

UNIVERSIDADE DE SÃO PAULO

FACULDADE DE CIÊNCIAS FARMACÊUTICAS DE RIBEIRÃO PRETO

Metabolômica para análise de cultivares, quimiotaxonomia e triagem de inibidores de lactato desidrogenase-5 (LDH-5) em espécies de Asteraceae

Felipe Antunes dos Santos

Ribeirão Preto

2019

RESUMO

SANTOS, F. A. **Metabolômica para análise de cultivares, quimiotaonomia e triagem de inibidores de lactato desidrogenase-5 (LDH-5) em espécies de Asteraceae** 2019. 178 f. Tese (Doutorado). Faculdade de Ciências Farmacêuticas de Ribeirão Preto – Universidade de São Paulo, Ribeirão Preto, 2019.

Metabolômica pode ser definida como o estudo de todos os metabólitos de um dado sistema biológico. Tal estudo possibilita abordagens holísticas e integradas de diversos ramos da ciência. Deste modo, problemas em delimitações taxonômicas, análise da influência ambiental sobre plantas, e triagem de metabólitos bioativos constituem exemplos de áreas onde a metabolômica pode atuar. A família Asteraceae, com aproximadamente 25.000 espécies, tem sido alvo de estudos taxonômicos por décadas. Assim sendo, a abordagem metabolômica aplicada à quimiotaonomia pode auxiliar em tais classificações, ou revelar novas descobertas. Por outro lado, discute-se que os metabólitos secundários sofrem muita influência de diversos fatores ambientais. Deste modo, este trabalho investigou também a influência da sazonalidade, luminosidade, adubação e ritmo circadiano no metaboloma de cultivares de *Aldama robusta* e *A. trichophylla* (Asteraceae). O cultivo de *A. robusta* foi também estudado quanto à sua germinação e desenvolvimento morfológico. A ampla variabilidade química em Asteraceae também a coloca como uma família com potencial para a triagem de metabólitos bioativos. Um alvo bioquímico relevante é a enzima lactato desidrogenase-5 (LDH-5), a qual esta relacionada com o aumento da sobrevivência de células cancerígenas. Deste modo, o estudo metabolômico em espécies de Asteraceae pode possibilitar a descoberta de substâncias ativas contra determinado alvo. O metaboloma de todas as espécies deste trabalho foi adquirido por UHPLC-ESI-UV/DAD-MS, com analisador Orbitrap™. Para a quimiotaonomia e estudos de inibição da LDH-5, foram estudadas 250 espécies de Asteraceae, provenientes de diversos gêneros de 16 tribos. Para os estudos da influência de determinados fatores ambientais no metaboloma de *A. robusta* e *A. trichophylla*, foram realizadas coletas em diferentes ritmos circadianos, estações do ano e diferentes cultivares com base na adubação do solo e luminosidade. Os dados químicos foram analisados por diversas técnicas quimiométricas e de *data mining* (mineração de dados), tais como PCA, HCABp, OPLS-DA e *decision tree*. Deste modo, a análise por OPLS-DA com espécies de Asteraceae foi capaz de separar algumas tribos e apontar metabólitos discriminantes em cada uma delas, não obstante a PCA e HCABp não tenham revelado agrupamento. Do mesmo modo, a OPLS-DA e a *decision tree* foram capazes de apontar alguns prováveis metabólitos correlacionados com extratos considerados “ativos” contra a LDH-5, bem como realizar a predição de extratos provenientes de outras espécies de Asteraceae que não haviam ainda sido ensaiados. Já os estudos com a germinação e o desenvolvimento dos cultivares pode revelar condições mais adequadas para o cultivo de *A. robusta* e *A. trichophylla*. Além disso, a metabolômica de tais cultivares também foi capaz de apontar acúmulo de metabólitos a depender da estação climática. Isso sugere uma necessidade maior em se padronizar a data da coleta do material vegetal para estudos quimiotaonômicos. Em breve, os cultivares terão suas atividades preditas frente à LDH-5, assim como serão isolados e ensaiados os metabólitos que foram considerados ativos nas espécies de Asteraceae.

Palavras-chave: Metabolômica, Quimiometria, *Decision Tree*, Asteraceae, Quimiotaonomia, Germinação, Cultivo, Lactato Desidrogenase-5 (LDH-5), Estacional.

ABSTRACT

SANTOS, F. A. **Metabolomics for cultivar analysis, chemotaxonomy and screening of lactate dehydrogenase-5 (LDH-5) inhibitors in Asteraceae species**. 2019. 178 f. Thesis (Doctoral). Faculdade de Ciências Farmacêuticas de Ribeirão Preto – Universidade de São Paulo, Ribeirão Preto, 2019.

Metabolomics can be defined as the study of all the metabolites of a given biological system. Such a study enables holistic and integrated approaches from several fields of science. Thus, problems in taxonomic delimitations, analysis of environmental influence on plants, and screening of bioactive metabolites are examples of areas where metabolomics can aid. The Asteraceae family, with approximately 25,000 species, has been the subject of taxonomic studies for decades. Thus, the metabolomics approach applied to chemotaxonomy may aid in such classifications, or reveal new findings. On the other hand, it is argued that secondary metabolites are influenced by many environmental factors. In this way, this work also investigated the influence of seasonality, luminosity, fertilization and circadian rhythm on the metabolome of cultivars of *Aldama robusta* and *A. trichophylla* (Asteraceae). The cultivation of *A. robusta* was also studied for its germination and morphological development. The wide chemical variability in Asteraceae also places it as a family with potential for the screening of bioactive metabolites. A relevant biochemical target is the enzyme lactate dehydrogenase-5 (LDH-5), which is related to the increased survival of cancer cells. In this way, the metabolomic study in Asteraceae species may allow the discovery of active compounds against LDH-5. The metabolome of all species of this work was acquired by UHPLC-ESI-UV/DAD-MS, with Orbitrap™ analyzer. For the chemotaxonomy and LDH-5 inhibition studies, 250 species of Asteraceae, from several genera of 16 tribes, were studied. For the studies of the influence of certain environmental factors on the metabolome of *A. robusta* and *A. trichophylla*, collections were made at different circadian rhythms, seasons of the year and different cultivars based on soil fertilization and luminosity. The chemical data were analyzed by several chemometric and data mining techniques, such as PCA, HCAbp, OPLS-DA and decision tree. Thus, OPLS-DA analysis with Asteraceae species was able to separate some tribes and point out discriminant metabolites in each of them, although PCA and HCAbp did not reveal clustering. Likewise, OPLS-DA and the decision tree were able to point out some probable metabolites correlated with extracts considered "active" against LDH-5, as well as to predict extracts from other Asteraceae species that had not yet been assayed. As for the germination and development studies of cultivars, more adequate conditions for the cultivation of *A. robusta* and *A. trichophylla* were described. In addition, the metabolomics of such cultivars were also able to indicate accumulation of metabolites depending on the climatic season. This suggests a greater need to standardize the date of collection of plant material for chemotaxonomic studies. Soon, the cultivars will have their activities predicted against the LDH-5, as well as the metabolites that were considered active in the species of Asteraceae will be isolated and tested.

Keywords: Metabolomics, Chemometrics, Decision Tree, Asteraceae, Chemotaxonomy, Germination, Cultivation, Lactate Dehydrogenase-5 (LDH-5), Seasonal.

INTRODUÇÃO GERAL

1 – METABOLÔMICA

Metabolômica é uma ciência emergente, que visa o estudo abrangente e integrado de uma ampla gama de metabólitos de baixo peso molecular, incluindo abordagens de triagem global e análise quantitativa e/ou qualitativa de todo o metaboloma presente em um dado sistema biológico, como uma célula específica, tecido ou organismo, ou o conjunto deles, seja oriundo de plantas, microrganismos, animais ou organismos marinhos (VILLAS-BÔAS, 2005; DE VOS et al., 2007; KRASTANOV, 2010). A metabolômica é conduzida por meio de técnicas de detecção e identificação de metabólitos de acordo com suas características físico-químicas. Há diversas formas de se analisar o metaboloma, mas as principais técnicas hoje empregadas são as hifenadas, que combinam dois ou mais equipamentos para se detectar e identificar os metabólitos. Como técnicas de separação podemos citar a Cromatografia Gasosa (GC), Cromatografia Líquida de Alta Performance (HPLC) e Cromatografia Líquida de Ultra Performance (UHPLC), por meio das quais pode-se obter o tempo de eluição de cada metabólito e assim tentar identificá-lo. Como técnicas de detecção, podemos citar a Espectrometria de Massas (MS), Ressonância Magnética Nuclear (RMN), Ultravioleta (UV), UV com de Arranjo de Diodos (UV/DAD) e Ionização em Chama (FID).

Cada equipamento possui vantagens e desvantagens, os quais irão nortear a escolha da técnica hifenada a ser empregada na análise do metaboloma. Por exemplo, no que diz respeito às técnicas cromatográficas descritas anteriormente, todas tem um mesmo princípio de funcionamento: as substâncias são conduzidas por uma fase móvel e, ao interagir com uma fase estacionária, são mais ou menos retidas e então eluídas; deste modo, cada substância terá um tempo de eluição, relacionado com suas descrições físico-químicas (tamanho molecular, polaridade, solubilidade, etc). Entretanto, a GC tem seu uso restrito a metabólitos que não são termolábeis, uma vez que ela requer altas temperaturas para volatilizá-los e separá-los através da fase estacionária. Na GC, a fase móvel é um gás inerte, e a fase estacionária é um líquido ou polímero sobre um sólido inerte. Nas HPLC e UHPLC, por outro lado, a fase móvel é um líquido, e a fase estacionária é composta por partículas microscópicas (geralmente sílica igual ou menor que 5 µm) que podem ser porosas (exemplo, ACE C18®, ACE), ou envoltas por uma película porosa (exemplo,

Kinetex® – Phenomenex), ou ainda são compostas por polímeros, como as colunas monolíticas (exemplo, Onyx monolithic C8– Phenomenex). As HPLC e UHPLC utilizam bombas que possibilitam gradientes de alta pressão e, deste modo, a utilização de fases estacionárias com partículas de menor tamanho ou porosidade. São justamente estas partículas de menor tamanho e/ou porosidade que possibilitam uma melhor separação dos metabólitos, uma vez que há maior interação físico-química entre a substância e a fase estacionária. As designações “C8” e “C18” expressas neste parágrafo informam que a sílica da fase estacionária é tratada com grupos alquilas que podem ser C_8H_{17} ou $C_{18}H_{37}$. Este tratamento torna a fase estacionária apolar (cromatografia de fase reversa), o que permite o uso de fases móveis polares de baixo custo (como água) e uma melhor separação de substâncias de polaridade intermediária, como metabólitos secundários.

Já no que se diz respeito aos métodos de detecção, sua escolha também depende de vantagens e desvantagens, bem como das finalidades. Por exemplo, o detector UV/DAD permite analisar um cromatograma em diferentes comprimentos de onda. O UV/DAD é importante, uma vez que permite inferir a classe química do metabólito que estamos visualizando. A espectroscopia por RMN é capaz de determinar com precisão a estrutura química das substâncias. O fenômeno de RMN ocorre quando núcleos alinhados a um campo magnético aplicado são induzidos a absorver energia e a mudar a orientação do “giro” (*spin*) em relação ao campo aplicado (PAVIA et al., 2010). Na RMN são estudados principalmente a ressonância dos átomos de hidrogênio e carbono. A RMN é útil pois cada núcleo de 1H (ou ^{13}C) da estrutura química possuirá frequência diferente, a depender de outros átomos próximos. Por exemplo, a densidade eletrônica de valência varia de um próton para o outro na estrutura da substância. Tais elétrons criam um campo magnético que se opõe ao efeito do campo magnético aplicado, “protegendo” (*blindagem diamagnética* ou *anisotropia*) tais núcleos (PAVIA et al., 2010). Tendo em vista que o “ambiente químico” (quantidade de elétrons próximos aos núcleos de 1H ou ^{13}C) é diferente dentro das estruturas químicas, é possível gerar uma impressão digital do metabólito, baseado em sua estrutura química. Entretanto, a técnica de RMN tem a desvantagem de, em geral, ter baixo limite de detecção/quantificação e baixa sensibilidade quando comparada às técnicas de MS. Por outro lado, por meio da MS é possível detectar e identificar uma substância com base em sua capacidade de ionização e peso molecular, bem como nos fragmentos produzidos quando é exposta a energias elevadas de ionização,

ou gases de colisão. Dentre as técnicas de ionização, podemos descrever: ionização por elétrons (EI), ionização química à pressão atmosférica (APCI), ionização por *electrospray* (ESI) e ionização por dessorção a laser assistida por matriz (MALDI). A ESI foi a técnica utilizada nos experimentos deste trabalho. Na ESI, um potencial elétrico é aplicado ao capilar onde há fase móvel contendo as substâncias que estão eluindo do cromatógrafo, formando um aerossol (cone de Taylor). Esse processo resulta na formação de gotas com superfícies carregadas. Um gás inerte secante aquecido é utilizado para evaporar o solvente, aumentando a repulsão eletrostática entre as cargas formais nas gotículas, promovendo a explosão *coulombica*, liberando os íons. A fonte de ESI pode oxidar ou reduzir a substância, gerando íons moleculares ($M^{+\circ}$ ou M°), mas também pode formar moléculas protonadas ($[M + H]^+$) ou desprotonadas ($[M + H]^-$), bem como coordenação com íons metálicos Na^+ , Cl^- , K^+ e NH_4^+ , por meio de reações ácido-base (CROTTI et al., 2006). A extensão da ionização das substâncias é um processo que depende não apenas da voltagem aplicada ao capilar, mas também dos grupamentos químicos das substâncias. Assim, substâncias que apresentam grupamentos básicos, como aminas, amidas e ésteres, são mais facilmente protonadas e, portanto, analisadas (visualizadas/detectadas) no modo positivo. Por outro lado, substâncias com grupos ácidos, como os fenóis e ácidos carboxílicos, são mais facilmente desprotonados e, portanto, analisadas no modo negativo (CROTTI et al., 2006). A técnica de ESI é considerada uma ionização “suave”, uma vez que ela gera pouca fragmentação. Sendo assim, ela é muito útil em metabolômica, quando acoplada a cromatógrafos e espectrômetros de alta resolução, como CG, HPLC ou UHPLC. A técnica torna-se ainda mais eficaz em metabolômica quando há espectrômetros de massas em *tandem* (em sequência, MS/MS). Por meio desta técnica, é possível selecionar e promover a fragmentação de íons específicos (de interesse). Do mesmo modo, há diversos analisadores que podem ser utilizados em MS. O espectrômetro utilizado neste trabalho foi o de alta resolução com transformada de Fourier (HRFT), com analisador do tipo Orbitrap. O orbitrap opera aprisionando o íon em uma armadilha de Kingdon (*Kingdon trap*), que consiste em um eletrodo externo em forma de barril e um eletrodo interno fusiforme. Assim, por meio de um campo eletrostático, o íon fica aprisionado e oscilando em órbita ao redor do eletrodo fusiforme. A frequência das oscilações harmônicas do íon independe de sua velocidade, mas é inversamente proporcional à relação massa carga m/z (VESSECCHI et al., 2011), segundo a equação:

$$\omega = \sqrt{k/(m/z)}$$

onde ω é a frequência angular e k é força constante do potencial. Os dados químicos (metaboloma) deste projeto foram adquiridos por UHPLC-UV/DAD-ESI-MS.

Tendo em vista os diferentes equipamentos e técnicas para se realizar a metabolômica, esta vem sendo subdividida em demais subáreas, cada qual com sua finalidade e propósito. De maneira geral, metabolômica pode ser subdividida em *targeted* e *untargeted*. Essas definições são mais conceituais do que analíticas. *Targeted* refere-se à metabolômica direcionada, ou seja, análises do metaboloma que objetivam a detecção/identificação, e mesmo a quantificação de metabólitos específicos, ao passo que *untargeted* seria a detecção/identificação de tantos metabólitos quanto for possível. Assim, desde a maneira como se coleta uma amostra, se extrai seus metabólitos, e se adquire os dados químicos, pode influenciar em se classificar uma análise como *targeted* ou *untargeted*.

Independente da abordagem de escolha e, embora tenha sido comentado no primeiro parágrafo da introdução deste capítulo que “há diversas formas de se realizar a metabolômica”, a sociedade científica de metabolômica, *Metabolomics Society*, vem realizando esforços para se padronizar as condições mínimas e aceitáveis para estudos metabolômicos (SUMNER et al., 2007, SALEK et al., 2013; SCHRIMPE-RUTLEDGE et al., 2016). Novamente, as técnicas, em geral, podem variar de acordo com os objetivos da análise, mas é de suma importância que elas sejam minuciosamente descritas a fim de se garantir a confiabilidade do experimento, bem como sua reprodutibilidade (SUMNER et al., 2007). Por exemplo, no caso da metabolômica de plantas, não há um procedimento único sobre como se deve processar o material coletado que se vai analisar (com N_2 líquido ou com gelo seco, por exemplo), ou como se deve realizar a extração das amostras com solvente (com ou sem subsequente etapa de *clean up*, por exemplo) (SUMNER et al., 2007), mas é necessário especificar tais condições, padronizar a técnica e evitar o máximo de variações externas à análise (DE VOS et al., 2007, SUMNER et al., 2007, NAZ et al., 2014). Não obstante, a metabolômica sempre gera uma grande quantidade de dados químicos, os quais necessitam de subsequentes passos para serem analisados. Assim, os

dados brutos provenientes das técnicas de LC-MS podem ser pré-tratados por *softwares* de processamento como o Metalign (LOMMEN, 2009; LOMMEN e KOOLS, 2012) e MZmine (KATAJAMMA et al., 2006; PLUSKAL et al., 2010), dentre outros, sendo que cada qual possui vantagens e desvantagens, como apresentado na Figura 1 (CASTILLO et al., 2011). Após o processamento, são geradas matrizes de dados que são normalizados pelas mais diferentes técnicas estatísticas, como *Pareto scaling*, *level scaling*, *log transformation*, etc (VAN DEN BERG et al., 2006). Após o pré-tratamento e a normalização, os dados do metaboloma podem ser analisados por diferentes técnicas quimiométricas (item 2, Capítulo 1).

Data preprocessing steps covered by the proteomic and metabolomic software tools discussed in this paper.

Software	Filtering	Peak detection	Deisotoping	Alignment	Gap filling	Visualization	Identification	Normalization	Batch mode
apLCMS	/	/	-	/	/	-	-	-	/
Decon2LS	/	/	/	-	-	/	-	-	-
Madison Metabolomics Consortium Database	-	-	-	-	-	-	/	-	-
MAVEN	/	/	-	/	-	/	-	-	-
MetAlign	/	/	-	/	-	-	-	-	/
MetaboAnalyst	-	/	-	/	/	-	-	/	-
MSight	-	-	-	-	-	/	-	-	-
MZedDB	-	-	-	-	-	-	/	-	-
MZmine 2	/	/	/	/	/	/	/	/	/
OpenMS/TOPT	-	-	/	-	-	/	/	-	/
Sirius	-	-	-	-	-	-	-	-	-
SuperHlim	-	/	/	/	-	-	-	/	-
XCMS	/	/	-	/	/	/	/	-	-

Figura 1 - Comparação de diferentes softwares de processamento de dados (Extraído de Castillo *et al.*, 2011).

Por fim, as diferentes abordagens metabolômicas podem ser utilizadas para diversas finalidades, em estudos envolvendo conceitos de ecologia química (ecoquímica ou ecofisiologia), quimiotaxonomia, triagem de extratos para a busca de substâncias ativas, controle de qualidade de amostras complexas, biogeografia, dentre outros. De maneira geral, em estudos de ecologia a metabolômica pode ser útil para identificar substâncias altamente correlacionadas com sazonalidade, ritmo circadiano, condições ambientais, interações ecoquímicas entre os seres vivos (como alelopatia), dentre outros (GOBBONETO e LOPES, 2007; SAMPAIO et al., 2016; PADILLA-GONZÁLEZ et al., 2017). Já estudos de quimiotaxonomia (e por vezes de quimiossistemática) buscam agrupar e caracterizar *taxa* (plural de *taxon*) de acordo com seus respectivos metabólitos biossintetizados. Tais estudos podem ser úteis na sistemática de plantas (PADILLA-GONZÁLEZ et al., 2016; GALLON et al., 2018) ou na triagem de rotas biossintéticas específicas de certos *taxa*, estudos de distribuição biogeográfica ou mesmo no controle de

qualidade de amostras biológicas, como por exemplo, plantas morfológicamente similares, mas quimicamente distintas (PADILLA-GONZÁLEZ et al., 2017). A metabolômica para triagem de substâncias biologicamente ativas é útil para apontar substâncias discriminantes e altamente correlacionadas com determinada atividade biológica em amostras complexas (CHAGAS-PAULA et al., 2015; SANTOS et al., 2018).

2 – TÉCNICAS QUIMIOMÉTRICAS

A enorme quantidade de dados gerados em metabolômica pode ser analisada pelas mais diversas técnicas quimiométricas, a fim de se extrair informações relevantes. Os dados obtidos são usualmente multidimensionais e, neste caso, são chamados de “multivariados”. Os diferentes métodos matemáticos e estatísticos aplicados aos dados podem ser a Análise das Componentes Principais (*Principal Component Analysis* – PCA), Análise de Agrupamento Hierárquico (*Hierarchical Cluster Analysis* – HCA), Mínimos Quadrados Parciais (*Partial Least Squares* – PLS) e sua variante ortogonal (OPLS), sendo que estas duas últimas podem ser utilizadas para Análises Discriminantes (*Discriminant Analysis* – DA) (WOLD et al., 2001; TRYGG e WOLD, 2002), compondo as conhecidas técnicas PLS-DA e OPLS-DA. Assim sendo, a HCA e PCA são análises multivariadas não supervisionadas, pois elas são úteis para o reconhecimento de padrões, tendências e agrupamentos não direcionados das amostras. Já o PLS-DA é considerado uma análise multivariada supervisionada, pois o agrupamento observado se correlacionará com certa resposta previamente informada à análise, as quais podem ser numérica ou descritiva, como o valor de dada bioatividade ou os *clusters* formados por gêneros ou origem biogeográfica.

Melhor explicando, tanto a PCA quanto o PLS-DA são técnicas de redução de dimensionalidade (RENCHER, 2002), onde um conjunto de amostras com variáveis multidimensionais são reduzidas no que se chama de “componentes principais”, no caso da PCA, e “variáveis latentes”, no caso do PLS-DA. Ambas técnicas projetam em um hiperplano bidimensional o que seria impossível de se observar em um espaço multidimensional. A PCA sempre traça a primeira componente na direção de maior variância encontrada entre as variáveis do sistema. Uma segunda componente, ortogonal à segunda, explica uma quantidade de variância menor que a primeira componente, mas

maior que a terceira componente, à qual é ortogonal às duas primeiras. Quanto maior a complexidade de um conjunto de dados multidimensionais, tão mais difícil é de se explicar com apenas duas componentes a maior variância encontrada, sendo, portanto, necessários mais componentes. O PLS funciona da mesma forma, e o termo componentes principais é utilizado às vezes para designar as variáveis latentes. Entretanto, a diferença é que as variáveis latentes sempre apontam na direção de maior variância sobre uma determinada informação (a resposta em Y) fornecida ao sistema. Assim sendo, o PLS-DA correlaciona a variância do espaço multidimensional (a matriz X) com uma resposta informada (matriz Y) (WOLD et al., 2001). Já o OPLS-DA é considerado uma modificação do PLS-DA, onde é realizada uma separação das variações da matriz X em duas partes, uma linearmente relacionada à matriz Y, chamada de componentes preditivas, e outra que é ortogonal a Y, chamada de não preditiva com a resposta Y e usualmente importante para detectar variações internas dentro dos grupos, mas não correlacionadas com a resposta (TRYGG e WOLD, 2002). Isso gera uma maior interpretabilidade, visto que as variações ortogonais a Y nos informam as diferenças dentro de X, mas que não se correlacionam com Y. Todas estas técnicas de redução dimensional do espaço multidimensional geram um novo gráfico (projeção do hiperplano bidimensional) de amostras, chamado agora de *scores*, os quais se relacionam diretamente com o novo gráfico gerado para as variáveis, chamado agora de *loadings*.

Já a HCA é um método de “*Cluster Analysis*” propriamente dito. Em outras palavras, a HCA necessariamente agrupa por similaridade as amostras do estudo, procurando por padrões e tendências dentro dos agrupamentos (*clusters*). Ela é representada na forma de uma “árvore” onde os *clusters* se formam ao longo de uma distância. Quanto mais próximo a 0, mais similar. Uma distância comum bastante utilizada é a Euclidiana, a qual se trata da distância entre dois vetores, um x (x1, x2, ... , xp) e outro y (y1, y2, ... , yp), definido como:

$$d(x, y) = \sqrt{x - y'(x - y)}$$

$$= \sqrt{\sum_p (x_j - y_j)(x_j - y_j)}$$

Há muitas formas de se fazer uma HCA, sendo as mais utilizadas o *Single Linkage*, o qual se utiliza da distância entre as amostras mais próximas para construir os

clusters, e o método de Ward (*Ward's Method*) (WARD, 1963), o qual usa a variância total dentro das amostras para definir os *clusters*. O resultado é que o método de Ward minimiza a somatória da distância dentro dos *clusters* (RENCHER, 2002; LAROSE, 2006). A escolha de qual método se deve utilizar é empírica e depende das necessidades ou do que se pretende informar. Entretanto, a HCA não informa a confiabilidade dos *clusters* que se formam e, portanto, os resultados devem ser interpretados com cautela. Todavia, a HCA por reamostragem (*bootstrapping*), ou HCAbp, pode ser utilizada se obter não somente os agrupamentos hierárquicos, mas também seus graus de incertezas, fornecendo estimativas da probabilidade do *cluster*. A HCAbp pode ser realizada por meio do pacote *pvclust* no *software* R (The R Project website, <http://www.r-project.org/>) (SUZUKI, R. e SHIMODARA, H., 2006).

Por fim, é necessário introduzir aqui os conceitos de *Data Mining* (Mineração de Dados) e como ela tem sido utilizada pela quimiometria recentemente (BUYDENS et al., 1999; MUTIHAC, L. e MUTIHAC R., 2008) para a classificação, predição de amostras e descoberta de variáveis latentes. *Data Mining* trata da resolução de problemas presentes em bancos de dados, explorando-os por meio de técnicas computacionais, visando o reconhecimento de padrões e/ou tendências e classificações. Tais padrões nos permitem realizar predições em novos dados. Atualmente, muitas das técnicas desenvolvidas em *Data Mining* podem ser utilizadas pela quimiometria, tais como Análise Bayesiana (*NaiveBayes*), Redes Neurais Artificiais e *Decision Tree* (Árvores de Decisão). Na representação gráfica da *Decision Tree*, os “nós” da árvore representam um determinado atributo (uma variável) que foi testado por condições específicas (concentração, por exemplo) e apontado como uma variável latente, discriminante, de uma classe (WITTEN, 2011). Há vários algoritmos que calculam uma *decision tree* e não convém explicá-los aqui, embora os mais utilizados sejam o C4.5, desenvolvido por Ross Quinlan (QUINLAN, 1993), e o J.48 (BHARGAVA et al., 2013), implementação do C4.5 no *software* gratuito Weka (HALL et al., 2009).

Durante as técnicas quimiométricas, são apontados pelos modelos diversos picos do cromatograma (como o cromatograma de íons totais – TIC) relevantes ou discriminantes de amostras específicas. A etapa final da metabolômica é, portanto, identificar (e por vezes até mesmo quantificar) tais picos. Para tanto, faz-se uso de diversas técnicas de identificação, tal como o isolamento por meio de técnicas cromatográficas da

substância correspondente ao pico de interesse e sua identificação por meio de técnicas espectroscópicas e espectrométricas usuais. Entretanto, muitas vezes tais picos são de baixa intensidade (portanto podendo ser pouco concentrado na amostra) e, além disso, às vezes a resolução cromatográfica não permite um fácil isolamento e identificação da substância. Deste modo, a identificação necessita ser realizada sem o prévio isolamento do pico, mas no próprio extrato, o que é chamado de “desrepliação” (do inglês, *dereplication*) (HUBERT et al., 2017). A desrepliação é enormemente facilitada pela presença de substâncias previamente isoladas ou adquiridas comercialmente (biblioteca de substâncias padrão), quando se pode injetá-las no sistema LC-UV-MS/MS e confirmar a substância que se quer identificar por meio de similaridades (correlação/correspondência/*match*) com o tempo de retenção, classe química, massa, e o perfil de ligações intramoleculares. Para tanto, é importante que se tenha as informações físico-químicas (banco de dados) de tal biblioteca. Ademais, se o usuário dispuser de um conjunto de substâncias químicas estruturalmente diversificadas e que foram analisadas pelo mesmo gradiente de eluição, é possível gerar modelos de tempos de retenção cromatográfico com base na estrutura química e prever o tempo de retenção de outras substâncias que não fazem parte da biblioteca de padrões do laboratório (EUGSTER et al., 2014; OLIVEIRA et al., 2015). Além disso, se o laboratório dispuser de espectros MS/MS ou MSⁿ em seu banco de dados, pode-se também construir modelos de predição de fragmentação de substâncias (inclusive de outros bancos de dados, de outros laboratórios) e aumentar a quantidade picos identificados (YANG et al., 2013; ALLARD et al., 2016; WANG et al., 2016). A identificação/desrepliação de substâncias químicas em metabolômica também segue parâmetros e níveis de identificação estabelecidos pela *Metabolomics Society* (SUMNER et al., 2007; SALEK et al., 2013). A equipe AsterBioChem (www.asterbiochem.org/), junto à qual este trabalho foi realizado, possui o banco de dados AsterDB, que pode ser acessado por meio do *link* “AsterDB” no site ou do catálogo da ZINC (<http://zinc15.docking.org/catalogs/asternp/>), banco de dados gratuito com milhões de estruturas químicas de substâncias.

3 – FAMÍLIA ASTERACEAE

A família Asteraceae (Compositae) compreende um dos maiores grupos de plantas das Angiospermas, com aproximadamente 25.000 espécies organizadas em 1.600

gêneros (VERDI et al., 2005; FUNK et al., 2009), 44 tribos e 13 subfamílias (Figura 2) (PANERO et al., 2014), representando 10% de todas as Angiospermas (PAVARINI et al., 2018). Esta família de plantas apresenta ampla distribuição geográfica em todos os continentes, sendo mais presente em climas tropicais e subtropicais (PAVARINI et al., 2018). A classificação mais recente (PANERO et al., 2014) das subfamílias de Asteraceae a subdivide em 13 subfamílias (Figura 2), sendo descrita a mais nova subfamília Famatinanthoideae, um grupo “irmão” das subfamílias Multisoideae-Asteroideae, as quais correspondem a mais de 99% das Asteraceae (PANERO et al., 2014).

A taxonomia de Asteraceae se inicia em 1816, com o botânico Henri Cassini agrupando as espécies em 19 tribos (CASSINI, 1816). Desde então, a família passou por diversos reagrupamentos e classificações. Destas, talvez ainda a mais conhecida seja a de Bremer (BREMER, 1996), o qual, baseado em critérios morfológicos, subdividiu a família em 4 subfamílias (Barnadesioideae, Carduoideae, Cichorioideae e Asteroideae) e 14 tribos (Figura 3), sendo a subfamília Barnadesioideae, com uma única tribo, considerada um clado basal dentro da família. O *taxon* Mutiseae foi considerado indefinido (Figura 3).

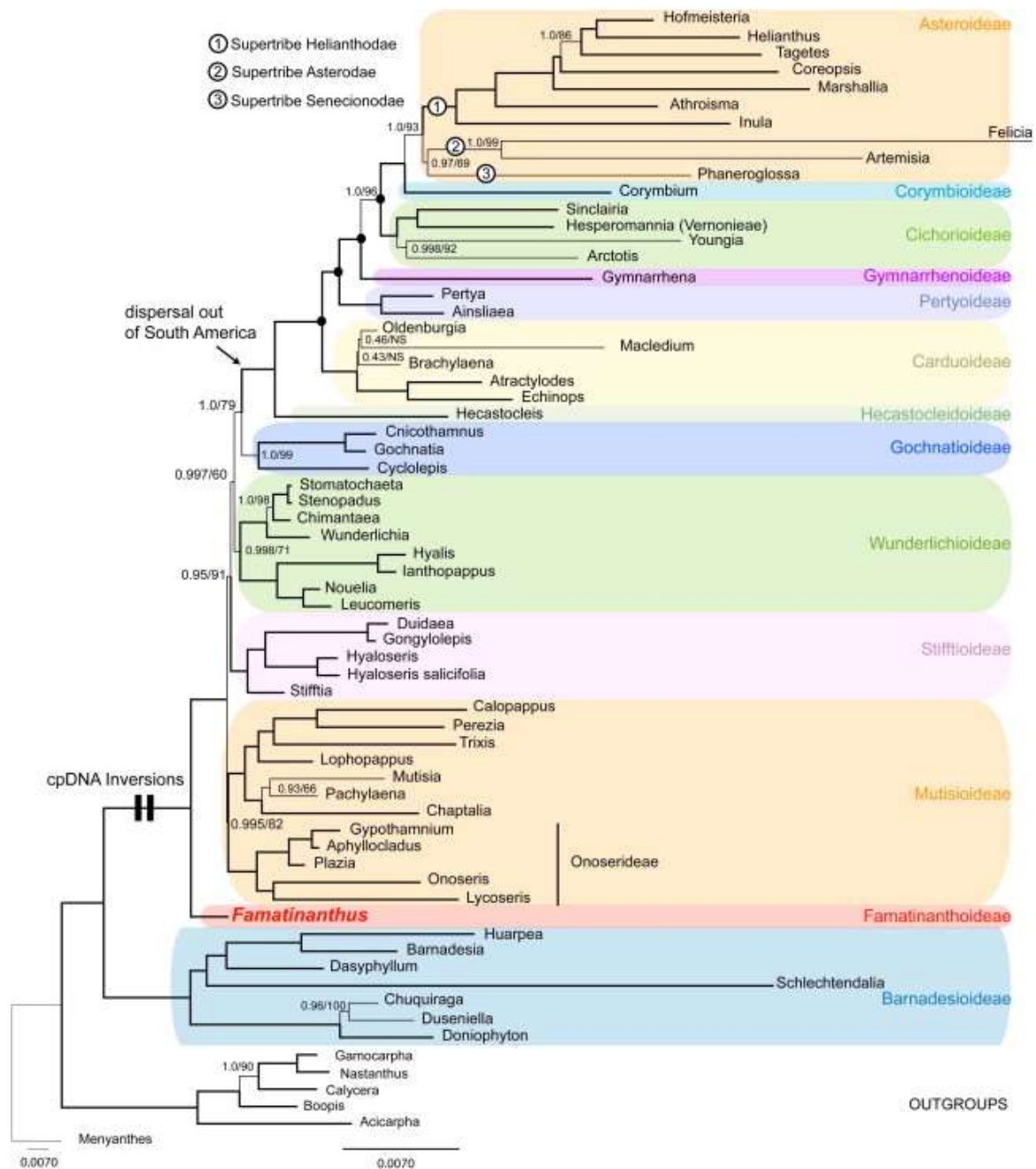


Figura 2 - Árvore filogenética de Asteraceae construída com base em dados de DNA e inferência Bayesiana. Destacam-se as 13 subfamílias. Figura extraída de Panero e colaboradores (PANERO et al., 2014).

Uma característica marcante de Asteraceae é a presença de capítulos (*capitula* plural em latim), que é uma inflorescência especializada que pode conter centenas de pequenas flores em um receptáculo discoide ou arredondado contendo brácteas. Muitos membros de Asteraceae possuem dois tipos de flores: flores do disco, localizadas na porção central do capítulo, e flores do raio, localizadas na periferia. O ovário das

Asteraceae é ífero (se encontra dentro do receptáculo) e apresenta um único óvulo. A família é frequentemente conhecida como “família do girassol” (*Helianthus annus* L.), espécie da tribo Heliantheae, subfamília Asteroideae.

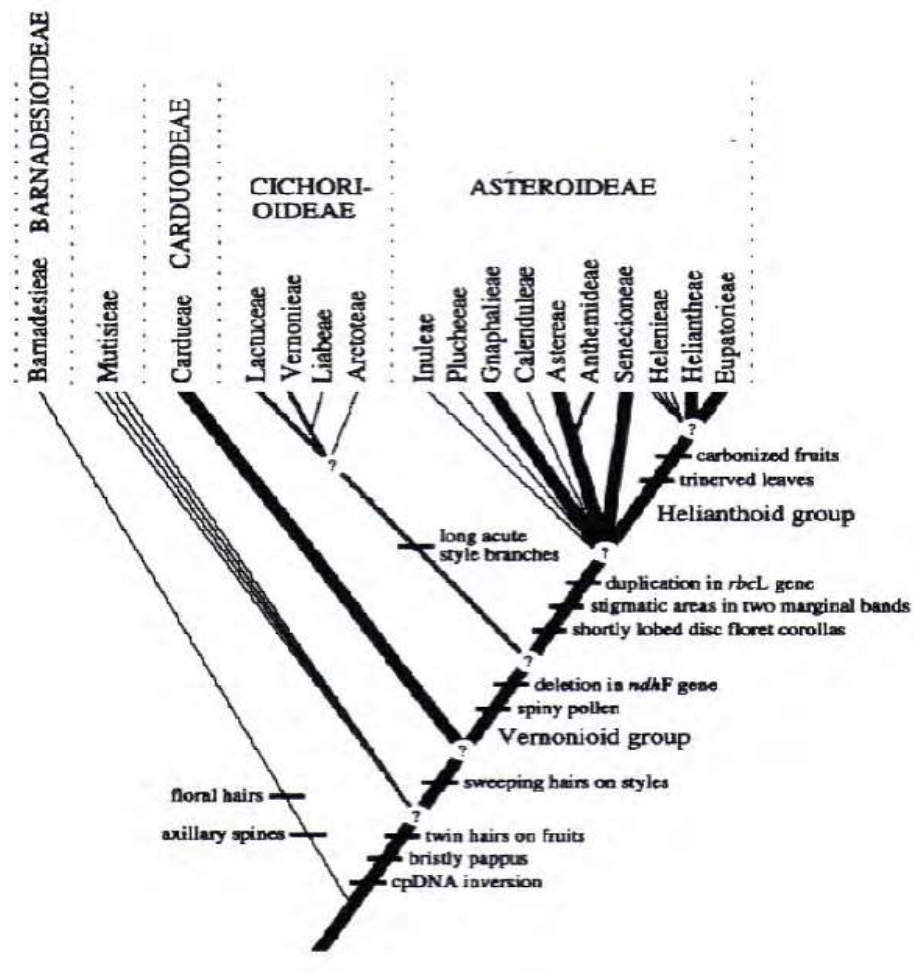


Figura 3 - Diagrama da família Asteraceae construído com base em caracteres morfológicos (BREMER et al., 1996). Evidenciam-se as 4 subfamílias (Barnadesioideae, Carduoideae, Cichorioideae e Asteroideae, e suas respectivas tribos). Figura oriunda de Calabria e coautores (CALABRIA et al., 2006).

Do ponto de vista químico, a família Asteraceae é caracterizada pela presença de lactonas sesquiterpênicas (LST) e sesquiterpenos furanos (SEAMAN, 1982; ALVES et al., 2018, in: SÜLSEN e MARTINO, 2018). Deste modo, elas poderiam ser estudadas computacionalmente na classificação das espécies de Asteraceae (DA COSTA et al., 2005; SCOTTI et al., 2012) (Figura 4). Além destes compostos, a família também possui outras classes de metabólitos secundários, tais como os diterpenos (labdanos, cauranos, pimaranos, clerodanos), triterpenos, derivados do ácido quínico (ácidos gálico e quínico,

por exemplo), flavonoides (flavonois, flavonas, flavanonas, flavanonois, isoflavonas, flavan-3-ois, chalconas, antocianinas, auronas e isoflavonas), derivados do ácido *trans*-cinâmico e poliacetilenos (Figura 4) (ZDERO e BOHLMANN, 1990; BOHM e STUESSY, 2001). Compostos voláteis das classes dos monoterpenos, sesquiterpenos e fenilpropanoides também estão presentes em óleos essenciais. Por outro lado, não é comum encontrar na família alcaloides, iridoideis, estilbenos, policetídeos e taninos.

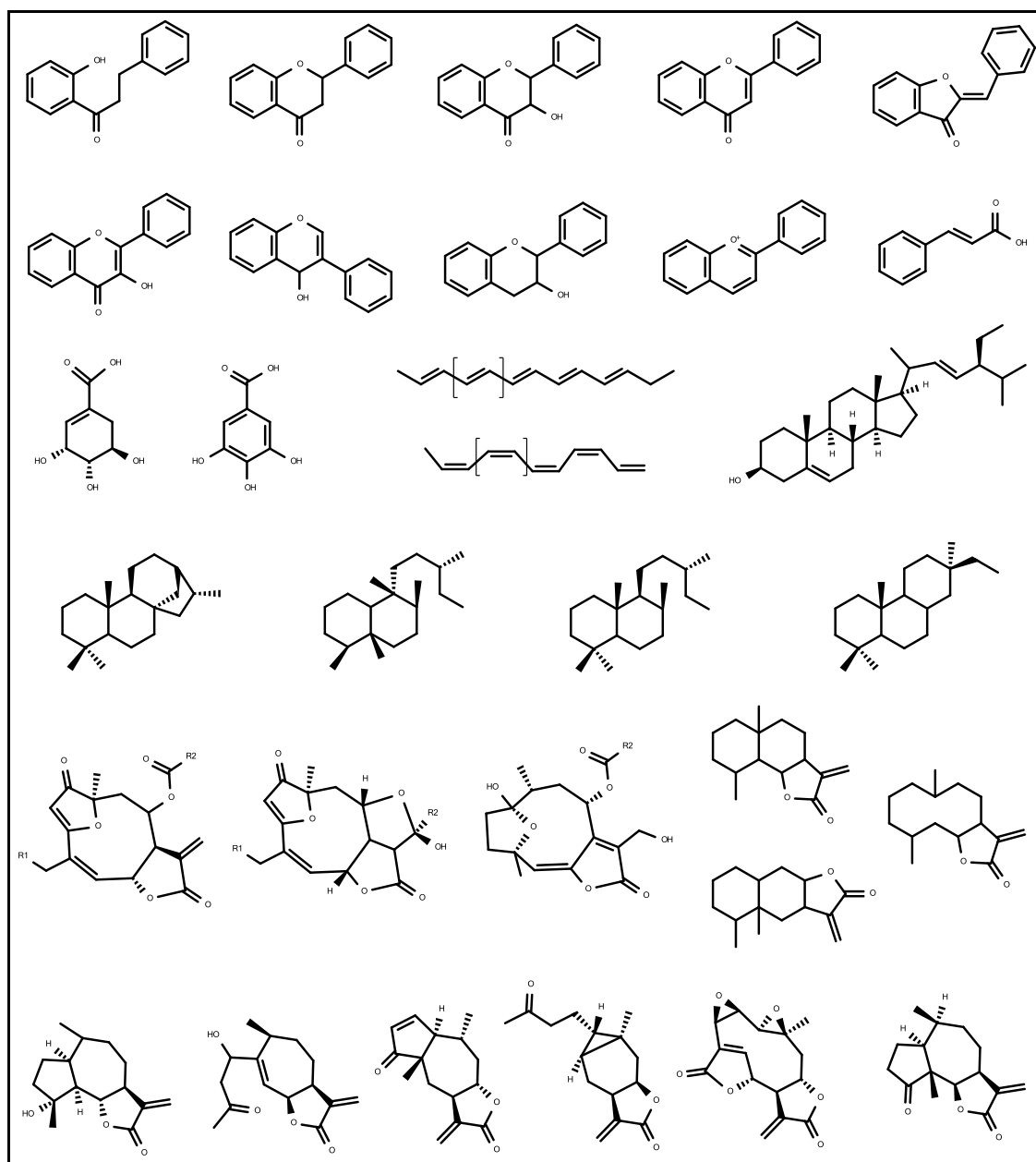


Figura 4 – Algumas das principais classes químicas de metabólitos secundários em Asteraceae. Da esquerda para a direita e de cima para baixo: flavonoides chalcona, flavanona, flavananol, flavona, aurona, flavonol, isoflavona, flavan-3-ol, antocianina; ácido *trans*-cinâmico; ácido quínico, ácido gálico; poliacetilenos; triterpeno

(estigmasterol); diterpenos caurano, clerodano, labdano, pimarano; LSTs furanoheliangolido, eremantolido, hirsutinolido, eudesmanolido, eremofilanolido, germacrolido, guaianolido, xantanolido, helenanolido, carabrone, mikanolido e ambrosanolido.

Neste trabalho, foi realizado um estudo quimiotaxonômico cerca de 250 espécies de Asteraceae, distribuídas em 16 tribos (ver Apêndice A), sendo que algumas delas são de grande importância do ponto de vista fitoquímico e, portanto, pode-se comentar com mais detalhes suas constituições químicas. O gênero *Baccharis* L. (tribo Astereae, subfamília Asteroideae), por exemplo, foi amplamente estudado, tendo sido descrito diversas atividades biológicas para seus constituintes químicos, dentre as quais efeito antioxidante, antibacteriano, antiprotozoário, antiviral, antifúngico, anti-inflamatório, antimutagênico, dentre outras (CAMPOS et al., 2016). Pertence a este gênero, inclusive, a planta medicinal, de amplo uso etnofarmacológico, carqueja (*B. trimera*), conhecida por melhorar problemas gastrointestinais. De maneira geral, as principais classes químicas do gênero são: flavonoides (flavonois, flavonas, flavanonas, flavanonois, glicosilados ou não), derivados do ácido *trans*-cinâmico (derivados cumarínicos e ácidos cafeoilquínicos, principalmente), fenilpropanoides (ácido ferúlico, por exemplo), monoterpenos, sesquiterpenos, diterpenos (clerodanos, labdanos, cauranos, por exemplo), triterpenos e esteróis. Outro gênero de grande importância, *Mikania* Willd. (tribo Eupatorieae, subfamília Asteroideae), merece ser sucintamente comentado, tendo em vista inclusive a importância etnofarmacológica e econômica de duas de suas espécies, *M. laevigata* e *M. glomerta*, popularmente conhecidas como “guaco”. Ambas espécies foram oficializadas como plantas medicinais (Formulário de Fitoterápicos da Farmacopeia Brasileira, 1ª edição). O xarope de guaco possui o marcador químico cumarina (1,2-benzopirona, derivada do ácido *trans*-cinâmico) (NAPIMOGA e YASTUDA, 2010) e é amplamente utilizado pelo seu efeito antitussígeno. Outras classes químicas do gênero são: diterpenos (labdanos, pimaranos e cauranos), derivados do ácido *trans*-cinâmico (ácidos cafeoilquínicos), flavonoides, triterpenos, esteróis e LST, em especial as do tipo mikanolido (RUFATTO et al., 2012). Outros gêneros importantes em Asteraceae são a *Tithonia* Desf. ex Juss. e *Helianthus* L. (ambos da tribo Heliantheae e subfamília Asteroideae), bem representados pelas suas espécies *T. diversifolia* e *H. annuus*. Em ambas *T. diversifolia* e *H. annuus* foi relatada a presença de derivados cumarínicos, flavonois, flavonas, diterpenos caurenóicos, LST dos tipos guaianolido, eudesmanolido, germacrolido, heliangolido and furanoheliangolidos,

triterpenos e esteróis (CHAGAS-PAULA et al., 2012; MARSNI et al., 2015). As atividades biológicas relatadas para *T. diversifolia* (extratos brutos ou substâncias isoladas) são diversas, dentre as quais efeito antimalárico, antiviral, antimicrobiano, antiespasmódico, anti-inflamatório e citotóxico (CHAGAS-PAULA et al., 2012). *H. annuus* é mais estudado por causa do óleo fixo extraído de suas sementes, tendo grande importância agrônômica e econômica. Ainda assim, foram relatadas algumas atividades para seus metabólitos secundários, tais como efeitos herbicida, antimicrobiano, antifúngico e regulares de crescimento (MARSNI et al., 2015). Outro gênero de constituição química muito semelhante à de *Helianthus* e *Tithonia* é *Aldama* La Llave (outrora *Viguiera* Kunth). De fato, não apenas são relativamente semelhantes em suas constituições químicas, como também pertencem à mesma tribo e subtribo (Helianthinae). Para os constituintes químicos e extratos brutos deste gênero, foram relatadas diversas atividades biológicas, como inseticida (MARQUINA et al., 2001), antimicrobiana (CANALES et al., 2008; PORTO et al., 2009), anti-tripanosoma (NOGUEIRA et al., 2016), anti-inflamatória (SHORR et al., 2002; ARAKAWA et al., 2008; CHAGAS-PAULA et al., 2015), muitas das quais principalmente relacionadas à presença de LST budleína A e ao diterpeno do tipo caurano ácido *ent-caur-16-en-19-oico*. Por fim, outros dois gêneros muito importantes em Asteraceae serão brevemente descritos aqui, *Vernonia* Schreb. e *Lychnophora* Mart., ambos da tribo Vernonieae (subfamília Cichorioideae), mas respectivamente das subtribos Vernoniinae e Lychnophorinae. Uma planta medicinal de *Vernonia* bem conhecida é a assa-peixe (*V. polyanthes*), sendo utilizada para o tratamento de gripe e como expectorante. De maneira geral, em ambas as subtribos foram relatados os mesmos constituintes químicos, como fenilpropanoides, flavonoides (principalmente flavonóis, favanonas, flavonas e chalconas), derivados do ácido *trans*-cinâmico, derivados do ácido quínico, sesquiterpenos, diterpenos (pouco comum) e LST germacrolidos, eudesmanolido e guaianolido. A diferença sutil se encontra nas LST do tipo furanoheliangolido, com os tipos eremantolido e goyazensolido como marcadores de Lychnophorinae (KELES et al., 2010), e hirsutinolido como marcador de Vernoniinae (BOHLMAN et al., 1981; KOS et al., 2006). Para a subtribo Lychnophorinae, são descritas na literatura atividades antioxidante, antineoplásica, genotóxica, citotóxica, tripanomicida, antimicrobiana e anti-inflamatória (KELES et al., 2010).

Pelo o que foi exposto neste item do capítulo 1 deste trabalho, fica evidente a grande importância taxonômica, farmacológica e econômica de muitas espécies de Asteraceae. Entretanto, muitas das espécies destes gêneros estão sendo extintas, tendo em vista as atividades econômicas predatórias no campo, a inexistência de estudos de cultivo, a negligência e o descaso das autoridades políticas. De maneira geral, a população não vê fins práticos para se preservar tais espécies, sendo algumas inclusive consideradas como ervas daninhas e pragas, a exemplo da *T. diversifolia*. Deste modo, cabe à comunidade científica demonstrar os ganhos econômicos e ambientais que se pode conseguir por meio da preservação e até mesmo do cultivo de tais espécies. Para tanto, investigações a respeito das atividades biológicas dos metabólitos de tais gêneros, seja em termos farmacológicos ou ecoquímicos, são de grande importância para que se justifiquem suas preservações.

CONCLUSÕES

O presente trabalho objetivou utilizar a metabolômica e técnicas quimiométricas para auxiliar na compreensão da quimiotaxonomia de Asteraceae, ao nível da classificação de tribos, realizar a triagem em extratos de metabólitos ativos na inibição de LDH-5, e analisar a influência de determinados fatores ambientais no metaboloma de plantas. De maneira geral, podemos concluir que as abordagens metabolômicas apresentadas neste trabalho foram capazes de:

- a) apontar metabólitos discriminantes de cada tribos, os quais podem ser considerados marcadores quimiotaxonômicos (capítulo 1);
- b) prever a atividade de extratos de outras espécies de Asteraceae ainda não ensaiados, bem como apontar prováveis substâncias correlacionadas com extratos ativos (capítulo 2), o que possibilita menor tempo e custo de análise, assim como o isolamento direcionado dos metabólitos;
- c) apontar marcadores químicos reacionados com estações do ano, o que permite um melhor controle de qualidade de amostras vegetais, bem como advertir para uma padronização das datas das coletas e secagem de folhas que antecedem os estudos quimiotaxonômicos (capítulo 3).

Além disso, o presente trabalho também traz informações de condições ideais de germinação e influência da adubação e sombreamento sobre a morfologia de *A. robusta* (e *A. trichophylla* e *T. diversifolia*, para o caso da germinação), o que permite um melhor delineamento experimental que vise a obtenção de uma maior quantidade de cultivares, bem como de um maior número de folhas e maiores ramos por planta.

REFERÊNCIAS

- Agati, G., Azzarello, E., Pollastri, S., Tattini, M. Flavonoids as antioxidants in plants: location and functional significance. *Plant science*. 2012, 196, 67-76.
- Agboola, D. A., Idowu, W. F., Kadiri, M. Seed germination and seedling growth of the Mexican sunflower *Tithonia diversifolia* (Compositae) in Nigeria, Africa. *Revista de biología tropical*, 2006. 395-402.
- Ajao, A. A., Moteetee, A. N. *Tithonia diversifolia* (Hemsl) A. Gray.(Asteraceae: Heliantheae), an invasive plant of significant ethnopharmacological importance: A review. *South African Journal of Botany*. 2017, 113, 396-403.
- Alcantara, G. B., Honda, N. K., Ferreira, M. M. C., Ferreira, A. Chemometric analysis applied in ¹H HR-MAS NMR and FT-IR data for chemotaxonomic distinction of intact lichen samples. *Analytica Chimica Acta*, 2007, 595, 3–8
- Allard, P. M., Péresse, T., Bisson, J., Gindro, K., Marcourt, L., Pham, V. C., ... Wolfender, J. L. Integration of molecular networking and in-silico MS/MS fragmentation for natural products dereplication. *Analytical chemistry*. 2016, 88, 3317-3323.
- Alvarenga, S. A. V., Ferreira, M. J. P., Emerenciano, V. P., Cabrol-Bass, D. Chemosystematic studies of natural compounds isolated from Asteraceae: characterization of tribes by principal component analysis. *Chemometric and Intelligent Laboratory Systems*. 2001, 56, 27-37.
- Ambrósio, S. R., Arakawa, N. S., Esperandim, V. R., de Albuquerque, S., Da Costa, F. B. Trypanocidal activity of pimarane diterpenes from *Viguiera arenaria* (Asteraceae). *Phytotherapy Research: An International Journal Devoted to Pharmacological and Toxicological Evaluation of Natural Product Derivatives*. 2008, 22, 1413-1415.
- Arakawa, N. S., Schorr, K., Ambrosio, S. R., Merfort, I., Da Costa, F. B. Further sesquiterpene lactones from *Viguiera robusta* and the potential anti-inflammatory activity of a heliangolide: inhibition of human neutrophil elastase release. *Zeitschrift für Naturforschung C*. 2008, 63, 533-538.
- Augoff, K., Hryniewicz-Jankowska, A., Tabola, R. Lactate dehydrogenase 5: An old friend and a new hope in the war on cancer. *Cancer Letters*. 2015, 358, 1-7.
- Barbosa, A. R., Yamamoto, K., Valio, I. F. M. Effect of light and temperature on germination and early growth of *Vochysia tucanorum* Mart., Vochysiaceae, in cerrado and forest soil under different radiation levels. *Revista Brasileira de Botânica*. 1999, 22, 275-280,.
- Baskin, C. C., Baskin, J. M. Role of temperature and light in the germination ecology of buried seeds of weedy species of disturbed forests. II. *Erechtites hieracifolia*. *Canadian Journal of Botany*. 1996, 74, 2002-2005.
- Beaman, J. H. Plant Taxonomy. *Clin. Dermatol.*, 1986, 4, 23-30.
- Bhargava, N., Sharma, G., Bhargava, R., Mathuria, M. Decision tree analysis on J48 algorithm for data mining. 2013, 3, 1114-1119.
- Blaženović, I., Kind, T., Sa, M. R., Ji, J., Vaniya, A., Wancewicz, B., ... , Showalter, M. R. Structure annotation of all mass spectra in untargeted metabolomics. *Analytical chemistry*. 2019, 91, 2155-2162.

Bohlmann, F., Zdero, C., King, R. M., Robinson, H. Hirsutinolides and other sesquiterpene lactones from *Vernonia* species. *Phytochemistry*. 1982, 21, 695-699.

Bohm, B. A., Stuessy, T. F. *Flavonoids of the sunflower family (Asteraceae)*. Springer Science Business Media. 2001

Bombo, A. B., Oliveira, T. S. D., Appezzato-da-Glória, B., Novembre, A. D. D. L. C. Seed germination of Brazilian *Aldama* species (Asteraceae). *Journal of Seed Science*, (AHEAD), 2015.

Bremer, K. Major clades and grades of the Asteraceae. In: Hind DN, Beebtj HJ (es) *Compositae: systematic. Proceedings of the international Compositae conference*, Kew. Hind DJN (Editor-in-Chief), v 1, Royal Botanic Gardens, Kew, 1994, 1-7.

Buydens, L.M.C., Reijmers, T.H., Beckers, M.L.M., Wehrens, R. Molecular data-mining: a challenger for chemometrics. *Chemometrics and Intelligent Laboratory Systems*. 1999, 49, 121-133.

Calabria, L. M., Emerenciano, V. P., Ferreira, M. J., SCotti, M. T., Mabry, T. J. A phylogenetic analysis of tribes of the Asteraceae based on phytochemical data. *Natural Product Communications*. 2007, 2, 277-285.

Calabria, L. M., Emerenciano, V. P., Ferreira, M. J. P. et al. A phylogenetic analysis of tribes of the Asteraceae based on phytochemical data. *Natural Products Communications*. 2007, 2, 277-285.

Campos, F., R., Bressan, J., Godoy Jasinski, V. C., Zuccolotto, T., da Silva, L. E., Bonancio Cerqueira, L. *Baccharis* (Asteraceae): chemical constituents and biological activities. *Chemistry biodiversity*. 2016, 13, 1-17.

Canales, M. et al. Antimicrobial Activity of the Extracts and Essential Oil of *Viguiera dentata*. *Pharmaceutical Biology*. 2008, 46, 719–723.

Chagas-Paula, D. A., de Oliveira, R. B., da Silva, V. C., Gobbo-Neto, L., Gasparoto, T. H., Campanelli, A. P., ... Da Costa, F. B. Chlorogenic acids from *Tithonia diversifolia* demonstrate better anti-inflammatory effect than indomethacin and its sesquiterpene lactones. *Journal of Ethnopharmacology*. 2011, 136, 355-362.

Chagas-Paula, D. A., Oliveira, R. B., Rocha, B. A., Da Costa, F. B. Ethnobotany, chemistry, and biological activities of the genus *Tithonia* (Asteraceae). *Chemistry biodiversity*. 2012, 9, 210-235.

Chagas-Paula, D. A., Oliveira, T. B., Zhang, T., Edrada-Ebel, R., Da Costa, F. B. Prediction of anti-inflammatory plants and discovery of their biomarkers by machine learning algorithms and metabolomic studies. *Planta medica*. 2015, 81, 450-458.

Chagas-Paula, D., Zhang, T., Da Costa, F., Edrada-Ebel, R. A metabolomic approach to target compounds from the Asteraceae family for dual COX and LOX inhibition. *Metabolites*. 2015, 5, 404-430.

Çirak, C., Radušienė, J., Ivanauskas, L., Janulis, V. Variation of bioactive secondary metabolites in *Hypericum organifolium* during its phenological cycle. *Acta Physiol. Plant*. 2007, 29, 197–203.

Correia, M. V., Fokou, H. H., Marcelo, J. et al., Self-organizing maps a good tool for classification of subfamily Asteroideae. *Journal of Medicinal Plants Research*. 2012, 6, 1207-1218.

Correia, M. V., Scotti, M. T., Ferreira, M. J. P., et al. Self-organizing maps as a new tool for classification of plants at lower hierarchical levels. *Natural Products Communications*. 2008, 3, 1723-1730.

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

Dantas, B. F., Alves, E., Aragão, C. A., Tofanelli, B. D., Corrêa, M. R., Rodrigues, J. D., ... Nakagawa, J. Germinação de sementes de capim-marmelada (*Brachiaria plantaginea* (link) hitc.) tratadas com ácido giberélico. *Revista Brasileira de Sementes*. 2001, 23, 27-34.

Darwin, C. *On the Origin of Species by Means of Natural Selection* (J. Murray, London). 1859.

De Toledo, J., Ambrósio, S., Borges, C., Manfrim, V., Cerri, D., Cruz, A., Da Costa, F. In vitro leishmanicidal activities of sesquiterpene lactones from *Tithonia diversifolia* against *Leishmania braziliensis* promastigotes and amastigotes. *Molecules*. 2014, 19, 6070-6079.

De Vos, R. C., Moco, S., Lommen, A., Keurentjes, J. J., Bino, R. J., & Hall, R. D. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature Protocols*, 2007, 2, 778-791.

Deiab, S., Mazzi, E., Messeha, S., Mack, N., & Soliman, K. F. A. High-throughput screening to identify plant derived human LDH-A inhibitors. *European journal of medicinal plants*. 2013, 3, 603.

Delgado, G., de Vivar, A. R., Herz, W. Sesquiterpene lactones from *Viguiera* species. *Phytochemistry*. 1982, 21, 1305-1308.

El Marsni, Z., Torres, A., Varela, R. M., Molinillo, J. M., Casas, L., Mantell, C., ... Macias, F. A. Isolation of bioactive compounds from sunflower leaves (*Helianthus annuus* L.) extracted with supercritical carbon dioxide. *Journal of agricultural and food chemistry*. 2015, 63, 6410-6421.

Emerenciano, V. D. P., Ferreira, Z. S., Kaplan, M. A. C., Gottlieb, O. R. A chemosystematic analysis of tribes of Asteraceae involving sesquiterpene lactones and flavonoids. *Phytochemistry*. 1987, 26, 3103-3115.

Emerenciano, V. P., Cabrol-Bass, D., Ferreira, M. J. P. et al. Chemical evolution in the Asteraceae. The oxidation-reduction mechanism and production of secondary metabolites. *Nat Prod Comm*. 2006, 1, 495-507.

Scotti, M. T., Emerenciano, V., Ferreira, M. J. P. et al. Self-organizing maps of molecular descriptors for sesquiterpene lactones and their application to the chemotaxonomy of the Asteraceae family. *Molecules*. 2012, 17, 468-4702.

Eugster, P.J., Boccard, J., Debrus, B., Bréant, L., Wolfender, J-L., Martel, S., Carrupt, P-A. Retention time prediction for dereplication of natural products (C_xH_yO_z) in LC-MS metabolite profiling. *Phytochemistry*. 2014, 108, 196-207.

Fariman, Z. K., Azizi, M., Noori, S. Seed germination and dormancy breaking techniques for *Echinacea purpurea* L. *Journal of Biology Environmental Science* 2011, 5, 7-10.

Funk, V. A., Susanna, A., Stuessy, T. F., Bayer, R. J. Systematics, evolution, and biogeography of Compositae. *International Association for Plant Taxonomy*. Vienna, 2009.

Gallon, M. E., Monge, M., Casoti, R., Da Costa, F. B., Semir, J., Gobbo-Neto, L. Metabolomic analysis applied to chemosystematics and evolution of megadiverse Brazilian *Vernoniae* (Asteraceae). *Phytochemistry*. 2018, 150, 93-105.

Gasteiger J. Engel, T. (Eds.), *Cheminformatics*. WILEY-VCH Verlag GmbH Co. KgaA, Weinheim, 2003.

Gobbo-Neto, L., Lopes, N. P. Plantas medicinais: fatores de influência no conteúdo de metabólitos secundários. *Química nova*. 2007, 30, 374.

Goffin, E., Ziemons, E., De Mol, P., de Madureira, M. D. C., Martins, A. P., da Cunha, A. P., ... Frederich, M. In vitro antiplasmodial activity of *Tithonia diversifolia* and identification of its main active constituent: tagitinin C. *Planta medica*. 2002, 68, 543-545.

Granchi, C., Roy, S., Giacomelli, C., Macchia, M., Tuccinardi, T., Martinelli, A., ... & Funel, N Discovery of N-hydroxyindole-based inhibitors of human lactate dehydrogenase isoform A (LDH-A) as starvation agents against cancer cells. *Journal of medicinal chemistry*. 2011, 54, 1599-1612.

Guo, S., Ge, Y., Jom, K. N. A review of phytochemistry, metabolite changes, and medicinal uses of the common sunflower seed and sprouts (*Helianthus annuus* L.). *Chemistry Central Journal*. 2017, 11, 95.

Hall, M., Frak, E., Holmes, G., Pfahinger, B., Reutemann, P., Witten, I.H. The weka data mining software: An update. 2009, 11, 10-18.

Harmer, S. L. The circadian system in higher plants. *Annual Review of Plant Biology*. 2009, 60, 357-377.

Hegnauer, R. Chemical characters in plant taxonomy: some possibilities and limitations. *Pure and Applied Chemistry*. 1967, 14, 173-188.

Hooley, R. Gibberellins: perception, transduction and responses. In *Signals and Signal Transduction Pathways in Plants*. 1994, 293-319. Springer, Dordrecht.

Hubert, J., Nuzillard, J. M., & Renault, J. H. Dereplication strategies in natural product research: How many tools and methodologies behind the same concept? *Phytochemistry Reviews*. 2017, 16: 55-95.

INCA (2017) Mortes por câncer aumentaram 31% no Brasil em 15 anos. Disponível em: <http://www2.inca.gov.br/wps/wcm/connect/agencianoticias/site/home/noticias/2017/mortes_por_cancer_aumentaram_31_no_brasil_em_15_anos_diz_oms>. Acessado em 20/fevereiro/2019.

Isejima, E. M., Figueiredo-Rrieiro, R. C. L., Zaidan, L. B. P. Fructan composition in adventitious tuberous roots of *Viguiera discolor* Baker (Asteraceae) as influenced by daylength. *New Phytologist*. 1991, 119, 149-154.

Panero, J. L., Freire, S. E., Espinar, L. A., Crozier, B. S., Crozier, G. E., Barboza, J. J. C. Resolution of deep nodes yields an improved backbone phylogeny and a new basal lineage to study early evolution of Asteraceae. *Molecular Phylogenetics and Evolution*. 201, 80, 43-53

Kanani, M. R., Rahiminejad, M. R., Sonboli, A., Mozaffarian, V., Osaloo, S. K., Ebrahimi, S. N. Chemotaxonomic significance of the essential oils of 18 *Ferula* species (Apiaceae) from Iran. *Chemistry & Biodiversity*, 2011, 8, 503-517.

Kaneko, M., Itoh, H., Ueguchi-Tanaka, M., Ashikari, M., Matsuoka, M. The α -amylase induction in endosperm during rice seed germination is caused by gibberellin synthesized in epithelium. *Plant Physiology*. 2002, 128, 1264-1270.

Katajamaa, M., Miettinen, J., Oresic, M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics*. 2006, 22, 634–636.

Keles, L. C., Melo, N. I. D., Aguiar, G. D. P., Wakabayashi, K. A. L., Carvalho, C. E. D., Cunha, W. R., ... Lopes, N. P. Lychnophorinae (Asteraceae): a survey of its chemical constituents and biological activities. *Quimica Nova*. 2010, 33, 2245-2260.

Kessner, D., Chambers, M., Burke, R., Agus, D., Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics*. 2008, 24, 2534-2536.

Kim, H.K., Khan, S., Wilson, E.G., Kricun, S.D.P., Meissner, A., Goraler, S., Deelder, A.M., Choi, Y.H., Verpoorte, R. Metabolic classification of South American *Ilex* species by NMR-based metabolomics. *Phytochemistry*. 2010, 71, 773–784.

Kirmizi, S., Guleryuz, G., Arslan, H. Germination responses to GA3 and short-time chilling of three endemic species: *Tripleurospermum picheleri*, *Cirsium leucopis* and *Senecio olympicus* (Asteraceae). *Plant Species Biology*. 2011, 26, 51-57.

Kos, O., Castro, V., Murillo, R., Poveda, L., Merfort, I. Ent-kaurane glycosides and sesquiterpene lactones of the hirsutinolide type from *Vernonia triflosculosa*. *Phytochemistry*. 2006, 67, 62-69.

Krastanov A. Metabolomics – The state of art. *Biotechnol. Biotechnol. Eq.* 2010, 1, 1537-1543.

Lankadurai, B. P., Nagato, E. G., Simpson, M. J. Environmental metabolomics: an emerging approach to study organism responses to environmental stressors. *Environmental Reviews*. 2013, 21, 180-205.

Larose, D. T. (2006). *Data Mining – Methods and Models*. Central Connecticut State University.

Lee, S., Oh, D. G., Lee, S., Kim, G., Lee, J., Son, Y., ... Lee, C. Chemotaxonomic metabolite profiling of 62 indigenous plant species and its correlation with bioactivities. *Molecules*. 2015, 20, 19719-19734.

Lommen, A. Kools, H.J. MetAlign 3.0: performance enhancement by efficient use of advances in computer hardware. *Metabolomics*. 2012, 8, 719-726.

Lommen, A. MetAlign: Interface-Driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Analytical Chemistry*. 2009, 81, 3079-3086.

Magenta, M.A.G. *Viguiera* Kunth (Asteraceae , Heliantheae) na América do Sul e sistemática das espécies do Brasil. Tese de Doutorado. 339. Universidade de São Paulo, 2006.

Maguire, J. D. Speed of Germination—Aid In Selection And Evaluation for Seedling Emergence And Vigor 1. *Crop science*. 1962, 2, 176-177.

Manerba, M., Di Ianni, L., Govoni, M., Roberti, M., Recanatini, M., Di Stefano, G. Lactate dehydrogenase inhibitors can reverse inflammation induced changes in colon cancer cells. *European Journal of Pharmaceutical Sciences*. 2017, 96, 37-44.

Marquina, S., Maldonado, N., Garduño-Ramírez, M. L., Aranda, E., Villarreal, M. L., Navarro, V., ... Alvarez, L. Bioactive oleanolic acid saponins and other constituents from the roots of *Viguiera decurrens*. *Phytochemistry*. 2001, 56, 93-97.

Martucci, M. E. P., Gobbo-Neto, L. Differential secondary metabolite accumulation and performance of *Chlosyne lacinia* fed with *Tithonia diversifolia* or *Vernonia polyanthes*. *Biochemical systematics and ecology*. 2016, 68, 156-162.

Martucci, M. E. P., De Vos, R. C., Carollo, C. A., Gobbo-Neto, L. Metabolomics as a potential chemotaxonomical tool: application in the genus *Vernonia* Schreb. *PLoS One*. 2014, 9, e93149.

McDonald, M. B., Khan, A. A. Acid Scarification and protein synthesis during Seed Germination I. *Agronomy journal*. 1983, 75, 111-114.

Moreira, E. A., Pilon, A. C., Andrade, L. E., Lopes, N. P. New Perspectives on Chlorogenic acid accumulation in harvested leaf tissue: impact on traditional medicine Preparations. *ACS Omega*. 2018, 3, 18380-18386.

Mutihac, L. Mutihac, R. Mining in chemometrics. *Analytic Chimica ACTA*. 2008, 612, 1-18.

Napimoga, M. H., Yatsuda, R. Scientific evidence for *Mikania laevigata* and *Mikania glomerata* as a pharmacological tool. *Journal of Pharmacy and Pharmacology*. 2010, 62, 809-820.

Naz, S., Vallejo, M., García, A., Barbas, C. Method validation strategies involved in non-targeted metabolomics. *Journal of Chromatography A*. 2014, 1353, 99-105.

Nicolete, R. et al. Budlein A from *Viguiera robusta* inhibits leukocyte-endothelial cell interactions, adhesion molecule expression and inflammatory mediators release. *Phytomedicine*. 2009, 16, 904-915.

Nogueira, M., Da Costa, F., Brun, R., Kaiser, M., Schmidt, T. ent-pimarane and ent-kaurane diterpenes from *Aldama discolor* (Asteraceae) and their antiprotozoal activity. *Molecules*. 2016, 21, 1237.

Oliveira, A. B., Oliveira, T. S., Glória, B. A., Novembre, A. D. L. C. Seed germination of Brazilian *Aldama* species (Asteraceae). *Journal of Seed Science*. 2015, 37, 185-191.

Oliveira, T. B., Gobbo-Neto, L., Schmidt, T. J., Da Costa, F. B. Study of chromatography retention of natural terpenoids by cheminformatic tools. *Journal of Chemical Information and Modeling*. 2015, 55, 26-38, 2015.

ONU BRASIL (2017) OMS: câncer mata 8,8 milhões de pessoas anualmente no mundo. Disponível em: < <https://nacoesunidas.org/oms-cancer-mata-88-milhoes-de-pessoas-anualmente-no-mundo/>>. Acessado em 20/fevereiro/2019.

Padilla-González, G. F., Diazgranados, M., Da Costa, F. B. Biogeography shaped the metabolome of the genus *Espeletia*: a phytochemical perspective on an Andean adaptive radiation. *Scientific reports*. 2017, 7, 8835.

Padilla-González, G.F., Aldana, J.A., Da Costa, F.B. Chemical characterization of two morphologically related *Espeletia* (Asteraceae) species and chemometric analysis based on essential oil components. *Revista Brasileira de Farmacognosia*. 2016, 1-7.

Patel, R. G., Mankad, A. U. Effect of gibberellins on seed germination of *Tithonia rotundifolia* Blake. International Journal of Innovative Research in Science, Engineering and Technology. 2014, 3, 10680-10684.

Pavarini, D., Bauermeister, A., Semir, J., Monge, M., Lopes, J., Peporine, N. P. (2018) Phenolic Compounds from the Brazilian genus *Lychnophora* Mart. (Asteraceae). Chapter in ACS Symposium Series · January 2018

Pluskal, T., Castillo, S., Villar-Briones, A. Oresic, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC bioinformatics. 2010, 11, 395–406.

Porto, T. S., Furtado, N. A., Heleno, V. C., Martins, C. H., Da Costa, F. B., Severiano, M. E., ... Ambrosio, S. R. Antimicrobial ent-pimarane diterpenes from *Viguiera arenaria* against Gram-positive bacteria. Fitoterapia. 2009, 80, 432-436.

Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers. 1993.

Rencher, A. C. (2002). Methods of Multivariate Analysis. Bringham Young University, 2nd Ed.

Reynolds, T. The evolution of chemosystematics. Phytochemistry. 2007, 68, 22-24.

Rivas-ubach, A., Sardans, J., Pérez-trujillo, M., Estiarte, M., Peñuelas, J. Strong relationship between elemental stoichiometry and metabolome in plants. Proceedings of the National Academy of Sciences. 2012, 109, 4181–4186.

Rosa, A. L. Aplicação de ferramentas computacionais e analíticas na construção de bibliotecas de produtos naturais da família Asteraceae. Tese de Doutorado. Universidade de São Paulo, 2018.

Rozema, J., Van de Staaij, J., Björn, L. O., Caldwell, M. UV-B as an environmental factor in plant life: stress and regulation. Trends in Ecology and Evolution. 1997, 12, 22–28.

Rufatto, L. C., Gower, A., Schwambach, J., Moura, S. Genus *Mikania*: chemical composition and phytotherapeutical activity. Revista Brasileira de Farmacognosia. 2012, 22, 1384-1403.

Ruggiero, P. G., Zaidan, L. B. Estudo de desenvolvimento de *Viguiera robusta* Gardn., uma Asteraceae do cerrado. Brazilian Journal of Botany. 1997, 20, 01-09.

Salek, R. M., Steinbeck, C., Viant, M. R., Goodacre, R., Dunn, W. B. The role of reporting standards for metabolite annotation and identification in metabolomic studies. GigaScience. 2013, 2, 13.

Santos, F. A. D., Sousa, I. P., Furtado, N. A., Costa, F. B. D. Combined OPLS-DA and decision tree as a strategy to identify antimicrobial biomarkers of volatile oils analyzed by gas chromatography–mass spectrometry. Revista Brasileira de Farmacognosia. 2018, 28, 647-653.

Sartor, T., Xavier, V. B., Falcão, M. A., Mondin, C. A., Dos Santos, M. A., Cassel, E., ... Santarém, E. R. Seasonal changes in phenolic compounds and in the biological activities of *Baccharis dentata* (Vell.) GM Barroso. Industrial crops and products. 2013, 51, 355-359.

Schilling, E. E. Panero, J. L. A revised classification of subtribe Helianthinae (Asteraceae: Heliantheae) II. Derived lineages. Botanical Journal of the Linnean Society. 2011, 167, 311–331.

Schorr, K., García-Piñeres, A.J., Siedle, B., Merfort, I., Da Costa, F.B. Guaianolides from *Viguiera gardneri* inhibit the transcription factor NF-kappaB. *Phytochemistry*. 2002, 60, 733–40.

Schweiger, R., Heise, A., Persicke, M., Müller, C. Interactions between the jasmonic and salicylic acid pathway modulate the plant metabolome and affect herbivores of different feeding types. *Plant Cell and Environment*. 2014, 37, 1574–1585.

Scognamiglio, M., D'Abrosca, B., Esposito, A., Fiorentino, A. Chemical composition and seasonality of aromatic Mediterranean plant species by NMR-based metabolomics. *Journal of Analytical Methods in Chemistry*. 2015, 1–9.

Scognamiglio, M., Fiumano, V., D'Abrosca, B., Esposito, A., Choi, Y. H., Verpoorte, R., Fiorentino, A. Chemical interactions between plants in Mediterranean vegetation: The influence of selected plant extracts on *Aegilops geniculata* metabolome. *Phytochemistry*. 2014, 106, 69–85.

Seaman, F. C. Sesquiterpene lactones as taxonomic characters in the Asteraceae. *The Botanical Review*. 1982. 48, 121-595.

Sneath, P. H., Sokal, R. R. Numerical taxonomy. The principles and practice of numerical classification. 1973.

Sokal, R. R., Sneath, P. H. Principles of numerical taxonomy. Principles of numerical taxonomy. 1963

Spring, O. Schilling, E.E. The sesquiterpene lactone chemistry of *Helianthus* sect. *Atrorubentes* (Asteraceae: Heliantheae). *Biochemical Systematics and Ecology*. 1991, 19, 59–79.

Spring, O. Schillingt, E.E. Chemosystematic investigation of the annual species of *Helianthus* (Asteraceae)*. *Biochemical Systematics and Ecology*. 1989, 17, 519–528.

Spring, O. Schillingt, E.E. The sesquiterpene lactone chemistry of *Helianthus* Sect . *Ciliares* (Asteraceae: Heliantheae)*. *Biochemical Systematics and Ecology*. 1990, 18, 139–143.

Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., Fan, T.W., Fiehn, O., Goodacre, R., Griffin, J.L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R. Proposed minimum reporting standards for chemical analysis. *Metabolomics*. 2007, 3, 211-221.

Suzuki, R., Shimodaira, H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*. 2006, 22, 1540-1542.

Trygg, J., Wold, S. Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics*. 2002, 16, 119-128.

Uehara, A., Nakata, M., Uchida, A., Iwashina, T. Chemotaxonomic consideration of flavonoids from the leaves of *Chrysanthemum arcticum* subsp. *arcticum* and *yezoense*, and related species. *Biochemical Systematics and Ecology*. 2017, 73, 11-15.

Valério, D. A., Cunha, T. M., Arakawa, N. S., Lemos, H. P., Da Costa, F. B., Parada, C. A., ... Verri Jr, W. A. Anti-inflammatory and analgesic effects of the sesquiterpene lactone budlein A in mice: inhibition of cytokine production-dependent mechanism. *European journal of pharmacology*. 2007, 562, 155-163.

Van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., Van der Werf, M. J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC genomics*. 2006, 7, 142.

Verdi, L.G., Brighente, I.M.C., Pizzolatti, M.G. Gênero *Baccharis* (Asteraceae): aspectos químicos, econômicos e biológicos. *Química Nova*. 2005, 28, 85-94.

Villas-Bôas, S.G., Rasmussen, S., Lane, G.A. Metabolomics or metabolite profiles? *Trends in biotechnology*. 2005, 23, 385–386.

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., ... Porto, C. Sharing and community curation of mass spectrometry data with GNPS. *Nature biotechnology*. 2016, 34, 828.

Wang, G. L., Semenza, G. L. General involvement of hypoxia-inducible factor 1 in transcriptional response to hypoxia. *Proceedings of the National Academy of Sciences*. 1993, 90, 4304-4308.

Warburg, O., *Über den Stoffwechsel der Tumore*. 1926. Springer, Berlin. Translated: *The metabolism of tumors*, 1930. Arnold Constable, London.

Warburg, O. On the origin of cancer cells. *Science*. 1956, 123: 309-314.

Ward, P. S., Thompson, C. B. Metabolic reprogramming: a cancer hallmark even warburg did not anticipate. *Cancer cell*. 2012, 21, 297-308.

WHO – World Health Organization. Cancer. Disponível em < <http://www.who.int/news-room/fact-sheets/detail/cancer>>. Acessado em 19/fevereiro/2019.

Wink, M. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry*. 2003, 64, 3-19.

Witten, I. H., Eibe, F., Hall, M. A. (2011). *Data Mining – practical machine learning tools and techniques*. 3rd Ed., Morgan Kaufmann Publishers.

Wold, S., Sjöström, M., Eriksson, L. Partial least squares projections to latent structures (PLS) in chemistry. *Encyclopedia of computational chemistry*. 2002, 3.

Xie, H., Hanai, J. I., Ren, J. G., Kats, L., Burgess, K., Bhargava, P., ... Wang, X. Targeting lactate dehydrogenase-a inhibits tumorigenesis and tumor progression in mouse models of lung cancer and impacts tumor-initiating cells. *Cell metabolism*. 2014, 19, 795-809.

Yang, J. Y., Sanchez, L. M., Rath, C. M., Liu, X., Boudreau, P. D., Bruns, N., ... & Wong, W. R. Molecular networking as a dereplication strategy. *Journal of natural products*. 2013, 76, 1686-1699.

Ward Jr, J. H. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 1963, 58, 236-244.

Zarpelon, A. C., Fattori, V., Souto, F. O., Pinto, L. G., Pinho-Ribeiro, F. A., Ruiz-Miyazawa, K. W., ... Casagrande, R. The sesquiterpene lactone, budlein A, Inhibits antigen-induced arthritis in mice: Role of NF- κ B and Cytokines. *Inflammation*. 2017, 40, 2020-2032.

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