelo e 0 caso contrário). Por isso, cada cromossomo representa uma equação definida por um subconjunto de variáveis (esqueletos). No programa MobyDigs v1.0, pode-se determinar o tamanho da população, ou seja o número de equações retidas e o número máximo permitido de variáveis (esqueletos) em um modelo. Em uma população, os operadores evolutivos de “crossing-over” e mutação são repetidos até um número máximo de iterações, ou até que o processo seja encerrado arbitrariamente. Na inicialização do algoritmo genético uma população aleatória com um número de variáveis (esqueletos) entre 1 e o número máximo determinado pelo usuário é formada. O número de indivíduos (cromossomos), também é definido pelo usuário, para este trabalho utilizou-se 73 indivíduos (14 esqueletos de monoterpenos, 26 de sesquiterpenos, 13 de sesquiterpenos lactonizados, 12 de diterpenos e 8 de triterpenos). Para as análises, o número máximo de variáveis (esqueletos) permitidas nas equações foi de 3 e 6.

O valor da função selecionada (Q_{cv}^2), utilizada para classificar os indivíduos (cromossomos), é calculada em um processo denominado avaliação, em que os modelos (cromossomos) são ordenados no que diz respeito ao valor do coeficiente de previsão pelo método de “*full cross-validation*” (Q_{cv}^2) (Leardi *et al.* 1992). Em uma

população, os pares de modelos são selecionados com uma probabilidade que é calculada em função da sua qualidade, ou seja, pelo seu valor de Qcv^2 , utilizando o processo de seleção por roleta. Na seleção por roleta, quanto maior o valor de Qcv^2 da equação (indivíduo), maior a sua chance de ser selecionado. Em uma “roleta” são colocados todos os indivíduos da população. A faixa de cada seção da roleta é proporcional ao valor de Qcv^2 de cada indivíduo: quanto maior for esse valor, mais larga a seção. Em seguida, a partir de cada par de modelos selecionados (pais), é gerado um novo modelo, preservando as características comuns destes, misturando-os de acordo com a probabilidade de “crossing-over”. Se o “filho” gerado coincide com um dos indivíduos já presentes na própria população, este simplesmente é rejeitado (Leardi *et al.* 1992), caso contrário, a equação é avaliada. Se o valor da função objetivo (Qcv^2) da equação é melhor do que o pior valor presente na população, esta será incluída no lugar correspondente na população à sua classificação, caso contrário, não é considerada. Cada indivíduo da população pode ter seu gene aleatoriamente mudado, alterando o valor de 0 para 1, de 1 para 0. Os indivíduos alterados por mutação são avaliados da mesma forma que os “filhos” no processo “crossing-over”. Este processo é controlado por probabilidade que é normalmente fixado em valores baixos, permitindo assim poucas mutações. As equações que foram geradas são modelos lineares (MLR – regressão linear múltipla) com modelos até 3 variáveis e modelos com até 6 variáveis, selecionando-se a equação com maior valor do coeficiente de previsão gerado pelo método *full cross-validation* (Qcv^2). Nesse estudo, para cada análise realizada foram salvos dez modelos de MLR, no total foram realizadas quatro análises, portanto foram criados 40 modelos de regressão múltipla linear.

Na tabela 3.7.1 encontram-se a descrição dos índices estatísticos utilizados nos modelos de MLR criados no âmbito desse trabalho.

Tabela 3.7.1. Índices estatísticos utilizados nos modelos de MLR utilizando algoritmos genéticos.

Expressão Matemática	Explicação, Incógnitas, onde:
$r^2 = 1 - \frac{\sum_{i=1}^n (y_{pi} - \hat{y}_{pi})^2}{\sum_{i=1}^n (y_{pi} - \bar{y}_p)^2}$	- r^2 é o coeficiente de correlação entre os y calculados e os y observados; - y_{pi} é o valor calculado da variável dependente do composto através do modelo; - $\hat{y}_{pi}$ é o valor calculado da variável dependente através da equação da reta de ajuste entre os valores observados e os valores calculados ; - $\bar{y}_p$ é a média dos valores calculados de y pelo modelo
$r^2(adj) = 1 - \frac{(n-1)}{(n-k-1)} * (1 - r^2)$	- $r^2(adj)$ é coeficiente de determinação ajustado que leva em conta o número de variáveis explicatórias em relação ao número de observações para facilitar a comparação de diversos modelos de regressões. - K é o número de regressões. - n é o tamanho da amostra.
$Q_{cv}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i^*)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$	- Q_{cv}^2 é o coeficiente de predição pelo método de <i>full crossvalidation</i>; - y_i é o valor observado da variável dependente do composto da série de treinamento o qual não participou da equação de regressão; - $\hat{y}_i^*$ é o valor da calculado variável dependente através do modelo de regressão do respectivo composto; - $\bar{y}$ é a média dos valores observados da variável dependente ;

4. RESULTADOS

4.1. Banco de Dados Químico

A planilha em formato Excel, na qual os dados químicos da família Asteraceae foram organizados, foi a fonte de dados para todas as análises realizadas no âmbito desse trabalho. Um fragmento dessa planilha, que chamaremos a seguir de banco de dados químico, está representada na tabela 4.1.1. O banco de dados químico apresenta 36.496 ocorrências de metabólitos secundários, os quais estão divididos nas doze classes já mencionadas: Monoterpenos (MONO), Sesquiterpenos (SESQ), Sesquiterpenos Lactonizados (LACT), Diterpenos (DITE), Triterpenos (TRITE), Cumarinas (CUMA), Flavonóides (FLAV), Poliacetilenos (POLI), Benzofuranos (BENZF), Benzopiranos (BENZP), Acetofenonas (ACET) e Fenilpropanóides (FENIL).

Como é possível observar na tabela 4.1.2, no banco de dados químicos estão presentes 4772 espécies, o que corresponde a 20,4% do total de espécies presentes no trabalho de Bremer. Quando observamos os gêneros, o banco de dados químico possui 673 gêneros, o que corresponde a 44,4% dos gêneros descritos no trabalho de Bremer (tabela 4.1.3). Tais dados mostram que o banco de dados químico construído apresenta uma boa amostragem da família Asteraceae, quando se faz a comparação das espécies e gêneros existentes na família segundo a filogenia de Bremer e os presentes no banco.

Tabela 4.1.1. Fragmento do Banco de dados químicos de Asteraceae.

TRIBO	SUBTRIBO	GÊNERO	ESPECIE	PO	POB	S	ESQUEL	CLASSE	REA
ANT	ACH	ACHILLEA	ABROTANOIDES	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	ABROTANOIDES	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	ABROTANOIDES	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	BIEBERSTEINII	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	BIEBERSTEINII	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	BISERRATA	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	BISERRATA	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	BISERRATA	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	CLYPEOLATA	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	CLYPEOLATA	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	CLYPEOLATA	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	CORABENSIS	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	CRITHMIFOLIA	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	CRITHMIFOLIA	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	CRITHMIFOLIA	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	CRITHMIFOLIA	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	ERBA-ROTTA	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	ERBA-ROTTA	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	ERBA-ROTTA	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	ERBA-ROTTA	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	FILIPENDULINA	0,00	0,00	0,00	1	1	4
ANT	ACH	ACHILLEA	FILIPENDULINA	0,00	0,00	0,00	1	1	4

Tabela 4.1.2 Quantidade de espécies presentes na família Asteraceae segundo Bremer [1], quantidade de espécies presentes no banco de dados químicos de Asteraceae do grupo, e porcentagem de espécies presentes no banco em relação à filogenia de Bremer.

Tribos	Código	Quantidade Espécies	Quantidade de Espécies no Banco	Porcentagem
Barnadesieae	BAR	92	36	39,1
Mutisieae	MUT	318	125	39,3
Cardueae	CAR	2513	357	14,2
Arctoteae	ARC	139	48	34,5
Lactuceae	LAC	2486	184	7,4
Liabeae	LIA	159	15	9,4
Vernonieae	VER	1346	213	15,8
Anthemideae	ANT	1737	600	34,5
Astereae	AST	2846	542	19,0
Calenduleae	CAL	113	60	53,1
Eupatorieae	EUP	2396	467	19,5
Gnaphalieae	GNA	1728	299	17,3
Helenieae	HLN	835	329	39,4
Heliantheae	HLT	2449	982	40,1
Inuleae	INU	480	142	29,6
Plucheeae	PLU	480	47	9,8
Senecioneae	SEN	3247	326	10,0
Total		23364	4772	20,4

Tabela 4.1.3. Quantidade de gêneros presentes na família Asteraceae segundo Bremer, quantidade de gêneros presentes no banco de dados químicos de Asteraceae do grupo, e porcentagem de gêneros presentes no banco em relação à filogenia de Bremer.

Tribos segundo Bremer	Quantidade de Gêneros	Quantidade de Gêneros no Banco	Porcentagem (%)
ANT	109	38	34,86
ARC	16	11	68,75
AST	173	70	40,46
BAR	9	7	77,78
CAL	9	8	88,89
CAR	83	39	46,99
EUP	169	80	47,34
GNA	180	49	27,22
HLN	111	52	46,85
HLT	188	109	57,98
INU	38	34	89,47
LAC	98	31	31,63
LIA	14	6	42,86
MUT	75	39	52,00
PLU	28	7	25,00
SEN	121	35	28,93
VER	98	31	31,63
Total	1519	674	44,30

Em relação à filogenia mais recente da família proposta por Funk et.al. 2009, não foi possível comparar a quantidade de espécies, pois na referência consultada (Funk et al. 2009) não há detalhamento do número de espécies considerado no estudo. Quanto aos gêneros o banco de dados químicos possui 324 gêneros (tabela 4.1.4), que representa 48,1% dos gêneros utilizados na construção da Meta-tree. Essa porcentagem mostra que o banco de dados químicos também possui uma amostragem significativa dos gêneros utilizados nessa filogenia.

Tabela 4.1.4. Quantidade de gêneros presentes na família Asteraceae segundo Meta-tree, quantidade de gêneros presentes no banco de dados químicos de Asteraceae do grupo, e porcentagem de gêneros presentes no banco em relação à filogenia da Meta-tree.

Tribos segundo Meta-tree	Quantidade de Gêneros Total	Quantidade de Gêneros no Banco	Porcentagem (%)
ANT	58	21	36,21
ARC	14	9	64,29
AST	78	38	48,72
ATH	2	2	100,00
BAH	17	8	47,06
BAR	10	6	60,00
CAL	12	7	58,33
CAR	43	27	62,79
CHA	3	1	33,33
CIC	55	18	32,73
COR	20	12	60,00

Tabela 4.1.4. Continuação.

Tribos segundo Meta-tree	Quantidade de Gêneros Total	Quantidade de Gêneros no Banco	Porcentagem (%)
COY	1	1	100,00
DIC	6	2	33,33
ERE	1	0	0,00
EUP	26	21	80,77
FED	1	0	0,00
GNA	47	19	40,43
GOC	4	3	75,00
GYM	1	0	0,00
HEC	1	0	0,00
HLN	13	10	76,92
HLT	6	4	66,67
HYA	4	1	25,00
INU	30	19	63,33
LIA	16	4	25,00
MAD	34	15	44,12
MIL	3	3	100,00
MOQ	2	2	100,00
MUT	9	7	77,78
NAU	13	9	69,23
NEU	1	1	100,00
OLD	1	1	100,00
ONO	5	2	40,00
PER	4	3	75,00
PET	3	2	66,67
PLA	1	1	100,00
POL	1	1	100,00
SEM	64	21	32,81
STI	5	1	20,00
TAG	17	6	35,29
TAR	2	2	100,00
VER	34	13	38,24
WUN	4	1	25,00
Total	673	324	48,14

A tabela 4.1.5 e a figura 4.1.1 mostram como as ocorrências das doze classes de metabólitos secundários estão distribuídas nas 17 tribos da família Asteraceae segundo a filogenia de Bremer, já a tabela 4.1.6 e a figura 4.1.2 mostram como as ocorrências das doze classes de metabólitos secundários estão distribuídas nas 39 tribos da família Asteraceae segundo a filogenia proposta na Meta-tree.

Tabela 4.1.5. Ocorrência e porcentagem das doze classes de metabólitos secundários nas 17 tribos (propostas por Bremer) da família Asteraceae. Monoterpenos (MONO), Sesquiterpenos (SESQ), Sesquiterpenos lactonizados (LACT), Diterpenos (DITE), Triterpenos (TRITE), Cumarinas (CUMA), Flavonóides (FLAV), Benzofuranos (BENZF), Benzopiranos (BENZP), Acetofenonas (ACET), Fenilpropanóides (FENIL).

TRIBOS	MONO		SESQ		LACT		DITE		TRITE		CUMA		FLAV		POLI		BENZF		BENZP		ACET		FENIL		Total	
	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%
ANT	3512	56,8	1443	20,6	1168	19,9	12	0,3	156	13	223	22,8	878	16,3	141	4	4	0,6	10	2,5	36	15,6	46	9,9	7629	20,9
ARC	1	0	37	0,5	23	0,4	18	0,4	35	2,9	1	0,1	0	0	23	0,6	0	0	0	0	0	0	0	0	138	0,4
AST	622	10,1	697	9,9	13	0,2	1521	33,4	127	10,6	187	19,1	520	9,7	377	10,6	65	10	16	4	5	2,2	102	21,9	4252	11,7
BAR	0	0	0	0	0	0	0	0	1	0,1	0	0	174	3,2	0	0	0	0	0	0	0	0	0	0	175	0,5
CAL	9	0,1	25	0,4	3	0,1	93	2	21	1,8	3	0,3	1	0	24	0,7	7	1,1	0	0	5	2,2	3	0,6	194	0,5
CAR	24	0,4	430	6,1	462	7,9	5	0,1	81	6,8	25	2,6	327	6,1	1055	29,6	0	0	0	0	1	0,4	26	5,6	2436	6,7
EUP	393	6,4	661	9,4	628	10,7	860	18,9	100	8,4	41	4,2	409	7,6	15	0,4	205	31,5	184	45,5	47	20,3	12	2,6	3555	9,7
GNA	155	2,5	570	8,1	27	0,5	415	9,1	31	2,6	52	5,3	367	6,8	112	3,1	33	5,1	29	7,2	60	26	27	5,8	1878	5,1
HLN	377	6,1	203	2,9	465	7,9	138	3	30	2,5	29	3	944	17,6	132	3,7	10	1,5	35	8,7	0	0	30	6,5	2393	6,6
HLT	473	7,6	822	11,7	1513	25,8	1274	28	103	8,6	40	4,1	891	16,6	1578	44,2	198	30,4	116	28,7	35	15,2	114	24,5	7157	19,6
INU	290	4,7	235	3,3	374	6,4	75	1,6	24	2	8	0,8	195	3,6	46	1,3	0	0	2	0,5	0	0	7	1,5	1256	3,4
LAC	15	0,2	17	0,2	200	3,4	1	0	152	12,7	77	7,9	433	8,1	0	0	6	0,9	0	0	1	0,4	21	4,5	923	2,5
LIA	5	0,1	17	0,2	52	0,9	0	0	9	0,8	0	0	8	0,1	0	0	0	0	0	0	0	0	6	1,3	97	0,3
MUT	58	0,9	184	2,6	119	2	35	0,8	64	5,3	160	16,3	86	1,6	5	0,1	6	0,9	2	0,5	13	5,6	2	0,4	734	2
PLU	80	1,3	226	3,2	18	0,3	0	0	5	0,4	40	4,1	33	0,6	2	0,1	0	0	0	0	0	0	7	1,5	411	1,1
SEN	130	2,1	1304	18,6	131	2,2	77	1,7	57	4,8	28	2,9	26	0,5	5	0,1	113	17,4	10	2,5	21	9,1	45	9,7	1947	5,3
UNA	15	0,2	39	0,6	85	1,4	16	0,4	2	0,2	1	0,1	10	0,2	37	1	3	0,5	0	0	0	0	5	1,1	213	0,6
VER	26	0,4	109	1,6	586	10	11	0,2	199	16,6	64	6,5	76	1,4	17	0,5	1	0,2	0	0	7	3	12	2,6	1108	3
Total	6185	100	7019	100	5867	100	4551	100	1197	100	979	100	5378	100	3569	100	651	100	404	100	231	100	465	100	36496	100

• Ocor. = Número de Ocorrência.

Tabela 4.1.6. Ocorrência das doze classes de metabólitos secundários nas 39 tribos presentes no banco (propostas na Meta-tree) da família Asteraceae. Monoterpenos (MONO), Sesquiterpenos (SESQ), Sesquiterpenos lactonizados (LACT), Diterpenos (DITE), Triterpenos (TRITE), Cumarinas (CUMA), Flavonóides (FLAV), Benzofuranos (BENZF), Benzopiranos (BENZP), Acetofenonas (ACET), Fenilpropanóides (FENIL).

TRIBOS	MONO		SESQ		LACT		DITE		TRITE		CUMA		FLAV		POLI		BENZF		BENZP		ACET		FENIL		Total	
	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%
ANT	2214	35,8	852	12,1	949	16,2	7	0,2	132	11	183	18,7	307	5,7	23	0,6	4	0,6	10	2,5	31	13,4	35	7,5	4747	13
ARC	1	0	33	0,5	20	0,3	7	0,2	35	2,9	1	0,1	0	0	23	0,6	0	0	0	0	0	0	0	0	120	0,3
AST	528	8,5	597	8,5	12	0,2	1073	23,6	101	8,4	185	18,9	255	4,7	292	8,2	50	7,7	14	3,5	5	2,2	79	17	3191	8,7
ATH	2	0	0	0	0	0	12	0,3	0	0	0	0	0	0	0	0	0	0	5	1,2	0	0	0	0	19	0,1
BAH	32	0,5	8	0,1	66	1,1	77	1,7	5	0,4	4	0,4	11	0,2	14	0,4	0	0	0	0	0	0	2	0,4	219	0,6
BAR	0	0	0	0	0	0	0	0	1	0,1	0	0	170	3,2	0	0	0	0	0	0	0	0	0	0	171	0,5
CAL	7	0,1	25	0,4	0	0	93	2	21	1,8	3	0,3	1	0	24	0,7	7	1,1	0	0	5	2,2	3	0,6	189	0,5
CAR	22	0,4	422	6	400	6,8	3	0,1	72	6	25	2,6	325	6	1040	29,1	0	0	0	0	1	0,4	26	5,6	2336	6,4
CHA	0	0	0	0	10	0,2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0
CIC	15	0,2	17	0,2	187	3,2	1	0	103	8,6	69	7	280	5,2	0	0	6	0,9	0	0	1	0,4	20	4,3	699	1,9
COR	13	0,2	38	0,5	10	0,2	0	0	12	1	4	0,4	191	3,6	827	23,2	2	0,3	0	0	0	0	41	8,8	1138	3,1
COY	0	0	0	0	0	0	8	0,2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	9	0
DIC	0	0	2	0	41	0,7	0	0	16	1,3	0	0	5	0,1	0	0	0	0	0	0	0	0	0	0	64	0,2
EUP	353	5,7	526	7,5	427	7,3	576	12,7	68	5,7	35	3,6	383	7,1	13	0,4	92	14,1	163	40,3	31	13,4	7	1,5	2674	7,3
GNA	141	2,3	442	6,3	21	0,4	299	6,6	31	2,6	46	4,7	255	4,7	74	2,1	19	2,9	27	6,7	58	25,1	22	4,7	1435	3,9
GOC	0	0	8	0,1	16	0,3	14	0,3	4	0,3	3	0,3	29	0,5	0	0	0	0	0	0	0	0	0	0	74	0,2
HLN	78	1,3	18	0,3	363	6,2	0	0	0	0	1	0,1	43	0,8	40	1,1	0	0	0	0	0	0	9	1,9	552	1,5
HLT	68	1,1	90	1,3	334	5,7	174	3,8	26	2,2	11	1,1	104	1,9	148	4,1	0	0	4	1	0	0	0	0	959	2,6
HYA	0	0	0	0	0	0	8	0,2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0
INU	216	3,5	328	4,7	283	4,8	67	1,5	18	1,5	45	4,6	159	3	28	0,8	0	0	1	0,2	0	0	13	2,8	1158	3,2
LIA	3	0	16	0,2	19	0,3	0	0	9	0,8	0	0	8	0,1	0	0	0	0	0	0	0	0	6	1,3	61	0,2
MAD	44	0,7	33	0,5	19	0,3	39	0,9	25	2,1	9	0,9	682	12,7	45	1,3	6	0,9	29	7,2	0	0	3	0,6	934	2,6
MIL	11	0,2	44	0,6	104	1,8	53	1,2	0	0	2	0,2	32	0,6	34	1	2	0,3	1	0,2	2	0,9	13	2,8	298	0,8
MOQ	0	0	1	0	3	0,1	0	0	6	0,5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0
MUT	29	0,5	0	0	0	0	1	0	4	0,3	70	7,2	15	0,3	0	0	6	0,9	2	0,5	10	4,3	1	0,2	138	0,4

Tabela 4.1.6. Continuação.

TRIBOS	MONO		SESQ		LACT		DITE		TRITE		CUMA		FLAV		POLI		BENZF		BENZP		ACET		FENIL		Total	
	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%	Ocor.	%
NAU	1	0	138	2	10	0,2	1	0	19	1,6	54	5,5	12	0,2	0	0	0	0	0	0	2	0,9	0	0	237	0,6
NEU	11	0,2	2	0	9	0,2	0	0	0	0	0	0	13	0,2	0	0	0	0	0	0	0	0	0	0	35	0,1
OLD	0	0	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0
ONO	0	0	0	0	0	0	3	0,1	8	0,7	8	0,8	10	0,2	0	0	0	0	0	0	1	0,4	0	0	30	0,1
PER	7	0,1	0	0	3	0,1	0	0	0	0	0	0	62	1,2	0	0	0	0	0	0	0	0	0	0	72	0,2
PET	0	0	0	0	4	0,1	0	0	1	0,1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0
PLA	0	0	1	0	3	0,1	11	0,2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0
POL	0	0	0	0	2	0	18	0,4	0	0	0	0	2	0	10	0,3	1	0,2	0	0	2	0,9	0	0	35	0,1
SEN	121	2	1183	16,9	88	1,5	47	1	44	3,7	5	0,5	19	0,4	4	0,1	70	10,8	3	0,7	14	6,1	11	2,4	1609	4,4
STI	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
TAG	203	3,3	144	2,1	0	0	4	0,1	0	0	15	1,5	133	2,5	12	0,3	4	0,6	1	0,2	0	0	14	3	530	1,5
TAR	28	0,5	5	0,1	40	0,7	0	0	5	0,4	0	0	2	0	0	0	0	0	0	0	0	0	0	0	80	0,2
VER	24	0,4	64	0,9	439	7,5	10	0,2	148	12,4	63	6,4	34	0,6	15	0,4	1	0,2	0	0	7	3	7	1,5	812	2,2
WUN	0	0	1	0	5	0,1	0	0	2	0,2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	9	0
UNA	2013	32,5	1978	28,2	1978	33,7	1945	42,7	281	23,5	138	14,1	1835	34,1	902	25,3	381	58,5	144	35,6	61	26,4	153	32,9	11809	32,4
Total	6185	100	7019	100	5867	100	4551	100	1197	100	979	100	5378	100	3569	100	651	100	404	100	231	100	465	100	36496	100

- Ocor. = Número de Ocorrência.

Figura 4.1.1. Distribuição das doze classes de metabólitos secundários nas 17 tribos (propostas por Bremer) da família Asteraceae. Monoterpenos (MONO), Sesquiterpenos (SESQ), Sesquiterpenos lactonizados (LACT), Diterpenos (DITE), Triterpenos (TRITE), Cumarinas (CUMA), Flavonóides (FLAV), Benzofuranos (BENZF), Benzopiranos (BENZP), Acetofenonas (ACET), Fenilpropanóides (FENIL).

Figura 4.1.2. Distribuição das doze classes de metabólitos secundários nas 41 tribos presentes no banco (propostas na Meta-tree) da família Asteraceae. Monoterpenos (MONO), Sesquiterpenos (SESQ), Sesquiterpenos lactonizados (LACT), Diterpenos (DITE), Triterpenos (TRITE), Cumarinas (CUMA), Flavonóides (FLAV), Benzofuranos (BENZF), Benzopiranos (BENZP), Acetofenonas (ACET), Fenilpropanóides (FENIL).

Na tabela 4.1.7 estão mostradas as quantidades e porcentagens dos esqueletos nas doze classes de metabólitos secundários, e as tabelas 4.1.8 – 4.1.19 detalham os números de ocorrência e as porcentagens dos cinco mais representativos esqueletos para cada uma das doze classes de metabólitos secundários, já as figuras 4.1.3 – 4.1.14 apresentam os cinco esqueletos mais representativos para cada uma das doze classes de metabólitos secundários.

Tabela 4.1.7. Quantidade e porcentagem de esqueletos em classe de metabólito secundário na família Asteraceae.

Classe	Quantidade de Esqueletos	%
Monoterpeno	48	8,9
Sesquiterpeno	163	30,4
Sesquiterpeno Lactonizado	58	10,8
Diterpeno	70	13,0
Triterpeno	33	6,1
Cumarina	40	7,4
Flavonóide	20	3,7
Poliacetileno	9	1,7
Benzofurano	25	4,7
Benzopirano	30	5,6
Acetofenona	36	6,7
Fenilpropanóide	5	0,9
Total	537	100,0

Tabela 4.1.8. Descrição dos esqueletos com maior número de ocorrência na classe dos monoterpenos.

Principais Esqueletos de Monoterpenos	Número de Ocorrência	Porcentagem (%)
Esqueleto 15	2731	38,33
Esqueleto 1	969	15,67
Esqueleto 44	927	14,99
Esqueleto 42	607	9,81
Esqueleto 37	484	7,83
Total	5718	86,63

Figura 4.1.3. Esqueletos com maior número ocorrência na classe dos monoterpenos.**Tabela 4.1.9.** Descrição dos esqueletos com maior número de ocorrência na classe dos sesquiterpenos.

Principais Esqueletos de Sesquiterpenos	Número de Ocorrência	Porcentagem (%)
Esqueleto 153	885	12,61
Esqueleto 150	753	10,73
Esqueleto 133	715	10,19
Esqueleto 195	710	10,12
Esqueleto 101	580	8,26
Total	3643	51,91

Figura 4.1.4. Esqueletos com maior número ocorrência na classe dos sesquiterpenos.

Tabela 4.1.10. Descrição dos esqueletos com maior número de ocorrência na classe dos sesquiterpenos lactonizados.

Principais Esqueletos de Sesquiterpenos lactonizados	Número de Ocorrência	Porcentagem (%)
Esqueleto 301	1741	29,67
Esqueleto 302	1520	25,91
Esqueleto 311	830	14,15
Esqueleto 303	528	9
Esqueleto 314	325	5,54
Total	4944	84,27

Figura 4.1.5. Esqueletos com maior número ocorrência na classe dos sesquiterpenos lactonizados.

Tabela 4.1.11. Descrição dos esqueletos com maior número de ocorrência na classe dos diterpenos.

Principais Esqueletos de Diterpenos	Número de Ocorrência	Porcentagem (%)
Esqueleto 471	1443	31,71
Esqueleto 430	1300	28,57
Esqueleto 439	723	15,89
Esqueleto 449	264	5,8
Esqueleto 412	253	5,56
Total	3983	87,53

Figura 4.1.6. Esqueletos com maior número ocorrência na classe dos diterpenos.**Tabela 4.1.12.** Descrição dos esqueletos com maior número de ocorrência na classe dos triterpenos.

Principais Esqueletos de Triterpenos	Número de Ocorrência	Porcentagem (%)
Esqueleto 525	481	40,18
Esqueleto 524	226	18,88
Esqueleto 521	163	13,62
Esqueleto 529	72	6,02
Esqueleto 528	51	4,26
Total	993	82,96

Figura 4.1.7. Esqueletos com maior número ocorrência na classe dos triterpenos.

Tabela 4.1.13. Descrição dos esqueletos com maior número de ocorrência na classe das cumarinas.

Principais Esqueletos de Cumarinas	Número de Ocorrência	Porcentagem (%)
Esqueleto 561	738	76,24
Esqueleto 563	67	6,92
Esqueleto 583	24	2,48
Esqueleto 581	18	1,86
Esqueleto 575	14	1,45
Total	861	88,95

Figura 4.1.8. Esqueletos com maior número ocorrência na classe das cumarinas.

Tabela 4.1.14. Descrição dos esqueletos com maior número de ocorrência na classe dos flavonóides.

Principais Esqueletos de Flavonóides	Número de Ocorrência	Porcentagem (%)
Esqueleto 657	2567	47,73
Esqueleto 656	2155	40,07
Esqueleto 655	311	5,78
Esqueleto 652	192	3,57
Esqueleto 651	107	1,99
Total	5332	99,14

Figura 4.1.9. Esqueletos com maior número ocorrência na classe dos flavonóides.**Tabela 4.1.15.** Descrição dos esqueletos com maior número de ocorrência na classe dos poliacetilenos.

Principais Esqueletos de Poliacetilenos	Número de Ocorrência	Porcentagem (%)
Esqueleto 694	2149	60,33
Esqueleto 691	437	12,27
Esqueleto 698	361	10,13
Esqueleto 695	238	6,68
Esqueleto 693	209	5,87
Total	3394	95,28

Figura 4.1.10. Esqueletos com maior número ocorrência na classe dos poliacetilenos.

Tabela 4.1.16. Descrição dos esqueletos com maior número de ocorrência na classe dos benzofuranos.

Principais Esqueletos de Benzofuranos	Número de Ocorrência	Porcentagem (%)
Esqueleto 712	477	72,49
Esqueleto 711	29	4,41
Esqueleto 713	22	3,34
Esqueleto 716	21	3,19
Esqueleto 718	19	2,89
Total	568	86,32

Figura 4.1.11. Esqueletos com maior número ocorrência na classe dos benzofuranos.

Tabela 4.1.17. Descrição dos esqueletos com maior número de ocorrência na classe dos benzopiranos.

Principais Esqueletos de Benzopiranos	Número de Ocorrência	Porcentagem (%)
Esqueleto 751	260	64,36
Esqueleto 752	48	12,13
Esqueleto 753	26	6,44
Esqueleto 754	19	4,7
Esqueleto 763	6	1,49
Total	359	89,12

Figura 4.1.12. Esqueletos com maior número ocorrência na classe dos benzopiranos.

Tabela 4.1.18. Descrição dos esqueletos com maior número de ocorrência na classe das Acetofenonas.

Principais Esqueletos de Acetofenonas	Número de Ocorrência	Porcentagem (%)
Esqueleto 792	56	24,24
Esqueleto 797	32	13,85
Esqueleto 790	31	13,42
Esqueleto 796	29	12,55
Esqueleto 793	5	2,16
Total	153	66,22

Figura 4.1.13. Esqueletos com maior número ocorrência na classe das acetofenonas.

Tabela 4.1.19. Descrição dos esqueletos com maior número de ocorrência na classe dos fenilpropanóides.

Principais Esqueletos de Fenilpropanóides	Número de Ocorrência	Porcentagem (%)
Esqueleto 831	408	85,71
Esqueleto 833	36	7,56
Esqueleto 834	28	5,88
Esqueleto 832	3	0,63
Esqueleto 835	1	0,22
Total	476	100

Figura 4.1.14. Esqueletos com maior número ocorrência na classe dos fenilpropanóides.

4.2. Redes Neurais

4.2.1. Redes Neurais não Supervisionadas

Para a obtenção dos resultados dessa seção, como já foi mencionando, foi utilizado o arquivo do banco de dados químico, com as ocorrências comuns as filogenias de Bremer e a proposta na Meta-tree (um total de 24.687 ocorrências de metabólitos secundários), como dado de entrada no software SOM toolbox 2.0. Das análises foram gerados os mapas auto-organizáveis que estão apresentados nas figuras 4.2.1.1 – 4.2.1.3. Os valores de acertos para cada tribo ou gênero e as dimensões dos mapas são mostrados na tabelas 4.2.1.1 – 4.2.1.6.

Tabela 4.2.1.1. Resultados dos Mapas Auto-Organizáveis, e suas respectivas dimensões, com os valores das ocorrências, os números de acertos absolutos e relativos para cinco Tribos segundo a filogenia de Bremer (figura 4.2.1.1 a) e sete tribos segundo a filogenia da Meta-tree (figura 4.2.1.1 b).

Tribos (Bremer)	Dimensão do mapa 82x7			Tribos (Meta-tree)	Dimensão do mapa 80x7		
	Nº de acertos	Ocorrência	% de acerto		Nº de acertos	Ocorrência	% de acerto
ANT	4216	4747	88,81	ANT	3972	4747	83,67
AST	1810	3185	56,83	AST	1876	3191	58,79
HLT	1646	2465	66,77	COR	968	1138	85,06
SEN	1149	1609	71,41	MAD	643	934	68,84
VER	490	822	59,61	NAU	161	237	67,93
Total	9311	12828	72,58	SEN	1174	1609	72,96
				VER	488	812	60,10
				Total	9282	12668	73,27

Tabela 4.2.1.2. Resultados do Mapa Auto-Organizável, e sua dimensão, com os valores das ocorrências, os números de acertos absolutos e relativos para nove gêneros da família Asteraceae (figura 4.2.1.1 c).

Dimensão do mapa 36x4				
Tribos	Gêneros	Nº de acertos	Ocorrência	% de acerto
EUP	ACRITOPAPPUS	70	100	70
AST	BELLIS	59	82	71,95
INU	GEIGERIA	100	132	75,76
MUT	MUTISIA	56	68	82,35
CAL	OSTEOSPERMUM	63	87	72,41
SEN	PETASITES	54	73	73,97
GNA	RELHANIA	60	90	66,67
HLT	RUDBECKIA	82	99	82,83
ANT	URSINIA	71	91	78,02
	Total	615	822	74,82

Tabela 4.2.1.3. Resultados dos Mapas Auto-Organizáveis, e suas respectivas dimensões, com os valores das ocorrências, os números de acertos absolutos e relativos para quatro tribos da família Asteraceae segundo Bremer (figura 4.2.1.2 a) e quatro tribos da família Asteraceae segundo a Meta-tree (figura 4.2.1.2 b).

Dimensão do mapa 76x5				Dimensão do mapa 68x5			
Tribos (Bremer)	Nº de acertos	Ocorrência	% de acerto	Tribos (Meta-tree)	Nº de acertos	Ocorrência	% de acerto
CAR	2023	2436	83,05	CAR	1960	2336	83,90
LAC	642	923	69,56	CIC	518	699	74,11
MUT	354	734	48,23	MUT	62	138	44,93
SEN	1762	1947	90,50	SEN	1444	1609	89,75
Total	4781	6040	79,16	Total	3984	4782	83,31

Tabela 4.2.1.4. Resultados dos Mapas Auto-Organizáveis, e suas respectivas dimensões, com os valores das ocorrências, os números de acertos absolutos e relativos para quatro gêneros da família Asteraceae. Gêneros: Saussurea, Lactuca, Trixis e Doronicum. Mapa na figura 4.2.1.3 a. Gêneros: Carthamus, Picris, Mutisia e Petasites, mapa na figura 4.2.1.3 b.

Tribos			Dimensão do mapa 10x9			Tribos			Dimensão do mapa 10x9		
Bremer	Meta-tree	Gêneros	Nº de acertos	Ocorrência	% de acerto	Bremer	Meta-tree	Gêneros	Nº de acertos	Ocorrência	% de acerto
CAR	CAR	SAUSSUREA	85	109	77,98	CAR	CAR	CARTHAMUS	77	88	87,5
LAC	CIC	LACTUCA	69	95	72,63	LAC	CIC	PICRIS	66	83	79,52
MUT	MUT	TRIXIS	61	71	85,92	MUT	NAU	MUTISIA	55	68	80,88
SEN	SEN	DORONICUM	56	68	82,35	SEN	SEN	PETASITES	71	73	97,26
		Total	271	343	79,01			Total	269	312	86,22

Tabela 4.2.1.5. Resultados dos Mapas Auto-Organizáveis, e suas respectivas dimensões, com os valores das ocorrências, os números de acertos absolutos e relativos para três tribos da subfamília Asteroideae (Asteraceae) segundo Bremer (figura 4.2.1.2 c) e três tribos da subfamília Asteroideae (Asteraceae) segundo a Meta-tree (figura 4.2.1.2 d).

Dimensão do mapa 73x5				Dimensão do mapa 65x5			
Tribos (Bremer)	Nº de acertos	Ocorrência	% de acerto	Tribos (Meta-tree)	Nº de acertos	Ocorrência	% de acerto
GNA	1441	1878	76,73	GNA	1101	1435	76,72
INU	820	1256	65,29	INU	707	1158	61,05
SEN	1566	1947	80,43	SEN	1315	1609	81,73
Total	3827	5081	75,32	Total	3123	4202	74,32

Tabela 4.2.1.6. Resultados dos Mapas Auto-Organizáveis, e suas respectivas dimensões, com os valores das ocorrências, os números de acertos absolutos e relativos para três gêneros da subfamília Asteroideae (Asteraceae) (figura 4.2.1.3 c) e quatro gêneros da tribo Astereae (figura 4.2.1.3 d).

Dimensão do mapa 3x30					Dimensão do mapa 13x12			
Tribos	Gêneros	Nº de acertos	Ocorrência	% de acerto	Gêneros	Nº de acertos	Ocorrência	% de acerto
GNA	RELHANIA	80	90	88,89	ASTER	239	355	67,32
INU	GEIGERIA	120	132	90,91	ERIGERON	174	219	79,45
SEN	PETASITES	68	73	93,15	HETEROTHECA	93	134	69,4
	Total	268	295	90,85	PSIADIA	138	204	67,65
					Total	644	912	70,61

Figura 4.2.1.1. Mapas auto-organizáveis. a) Mapa obtido da classificação de cinco tribos da família Asteraceae (segundo Bremer): ANT (Azul), AST (Amarelo), HLT (Vermelho), SEN (Rosa), VER (Verde). b) Mapa obtido da classificação de sete tribos da família Asteraceae (segundo Meta-tree): ANT (Azul), AST (Amarelo), COR (Vermelho), MAD (Marrom), NAU (Azul Claro), SEN (Rosa), VER (Verde). c) Mapa obtido da classificação de nove gêneros da família Asteraceae: *Acratopappus* (Azul), *Bellis* (Verde), *Geigeria* (Azul claro), *Mutisia* (Vermelho), *Osteospermum* (Amarelo), *Petasites* (Rosa), *Relhania* (Cinza), *Rudbeckia* (Preto), *Ursinia* (Marrom).

Figura 4.2.1.2. Continuação Mapas. a) Mapa obtido da classificação de quatro Tribos da família Asteraceae segundo Bremer: CAR (Azul), LAC (Amarelo), MUT (Vermelho), SEN (Rosa). b) Mapa obtido da classificação quatro Tribos da família Asteraceae segundo Meta-tree: CAR (Azul), CIC (Amarelo), MUT (Vermelho), SEN (Rosa). c) Mapa obtido da classificação de três tribos da subfamília Asteroideae (segundo Bremer): GNA (Azul), INU (Vermelho), SEN (Amarelo). d) Mapa obtido da classificação de três tribos da subfamília Asteroideae (segundo Meta-tree): GNA (Azul), INU (Vermelho), SEN (Amarelo).

a)

b)

c)

d)

Figura 4.2.1.3. Continuação Mapas. a) Mapa obtido da classificação de quatro gêneros da família Asteraceae: Saussurea (Azul), Lactuca (Vermelho), Trixis (Amarelo) e Doronicum (Verde). b) Mapa obtido da classificação de quatro gêneros da família Asteraceae: Carthamus (Azul), Picris (Vermelho), Mutisia (Amarelo), Petasites (Verde). c) Mapa obtido da classificação de três gêneros da Subfamília Asteroideae: Geigeria (Azul), Petasites (Vermelho), Relhania (Amarelo). d) Mapa obtido da classificação de quatro Gêneros da Tribo Astereae: Aster (Azul), Erigeron (Vermelho), Heterotheca (Amarelo), Psidia (Verde).

4.2.2 Redes Neurais Supervisionadas

A partir do banco de dados químico completo (36.496 ocorrências) foi realizada a contagem do número de ocorrência das classes de metabólitos por gênero, utilizando o programa GERTAXO (Gertaxo 2004). A tabela 4.2.2.1 apresenta um fragmento da planilha gerada pela contagem.

Tabela 4.2.2.1. Fragmento da planilha de contagens do número de ocorrência das classes de metabólitos secundários por gêneros, gerada pelo programa Gertaxo (Gertaxo 2004)0,...

[illegible]

A partir da planilha que está mostrada na tabela 4.2.2.1, foram calculados a média e o desvio padrão para cada uma das doze classes de metabólitos (tabela 4.2.2.2). Com os cálculos de média e desvio padrão foram criadas cinco faixas para representar a probabilidade relativa de cada classe de composto produzido por determinado gênero. A tabela 4.2.2.3 mostra as probabilidades criadas e como as médias e os desvios padrão foram utilizados para gerar essa classificação heurística da distribuição dos metabólitos secundários por gêneros.

Tabela 4.2.2.2. Média e Desvio padrão das ocorrências das doze classes metabólitos nos gêneros da família Asteraceae.

Classe de Metabólito	Média ($\bar{x}$)	Desvio Padrão (DP)
Monoterpeno	4,61	15,9
Sesquiterpeno	6,79	19,6
Sesquiterpeno lactonizado	7,32	21,2
Diterpeno	6,18	20,6
Triterpeno	1,43	4,7
Cumarinas	1,05	4,2
Flavonóides	7,22	26,6
Poliacetileno	4,79	20,8
Benzofurano	0,96	4,8
Benzopirano	0,60	4,5
Acetofenona	0,26	1,6
Fenilpropanóides	0,69	3,1

Tabela 4.2.2.3. Classificação heurística da probabilidade relativa de ocorrência das doze classes de metabólitos nos gêneros da família Asteraceae.

Presença do metabólito	Probabilidade relativa	Faixa
Raro	0,2	0 até a média ($\bar{x}$)
Pouco Provável	0,4	$\bar{x} + 2$ Desvios Padrão (DP)
Provável	0,6	$\bar{x} + 4$ DP
Muito Provável	0,8	$\bar{x} + 6$ DP
Presente	1	acima de $\bar{x} + 6$ DP

Na planilha obtida da contagem no programa GERTAXO (Gertaxo 2004) (tabela 4.2.2.1) os valores foram substituídos pelas probabilidade relativas,

obedecendo a faixa presente na tabela 4.2.2.3. A substituição gerou uma planilha, que está amostrada na tabela 4.2.2.4.

Tabela 4.2.2.4. Fragmento da planilha gerada com a substituição dos valores obtidos na contagem do Gertaxo (Gertaxo 2004) pelas probabilidades relativas para cada classe de composto.

GENERO	MONO	SESQ	LACT	DITE	TRITE	CUMA	FLAV	POLI	BENZF	BENZP	ACET	FENIL
DYSSODIA	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
PENTANEMA	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
LESSINGIANTHUS	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
CARPHEPHORUS	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
ATHROISMA	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
DICORIA	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
SYNEDRELLOPSIS	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
SYNEILESIS	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
CRITHMUM	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
ANTIPHIONA	0,4	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
KIPPISTIA	0,4	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
SCHIZOGYNE	0,4	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
SERIPHIDIUM	0,4	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
AUTROBRICKELLIA	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
CONOCLINIUM	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
LUCILIA	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
KIGELIA	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
YOUNGIA	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
PSEUDOSTIFFTIA	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
ARGYRANTHEMUM	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
SCLEROLEPIS	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
CHILIADENUS	0,4	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
SCLERORHACHIS	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
COREOCARPUS	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
NAUPLIUS	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
ARTCTOTIS	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
ASTERISCUS	0,4	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
GONOSPERMUM	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
CHAMOMILA	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2

A tabela 4.2.2.4 foi utilizada como entrada no programa Statistica (Statsoft 2000). Para realizar as análises com o intuito de prever a probabilidade relativa de metabólitos de uma classe utilizando os dados de ocorrências das demais classes. Para as análise foi utilizada a ferramenta “Neural Networks”, a opção de análise para todos os casos foi a “Intelligent

Problem Solver”, foram testadas 1000 redes em cada ensaio, com retenção das cinco melhores redes, as redes foram perceptron de três camadas (Three layer perceptron), as funções de ativação utilizadas foram linear e logística.

As análises foram realizadas utilizando onze classes como entrada e uma classe como dado de saída. Ao total foram realizadas doze análises, cada uma gerou cinco redes diferentes. Para a avaliação do erro ou acerto nas redes adotou-se o critério presente na função 4.2.2.1.

Se, $ABS(\text{valor predito} - \text{valor real}) < 0,2 \times \text{valor predito}$, então é acerto
Se, $ABS(\text{valor predito} - \text{valor real}) \geq 0,2 \times \text{valor predito}$, então é erro **Função 4.2.2.1**

A tabela 4.2.2.5, mostra a série de treino (531 gêneros) e teste (143 gêneros) para cada uma das análises e a média de acertos das cinco redes geradas para cada variável (classe de metabólito).

As figuras 4.2.2.1 – 4.2.2.8 mostram gráficos dos valores reais versus os valores previstos da melhor rede entre as cinco redes retidas nas análises para cada uma das doze classes de metabólitos secundários utilizadas.

Tabela 4.2.2.5. Médias dos acertos da série de treino e série de teste na previsão da variável utilizando como entrada para a rede os dados das outras 11 variáveis.

Dados de Entrada (Metabólitos Secundários)	Série de Treino (Número de gêneros)	Média de acerto do treino (%)	Série de Teste (Número de gêneros)	Média de acerto do teste (%)
Monoterpenos	531	84	143	85
sesquiterpenos	531	81	143	79
sesquiterpenos lactonizados	531	78	143	78
Diterpenos	531	83	143	78
Triterpenos	531	81	143	75
Cumarinas	531	89	143	83
Flavonóides	531	84	143	87
Poliacetilenos	531	84	143	81
Benzofuranos	531	88	143	86
Benzopiranos	531	92	143	95
Acetofenonas	531	93	143	90
Fenilpropanóides	531	85	143	85

A tabela 4.2.2.6 mostra a distribuição das cinco faixas de probabilidade relativa de ocorrência para cada uma das doze classes de metabólitos nos gêneros da família Asteraceae presentes no banco de dados químico.

Tabela 4.2.2.6. Distribuição das cinco faixas probabilidades relativas das ocorrências nas doze classes de substâncias nos 674 gêneros da família Asteraceae.

MONO-PR	Quantidade	%	SESQ-PR	Quantidade	%	LACT-PR	Quantidade	%
0,2	568	84,3	0,2	543	80,6	0,2	543	80,6
0,4	78	11,6	0,4	97	14,4	0,4	103	15,3
0,6	15	2,2	0,6	18	2,7	0,6	16	2,4
0,8	6	0,9	0,8	7	1,0	0,8	4	0,6
1,0	7	1,0	1	9	1,3	1	8	1,2
Total	674	100,0		674	100,0		674	100,0
DITE-PR	Quantidade	%	TRITE-PR	Quantidade	%	CUMA-PR	Quantidade	%
0,2	563	83,5	0,2	546	81,0	0,2	591	87,7
0,4	87	12,9	0,4	104	15,4	0,4	61	9,1
0,6	11	1,6	0,6	11	1,6	0,6	6	0,9
0,8	6	0,9	0,8	6	0,9	0,8	9	1,3
1,0	7	1,0	1	7	1,0	1	7	1,0
Total	674	100,0	674	674	100,0	674	674	100,0
FLAV-PR	Quantidade	%	POLI-PR	Quantidade	%	BENZF-PR	Quantidade	%
0,2	566	84,0	0,2	572	84,9	0,2	589	87,4
0,4	87	12,9	0,4	84	12,5	0,4	70	10,4
0,6	11	1,6	0,6	10	1,5	0,6	6	0,9
0,8	4	0,6	0,8	3	0,4	0,8	4	0,6
1,0	6	0,9	1	5	0,7	1	5	0,7
Total	674	100,0	674	674	100,0	674	674	100,0
BENZP-PR	Quantidade	%	ACET-PR	Quantidade	%	FENIL-PR	Quantidade	%
0,2	628	93,2	0,2	628	93,2	0,2	585	86,8
0,4	35	5,2	0,4	31	4,6	0,4	69	10,2
0,6	7	1,0	0,6	6	0,9	0,6	13	1,9
0,8	1	0,1	0,8	5	0,7	0,8	3	0,4
1,0	3	0,4	1	4	0,6	1	4	0,6
Total	674	100,0	674	674	100,0	674	674	100,0

- -PR = Probabilidade relativa segundo a classificação heurística.

Figura 4.2.2.1. Gráficos dos valores reais versus previsto obtidos no treino da rede neural com algoritmo de retro propagação. a) Gráfico dos monoterpenos, b) Gráfico dos sesquiterpenos, c) Gráfico dos sesquiterpenos lactonizados.

Figura 4.2.2.2. Gráficos dos valores reais versus previsto obtidos no teste da rede neural com algoritmo de retro propagação. a) Gráfico dos monoterpenos, b) Gráfico dos sesquiterpenos, c) Gráfico dos sesquiterpenos lactonizados.

Figura 4.2.2.3. Gráficos dos valores reais versus previsto obtidos no treino da rede neural com algoritmo de retro propagação. a) Gráfico dos diterpenos, b) Gráfico dos triterpenos, c) Gráfico das cumarinas.

Figura 4.2.2.4. Gráficos dos valores reais versus previsto obtidos no teste da rede neural com algoritmo de retro propagação. a) Gráfico dos diterpenos, b) Gráfico dos triterpenos, c) Gráfico das cumarinas.

Figura 4.2.2.5. Gráficos dos valores reais versus previsto obtidos no treino da rede neural com algoritmo de retro propagação. a) Gráfico dos flavonóides, b) Gráfico dos poliacetilenos, c) Gráfico dos benzofuranos.

Figura 4.2.2.6. Gráficos dos valores reais versus previsto obtidos no teste da rede neural com algoritmo de retro propagação. a) Gráfico dos flavonóides, b) Gráfico dos poliacetilenos, c) Gráfico dos benzofuranos

Figura 4.2.2.7. Gráficos dos valores reais versus previsto obtidos no treino da rede neural com algoritmo de retro propagação. a) Gráfico dos benzopiranos, b) Gráfico das acetophenona, c) Gráfico dos fenilpropanóides.

Figura 4.2.2.8. Gráficos dos valores reais versus previsto obtidos no teste da rede neural com algoritmo de retro propagação. a) Gráfico dos benzopiranos, b) Gráfico das acetofenonas, c) Gráfico dos fenilpropanóides.

4.3 Algoritmos genéticos

Dos dois grupos de análises em que a ferramenta de algoritmos genéticos foi utilizada para a previsão dos esqueletos 1 e 101, obteve-se os modelos de regressão múltipla linear (MLR) descritos na tabela 4.3.1 e nas equações 4.3.1 – 4.3.40 (presentes no apêndice B).

As figuras 4.3.1 – 4.3.4 mostram os esqueletos utilizados na criação dos modelos de MLR.

Tabela 4.3.1. Resumo dos modelos de regressão da previsão dos esqueletos 1 e 101.

Modelos	Esqueleto Previsto (variável dependente)	Variáveis independentes (Esqueletos)	r^2	Qcv ²	r^2_{adj}
Equação 4.3.1	1	15 41 128	0,9	0,9	0,88
Equação 4.3.2	1	15 41 207	0,9	0,9	0,88
Equação 4.3.3	1	41 44 134	0,9	0,8	0,86
Equação 4.3.4	1	41 44 128	0,9	0,8	0,87
Equação 4.3.5	1	15 41 134	0,9	0,8	0,86
Equação 4.3.6	1	15 41 196	0,9	0,8	0,86
Equação 4.3.7	1	15 41 153	0,9	0,8	0,86
Equação 4.3.8	1	41 44 207	0,9	0,8	0,87
Equação 4.3.9	1	41 44 153	0,9	0,8	0,86
Equação 4.3.10	1	15 39 134	0,9	0,8	0,85
Equação 4.3.11	1	15 23 41 114 128 149	0,9	0,9	0,89
Equação 4.3.12	1	15 41 114 128 149 301	0,9	0,9	0,89
Equação 4.3.13	1	15 41 114 128 149 302	0,9	0,9	0,89
Equação 4.3.14	1	15 41 128 149 207 212	0,9	0,9	0,9
Equação 4.3.15	1	15 41 114 128 149 317	0,9	0,9	0,89
Equação 4.3.16	1	15 41 114 128 149 416	0,9	0,9	0,89
Equação 4.3.17	1	15 41 114 128 149	0,9	0,9	0,89
Equação 4.3.18	1	15 41 114 128 149 253	0,9	0,9	0,89
Equação 4.3.19	1	15 41 114 128 149 456	0,9	0,9	0,89
Equação 4.3.20	1	15 41 114 128 149 319	0,9	0,9	0,89
Equação 4.3.21	101	42 133 203	0,9	0,9	0,89
Equação 4.3.22	101	42 133 234	0,9	0,9	0,9
Equação 4.3.23	101	37 133 203	0,9	0,9	0,89
Equação 4.3.24	101	37 133 234	0,9	0,9	0,9
Equação 4.3.25	101	42 149 234	0,9	0,9	0,9
Equação 4.3.26	101	37 149 234	0,9	0,9	0,89
Equação 4.3.27	101	1 8 234	0,9	0,9	0,88
Equação 4.3.28	101	39 133 234	0,9	0,9	0,89

Tabela 4.3.1. Continuação

Modelos	Esqueleto Previsto (variável dependente)	Variáveis independentes (Esqueletos)	r^2	Q_{cv}^2	r^2_{adj}
Equação 4.3.29	101	8 149 234	0,9	0,9	0,9
Equação 4.3.30	101	15 149 234	0,9	0,9	0,9
Equação 4.3.31	101	1 42 113 149 212 234	0,9	0,9	0,92
Equação 4.3.32	101	1 42 113 149 234 302	0,9	0,9	0,92
Equação 4.3.33	101	1 8 149 171 212 234	0,9	0,9	0,92
Equação 4.3.34	101	1 8 133 149 212 234	0,9	0,9	0,92
Equação 4.3.35	101	1 8 42 149 212 234	0,9	0,9	0,92
Equação 4.3.36	101	1 8 23 149 212 234	0,9	0,9	0,92
Equação 4.3.37	101	1 8 149 212 234 303	0,9	0,9	0,92
Equação 4.3.38	101	1 8 113 149 234 311	0,9	0,9	0,92
Equação 4.3.39	101	1 8 149 212 234 301	0,9	0,9	0,92
Equação 4.3.40	101	1 8 149 212 234 528	0,9	0,9	0,92

Figura 4.3.1. Esqueletos de Monoterpenos que aparecem nos modelos de regressão das equações 4.3.1 a 4.3.40.

Figura 4.3.2. Esqueletos de Sesquiterpenos que aparecem nos modelos de regressão das equações 4.3.1 a 4.3.40.

Figura 4.3.3. Esqueletos de Sesquiterpenos lactonizados que aparecem nos modelos de regressão das equações 4.3.12-4.3.13, 4.3.2.15, 4.3.2.20, 4.3.2.32 e, 4.3.37 – 4.3.39.

Figura 4.3.4. Esqueletos de diterpenos e triterpeno que aparecem nos modelos de regressão das equações 4.3.16, 4.3.19 e 4.3.40.

5. DISCUSSÃO

5.1. Banco de Dados Químico

No decorrer do trabalho a maior parte do tempo foi dedicada à construção do Banco de dados químico, o que é explicado pelo fato de esse banco ter sido a base de dados para realizar todas as análises desse trabalho.

O banco apresenta um número considerável de espécies (20,4% das espécies estudadas por Bremer) e gêneros (44,4% e 48,1% dos gêneros presentes na família segundo as filogenias de Bremer e da Meta-tree respectivamente). Tais valores são bastante representativos para uma amostragem da química da família.

Apesar da alta porcentagem de gêneros e espécies presentes no banco de dados químico, existem algumas distorções quando são consideradas individualmente as tribos da família, seja na filogenia de Bremer ou na proposta na Meta-tree, a amostragem não é homogênea. Ao se observar a tabela 4.1.2 pode-se perceber que existem grandes diferenças entre as porcentagens relativas de espécies no banco com relação às estudadas por Bremer nas diferentes tribos. Como exemplo pode-se analisar as tribos Calenduleae e Liabeae, que possuem respectivamente 53,1% e 9,4% das espécies estudadas por Bremer amostradas no Banco. Segundo Bremer as duas tribos possuem semelhante quantidade de espécies, mas no banco são amostradas de maneira desigual.

Quando a amostragem do banco em relação aos gêneros da família é considerada, observa-se que as espécies presentes no banco (segundo a filogenia de Bremer) estão relativamente bem distribuídas entre os gêneros. Em geral a porcentagem de gêneros por tribos presentes no banco é maior que a porcentagem

de espécies, o que permite inferir que essas espécies estão espalhadas entre os gêneros da tribo a qual pertencem. Esse dado, em oposição ao dado anterior, que foi ilustrado pelo exemplo das tribos Calenduleae e Liabeae, indica que a amostragem do banco cobre uma parcela significativa do universo da família Asteraceae.

Com relação à amostragem segundo a filogenia proposta na Meta-tree, não foi possível fazer considerações a respeito das espécies, uma vez que a referência consultada (Funk et al. 2009) não continha o dado da quantidade de espécies em cada tribo. Entretanto, em relação à amostragem dos gêneros (tabela 4.1.4), é possível observar que o banco de dados químicos amostra significativamente a maior parte das tribos propostas nessa filogenia, sendo que 9 das 39 tribos presentes no banco de dados químicos possuem 100% dos gêneros utilizados na filogenia da Meta-tree.

Outro aspecto a ser discutido sobre o banco de dados químicos diz respeito à distribuição do número de ocorrência das doze classes de metabólitos secundários nas 17 tribos da filogenia de Bremer e nas 39 tribos da filogenia apresenta na Meta-tree.

A distribuição dos metabólitos nas 17 tribos propostas por Bremer é bastante variável, o que pode ser visto na tabela 4.1.5 e figura 4.1.1. Da tabela 4.1.5 vale ressaltar alguns pontos sobre a distribuição dos metabólitos: A tribo Antemideae (ANT) possui o maior número de ocorrência de metabólitos (7.629), e concentra 46% dessas ocorrências nas classes dos monoterpenos. A tribo também possui alta porcentagem de ocorrência, em relação às demais tribos, nas classes dos sesquiterpenos (20,6%), sesquiterpenos lactonizados (19,9%), triterpenos (13%), cumarinas (22,8%), flavonóides (16,3%) e acetofenonas (15,6%). Curiosamente, a

classe dos diterpenos não apresenta destaque na porcentagem de ocorrências, sendo essa uma das tribos com menor porcentagem de diterpenos. Outra tribo que merece destaque é a tribo Heliantheae (HLT), a segunda em número de ocorrência dentre as 17 da filogenia de Bremer. A tribo HLT possui número de ocorrência de 7.157 (19,6%) e a distribuição das ocorrências entre as doze classes é mais homogênea que a observada na tribo ANT, a única classe em que há um destaque na tribo é a classe dos sesquiterpenos lactonizados que responde por 22,5% das ocorrências de metabólitos na tribo e 25,8% das ocorrências da classe entre as 17 tribos. E por fim, a tribo Barnadesiae merece atenção por apresentar todo seu número de ocorrência concentrado somente em duas classes de metabólitos, as classes dos flavonóides (99,4%) e a dos triterpenos (0,6%).

Quando se analisa o número de ocorrência e as porcentagens entre as 39 tribos presentes na filogenia da Meta-tree é possível perceber que existe maior homogeneidade na distribuição dos metabólitos, tal constatação pode ser explicada por nessa filogenia existir um maior número de tribos, assim sendo muitas das tribos consideradas na filogenia de Bremer foram fragmentadas em outras tribos na filogenia proposta na Meta-tree. As exceções continuam nas tribos Antemideae e Barnadesieae, na primeira ainda existe concentração dos metabólitos já citados anteriormente, a segunda permanece inalterada, tendo somente ocorrência de flavonóides e triterpenos.

O último aspecto a ser discutido do banco de dados químico diz respeito aos tipos de esqueletos (ESQUEL) das doze classes de substâncias. A tabela 4.1.7 traz a quantidade e a porcentagem dos tipos de esqueletos nas doze classes. Os sesquiterpenos são a classe que possui a maior quantidade de tipos de esqueletos (163 tipos de esqueletos), assim, é a classe com maior diversidade estrutural entre

as doze estudadas da família Asteraceae. Em seguida os diterpenos (70 tipos de esqueletos), sesquiterpenos lactonizados (58 tipos de esqueletos) e monoterpenos (48 tipos de esqueletos) são as classes que possuem as maiores quantidades de tipos de esqueletos. Juntamente com os triterpenos (33 tipos de esqueletos) o grupo dos terpenóides apresenta 69,2% dos tipos de esqueletos, o que mostra que na família Asteraceae existe grande uma grande diversidade estrutural nesse grupo de substâncias.

Os poliacetilenos e os sesquiterpenos lactonizados são as classes que diferenciam quimicamente a família Asteraceae das demais famílias do grupo das Angiospermas (Heywood et al. 1977). E como pode-se observar no número de ocorrência no banco de dados químico (tabelas 4.1.5 e 4.1.6), as duas classes apresentam-se bem representadas (16% de sesquiterpenos lactonizados e 9,8 de poliacetilenos).

Apesar de ser uma classe característica da família Asteraceae e possuir um número de ocorrência alto, observa-se que os poliacetilenos possuem uma baixa variabilidade estrutural, a classe apresenta somente 9 tipos de esqueletos.

No que tange as classes de substâncias aromáticas presentes no banco de dados químicos é possível observar uma menor diversidade estrutural, em relação ao grupo dos terpenos, a classe com maior quantidade de tipos de esqueletos é a classe das cumarinas (40 tipos de esqueletos) e a com a menor quantidade de tipos de esqueletos é a classe dos fenilpropanóides (5 tipos de esqueletos).

Para entender um pouco melhor a distribuição das substâncias em tipos de esqueletos, pode-se consultar as tabelas 4.1.8 – 4.1.19, que trazem o número de ocorrência e as porcentagens de ocorrência para os cinco mais representativos tipos de esqueletos para cada classe, já as figuras 4.1.3 – 4.1.14 ilustram os cinco

principais tipos de esqueletos para cada classe. Como regra, observa-se que os cinco principais esqueletos compreendem mais de 80% das ocorrências dos tipos de esqueletos, as exceções estão na classe do sesquiterpenos, em que os cinco principais tipos de esqueletos representam 51,9% das ocorrências dos tipos de esqueletos, e na classe das acetofenonas (66,2% das ocorrências dos tipos de esqueletos).

O exposto no parágrafo anterior permite afirmar que apesar de algumas classes apresentarem uma grande variedade de tipos de esqueletos, a maior parte das ocorrências está concentrada em apenas alguns tipos. A classe dos flavonóides pode servir de exemplo para essa afirmação, somente dois tipos de esqueletos da classe representam 87% das ocorrências dos tipos de esqueletos da classe. Como já mencionado os sesquiterpenos e as acetofenonas apresentam uma distribuição menos concentrada dos tipos de esqueleto, tal fato revela que um maior número de espécies da família Asteraceae produz diferentes tipos de esqueletos dessas duas classes em relação às demais classes.

5.2. Redes Neurais

5.2.1. Redes Neurais não supervisionadas

Os resultados apresentados nas tabelas 4.2.1.1 – 4.2.1.2 e nas figuras 4.2.1 a, b e c, dizem respeito às análises em que se pretendia separar o maior número de tribos da família Asteraceae. Utilizando como entrada os dados do banco de dados químico organizado segundo a filogenia de Bremer, foi possível separar somente cinco tribos com taxas de acertos superiores a 56%. Quando se utilizou o mesmo

banco, agora organizado segundo a filogenia proposta na Meta-tree, foi possível separar sete tribos com taxas de acertos superiores a 58%. Por fim, ao utilizar como dados de entrada na rede um gênero como representante de cada tribo, obteve-se a separação de nove tribos com taxas de acertos superiores 66%.

O exposto acima mostra que a filogenia proposta na Meta-tree é mais concordante com a informação química e também permite dizer que ao utilizar gêneros representando uma tribo tem-se melhor separação.

A segunda parte das análises concentrou-se na separação de uma tribo ou um gênero, pertencente cada um a um dos quatro principais grupos em que a família se dividia até a filogenia de Bremer.

Primeiramente, separaram-se quatro tribos, uma de cada um dos quatro grupos da família Asteraceae, e os resultados estão apresentados na tabela 4.2.1.3 e figuras 4.2.1.2 a e b. Excetuando-se a tribo Mutisiae (MUT), a qual, tanto no ensaio utilizando a filogenia de Bremer quanto no ensaio utilizando a filogenia da Meta-tree, apresentou valores de acerto inferiores a 50%. As demais tribos apresentaram taxas de acertos melhores com valores superiores a 69%. Em seguida separaram-se, em duas análises (tabela 4.2.1.4 e figuras 4.2.1.3 a e b), quatro tribos utilizando como dados de entrada para a rede um gênero representando cada uma das tribos. Ambas as análises apresentaram taxas de acertos superiores a 72%. De maneira análoga aos ensaios anteriores temos nessa segunda parte que a utilização dos gêneros para a separação das tribos permite uma melhor separação.

Depois de realizados ensaios com o intuito de separar o maior número de tribos e em seguida ter realizado ensaios para separar quatro tribos, uma de cada um dos quatro principais grupos da família, decidiu-se realizar ensaios para separar três tribos de um mesmo grupo (subfamília Asteroideae) e por fim separar o maior

número de gêneros na tribo Astereae (AST). Com respeito a separação de três tribos na subfamília Asteroideae três análises foram realizadas, as tribos escolhidas foram: Gnaphalieae (GNA), Inuleae (INU) e Senecioneae (SEN). As duas primeiras análises foram realizadas utilizando os dados totais das tribos citadas, nas filogenias de Bremer e da Meta-tree. Os resultados estão descritos na tabela 4.2.1.5 e figuras 4.2.1.2 c e d.

Os resultados de acertos entre as duas classificações não apresentou diferenças consideráveis, em ambas as análises a tribo com menor taxa de acerto foi a tribo INU (65,3% em Bremer e 61% na Meta-tree). As outras duas tribos apresentaram acertos acima de 75%. Na terceira análise utilizaram-se dados de um gênero para cada uma das tribos como entrada na rede, os resultados estão mostrados na tabela 4.2.1.6 e figura 4.2.1.3 c. As taxas de acerto para essa análise foram superiores a 88%, valor bastante superior ao obtido utilizando os dados totais de cada tribo. Esse dado que reforça a idéia de que gêneros representando as tribos permitem melhores separações entre as tribos. A última análise realizada, na qual foram separados quatro gêneros da tribo Astereae (AST) (tabela 4.2.1.6 e figura 4.2.1.3 d), mostrou taxas de acertos com valores acima de 67%.

5.2.2. Redes Neurais supervisionadas

Trabalhos sobre previsão das ocorrências de metabólitos secundários na família Asteraceae utilizando classificação heurística das ocorrências em duas faixas, ausência e presença, mostraram que através da rede de perceptron de múltiplas camadas utilizando o algoritmo de retropropagação de erros foi possível

realizar a previsão da ocorrência dos metabólitos com sucesso (Ferreira et al. 2004; Emerenciano et al. 2005).

Como exposto nos objetivos desse trabalho tentou-se prever as ocorrências classificadas heurísticamente, em cinco faixas (tabela 4.2.2.3), das classes de substâncias da família Asteraceae utilizando-se a rede de perceptron de múltiplas camadas utilizando o algoritmo de retropropagação de erros. Das análises realizadas obtiveram-se os resultados presentes na tabela 4.2.2.5, onde estão apresentadas as médias de acertos nas séries de treino e teste da previsão das doze classes de metabólitos presentes nesse estudo. Na tabela citada observa-se que na série de treino a menor média de acerto é de 78% (classe dos sesquiterpenos lactonizados) e na série de teste a menor média de acerto é de 75% (classe dos triterpenos). Esses dados isoladamente levam a crer que a rede foi capaz de prever as probabilidades relativas das classes de metabólitos com sucesso. No entanto as figuras 4.2.2.1 – 4.2.2.8 (gráficos dos valores previstos pela rede versus os valores reais para as doze classes) mostram que não é essa uma realidade. Ao observar os gráficos é facilmente notado que a rede não foi capaz de prever com sucesso as ocorrências relativas dos metabólitos. A contradição entre os dois resultados está no fato de os dados de entrada na rede não serem homogêneos, a maior parte dos gêneros possui probabilidade relativa de ocorrência rara (0,2), assim sendo, a rede tem grande quantidade de dados nessa faixa para o treino e acerta bastante para essa faixa, mas nas outras faixas como a quantidade de dados é menor o treino é ineficiente e acaba por resultar em muitos erros. A tabela 4.2.2.4, construída a partir da função 4.2.2.1 leva em conta somente a quantidade de acertos total e não leva em conta a distribuição dos acertos entre as

diversas faixas, o que é visto nos gráficos dos valores previstos versus valores reais (figuras 4.2.2.1 – 4.2.2.8), fato que explica a aparente contrariedade dos resultados.

Em alguns casos os valores previstos em relação aos valores reais na série de treinos foram até bastante satisfatórios (benzopiranos, acetofenonas e fenilpropanóides), considerando-se as informações presente nos gráficos, no entanto, a informação nos gráficos da série de teste é bastante diferente, e não há concordância satisfatória entre os valores previstos e os reais mesmo nas classes em que os valores de treino são satisfatórios.

Do exposto acima se extrai que a rede não foi capaz de prever com sucesso as probabilidades de ocorrência relativa dos metabólitos secundários. Mesmo as classes que apresentaram melhores resultados na série de treino (benzopiranos, acetofenonas e fenilpropanóides), na série de teste apresentaram valores insatisfatórios. Esses dados mostram que a rede não foi hábil em criar um modelo robusto, capaz de fazer previsões para os dados não utilizados durante o seu treinamento.

Também nesse trabalho, foi feita uma tentativa de criação de um modelo para a classe dos sesquiterpenos lactonizados com uma heurística de presença e ausência, para repetir o experimento de Ferreira et al. 2001, pois o banco de dados teve um aumento de 8.000 ocorrências. Os resultados dessa rede estão no apêndice C (tabelas 5.2.1 – 5.2.2). A média de acertos do treino foi de 68% e no teste 58%.

Como a distribuição de presença e ausência entre os gêneros é homogênea, significa que a rede acerta no teste mais ou menos a mesma proporção de 1 e 0, não existindo as distorções como as encontradas nas redes com cinco faixa heurísticas.

Esses resultados são preliminares, muitos parâmetros da rede ainda estão sendo ajustados. Ainda assim, nota-se que a série de teste é melhor que nas redes de cinco faixas heurísticas.

5.3. Algoritmos genéticos

Existe consenso na literatura (Kirby e Keasling 2009; McGarvey e Croteau 1995) de que monoterpenos e diterpenos são biossintetizados em grande extensão pela via do 1-desoxi-D-xilose-5-fosfato (DOXP) (figura 3.2.4) no interior dos cloroplastos, e que sesquiterpenos, sesquiterpenos lactonizados e triterpenos são biossintetizados preferencialmente pela via do ácido mevalônico (figura 3.2.3) no citoplasma celular.

O objetivo dessa parte do trabalho foi verificar se os dados de ocorrências do grupo dos terpenos no banco de dados químicos permitiriam encontrar algum equilíbrio na produção dos esqueletos desse grupo de metabólitos, e se poderiam confirmar o exposto acima. A tabela 4.3.1 e as equações 4.3.1 – 4.3.40 (apêndice B) apresentam os resultados das análises de regressão múltipla linear (MLR) utilizando algoritmo genético como método para a seleção das variáveis.

Foram criados 40 modelos de MLR, 20 modelos utilizaram o esqueleto 1, que é o precursor biossintético dos monoterpenóides e demais terpenóides, como variável dependente (dez dos modelos com três variáveis independentes e outros dez modelos com seis variáveis independentes) e 20 modelos com o esqueleto 101, que é o precursor biossintético de todos os sesquiterpenos e triterpenos, como

variável dependente (dez dos modelos com três variáveis independentes e outros dez modelos com seis variáveis independentes).

Os esqueletos presentes nos modelos onde o esqueleto 1 foi utilizado como variável dependente, nos modelos com três variáveis independentes, são exclusivamente esqueletos da classe dos monoterpenos e sesquiterpenos, já nos modelos com seis variáveis independentes aparecem além dos esqueletos de monoterpenos e sesquiterpenos, esqueletos de sesquiterpenos lactonizados em quatro modelos e esqueletos de diterpenos em dois dos modelos. Esse resultado indica que apesar de serem preferencialmente sintetizados em organelas diferentes em vegetais superiores, os monoterpenos e sesquiterpenos estão bastante relacionados biossinteticamente. O resultado também aponta que os monoterpenos têm sua síntese relacionada à síntese dos diterpenos, resultado esperado, uma vez que se postula que ambos os metabólitos são biosintetizados no interior dos cloroplastos (Kirby e Keasling 2009; McGarvey e Croteau 1995).

Nos modelos em que o esqueleto 101 foi utilizado como variável dependente, os modelos com três variáveis, como no caso dos modelos de três variáveis da previsão do esqueleto 1, apresentaram exclusivamente esqueletos de monoterpenos e sesquiterpenos, já nos modelos com seis variáveis em quatro modelos apareceram esqueletos de sesquiterpenos lactonizados e em um modelo apareceu um esqueleto de triterpeno. Tal resultado é semelhante ao anterior, e confirma a íntima relação na produção de monoterpenos e sesquiterpenos, e mostra também, que existe equilíbrio na produção de sesquiterpenos e triterpenos, o que era esperado, pois, segundo as referências já citadas, ambos os metabólitos são sintetizados no citoplasma.

6. CONCLUSÃO

Apesar dos problemas de amostragem, discutido na seção anterior, o banco de dados químicos pode ser considerado adequado para as análises que foram realizadas, uma vez que uma porcentagem considerável de gêneros está amostrada para cada tribo, quer seja considerada a filogenia de Bremer ou a filogenia proposta na Meta-tree.

Quanto à amostragem das doze classes de substâncias, vale ressaltar que o banco foi construído através de extensiva busca na literatura entre os anos de 1960 e 2008, milhares de artigos foram consultados, artigos esses de pesquisadores de áreas diversas. Assim, apesar de em determinada época existir concentração na busca de uma dada classe de composto em outra época outra classe teve seu auge, e também como foram consultados trabalhos de pesquisadores de diferentes áreas é bastante provável que o foco de pesquisa tenha sido bastante variado no período citado.

Por fim, um grande número de trabalhos foram publicados utilizando versões cada vez mais completas desse banco (Alvarenga et al. 2001; Ferreira et al. 2004, 2005; Emerenciano et al. 2006a, 2006b; Correia et al. 2008), o que é um bom indicativo da qualidade do banco para o estudo químico da família Asteraceae.

Em relação à distribuição dos metabólitos secundários no banco (tabelas 4.1.5 – 4.1.7) pode-se concluir que a família Asteraceae é especializada na produção de metabólitos do grupo dos terpenos, pois concentra quase 70% das ocorrências e tipos de esqueletos de metabólitos nesse grupo de metabólitos. Outras classes que merecem destaque na família são os flavonóides e poliacetilenos, que possuem número de ocorrência bastante semelhante aos encontrados entre os

terpenos, mas não apresentam grande diversidade estrutural, apresentando 20 e 9 tipos de esqueletos respectivamente. As cumarinas não são tão abundantes quanto as demais classes, seu número de ocorrência é de 979, mas é uma classe de grande diversidade, apresentando 7,4% dos tipos de esqueletos do banco de dados químico. Os demais metabólitos possuem baixo número de ocorrência, não sendo sua produção tão importante na família em relação aos demais metabólitos, mas apresentam alta variabilidade estrutural, com exceção dos fenilpropanóides que possuem somente 5 tipos de esqueletos. A última conclusão a respeito da distribuição das doze classes no banco é a de que apesar da grande diversidade de tipos de esqueletos apresentada na maioria das classes, no geral somente alguns poucos esqueletos são responsáveis por cerca de 80% do número de ocorrência das classes (4.1.8-4.1.18), o que indica que existe na família Asteraceae ampla distribuição desses poucos tipos de esqueletos, e que os demais são específicos de um determinado gênero ou espécie, e assim, sua produção deve se dar em decorrência de um estímulo de algum agente particular ao ambiente em que as espécies ou os gêneros estão localizados.

No que tange aos resultados dos mapas auto-organizáveis, a conclusão é de que a ferramenta utilizada é adequada para a taxonomia da família Asteraceae, como já se tinha observado em estudos anteriores (Costa *et al.* 2005; Hristov *et al.* 2007; Scotti 2008; Correia *et al.* 2008), já que em diversos níveis foi possível realizar a separação de tribos e gêneros.

Mais especificamente pode se dizer que os resultados permitiram determinar que a informação química está em maior concordância com a filogenia proposta na Meta-treee (Tabela 4.2.2.1 e figuras 4.2.1.1 a e b), também permitiram verificar que a utilização de um gênero como representante de uma dada tribo torna a separação

entre as tribos mais eficiente. Tal ocorrência está ligada ao fato de os gêneros de uma dada tribo não serem homogêneos com relação a sua química e por conta disso a soma dos gêneros na tribo dificulta a identificação de padrões separáveis para cada tribo pela rede SOM. Finalmente, pôde ser constatado que ao separar gêneros de tribos diferentes a taxa de acerto é maior do que ao separar gêneros de uma mesma tribo, resultado que indica que apesar de os gêneros de uma mesma tribo serem distintos quimicamente, eles são mais próximos entre si do que entre gêneros de tribos diferentes.

No tópico de previsão das ocorrências de metabólitos secundários utilizando rede neural MLP, os resultados não foram satisfatórios. Não foi possível criar modelos adequados de previsão para nenhuma das classes quando se utilizou a classificação heurística dos dados em cinco faixas (tabela 4.2.2.3). O fracasso na construção dos modelos está ligado à falta de homogeneidade na distribuição das probabilidades relativas de ocorrência nos gêneros (tabela 4.2.2.6). Nas doze classes de metabólitos mais do que 80% das ocorrências são classificadas na faixa heurística de 0,2 (probabilidade relativa rara), assim sendo a rede não tem exemplos suficientes das outras faixas de probabilidades no treino e assim a previsão não é realizada de maneira satisfatória.

Em relação à rede de presença e ausência, os resultados ainda não são muito bons, mas os experimentos estão sendo aperfeiçoados, os resultados mostrados são preliminares.

No último tópico desse trabalho, que foi a construção de modelos de MLR, utilizando algoritmos genéticos, os resultados permitem indicar que monoterpenos e sesquiterpenos são bastante correlacionados biossinteticamente, resultado que vai de encontro a alguns trabalhos da literatura (Adam e Zapp, 1998; Dudareva et al.

2005; Bartram et al. 2006; Laule et al. 2003), onde há indicação de que alguns sesquiterpenos são produzidos, em condições normais, no cloroplasto pela via DOXP, e também que em condições onde há inibição da via do mevalonato a síntese dos sesquiterpenos é realizada também nos cloroplastos. Esses dados e os obtidos aqui indicam que possa existir algum tipo de equilíbrio entre as duas vias para a produção de sesquiterpenos, já com relação aos monoterpenos, nos trabalhos citados a indicação é de que essa classe de substâncias deva ser exclusivamente produzida pela via do DOXP, nos cloroplastos. Assim, conclui-se que deve existir algum mecanismo capaz de regular a síntese dos sesquiterpenos entre o citoplasma e os cloroplastos, e ajustar sua síntese à dos monoterpenos nos cloroplastos. Também nessa parte do estudo foi possível revelar que existe uma relação entre monoterpenos e diterpenos e entre os sesquiterpenos e triterpenos, resultado consoante com a literatura (Kirby e Keasling 2009; McGarvey e Croteau 1995; Umlauf et al. 2004) e que reforça a proposição de que o banco de dados químico criado é adequado para obter informações úteis da química da família Asteraceae.

7. Referências

- Adam KP, Zapp J. Biosynthesis of the isoprene units of chamomile sesquiterpenes. *Phytochemistry* 1998; 48(6): 953-959.
- Alvarenga SAV, Ferreira MJP, Emerenciano VP, Cabrol-Bass D. Chemometrics and Intelligent Laboratory Systems 2001; 56: 27-37.
- The Angiosperm Phylogeny Group (APG). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* 2003;141: 399-436.
- Bartram S, Jux A, Gleixner G, Boland W. Dynamic pathway allocation in early terpenoid biosynthesis of stress-induced lima bean leaves. *Phytochemistry* 2006; 67:1661-1672.
- Bentham, G. Compositae. Pp. 163–533 in: Bentham, G. & Hooker, J.D. (eds.), *Genera Plantarum*, vol. 2(1). London: Reeve, 1873.
- Berkhey, JLFV. *Expositio Characteristica Structurae Florum qui Dicuntur Compositi*. Van der Eyk, Leiden: 1760.
- Brant AJC. Flavonóides, Cumarinas e Benzofuranos como marcadores quimiotaxonômicos em Asteraceae. São Paulo, Dissertação de Mestrado – Instituto de Química – Universidade de São Paulo 2003; 169p.
- Bremer K. Tribal interrelationships of the Asteraceae. *Cladistics*. 1987;3:210-253.
- Bremer K, Jansen RK, Karis PO, Kallersjö M, Keeley SC, Kim KJ, Michaels HJ, Palmer JD, Wallace RS. A review of the phylogeny and classification of Asteraceae. *Nordical Journal of Botany*. 1992;12:141-148.
- Bremer K. Asteraceae - Cladistic & classification. Portland (OR): Timber Press; 1994.
- Bremer K Major clades and grades of the Asteraceae. In Compositae: Systematics In Hind DJN, Beentje H, editors. *Compositae: Systematics. Proceedings of the International Compositae Conference*. Vol. 1. Kew: Royal Botanic Garden; 1996: 1-7.
- Calabria, ML. Emerenciano, VP. Ferreira, MJP. Scotti, MT. Mabry, TJ. Phylogenetic analysis of tribes of the Asteraceae based on phytochemical data. *Natural Products Communications, Estados Unidos*. 2007; 2(3): 277-85
- Carlquist S. Tribal interrelationships and phylogeny of the Asteraceae. *Aliso* 1976; 8: 465–492.
- Carvalho ACPLF, Braga AP, Ludemir TB. Computação Evolutiva. In Rezende SO, editor. *Sistemas Inteligentes – Fundamentos e Aplicações*, 1º Ed. São Paulo: Ed. Manole 2003.

Cassini H. Tableau exprimant les affinités des tribus naturelles de famille des Synanthérées. In Cloquet MH, editor. Dictionnaire des Sciences Naturelles, vol 3. 2nd. Le Normant (Paris): Ed. G. Cuvier; 1816.

Cleva C, Cachet C, Cabrol-bass D. Clustering of infrared spectra with Kohonen networks. *Analisis*. 1999;27: 81-90.

Corne AS. Artificial Neural Networks for Pattern Recognition. *Concepts in Magnetic Resonance* 1996; 8(5): 303-324.

Correia MV, Scotti MT, Ferreira MJP, Alvarenga SAV, Rodrigues GV, Emerenciano VP. Self-Organizing Maps as a New Tool for Classification of Plants at Lower Hierarchical Levels. *Natural Products Communications, Estados Unidos*. 2008; 3(10): 1723-1730.

Cronquist A. Phylogeny and taxonomy of the Compositae. *American Midland Naturalist* 1955; 53: 488–511.

Cronquist A. The Compositae revisited. *Brittonia* 1997; 29:137–153.

Da Costa FB, Terfloth L, Gasteiger J. Sesquiterpene lactone-based classification of three Asteraceae tribes: a study based on self-organizing neural networks applied to chemosystematics. *Phytochemistry* 2005; 66: 345-353.

Darwin C. On the Origin of Species by Means of Natural Selection, or the Preservation Favoured Races in the Struggle for Life (1st ed.), London: John Murray 1859.

Dewick, M. P. Medicinal Natural Products. England: John Wiley & Sons, Ltd 2002.

Disch A, Schwender J, Muller C, Lichtenthaler HK, Rohmer M. Distribution of the mevalonate and glyceraldehyde phosphate/pyruvate pathways for isoprenoid biosynthesis in unicellular algae and the cyanobacterium *Synechocystis* PCC 6714. *Biochemical Journal* 1998; 333(2): 381-388.

Doucet J, Panaye A, Feuilleaubois E, Ladd P. Neural networks and ¹³C NMR chemical shift prediction. *Journal of Chemical Information and Modeling* 1993; 33: 320-324.

Dudareva N, Andersson S, Orlova I, Gatto N, Reichelt M, Rhodes D, Boland W, Gershenzon J. The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. *Proc. Natl. Acad. Sci. USA* 2005; 102(3): 933-938.

Emerenciano VP, Kaplan MAC, Gottlieb OR. The evolution of sesquiterpene lactones in angiosperms. *Biochemical Systematics and Ecology* 1985; 3: 145-166.

Emerenciano VP, Ferreira ZS, Kaplan MAC, Gottlieb OR. A chemosystematic analysis of tribes of Asteraceae involving sesquiterpene lactones and flavonoids. *Phytochemistry* 1987; 26: 3103-3115.

Emerenciano VP, Ferreira MJP, Branco MD, Dubois JE. The applications of Bayes theorem in natural products as a guide for skeletons identifications. *Chemometrics and Intelligent Laboratory Systems* 1998a; 4: 83-92.

Emerenciano VP, Rodrigues GV, S. Alvarenga SAV, Macari PAT, Kaplan MAC. Um método para união de vários marcadores quimiotaxonômicos. *Química Nova* 1998b; 21: 125-129.

Emerenciano VP, Schwabe T, Ferreira MJP, Alvarenga SAV. Neural Networks for secondary metabolites prediction in *Artemisia* genus Astereaceae. *Internet Eletronical Journal of Molecular Design* 2005; 4: 9-16.

Emerenciano VP, Scotti MT, Stefani R, Alvarenga SAV, Nuzillard JM, Rodrigues GV. Diterpene Skeletal Type Classification and Recognition using Self-Organizing Maps. *Internet Electronic Journal of Molecular Design* 2006; 5(4): 213-23.

Emerenciano VP, Alvarenga SAV, Scotti MT, Ferreira MJP, Stefani R, Nuzillard J-M. Automatic identification of terpenoid skeletons by feed-forward neural networks. *Analytica Chimica Acta*. 2006a; 579:217–26.

Emerenciano VP, Scotti MT, Stefani R, Alvarenga SAV, Nuzillard JM, Rodrigues GV.. Diterpene Skeletal type classification and recognition using Self-Organization Maps. *Internet Eletronical Journal of Molecular Design*, EUA 2006b; 5: 213-223.

Fernandes MB, Scotti MT, Ferreira MJP, Emerenciano VP. Use of selforganizing maps and molecular descriptors to predict the cytotoxic activity of sesquiterpene lactones. *European Journal of Medicinal Chemistry* 2008; 43(10): 2197-2205.

Ferreira MJP, Brant AJC, Rufino AR, Alvarenga SAV, Magri FMM, Emerenciano VP. Prediction of occurrences of divers chemical classes in Asteraceae Family through artificial neural networks. *Phytochemical Analysis* 2004; 15: 389-396.

Ferreira ZS, Gottlieb OR. Polyacetylenes as systematic markers in dicotyledons. *Biochemical Systematics and Ecology* 1982; 10: 155-160.

Fraser LA, Mulholland A, Fraser DD. Classification of limonoids and protolimonoids using neural networks. *Phytochemical analysis*. 1997;8:301-11.

Funk VA, Bayer RJ, Keeley S, Chan R, Watson L, Gemeinholzer B, Schilling E, Panero JL, Baldwin BG, Garcia-Jacas N, Susanna A, Jansen RK. Everywhere but Antarctica: using a supertree to understand the diversity and distribution of the Compositae. *Biologische Skrifter*. 2005; 55: 343–74.

Funk VA, Susanna A, Stuessy T, Bayer R. Systematics Evolution and Biogeography of the Compositae, Washington: IAPT; 2009.

Gasteiger J, Teckentrup A, Terfloth L, Spycher S. Neural networks as data mining tools in drug design. Journal of Physical Organic Chemistry 2003;16:232-245.

Gershenson J, Mabry TJ. Secondary Metabolites and the Higher classification of angiosperms Nord. J. Bot. 1983; 3: 5-34.

GERTAXO, Rodrigues GV. Software não registrado; 2004.

Gottlieb OR. Micromolecular Evolution Systematics and Ecology, springer- verlag, Berlim;1982.

Gottlieb OR. The role of oxygen in phytochemical evolution towards diversity. Phytochemistry. 1989;28:2545-2558.

Gottlieb OR. Phytochemicals: Differentiation and function. Phytochemistry 1990; 29: 1715-1729.

Gottlieb OR, Kaplan MAC, Borin, MRMB. Biodiversidade: um enfoque químicobiológico. Rio de Janeiro: Editora da UFRJ; 1996.

Grayer RJ, Chase MW, Simmonds MSJ. A comparison between chemical and molecular characters for the determination of phylogenetic relationships among plant families: An appreciation of Hegnauer's "Chemotaxonomie der Pflanzen". Biochemical Systematics and Ecology 1999; 27: 369-393.

Gulbag A, Temurtas F. A study on quantitative classification of binary gas mixture using neural networks and adaptive neuro-fuzzy inference systems. Sensors and Actuators B 2006; 115: 252-262.

Gupta S, Aires-de-Sousa J. Comparing the chemical spaces of metabolites and available chemicals: models of metabolite-likeness. Molecular Diversity 2007;11(1):23-36.

Harborne JB. Ecological Biochemistry. 3rd. London: Academic Press; 1988.

Haykin S. Redes Neurais Principios e prática, 2º Ed. Porto Alegre: Bookman; 2001.

Hebb DO. The Organization of Behavior: A neuropsychological Theory, New York: Wiley; 1949.

Hegnauer, R. Chemotaxonomie der Planzen, vol. 7. Birkhäuser-Verlag, Basel;1986.

Hendrickson JB, Cram DJ, Hammond GS. Organic Chemistry. 3º ed. New York: Ed. McGraw-Hill; 1970; 1-1280.

Herman PPJ, Retief E, Koekemoer M, Welman WG. "Asteraceae (Compositae)". In: Leistner OA, editor. Seed Plants of Southern Africa. Strelitzia vol 10. National Botanical Institute, Pretoria; 2000: 101-170.

Heywood VH, Harborne JB, Turner BL. The Biology and Chemistry of the Compositae, Vols I and II New York: Academic Press; 1977.

Hoffman O. Compositae. In : Engler A, Prantl K, editors. Die Natürlichen Pflanzenfamilien . Vol . 4 Leipzig: Engelmann; 1890 . P. 87–381.

Hopfield JJ. Neural Networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA 1982; 79: 2554-2558.

Hristozov D, Da Costa FB, Gasteiger J. Sesquiterpene Lactones-Based Classification of the Family Asteraceae Using Neural Networks and k-Nearest Neighbors. Journal of Chemical Information and Modeling. 2007;47(1):9-19.

Ito M, Watanabe K, Kita Y, Kawahara T, Crawford DJ, Yahara T. Phylogeny and phytogeography of Eupatorium (Eupatorieae, Asteraceae): insights from sequence data of the nrDNA ITS regions and cpDNA RFLP. J. Plant Res. 2000 113(1109): 79-89

Jeffrey C. Compositae. In Heywood VH, editor Flowering Plants of the World. New York: Mayflower Books; 1978: 263-268.

Jansen RK, Kim K. Implications of chloroplast DNA data for the classification and phylogeny of the Asteraceae. In Hind DJN, Beentje H, editors. Compositae Systematics. Proceedings of the International Compositae Conference. Vol. 1. Kew: Royal Botanic Garden; 1996 p.317-339.

Junqueira LC, Carneiro J. Histologia Básica. Rio de Janeiro: Editora Guanabara Koogan S.A.; 2004.

Kim K, Jansen RK. ndhF sequence evolution and the major clades of the sunflower family. Proceedings of the National Academy of Science USA 1995;92:10379-10383.

Kirby J, Keasling JD. Biosynthesis of Plant Isoprenoids: Perspectives for Microbial Engineering. Annu. Rev. Plant Biol. 2009; 60: 335-355.

Kohonen T. Self-Organized formation of topologically correct feature maps. Biological Cybernetics 1982; 43: 59-69.

Kohonen T. Self-Organizing Maps, volume 30 of Springer Series in Information Sciences. 3rd. Heidelberg (Berlin): Springer; 2001.

Kovács LZ. Redes neurais artificiais: fundamentos e aplicações. 3ed, São Paulo: Editora da Livraria da Física; 2002.

Laule O, Fürholz A, Chang HS, Zhu T, Wang X, Heifetz PB, Grissem W, Lange Markus. Crosstalk between cytosolic and plastidial pathways of isoprenoid

biosynthesis in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* 2003; 100(11): 6866-6871.

Lawrence J. *Introduction to Neural Networks – Design, Theory and Applications*. Nevada City: California Scientific Software; 1994.

Leardi R, Boggia R, Terrile M. Genetic algorithms as a strategy for feature selection. *Journal Of Chemometrics*. 1992;6(5):267-81.

Lessing, C.F. *Synopsis Generum Compositarum*. Berlin: Duncker & Humblot; 1832.

Lichtenthaler HK. The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annual Review of Plant Physiology and Plant Molecular Biology*. 1999;50:47-65.

Linden R. *Algoritmos Genéticos*. Rio de Janeiro: Brasport; 2006.

Lyova L, Martinelli E, Dini F, Bergamini A, Paolesse R, Di Natale C, D'Amico A. Clinical analysis of human urine by means of potentiometric Electronic tongue. *Talanta* 2009; 77: 1097-1104.

Mann J., Davidson S., Hobbs J., Banthorpe D., Harborne J. *Natural Products – Their Chemistry and Biological Significance*, Prentice Hall: 1994 .

MATLAB® The MathWorks, Natick, Massachusetts, 2004.

Mcculloch WS, Pitts W. A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics*. 1943;5:115-37.

McGarvey DJ, Croteau R. Terpenoid Metabolism. *The Plant Cell* 1995; 7: 1015-1026.

Minsky M, Papert S. *Perceptrons*. Cambridge: MIT Press; 1969.

Panero JL, Funk VA. The value of sampling anomalous taxa in phylogenetic studies: major clades of the Asteraceae revealed. *Molecular Phylogenetics and Evolution* 2008; 47: 757–782.

Poljakov PP. (Systematics and Origin of the Compositae) Alma-atka: Nauka (In Russian.); 1967.

Raven PH, Evert RF, Eichhorn SE. *Biologia Vegetal*. Brasil: Guanabara Koogan; 2001.

Rechenberg, I. Cybernetic solution path of an experimental problem. Royal Aircraft Establishment, Library translation no. 1122, Farnborough, Hants, UK, ago. 1965.

Robinson H. A revision of the tribal and subtribal limits of Heliantheae (Asteraceae). *Smithsonian Contributions to Botany* 1981; 51: 1-102.

Robinson H. A generic review of the tribe Liabeae(Asteraceae). Smithsonian Contributions to Botany 1983; 54: 1-69.

Rohmer M, Knani M, Simonin P, Sutter B, Sahm H. Isoprenoid biosynthesis in bacteria - a novel pathway for the early steps leading to isopentenyl diphosphate. Biochemical journal. 1993;295:517-24.

Rosenblat F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review 1958; 65: 386-408.

Rumelhart DE, McClelland, editors. Parallel Distributed Processing: Explorations in the Microstructure of Cognition vol.1, Cambridge, MA: MIT Press; 1986.

Scotti MT. Emprego de Redes Neurais e de Descritores Moleculares em Quimiotaxonomia da Família Asteraceae. São Paulo, Tese de Doutorado – Instituto de Química – Universidade de São Paulo 2008; 165p.

Soltis DE, Soltis PS, Endress PK, Chase MW. Phylogeny and Evolution of Angiosperms. USA: Sinauer Associates; 2005.

Statistica Neural Networks (2000). Statsoft, Inc. Tulsa, OK, EUA.

Stuessy TF. Plant taxonomy the systematic evaluation of comparative data. New York: Columbia university press; 1990.

Talete, S. R. L. Mobydigs Academic version - Version 1.0 – 2004 – <http://www.talete.mi.it>.

Thorne RF. Proposed new realignments in the Angiosperms. Nord. J. Bot 1983; **3**: 85–117.

Umlauf D, Zapp J, Becker H, Adam KP. Biosynthesis of irregular monoterpene Artemisia ketone, the sesquiterpene germacrene D and other isoprenoids in *Tanacetum vulgare* L. (Asteraceae). Phytochemistry 2004; 65: 2463-2470.

Vesanto J, Himberg J, Alhoniemi E, Parhankangas J. Self-Organizing Map in Matlab: the SOM Toolbox. Proceeding of the Matlab DSP Conference. 1999; 35–40.

Vesanto, J. Himberg, E. Alhoniemi and J. Parhankangas, SOM Toolbox 2.0 for Matlab 5, 2005, <http://www.cis.hut.fi/projects/somtoolbox>.

Wagenitz G. Systematics and phylogeny of the Compositae (Asteraceae). Plant Systematics and Evolution. 1976;125:29-46.

Wagner S, Hofmann A, Siedle B, Terfloth L, Merfort I, Gasteiger J. Development of a Structural Model for NF- κ B Inhibition of Sesquiterpene Lactones Using Self-Organizing Neural Networks. Journal of Medicinal Chemistry. 2006;49(7):2241-52.

Waterman PG. The current status of chemical systematic. Phytochemistry 2007; 68: 2896-2903.

Zdero C, Bohlmann F. Systematics and evolution within the Compositae, seen with the eyes of a chemist. *Plant Systematics Evolution*. 1990;171:1-14.

Zupan E, Gasteiger J. *Neural Networks for Chemists – An Introduction*. Weinheim: VCH; 1993.

APÊNDICES

APÊNDICE A - Esqueletos

MONOTERPENOS

MONOTERPENOS

37

38

39

40

41

42

43

44

45

46

47

48

SESQUITERPENOS

SESQUITERPENOS

SESQUITERPENOS

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

SESQUITERPENOS

SESQUITERPENOS

SESQUITERPENOS

SESQUITERPENOS

SESQUITERPENOS

SESQUITERPENOS

SESQUITERPENOS LACTONIZADOS

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

SESQUITERPENOS LACTONIZADOS

SESQUITERPENOS LACTONIZADOS

SESQUITERPENOS LACTONIZADOS

DITERPENOS

DITERPENOS

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

DITERPENOS

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

DITERPENOS

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

TRITERPENOS

TRITERPENOS

TRITERPENOS

CUMARINAS

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

CUMARINAS

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

CUMARINAS

592

593

594

595

596

597

598

599

600

FLAVONÓIDES

651

652

653

654

655

656

657

658

659

660

661

FLAVONÓIDES

662

663

664

665

666

667

668

669

670

BENZOFURANOS

711

712

713

714

715

716

717

718

719

720

721

722

BENZOFURANOS

723

724

725

726

727

728

729

730

731

732

733

734

735

BENZOPIRANOS

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

BENZOPIRANOS

ACETOFENONAS

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

ACETOFENONAS

808

809

810

811

812

813

814

815

816

ACETOFENONAS

817

818

819

820

821

822

823

824

825

FENILPROPANÓIDES

831

832

833

834

835

APÊNDICE B – Equações (MLR)

$$\text{Esq1} = 0,18 (\pm 0,01)\text{Esq15} + 4,07(\pm 0,43)\text{Esq41} + 1,63(\pm 0,15)\text{Esq128} + 0,16(\pm 0,19)$$

Equação 4.3.1

$$\text{Esq1} = 0,16 (\pm 0,01)\text{Esq15} + 4,48(\pm 0,43)\text{Esq41} + 0,68(\pm 0,06)\text{Esq207} + 0,19(\pm 0,19)$$

Equação 4.3.2

$$\text{Esq1} = 4,46 (\pm 0,47)\text{Esq41} + 0,36(\pm 0,03)\text{Esq44} + 0,64(\pm 0,08)\text{Esq134} + 0,37(\pm 0,20)$$

Equação 4.3.3

$$\text{Esq1} = 4,13 (\pm 0,45)\text{Esq41} + 0,37(\pm 0,03)\text{Esq44} + 1,66(\pm 0,16)\text{Esq128} + 0,31(\pm 0,19)$$

Equação 4.3.4

$$\text{Esq1} = 0,17 (\pm 0,01)\text{Esq15} + 4,52(\pm 0,46)\text{Esq41} + 0,58(\pm 0,09)\text{Esq134} + 0,26(\pm 0,20)$$

Equação 4.3.5

$$\text{Esq1} = 0,14 (\pm 0,02)\text{Esq15} + 5,23(\pm 0,48)\text{Esq41} + 0,34(\pm 0,06)\text{Esq196} + 0,33(\pm 0,20)$$

Equação 4.3.6

$$\text{Esq1} = 0,17 (\pm 0,01)\text{Esq15} + 4,45(\pm 0,46)\text{Esq41} + 0,17(\pm 0,03)\text{Esq153} + 0,34(\pm 0,20)$$

Equação 4.3.7

$$\text{Esq1} = 4,64 (\pm 0,45)\text{Esq41} + 0,32(\pm 0,03)\text{Esq44} + 0,69(\pm 0,06)\text{Esq207} + 0,33(\pm 0,19)$$

Equação 4.3.8

$$\text{Esq1} = 4,52 (\pm 0,48)\text{Esq41} + 0,36(\pm 0,03)\text{Esq44} + 0,18(\pm 0,03)\text{Esq153} + 0,48(\pm 0,20)$$

Equação 4.3.9

$$\text{Esq1} = 0,14 (\pm 0,02)\text{Esq15} + 1,22(\pm 0,18)\text{Esq39} + 0,70(\pm 0,10)\text{Esq134} + 0,29(\pm 0,21)$$

Equação 4.3.10

$$\text{Esq1} = 0,19 (\pm 0,01)\text{Esq15} - 1,05(\pm 0,34)\text{Esq23} + 3,50(\pm 0,42)\text{Esq41} - 1,39$$

Equação 4.3.11

$$(\pm 0,26)\text{Esq114} + 2,08(\pm 0,16)\text{Esq128} + 0,20(\pm 0,04)\text{Esq149} + 0,14(\pm 0,18)$$

$$\text{Esq1} = 0,19 (\pm 0,01)\text{Esq15} + 3,76(\pm 0,42)\text{Esq41} - 1,34(\pm 0,26)\text{Esq114} + 2,09 (\pm 0,16)\text{Esq128} + 0,19(\pm 0,04)\text{Esq149} - 0,02(\pm 0,01)\text{Esq301} + 0,13(\pm 0,18)$$

Equação 4.3.12

$$\text{Esq1} = 0,19 (\pm 0,01)\text{Esq15} + 3,91(\pm 0,43)\text{Esq41} - 1,10(\pm 0,31)\text{Esq114} + 2,06 (\pm 0,16)\text{Esq128} + 0,18(\pm 0,04)\text{Esq149} - 0,03(\pm 0,02)\text{Esq302} + 0,14(\pm 0,18)$$

Equação 4.3.13

$$\text{Esq1} = 0,13 (\pm 0,01)\text{Esq15} + 4,72(\pm 0,39)\text{Esq41} + 1,71(\pm 0,22)\text{Esq128} + 0,22 (\pm 0,03)\text{Esq149} + 0,51(\pm 0,09)\text{Esq207} - 3,94(\pm 0,44)\text{Esq212} + 0,07(\pm 0,17)$$

Equação 4.3.14

$$\text{Esq1} = 0,19 (\pm 0,01)\text{Esq15} + 3,68(\pm 0,42)\text{Esq41} - 1,40(\pm 0,26)\text{Esq114} + 2,09 (\pm 0,16)\text{Esq128} + 0,19(\pm 0,04)\text{Esq149} - 0,09(\pm 0,07)\text{Esq317} + 0,11(\pm 0,18)$$

Equação 4.3.15

$$\text{Esq1} = 0,19 (\pm 0,01)\text{Esq15} + 3,68(\pm 0,41)\text{Esq41} - 1,35(\pm 0,26)\text{Esq114} + 2,04 (\pm 0,16)\text{Esq128} + 0,19(\pm 0,04)\text{Esq149} + 0,54(\pm 0,20)\text{Esq416} + 0,05(\pm 0,18)$$

Equação 4.3.16

$$\text{Esq1} = 0,19 (\pm 0,01)\text{Esq15} + 3,72(\pm 0,42)\text{Esq41} - 1,37(\pm 0,26)\text{Esq114} + 2,09 (\pm 0,16)\text{Esq128} + 0,19(\pm 0,04)\text{Esq149} + 0,09(\pm 0,18)$$

Equação 4.3.17

$$\text{Esq1} = 0,19 (\pm 0,01)\text{Esq15} + 3,72(\pm 0,42)\text{Esq41} - 1,38(\pm 0,26)\text{Esq114} + 2,07 (\pm 0,16)\text{Esq128} + 0,19(\pm 0,04)\text{Esq149} - 0,00(\pm 0,08)\text{Esq253} + 0,09(\pm 0,18)$$

Equação 4.3.18

$$\text{Esq1} = 0,19 (\pm 0,01)\text{Esq15} + 3,72(\pm 0,42)\text{Esq41} - 1,38(\pm 0,26)\text{Esq114} + 2,07 (\pm 0,16)\text{Esq128} + 0,19(\pm 0,04)\text{Esq149} + 0,01(\pm 0,46)\text{Esq456} + 0,13(\pm 0,18)$$

Equação 4.3.19

$$\text{Esq1} = 0,19 (\pm 0,01)\text{Esq15} + 3,69(\pm 0,42)\text{Esq41} - 1,39(\pm 0,26)\text{Esq114} + 2,08 (\pm 0,16)\text{Esq128} + 0,19(\pm 0,04)\text{Esq149} - 0,15(\pm 0,11)\text{Esq319} + 0,10(\pm 0,18)$$

Equação 4.3.20

$$\text{Esq101} = 0,23 (\pm 0,00)\text{Esq42} + 0,18(\pm 0,01)\text{Esq133} + 2,11(\pm 0,10)\text{Esq203} + 0,13(\pm 0,07)$$

Equação 4.3.21

$$\text{Esq101} = 0,18 (\pm 0,00)\text{Esq42} + 0,16(\pm 0,01)\text{Esq133} + 1,24(\pm 0,06)\text{Esq234} + 0,09(\pm 0,07)$$

Equação 4.3.22

$$\text{Esq101} = 0,29 (\pm 0,01)\text{Esq37} + 0,19(\pm 0,01)\text{Esq133} + 2,11(\pm 0,11)\text{Esq203} + 0,11(\pm 0,07)$$

Equação 4.3.23

$$\text{Esq101} = 0,23 (\pm 0,01)\text{Esq37} + 0,17(\pm 0,01)\text{Esq133} + 1,24(\pm 0,06)\text{Esq234} + 0,07(\pm 0,07)$$

Equação 4.3.24

$$\text{Esq101} = 0,19 (\pm 0,00)\text{Esq42} + 0,17(\pm 0,01)\text{Esq149} + 1,43(\pm 0,06)\text{Esq234} + 0,20(\pm 0,07)$$

Equação 4.3.25

$$\text{Esq101} = 0,25 (\pm 0,01)\text{Esq37} + 0,18(\pm 0,01)\text{Esq149} + 1,44(\pm 0,06)\text{Esq203} + 0,19(\pm 0,07)$$

Equação 4.3.26

$$\text{Esq101} = 0,13 (\pm 0,01)\text{Esq1} + 0,32(\pm 0,03)\text{Esq8} + 1,28(\pm 0,06)\text{Esq234} + 0,18(\pm 0,07)$$

Equação 4.3.27

$$\text{Esq101} = 0,64 (\pm 0,02)\text{Esq39} + 0,16(\pm 0,01)\text{Esq133} + 1,25(\pm 0,06)\text{Esq234} + 0,05(\pm 0,07)$$

Equação 4.3.28

$$\text{Esq101} = 0,56 (\pm 0,01)\text{Esq8} + 0,18(\pm 0,01)\text{Esq149} + 1,50(\pm 0,06)\text{Esq234} + 0,22(\pm 0,06)$$

Equação 4.3.29

$$\text{Esq101} = 0,08 (\pm 0,02)\text{Esq15} + 0,16(\pm 0,01)\text{Esq149} + 1,31(\pm 0,06)\text{Esq203} + 0,06(\pm 0,06)$$

Equação 4.3.30

$$\text{Esq101} = 0,09 (\pm 0,01)\text{Esq1} + 0,14(\pm 0,01)\text{Esq42} + 0,16(\pm 0,03)\text{Esq113} + 0,10 (\pm 0,01)\text{Esq149} + 1,03(\pm 0,15)\text{Esq212} + 0,78(\pm 0,08)\text{Esq234} + 0,09(\pm 0,06)$$

Equação 4.3.31

$$\text{Esq101} = 0,09 (\pm 0,01)\text{Esq1} + 0,11(\pm 0,01)\text{Esq42} + 0,19(\pm 0,03)\text{Esq113} + 0,10 (\pm 0,01)\text{Esq149} + 0,94(\pm 0,08)\text{Esq234} + 0,03(\pm 0,01)\text{Esq234} + 0,04(\pm 0,06)$$

Equação 4.3.32

$$\text{Esq101} = 0,10 (\pm 0,01)\text{Esq1} + 0,42(\pm 0,02)\text{Esq8} + 0,14(\pm 0,01)\text{Esq149} + 0,27 (\pm 0,08)\text{Esq171} + 1,10(\pm 0,15)\text{Esq212} + 0,85(\pm 0,08)\text{Esq234} + 0,11(\pm 0,06)$$

Equação 4.3.33

$$\text{Esq101} = 0,10 (\pm 0,01)\text{Esq1} + 0,41(\pm 0,02)\text{Esq8} + 0,44(\pm 0,02)\text{Esq133} + 0,10 (\pm 0,22)\text{Esq149} + 1,16(\pm 0,15)\text{Esq212} + 0,95(\pm 0,07)\text{Esq234} + 0,10(\pm 0,06)$$

Equação 4.3.34

$$\text{Esq101} = 0,09 (\pm 0,01)\text{Esq1} + 0,31(\pm 0,05)\text{Esq8} + 0,04(\pm 0,02)\text{Esq42} + 0,14 (\pm 0,01)\text{Esq149} + 1,17(\pm 0,15)\text{Esq212} + 0,99(\pm 0,06)\text{Esq234} + 0,14(\pm 0,06)$$

Equação 4.3.35

$$\text{Esq101} = 0,10 (\pm 0,01)\text{Esq1} + 0,42(\pm 0,02)\text{Esq8} + 0,25(\pm 0,01)\text{Esq23} + 0,14 (\pm 0,01)\text{Esq149} + 1,16(\pm 0,15)\text{Esq212} + 1,00(\pm 0,06)\text{Esq234} + 0,12(\pm 0,06)$$

Equação 4.3.36

$$\begin{aligned} \text{Esq101} = & 0,10 (\pm 0,01) \text{Esq1} + 0,41 (\pm 0,02) \text{Esq8} + 0,14 (\pm 0,01) \text{Esq149} + 1,19 \\ & (\pm 0,15) \text{Esq212} + 0,99 (\pm 0,07) \text{Esq234} - 0,02 (\pm 0,01) \text{Esq303} + 0,15 (\pm 0,06) \end{aligned}$$

Equação 4.3.37

$$\begin{aligned} \text{Esq101} = & 0,08 (\pm 0,01) \text{Esq1} + 0,27 (\pm 0,03) \text{Esq8} + 0,16 (\pm 0,03) \text{Esq113} + 0,12 \\ & (\pm 0,01) \text{Esq149} + 1,12 (\pm 0,07) \text{Esq234} + 0,04 (\pm 0,01) \text{Esq311} + 0,08 (\pm 0,06) \end{aligned}$$

Equação 4.3.38

$$\begin{aligned} \text{Esq101} = & 0,10 (\pm 0,01) \text{Esq1} + 0,40 (\pm 0,02) \text{Esq8} + 0,15 (\pm 0,01) \text{Esq149} + 1,14 \\ & (\pm 0,15) \text{Esq212} + 0,98 (\pm 0,07) \text{Esq234} + 0,02 (\pm 0,00) \text{Esq301} + 0,10 (\pm 0,06) \end{aligned}$$

Equação 4.3.39

$$\begin{aligned} \text{Esq101} = & 0,10 (\pm 0,01) \text{Esq1} + 0,42 (\pm 0,02) \text{Esq8} + 0,14 (\pm 0,01) \text{Esq149} + 1,19 \\ & (\pm 0,15) \text{Esq212} + 1,00 (\pm 0,07) \text{Esq234} - 0,11 (\pm 0,07) \text{Esq528} + 0,15 (\pm 0,06) \end{aligned}$$

Equação 4.3.40

APÊNDICE C – Tabelas de Acertos e Erros - Rede Supervisionada.

Tabela 5.2.1. Acertos e erros da rede supervisionada por gêneros para os sesquiterpenos lactonizados, com heurística de 0 e 1, na fase de treino.

Gêneros	Valor Real	Valor Previsto	Acertos
ARTEMISIA	1.0	1.1	1.0
HELICHRYSUM	1.0	0.8	1.0
EUPATORIUM	1.0	1.0	1.0
LIATRIS	1.0	1.0	1.0
ACHILLEA	1.0	1.0	1.0
CENTAUREA	1.0	0.8	1.0
SENECIO	1.0	0.9	1.0
TANACETUM	1.0	1.0	1.0
VIGUIERA	1.0	0.8	1.0
MIKANIA	1.0	0.8	1.0
INULA	1.0	1.0	0.0
CALEA	1.0	1.0	0.0
HAPLOPAPPUS	0.0	0.3	1.0
AGERATINA	1.0	0.7	1.0
VERBESINA	1.0	1.0	1.0
SANTOLINA	1.0	0.9	0.0
HELIANTHUS	1.0	1.0	1.0
ASTER	1.0	0.2	1.0
WEDELIA	1.0	1.0	1.0
CASSINIA	0.0	0.1	1.0
FLOURENSIA	1.0	0.9	1.0
CHRYSANthemUM	1.0	0.9	1.0
SOLIDAGO	0.0	-0.1	1.0
AMBROSIA	1.0	1.0	1.0
PARTHENIUM	1.0	0.8	1.0
PULICARIA	1.0	1.0	1.0
BRICKELLIA	1.0	1.0	1.0
ICHTHYOTHERE	1.0	1.0	1.0
MONTANOA	1.0	0.9	0.0
XANTHIUM	1.0	1.0	1.0
RELHANIA	0.0	0.3	1.0
CHROMOLAENA	1.0	1.0	0.0
OPHRYOSPORUS	0.0	0.2	1.0
GOCHNATIA	1.0	-0.1	1.0
STEIRACTINIA	1.0	1.0	0.0
TAGETES	0.0	0.0	1.0
BIDENS	0.0	0.8	1.0
GRINDELIA	0.0	0.2	1.0
HELENIUM	1.0	0.8	1.0
HEMIZONIA	0.0	0.1	0.0
SONCHUS	1.0	1.0	1.0
CIRSIUM	1.0	0.5	1.0
JASONIA	0.0	0.2	1.0
CHRYSOTHAMNUS	0.0	0.2	1.0
SCHISTOSTEPHIUM	1.0	1.0	1.0
TITHONIA	1.0	1.0	1.0
TETRAGONOTHECA	1.0	0.8	0.0
OSTEOSPERMUM	0.0	0.1	1.0
SMALLANTHUS	1.0	0.7	1.0
BAHIA	1.0	0.9	1.0
SCHKUHRIA	1.0	0.9	1.0

Tabela 5.2.1. Continuação

DORONICUM	0.0	-0.1	1.0
LASIOLAENA	1.0	0.8	0.0
ACANTHOSPERMUM	1.0	1.0	0.0
LASTHENIA	0.0	0.3	1.0
OTHONNA	0.0	1.0	1.0
CALENDULA	0.0	0.0	1.0
TARAXACUM	1.0	1.0	1.0
HYMENOPAPPUS	1.0	0.9	0.0
DUGALDIA	1.0	0.8	1.0
NEOMIRANDEA	1.0	0.6	1.0
EURYOPS	1.0	1.0	1.0
GUTIERREZIA	0.0	0.1	1.0
ECHINOPS	0.0	0.1	1.0
ACHYROCLINE	0.0	0.0	0.0
GAILLARDIA	1.0	0.9	1.0
CARTHAMUS	0.0	0.5	1.0
PENTZIA	1.0	1.0	1.0
LYCHNOPHORA	1.0	0.9	1.0
SERRATULA	1.0	0.8	1.0
HETEROTHECA	0.0	0.0	0.0
LACTUCA	1.0	1.0	0.0
ENCELIA	1.0	0.6	0.0
ONOPORDUM	1.0	0.5	1.0
PEREZIA	0.0	0.3	0.0
TRICHOGONIA	1.0	1.0	1.0
DICOMA	1.0	0.3	1.0
MICROGLOSSA	0.0	0.0	1.0
DIMEROSTEMMA	1.0	0.9	1.0
HELIOPSIS	1.0	0.9	1.0
PAPPOBOLUS	1.0	1.0	0.0
KOANOPHYLLON	1.0	1.0	1.0
POROPHYLLUM	0.0	0.3	0.0
CYNARA	1.0	1.0	1.0
SILPHIUM	1.0	0.7	1.0
GREENMANIELLA	1.0	0.9	1.0
OYEDAEA	1.0	1.0	0.0
WYETHIA	0.0	0.1	1.0
MILLERIA	1.0	0.7	0.0
COTULA	1.0	0.9	1.0
PERYMENIUM	1.0	0.7	1.0
ANAPHALIS	1.0	0.8	1.0
XANTHOCEPHALUM	0.0	0.0	0.0
LIABUM	1.0	0.9	1.0
CENTRATHERUM	1.0	0.6	1.0
BRASILIA	1.0	1.0	1.0
FLEISCHMANNIA	1.0	1.0	1.0
KAUNIA	0.0	0.1	0.0
ATHANASIA	1.0	1.0	1.0
DISYNAPHIA	1.0	0.2	0.0
CRITONIA	1.0	0.8	1.0
OXYLOBUS	1.0	0.6	1.0
ONOSERIS	0.0	0.0	1.0
PIQUERIA	0.0	0.2	1.0
STYLOTRICHIMUM	1.0	1.0	0.0
CENTIPEDA	1.0	1.0	1.0
FLORESTINA	1.0	0.4	1.0
AJANIA	1.0	0.9	1.0

Tabela 5.2.1. Continuação

INULANTHERA	1.0	0.9	1.0
COREOPSIS	0.0	0.0	1.0
DAHLIA	0.0	0.0	0.0
CARDUUS	0.0	0.0	1.0
DENDRATHEMA	0.0	1.0	1.0
ECHINACEA	0.0	0.1	1.0
COESPELETIA	0.0	0.1	0.0
BELLIS	0.0	0.0	1.0
SPILANTHES	0.0	0.6	1.0
PSIADIA	0.0	0.0	0.0
EPALTES	0.0	-0.1	0.0
CARLINA	0.0	1.0	0.0
COUSINIA	1.0	0.7	0.0
SIGESBECKIA	1.0	0.7	1.0
PERITYLE	1.0	0.3	1.0
PTERONIA	0.0	0.1	0.0
GNAPHALIUM	0.0	0.1	0.0
TRIXIS	0.0	0.3	1.0
ELEPHANTOPUS	1.0	0.7	0.0
IPHIONA	0.0	0.1	0.0
HELOGYNE	1.0	0.3	1.0
NEUROLAENA	1.0	0.7	1.0
ADENOSTEMMA	0.0	0.2	0.0
PTEROCAULON	0.0	0.0	1.0
TRIDAX	0.0	0.7	0.0
NASSAUVIA	0.0	0.1	0.0
PIPTOTHRIX	1.0	0.3	0.0
ZALUZANIA	1.0	0.7	1.0
TELEKIA	1.0	0.7	1.0
GRAZIELIA	1.0	1.0	1.0
GRANGEA	0.0	-0.1	1.0
ERIOPHYLLUM	1.0	0.8	0.0
GYPOTHAMNIUM	0.0	0.1	0.0
AYAPANA	0.0	0.6	1.0
ETHULIA	0.0	0.6	0.0
FERREYRANTHUS	1.0	1.0	1.0
CARPESIUM	1.0	0.7	1.0
ATHRIXIA	0.0	0.0	0.0
BEJARANOA	1.0	1.0	1.0
CRATYSTYLIS	1.0	0.4	1.0
LESSINGIA	0.0	0.0	0.0
ATRACTYLIS	0.0	0.0	0.0
ISOCARPHA	1.0	0.7	0.0
TRICHOGONIOPSIS	1.0	0.2	1.0
OTANTHUS	1.0	0.7	1.0
ALLAGOPAPPUS	0.0	0.1	1.0
LAYIA	0.0	0.1	0.0
STOEBE	0.0	0.0	1.0
VANILLOSMOPSIS	1.0	0.7	0.0
GUIZOTIA	0.0	0.1	0.0
PYRETHRUM	1.0	0.7	1.0
CALOCEPHALUS	1.0	0.5	1.0
HETEROMMA	0.0	-0.1	1.0
ERLANGEA	1.0	0.9	0.0
ACTINOSERIS	1.0	0.9	1.0
LINDHEIMERA	0.0	0.7	0.0
WERNERIA	0.0	0.0	0.0

Tabela 5.2.1. Continuação

WUNDERLICHIA	1.0	0.2	1.0
KINGIANTHUS	1.0	0.1	1.0
CHRYSOLAENA	0.0	0.0	0.0
PARASTREPHIA	0.0	0.0	1.0
BAHIANTHUS	0.0	0.8	0.0
MADIA	0.0	0.1	0.0
CRONQUISTIANTHUS	1.0	0.3	0.0
INEZIA	1.0	0.7	1.0
PICRADENIOPSIS	1.0	0.6	1.0
BRACHYCOME	0.0	0.0	0.0
LOURTEIGIA	1.0	1.0	0.0
ARGYRANTHEMUM	0.0	0.7	1.0
URSINIA	1.0	0.7	0.0
ECLIPTA	0.0	0.0	0.0
JURINEA	1.0	0.2	1.0
ONOPORDON	1.0	0.4	0.0
ANACYCLUS	0.0	0.0	0.0
PALAFIXIA	0.0	0.4	1.0
ARCTIUM	1.0	0.5	1.0
BROCCHIA	1.0	0.9	1.0
BOTHRIOCLINE	1.0	0.9	0.0
EGLITES	0.0	0.1	1.0
BRACHYLAENA	1.0	0.6	1.0
VERNONANTHURA	1.0	0.9	0.0
GEIGEIRA	0.0	0.1	0.0
LAUNAEA	0.0	0.3	0.0
ZEXMENIA	1.0	0.1	0.0
HIERACIUM	0.0	0.3	1.0
ONDETIA	1.0	0.3	0.0
FELICIA	0.0	-0.1	1.0
XERANTHEMUM	1.0	0.2	1.0
IOSTEPHANE	0.0	0.1	1.0
AUSTROEUPATORIUM	0.0	0.2	0.0
HYPOCHAERIS	0.0	0.0	1.0
UROSPERMUM	1.0	0.3	0.0
GYMNOSPERMA	0.0	-0.1	1.0
IXERIS	1.0	0.4	0.0
PALLENIS	0.0	0.0	1.0
LECOCARPUS	1.0	0.1	0.0
ROLANDRA	1.0	0.9	1.0
BAILEYA	1.0	0.3	1.0
ANISOPAPPUS	1.0	0.9	1.0
PSEUDOGNAPHALIUM	0.0	0.1	1.0
GERAEA	0.0	0.0	1.0
HERTIA	0.0	0.1	0.0
WAITZIA	0.0	0.0	0.0
BEDFORDIA	1.0	0.4	0.0
PLATYCARPHA	1.0	0.5	1.0
PROUSTIA	1.0	0.4	1.0
GERBERA	0.0	0.2	1.0
LEPIDOPHYLLUM	0.0	0.1	1.0
ASTERISCUS	0.0	0.1	1.0
OZOTHAMNUS	0.0	0.0	1.0
RUTIDOSIS	0.0	0.0	0.0
VANCLEVEA	0.0	-0.1	1.0
PETERAVENIA	0.0	1.0	0.0
GUTENBERGIA	1.0	0.9	0.0

Tabela 5.2.1. Continuação

CAMPOVASSOURIA	1.0	0.3	0.0
SILYBUM	0.0	0.5	0.0
LOXOTHYSANUS	1.0	0.5	1.0
REICHARDIA	1.0	0.3	1.0
BOLTONIA	0.0	0.0	0.0
HELIANTHOPSIS	0.0	0.1	0.0
PEYROUSEA	1.0	0.7	1.0
ANDRYALA	1.0	0.3	1.0
ACRISIONE	0.0	0.1	1.0
APHYLLOCLADOS	0.0	0.2	1.0
DIMORPHOTHECA	0.0	0.0	0.0
GOYAZIANTHUS	0.0	0.2	0.0
HELIPTERUM	0.0	0.5	0.0
ISOCARPHA	1.0	0.2	0.0
OSMITOPSIS	1.0	0.7	1.0
CHRYSACTINIA	0.0	0.3	1.0
CAESULIA	0.0	0.1	1.0
CORYMBIUM	0.0	0.1	1.0
EIRMOCEPHALA	1.0	0.9	1.0
PIPTOCARPHA	1.0	0.9	1.0
CACOSMIA	1.0	0.9	0.0
ABROTANELLA	0.0	0.0	1.0
HELIANTHELLA	0.0	0.6	1.0
IXIOLAENA	0.0	0.0	1.0
STOMATANTHES	0.0	0.0	0.0
CALOSTEPHANE	1.0	0.8	1.0
LEONTODON	1.0	0.3	1.0
MOQUINIA	1.0	0.9	0.0
HETEROPAPPUS	0.0	-0.1	1.0
VIEREA	0.0	0.7	1.0
PROTEOPSIS	1.0	0.9	1.0
CARTEROTHAMNUS	0.0	0.2	1.0
HOFMEISTERIA	0.0	0.2	1.0
ALCANTARA	1.0	0.9	0.0
CICHORIUM	1.0	0.8	0.0
AUSTROBRICKELLIA	1.0	0.1	0.0
HETEROCOMA	1.0	0.2	1.0
OLDENBURGIA	1.0	0.4	0.0
CALYCADENIA	0.0	0.1	1.0
HYMENOTHRIX	0.0	0.4	0.0
VITTADINIA	0.0	-0.1	0.0
GAZANIA	1.0	0.5	1.0
HARTWRIGHTIA	1.0	0.3	0.0
KLEINIA	0.0	0.1	1.0
KYRSTENIOPSIS	0.0	0.3	0.0
PLANALTOA	0.0	0.2	1.0
TAMAULIPA	0.0	0.3	1.0
CYATHOCLINE	1.0	1.0	0.0
ELYTROPAPPUS	0.0	0.1	1.0
BISHOVIA	0.0	0.3	0.0
PLATYCHAETE	0.0	0.1	0.0
ASANTHUS	1.0	0.1	0.0
CLADANTHUS	1.0	0.7	1.0
CONOCLINIOPSIS	1.0	0.1	0.0
CARELIA	0.0	0.1	0.0
PLEUROCORONIS	0.0	0.3	1.0
AGRIANTHUS	1.0	0.5	1.0

Tabela 5.2.1. Continuação

CICERBITA	1.0	0.8	1.0
MUNNOZIA	1.0	0.9	1.0
GIBBARIA	0.0	0.0	1.0
LEIBNITZIA	0.0	0.1	0.0
METALASIA	0.0	0.0	0.0
PERTYA	1.0	0.6	1.0
TUSSILAGO	1.0	0.5	0.0
ACOURTIA	0.0	0.1	1.0
ALOMIA	0.0	0.3	1.0
CAMPULOCLINIUM	0.0	0.0	1.0
RAOULIA	0.0	0.1	1.0
PECTIS	0.0	0.0	0.0
CHUQUIRAGA	0.0	0.0	1.0
ARISTEGUIETIA	0.0	0.3	1.0
DASYPHYLLUM	0.0	0.0	0.0
LIBANOTHAMNUS	0.0	0.1	1.0
SERIPHIDIUM	0.0	0.7	1.0
CHEIROLOPHUS	1.0	0.9	1.0
GARULEUM	0.0	0.0	1.0
HOLOCARPHA	0.0	0.0	1.0
HYMENOCLEA	1.0	0.9	1.0
PARPPOBOLUS	1.0	0.9	1.0
ACMELLA	0.0	0.2	1.0
GUARDIOLA	0.0	0.1	1.0
HELIOMCRIS	1.0	0.9	1.0
BARNADESIA	0.0	0.0	1.0
GROSVENORIA	0.0	0.0	0.0
ZOEGEA	0.0	0.2	0.0
BAERIA	0.0	0.3	1.0
SCHIZOGYNE	0.0	0.3	0.0
WILKESIA	0.0	0.0	1.0
OXYPHYLLUM	0.0	0.4	1.0
AUSTROLIABUM	1.0	0.9	0.0
BRACTEANTHA	0.0	0.1	1.0
SYMPHYOPAPPUS	0.0	0.3	1.0
STILPNOPAPPUS	1.0	0.9	0.0
CALLILEPIS	0.0	0.0	1.0
CHILIADENUS	0.0	0.3	0.0
CHILIOTRICHUM	0.0	-0.1	1.0
GLOSSOPAPPUS	0.0	0.7	1.0
LASIANTHAEA	0.0	0.1	0.0
NARDOPHYLLUM	0.0	-0.1	1.0
CARDUNCCELLUS	0.0	0.5	1.0
PIRETHRUM	1.0	0.9	1.0
ADENOTHAMNUS	0.0	0.0	1.0
BIOTIA	0.0	0.2	0.0
BLEPHARISPERMUM	0.0	0.1	0.0
BRINTONIA	0.0	0.4	1.0
TAECKHOLMIA	0.0	0.3	0.0
DIPLOPAPPUS	0.0	0.2	1.0
HEBECLINIUM	0.0	0.3	1.0
HETEROTHALAMUS	0.0	0.0	0.0
KLENIA	0.0	0.1	1.0
POLYACHYRUS	0.0	0.4	0.0
KIPPISTIA	0.0	0.0	0.0
TRAGOPOGON	0.0	0.3	1.0
BRACHYCLADOS	0.0	0.6	1.0

Tabela 5.2.1. Continuação

TAMANIA	0.0	0.1	0.0
PODANTHUS	1.0	0.9	1.0
VENIDIUM	1.0	0.5	1.0
ACAMPTOPAPPUS	0.0	-0.1	0.0
LEUCOCYCLUS	0.0	0.2	1.0
ROBINSONIA	0.0	0.4	1.0
SALMEA	0.0	0.0	1.0
SPHAEROCLINIUM	0.0	0.2	1.0
SYNEDRELLA	0.0	0.0	1.0
LEUCANTHEMOPS	1.0	0.9	1.0
STOKESIA	1.0	0.9	1.0
ANVILLEA	0.0	0.2	1.0
MINURIA	0.0	0.0	1.0
POLYANTHINA	0.0	0.0	1.0
PRINTZIA	0.0	0.0	1.0
SCHISTOCARPHA	0.0	0.0	0.0
TETRACHYRON	0.0	0.1	1.0
TRIPTILION	0.0	0.6	1.0
ANDRIALA	1.0	0.9	1.0
BRACHYLOENA	1.0	0.9	1.0
FULCALDEA	0.0	0.0	1.0
PHOEBANTHUS	0.0	0.1	1.0
AMELLUS	0.0	-0.1	1.0
ANISOTHRIX	0.0	0.0	1.0
BLEPHARIZONIA	0.0	0.1	0.0
LIABIUM	0.0	0.1	1.0
MORITHAMNUS	0.0	0.3	0.0
ROLDANA	0.0	0.1	1.0
CARPHOCHAETE	1.0	0.5	1.0
PICRIDIMUM	1.0	0.8	1.0
CHRYOSPLENIUM	0.0	0.1	1.0
BADILLOA	0.0	0.1	1.0
CARAMBOA	0.0	0.1	1.0
CARRAMBOA	0.0	0.1	1.0
DENEKIA	0.0	0.0	1.0
DITRICHIA	0.0	0.0	1.0
GALINSOGA	0.0	0.0	1.0
GYNOXYS	0.0	0.1	1.0
HYSTERIONICA	0.0	-0.1	1.0
RHYNCHOSPERMUM	0.0	-0.1	1.0
JURINELLA	1.0	0.9	1.0
LAUNAELA	1.0	0.8	1.0
AMAURIA	0.0	0.0	1.0
ARNALDOA	0.0	0.0	0.0
SCHLECHTENDAHLIA	0.0	0.0	1.0
ACHYRACHAENA	0.0	0.3	0.0
CALLISTEPHUS	0.0	-0.1	1.0
CHAPTALIA	0.0	0.6	1.0
CONDYLIDIUM	0.0	0.0	0.0
DELILIA	0.0	0.0	1.0
IREXIS	0.0	0.4	1.0
LAENNECIA	0.0	0.1	1.0
LEONTOPODIUM	0.0	0.0	0.0
TOWNSENDIA	0.0	-0.1	0.0
VARILLA	0.0	0.5	0.0
AINSLIAEA	1.0	0.6	1.0
CYCLOLEPIS	1.0	0.6	1.0

Tabela 5.2.1. Continuação

ERPANTOPUS	1.0	0.9	1.0
OLIGANTHES	1.0	0.9	0.0
PSEUDOELEPHANT	1.0	0.9	1.0
CHONDRILLA	0.0	0.3	1.0
AMMOBIUM	0.0	0.0	0.0
BEBBIA	0.0	0.0	1.0
CELMISIA	0.0	0.3	1.0
COUSINA	0.0	0.1	1.0
CRITHMUM	0.0	0.2	1.0
DIOTIS	0.0	0.2	1.0
GLOSSOCARDIA	0.0	0.0	0.0
HINTERHUBERA	0.0	-0.1	1.0
KOELPINIA	0.0	0.8	0.0
PHAGNALON	0.0	0.0	1.0
PSEUDOBICKELLIA	0.0	0.5	1.0
SYNEDRELLOPSIS	0.0	0.2	1.0
ABUTILON	1.0	0.9	1.0
GROSSHEIMIA	1.0	0.9	1.0
PECHVEL-LOESCH	1.0	0.8	1.0
PTILOSTEMMON	1.0	0.9	0.0
ARGYROXIPHUM	0.0	0.0	1.0
EMBERGERIA	0.0	0.3	1.0
GYMNOPENTZIA	0.0	0.0	0.0
SCALESIA	0.0	0.1	0.0
ALBERTINIA	0.0	0.9	1.0
ARCTOTHECA	0.0	0.5	1.0
ATHROISMA	0.0	0.0	1.0
CARPHEPHORUS	0.0	0.1	1.0
CINERRIA	0.0	0.1	1.0
CRUPINA	0.0	0.2	0.0
ELVIRA	0.0	0.0	1.0
FERULA	0.0	0.4	1.0
GALACTITES	0.0	0.2	0.0
GORTERIA	0.0	0.0	0.0
HAPLOCARPHA	0.0	0.5	1.0
HIRPICUM	0.0	0.5	1.0
ISOSTIGMA	0.0	0.0	1.0
LASIANTHEA	0.0	0.1	1.0
MELANTHERA	0.0	0.0	1.0
MICROPUS	0.0	0.0	0.0
OSTEOSPERMUM	0.0	0.2	1.0
PECHUEL-LOESCHIA	0.0	0.3	1.0
PERYMENIOPSIS	0.0	0.1	1.0
PLAGIOCHEILUS	0.0	-0.1	0.0
PLATYPODANTHERA	0.0	0.0	1.0
SPIRACANTHA	0.0	0.9	1.0
TUGARINOVIA	0.0	0.1	0.0
WULFFIA	0.0	0.1	0.0
ANGIANTHUS	1.0	0.4	1.0
BISHOPIELLA	1.0	0.5	1.0
CHARTOLEPIS	1.0	0.9	1.0
DUGESIA	1.0	0.9	1.0
IXCRIS	1.0	0.9	1.0
LEUCERIA	1.0	0.9	1.0
MATTFELDANTHUS	1.0	0.9	1.0
PARTHENICE	1.0	0.9	1.0
RUDBEKIA	1.0	0.9	0.0

Tabela 5.2.1. Continuação

TAECKOLMIA	1.0	0.8	1.0
VENEGASIA	1.0	0.5	1.0
VICOA	1.0	0.9	1.0
AARONSOHNIA	0.0	0.0	1.0
CHRYSOPOLENIUM	0.0	0.1	1.0
GALEANA	0.0	0.0	0.0
HESPEROMANNIA	0.0	0.2	1.0
HISPIDELLA	0.0	0.3	1.0
PLEOCARPHUS	0.0	0.2	0.0
PSACALIUM	0.0	0.1	1.0
PYRRHOPAPPUS	0.0	0.3	1.0
STAEHELINA	0.0	0.1	1.0
STIZOLOPHUS	0.0	0.1	0.0
ACTINOBLE	0.0	0.1	1.0
AMBLYOLEPIS	0.0	0.5	1.0
ANCHYRACHAMA	0.0	0.2	0.0
ANCHYRACHANA	0.0	0.2	0.0
APHYLLOCLADUS	0.0	0.9	1.0
BARTLETTINA	0.0	0.3	1.0
CHARDINIA	0.0	0.2	1.0
CINERARIA	0.0	0.1	0.0
CINERARAI	0.0	0.1	1.0
COUSINIAECA	0.0	0.4	1.0
CRITONIELLA	0.0	0.0	1.0
DESMANTHODIUM	0.0	0.1	1.0
DYMONDIA	0.0	0.0	1.0
ELITROPAPPUS	0.0	0.1	0.0
EVAX	0.0	0.0	1.0
GARBERIA	0.0	0.3	1.0
HETEROLEPIS	0.0	0.0	1.0
HOPLOPHYLLUM	0.0	0.2	1.0
HUMEA	0.0	0.0	1.0
HYRSTENIOPSIS	0.0	0.1	1.0
JAEGERIA	0.0	0.0	0.0
LACHNOPHYLLUM	0.0	-0.1	0.0
MATTFELDIA	0.0	0.5	0.0
MONOLOPIA	0.0	0.3	0.0
PEYROUSIA	0.0	0.9	1.0
PHAENOCOMA	0.0	0.4	1.0
PLAGIOCEILUS	0.0	0.1	0.0
PODOLOPIS	0.0	0.1	0.0
POLYACHURUS	0.0	0.4	1.0
PSEUDOKYRSTENIOPSIS	0.0	0.5	0.0
RIENCOURTIA	0.0	0.1	1.0
SCIADOCEPHALA	0.0	0.3	0.0
SOLIVA	0.0	0.2	1.0
STRUCHIUM	0.0	0.7	1.0
TRICHOLEPIS	0.0	0.2	0.0
TRIGONOSPERMUM	0.0	0.0	1.0
TRILISA	0.0	0.5	1.0
Total de acertos			338.0
% de acertos			67.6

Tabela 5.2.2. Acertos e erros da rede supervisionada por gêneros para os sesquiterpenos lactonizados, com heurística de 0 e 1, na fase de teste.

Gêneros	Valor Real	Valor Previsto	Acertos
BACCHARIS	1.0	0.3	1.0
VERNONIA	1.0	0.4	1.0
STEVIA	1.0	1.0	1.0
MELAMPODIUM	1.0	1.0	0.0
SAUSSUREA	1.0	0.8	1.0
BALSAMORHIZA	1.0	0.1	1.0
ANTHEMIS	1.0	1.0	1.0
MATRICARIA	1.0	1.0	1.0
ERIOCEPHALUS	1.0	1.0	0.0
BLAINVILLEA	1.0	0.9	1.0
BLUMEA	1.0	0.2	1.0
PICRIS	1.0	1.0	1.0
POLYMNIA	1.0	0.9	1.0
RATIBIDA	1.0	0.8	0.0
ZINNIA	1.0	0.9	1.0
ACRITOPAPPUS	0.0	0.8	1.0
CHAMAEMELUM	1.0	1.0	0.0
NIDORELLA	0.0	0.0	0.0
PEGOLETTIA	1.0	0.1	0.0
ARCTOTIS	1.0	0.2	0.0
COSMOS	1.0	0.6	0.0
TESSARIA	1.0	0.0	1.0
BERKHEYA	1.0	0.0	1.0
PODACHAENIUM	1.0	1.0	1.0
LEUCANTHEMUM	0.0	0.0	0.0
ISOCOMA	0.0	-0.1	0.0
IVA	1.0	0.0	0.0
LEUZEA	1.0	0.2	0.0
HYPOCHOERIS	1.0	0.4	0.0
TRICHOCLINE	0.0	0.9	0.0
DECACHAETA	1.0	0.1	0.0
CACALIA	1.0	0.0	0.0
FITCHIA	1.0	0.6	0.0
VILLANOVA	0.0	0.9	1.0
MACOWANIA	1.0	0.1	0.0
UROLEPIS	0.0	0.1	0.0
LAGASCEA	0.0	0.9	0.0
ASPILIA	0.0	0.7	1.0
STEPHANOMERIA	0.0	0.3	1.0
OLEARIA	0.0	0.1	0.0
AMPHIACHYRIS	0.0	0.1	0.0
HIPPIA	0.0	0.7	0.0
TARCHONANTHUS	0.0	0.9	1.0
MYRIOCEPHALUS	0.0	0.5	0.0
FLAVERIA	0.0	0.1	1.0
GNEPHOSIS	1.0	0.1	0.0
ERICAMERIA	0.0	-0.1	1.0
LEUCHERIA	1.0	0.2	1.0
CHRESTA	1.0	0.9	1.0
LEYSERA	0.0	0.1	0.0
PTILOSTEMON	1.0	0.9	0.0
MARSHALLIA	0.0	0.3	0.0
PEGOLETIA	0.0	0.9	0.0
MOSCHARIA	0.0	0.4	0.0
SCORZONERA	0.0	0.6	1.0

Tabela 5.2.2. Continuação

CROPTILON	1.0	-0.1	1.0
HETEROCONDYLUS	0.0	0.2	1.0
ESPELETIOPSIS	0.0	0.1	1.0
HAZARDIA	0.0	0.1	0.0
THELESERMA	0.0	0.0	1.0
LAGGERA	0.0	0.3	1.0
APALOCHLAMYS	0.0	0.1	1.0
CALOTIS	0.0	0.2	1.0
BELLIUM	0.0	-0.1	1.0
LEUCANTHEMELLA	0.0	0.2	0.0
ENYDRA	1.0	0.9	1.0
ANTIPHIONA	0.0	0.3	1.0
HYALIS	0.0	0.1	1.0
HETEROSPERMA	0.0	0.0	1.0
PLAGIUS	0.0	0.2	1.0
IFOGA	0.0	0.1	1.0
CRASPEDIA	0.0	0.0	0.0
ODIXIA	0.0	0.1	1.0
STEYMARKINA	0.0	0.3	0.0
ERODIOPHYLLUM	0.0	-0.1	0.0
BALDUINA	1.0	0.5	1.0
BABCOCKIA	0.0	0.3	0.0
LONAS	0.0	0.2	1.0
HOMOGYNE	1.0	0.5	1.0
ASTEROMOEIA	0.0	-0.1	1.0
CHRYSOPTIS	0.0	-0.1	1.0
TRIPTERIS	0.0	0.0	0.0
TYRIMNUS	0.0	0.2	1.0
ANAXETON	0.0	0.4	1.0
FLEISCHMANNIOPSIS	0.0	0.0	0.0
PEGOLITTIA	0.0	0.0	0.0
PENTACALIA	0.0	0.5	1.0
GUEVARIA	1.0	0.5	1.0
MONACTIS	1.0	0.9	1.0
ONOPORDUN	1.0	0.9	1.0
CERUANA	0.0	0.1	0.0
PSATHYROTES	0.0	0.0	1.0
CHRYSANTHELLUM	0.0	0.9	0.0
DYSSODIA	0.0	0.0	1.0
LESSINGIANTHUS	0.0	0.9	0.0
MYRIACTIS	0.0	-0.1	1.0
PENTANEMA	0.0	0.3	1.0
PODOCOMA	0.0	-0.1	1.0
VERSINA	0.0	0.1	1.0
XANTHISMA	0.0	-0.1	1.0
Total de acertos			58.0
% de acertos			58.0

SÚMULA CURRICULAR

DADOS PESSOAIS

Nome: Mauro Vicentini Correia

Local e data de nascimento: São Paulo – SP, 28/09/1981.

EDUCAÇÃO

Colégio Piaget, São Paulo, 1999.

Universidade de São Paulo, São Paulo, 2008.

Farmacêutico - Bioquímico

OCUPAÇÃO

Bolsista de Mestrado Agência, FAPESP, 2010

PUBLICAÇÕES

Artigos Completos

1. Correia, M.V. ; SCOTTI, M. T.; FERREIRA, J. P. M.; ALVARENGA, S. A. V.; RODRIGUES, G. V.; Emerenciano, V. P.. Self-Organizing Maps as a New Tool for Classification of Plants at Lower Hierarchical Levels. Natural Product Communications (Online), v. 3, p. 1723-1730, 2008.

Resumos em Congressos

1. Correia MV, Fokoue HH, Scotti MT, Ferreira MJP, Emerenciano VP. Molecular Descriptors to Predict Inhibitor Activity of Human Neutrophil Elastase (HNE) From Plants's Phenolics Compounds. In: 2nd Brazilian Conference on Natural Products (2nd BCNP), 2009, São Pedro.
2. Correia MV, Fokoue HH, Scotti MT, Emerenciano VP. Self-Organizing Maps to Predict Inhibitor Human Neutrophil Elastase Activity of Phenolics Compounds. In: 2nd Brazilian Conference on Natural Products (2nd BCNP), 2009, São Pedro.
3. Fokoue HH, Correia MV, Ferreira MJP, Alvarenga SAV, Rodrigues GV, Emerenciano VP. A New Parameter for Chemosystematic Analysis of Asteraceae Tribes. In: 2nd Brazilian Conference on Natural Products (2nd BCNP), 2009, São Pedro.
4. Fokoue HH, Correia MV, Ferreira MJP, Alvarenga SAV, Rodrigues GV, Emerenciano VP. Chemosystematics of Heliantheae (Asteraceae) Subtribes Using Cluster Analysis. In: 2nd Brazilian Conference on Natural Products (2nd BCNP), 2009, São Pedro.
5. Yamaguchi LF, Salazar KM, Freitas GC, Yoshida NC, Santos EL, Marques JV, Oliveira A, Navarro LB, Silva RA, Ferreira EA, Gaia AM, Rodrigues CA, Benedetti AM, Valero YG, Scotti MT, Correia MV, Fokoue HH, Emerenciano VP, Kato MJ. Metabolomic analysis of Piperaceae species using ESI/¹H NMR and PCA. In: 2nd Brazilian Conference on Natural Products (2nd BCNP), 2009, São Pedro.

6. Rossini M, Correia MV, Scotti MT, Ferreira JPM, Emerenciano VP. Volsurf Descriptors to Analyze the Anti-Viral Activity of Sesquiterpene Lactone. In: 4º Simpósio Brasileiro em Química Medicinal, 2008, Porto de Galinhas.
7. Ferreira JPM, Romoff P, Fokoue HH, Scotti MT, Speck-Planche A, Scotti L, Correia MV, Lago JHG, Emerenciano VP. Molecular Descriptors to Predict the Cytotoxic Activity of Caffeic Ester Analogues. In: 4º Simpósio Brasileiro em Química Medicinal, 2008, Porto de Galinhas.
8. Scotti MT, Correia MV, Ferreira JPM, Emerenciano VP. Sesquiterpenos Lactonizado, Descritores Moleculares e Mapas Auto-Organizáveis Utilizados no Estudo Quimiotaxonômico da Tribo Heliantheae. In: 31ª Reunião Anual da Sociedade Brasileira de Química, 2008, Águas de Lindóia.
9. Emerenciano VP, Scotti MT, Ferreira JPM, Correia MV, Alvarenga SAV, Rodrigues GV. Self Organizing Maps as Tool for Taxonomic Classification at Lower Hierarchical Levels. In: 1st Brazilian Conference on Natural Products (1st BCNP), 2007, São Pedro.
10. Scotti MT, Ishiki HM, Correia MV, Ferreira JPM, Emerenciano VP Use Self-Organizing Maps and ¹³C NMR Spectral Data to Predict Aldose Reductase Activity of Flavonoids. In: 1st Brazilian Conference on Natural Products (1st BCNP), 2007, São Pedro.
11. Emerenciano VP, Scotti MT, Ferreira JPM, Correia MV, Alvarenga SAV, Rodrigues GV. Chemosystematics of Asteraceae Tribes Using Principal Component Analysis. In: 1st Brazilian Conference on Natural Products (1st BCNP), 2007, São Pedro.
12. Emerenciano VP, Ferreira JPM, Scotti MT, Correia MV, Alvarenga SAV, Rodrigues GV. Use of Backpropagation Artificial Neural Networks to Predict the Occurrences of Chemical Classes in Asteraceae. In: 1st Brazilian Conference on Natural Products (1st BCNP), 2007, São Pedro.
13. Correia MV, Scotti MT, Emerenciano VP. Redes Neurais não Supervisionadas Utilizadas no Estudo Quimiotaxonômico da Tribo Heliantheae (Asteraceae). In: 30ª Reunião Anual da Sociedade Brasileira de Química, 2007, Águas de Lindóia.