UNIVERSITY OF SÃO PAULO

FACULTY OF PHARMACEUTICAL SCIENCES OF RIBEIRÃO PRETO

Metabolomics in plant taxonomy: The Arnica model

Madeleine Ernst

Corrected version of the master's thesis presented to the Post-Graduate Program in Pharmaceutical Sciences on 23/08/2013. The original version is available at the Faculty of Pharmaceutical Sciences of Ribeirão Preto/USP.

Ribeirão Preto

2013

RESUMO

ERNST, M. **Metabolômica como ferramenta em taxonomia: O modelo em Arnica** . 2013. 250f. Dissertação (Mestrado). Faculdade de Ciências Farmacêuticas de Ribeirão Preto - Universidade de São Paulo, Ribeirão Preto, 2013.

Taxonomia vegetal é a ciência que trata da descrição, identificação, nomenclatura e classificação de plantas. O desenvolvimento de novas técnicas que podem ser aplicadas nesta área de conhecimento é essencial para dar suporte às decisões relacionadas a conservação de *hotspots* de biodiversidade. Nesta dissertação de mestrado foi desenvolvido um protocolo de *metabolic fingerprinting* utilizando MALDI-MS (*matrix-assisted laser desorption/ionisation mass spectrometry*) e subsequente análise multivariada utilizando *scripts* desenvolvidos para o pacote estatístico R. Foram classificadas, com base nos seus metabólitos detectados, 24 plantas de diferentes famílias vegetais, sendo todas elas coletadas em áreas da Savana Brasileira (Cerrado), que foi considerada um *hotspot* de biodiversidade. *Metabolic fingerprinting* compreende uma parte da Metabolômica, i.e., a ciência que objetiva analisar todos os metabólitos de um dado sistema (celula, tecído ou organismo) em uma dada condição. Comparada com outros métodos de estudo do metaboloma MALDI-MS apresenta a vantagem do rápido tempo de análise. A complexidade e importância da correta classificação taxonômica é ilustrada no exemplo do gênero *Lychnophora*, o qual teve diversas espécies incluídas neste estudo. No Brasil espécies deste gênero são popularmente conhecidas como "arnica da serra" ou "falsa arnica". Os resultados obtidos apontam similaridades entre a classificação proposta e a classificação taxonômica atual. No entanto ainda existe um longo caminho para que a técnica de *metabolic fingerprinting* possa ser utilizada como um procedimento padrão em taxonomia. Foram estudados e discutidos diversos fatores que afetaram os resultados como o preparo da amostra, as condições de análise por MALDI-MS e a análise de dados, os quais podem guiar futuros estudos nesta área de pesquisa.

Palavras-chave: metabolic fingerprinting, taxonomia vegetal, MALDI-MS, análise multivariada

ABSTRACT

ERNST, M. **Metabolomics in plant taxonomy: The Arnica model**. 2013. 250p. Master's thesis. Faculty of Pharmaceutical Sciences of Ribeirão Preto - University of São Paulo, Ribeirão Preto, 2013.

Plant taxonomy is the science of description, identification, nomenclature and classification of plants. The development of new techniques that can be applied in this field of research are essential in order to assist informed and efficient decision-making about conservation of biodiversity hotspots. In this master's thesis a protocol for metabolic fingerprinting by matrix-assisted laser desorption/ionisation mass spectrometry (MALDI-MS) with subsequent multivariate data analysis by in-house algorithms in the R environment for the classification of 24 plant species from closely as well as from distantly related families and tribes was developed. Metabolic fingerprinting forms part of metabolomics, a research field, which aims to analyse all metabolites, i.e., the metabolome in a given system (cell, tissue, or organism) under a given set of conditions. Compared to other metabolomics techniques MALDI-MS shows potential advantages, mainly due to its rapid data acquisition. All analysed species were collected in areas of the Brazilian Savanna (*Cerrado*), which was classified as "hotspot for conservation priority". The complexity and importance of correct taxonomic classification is illustrated on the example of the genus *Lychnophora*, of which several species also have been included into analysis. In Brazil species of this genus are popularly known as "arnica da serra" or "falsa arnica". Similarities to taxonomic classification could be obtained by the proposed protocol and data analysis. However there is still a long way to go in making metabolic fingerprinting by MALDI-MS a standard procedure in taxonomic research. Several difficulties that are inherent to sample preparation, analysis of plant's metabolomes by MALDI-MS as well as data analysis are highlighted in this study and might serve as a basis for further research.

Keywords: metabolic fingerprinting, plant taxonomy, MALDI-MS, multivariate data analysis

1 INTRODUCTION

Plant taxonomy is the science of description, identification, nomenclature and classification of plants. It is a research field, which suffered from a decline in expertise during the last few years, and is therefore recognized as a "science in crisis" by various researchers (COLLEVATTI, 2011; HOPKINS and FRECKLETON, 2002; WHEELER *et al.*, 2004). Plant taxonomy plays an important role in conservation of biodiversity, as there is a need for exact characterization of distribution of species as well as localization of areas of high species richness in order to make efficient and informed decisions about conservation (HOPKINS and FRECKLETON, 2002; WHEELER *et al.*, 2004). It is therefore essential to develop new taxonomic tools, in order to assist in a better understanding of taxonomic relationships between plant species.

In the present master's thesis metabolic fingerprints of 24 plant species belonging to four different tribes, three subfamilies, two families and two orders were acquired by matrix-assisted laser desorption/ionisation mass spectrometry (MALDI-MS) and were subsequently classified by multivariate data analysis by in-house algorithms in the R environment. All analysed plants were collected in areas of the Brazilian Savanna (*Cerrado*), which was classified as "hotspot for conservation priority" by Myers and collaborators (2000), an area, which features exceptional concentrations of endemic species and experiences an exceptional loss of habitat. The analysed samples included species from the tribes Vernonieae, Eupatorieae, Heliantheae and Microlicieae. Out of these, 12 belong to the genus *Lychnophora*, a genus which showed serious problems with correct botanical identification. In Brazil species of this genus are popularly known as "arnica da serra" or "falsa arnica" and are used in folk medicine as analgesic and anti-inflammatory agents (BASTOS *et al.*, 1987; GUZZO *et al.*, 2008; KELES *et al.*, 2010; SEMIR *et al.*, 2011).

Metabolic fingerprinting is a term that emerged in the field of metabolomics. By definition after Rochfort (2005) metabolomics is the measurement of all metabolites; i.e., the metabolome, in a given system (cell, tissue, or organism) under a given set of conditions. Corresponding definitions are also given in other sources (FIEHN, 2001; GOODACRE *et al.*, 2004; KOPKA *et al.*, 2004; TOMITA, 2005). It is still a recent field that has developed over the last ten years and which has found application in a variety of research areas (GOODACRE, 2005;

MORITZ and JOHANSSON, 2008; ROCHFORT, 2005; ROESSNER *et al.*, 2001; TOMITA, 2005; VERPOORTE *et al.*, 2008; VILLAS-BÔAS *et al.*, 2007). An international Metabolomics society was established in 2004, which now publishes its own journal "Metabolomics", thus demonstrating the growing interest in and the potential importance of this new research field (GOODACRE, 2005).

Together with other "omic" fields, such as transcriptomics, proteomics, and genomics, the potential of metabolomics has mainly been regarded as a functional genomics tool that should serve for a better understanding of systems biology (ROCHFORT, 2005; TOMITA, 2005; VILLAS-BÔAS *et al.*, 2007). Examples of the application of metabolomics in the determination of gene function are given in the works of Allen and coworkers (2003) and Raamsdonk and coworkers (2001). However, metabolomics has gone far beyond that and is useful whenever changes in metabolite levels are of interest (SHULAEV, 2006). Nowadays, metabolomics is applied in a variety of research fields, in medicinal research such as drug toxicity (GRIFFIN and BOLLARD, 2004; LINDON *et al.*, 2004; LINDON *et al.*, 2003; NICHOLSON and WILSON, 2003; NICHOLSON *et al.*, 2002; NICHOLSON *et al.*, 1999; ROBERTSON *et al.*, 2005), drug discovery (KELL, 2006; WATKINS and GERMAN, 2002b), disease diagnosis (BRINDLE *et al.*, 2002; KUHARA, 2005; MOOLENAAR *et al.*, 2003; OOSTENDORP *et al.*, 2006; WISHART *et al.*, 2001), or research into diseases like cancer (GRIFFITHS and STUBBS, 2003) or diabetes (WATKINS *et al.*, 2002c), in nutrition and nutritional genomics (GERMAN *et al.*, 2003; GERMAN *et al.*, 2002; GIBNEY *et al.*, 2005; TRUJILLO *et al.*, 2006; WATKINS *et al.*, 2001, WATKINS and GERMAN, 2002a), in natural product discovery (FIEHN *et al.*, 2000), in reserach of bacteria, fungi and yeast (KANG *et al.*, 2011, POPE *et al.*, 2007; TWEEDDALE *et al.*, 1998; MARTINS *et al.*, 2004), and diverse applications in plant sciences (TAYLOR *et al.*, 2002; RIZHSKY *et al.*, 2004; CATCHPOLE *et al.*, 2005). The list of examples has been collected and replenished from Brown and collaborators (2005), Shulaev (2006) and Wishart (2007).

In particular, plant metabolomics has a potentially broad field of applications (HALL, 2006). A wide range of applications for plant metabolomics can be found in literature. In a recent review on plant metabolomics, Wolfender and collaborators (2013) subdivide plant metabolomic studies into seven subareas: (i) fingerprinting of species, genotypes or ecotypes for taxonomic, or biochemical (gene discovery) purposes, (ii) comparing and contrasting the metabolite content of mutant or transgenic plants with that of their wild-type counterparts, (iii) monitoring the behaviour of specific classes of metabolites in relation to applied exogenous chemical and/or physical stimuli, (iv) interaction of plants with the environment or herbivores/pathogens, (v) studying developmental processes, such as the establishment of symbiotic associations, fruit ripening, or germination, (vi) quality control of medicinal herbs and phy-

topharmaceuticals, and (vii) determining the activity of medicinal plants and health-affecting compounds in food.

Compared to animals, plants contain a remarkable wide variety of metabolites. Indeed, the total number of metabolites present in the plant kingdom is estimated at 200,000 or more (OKSMAN-CALDENTEY and SAITO, 2005). The large variety of plant metabolites has always received special attention in various research fields (WINK *et al.*, 2010). In plant systematics; i.e., the biological classification of plants, secondary metabolites have been used as taxonomic markers for nearly 200 years (VILLAS-BÔAS *et al.*, 2007; WINK *et al.*, 2010). Targeted analysis of plant metabolites for various purposes dates back as far as the analysis of essential oils, which has been performed since the introduction of gas chromatography in the early 1950s (ROCHFORT, 2005; RYAN, D. and ROBARDS, 2006; VERPOORTE *et al.*, 2008). As the field of chemotaxonomy emerged in the 1960s, analysis of plant secondary metabolites became a common attempt for the taxonomic classification of plants. Moreover, different analytical methods for the comparison of chemical data in plants and for the establishment of a "degree of similarity" by computational data analysis became a common technique in the 1970s (WINK *et al.*, 2010). Hence, neither the analysis of metabolites itself nor the use of metabolites as taxonomic markers in plant systematics is new (VERPOORTE *et al.*, 2008; WINK *et al.*, 2010). What is it then that distinguishes plant metabolomics from traditional metabolite analysis? Whereas traditional metabolite analysis has focused on a small number of targeted analytes that have been preselected by the researcher according to their assumed importance or due to technical limitations of the experiment, metabolomics seeks to simultaneously measure all the metabolites; i.e., the metabolome (GOODACRE *et al.*, 2004; KOPKA *et al.*, 2004; TOMITA and NISHIOKA, 2005). This enables the visualization of the changes in plant metabolism caused either by environmental, genetic, or developmental alterations (TRYGG *et al.* 2006).

In the special case of plant taxonomy, metabolomics may be able to circumvent several problems inherent to traditional chemotaxonomic investigations of plant metabolites. Chemotaxonomic investigations focus mainly on the analysis and comparison of the presence, absence, or amount of one group of secondary metabolites (WINK *et al.*, 2010). Sesquiterpene lactones for example are used as taxonomic characters for differentiating species of the Asteraceae family (SEAMAN, 1982). Metabolomics however provides a picture of the metabolome as a whole, and therefore is a more holistic approach. Furthermore, several secondary metabolites assessed by chemotaxonomy have turned out to be useless as taxonomic markers, since they are distributed over various unrelated plant families and might have developed as a convergent trait (WINK *et al.* 2010).

Compared with more recently developed DNA sequence-based taxonomic methods, metabolomics may also offer advantages. The metabolome is further down the line from gene

to function. Therefore, it reflects the activities of the cell at a functional level more closely. Changes in the metabolome are thus expected to be amplified relative to alterations in the genome, transcriptome, and proteome (GOODACRE *et al.*, 2004; POPE *et al.*, 2007). DNA barcoding for example, which was introduced by Herbert and collaborators (2003) and aimed to identify animal species based on a short DNA sequence could not be applied in most plant species, as the gene *coI* used in DNA barcoding evolves relatively slowly in plants and due to the instability of the structure of the mitochondrial genome. Finding an alternative gene sequence appropriate for DNA barcoding further showed to be problematic and is still being discussed (COLLEVATTI, 2011).

Successful metabolomics approaches for chemotaxonomic purposes in plants belonging to the same genus by ^{1}H-NMR, LC-MS and GC-MS have already been published (FARAG *et al.*, 2012; GAO *et al.*, 2012; GEORGIEV *et al.*, 2011; KIM *et al.*, 2010, XIANG *et al.*, 2011), however, at present there neither exist studies applying metabolomics techniques for the classification of species from different genera nor do there exist studies applying MALDI-MS for plant taxonomic purposes.

MALDI-MS has originally been developed for the analysis of high molecular weight compounds and was only recently introduced to the analysis of low molecular weight compounds (WANG *et al.*, 2011). At the current state of art of MALDI-MS, compared to other more traditional analytical methods applied in plant metabolomics research mainly the rapid acquisition time, which is of central importance in metabolomics experiments may be named.

In the following an introduction to plant metabolomic research (Section 1.1), MALDI-MS (Section 1.2), the genera of all analysed plants (Section 1.3) and to metabolomics data analysis (Section1.4) is given. The complexity of taxonomic classification by traditional methods is illustrated on the example of the genus *Lychnophora* in Section 1.3.1. Chapter 3 introduces materials and methods that were used throughout all experiments. Results, including more specific descriptions of material and methods are then presented in Chapters 4 to 9. Chapter 8 forms the main part, where statistical data analysis and obtained classification results are discussed. Finally, Chapter 9 gives an outlook on realized experiments that could guide future reserach. Some definitions of terms used throughout the text are given in Table 1. For simplicity, the terms plant systematics and plant taxonomy will be used interchangeably.

Table 1: Definitions.

Term	Explanation	Ref.
Chemotaxonomy	Uses chemical features of plants for taxonomical classification or to solve other taxonomic problems. Synonyms are Chemosystematics, Chemical Taxonomy, Chemical Plant Taxonomy, or Plant Chemotaxonomy.	5
Functional Genomics	Strategies developed for a better understanding of the correlation between genes and the functional phenotype of an organism.	1
Phylogenetic	Related through common ancestry Phylon (gr.) = tribe, family, genus; genesis (gr.) = genesis, origin, development.	2, 8
Plant Taxonomy	Science of description, identification, nomenclature, and classification of organisms. It is a synthetic science that uses data from diverse research fields such as morphology, anatomy, cytology, genetics, chemistry, and molecular biology. The term is used as a synonym of "plant systematics", other sources define taxonomy and systematics as overlapping but separate areas.	3, 6, 7
Plant Systematics	Includes plant taxonomy, but has as primary aim the identification of relationships based on phylogeny.	3, 6
Systems biology	Techniques, such as the "omic" sciences and mathematical modelling, which are applied in order to understand systems "as a whole". Data that are derived from genes, proteins, or metabolites shall be combined, in order to obtain a more holistic view of the state of the "system". Accordingly, a more fundamental understanding of biology shall be achieved.	4, 9
Taxa	Term that describes all the categories in the plant kingdom. Classes, orders, tribes, and species are all taxa.	3

Rf: *1* BINO *et al.*, 2004; *2* DE QUEIROZ and GAUTHIER, 1990; *3* FROHNE and JENSEN, 1998; *4* GOODACRE *et al.*, 2004; *5* SHARMA, 2007; *6* SIMPSON, 2006; *7* STUESSY, 2009; *8* WIESEMÜLLER *et al.*, 2003; *9* WILSON and NICHOLSON, 2008.

1.1 Introduction to metabolomics

1.1.1 Classification of metabolomic approaches

The quantification of all the metabolites in a given system is at present impossible, due to the lack of simple automated and sufficiently sensitive analytical strategies (DETTMER *et al.*, 2007; GOODACRE *et al.* 2004; KOPKA *et al.* 2004; SUMNER *et al.*; 2003; TOMITA, 2005). In order to distinguish different approaches of metabolite analysis, several terms have been introduced. Because metabolomics is a developing research field, terminologies are still evolving (GOODACRE *et al.*, 2004; OLDIGES *et al.*, 2007) and different sources consider slightly different definitions. In Table 2 some of the most common definitions are listed.

Table 2: Classification of metabolomic approaches.

Term	Explanation	Ref.
Metabolic fingerprinting	Rapid high throughput screening of all detectable analytes in a sample. Detected metabolites are not necessarily identified. The aim is to find large differences between samples by comparing spectra with unidentified peaks, and classify them on the basis of multivariate data analysis. Prominent metabolites, which define the sample classes, are eventually identified in a further step. Sample preparation, separation, and detection should be as fast and as simple as possible. Sometimes it is performed before metabolite profiling, in order to guide a research project. In Fiehn (2002) this method is mentioned specifically in relation to analytical techniques that do not use previous chromatographic separation, such as mass spectrometry, infrared spectroscopy, or nuclear magnetic resonance spectroscopy.	2, 3, 6, 8, 14
Metabolite profiling/ Metabolic profiling	Analysis is restricted to the identification and quantification of a number of predefined metabolites. These metabolites may all be associated with the same pathway, or belong to the same class of compounds. Similarly, the analytical procedure may be matched to the specific chemical properties of these substances. Computational methods are then employed for transformation of the spectra into lists of metabolites and their concentrations. It is not only a qualitative analysis like metabolite fingerprinting, but it also quantitatively analyses metabolites. Ideally, results should be independent of the technology applied for the measurement of metabolite concentrations.	2, 3, 4, 14
Metabolite target analysis	Analysis restricted to metabolites that have been selected (targeted) before, for example, by optimized extraction or specific separation and/or detection. Metabolite target analysis may follow broad-scale metabolomic analysis or be based upon previous knowledge.	4, 6

Continued on next page

Table 2 – *Continued from previous page*

Term	Explanation	Ref.
Metabolome	The term metabolome was first introduced by Oliver and coworkers in 1998. At present, according to Goodacre and coworkers (2004) there is a disagreement on the exact definition of the metabolome in the research community: Hall (2006) and Pope and coworkers (2007) define it as all the small molecules present in living things, excluding larger molecules (> 1500 Da) and typical amino acids and sugar polymers out of reasons of practicality in terms of extraction and detection. Hollywood and collaborators (2006) add that the metabolome comprises all the measurable metabolites under a particular physiological or developmental state of the cell, and that these have a typical mass-to-charge ratio < 3000 *m/z*. Another source, however, defines the metabolome as including only the small molecules that participate in general metabolic reactions and are essential for the normal living functions of a cell.	1, 5, 6, 7, 11, 10
Metabolomics	Metabolomics is defined as a comprehensive, qualitative, and quantitative analysis of all the metabolites in a biological system. Last and coworkers (2007) add that, in order to identify and measure as many metabolites as possible (not all, due to the lack of appropriate detection methods), high-throughput analytical strategies must be applied. It remains unclear whether metabolomics, by definition, comprises the quantification and identification of all the metabolites or of as many metabolites as possible.	2, 3, 4, 6, 7, 9, 12, 13, 15

Rf: *1* BEECHER, 2003; *2* DETTMER *et al.*, 2007; *3* FIEHN, 2002; *4* FIEHN, 2001; *5* GOODACRE *et al.*, 2004; *6* HALL, 2006; *7* HOLLYWOOD *et al.*, 2006; *8* KOPKA *et al.*, 2004; *9* LAST *et al.* 2007; *10* OLIVER *et al.*, 1998; *11* POPE *et al.*, 2007; *12* TOMITA, 2005; *13* VERPOORTE *et al.*, 2007; *14* VIANT *et al.*, 2008; *15* VILLAS-BÔAS, 2007.

1.1.2 Sampling and extraction of metabolites

Sampling of the metabolites is a critical step in every metabolomics experiment and has to be treated with special care. The first difficulties already arise with the harvesting of the plants (ROESSNER and PETTOLINO, 2007; VERPOORTE *et al.*, 2008). There are several factors, that have an influence on a plant's metabolome. According to Villas-Bôas and coworkers (2007) the main sources of variability during the sampling of plant material is light. The reason for this is photosynthesis, which makes the intensity of metabolic processes depend heavily upon the availability of light (ROESSNER and PETTOLINO, 2007). In addition, the wavelength of the light also influences the metabolite profiles. The upper leaves may have different metabolite profiles than the lower leaves of the same plant, because light does not reach each leaf to the same extent (VILLAS-BÔAS *et al.*, 2007). This has not only an influence on leaves but also on subterranean parts of the plant (ROESSNER and PETTOLINO, 2007). A further factor that has an influence on a plant's metabolome is the time of harvesting. Due to diurnal changes a plants metabolome differs depending on the time of the day (MORITZ and JOHANSSON, 2008; ROESSNER and PETTOLINO, 2007; VERPOORTE *et al.*, 2008; VILLAS-BÔAS *et al.*, 2007).

Further factors according to Villas-Bôas and collaborators (2007) are the atmospheric O_2/CO_2 ratio during sampling and nutrient/substrate supply. Moreover also the stage of plant development at the time of harvesting affects the metabolite profile (ROESSNER and PETTOLINO, 2007; VERPOORTE *et al.*, 2008; VILLAS-BÔAS *et al.*, 2007). A series of samples harvested at different times of the day and in distinct stages of development should therefore ideally be measured, in order to determine the biological variation and set standard conditions for the experiments (VERPOORTE *et al.*, 2008). Another possible solution for the minimization of variability is to harvest all plants under the same light intensity (the same period of day/night) within a very small timeframe and select leaves or other parts of the plant that are under similar light-exposure. Of course this method is only applicable if only a small amount of plant material is needed (ROESSNER, 2007; VILLAS-BÔAS *et al.*, 2007). To illustrate the large influence of biological variance, one might consider unpublished data from Sumner and coworkers (2003), which suggest an average biological variance of 50% for *Medicago truncatula*.

Harvesting also means stressing and wounding the plant, which causes alterations in the metabolome, as well (VERPOORTE *et al.*, 2008; VILLAS-BÔAS *et al.*, 2007). Because changes in the plant metabolism take place within seconds up to a few minutes, harvesting must be performed quickly, and metabolism must be stopped right after the harvesting procedure (VERPOORTE *et al.*, 2008). The most frequently used and, at present best method for stop-

ping metabolism after harvesting, according to Moritz and Johansson (2008) is to freeze the plant material in liquid nitrogen. However, as the cells are destroyed during freezing, thawing may induce all kinds of biochemical conversions. Enzyme activity must be halted either by extraction of the frozen material by means of a denaturing solvents or by brief treatment with microwave (VERPOORTE *et al.*, 2008). It should be borne in mind that modifications in the metabolite profiles will nevertheless still occur during the short period between sampling of the plant material and its placement into liquid nitrogen (MORITZ and JOHANSSON, 2008).

It should further be noted that each plant organ, tissue, or cell type contains different, characteristic metabolites because of different external stimuli. The analytical techniques that are currently employed in metabolomics still lack sensitivity and therefore, many different cell types and tissues have to be extracted together, so that sufficient levels of all the metabolites are simultaneously obtained. Hence, the results of metabolomic studies in plants only show an average of the metabolite content distributed over different plant organs and tissues. In this context, research is also being conducted on single cell metabolite analysis (ROESSNER, 2007; ROESSNER and PETTOLINO, 2007). A first successful attempt of single cell metabolomics in *Arabidopsis thaliana* has been presented by Schad and collaborators (2005).

For the same reasons that sampling needs to be rapid, extraction of the plant material must also occur within a short timeframe (MORITZ and JOHANSSON, 2008). Extraction methods for metabolomic experiments should be as simple and fast as possible (MORGENTHAL *et al.*, 2007). Common are solvent extraction, steam distillation, and supercritical fluid extraction or the use of ionic liquids. According to Verpoorte and collaborators (2008), solvent extraction is applied most frequently in metabolomics experiments. Steam distillation is utilized for volatile compounds and supercritical fluid extraction and the use of ionic liquids is still not very common, since there is little experience of applying these extraction methods in metabolomic high-throughput analytical techniques (VERPOORTE *et al.*, 2008). Degradation, modification, and loss of metabolites during the extraction must be minimized (MORITZ and JOHANSSON, 2008; VERPOORTE *et al.*, 2008; VILLAS-BÔAS *et al.*, 2007). However, to date, no method for the extraction of all the metabolites without artefact formation or degradation has been reported (MORITZ and JOHANSSON, 2008; VERPOORTE *et al.*, 2008). Because metabolites in plant tissue are highly diverse and may contain non-polar compounds (terpenoids and fatty acids from cell membranes), compounds of medium polarity (most of the secondary metabolites), and polar compounds (most of the primary metabolites such as sugars and amino acids), no solvent is able to extract all the compounds at the same time (VERPOORTE *et al.*, 2008). In order to obtain good reproducibility of a certain class of metabolites, others have to be sacrificed (VILLAS-BÔAS *et al.*, 2007). Common solvent extraction methods usually extract medium-polar or polar compounds. A sub-metabolomic approach focused on the

analysis of non-polar metabolites, namely lipidomics, has been developed for the extraction of non-polar metabolites. Several considerations have to be made previous to the selection of a solvent. The choice of solvent is extremely important for the achievement of reliable results, since it needs to be adequate for the metabolites targeted for extraction as well as the analytical method. Therefore, the selectivity and polarity of the solvent, its boiling point (in case solvents have to be evaporated), toxicity and environmental considerations, interference with the analytical procedure, and possible contaminants must be taken into account during the selection (VERPOORTE *et al.*, 2008). The most common way of extracting metabolites is to shake the previously homogenized plant tissue at high or low temperatures in either a pure organic solvent, in the case of non-polar compounds, or a mixture of solvents, for more polar compounds. Solvents employed for the extraction of polar metabolites are methanol, ethanol, and water, while chloroform is most often used for lipophilic compounds. Methanol/water/chloroform 1:3:1 is a common mixture for the extraction of compounds of medium polarity (MORITZ and JOHANSSON, 2008). More detailed information on extraction methods for metabolomics can be found in Villas-Bôas and coworkers (2007). A possible standard method for the optimization of extraction methods by design of experiments for metabolomic studies has been proposed by Gullberg and collaborators (2004).

Well-known extraction methods still are most suitable for targeted analysis, whereas an ideal extraction method for research in metabolomics has yet to be developed (VILLAS-BÔAS *et al.*, 2007). Because a completely non-compound specific sample preparation as the one required for a metabolomic experiment is impossible, focus should therefore be placed mainly on the reproducibility of the sample processing protocol. In order to compare different samples, their preparation must be identical, although some metabolites might be excluded (OLDIGES *et al.*, 2007).

1.1.3 Analytical methods used in plant metabolomics

Different analytical methods are used for the analysis of plant's metabolomes. The most widely employed methods are mass spectrometry coupled to gas or liquid chromatography (GC-MS and LC-MS) and nuclear magnetic resonance (NMR), but also other methods such as capillary electrophoresis mass spectrometry (CE-MS), high-performance liquid chromatography with photodiode array detection (HPLC-PDA), thin layer chromatography with UV detection (TLC-UV) and Fourier transform-ion cyclotron resonance mass spectrometry (FT-ICR-MS) have been described (HAGEL and FACCHINI, 2008; MORITZ and JOHANSSON, 2008). Each method has its own advantages and disadvantages, an appropriate combination of analytical tools with respect to the analysed plant material is therefore essential (MORITZ and

JOHANSSON, 2008).

In the following, the most commonly cited analytical methods for plant metabolomic studies shall be briefly introduced, and the advantages and disadvantages of each technique will be outlined. Bino and collaborators (2004) state that a combination of LC, NMR, and MS systems might be preferably applied in the future, because it increases the number of quantifiable and identifiable metabolites, but will probably only be limited to few laboratories due to the high costs. Since all techniques used in metabolomics research listed here are well-established and common analytical methods, they shall only be briefly discussed.

Hyphenated mass spectrometry methods: GC/LC-MS

The techniques that are most often utilized are the hyphenated mass spectrometry methods, so-called separation-based methods coupled with mass spectrometry. In these methods, the metabolites are separated via gas chromatography, liquid chromatography, or capillary electrophoresis (CE) prior to mass spectrometric analysis. GC and LC or HPLC (high-performance liquid chromatography) separate compounds due to different interactions of the substances with the stationary phase, while CE separation is based on the size-to-charge ratio of the ionic molecules (HAGEL and FACCHINI, 2008). CE-MS will not be further discussed at this point. A CE-MS approach and the advantages of using it in metabolomics studies of bacteria are given in Soga (2007), for instance. The separation technique is chosen on the basis of the type of molecules present in the target sample. GC is applicable in the case of hydrophobic, low molecular weight compounds such as essential oils, hydrocarbons, esters, and metabolite derivatives with reduced polarity (HAGEL and FACCHINI, 2008). In order to perform a GC analysis the analytes must be heat-stable and volatile. Otherwise they have to be derivatized (HAGEL and FACCHINI, 2008; MORITZ and JOHANSSON, 2008).

Different ionisation methods and mass detectors can be applied in mass spectrometry, depending on the kind of molecules targeted for analysis. The ionisation method that is usually utilized for GC is electron impact (EI) ionisation. EI belongs to the hard ionisation methods. It transfers an excess amount of energy, which causes strong molecular fragmentation (KOPKA *et al.*, 2004). Soft ionisation technologies are more commonly employed when MS is coupled with LC. Due to their lower energy transfer, less molecular fragmentation is induced, so that fewer molecules are ionised (KOPKA *et al.*, 2004; VERPOORTE *et al.*, 2008). Atmospheric pressure chemical ionisation MS coupled to LC (APCI-LC-MS), can be applied in the case of polar metabolites with low to moderate molecular weight (i.e., MW$<$2000), such as sterols; fatty, organic, and amino acids; metal ions; and alkaloids. Electrospray ionisation LC-MS, on the other hand, is used for polar molecules, such as proteins, peptides, polysaccharides, or DNA, within a broad molecular weight range (i.e. MW 10 - 300,000) (HAGEL and FACCHINI, 2008;

KOPKA *et al.*, 2004; VERPOORTE *et al.*, 2008).

In order to obtain better peak resolutions, tandem mass spectrometry (MS/MS) or MS^n can be applied instead of simple MS (VILLAS-BÔAS *et al.*, 2007). In tandem mass spectrometry two mass spectrometers are coupled to each other. The first spectrometer chooses ions of a specific mass, whereas the second mass spectrometer causes further decay to the ions produced before (MCLAFFERTY, 1981). In this way, these methods allow the identification of individual metabolites of a specific *m/z* value (TAGUCHI, 2005). A previous chromatographic separation can even be left out in order to differentiate between isomers of two samples when various mass spectrometers (MS^n) are combined (MCLAFFERTY, 1981), given that the fragmentation patterns of the isomers are different.

The combination of many different ionisation methods and mass-detection approaches has given rise to several types of mass spectrometers. Features of each mass analyser are rather different; therefore, it is essential to choose the mass spectrometer that best suits the specific research requests (TAGUCHI, 2005). For major advantages and disadvantages of employing hyphenated mass spectrometry methods in metabolomics see Table 3.

Table 3: Advantages and disadvantages of hyphenated mass spectrometry methods.

Hyphenated MS methods			
Advantages	Ref.	Disadvantages	Ref.
High sensitivity (detection limit at 10^{-12}-10^{-15} moles)	HAGEL and FACCHINI, 2008; SUMNER *et al.*, 2003; VERPOORTE *et al.*, 2008	Time-consuming	HAGEL and FACCHINI, 2008
Relative low cost	HAGEL and FACCHINI, 2008	No absolute quantitation of metabolites possible (acquisition of calibration curves of all individual, often unidentified compounds is unrealistic)	VERPOORTE *et al.*, 2008
Good availability	HAGEL and FACCHINI, 2008		
Enhanced metabolite identification: Information about the chemical nature (by separation technique) and mass (by MS) of the compound	HAGEL and FACCHINI, 2008		
Good relative quantitation	VERPOORTE *et al.*, 2008		
GC-MS			

Continued on next page

Table 3 – *Continued from previous page*

Advantages	Ref.	Disadvantages	Ref.
Analysis of low molecular weight, hydrophobic compounds (essential oils, hydrocarbons, esters)	HAGEL and FACCHINI, 2008	Only suitable for low molecular weight, hydrophobic compounds; derivatization may mask the result	VERPOORTE *et al.*, 2008
Direct measurement of volatile compounds	HAGEL and FACCHINI, 2008	Derivatization needed depending on the type of molecules that one wishes to analyse	HAGEL and FACCHINI, 2008
Relative predictable ionisation behaviour of neutral compounds and thus available databases (Golm Metabolome Database, `http://csbdb.mpimp-golm.mpg.de` and Fiehn Laboratory Database, `http://fiehnlab.ucdavis.edu/`)	HAGEL and FACCHINI, 2008		
Suited for analysis of mixtures as almost all non-polar compounds do ionise in positive ion mode	HAGEL and FACCHINI, 2008		
LC-MS			
Analysis of polar compounds with low, moderate or high molecular weight (MW 10-300,000) depending on the ionisation method applied	FRASER *et al.* 2007; HAGEL and FACCHINI, 2008; ROESSNER, 2007	Low reproducibility of fragmentation patterns, which consequently makes the elaboration of databases more problematic	HAGEL and FACCHINI, 2008

Continued on next page

Table 3 – *Continued from previous page*

Advantages	Ref.	Disadvantages	Ref.
No derivatization required	ROESSNER, 2007	More suited for the targeted profiling of molecules with similar ionisation behavior as not all compounds in a plant extract do ionise under same conditions	HAGEL and FACCHINI, 2008
Analysis of thermo-unstable compounds	ROESSNER, 2007		

Direct injection MS (DIMS)

Direct injection MS describes MS analyses that are done by direct injection of a sample into the ionisation source of a mass spectrometer. Usually atmospheric pressure ionisation techniques are applied in DIMS, most commonly ESI (DETTMER *et al.*, 2007). The soft ionisation technologies are often utilized in metabolomics, because they give rise to less fragmentation than hard ionisation technologies. This results in a smaller number of signals in the spectra, which in turn are less complex (HAGEL and FACCHINI, 2008).

A large variety of mass analysers have been applied in DIMS analyses in metabolomics research such as single-stage quadrupole, triple quadrupole, Orbitrap, time of flight (TOF) and Fourier transform ion cyclotron mass spectrometers (FT-ICR-MS) (DETTMER *et al.*, 2007; WOLFENDER *et al.*, 2013). According to Dettmer and collaborators (2007) and Wolfender and collaborators (2013), high-resolution mass spectrometers such as TOF and FT-ICR should be preferably used, in order to be able to distinguish between isobars (compounds with the same nominal mass). FT-ICR-MS shows a very high resolution and mass accuracy and is therefore described as method of choice for DIMS in most plant metabolomics reviews (HAGEL and FACCHINI, 2008; HALL, 2006; KOPKA *et al.*, 2004; WOLFENDER *et al.*, 2013). However, since its first application in plant metabolomics by Aharoni and coworkers (2002), not many studies have been reported mainly due to the high instrument costs (DETTMER *et al.*, 2007; HAGEL and FACCHINI, 2008).

A main disadvantage of direct infusion FT-ICR-MS is that isomers cannot be distinguished and less chemical information than in hyphenated MS methods is obtained (HAGEL and FACCHINI, 2008). In order to distinguish between isomers, MS^n analyses can be performed (HAGEL and FACCHINI, 2008). MS^n can avoid a previous chromatographic separation in order to differentiate between isomers of different samples, because enough information is collected from the obtained fragmentation patterns (MCLAFFERTY, 1981), given that the fragmentation patterns of the two isomers are not identical. And chemical information can be increased by performing analyses in negative as well as in positive ion mode (WOLFENDER *et al.*, 2013).

Mass accuracy, resolution, and detection limits of all mass spectrometric methods depend highly on the type of mass analyser and ionisation. Advantages and disadvantages inherent to direct infusion with FT-ICR-MS are summarized in Table 4. Specific advantages depend on the chosen instrument type.

Table 4: Advantages and disadvantages of DIMS by FT-ICR-MS.

DIMS by FT-ICR-MS			
Advantages	Ref.	Disadvantages	Ref.
High-throughput	HAGEL and FAC-CHINI, 2008	No distinction between chemical isomers	KOPKA *et al.*, 2004
Very sensitive (detection limit at 10^{-15}-10^{-18} moles)	DETTMER *et al.*, 2007	Less chemical information obtained than in hyphenated MS techniques	HAGEL and FAC-CHINI, 2008
High mass resolution	HAGEL and FAC-CHINI, 2008	High cost, limited instrument availability	HAGEL and FAC-CHINI, 2008
High mass accuracy	HAGEL and FAC-CHINI, 2008	Ion suppression if ESI is used as ionisation method	DETTMER *et al.*, 2007
Increase of sample reproducibility due to short analysis time	DETTMER *et al.*, 2007		
MS^n capabilities	DETTMER *et al.*, 2007		

NMR-based methods

Another analytical approach used in plant metabolomics is nuclear magnetic resonance (NMR). Krishnan and coworkers stated in 2005 that MS-based techniques were more suited for plant metabolomic experiments, but NMR could offer interesting advantages when employed in combination with MS. However, nowadays, NMR-based metabolomics, especially one-dimensional (1D) ^{1}H-NMR is a major analytical tool for many applications in plant metabolomics from quality control, to chemotaxonomy, to comparison of genetically modified plants, interaction with other organisms and many more (KIM *et al.*, 2011). A major drawback of 1D ^{1}H-NMR besides the low sensitivity is the signal overlap. To obtain better signal resolution 2D NMR spectroscopy, LC-NMR or LC-NMR-MS are also used (HAGEL and FACCHINI, 2008; KIM *et al.*, 2011). As a detailed description of NMR-based metabolomics would exceed the scope of this master's thesis, merely some pros and cons are given in Table 5.

Table 5: Advantages and disadvantages of 1D ^{1}H-NMR.

1D ^{1}H NMR			
Advantages	Ref.	Disadvantages	Ref.
Rapid and simple sample preparation	KIM *et al.*, 2011	Low sensitivity (number of metabolites that can be detected by NMR covers only about 10% of a plant's metabolome)	KIM *et al.*, 2011; KOPKA *et al.*, 2004; VIANT *et al.*, 2008
Signal intensity is only dependent on molar concentration, and absolute quantitation is therefore possible	KIM *et al.*, 2011; KOPKA *et al.*, 2004	Signal overlap	KIM *et al.*, 2011
Highly reproducible spectra (the physical characteristics of the compounds are measured, and the produced data are valid for ever if the same extraction method and NMR solvent are employed)	VERPOORTE *et al.*, 2008		
No derivatization is required	KRISHNAN *et al.*, 2005		
Noninvasive analysis, *in vivo* analyisis is possible and samples can still be used for the extraction of other cell products in a further analysis	KRISHNAN *et al.*, 2005; ROESSNER, 2007		
Most comprehensive structural information for elucidation of small organic compounds	SEGER and STURM, 2007		

1.2 Introduction to matrix-assisted laser desorption/ionisation mass spectrometry (MALDI-MS)

Lasers have been used in mass spectrometers since the early 1960s. Though, none of the existing laser ionisation methods were able to completely solve the problem of measuring thermally instable and difficult to ionise high molecular weight compounds (HILLENKAMP *et al.*, 1991). The main breakthrough in measuring high molecular weight compounds was only when Karas and coworkers (1987) discovered that laser desorption (LD) combined with the use of a matrix could solve the problem. With their newly developed technique Hillenkamp and Karas were able to measure proteins with molecular masses exceeding 10,000 Da (KARAS *et al.*, 1988). Later, Tanaka (1988) applied the same technique with a special matrix and was able to measure biomolecules having a molecular weight up to 100,000 Da. For this discovery he was awarded the Nobel Prize in Chemistry in 2002 (TANAKA, 2003).

The mechanisms of ionisation in MALDI-MS are not completely understood yet (EL-ANEED *et al.*, 2009; THOLEY and HEINZLE, 2006). The general idea though is that by dissolving the analyte in a solvent and mixing it with a solid matrix (compound that absorbs laser radiation) a matrix-analyte crystal results after evaporation of the solvents. The bombardment of the matrix-analyte crystal with a laser beam excites the matrix molecules and those transfer energy to the analytes leading to ionisation and desorption of the same (EL-ANEED *et al.*, 2009). Gates and collaborators (2006) summarize the ionisation process into three basic steps:

(i) *Formation of a 'solid solution'*. Formation of the matrix-analyte crystal after evaporation of the solvents.

(ii) *Matrix excitation*. Absorption of photons from the laser beam by the chromophore of the matrix substance, which causes rapid vibrational excitation and leads to disintegration of the solid solution. The formed clusters on the surface consist of analyte molecules, which are surrounded by matrix and salt ions. The matrix molecules evaporate way and as they do so transfer their charge to the analyte.

(iii) *Analyte ionisation*: Stabilisation of the photo-excited matrix molecules occurs via proton transfer to the analyte during which also cation attachment is encouraged, leading to the characteristic $[M+X]^+$ (X = H, Na, K etc.) analyte ions in positive ion mode. The described ionisation reactions occur in the desorbed matrix-analyte cloud close to the surface. From there analyte ions are extracted into the mass spectrometer for analysis.

More detailed reviews on ionisation mechanisms can be found in Zenobi and Knochenmuss (1998), Karas and collaborators (2000) and Knochenmuss and Zenobi (2003).

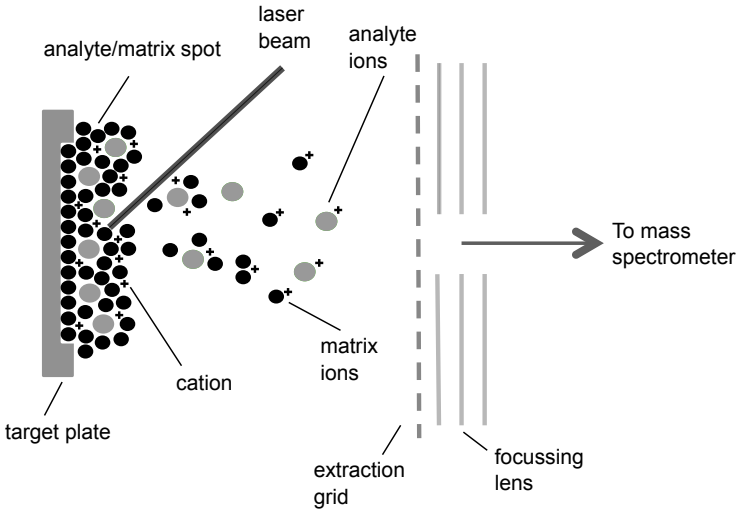


Figure 1: A schematic representation of the mechanism of matrix-assisted laser desorption/ionisation in positive ion mode. Redrawn after Gates and collaborators (2006).

During the last few years, despite its original purpose to analyse big, high molecular weight and non-volatile biopolymers, MALDI-MS has been introduced to the analysis of low molecular weight compounds (LMW) (WANG *et al.*, 2011). As advantages for using MALDI-MS instead of other mass spectrometric methods, that are undoubtedly as suitable, or even more suitable for the analysis of LMW compounds, the following advantages may be mentioned (COHEN *et al.*, 2002; THOLEY and HEINZLE, 2006; VAN KAMPEN *et al.*, 2011; YANG *et al.*, 2007):

- simple and rapid sample preparation
- very low sample consumption
- high-sensitivity
- high throughput
- samples can be stored directly on the target plate for a defined time interval (see Section 9.6)
- relative tolerance to impurities, salts and buffers
- less ion suppression effects in compound mixtures observed compared to electrospray ionisation mass spectrometry (ESI-MS)

For metabolomics studies, where a large number of samples are analysed, the extremely fast analysis time and consequently high-throughput may be named as the main advantages. Analysis of 50 *Arabidopsis* samples by liquid chromatography coupled to mass spectrometry (including spectra preprocessing) takes about 4 days (DE VOS *et al.*, 2007), while in MALDI-MS, assuming a total of 8 technical replicates per sample, a capacity of 384 sample spots per MALDI-plate and using any of the MALDI parameters that are described in this master's thesis, analysis would take 2 days (it was observed during own laboratory experience that after having acquired spectra of one whole target plate the MALDI equipment had to be cleaned. During this process the vacuum is lost and analyses were only continued on the next day in order to allow the vacuum pressures to stabilize). Acquisition of one single sample with 8 technical replicates in MALDI-MS takes one to five minutes, depending on the chosen acquisition parameters. Furthermore a previous sample cleanup of the plant extracts with hexane was not necessary in analysis by MALDI-MS as there is no source contamination as in ESI.

Disadvantages of MALDI-MS applied in metabolomic studies correspond to a major part to those mentioned for DIMS (see Section 1.1.3) including that no distinction between chemical isomers can be made, less chemical information is obtained than in hyphenated mass spectrometry methods and the high instrument cost and limited availability of the MALDI-MS equipment. In order to compensate for the lack of chemical information compared to hyphenated techniques, spectra of plants analysed in this master's thesis were acquired with two different matrix substances in positive ion mode as well as with one matrix substance in negative ion mode. It was assumed that in this way the probability of different types of compounds to be ionised would raise and a broader insight into the plants metabolome would be enabled.

A further bottle-neck in applying MALDI-MS for metabolomics studies is the analysis of LMW compounds. Selection of an appropriate matrix is relatively difficult since traditional matrix substances interfere with analyte ions and further MALDI-MS has a poor reproducibility of signal intensities (VAN KAMPEN *et al.*, 2011). As the process of ion formation in MALDI is still not fully understood, the choice of an appropriate matrix is mainly experimental (EL-ANEED *et al.*, 2009). The only requirements of a matrix substance is its absorption of laser light at the applied wavelength, solubility in a solvent in which also the analyte may be dissolved, inertness, vaccuum stability, and absence of overlap of matrix and analyte ions. Most common used matrices are small organic molecules that absorb laser light in the range of 266-355 nm, typically having hydroxyl- or amino groups in ortho- or para-position (-OH, -NH) and facultatively exhibit acidic groups or carbonyl functions (carboxyl group, amides, ketones). Some of the frequently used matrices are 2, 5-dihydroxybenzoic acid (2,5-DHB), which besides others is thought to be especially suited for the analysis of LMW compounds, and α-cyano-4-hydroxycinnamic acid (CCA), a matrix mostly used for peptide analysis (THOLEY

and HEINZLE, 2006).

Various approaches already have been proposed in order to improve analysis of low molecular weight compounds by MALDI-MS including desorption/ionisation on silicon (FINKEL *et al.*, 2005; GO *et al.*, 2005; WEI *et al.*, 1999); using inorganic compounds as matrix substances (such as porous alumninium, zinc oxide nanoparticles, carbon nanotube, graphite, graphene and graphene flakes) (DONG *et al.*, 2010; LANGLEY *et al.*, 2007; LU *et al.*, 2011; NAYAK *et al.*, 2007; WATANABE *et al.*, 2008; XU *et al.*, 2003), which were also classified as matrix-free approaches in Guo and coworkers (2002); high-mass matrix molecules as matrix substances (such as *meso*-tetrakis(pentafluorophenyl)porphyrin) (AYORINDE *et al.*, 1999); the addition of a surfactant to traditional crystalline matrices (2,5-DHB or CCA) (GUO *et al.*, 2002) or the optimization of the matrix suppression effect (MSE) by testing different laser intensities (McCOMBIE and KNOCHENMUSSS, 2004). The list of examples was adapted and replenished from Liu and collaborators (2012).

As traditional MALDI matrices often show big differences in intensity and resolution at different positions of the sample spot (also called hot/sweet spot formation), also several techniques have been proposed to improve sample homogeneity. One of those is the use of ionic liquid matrices (ILM) (THOLEY and HEINZLE, 2006). ILM are so called class II ionic liquids and were especially developed for MALDI-MS, as classical ionic liquids weren't suited as MALDI matrices. They are equimolar mixtures of crystalline MALDI matrices (e.g. CCA, 2,5-DHB) with organic bases (tributylamine, pyridine, or 1-methylimidazole) forming organic salts. ILM form viscous liquids (sometimes also solids), that have a highly homogeneous surface. Hot spot formation is therefore prevented, leading to more homogeneous samples and consequently higher reproducibility of intensities and resolution, compared to traditional crystalline matrices (THOLEY and HEINZLE, 2006). Besides traditional crystalline matrices, two forms of ILM were also tested in this master's thesis (see Section 9.4).

Due to the difficulties mentioned above, application of MALDI-MS in plant metabolomics studies up to this day has been very limited. Fraser and collaborators (2007) describe a targeted approach for carotenoids. And other studies propose new matrix substances for MALDI-MS with potential to be applied in metabolomics studies (SHROFF *et al.*, 2009; LIU *et al.*, 2012). Furthermore, the potential of MALDI-MS applied in plant metabolomics studies is also seen in MALDI imaging in order to access the spatial distribution of metabolites in a plant organ (WOLFENDER *et al.*, 2013; LEE *et al.*, 2012). MALDI-MS applied in metabolomics studies for plant taxonomy as described in this master's thesis however, has not been published yet.

1.3 Introduction to the genera of the analysed plants and their chemical constituents

In the following a brief introduction to the genus of each plant analysed, *Lychnophora*, *Vernonia*, *Ageratum*, *Bidens*, *Calea*, *Porphyllum*, *Lavoisiera* and *Microlicia* including their chemical constituents shall be given. The descriptions are far from being complete and merely shall serve as a brief overview. As the present study did only include leaves of the named plants, only chemical compounds isolated from extracts of aerial parts, or of the whole plant were considered. Metabolites isolated from other plant parts or from essential oils were excluded. Solvents that were used for extraction varied and have to be checked in the corresponding references.

In order to illustrate the complexity of taxonomic classification by traditional means a more extensive introduction to the genus *Lychnophora*, which showed serious problems in correct botanical classification is given in Sections 1.3.1, 1.3.2 and 1.3.3 .

1.3.1 Taxonomic classification of the genus *Lychnophora*

The genus *Lychnophora* forms part of the Asteraceae (also referred to as Compositae) family within the tribe Veronieae (NAKAJIMA *et al.*, 2012) and the subtribe Lychnophorinae (ROBINSON, 1999).

Numerous changes considering taxonomic classification were made beginning from the family level down to the genus level. The family Asteraceae was first described by Cassini (1817, 1819). Further authors that Semir (2011) and collaborators name, which contributed to morphology-based classification of the Asteraceae family are Lessing (1829, 1831a, b), De Candolle (1836), Bentham (1873a,b), Hoffmann (1894), Cronquist (1977) and Jeffrey (1978). Subsequently, studies based on molecular taxonomy and DNA analysis by Jansen and Palmer (1987), Bremer (1994,1996) and Panero and Funk (2002) lead to further reclassifications. The last paper published in 2008 by Panero and Funk proposes 12 subfamilies, which are divided in 43 tribes. The same subdivision into 12 subfamilies also corresponds to the classification found on the Angiosperm Phylogeny Website (STEVENS, 2012) (see Figure 2).

Various scientist also contributed to diverse reclassifications of subtribes and genera in the tribe Vernonieae (SEMIR *et al.*, 2011). According to Semir and coworkers (2011) Brazilian plants of the subtribe Vernonieae have mainly been studied by Lessing (1829, 1831a, b), De Candolle (1836), Gardner (1842, 1845), Schultz-Bipontinus (1861, 1863) and Baker (1873). Doubts considering the tribe Vernonieae exist mainly in the subdivision into subtribes. The subtribe Lychnophorinae as represented in Figure 2 corresponds to the most recent classifica-

tion by Robinson (1999). Robinson described 61 genera for the tribe Vernonieae, Semir and coworkers (2011) summarize the tribe to approximately 40 genera that are represented in Brazil and on the official and regularly updated website of the Species List of the Flora of Brasil (NAKAJIMA *et al.*, 2012) 55 genera are listed. Latter is also represented in Figure 2.

The genus *Lychnophora* was first described by Martius in 1822 (MARTIUS, 1822). He collected a total of 12 species, of which he describes 8 and illustrates 7 in his publication in 1822 (MARTIUS, 1822): *L. brunioides, L. ericoides, L. pinaster, L. staavioides, L. rosmarinifolia, L. hakeaefolia*, *L. salicifolia* and *L.villosissima*. In the following years various botanists contributed to revisions, new classifications and additions of new species to the genus. Major publications include Sprengel (1826, 1827), Lessing (1829, 1832), De Candolle (1836), Gardner (1846), Schultz-Bipontinus (1863) and Baker (1873). For a short description of major publications in chronological order see Table 6.

Semir, the only Brazilian botanist that made a revision on taxonomic classification of the genus disagrees with the classifications of all previous authors and proposes 68 species in his unpublished doctoral thesis, which he divides in six sections: Lychnophora, Lychnophoriopsis, Lychnophorioides, Lychnocephaliopsis, Sphaeranthus and Chronopappus. Separation of those he bases on inflorescence morphology and absence or presence of sheath or petiole (SEMIR, 1991). In Semir and coworkers (2011) he heavily critisizes the review by Coile and Jones (1981), see Table 6. The review publicated in 1981 strongly disagrees with the two previous reviews on the genus by Schultz-Bipontinus (1863) and Baker (1873). They reduce the number of species of the genus *Lychnphora* to 11: *L. diamantinana* Coile and Jones, sp. nov., *L. heterotheca* (Schultz-Bip.) Jones and Coile, comb. nov., *L. tomentosa* (Mart. ex DC.) Schultz-Bip., *L. humillima* Schultz-Bip, *L. sellowii* Schultz-Bip., *L. salicifolia* Mart., *L. villosissima* Mart., *L. staavioides* Mart., *L. ericoides* Mart., *L. phylicifolia* DC., *L. uniflora* Schultz-Bip. As critical points Semir and collaborators (2011) mention the imprecise circumscription of the genus, the use of inadequate synonyms, apparent ignorance of the natural habitat of the plants, and incomplete description of morphological characteristics of the genus. Out of this reason, Semir and collaborators (2011) further emphasize that morphological and molecular taxonomic studies only based on voucher specimens clearly show to be insufficient to classify species of the genus *Lychnophora*. Instead, more detailed observations of the plants in their natural habitat together with the selection of appropriate morphological characteristics for their differentiation from other genera are essential for the correct classification (SEMIR *et al.*, 2011).

Besides the morphological based taxonomy studies on the genus *Lychnophora* mentioned above, also cytotaxonomic studies have been realized (MANSANARES *et al.*, 2002). However they turned out to be problematic for classification: chromosome numbers of different species of the genus *Lychnophora* were compared, and didn't show to be an important character

for separation (MANSANARES *et al.*, 2002).

The above described complexity of correct taxonomic classification of species of the genus *Lychnophora* based on morphology as well as cytotaxonomy emphasizes the need for further methodologies supporting traditional taxonomic methods.

Table 6: Publications on the genus *Lychnophora*. Major publications on the genus *Lychnophora* according to Semir and collaborators (2011).

Description	**Ref.**
Collects 12 species, describes 8 of them and illustrates 7	MARTIUS, 1822
First scientist to recognize the genus *Lychnophora* Mart., unites *L. salicifolia* Mart. and *L. hakeaefolia* Mart. and transfers *Vernonia trichocarpha* to *Lychnophora*, now named *Lychnophora trichocarpha*	SPRENGEL, 1826, 1827
Unites *Vernonia* with *Lychnophora* with exception of *L. rosmarinifolia* and transfers *L.trichocarpha* to the genus *Piptocoma* Cass.	LESSING 1829, 1832
Accepts Martius's *Lychnophora* and adds one new species *L. phylicaefolia*, which he found in Martius's Herbarium as *Vernonia*. Describes the genus *Lychnocephalus*.	DE CANDOLLE, 1836
Describes 4 new species: *L. martiana* Gard., *L. rosmarinus* Pohl (as *L. affinis* Gard.), *L. reticulata* Gard., *L. albertinioides* Gard.. Unites *Haplostephium* with *Lychnophora* and adds 2 new species.	GARDNER, 1846
Describes 11 new species, and includes a total of 27 species into the genus. Accepts in general treatment of De Candolle (1836) and heavily refuses treatment of Lessing (1829, 1832). However he unites the genus *Lychnocephalus* described by De Candolle (1836) with *Lychnophora* and names it *L. tomentosa*.	SCHULTZ-BIPONTINUS 1863
Description of 17 species for the genus. Follows in general treatment of Schultz-Bipontinus (1863), but introduces several new infraspecific taxa.	BAKER, 1873
Comments or descriptions of new species	WAWRA, 1888; BEAUVERD, 1913; KRASCHENINNIKOV, 1922; MATTFELD, 1923; BARROSO, 1956, ROBINSON, 1980, 1983

Continued on next page

Table 6 – *Continued from previous page*

Description	**Ref.**
Transfer of some species of *Lychnophora* to the genus *Vernonia*	LEITÃO FILHO and SEMIR, 1979; JONES and COILE, 1981; MACLEISH, 1984
Description of 11 species for the genus *Lychnophora*.	JONES and COILE, 1981
Description of 68 species for the genus *Lychnophora* whereof 27 are new.	SEMIR, 1991
Description of 34 species for the genus *Lychnophora* including *Haplostephium* Mart. ex DC..	ROBINSON, 1999
Description of a new species, *L. sericea*.	HIND, 2000

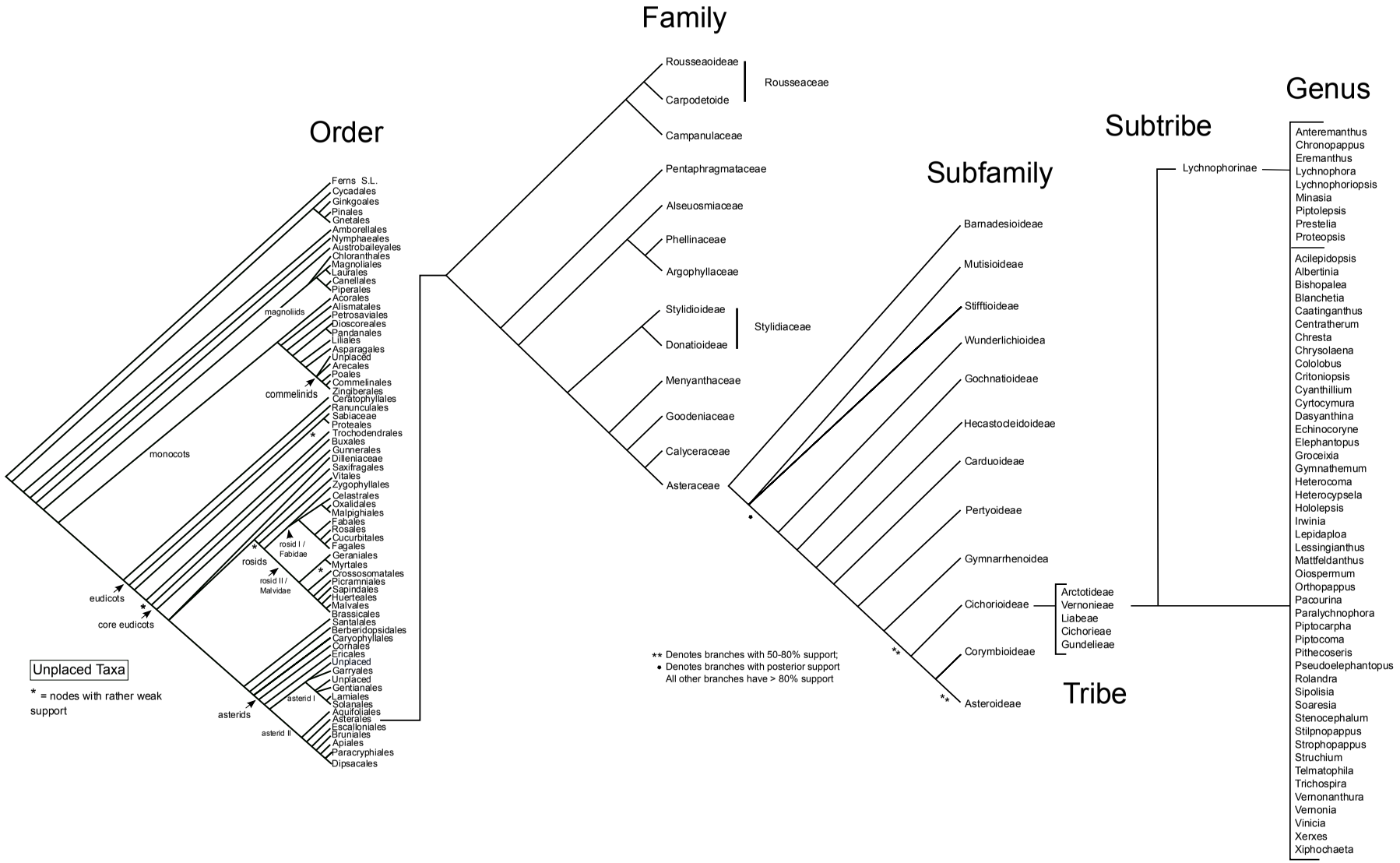


Figure 2: Phylogeny of *Lychnophora*. Up to subfamilial classification the figure is redrawn after STEVENS (2012), tribe classification is according to PANERO and FUNK (2008), subtribal classification according to ROBINSON (1999) and genera classification according to NAKAJIMA *et al.* (2012).

1.3.2 Distribution of species of the genus *Lychnophora*

Martius (1824) was the first scientist to classify Brazilian vegetation. He divided Brazil into five floristic domains, naming those domains after Greek nymphs: Amazon (Nayades), the *Cerrados* of central Brazil (Oreades), the Atlantic rainforests (Dryades), the Araucaria forests and southern grasslands (Napeias), and the northeastern *Caatinga* (Hamadryades) (FIASCHI and PIRANI, 2009).

Although altered several times, up today scientists use similar classification systems for Brazilian vegetation. According to Fiaschi and Pirani (2009) the most widely accepted classification system is that of Veloso and coworkers (1991). According to those authors the country is subdivided into four biomes [1]: the Amazon Forest, the Atlantic Forest, the Savanna (*Cerrado*) and the Steppe (*Caatinga* and *Campos sulinos*). The IBGE (Brazilian Institute of Geography and Statistics) subdivides the country into six biomes: Amazon Forest, Atlantic Forest, *Caatinga*, Savanna (*Cerrado*), Pantanal and Pampa.

According to the IBGE the biome *Cerrado* is after the Amazon the second largest biome and covers 23.92% of the total area of Brazil (see Figure 3). The vegetation type of the *Cerrado* was denominated differently by different authors, such as *Estepe*, *Savana* and *Savannas* between others. Due to the fact that the phytophysionomy of the Brazilian *Cerrado* is similar to that of the African and Asian savannas, the official term accepted by the IBGE today is Savanna with *Cerrado* as regional synonym (IBGE, 2012). Determinig factors for the savannic phytophysionomy are climate, soil, and fire. The climate is sub-humid tropical with dry and wet seasons with annual rainfall ranging from 600 to 2200 mm and temperatures from 22 to 27°C. The geology of the *Cerrado* is one of the most diverse and complex of Brazil with rocks formed during the precambrian and cenozoic era. More than ten different soil types are encountered and morphological features vary according to the altitude, which ranges from 50m to 2000m. Three well-defined units are distinguished by the IBGE: the plains, depressions and plateaus. Plateaus are the most widespread unities, charcteristically they form vast plain surfaces called *Chapadas* (IBGE, 2012).

[1] The term biome (Gr. *bios* = life; *oma* = suffix for generalization (group, set)) was defined differently since its introduction in the mid 20th century. Good reviews are given in Coutinho (2006) and IBGE (2012). The IBGE (2004) states that a biome is defined by its predominant physical conditions, such as climate, lithology, geomorphology, pedology and common evolutionary history. It is composed of a group of spatially close vegetation types, which results like this in an own biological diversity. The authors further summarize the definition as the biome being a community (plants and animals) with a spatially close type of vegetation, similar geo-climatic conditions and a similar evolutionary history, which results in an own biological diversity.

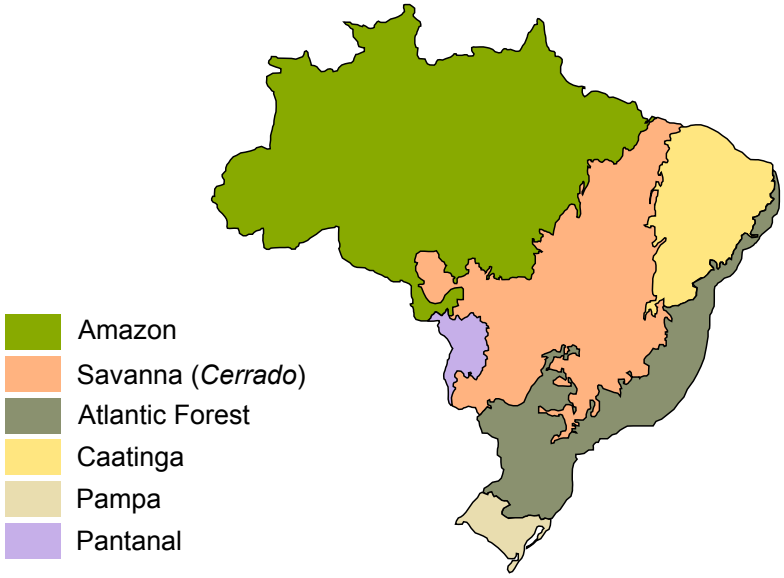


Figure 3: Map of Brazilian biomes (IBGE, 2004).

Species of the genus *Lychnophora* occur in the *Cerrado* biome described above. They are encountered in the Brazilian rupestrian fields, the so-called *campos rupestres* [2] of four Brazilian states: Bahia, Goiás, Minas Gerais and São Paulo. However between places in São Paulo, where *Lychnophora* were found and the state Minas Gerais no ecological differences exist. Therefore it is generally said that *Lychnophora* is endemic to three states: Bahia, Goiás and Minas Gerais (SEMIR *et al.*, 2011). More specifically Semir and collaborators (2011) name three places, where species of the genus *Lychnophora* can be found:

1. Serra do Espinhaço (MG)
2. Chapada Diamantina (BA)
3. Serra dos Pireneus with branches into the Serra Geral do Paraná, Serra Dourada and Serra dos Cristais (GO)

The Serra do Espinhaço and the Chapada Diamantina form together the so-called Cadeia do Espinhaço and places named in item 3. can also be summarized as Maciço Goiano. Both mountain chains are thought to have originated at the same time, during the Precambrian (SEMIR *et al.*, 2011). Figure 4, which was adapted from Semir and collaborators (2011) illustrates the distribution described above of species of the genus *Lychnophora*.

[2]The term was defined differently by several scientists, by botanists and phytogeographs the term is most commonly used to describe the vegetation of the quartz gravel fields along the Cadeia do Espinhaço, a mountain chain in the state of Bahia and Minas Gerais (DE VASCONCELOS, 2011). Here *campos rupestres* describes the vegetation on the top of mountain chains in central and southeastern regions of Brazil (KELES *et al.*, 2010).

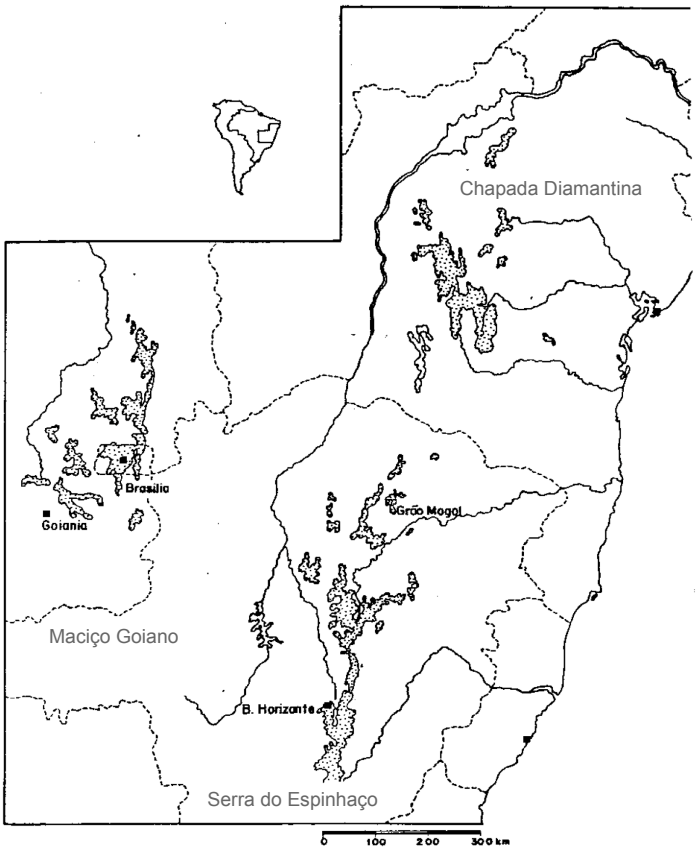


Figure 4: Map of distribution of *Lychnophora* adapted from Semir and collaborators (2011). The dotted area represents the Cadeia do Espinhaço on the right side and Maciço Goiano on the left side, which basically corresponds to the distribution of *Lychnophora*.

1.3.3 Chemical constituents of Lychnophorinae

Species of the genus *Lychnophora* are popularly known as "arnica da serra" or "falsa arnica" (BASTOS *et al.*, 1987). In Brazilian folk medicine extracts of aerial parts or roots of species of the genus *Lychnophora* in water, ethanol or "cachaça" (sugar cane spirit), which are administered orally or topically, are used as analgesic and anti-inflammatory treatments (GUZZO *et al.*, 2008; KELES *et al.*, 2010; SEMIR *et al.*, 2011). Reported biological activities include *in vitro* antitumor, analgesic, antimicrobial, cytotoxic, analgesic and anti-inflammatory activity (SEMIR *et al.*, 2011; KELES *et al.*, 2010). *L.ericoides* was the first species being used as a medicinal plant, though, over time, almost all species having any morphological similarity with *L.ericoides* were utilized. Emphasizing by this means the importance of a correct taxonomic classification of species of this genus (SEMIR *et al.*, 2011).

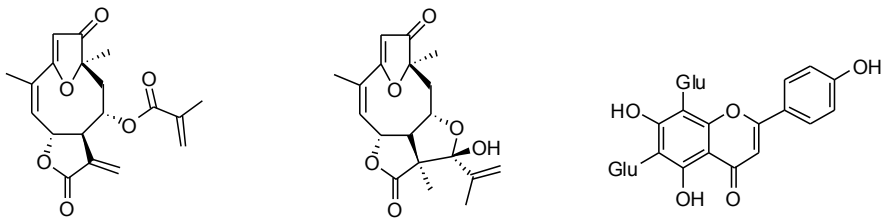
An extensive review on chemical compounds isolated from species of the subtribe Lychnophorinae, to which *Lychnophora* belongs to (besides *Eremanthus* the most numerous genus of the subtribe) can be found in Keles and collaborators (2010). Data reported in the following

is based on this review.

Based on literature, the main secondary metabolites reported for species of the subtribe Lychnophorinae are terpenoids (70.2 %) and flavonoids (16.9%), furthermore, also acetylene derivatives, quinic acid derivatives, benzoic acid derivatives and phenylpropanoids are mentioned. Among the terpenoids, sesquiterpenes are most abundant. Triterpenes, steroids, saponins and monoterpenes were reported. Diterpenes were only found in *L. sellowii* (BOHLMANN *et al.*, 1982a). Sesquiterpene lactones were reported for 90% of the species of the subtribe Lychnophorinae. Sesquiterpene lactones of a major part of the investigated species were furanoheliangolides of the goyazensolide and eremantholide type. Only two species of *Lychnophora*, *L. pseudovillosissima* and *L. reticulata* showed guaianolides and eudesmanolides.

In flavonoids, flavonols and flavones are most commonly found, whereas dihydroflavones and dihydroflavanols are less common. Chalcones could only be isolated from *L. ericoides* (GOBBO-NETO and LOPES, 2008).

Quinic acid derivatives up to now have only been reported in the species *L. ericoides* (DOS SANTOS *et al.*, 2005; GOBBO-NETO *et al.*; 2008), *L.pinaster* (DE MORAES *et al.* 2009), *L.pohlii* (DE MORAES *et al.* 2009) and *L. villosissima* (DE MORAES *et al.* 2009). Figure 5 shows some substances isolated from aerial parts of species of *Lychnophora*. A furanoheliangolide of the goyazensolide type (**1**) (ALVES *et al.*, 2008) and of the eremantholide type (**2**) (BOHLMANN *et al.*, 1981d) and the flavone vicenin-2 (**3**) (GRAEL *et al.*, 2005), which showed significant anti-inflammatory activity (GOBBO-NETO *et al.*, 2005).



1 **2** **3**

Figure 5: Structures of compounds isolated from aerial parts of species of the genus *Lychnophora*.

1.3.4 Chemical constituents of the genus *Vernonia*

With almost 1000 species the genus *Vernonia* is the largest genus in the tribe Vernonieae, which comprises a total of 1500 species. Species of this genus can adapt to a variety of habitats, however they are most frequently encountered in tropical regions, mainly in Africa and South America (BREMER, 1994; KEELEY and JONES, 1979). Also the morphological variation observed in the genus is wide. Species of the genus *Vernonia* are encountered as annuals,

herbaceous perennials, lianas, shrubs and trees (KEELEY and JONES, 1979). They are widely used as food and medicine and various bioactive compounds have been isolated so far. A good review of ethnomedicinal uses of plants of the *Vernonia* genus and *in vivo* and *in vitro* studies of isolated compounds can be found in Toyang and Verpoorte (2013).

Secondary metabolites described for species of the genus *Vernonia* include triterpenes (ALVES *et al.*, 1997), diterpenes (KOS *et al.*, 2006), steroids (DA COSTA *et al.*, 2008; JISAKA *et al.*, 1993; JISAKA *et al.*, 1992; MACHADO *et al.*, 2013; OHIGASHI *et al.*, 1991; TCHINDA *et al.*, 2003; TCHINDA *et al.*, 2002), sesquiterpene lactones (ABEGAZ *et al.*, 1994; BAZON *et al.*, 1997; BUSKUHL *et al.*, 2010; JAKUPOVIC *et al.*, 1986; PERDUE *et al.*, 1993; ZDERO *et al.*, 1991) and flavonoids (ABEGAZ *et al.*, 1994; BUSKUHL *et al.*, 2010; IGILE *et al.*, 1994; MORALES-ESCOBAR *et al.*, 2007). Most abundant compounds found in the genus are the flavonoids and sesquiterpene lactones. In sesquiterpene lactones the family of germacranolides, the glaucolides and hirsutinolides are predominantly found (BUSKUHL *et al.*, 2010). In African species also elemanolides have been described (ABEGAZ *et al.* 1994; ZDERO *et al.*, 1991) and Da Costa and coworkers (2005) and Buskuhl and collaborators (2010) also reported cadinanolides in Brazilian species. However, there is an ongoing discussion on whether hirsutinolides and cadinanolides are artefacts formed during isolation or whether they are natural products contained in the plant (BAZON *et al.*, 1997; BUSKUHL *et al.*, 2010).

Table 7 gives a short summary of compounds isolated and identified in the articles cited above. A representative structure of each mentioned class can be found in Figure 6. The triterpene lupeol (**4**) (ALVES *et al.* 1997; DA COSTA *et al.*, 2008; MACHADO *et al.*, 2013), the diterpene *ent*-kaurane glycoside (**5**) (KOS *et al.*, 2006), a stigmastane-type steroid glucoside (vernonioside) (**6**) (OHIGASHI *et al.*, 1991), the flavonoid luteolin (**7**) (ALVES *et al.*1997; BUSKUHL *et al.*, 2010; IGILE *et al.* 1994; MACHADO *et al.*, 2013), a hirsutinolide (**8**) (ZDERO *et al.*, 1991), and a glaucolide (**9**) (ZDERO *et al.*, 1991).

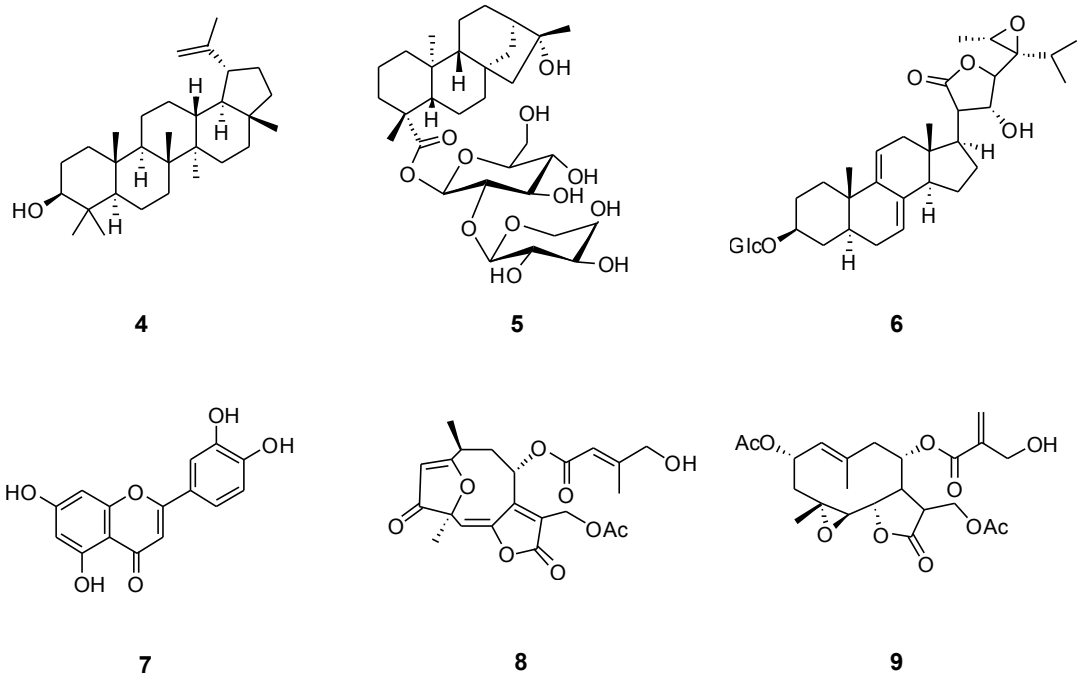


Figure 6: Structures of compounds isolated from aerial parts of species of the genus *Vernonia*.

Table 7: Chemical constituents identified in species of the genus *Vernonia*.

Species	Chemical constituents	Ref.
V. amygdalina	7 stigmastane-type steroid glucosides (vernoniosides) and one of its aglycones	JISAKA *et al.*, 1992 and 1993; OHIGASHI *et al.*, 1991
V. brasiliana	3 triterpenes (lupeol, β-amyrin, germanicol)	ALVES *et al.*, 1997
V. chalybaea	8 triterpenes, 2 steroids, 5 flavonoids, 1 trihydroxy alcohol, 1 tetrahydroxy aliphatic ether	DA COSTA *et al.*, 2008
V. cinerascens	2 sesquiterpene lactones	ABEGAZ *et al.*, 1994
V. filigera	2 elemanolides (vernolepin and vernodalin) and 2 flavonoids	ABEGAZ *et al.*, 1994
V. fruticulosa	1 glaucolide, 3 cadinanolides and 3 flavonoids (isorhamnetin, acacetin and tamarixetin)	BAZON *et al.*, 1997
V. galamensis	5 glaucolides, one aromatic compound (benzylsenecioate)	PERDUE *et al.*, 1993
V. guineensis	2 stigmastane derivatives (vernoguinosterol and vernoguinoside), 2 sucrose esters	TCHINDA *et al.*, 2003; TCHINDA *et al.*, 2002
V. hindei	2 flavnonoids (eriodictyol and isorhamnetin) and 1 elemanolide (vernodalin)	ZDERO *et al.*, 1991
V. leopoldi	3 sesquiterpene lactones, 1 coumarin glucoside	ABEGAZ *et al.*, 1994
V. mapirensis	8 flavonoids, 2 benzofuranones	MORALES-ESCOBAR *et al.*, 2007
V. natalensis	3 glaucolides	ZDERO *et al.*, 1991
V. scorpioides	2 glaucolides and 4 hirsutinolides, 2 flavonoids (luteolin and apigenin) and ethyl caffeate, 2 triterpenes (lupenone, lupeol, taraxasteryl acetate), 2 steroids (β-sitosterol, stigmasterol), 1 polyacetlyene lactone	BUSKUHL *et al.*, 2010; MACHADO *et al.*, 2013
V. steetziana	2 glaucolides, 3 hirsutinolides	ZDERO *et al.*, 1991
V. syringifolia	6 glaucolides, 1 hirsutinolide, 1 flavanone (eriodictyol)	ABEGAZ *et al.*, 1994; ZDERO *et al.*, 1991
V. trifloscula	3 hirsutinolides, 2 *ent*-kaurane glycosides, one kaurane acid, two flavonoids	KOS *et al.*, 2006

1.3.5 Chemical constituents of the genus *Ageratum*

The genus *Ageratum* forms part of the tribe Eupatoriae and consists of approximately 30 species of which very few have been phytochemically investigated (BURKILL, 1985). Table 8 lists some chemical compounds identified in species of the genus *Ageratum*. The most studied species is *Ageratum conyzoides*. A review on its chemical constituents and biological activities was published by Okunade (2002). Four main classes are described, mono- and sesquiterpenes (from essential oils); chromones, chromenes and coumarin; polyoxygenated flavonoids and triterpenes and sterols. Furthermore studies of anti-inflammatory, analgesic, antipyretic and insecticidal activities are described.

In Brazil *Ageratum fastigiatum* is also known as "matapasto" or "enxota" and aqueous extracts are used to treat pain and inflammations (GONÇALVES *et al.*, 2011). Substances that were isolated from this species include triterpenes, diterpenes, coumarin, eudesmane derivatives, labdane derivatives and steroids (see Table 8). Figure 8 shows some substances isolated from species of the genus *Ageratum*. An eudesmane derivative (**10**) (BOHLMANN *et al.*, 1983a), a *seco*-kaurene (**11**) (BOHLMANN *et al.*, 1981a), the triterpene taraxasterol (**12**) (GONÇALVES *et al.*, 2011), the chromone conyzorigun (**13**) (ADESOGAN and OKUNADE, 1978) and a flavone (**14**) (QUIJANO *et al.*, 1980).

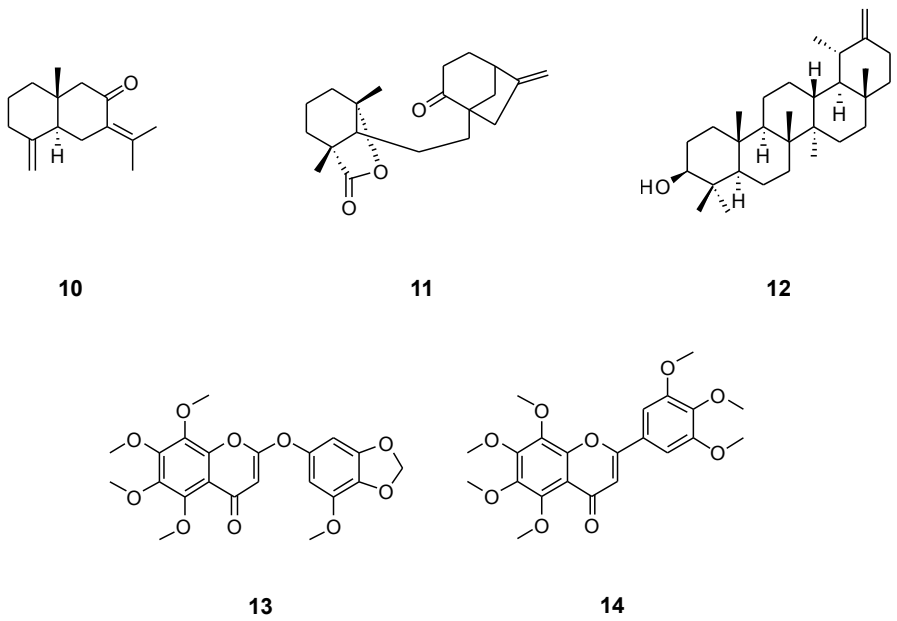


Figure 7: Structures of compounds isolated from aerial parts of species of the genus *Ageratum*.

Table 8: Chemical constituents identified in species of the genus *Ageratum*.

Species	Chemical constituents	Ref.
A. conyzoides	1 flavone, 2 steroids, 1 paraffin hydrocarbon, 2 chromones, 2 chromenes, 1 triterpene (friedelin)	ADESOGAN and OKUNADE, 1979; ADESOGAN and OKUNADE, 1978; HUI and LEE, 1971; KASTURI and MANITHOMAS, 1967
A. corymbosum	1 coumarin, 2 steroids, 1 triterpene, 4 flavones, 2 flavanones	QUIJANO *et al.*, 1980
A. fastigiatum	3 steroids, squalene, 10 triterpenes, 4 eudesmane derivatives, 1 angelate, 2 diterpenes (*seco*-kaurene and -kaurane derivative), 2 labdane derivatives, 1 coumarin, 4 lactones, 3 dehydronerolidol derivatives	BOHLMANN *et al.*, 1981a; BOHLMANN *et al.*, 1983a; GONÇALVES *et al.*, 2011
A. houstonianum	2 steroids, 2 triterpenes (friedelin, friedelan-3β-ol), 2 chromenes	BOWERS *et al.*, 1976, HUI and LEE, 1971

1.3.6 Chemical constituents of the genus *Bidens*

The genus *Bidens* includes more than 200 species and is represented in the Americas, Africa, Polynesia and to a smaller extent in Europe and northern Asia (GANDERS *et al.*, 2000). *Bidens alba* forms part of the *Bidens pilosa* complex, a species complex comprised of *Bidens odorata*, *Bidens alba* and *Bidens pilosa*. Species of this complex are widely distributed subtropical and tropical weeds with a center of diversification in Mexico (BALLARD, 1986). They are encountered in agricultural areas and along roadsides. Grombone-Guaratini and collaborators (2006) report that *Bidens alba* is the only species of the complex that is restricted to the seacoast.

Da Silva (2009) reports that compared to the amount of species of the genus *Calea* only very few studies of chemical constituents exist. Based on literature, compounds that were found most frequently included polyacetylenes (34%), chalcones (12%), phenylpropanoids (9%), flavonoids (9%), thiophene derivatives (9%) and aurones (5%) (DA SILVA, 2009). Several biological activity studies for species of the genus *Bidens* can be found in literature including antiulcer, antioxidant, anti-inflammatory, immunomodulatory, anti-hypertensive, antimicrobial, anti allergic, anti-diabetic, antiviral, antitumoral and antimalarial activities, which were related to the presence of flavonoids and polyacetylenes (DA SILVA, 2009).

Table 9 was adapted from Da Silva (2009), only substances that have been isolated or identified from aerial parts or the whole plant were included. In Figure 8 a representative substance for each described class is shown. The polyacetylene 1-phenylhepta-1,3,5-triyne (**15**) (CANTONWINE and DOWNUM, 2001), a chalcone (**16**) (WANG *et al.*, 2007a), a phenylpropanoid (**17**) (SASHIDA *et al.*, 1991), the flavonoid luteolin (**18**) (WOLNIAK *et al.*, 2007), a tiophene acetylene (**19**) (CHRISTENSEN, 1990) and the aurone sulfurein (**20**) (ZHU *et al.*, 2009). For a recent review article about botanical properties, phytochemistry and pharmacology of *Bidens pilosa* see Bartolome and collaborators (2013).

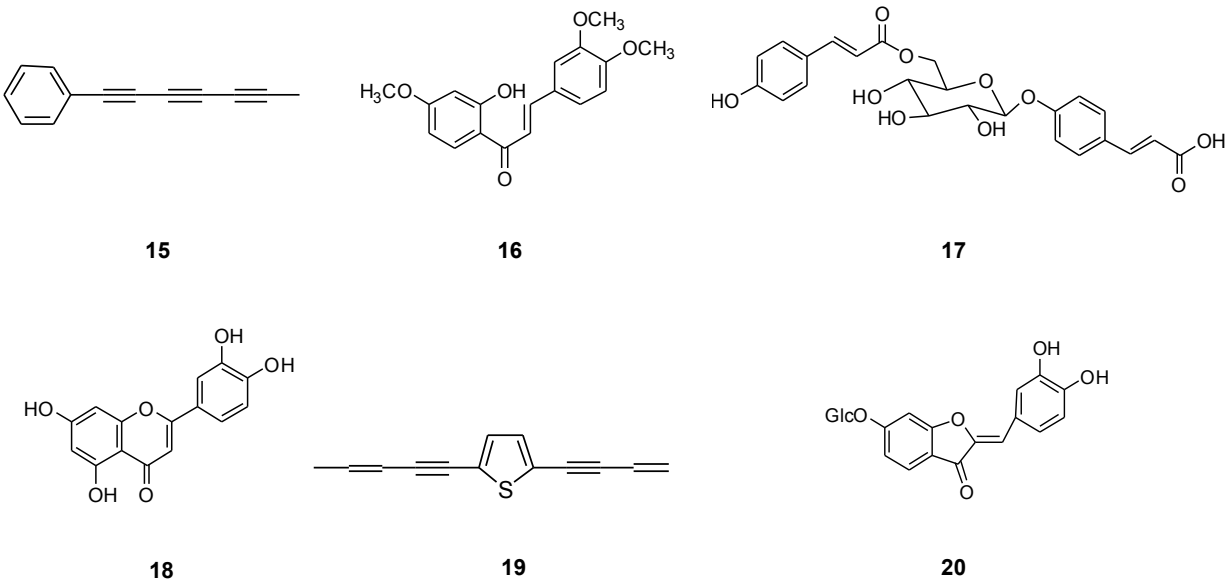


Figure 8: Structures of compounds isolated from aerial parts of species of the genus *Bidens*.

Table 9: Chemical constituents identified in species of the genus *Bidens*.

Species	Chemical constituents	Ref.
B. alba	1 polyacetylene	CANTONWINE and DOWNUM, 2001
B. amplectens	2 polyacetylenes, 1 tiophene derivative	MARCHANT *et al.*, 1984
B. assymetrica	5 polyacetylenes, 1 tiophene derivative	MARCHANT *et al.*, 1984
B. aurea	2 steroids and 1 chalcone	ORTEGA *et al.*, 2000
B. bipinnata	2 polyacetylenes	LI *et al.*, 2004
B. campylotheca	3 chalcones, 17 polyacetylenes, 2 tiophene derivatives	BAUER *et al.*, 1992; MARCHANT *et al.*, 1984; REDL *et al.*, 1994; REDL *et al.*, 1993
B. ceruna	1 flavone, 2 flavonols, 4 chalcones, 1 aurone	ZHU *et al.*, 2009
B. cervicata	5 polyacetylenes, 2 thiophene derivatives	MARCHANT *et al.*, 1984
B. conjuncta	5 polyacetylenes, 1 thiophene derivative	MARCHANT *et al.*, 1984
B. cosmoides	7 polyacetylenes, 2 thiophene derivatives	MARCHANT *et al.*, 1984
B. forbesii	6 polyacetylenes, 2 thiophene derivatives	MARCHANT *et al.*, 1984
B. hawaiensis	4 polyacetylenes, 2 thiophene derivatives	MARCHANT *et al.*, 1984
B. leucantha	4 flavonols, 5 chalcones	TOMMASI and PIZZA, 1997
B. macrocarpa	5 polyacetylenes, 2 thiophene derivatives	MARCHANT *et al.*, 1984
B. maviensis	4 polyacetylenes, 1 thiophene derivative	MARCHANT *et al.*, 1984
B. menziesii	6 polyacetylenes, 4 thiophene derivatives	MARCHANT *et al.*, 1984
B. micrantha	1 polyacetylene, 1 thiophene derivative	MARCHANT *et al.*, 1984
B. molokaiensis	5 polyacetylenes, 4 thiophene derivatives	MARCHANT *et al.*, 1984
B. hillebrandiana	4 polyacetylenes, 2 thiophene derivatives	MARCHANT *et al.*, 1984

Continued on next page

Table 9 – *Continued from previous page*

Species	Chemical constituents	Ref.
B. parviflora	5 chalcones, 2 flavanones, 2 aurones, 2 monoterpenes, 1 jasmololone, 2 neolignans, 3 sucrose esters, 1 lignan, 5 polyacetylenes	WANG *et al.*, 2007a; WANG *et al.*, 2007b; WANG *et al.*, 2006; WANG *et al.*, 2003; WANG *et al.*, 2001
B. pilosa	2 sesquiterpenes, 3 diterpenes, 8 triterpenes, 7 steroids, 5 flavones, 7 flavonols, 1 flavanone, 9 chalcones, 7 aurones, 20 polyacetylenes, 13 phenylpropanoids, 1 chromene, 1 carotenoid, 8 pheophytins, 1 caffeine, 1 acetylacetone	ALVAREZ *et al.*, 1996; BENHURA *et al.*, 1997; BRANDÃO *et al.*, 1998; CHANG *et al.*, 2007; CHANG *et al.*, 2000; CHIANG *et al.*, 2007; CHIANG *et al.*, 2004; GEISSBERGER and SEQUIN, 1991; HOFFMANN and HOLZL, 1988a, 1988b, 1988c; KUMAR and SINHA, 2003; KUMARI *et al.*, 2009; LEE *et al.*, 2008; OGAWA and SASHIDA, 1992; PEREIRA *et al.*, 1999; SARG *et al.*, 1991; SARKER *et al.*, 2000; SASHIDA *et al.*, 1991; TOBINAGA *et al.*, 2009; UBILLAS *et al.*, 2000; WANG *et al.*, 1997; WU *et al.*, 2007; ; WU *et al.*, 2004; ZOLLO *et al.*, 1995; ZULUETA *et al.*, 1995
B. populifolia	3 polyacetylenes, 1 thiophene derivative	MARCHANT *et al.*, 1984
B. sandvicensis	4 polyacetylenes, 1 thiophene derivative	MARCHANT *et al.*, 1984
B. subalternans	3 polyacetylenes, 1 thiophene derivative	ORTEGA *et al.*, 2000
B. torta	5 chalcones, 10 polyacetylenes, 3 thiophene derivatives	ORTEGA *et al.*, 2000
B. tripartita	2 monoterpenes, 2 flavones, 1 flavanone, 1 chalcone, 3 polyacetylenes, 1 phenylpropanoid, 2 thiophene derivatives, 1 linoleic acid	CHRISTENSEN *et al.*, 1990; WOLNIAK *et al.*, 2007

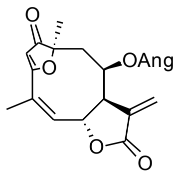
Continued on next page

Table 9 – *Continued from previous page*

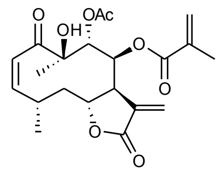
Species	**Chemical constituents**	**Ref.**
B. valida	6 polyacetylenes, 1 thiophene derivative	MARCHANT *et al.*, 1984
B. wiebkei	2 polyacetylenes, 1 thiophene derivative	MARCHANT *et al.*, 1984

1.3.7 Chemical constituents of the genus *Calea*

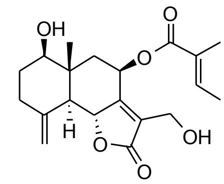
The genus *Calea* comprises about 110 species, which can be found in Mexico, Central and South America (KARIS and RYDING, 1994). According to Do Nascimento and coworkers (2002) approximately 37 species have been investigated chemically, some of them are listed in Table 10. Sesquiterpene lactones of the furanoheliangolide, germacranolide, eudesmanolide and guaianolide type, daucane and acorane sesquiterpenes, *p*-hydroxyacetophenone derivatives, thymol derivatives, benzofurans and chromenes were some of the main substances isolated from species of the genus *Calea*. Figure 9 shows some of the structures. A furanoheliangolide (**21**) (BOHLMANN *et al.*, 1982b), a germacranolide (**22**) (BOHLMANN and JAKUPOVIC, 1979), the eudesmanolide trichomatolide A (**23**) (OBER *et al.*, 1984c), a guaianolide (**24**) (CASTRO *et al.*, 1989), a daucane (**25**) (JAKUPOVIC and BOHLMANN, 1984), an acorane (**26**) (JAKUPOVIC and BOHLMANN, 1984), a *p*-hydrocyacetophenone (**27**) (GÓMEZ and GIL, 2011), a thymol derivative (**28**) (MALDONADO *et al.*, 1992) and the benzofuran derivative calebertin (**29**) (OBER *et al.*, 1985). Several biological activities of species of the genus *Calea* have already been reported, Do Nascimento and coworkers (2004) summarize them as anti-inflammatory, cytotoxic, larvicidal, antiplasmodial and antihypertensive.



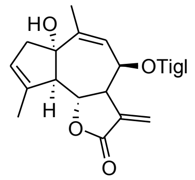
21



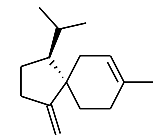
22



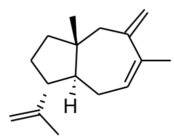
23



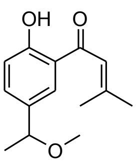
25



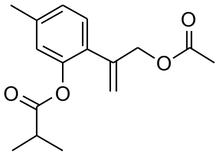
25



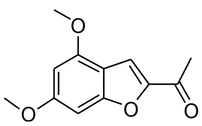
26



27



28



29

Figure 9: Structures of compounds isolated from aerial parts of species of the genus *Calea*.

Table 10: Chemical constituents identified in species of the genus *Calea*.

Species	Chemical constituents	Ref.
C. berteriana	1 benzofuran, 1,4-dioxin derivative	OBER *et al.*, 1985
C. clausseniana	5-deoxyflavone glycoside	DO NASCIMENTO and DE OLIVEIRA, 2007
C. hymnolepsis	2 benzofurans, 3 furanoheliangolides, 1 heliangolide, 2 sesquiterpene lactones	BOHLMANN *et al.*, 1982b
C. nelsonii	4 thymol derivatives	MALDONADO *et al.*, 1992
C. peckii	3 benzofurans, 7 chromene derivatives, 4 bisabolene derivatives, 3 guaianolides, 2 heliangolides	CASTRO *et al.*, 1989
C. pilosa	squalene, α-humulene, caryophyllene epoxide, spathulenol, thymohydroquinone dimethyl ether, 1 dehydro compound, nerolidol, 2 hydroxyderivatives, phytol, β-amyrin, 3 dithienyl derivatives, 9 heliangolides, 1 costunolide, 1 new lactone type (heliangolide substituted with a monoterpene residue)	BOHLMANN *et al.*, 1981c
C. platylepis	3 flavonoids, 1 sesquiterpene, 4 benzofurans, 2 steroid saponins, *p*-hydroxyacetophenone	DO NASCIMENTO *et al.*, 2002
C. prunifolia	3 acorane derivatives, 1 lactone, 1 norsesquiterpene ketone, 7 benzofurans, 1 acetate, 2 norsesquiterpenes, 2 cyclic ethers, 1 daucane derivative, 1 prenyl phenol, 2 lactones, 3 sesquiterpene lactones, 1 dibenzofurane, 1 aldehyde, 1 costol derivative, 2 *p*-hydroxyacetophenone derivatives, 1 flavonoid glycoside, 1 quinic acid derivative, 1 kaurane diterpenoid glycoside	CASTRO *et al.*, 1984, 1989; GÓMEZ and GIL, 2011; OBER *et al.*, 1985; PUEBLA *et al.*, 2011

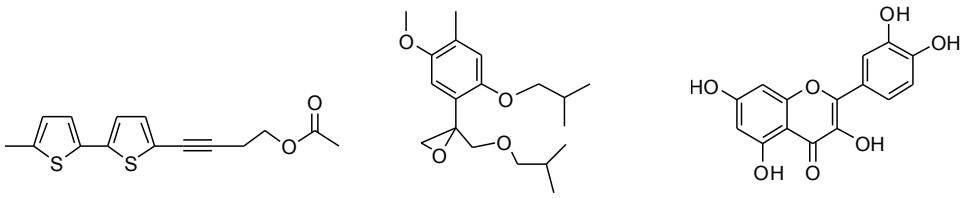
Continued on next page

Table 10 – *Continued from previous page*

Species	**Chemical constituents**	**Ref.**
C. serrata	2 chromenes	STEINBECK *et al.*, 1997
C. subcordata	3 guaianolides, 1 eudesmanolide	OBER *et al.*, 1984a, 1984b
C. teucrifolia	2 germacrene derivatives, 1 spathulenol, 1 monoterpene, 1 triterpene, 1 acid, 1 furanoheliangolide, 2 nerolidol derivatives, 4 benzofuran derivatives, 1 caryophyllene derivative	BOHLMANN *et al.*, 1981b
C. trichomata	5 eudesmanolides	OBER *et al.*, 1984c
C. urticifolia	8 sesquiterpene lactones, 9 germacranolides	BOHLMANN and JAKUPOVIC, 1979; FERREIRA *et al.*, 1980; HERZ and KUMAR, 1980
C. zacatechichi	2 germacrolides, 7 sesquiterpene lactones, 8 germacranolides	BOHLMANN and ZDERO, 1977; FERREIRA *et al.*, 1980; HERZ and KUMAR, 1980; QUIJANO *et al.*, 1979; WU *et al.*, 2011
New species	4 furanoheliangolides, 1 heliangolide	BOHLMANN *et al.*, 1981b

1.3.8 Chemical constituents of the genus *Porophyllum*

Species of the genus *Porophyllum* can be found in an area extending from the southwest of the United States to South America (RICHETT, 1966). Several studies on the essential oils of species of the genus *Porophyllum* and their insecticidal and antioxidative activity could be found (LOAYZA *et al.*, 1999; JIMENEZ *et al.*, 2012, GUILLET *et al.*, 1998; BEZERRA *et al.*, 2002; FONSCECA *et al.*, 2006; LABUCKAS *et al.*, 1999), however very few studies were found on non-volatile substances. Table 11 lists some of them. In Brazil *Porophyllum ruderale* is widely distributed and is used in traditional medicine as an antibacterial and anti-inflammatory, for the treatment of snakebites, stomach ache and wounds (LORENZI and MATOS, 2002). Structures of substances isolated from species of the genus *Porophyllum* can be found in Figure 10. A dithienyl derivative (**30**) (BOHLMANN *et al.*, 1980), a thymol derivative (**31**) (BOHLMANN *et al.*, 1983b) and the flavonol quercetin (**32**) (VAN BAREN *et al.*, 1994).



30 **31** **32**

Figure 10: Structures of compounds isolated from aerial parts of species of the genus *Porophyllum*.

Table 11: Chemical constituents identified in species of the genus *Porophyllum*.

Species	Chemical constituents	Ref.
P. gracile	4 thiophenes	DOWNUM *et al.*, 1985; DOWNUM and TOWERS, 1983
P. obscurum	2 flavonols and derivatives, chlorogenic acid, caffeic acid, 1 coumarin (escopoletin)	VAN BAREN *et al.*, 1994
P. riedelii	2 thiophene derivatives, 3 hydrocarbons, 4 thymol derivatives	BOHLMANN *et al.*, 1983b
P. ruderale	3 diethienyl derivatives, 1 thiophene	BOHLMANN *et al.*, 1980; DOWNUM *et al.*, 1985
P. scoparia	4 dithienyl derivatives, 1 flavanone	BOHLMANN *et al.*, 1985
P. scoparium	1 thiophene	DOWNUM *et al.*, 1985

1.3.9 Chemical constituents of the genus *Lavoisiera*

Renner (1993) describes 46 species belonging to the genus *Lavoisiera* that are encountered in south-central Brazil. Very few phytochemical studies of species of this genus were found. An extensive study on flavonoids contained in species of three genera of the tribe Microlicieae including *Lavoisiera* was done by Bomfim-Patrício and collaborators (2001). The 15 analysed species of *Lavoisiera* showed a large diversity of flavonoids. Flavones (mostly apigenin derivatives) and flavonols, as well as 6-oxygenated derivatives (mainly flavones) were identified. Jamal and coworkers (1999) reported antinociceptice activity of an extract of *Lavoisiera pulcherrima* and isolated the triterpenes betulin and betulinic acid, as well as ferulic acid esters. Cota and collaborators (2002) reported antimicrobial activity of extracts of *L. confertiflora* and *L. cordata* against *Staphylococcus aureus* and *Micrococcus luteus*.

1.3.10 Chemical constituents of the genus *Microlicia*

The genus *Microlicia* comprises 100 species, which are encountered in south-central Brazil, and a few also in Guayana, Peru and Bolivia (RENNER, 1993). In Brazil the herbs or shrubs of species of this genus are predominantly found in the *campos rupestres* (TOUDAHL *et al.*, 2012). Almost no studies on the chemistry of species of this genus could be encountered. Bomfim-Patrício and coworkers (2001) name tannins (very common) and alkaloids (rare) as being typical for the family of Melastomataceae to which the genus *Microlicia* belongs to.

The 17 species of *Microlicia* analysed by Bomfim-Patrício and coworkers (2001) showed highly homogeneous flavonoid patterns: flavonol derivatives were predominantly found with a wide structural diversity of quercetin and kaempferol derivatives. On the contrary to species of the genus *Lavoisiera*, which belongs to the same tribe as *Microlicia* no 6-oxygenated flavonoid derivatives were found. Chemical compounds isolated from the essential oil of one *Microlicia* species, *Microlicia graveolens* are described in Toudahl and coworkers (2012).

1.4 Introduction to metabolomics data analysis

As an enormous amount of data is produced by metabolomics studies, statistical and computational methods have been applied for the evaluation of the large data sets that are generally obtained (MEHROTRA and MENDES, 2006; ROESSNER *et al.*, 2002). Mathematical and statistical skills are a necessary prerequisite for the extraction of a maximum of information contained in metabolomics data (SHULAEV, 2006; TRYGG *et al.*, 2007). Research into metabolomics requires its own computational methods, which are routinely applied

in metabolomics experiments (WISHART, 2007). However, standard methods have not yet been defined, and analytical methods have been adapted from other 'omics' sciences that have been developed in the fields of chemometrics, chemoinformatics, and bioinformatics (TRYGG *et al.*, 2007; WISHART, 2007). Some sources denote the data processing tools employed in metabolomics as chemometric methods (BROWN *et al.*, 2005; TRYGG *et al.*, 2007; TRYGG *et al.*, 2006), whereas others designate the utilized methods as combined methods from bioinformatics and chemoinformatics, since metabolomics is not only about the identification and quantification of metabolites, but also about the contextualization of the data from a biological and metabolic viewpoint (WISHART, 2007). The terms are overlapping, and brief definitions are given in Table 12. In the case of metabolomics research in plant taxonomy, the use of chemoinformatic or chemometric methods (the terms are overlapping and in the present study describe one and the same tool, see definitions in Table 12) are sufficient.

At present, the field of computational metabolomics is still evolving, and the main developments in this field will be a key factor in making metabolomics a standard analytical procedure in different research areas (WISHART, 2007). This is because computational metabolomics will not only have a major impact on data analysis, but also on (i) metabolomics databases; (ii) metabolomics laboratory information management systems (LIMS) and data standards; (iii) spectral analysis tools for metabolomics ; and (iv) metabolic modeling (WISHART, 2007).

As these fields are not of particular interest regarding data analysis of metabolomics studies for plant taxonomy, they will not be further discussed at this point. However, it is suggested that whenever a metabolomic experiment is carried out, one should also seek familiarization with those aspects of computational informatics (WISHART, 2007). Storage of raw data plays a particularly important role, because there is always the possibility that better data processing methods will be developed in the future. This is particularly probable in the case of metabolomics research, as it is still a developing research field (MEHROTRA and MENDES, 2006). A good overview of computational metabolomics is given in Wishart (2007). Hansen (2007) and Steuer and coworkers (2007) both give introductions into data analysis as well as preprocessing in metabolomics research. Steuer and coworkers (2007) additionally provide scripts and implementations for the statistical program Matlab in a ready-to-use format on the webpage `http://bioinformatics.mpimp-golm.mpg.de/`. In the following a brief introduction into metabolomics data analysis including preprocessing and multivariate data analysis is given.

For the multivariate analysis various computational programs may be applied such as Excel, Matlab, SAS, Octave or R, some basic statistical analyses are also available in the preprocessing software Mzmine and XCMS (PLUSKAL *et al.*, 2010; WEI *et al.*, 2011). R (R Devel-

opment Core Team, 2011), a free software environment for statistical computing and graphics, was used in the present master's thesis, as it is available for free, widely used by the scientific community and furthermore open-source, meaning that exact mathematical algorithms applied for the analysis are accessible and can be retraced by any scientist, facilitating like this data exchange between different laboratories. Furthermore R is extensible through a wide varieties of packages, all spectra preprocessing steps, which will be described in more detail in the following could also be performed in R, through the packages MALDIquant (GIBB and STRIMMER, 2012) and MALDIquantForeign (GIBB, 2013). All R scripts used for data analysis for all described experiments were developed in collaboration with Ricardo Silva, doctoral student of Prof. Dr. Ricardo Z. N. Vêncio of the Department of Computing and Mathematics, Faculty of Philosophy, Sciences and Letters of Ribeirão Preto, University of São Paulo (FFCLRP, USP).

Table 12: Definitions of chemometrics, chemoinformatics and bioinformatics.

Term	Explanation	Ref.
Chemometrics	"Chemical discipline that uses mathematics, statistics and formal logic (a) to design or select optimal experimental procedures; (b) to provide maximum relevant chemical information by chemical data; and (c) to obtain knowledge about chemical systems." (MASSART, 1988). In contrast to chemoinformatics, the chemical structure is not of interest, but in the area of application of statistical methods (machine learning) the two disciplines, chemometrics and chemoinformatics, overlap. The term appeared in the early 1970s and was coined by Svante Wold.	MASSART, 1988; VARNEK and BASKIN, 2011; WOLD and SJÖSTRÖM, 1998; WOLD, 1995
Chemoinformatics	"Chemoinformatics is the application of informatics methods to solve chemical problems." (GASTEIGER, 2006) A good summary and more detailed definitions of chemoinformatics are given in VARNEK and BASKIN (2011). The term first appeared in 1998. Chemoinformatics includes chemical structure modelling as well as the application of statistical methods (machine learning).	GASTEIGER, 2006; VARNEK and BASKIN, 2011
Bioinformatics	The term first appeared in 1970. It was introduced by Hesper and Hogeweg (HOGEWEG, 2011) and defined as "the study of informatic processes in biotic systems". It provides a systematic approach to the treatment of experimental data from biological samples by statistical and computational tools.	HOGEWEG, 2011

1.4.1 Preprocessing of spectra

The aim of data preprocessing is to eliminate the variance and bias that is introduced during the analysis of samples, which cannot be attributed to real differences and may mask the outcome of chemometric methods (SMITH *et al.*, 2006).

Since data preprocessing software is usually proprietary or may only be applied to data produced by a vendor's specific instrument, efforts have been made towards the development of freely accessible, open-source software, in order to favour standardization, reproducibility, and exchangeability of metabolomics data (CASTILLO *et al.*, 2011; PEDRIOLI *et al.*, 2004; SMITH *et al.*, 2006). Various open-source software programs are currently in use (mainly for LC-MS), some of which have been developed especially for metabolomics, while others have been adapted from data processing software for proteomics (CASTILLO *et al.*, 2011). Because these programs are open-source, they are relatively easy to customize with few programming skills, dismissing the need for the development of a completely new algorithm for specific problems. Furthermore, open-source software also facilitates the exchange of algorithms and data with other laboratories involved in metabolomics research (CASTILLO *et al.*, 2011; SMITH *et al.*, 2006). Also the R environment, in which data analysis, including preprocessing, of this master's thesis was performed is a free and open-source software. For preprocessing R was extended with the MALDIquant R package (GIBB and STRIMMER, 2012).

As raw data are usually in the specific data format of a proprietary vendor, they have to be converted to an open format before preprocessing. Different tools are available and described in Castillo and coworkers (2011). MALDIquant was designed to be independent of any specific mass spectrometry hardware. However, through the package MALDIquantForeign (GIBB, 2013) input of binary data files or complete folder hierarchies from Bruker flex series instruments is possible (GIBB and STRIMMER, 2012). As all data of this master's thesis were acquired on a MALDI-TOF/TOF mass spectrometer (ultrafleXtreme, Bruker Daltonics), raw data could be imported into R without previously converting them into an open format. Import of raw data including preprocessing therefore took only a few minutes, which further facilitated data analysis and decreased the time used for it.

Preprocessing comprises all steps that have the reduction of complexity and extraction of the most important features of the raw data as objectives. This includes noise filtering, peak detection, deisotoping, normalization, alignment, and identification (CASTILLO *et al.*, 2011). Except for peak identification and deisotoping, an algorithm which groups isotopic peaks that originate from one and the same compound, which otherwise would be considered as various different compounds and would consequently falsify data interpretation (CASTILLO *et al.*, 2011), all preprocessing steps are included in MALDIquant. The preprocessing workflow that

was applied in this master's thesis (a spectrum acquired in negative ion mode from an extract of *Lychnophora candelabrum* was taken as an example) is illustrated in Figure 11. It included variance stabilization (by square root transformation of the absolute intensities), noise filtering (called smoothing in MALDIquant) by the moving average algorithm, baseline subtraction by the Top Hat algorithm, TIC normalization and peak detection.

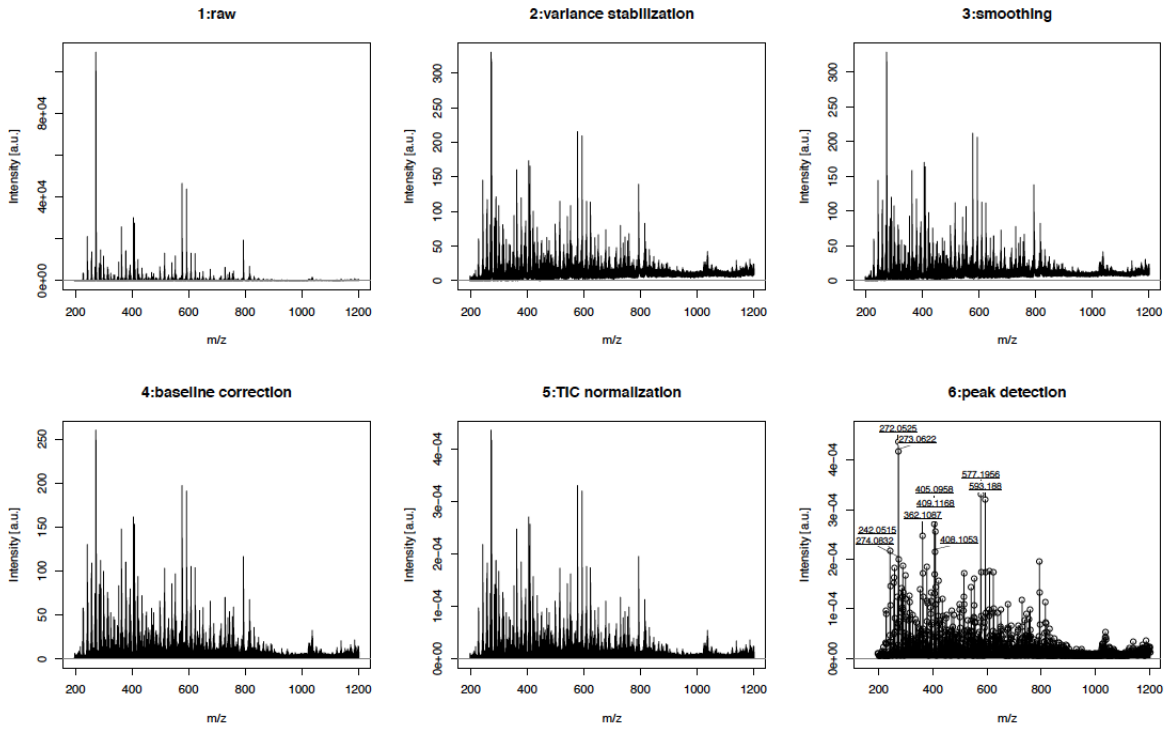


Figure 11: Example of MALDIquant output: **1** raw spectrum; **2** variance-stabilized spectrum; **3** smoothed spectrum; **4** baseline-corrected spectrum; **5** normalized spectrum; **6** preprocessed spectrum with detected and labeled peaks.

Noise filtering separates compound signals from background signals, whilst peak detection identifies the correct form of a compound signal and estimates its intensity. Normalization corrects systematic variation and enables the direct comparison of different samples (CASTILLO *et al.*, 2011). The normalization type that was applied in this master's thesis was the total ion current (TIC) normalization. According to Murray *et al.* (2013) the total ion current is the sum of all separate ion currents carried by the ions of different *m/z* contributing to a complete mass spectrum. The total ion current normalization is a global normalization type, which sets the TIC of every analysed spectra to one (GIBB and STRIMMER, 2012) and like this allows comparison of intensities between different spectra. One and the same compound may appear at slightly different *m/z* values. In order to compare different features between samples, peaks have to be aligned before interpretation. In MALDIquant landmark peaks are

identified that occur in most spectra and a non-linear warping function is then computed for each spectrum by fitting a local regression to the matched reference peaks (GIBB and STRIMMER, 2012). Figure 12 illustrates the alignment of four spectra of species of the genus *Lychnophora* by MALDIquant.

Both normalization and alignment can be performed with or without the addition of an internal standard (CASTILLO *et al.*, 2011; SMITH *et al.*, 2006). Indeed, a frequently discussed question in studies on metabolomics is whether or not an internal standard should be added. Smith and coworkers (2006) have highlighted various drawbacks of the addition of internal standards during alignment in metabolomic studies. According to these authors, working with internal standards wrongly assumes that deviations in retention time are linear. Furthermore, an additional step in sample preparation is required, which is more time-consuming and thus unfavorable for metabolomics experiments. Finally, internal standards may mask other experimentally relevant analytes (SMITH *et al.*, 2006). Instead, Smith and collaborators (2006) encourage the use of an algorithm in XCMS, a program developed for LC-MS data preprocessing that takes into consideration the nonlinearity of retention time deviation without using internal standards. Also in this master's thesis no internal standards were used due to the afore mentioned disadvantages.

Van den Berg and coworkers (2006) have tested different data preprocessing methods and have come to the conclusion that these methods greatly influence the outcome of data analysis. As in the case of all the other metabolomics areas, preprocessing of metabolomics data is still in its infancy, and new algorithms and/or software are continuously being developed by various research groups (SMITH *et al.*, 2006). It is therefore essential to store raw data, in order to be able to perform analysis at a later point in time using newly developed algorithms.

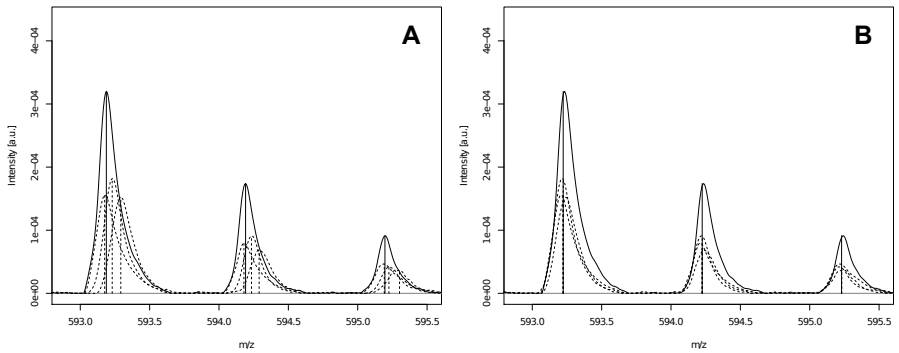


Figure 12: Example of peak alignment by MALDIquant: **A** four unaligned peaks; **B** four aligned peaks.

1.4.2 Multivariate data analysis

Metabolomics data sets are usually multivariate, which means they include more variables (e.g. ions) than samples (TRYGG *et al.*, 2006). For analysis of those highly collinear and complex data sets multivariate statistical methods are applied (TRYGG *et al.*, 2006). Multivariate statistical methods are procedures that observe and analyse more than one variable at a time (NORDSTRÖM, 2008). Univariate approaches (e.g., student's t-test) and traditional statistical methods (e.g., multiple linear regression) assume independent variables, and are therefore inappropriate for the analysis of data resulting from metabolomics studies (TRYGG *et al.*, 2006). There are two sub-classes of multivariate statistical methods, namely non-supervised classification and supervised classification methods (GOODACRE *et al.*, 2004; HALL, 2006; NORDSTRÖM, 2008). Non-supervised classification methods do not associate samples with labels; that is, no assumption about the existence of groups among the samples is made. In contrast, supervised classification methods use external information related to the data set to make predictions through multivariate calibration or by performance of discriminant analysis (NORDSTRÖM, 2008; WEBB, 2002). According to Hall (2006), unsupervised approaches for discriminatory analyses such as principal component analysis (PCA) and hierarchical clustering (HCA) or supervised approaches like partial least squares (PLS) or Soft Independent Modeling of Class Analogy (SIMCA) are the simplest and most widely employed classification methods in metabolomics studies (HALL, 2006). Trygg and coworkers (2007) further mention orthogonal-PLS (OPLS) as a more advanced method. They point out that more advanced multivariate methods such as SIMCA, PLS-DA (PLS discriminant analysis), and/or OPLS-DA are rarely utilized in studies in metabonomics (metabolomics for disease or treatment observation in tissues and biological fluids [DUNN and ELLIS, 2005]). Although PCA is an appropriate tool for any work on metabonomics (as well as metabolomics) and is always recommended for the attainment of a rapid overview of the information contained in the data, it is important that various and more advanced multivariate methods are also considered, because they provide important information that is not contained in PCA (Trygg *et al.*, 2007). In Johnson and collaborators (2003), for instance, it was not possible to discriminate between salt-stressed tomatoes and the control, non-salt-stressed tomatoes by PCA. With discriminant function analysis (DFA), a supervised method, instead it was possible to differentiate between salt-treated and non-treated fruits.

Van den Berg and coworkers (2009) have presented further variations of PCA, consensus principal component analysis (CPCA), and canonical correlation analysis (CCA), by means of which data analysis can be focused on a group of metabolites that is of special interest. It was shown that the application of these advanced statistical methods lead to better biological

interpretation of data in *E. coli*.

There are attempts in the metabolomics research community to uniform description of metabolomics experiments from the experimental part up to data analysis. Goodacre (2007) and coworkers report minimum reporting standards for data analysis in metabolomics.

Multivariate data analysis applied in this master's thesis included 4 types of non-supervised classification namely PCA and three types of cluster analyses. A short introduction to those methods is therefore given in the following.

Principal component analysis

Principal component analysis (PCA) as well as the partial least squares (PLS) method belong to the projection-based methods. In these methods, each row (individual sample) of data listed in a table (so called data matrix or table X) with *n* rows (plant extracts) and *k* columns (ions), is represented as a point in a *k*-dimensional space, where its position is given by the coordinates (values in each of the *k* columns) (NORDSTRÖM, 2008; STEUER *et al.*, 2003; TRYGG *et al.*, 2006). Each sample creates one point, thereby leading to a swarm of points in a multidimensional space. Samples (points) that lie close to each other are chemically more similar than points lying far apart. Projection-based methods convert this multidimensional data table into a low-dimensional model plane that usually consists of two to five dimensions and gives a summary of the variation in the data table X (swarm of points) (NORDSTRÖM, 2008; TRYGG *et al.*, 2006). Figure 13 was adapted from Trygg an collaborators (2006) and illustrates data analysis by multivariate projection-based methods.

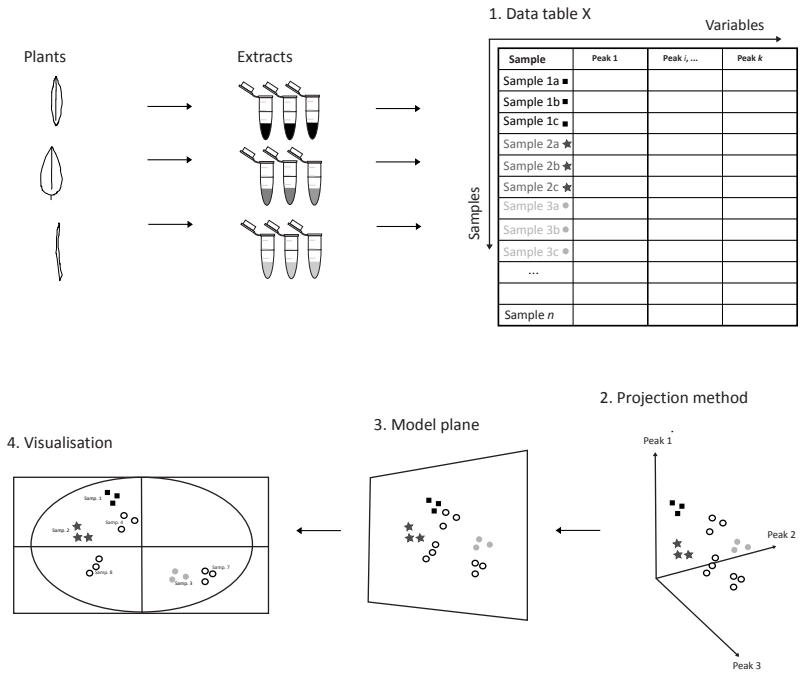


Figure 13: Illustration of projection-based methods based on Trygg and collaborators (2006): **1** each row in the data table X represents one sample, each of which contains three metabolites ($k = 3$); **2** in a $k = 3$ dimensional space each sample may be represented as one point. All samples together create a swarm of points; **3** projection-based methods such as PCA and PLS then convert the data into a representative low-dimenional model plane (here two-dimensional) that summarizes the variation in the data table X; **4** this model plane may further be visualized in scatter plots, which show groupings, trends, or outliers in the data.

In PCA, the largest variation is expressed in the first PC ($t1p1^T$), the second largest variation is expressed in the second PC ($t2p2^T$), and so on. Points in the data swarm have the largest distance from each other in the direction of the first principal component and principal components are mutually orthogonal (JOLLIFFE, 2002; TRYGG *et al.*, 2007). There are many algorithms for the accomplishment of PCA, each of which is suited for different applications. The most common algorithm used is the Eigenvector method (JOLLIFFE, 2002), which was also applied in this master's thesis. PCA results are presented in two different plots, namely the score plot and the loading plot (MORITZ and JOHANSSON, 2008; TRYGG *et al.*, 2006). The sample points on the low dimensional model plane described above are the weighted averages of all the variables (the ions) in the data table X and are designated scores T. This plot is therefore also called the score plot. It provides an overview of all the samples in X and of how they relate to each other, thus affording groupings of samples (clusters). Trends and outliers (deviating samples) can be clearly distinguished, because they show up as distinct clusters. Hence, in the score plot the relation among the samples are shown (KRZANOWSKI, 2000; MORITZ and JOHANSSON, 2008; TRYGG *et al.*, 2007; TRYGG *et al.*, 2006). To understand the reason for the observed patterns; i.e., to find out which variables (ions) are responsible for a distinct pattern, the loading plot P can be used. As directions in the loading plot correspond to directions

in the score plot, it is possible to identify which variables (loadings), or ions respectively, are responsible for the separation of objects (scores), the plant samples (KRZANOWSKI, 2000; MORITZ and JOHANSSON, 2008; TRYGG *et al.*, 2006). The part of the X matrix, which cannot be explained by the model plane is known as residuals E. E describes the distance between each sample point in the *k*-dimensional space and its projection on the plane (TRYGG *et al.*, 2007; TRYGG *et al.*, 2006). Consequently, all of the variation in the data table X can be described as (TRYGG *et al.*, 2007):

$$X = TP^T + E = t1p1^T + t2p2^T + E \quad (1)$$

Recently, Kuhnert and coworkers (2011) have compared several PCA processing parameters, in order to distinguish between different green coffee beans by LC-ESI-TOF-MS. They concluded that the grouping of samples does not depend on PCA parameters, but they found out that the identification of metabolites responsible for grouping patterns varies significantly between different PCA parameters. The partial least squares method (PLS) can be applied when additional knowledge about each sample exists. This additional information about the X matrix is denominated Y matrix. The Y matrix may include quantitative (e.g. , glucose concentration) as well as qualitative information (e.g. , morphological trait, such as leaf shape). PLS is analogous to PCA, except that the additional information in the Y matrix is used to show the variation that is related to Y in X instead of showing the overall variation (TRYGG *et al.*, 2007; TRYGG *et al.*, 2006).

Clustering

Clustering is defined as discipline, which aims to reveal groups or clusters of similar entities in the data (MIRKIN, 2005). There exist various clustering algorithms in the literature, which can be further subdivided in partitioning methods and hierarchical methods. In this master's thesis two partitioning methods, partitioning around medoids (PAM) and dynamic tree cut, a recently developed tree cut method (LANGFELDER *et al.*, 2008) and one hierarchical method, the hierarchical cluster analysis were performed. In both methods the input data is a dissimilarity matrix. In such a matrix distances between each pair of objects *i* and *j* are computed, in order to quantify their degree of dissimilarity. Distance measures, which are used to calculate the dissimilarity between two objects are numerous and depend upon the data type. In this master's thesis clustering was performed on a binary data set. Binary variables are variables that only have two possible outcomes. In the present case this would be either the presence (designated as 1) or the absence of a particular ion (designated as 0). There are two types of binary variables. Asymmetric and symmetric ones. In symmetric binary variables it is

assumed that both variables are equally valuable and carry the same weight. For example two plants showing the same ion in their mass spectrum would be accounted equally similar, as two plants that do not show this particular ion. In asymmetric binary data however it is assumed that not both variables carry the same weight. Two plants showing the same ion in their mass spectrum would be correctly accounted more similar to each other than two plants that don't show the ion. In the present study the Jaccard coefficient, which is most commonly used when working with asymmetric binary data was used to calculate the dissimilarity matrix of the data set. Given the association table in Table 13, the Jaccard Coefficient is defined as (KAUFMAN and ROUSSEEUW, 1990):

$$\frac{b+c}{a+b+c} \tag{2}$$

Partitioning methods classify data into k groups, where k is defined by the user. Hierarchical algorithms however deal with all values of k in the same run (KAUFMAN and ROUSSEEUW, 1990). Hall and collaborators (2006) state that the cluster algorithms that are most often employed in metabolomics belong to the hierarchical techniques. A hierarchy of a treelike structure can be produced by these methods, the so-called dendrogram. Such tree diagrams are also commonly seen in taxonomic studies and phylogenetics (NORDSTRÖM, 2008; GOODACRE *et al.*, 2004).

Table 13: Association table. Source: KAUFMAN and ROUSSEEUW, 1990.

		object *j*		
		1	0	
object *i*	1	*a*	*b*	*a* + *b*
	0	c	*d*	*c* + *d*
		a + *c*	*b* + *d*	

There exist two kinds of hierarchical algorithms, agglomerative methods and divisive methods. Agglomerative methods start when all objects are apart and two clusters are continuously merged until only one is left. Divisive methods however start when all objects are together and splits it continuously up (HAIR, 1995; KAUFMAN and ROUSSEEUW, 1990; NORDSTRÖM, 2008). In the hierarchical cluster analysis that was performed in this master's thesis an agglomerative cluster method was performed. There exist a variety of different agglomerative clustering methods including for example Ward's method, single linkage, complete linkage, unweighted pair-group average, McQuitty's similarity analysis, median method and centroid method besides others. Each method uses a different clustering algorithm and has its own advantages and disadvantages. There is no single "best" clustering method, and the cho-

sen clustering method depends highly on the data set (KAUFMAN and ROUSSEEUW, 1990). Out of this reason various clustering methods implemented in R were tested for the data set presented in this master's thesis.

10 CONCLUSIONS

To develop a protocol for metabolic fingerprinting by MALDI-MS, which can be used for taxonomic classification of a variety of plant species is a very ambitious goal. Several taxonomic tools, based on morphological traits up to genome analysis showed various drawbacks in correct taxonomic classification of plants. It is therefore not surprising that various difficulties also remain in the taxonomic classification of plants based on their metabolome. Hence, metabolomics applied to plant systematics may rather serve as additional tool besides others, than provide independent results. In order to judge whether results obtained from plant metabolomics studies provide a reliable approach for plant systematics limitations of chemotaxonomy as well as limitations present during sampling, in the chosen analytical method and data analysis should be considered.

The major challenge in chemotaxonomy has always been the fact that particular secondary metabolites can occur in plant groups that are not related in a phylogenetic context. Same secondary metabolites occur in non-related groups due to convergent evolution, due to the fact that genes encoding the enzymes of secondary metabolism might be widely distributed in the plant kingdom, but switched on or off in a certain context or due to the presence of endophytic fungi, which are able to produce secondary metabolites. As a consequence, similar metabolite profiles of two taxa cannot necessarily be taken to imply a monophyletic relationship. A further confounding factor is that external pressure may influence the production of secondary metabolites. Phylogenetic unrelated plants that are confronted with same environmental factors, may develop similar stratagems to combat the influencing factor, meaning they would produce similar chemical compounds (WINK *et al.*, 2010). Wink and collaborators state that these confounding factors have a greater impact at higher taxonomic levels and influence of those on lower taxonomic levels may be considered minor. Even if a broader picture of a plant's metabolites is obtained in metabolomics studies than in classical chemotaxonomic studies, where only few substances are analysed, it may not automatically be implied that classification based on plant's metabolomes lead to a taxonomy correspondent classification.

In this master's thesis a protocol for metabolic fingerprinting by MALDI-MS as well as a subsequent data analysis pipeline that is able to detect in a robust way ions present in all

replicate samples was developed. Classifications that were obtained from the metabolic fingerprints of 24 plant species belonging to four different tribes, three subfamilies, two families and two orders based on binary hierarchical cluster analyses showed similarity to taxonomic classifications of the plants. Despite this similarity however, it was also possible to point out several difficulties that are inherent to plant metabolomic studies and specifically, which are inherent to MALDI-MS applied in plant metabolomic studies.

A first essential step is the harvesting of the plant material. If the study is conducted with plant species that were collected in their natural habitat, at different points in time, with different developmental stages and at different places as was the case in this master's project, it should be kept in mind that the conditions under which these plants were grown could have varied to a great extent. A quantitative metabolomic analysis would therefore be questionable as the quantitative differences found in the metabolome also could be attributed to external factors such as climate, soil condition or water stress and not be inherent to the species genetic composition. In this master's thesis it was therefore opted to base data analysis merely on the presence or absence of specific ions.

Metabolomics aims to analyse the metabolome as a whole, however there are various steps during sample preparation, sample analysis and data analysis that considerably decrease the insight into a plant's metabolome. The first decrease already occurs during sample preparation. Depending on which solvent is used (as outlined in Section 9.5) different metabolites are extracted and hence a different view of the metabolome is obtained. Subsequently the chosen analytical method further decreases this amount as never all metabolites can be detected with one single analytical technique. As was shown MALDI-MS is able to detect different metabolites not only depending on the chosen ionisation mode but also on the chosen matrix substance. To capture an as large variety of metabolites as possible it was proposed to perform analysis in both ionisation modes and with various different matrix substances. On the other hand however this increases considerably the time that is used for data acquisition, which originally was thought to be the main advantage of MALDI-MS applied in plant metabolomics. Furthermore a major difficulty was to find an appropriate matrix substance for analysis of low molecular weight compounds, which can be applied to a variety of plant species and does not show ions that interfere with ions of the plant extract. The exploration of a matrix substance that can be used in both ionisation modes, which additionally shows no interfering ions, may be applied to various different plant species and with which an as wide variety of metabolites as possible may be detected will therefore be a key factor in making MALDI-MS feasible for metabolic fingerprinting.

Last but not least also data analysis further reduces the amount of ions that actually will be considered for classification. It was shown that several ions do not constantly show neither in

all technical replicates nor in all sampling replicates and therefore can't be considered for data analysis. Furthermore it was also shown that any kind of preprocessing has a large influence on the obtained classification when working with a binary data set. In order to get reproducible classification patterns a subset of ions would have to be defined with which standard ideal preprocessing methods could be chosen. However the total number of possible subsets out of the set of all detected ions is so big that not all possibilities can be tested. The in-house R-script that was developed in this master's thesis was only able to test a very small percentage of the total number of possibilities. It is therefore possible that the subset of ions that would have led to a complete taxonomy correspondent classification remained undetected. Application of further multivariate data analysis methods and the development of further algorithms in combinatorial optimization would therefore be necessary in order to improve obtained classifications and in order to increase reliability of the obtained results. Additionally a deisotoping algorithm during preprocessing of the spectra might further lead to more reliable results.

The ions that were found to be responsible for classifications that showed most similarity with taxonomy showed very low intensities, furthermore also no class specific ions could be detected. It remains unknown if the metabolites responsible for the taxonomic classification of the analysed plant species are only present at very low concentrations and compounds that are more abundant are equal in all plant species. If these compounds only could be ionised to a very low extent with the chosen matrix substances or if these substances are not at all detectable with MALDI-MS. If they were at all present in the plant extract, or if they could not be extracted with the chosen solvent, or at last, if they do exist at all.

As pointed out here, there are still various open questions and still a long way to go in making MALDI-MS a feasible, rapid and confident tool for metabolic fingerprinting applied for taxonomic classification of plants. Also with metabolic fingerprinting, as stated in Stuessy (2009) taxonomy remains dynamic, beautiful, frustrating, and challenging all at the same time. The results presented in this master's thesis are a first attempt in applying MALDI-MS for metabolic fingerprinting for plant taxonomic purposes and might present essential knowledge needed for further studies. Developed algorithms in the R environment are included in the appendix, proposed data analysis workflows can therefore be directly applied and further improved during future studies.

11 REFERENCES

ABEGAZ, B. M.; KEIGE, A. W.; DIAZ, J. D.; HERZ, W. Sesquiterpene lactones and other constituents of *Vernonia* species from Ethiopia. Phytochemistry, v. 37, pp. 191-196, 1994.

ADESOGAN, E. K.; OKUNADE, A. L. A new flavone from *Ageratum conyzoides*. Phytochemistry, v. 18, pp. 1863-1864, 1979.

ADESOGAN, E. K.; OKUNADE, A. L. Structure of conyzorigun, a new chromone from *Ageratum conyzoides*. Journal of the Chemical Society, Chemical Communications, p. 152, 1978.

ADLER, D. and MURDOCH, D. rgl: 3D visualization device system (OpenGL). R package version 0.92.894. `http://CRAN.R-project.org/package=rgl`. 2012.

AHARONI, A.; DE VOS, R. C. H.; VERHOEVEN, H. A.; MALIEPAARD, C. A.; KRUPPA, G.; BINO, R.; GOODENOWE, D. B. Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry. OMICS, v. 6, pp. 217-234, 2002.

ALLEN, J.; DAVEY, H. M.; BROADHURST, D.; HEALD, J. K.; ROWLAND, J. J.; OLIVER, S. G.; KELL, D. B. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. Nature Biotechnology, v. 21, pp. 692-696, 2003.

ALVAREZ, L.; MARQUINA, S.; VILLARREAL, M. L.; ALONSO, D.; ARANDA, E.; DELGADO, G. Bioactive polyacetylenes from *Bidens pilosa*. Planta Medica, v. 62, pp. 355-357, 1996.

ALVES, K. C. M.; GOBBO-NETO, L.; LOPES, N. P. Sesquiterpene lactones and flavonoids from *Lychnophora reticulata* Gardn. (Asteraceae). Biochemical Systematics and Ecology, v. 36, pp. 434-436, 2008.

ALVES, T. M. A.; NAGEM, T. J.; CARVALHO, L. H.; KRETTLI, A. U.; ZANI, C. L. Antiplasmodial Triterpene from *Vernonia brasiliana*. Planta Medica, v. 63, pp. 554-555, 1997.

APG III. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Botanical Journal of the Linnean Society, v. 161, pp.

105-121, 2009.

AYORINDE, F. O.; HAMBRIGHT, P.; PORTER, T. N.; KEITH, Q. L. Jr. Use of *meso*-Tetrakis(pentafluorophenyl)porphyrin as a Matrix for Low Molecular Weight Alkylphenol Ethoxylates in Laser Desorption/Ionization Time-of-flight Mass Spectrometry. Rapid Communications in Mass Spectrometry, v. 13, pp. 2474-2479, 1999.

BAKER, J. G. Compositae I. Vernoniaceae. In Flora Brasiliensis, edited by MARTIUS, C. F. P. and EICHLER, A. G., Lipsiae Monachii, v. 6, pp. 1-180, 1873.

BALLARD, R. Bidens pilosa complex (Asteraceae) in North and Central America. American Journal of Botany, v. 73, pp. 1452-1465, 1986.

BASTOS *et al.* (1987). In: KELES, L. C.; DE MELO.; N. I.; AGUIAR, G. de P.; WAKABAYASHI, A. L.; DE CARVALHO, C. E.; CUNHA, W. R.; CROTTI, A. E. M. Lychnophorinae (Asteraceae): A survey of its chemical constituents and biological activities. Química Nova, v. 33, pp. 2245-2260, 2010.

BARTOLOME, A. P.; VILLASEÑOR, I. M.; YANG, W. C. *Bidens pilosa* L. (Asteraceae): Botanical Properties, Traditional Uses, Phytochemistry, and Pharmacology. Evidence-Based Complementary and Alternative Medicine, v. 2013, 51 pages, 2013.

BAUER, R.; REDL, K.; DAVIS, B. Four polyacetylene glucosides from *Bidens campylotheca*. Phytochemistry, v. 31, pp. 2035-2037, 1992.

BAZON, J. N.; LOPES, J. L. C.; VICHNEWSKI, W.; DIAS, D. A.; NAGAMITI, K.; CUNHA, W. R.; HERZ, W. Cadinanolides and other constituents from *Vernonia fruticulosa* and *Vernonanthura discolor*. Phytochemistry, v. 44, pp. 1535-1536, 1997.

BEAUVERD, G. Un nouveau *Lychnophora* brésilien. Bulletin de la Société Botanique de Genève, v. 5, pp. 241-242, 1913.

BEECHER, C. W. W. The human metabolome. In Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis, edited by HARRIGAN, G. G.; GOODACRE, R. Kluwer Academic Publishers, pp. 311-319, 2003.

BENGTSSON, H. matrixStats: Methods that apply to rows and columns of a matrix. R package version 0.8.1. `http://CRAN.R-project.org/package=matrixStats`. 2013.

BENHURA, M. A. N.; CHITSIKU, I. C. The extractable β-carotene content of guku (*Bidens pilosa*) leaves after cooking, drying and storage. International Journal of Food Science and Technology, v. 32, pp. 495-500, 1997.

BENTHAM, G. Compositae. In Genera plantarum, edited by BENTHAM, G. and HOOKER J. D. Reeve & co, London, v. 2, pp. 163-533, 1873a.

BENTHAM, G. Notes on the Classification, History, and Geographical Distribution of Compositae. Journal of the Linnean Society of London, Botany, v.13, pp. 335-577, 1873b.

BEZERRA, M. Z. B.; ANDRADE-NETO, M.; DE FREITAS, R. M. The essential oil of *Porophyllum ruderale* Cass. (Asteraceae). Journal of Essential Oil Research, v. 14, pp. 14-15, 2002.

BINO, R. J.; HALL, R. D.; FIEHN, O.; KOPKA, J.; SAITO, K.; DRAPER, J.; NIKOLAU, B. J.; MENDES, P.; ROESSNER-TUNALI, U.; BEALE, M. H.; TRETHEWEY, R. N.; LANGE, B. M.; WURTELE, E. S.; SUMNER, L. W. Potential of metabolomics as a functional genomics tool. Trends in Plant Science, v. 9, pp. 418-424, 2004.

BOHLMANN, F.; BARUAH, R. N.; DOMINGUEZ, X. A further dithienyl derivative from *Porophyllum scoparia*. Planta Medica, pp. 77-78, 1985.

BOHLMANN, F.; LUDWIG, G. W.; JAKUPOVIC, J.; KING, R. M.; ROBINSON, H. A daucanolide and further farnesene derivatives from *Ageratum fastigiatum*. Phytochemistry, v. 22, pp. 983-986, 1983a.

BOHLMANN, F.; ZDERO, C.; KING, R. M.; ROBINSON, H. Thymol derivatives from *Porophyllum riedelii*. Phytochemistry, v. 22, pp. 1035-1036, 1983b.

BOHLMANN, F.; ZDERO, C.; ROBINSON, H.; KING, R. M. Germacranolides from *Lychnophora* species. Phytochemistry, v. 21, pp. 1087-1091, 1982a.

BOHLMANN, F.; MATHUR, R.; JAKUPOVIC, J.; GUPTA, R. K.; KING, R. M.; ROBINSON, H. Furanoheliangolides and other compounds from *Calea hymenolepis*. Phytochemistry, v. 21, pp. 2045-2048, 1982b.

BOHLMANN, F.; AHMED, M.; KING, R. M.; ROBINSON, H. Labdane and eudesmane derivatives from *Ageratum fastigiatum*. Phytochemistry, v. 20, pp. 1434-1435, 1981a.

BOHLMANN, F.; ZDERO, C.; KING, R. M.; ROBINSON, H. Heliangolides, and nerolidol and *p*-hydroxyacetophenone derivatives from *Calea* species. Phytochemistry, v. 20, pp. 1643-1647, 1981b.

BOHLMANN, F.; FRITZ, U.; KING, R. M.; ROBINSON, H. Fourteen heliangolides from *Calea* species. Phytochemistry, v. 20, pp. 743-749, 1981c.

BOHLMANN, F.; MÜLLER, L.; KING, R. M.; ROBINSON, H. A guaianolide and other con-

stituents from *Lychnophora* species. Phytochemistry, v. 20, pp. 1149-1151, 1981d.

BOHLMANN, F.; JAKUPOVIC, J.; ROBINSON, H.; KING, R. M. A dithienylacetylene from *Porophyllum ruderale*. Phytochemistry, v. 19, p. 2760, 1980.

BOHLMANN, F.; JAKUPOVIC, J. Neue germacranolide aus *Calea urticifolia*. Phytochemistry, v. 18, pp. 119-123, 1979.

BOHLMANN, F.; ZDERO, C. Neue Germacrolide aus *Calea zacatechichi*. Phytochemistry, v. 16, pp. 1065-1068, 1977.

BOMFIM-PATRÍCIO, M. C.; SALATINO, A.; MARTINS, A. B.; WURDACK, J. J. Flavonoids of *Lavoisiera*, *Microlicia* and *Trembleya* (Melastomataceae) and their taxonomic meaning. Biochemical Systematics and Ecology, v. 29, pp. 711-726, 2001.

BOWERS, W. S.; OHTA, T.; CLEERE, J. S.; MARSELLA, P. A. Discovery of Insect Anti-Juvenile Hormones in Plants. Science, v. 193, pp. 542-547, 1976.

BRANDÃO, M. G. L.; NERY, C. G. C.; MAMÃO, M. A. S.; KRETTLI, A. U. Two methoxylated flavone glycosides from *Bidens pilosa*. Phytochemistry, v. 48, pp. 397- 399, 1998.

BREMER, K. Major clades and grades of the Asteraceae. In Compositae: Systematics. Proceedings of the International Compositae Conference, edited by HIND, D. J. N. and BEENTJE H. J. Kew Royal Botanic Gardens, v.1, 1996.

BREMER, K. Asteraceae: Cladistics and Classification. Timber Press. Portland, Oregon, 1994.

BRINDLE, J. T.; ANTTI, H.; HOLMES, E.; TRANTER, G.; NICHOLSON, J. K.; BETHELL, H. W. L.; CLARKE, S.; SCHOFIELD, P. M.; MCKILLIGIN, E.; MOSEDALE, D. E.; GRAINGER, D. J. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. Nature Medicine, v. 8, pp. 1439-1444, 2002.

BROWN, M.; DUNN, W. B.; ELLIS, D. I.; GOODACRE, R.; HANDL, J.; KNOWLES, J. D.; O'HAGAN, S.; SPASIĆ, I.; KELL, D. B. A metabolome pipeline: from concept to data to knowledge. Metabolomics, v. 1, pp. 39-51, 2005.

Bruker Daltonics MALDI Theory. Mass spectrometry. 2004

Bruker Daltonics Application Note Number MT-97. MALDI Imaging of Proteins at 20 μm Resolution on the ultrafleXtreme.

BURKILL, H. M. (1985). In: OKUNADE, A. L. *Ageratum conyzoides* L. (Asteraceae). Fitoterapia, v. 73, pp. 1-16, 2002.

BUSKUHL, H.; DE OLIVEIRA, F. L.; BLIND, L. Z.; DE FREITAS, R. A.; BARISON, A.; CAMPOS, F. R.; CORILO, Y. E.; EBERLIN, M. N.; CARAMORI, G. F.; BIAVATTI, M. W. Sesquiterpene lactones from *Vernonia scorpioides* and their *in vitro* cytotoxicity. Phytochemistry, v.71, pp. 1539-1544, 2010.

CANTONWINE, E. G.; DOWNUM, K. R. Phenylheptatriyne variation in *Bidens alba* var. *radiata* leaves. Journal of Chemical Ecology, v. 27, pp. 313-326, 2001.

CASSINI, A. H. G. Suite de Sixième mémoire sur la famille des Synanthérées, contenant les caractères des tribus. Journal de Physique, de Chimie, d'Histoire Naturelle et des Arts, v. 88, pp. 189-204, 1819.

CASSINI, A. H. G. Bulletin des Sciences, par la Société Philomathique de Paris. p.151, 1817.

CASTILLO, S.; GOPALACHARYULU, P.; YETUKURI, L.; ORESIC, M. Algorithms and tools for the preprocessing of LC-MS metabolomics data. Chemometrics and Intelligent Laboratory Systems, v. 108, pp. 23-32, 2011.

CASTRO, V.; TAMAYO-CASTILLO, G.; JAKUPOVIC, J. Sesquiterpene lactones and other constituents from *Calea prunifolia* and *C.peckii*. Phytochemistry, v. 28, pp. 2415-2418, 1989.

CASTRO, V.; JAKUPOVIC, J.; BOHLMANN, F. A new type of sesquiterpene and acorane derivative from *Calea prunifolia*. Journal of Natural Products, v. 47, pp. 802-808, 1984.

CATCHPOLE, G. S.; BECKMANN, M.; ENOT, D. P.; MONDHE, M.; ZYWICKI, B.; TAYLOR, J.; HARDY, N.; SMITH, A.; KING, R. D.; KELL, D. B.; FIEHN, O.; DRAPER, J. Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. Proceedings of the National Academy of Sciences, v. 102, pp. 14458-14462, 2005.

CHAMBERS, J. M.; CLEVELAND, W. S.; KLEINER, B.; TUKEY, P. A. Graphical Methods for Data Analysis. Wadsworth and Brooks, Cole, 1983.

CHANG, S.; CHIANG, Y.; CHANG, C. L.; YEH, H.; SHYUR, L.; KUO, Y.; WU, T.; YANG, W. Flavonoids, centaurein and centaureidin from *Bidens pilosa*, stimulate IFN-γ expression. Journal of Ethnopharmacology, v. 112, pp. 232-236, 2007.

CHANG, M. H.; WANG, G. J.; KUO, Y. H.; LEE, C. K. The low polar constituents from *Bidens pilosa* L. var. minor (Blume) Sherff. Journal of the Chinese Chemical Society, v. 47, pp. 1131- 1136, 2000.

CHEN, H. VennDiagram: Generate high-resolution Venn and Euler plots. R package version

1.5.1. `http://CRAN.R-project.org/package=VennDiagram`. 2012.

CHIANG, Y.; CHANGA, C. L.; CHANGA, S.; YANG, W.; SHYUR, L. Cytopiloyne, a novel polyacetylenic glucoside from *Bidens pilosa*, functions as a T helper cell modulator. Journal of Ethnopharmacology, v. 110, pp. 532-538, 2007.

CHIANG, Y. M.; CHUANG, D. Y.; WANG, S. Y.; KUO, Y. H.; TSAI, P. W.; SHYUR, L. F. Metabolite profiling and chemopreventive bioactivity of plant extract from *Bidens pilosa*. Journal of Ethnopharmacology, v. 95, pp. 409-419, 2004.

CHRISTENSEN, L. P.; LAM, J.; THOMASEN, T. A chalcone and other constituents of *Bidens tripartitus*. Phytochemistry, v. 29, pp. 3155-3156, 1990.

COHEN, L. H.; GUSEV, A. I. Small molecule analysis by MALDI mass spectrometry. Analytical and Bioanalytical Chemistry, v. 373, pp. 571-586, 2002.

COILE, N. C.; JONES, Jr. S. B. *Lychnophora* (Compositae: Vernonieae), a genus endemic to the Brazilian planalto. Brittonia, v. 33, pp. 528-542, 1981.

COLLEVATTI, R. G. Nós realmente precisamos de *DNA Barcoding*? Natureza e Conservação, v.9, pp.105-110, 2011.

COTA, B. B.; DE OLIVEIRA, A. B.; VENTURA, C. P.; MENDONÇA, M. P.; BRAGA, F. C. Antimicrobial activity of plant species from a Brazilian Hotspot for Coservation Priority. Pharmaceutical Biology, v. 40, pp. 542-547, 2002.

COUTINHO, L. M. O conceito de bioma. Acta Botanica Brasilica, v. 20, pp. 13-23, 2006.

CRONQUIST, A. The compositae revisited. Brittonia, v. 29, pp. 137-153, 1977.

CVACKA, J.; SVATOS, A. Matrix-assisted laser desorption/ionization analysis of lipids and high molecular weight hydrocarbons with lithium 2,5-dihydroxybenzoate. Rapid Communications in Mass Spectrometry, v.17, pp. 2203-2207, 2003.

DA COSTA, F. J.; BANDEIRA, P. N.; ALBUQUERQUE, M. R. J. R. Constituintes químicos de *Vernonia chalybaea* MART. Química Nova, v. 31, pp. 1691-1695, 2008.

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 chemosystematics. Phytochemistry, v. 66, pp. 345-353, 2005.

DA SILVA, D. B. Atividade antialérgica e estudos químicos das espécies *Bidens gardneri* Bak. e *Bidens sulphurea* (Cav.) Sch. Bip. (Asteraceae). Doctoral thesis, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2009.

DE CANDOLLE, A. P. Prodromus systematis naturalis regni vegetabilis. Treuttel & Würz, v. 5, Paris, 1836.

DE MORAES, S. L.; GREGORIO, L. E.; TOMAZ, J. C.; LOPES, N. P. Rapid Screening and Identification of Polar Constituents from Brazilian Arnica *Lychnophora sp.* by LC-UV/DAD-ESI-MS and LC-UV/DAD-ESI-MS/MS Analysis. Chromatographia Supplement, v. 69, pp. 157-165, 2009.

DE QUEIROZ, K.; GAUTHIER, J. Phylogeny as a Central Principle in Taxonomy: Phylogenetic Definitions of Taxon Names. Systematic Biology, v. 39, pp. 307-322, 1990.

DE VASCONCELOS, M. F. O que são campos rupestres e campos de altitude nos topos de montanha de Leste do Brasil? Revista Brasileira de Botânica, v. 34, pp. 241-246, 2011.

DE VOS, R. C. H.; MOCO, S.; LOMMEN, A.; KEURENTJES, J. J. B.; BINO, R. J.; HALL, R. D. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. Nature Protocols, v. 2, pp.778-791, 2007.

DETTMER, K.; ARONOV, P. A.; HAMMOCK, B. D. Mass Spectrometry-based Metabolomics. Mass Spectrometry Reviews, v. 26, pp. 51-78, 2007.

DO NASCIMENTO, A. M.; SALVADOR, M. J.; CANDIDO, R. C.; DE ALBUQUERQUE, S.; DE OLIVEIRA, D. C. R. Trypanocidal and antifungal activities of *p*-hydroxyacetophenone derivatives from *Calea uniflora* (Heliantheae, Asteraceae). Journal of Pharmacy and Pharmacology, v. 56, pp. 663-669, 2004.

DO NASCIMENTO, A. M.; SOUSA e SILVA, F.; DE OLIVEIRA, D. C. R. Constituents of *Calea platylepis* Sch. Bip. ex Baker. Biochemical Systematics and Ecology, v. 30, pp. 993-996, 2002.

DONG, X.; CHENG, J.; LI, J.; WANG, Y. Graphene as a Novel Matrix for the analysis of Small Molecules by MALDI-TOF MS. Analytical Chemistry, v. 82, pp. 6208-6214, 2010.

DOS SANTOS, M. D.; GOBBO-NETO, L.; ALBARELLA, L.; DE SOUZA, G. E. P.; LOPES, N. P. Analgesic activity of di-caffeoylquinic acids from roots of *Lychnophora ericoides* (Arnica da serra). Journal of Ethnopharmacology, v.96, pp.545-549, 2005.

DOWNUM, K. R.; KEIL, D. J.; RODRIGUEZ, E. Distribution of acetylenic thiophenes in the Pectidinae. Biochemical Systematics and Ecology, v. 13, pp. 109-113, 1985.

DOWNUM, K. R.; TOWERS, G. H. N. Analysis of thiophenes in the Tageteae (Asteraceae) by HPLC. Journal of Natural Products, v. 46, pp. 98-103, 1983.

DUNN, W. B.; ELLIS, D. I. Metabolomics: Current analytical platforms and methodologies. Trends in Analytical Chemistry, v. 24, pp. 285-294, 2005.

EL-ANEED, A.; COHEN, A.; BANOUB, J. Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. Applied Spectroscopy Reviews, v. 44, pp. 210-230, 2009.

FARAG, M. A.; PORZEL, A.; WESSJOHANN, L. A. Comparative metabolite profiling and fingerprinting of medicinal licorice roots using a multiplex approach of GC-MS, LC-MS and 1D NMR techniques. Phytochemistry, v. 76, pp. 60-72, 2012.

FELLER, W. An Introduction to Probability Theory and Its Applications. 3rd ed., John and Wiley and Sons, New York, p. 34, 1968.

FERREIRA, Z. S.; ROQUE, N. F.; GOTTLIEB, O. R.; OLIVEIRA, F.; GOTTLIEB, H. E. Structural clarification of germacranolides from *Calea* species. Phytochemistry, v. 19, pp. 1481-1484, 1980.

FIASCHI, P.; PIRANI, J. R. Review of plant biogeographic studies in Brazil. Journal of Systematics and Evolution, v. 47, pp. 477-496, 2009.

FIEHN, O. Metabolomics - the link between genotypes and phenotypes. Plant Molecular Biology, v. 48, pp. 155-171, 2002.

FIEHN, O. Combining genomics, metabolome analysis, and biochemical modeling to understand metabolic networks. Comparative and Functional Genomics, v. 2, pp. 155- 168, 2001.

FIEHN, O.; KOPKA, J.; TRETHEWEY, R. N.; WILLMITZER, L. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. Analytical Chemistry, v. 72, pp. 3573-3580, 2000.

FINKEL, N. H.; PREVO, B. G.; VELEV, O. D.; HE, L. Ordered silicon nanocavity arrays in surface-assisted desorption/ionization mass sepctromtery. Analytical Chemistry, v. 77, pp. 1088-1095, 2005.

FONSCECA, M. C. M.; BARBOSA, L. C. A.; NASCIMENTO, E. A.; CASALI, V. W. D. Essential oils from leaves and flowers of *Porphyllum ruderale* (Jacq.) Cassini (Asteraceae). Journal of Essential Oil Research, v. 18, pp. 1-3, 2006.

FOX, J. and WEISBERG, S. An R Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage. `http://socserv.socsci.mcmaster.ca/jfox/Books/Companion`. 2011.

FRASER, D. P.; ENFISSI, E. M. A.; GOODFELLOW, M.; EGUCHI, T.; BRAMLEY, P. M. Metabolite profiling of plant carotenoids using the matrix-assisted laser desorption ionization time-of-flight mass spectrometry. The Plant Journal, v. 49, pp. 552-564, 2007.

FRISVAD, J. C.; ANDERSEN, B.; THRANE, U. The use of secondary metabolite profiling in chemotaxonomy of filamentous fungi. Mycological Research, v. 112, pp. 231-240, 2008.

FROHNE, D.; JENSEN, U. Systematik des Pflanzenreichs. Wissenschaftliche Verlagsgesellschaft GmbH, Stuttgart, 5. ed., 1998.

GANDERS, F. R.; BERBEE, M.; PIRSEYEDI, M. ITS base sequence phylogeny in *Bidens* (Asteraceae): evidence for the continental relatives of Hawaiian and Marquesan *Bidens*. Systematic Botany, v. 25, pp. 122-123, 2000.

GAO, W.; YANG, H.; QI, L. W.; LIU, E. H.; REN, M. T.; YAN, Y. T.; CHEN, J.; LI, P. Unbiased metabolite profiling by liquid chromatography-quadrupole time-of-flight mass spectrometry and multivariate data analysis for herbal authentication: Classification of seven *Lonicera* species flower buds. Journal of Chromatography A, v. 1245, pp. 109-116, 2012.

GARDNER, G. 1846. Contributions towards a flora of Brazil. London Journal of Botany, v. 5, pp. 235-237, 1846.

GARDNER, G. Contributions towards a flora of Brazil. London Journal of Botany, v. 4, pp. 113-129, 1845.

GARDNER, G. Characters of three new species of *Chresta*. London Journal of Botany, v. 1, pp. 238-241, 1842.

GASTEIGER, J. Chemoinformatics: a new field with a long tradition. Analytical and Bioanalytical Chemistry, v. 384, pp. 57-64, 2006.

GATES, P. J.; SANTOS, M. D.; GOBBO-NETO, L.; CAROLLO, C. A.; CROTTI, A. E. M.; VESSECCHI, R.; LOPES, N. P. Mass spectrometry ionisation techniques and applications for the analysis of organic compounds. In Current Methods in Medicinal Chemistry and Biological Physics, edited by CARLTON, A. and CARLOS, H. T. P. S. Research Signpost, Kerala, pp. 1-19, 2006.

GEISSBERGER, P.; SEQUIN, U. Constituents of *Bidens pilosa* L.: do the components found so far explain the use of this plant in traditional medicine? Acta Tropica, v. 48, pp. 251-261, 1991.

GEORGIEV, M. I.; ALI, K.; ALIPIEVA, K.; VERPOORTE, R.; CHOI, Y. H. Metabolic differentiations and classification of *Verbascum* species by NMR-based metabolomics. Phyto-

chemistry, v. 72, pp. 2045-2051, 2011.

GERMAN, J. B.; ROBERTS, M. A.; WATKINS, S. M. Genomics and metabolomics as markers for the interaction of diet and health: lessons from lipids. Journal of Nutrition, v. 133, pp. 2078S-2083S, 2003.

GERMAN, J. B.; ROBERTS, M. A.; FAY, L.; WATKINS, S. M. Metabolomics and individual metabolic assessment: the next great challenge for nutrition. Journal of Nutrition, v. 132, pp. 2486-2487, 2002.

GIBB, S. MALDIquantForeign: Import/Export routines for MALDIquant. R package version 0.4. `http://CRAN.R-project.org/package=MALDIquantForeign`. 2013.

GIBB, S. and STRIMMER, K. MALDIquant: a versatile R package for the analysis of mass spectrometry data. Bioinformatics, v. 28, pp. 2270-2271, 2012.

GIBNEY, M. J.; WALSH, M.; BRENNAN, L.; ROCHE, H. M.; GERMAN, B.; VAN OMMEN, B. Metabolomics in human nutrition: opportunities and challenges. American Journal of Clinical Nutrition, v. 82, pp. 497-503, 2005.

GO, E. P.; APON, J. V.; LUO, G. SAGHATELIAN, A.; DANIELS, R. H.; SAHI, V.; DUBROW, R.; CRAVATT, B. F.; VERTERS, A.; SIUZDAK, G. Desorption/Ionization on Silicon Nanowires. Analytical Chemistry, v. 77, pp. 1641-1646, 2005.

GOBBO-NETO, L.; GUARATINI, T.; PESSOA, C.; DE MORAES, M. O.; COSTA-LOTUFO, L.; VIEIRA, R. B.; COLEPICOLO, P.; LOPES, N. P. Differential Metabolic and Biological Profiles of Lychnophora ericoides Mart. (Asteraceae) from Different Localities in the Brazilian "campos rupestres". Journal of the Brazilian Chemical Society, v. 21, pp. 750-759, 2010.

GOBBO-NETO, L.; LOPES, N. P. Online identification of chlorogenic acids, sesquiterpene lactones, and flavonoids in the Brazilian Arnica *Lychnophora ericoides* Mart. (Asteraceae) leaves by HPLC-DAD-MS and HPLC-DAD-MS/MS and a validated HPLC-DAD method for their simultaneous analysis. Journal of Agricultural and Food Chemistry, v. 56, pp. 1193-1204, 2008.

GOBBO-NETO, L.; DOS SANTOS, M. D.; ALBARELLA, L.; ZOLLO, F.; PIZZA, C.; LOPES, N. P. Glycosides, caffeoylquinic acids and flavonoids from the polar extract of leaves of *Lychnophora ericoides* Mart. (Asteraceae). Biochemical Systematics and Ecology, v. 36, pp. 473-475, 2008.

GOBBO-NETO, L.; DOS SANTOS, M.; KANASHIRO, A.; ALMEIDA, M. C.; LUCISANO-

VALIM, Y. M.; LOPES, J. L. C.; SOUZA, G. E. P; LOPES, N. P. Evaluation of the anti-inflammatory and antioxidant activities of di-C-glucosylflavones from *Lychnophora ericoides* (Asteraceae). Planta Medica, v. 71, pp. 3-6, 2005.

GÓMEZ, M.; GIL, J. F. Topical anti-inflammatory activity of *Calea prunifolia* HBK (Asteraceae) in the TPA model of mouse ear inflammation. Journal of the Brazilian Chemical Society, v.22, pp. 2391-2395, 2011.

GONÇALVES, L. D.; ALMEIDA, H. R.; DE OLIVEIRA, P. M.; LOPES, N. P.; TURATTI, I. C. C.; ARCHANJO, F. C.; GRAEL, C. F. F. Contribution for the phytochemical studies of *Ageratum fastigiatum*. Brazilian Journal of Pharmacognosy, v. 21, pp. 936-942, 2011.

GOODACRE, R.; BROADHURST, D.; SMILDE, A. K.; KRISTAL, B. S.; BAKER, J. D.; BEGER, R.; BESSANT, C.; CONNOR, S.; CAPUANI, G.; CRAIG, A.; EBBELS, T.; KELL, D. B.; MANETTI, C.; NEWTON, J.; PATERNOSTRO, G.; SOMORJAI, R.; SJÖSTRÖM, M.; TRYGG, J.; WULFERT, F. Proposed minimum reporting standards for data analysis in metabolomics. Metabolomics, v. 3, pp. 231-241, 2007.

GOODACRE, R. Metabolomics - the way forward. Metabolomics, v.1, pp. 1-2, 2005.

GOODACRE, R.; VAIDYANATHAN, S.; DUNN, W. B.; HARRIGAN, G. G.; KELL, D. B. Metabolomics by numbers: acquiring and understanding global metabolite data. Trends in Biotechnology, v. 22, pp. 245-252, 2004.

GOUVEA, D. R.; MELONI, F.; RIBEIRO, A. de B. B.; LOPES, J. L. C.; LOPES, N. P. A new HPLC-DAD-MS/MS method for the simultaneous determination of major compounds in the crude extract of *Lychnophora salicifolia* Mart. (Brazilian arnicão) leaves: Application to chemical variability evaluation. Analytica Chimica Acta, v. 748, pp. 28-36, 2012.

GRAEL, C. F. F.; ALBUQUERQUE, S.; LOPES, J. L. C. Chemical constituents of *Lychnophora pohlii* and trypanocidal activity of crude plant extracts of isolated compounds. Fitoterapia, v. 76, pp. 73-82, 2005.

GRIFFIN, J. L.; BOLLARD, M. E. Metabonomics: its potential as a tool in toxicology for safety assessment and data integration. Current Drug Metabolism, v. 5, pp. 389-398, 2004.

GRIFFITHS, J. R.; STUBBS, M. Opportunities for studying cancer by metabolomics: preliminary observations on tumors deficient in hypoxia-inducible factor 1. Advances in Enzyme Regulation, v. 43, pp. 67-76, 2003.

GROMBONE-GUARATINI, M. T.; MANSANARES, M. E.; SEMIR, J.; SOLFERINI, V. N. Chromosomal studies of three species of *Bidens* (L.) (Asteraceae). Caryologia, v. 59, pp.

14-18, 2006.

GUILLET, G.; BÉLANGER, A.; ARNASON, J. T. Volatile monoterpenes in *Porophyllum gracile* and *P. ruderale* (Asteraceae): Identification, localization and insecticidal synergism with α-terthienyl. Phytochemistry, v. 49, pp. 423-429, 1998.

GULLBERG, J.; JONSSON, P.; NORDSTRÖM, A.; SJÖSTRÖM, M.; MORITZ, T. Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of Arabidopsis thaliana samples in metabolomic studies with gas chromatography/mass spectrometry. Analytical Biochemistry, v. 331, pp. 283-295, 2004.

GUO, Z.; ZHANG, Q.; ZOU, H.; GUO, B.; NI, J. A method for the analysis of low-mass molecules by MALDI-TOF mass spectrometry. Analytical Chemistry, v. 74, pp. 1637-1641, 2002.

GUZZO, L. S.; SAÚDE-GUIMARÃES, D. A.; SILVA, A. C. A.; LOMBARDI, J. A.; GUIMARÃES, H. N.; GRABE-GUIMARÃES, A. Antinociceptive and anit-inflammatory activities of ethanolic extracts of Lychnophora species. Journal of Ethnopharmacology, v. 116, pp. 120-124, 2008.

HAGEL, J. M. and FACCHINI, P. J. Plant metabolomics: analytical platforms and integration with functional genomics. Phytochemistry Reviews, v. 7, pp. 479-497, 2008.

HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. Multivariate data analysis. Simon & Schuster Company, New Jersey, 4th edition, 1995.

HALL, R. D. Plant metabolomics: from holistic hope, to hype, to hot topic. New Phytologist, v. 169, pp. 453-468, 2006.

HANSEN, M. A. E. Data Analysis. In Metabolome Analysis: An Introduction, edited by VILLAS-BÔAS, S.G.; ROESSNER, U.; HANSEN, M. A. E.; SMEDSGAARD, J.; NIELSEN, J. John Wiley & Sons, New Jersey, pp. 146-187, 2007.

HERBERT, P. D. N.; CYWINSKA, A.; BALL, S. L.; DEWAARD, J. R. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London B, v. 270, pp. 313-321, 2003.

HERZ, W.; KUMAR, N. Sesquiterpene lactones of *Calea zacatechichi* and *C. urticifolia*. Phytochemistry, v. 19, pp. 593-597, 1980.

HILLENKAMP, F.; KARAS, M.; BEAVIS, R. C.; CHAIT, B. T. Matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers. Analytical Chemistry, v. 63, pp. 1193-1201, 1991.

HIND, D. J. N. A New Species of *Lychnophora* (Compositae: Vernonieae: Lychnophorinae) from Bahia, Brazil. Kew Bulletin, v. 55, pp. 393-397, 2000.

HOFFMANN, O. Compositae. In Die natürlichen Pflanzenfamilien, edited by ENGLER, A. and PRANTL K. Wilhelm Engelmann, Leipzig, pp. 324-333, 1894.

HOFFMANN, B.; HOLZL, J. A methylated chalcone glucoside from *Bidens pilosa*. Phytochemistry, v. 27, pp. 3700-3701, 1988a.

HOFFMANN, B.; HOLZL, J. Further acylated chalcones from *Bidens pilosa*. Planta Medica, v. 54, pp. 450-451, 1988b.

HOFFMANN, B.; HOLZL, J. New chalcone from *Bidens pilosa*. Planta Medica, v. 54, pp. 52-54, 1988c.

HOGEWEG, P. The Roots of Bioinformatics in Theoretical Biology. PLoS Computational Biology, v. 7, pp. 1-5, 2011.

HOLLYWOOD, K.; BRISON, D. R.; GOODACRE, R. Metabolomics: current technologies and future trends. Proteomics, v. 6, pp. 4716-4723, 2006.

HOPKINS, G. W.; FRECKLETON, R. P. Declines in the numbers of amateur and professional taxonomists: implications for conservation. Animal Conservation, v. 5, pp. 245-249, 2002.

HOTHORN, T.; BRETZ, F.; WESTFALL, P. Simultaneous Inference in General Parametric Models. Biometrical Journal, v. 50, pp. 346-363, 2008.

HUI, W. H.; LEE, W. K. Triterpenoid and steroid constituents of some *Lactuca* and *Ageratum* species of Hong Kong. Phytochemistry, v. 10, pp. 899-901, 1971.

IBGE. Manual Técnico da Vegetação Brasileira. Manuais Técnicos em Geociências, n.1, IBGE, Rio de Janeiro, 2012.

IBGE. Mapas de Biomas do Brasil. Ministério do Planejamento e Orçamento, 2004. Available at: <`http://www.ibge.gov.br/home/presidencia/noticias/noticia_visualiza.php?id_noticia=169`>. Accessed on: 1. feb. 2013.

IGILE, G. O.; OLESZEK, W.; JURZYSTA, M.; BURDA, S.; FAFUNSO, M.; FASANMADE, A. A. Flavonoids from *Vernonia amygdalina* and their antioxidant activities. Journal of Agricultural and Food Chemistry, v. 42, pp. 2445-2448, 1994.

JAKUPOVIC, J.; SCHMEDA-HIRSCHMANN, G.; SCHUSTER, A.; ZDERO, C.; BOHLMANN, F.; KING, R. M.; ROBINSON, H.; PICKARDT, J. Hirsuntinolides, glaucolides and sesquiterpene lactone from *Vernonia* species. Phytochemistry, v. 25, pp. 145-158.

1986.

JAMAL, C. M.; TAVARES, K. G.; RASLAN, D. S.; MOREIRA, M. C. G.; MIRAGLIA, M. C. M.; DUARTE, I. D. G.; SANTOS, M. C. Antinociceptive activity of *Lavoisiera pulcherrima*. Fitoterapia, v. 70, pp. 178-180, 1999.

JANSEN, R. K.; PALMER, J. D. A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). Proceedings of the National Academy of Sciences of the United States of America, v. 84, pp. 5818-5822, 1987.

JEFFREY, C. Compositae. In Flowering plants of the world, edited by Heywood, V. H. Mayflower Books, New York, pp. 263-268, 1978.

JIMENEZ, M.; GUZMAN, A. P.; AZUARA, E.; GARCIA, O.; MENDOZA, M. R.; BERISTAIN, C. I. Volatile Compounds and Antioxidative Activity of *Porophyllum tagetoides* Extracts. Plant Foods for Human Nutrition, v. 67, pp. 57-63, 2012.

JISAKA, M.; OHIGASHI, H.; TAKEGAWA, K.; HIROTA, M.; IRIE, R.; HUFFMANN, M. A.; KOSHIMIZU, K. Steroid glucosides from *Vernonia amygdalina*, a possible chimpanzee medicinal plant. Phytochemistry, v. 34, pp. 409-413, 1993.

JISAKA, M.; OHIGASHI, H.; TAKAGAKI, T.; NOZAKI, H.; TADA, T.; HIROTA, M.; IRIE, R.; HUFFMAN, M. A.; NISHIDA, T.; KAJI, M.; KOSHIMIZU, K. Bitter Steroid Glucosides, Vernoniosides A1, A2 and A3, and Related B1 from a Possible Medicinal Plant, *Vernonia amygdalina*, used by Wild Chimpanzees. Tetrahedron, v. 48, pp. 625-632, 1992.

JOHNSON, H. E.; BROADHURST, D.; GOODACRE, R.; SMITH, A. R. Metabolic fingerprinting of salt-stressed tomatoes. Phytochemistry, v. 62, pp. 919-928, 2003.

JOLLIFFE, I. T. Principal component analysis. Springer-Verlag, New York, 2nd edition, 2002.

KABACOFF, R. I. R in Action. Manning Publications, New York, 2011.

KANG, D.; KIM, J.; CHOI, J. N.; LIU, K.; LEE, C. H. Chemotaxonomy of *Trichoderma spp.* using mass spectrometry-based metabolite profiling. Journal of Microbiology and Biotechnology, v. 21, pp. 5-13, 2011.

KARAS, M.; GLÜCKMANN, M.; SCHÄFER, J. Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. Journal of Mass Spectrometry, v. 35, pp.1-12, 2000.

KARAS, M.; HILLENKAMP, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10000 Daltons. Analytical Chemistry, v. 60, pp. 2299-2301, 1988.

KARAS, M.; BACHMANN, D.; BAHR, U.; HILLENKAMP, F. Matrix-assisted ultraviolet laser desorption of non-volatile compounds. International Journal of Mass Spectrometry and Ion Process, v. 78, pp. 53-68, 1987.

KARIS and RYDING (1994) In: DO NASCIMENTO, A. M.; SOUSA e SILVA, F.; DE OLIVEIRA, D. C. R. Constituents of *Calea platylepis* Sch. Bip. ex Baker. Biochemical Systematics and Ecology, v. 30, pp. 993-996, 2002.

KASTURI, T. R.; MANITHOMAS. Essential oil of *Ageratum conyzoides* - Isolation and structure of two new constituents. Tetrahedron Letters, v. 27, pp. 2573-2575, 1967.

KAUFMAN, L.; ROUSSEEUW, P. J. Finding Groups in Data - An Introduction to Cluster Analysis. John Wiley and Sons, New York, 1990.

KEELEY, S. C.; ROBINSON, H. Vernonieae. In Systematics, Evolution, and Biogeography of Compositae, edited by FUNK, V. A.; STUESSY, S. A.; BAYER, T. F.; RANDALL, J. IAPT, Vienna, pp.439-469, 2009.

KEELEY, S. C.; JONES, S. B. Distribution of pollen types in *Vernonia* (*Vernonieae*: Compositae). Systematic Botany, v. 4, pp. 195-202, 1979.

KELES, L. C.; DE MELO, N. I.; AGUIAR, G. P.; WAKABAYASHI, K. A. L.; DE CARAVALHO, C. E.; CUNHA, W. R.; CROTTI, A. E. M.; LOPES, J. L. C.; LOPES, N. P. Lychnophorinae (Asteraceae): a survey of its chemical constituents and biological activities. Química Nova, v. 33, pp. 2245-2260, 2010.

KELL, D. B. Systems biology, metabolic modelling and metabolomics in drug discovery and development. Drug Discovery Today, v. 11, pp. 1085-1092, 2006.

KIM, H. K.; CHOI, Y. H.; VERPOORTE, R. NMR-based plant metabolomics: where do we stand, where do we go? Trends in Biotechnology, v. 29, pp. 267-275 , 2011.

KIM, H. K.; SAIFULLAH, 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 *Illex* species by NMR-based metabolomics. Phytochemistry, v. 71, pp. 773-784, 2010.

KNOCHENMUSS, R.; ZENOBI, R. MALDI Ionization: The Role of In-Plume Processes. Chemical Reviews, v. 103, pp.441-452, 2003.

KOOPMANS, L. H.; OWEN, D. B.; ROSENBLATT, J. I. Confidence intervals for the coefficient of variation for the normal and log normal distributions. Biometrika, v.51, pp.25-32, 1964.

KOPKA, J.; FERNIE, A.; WECKWERTH, W.; GIBON, Y.; SITT, M. Metabolite profiling in plant biology: platforms and destinations. Genome Biology, v. 5, pp. 109.1-109.9, 2004.

KOS, O.; CASTRO, V.; MURILLO, R.; POVEDA, L.; MERFORT, I. *Ent*-kaurane glycosides and sesquiterpene lactones of the hirsutinolide type from *Vernonia triflosculosa*. Phytochemistry, v. 67, pp. 62-69, 2006.

KRASCHENINNIKOV, H. Compositae austro-americanae novae I. Notulae systematicae Herbario Hosti Botanici Petropolitoni, v. 3, pp. 157-162, 1922.

KRISHNAN, P.; KRUGER, N. J.; RATCLIFFE, R. G. Metabolite fingerprinting and profiling in plants using NMR. Journal of Experimental Botany, v. 56, pp. 255-265, 2005.

KRZANOWSKI, W.J. Principles of multivariate analysis. Oxford University Press, New York, revised edition, 2000.

KUHARA, T. Chemical diagnosis of inborn errors of metabolism and metabolome analysis of urine by capillary gas chromatography/mass spectrometry. In Metabolomics: The Frontier of Systems Biology, edited by TOMITA, M. and NISHIOKA, T. Springer Verlag, Tokyo, pp. 167-192, 2005.

KUHNERT, N.; JAISWAL, R.; ERAVUCHIRA, P.; EL-ABASSY, R. M.; VON DER KAMMER, B.; MATERNY, A. Scope and limitations of principal component analysis of high resolution LC-TOF-MS data: the analysis of the chlorogenic acid fraction in green coffee beans as a case study. Analytical Methods, v. 3, pp. 144-155, 2011.

KUMAR, J. K.; SINHA, A. K. A new disubstituted acetylacetone from the leaves of *Bidens pilosa* Linn. Natural Product Research, v. 17, pp. 71-74, 2003.

KUMARI, P.; MISRA, K.; SISODIA, B. S.; FARIDI, U.; SRIVASTAVA, S.; LUQMAN, S.; DAROKAR, M. P.; NEGI, A. S.; GUPTA, M. M.; SINGH, S. C.; KUMAR, J. K. A promising anticancer and antimalarial component from the leaves of *Bidens pilosa*. Planta Medica, v. 75, p. 59-61, 2009.

LABUCKAS, D. O.; ZYGADLO, J. A.; FAILLACI, S. M.; ESPINAR, L. A. Constituents of the volatile oil of *Porophyllum obscurum* (Spreng.) D. C. Flavour and Fragrance Journal, v. 14, pp. 107-108, 1999.

LANGFELDER, P.; HORVATH, S. moduleColor: Basic module functions. R package version 1.08-1. `http://CRAN.R-project.org/package=moduleColor`. 2009.

LANGFELDER, P.; ZHANG, B.; HORVATH, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics Application Note, v. 24, pp.

719-720, 2008.

LANGLEY, G. J.; HERNIMAN, J. M.; TOWNELL, M. S. 2B or not 2B, that is the question: further investigation into the use of pencil as a matrix for matrix-assisted laser desorption/ionisation. Rapid Communications in Mass Spectrometry, v. 21, pp. 180-190, 2007.

LAST, R. L.; JONES, A. D.; SHACHAR-HILL, Y. Towards the plant metabolome and beyond. Nature Reviews Molecular Cell Biology, v. 8, pp. 167-174, 2007.

LEE, Y. J.; PERDIAN, D. C.; SONG, Z.; YEUNG, E. S.; NIKOLAU, B. J. Use of mass spectrometry for imaging metabolites in plants. The Plant Journal, v. 70, pp. 81-95, 2012.

LEE, T.; LU, C.; KUO, Y.; LO, J.; LEE, C. Unexpected novel pheophytin peroxides from the leaves of *Bidens pilosa*. Helvetica Chimica Acta, v. 91, pp. 79-84, 2008.

LEITÃO FILHO, H. F.; SEMIR, J. Uma nova combinação para o gênero *Vernonia* Schreb. (Compositae): *Vernonia damazoi* (Beauv.) Leitão Filho e Semir. Revista Brasileira de Botânica, v. 2, pp. 113-116, 1979.

LESSING, C. F. Synopsis Generum Compositarum, Berlin, Duncker and Humboldt, 1832.

LESSING, C. F. De synanthereis dissertatio quarta. Linnaea, v. 6, pp. 240-399, 1831a.

LESSING, C. F. De synanthereis herbarii regii berolinensis dissertationes, IV. Vernoniearum mantissa. Linnaea, v. 6, pp. 624-721, 1831b.

LESSING, C. F. De synanthereis herbarii regii berolinensis dissertationes. I-Vernonieae. Linnaea, v. 4, pp. 240-356, 1829.

LI, S.; KUANG, H.; OKADA, Y.; OKUYAMA, T. New acetylenic glucosides from *Bidens bipinnata* Linne. Chemical and Pharmaceutical Bulletin, v. 52, po. 439-440, 2004.

LINDON, J. C.; HOLMES, E.; NICHOLSON, J. K. Metabonomics and its role in drug development and disease diagnosis. Expert Review of Molecular Diagnostics, v. 4, pp. 189-199, 2004.

LINDON, J. C.; NICHOLSON, J. K.; HOLMES, E.; ANTTI, H.; BOLLARD, M. E.; KEUN, H.; BECKONERT, O.; EBBELS, T. M.; REILY, M. D.; ROBERTSON, D.; STEVENS, G. J.; LUKE, P.; BREAU, A. P.; CANTOR, G. H.; BIBLE, R. H.; NIEDERHAUSER, U.; SENN, H.; SCHLOTTERBECK, G.; SIDELMANN, U. G.; LAURSEN, S. M.; TYMIAK, A.; CAR, B. D.; LEHMAN-MCKEEMAN, L.; COLET, J.; LOUKACI, A.; THOMAS, C. Contemporary issues in toxicology the role of metabonomics in toxicology and its evaluation by the COMET project. Toxicology and Applied Pharmacology, v.187, pp. 137-146, 2003.

LIU, J. A.; XIONG, L.; ZHANG, S.; WEI, J. C.; XIONG, S. X. C60 Fluorine Derivative as Novel Matrix for Small Molecule Analysis by MALDI-TOF MS. Metabolomics Open Access, v. 2, pp. 1-5, 2012.

LORENZI, H.; MATOS, F. J. A. (2002). In: FONSCECA, M. C. M.; BARBOSA, L. C. A.; NASCIMENTO, E. A.; CASALI, V. W. D. Essential Oils form Leaves and Flowers of *Porophyllum ruderale* (Jacq.) Cassini (Asteraceae). Journal of Essential Oil Research, v. 18, pp. 1-3, 2006.

LOAYZA, I.; DE GROOT, W.; LORENZO, D.; DELLACASSA, E.; MONDELLO, L.; DUGO, G. Composition of the essential oil of *Porophyllum ruderale* (Jacq.) Cass. from Bolivia. Flavour and Fragrance Journal, v. 14, pp. 393-398, 1999.

LU, M.; LAI, Y.; CHEN, G.; CAI, Z. Matrix Interference-Free Method for the Analysis of Small Molecules by Using Negative Ion Laser Desorption/Ionization on Graphene Flakes. Analytical Chemistry, v. 83, pp. 3161-3169, 2011.

MACHADO, A. L.; ARAGÃO, F. M.; BANDEIRA, P. N.; DOS SANTOS, H. S.; ALBUQUERQUE, M. R. J. R. Constituintes químicos de *Vernonia scorpioides* (Lam) Pers. (Asteraceae). Química Nova, v. 36, pp. 540-543, 2013.

MACLEISH, N. F. F. Eight new combinations in *Vernonia* (Compositae: Vernonieae). Systematic Botany, v. 9, pp. 133-136, 1984.

MAECHLER, M.; ROUSSEEUW, P.; STRUYF, A.; HUBERT, M.; HORNIK, K. cluster: Cluster Analysis Basics and Extensions. R package version 1.14.4, 2013.

MALDONADO, E.; MÁRQUEZ, C. L.; ORTEGA, A. A thymol derivative from *Calea nelsonii*. Phytochemistry, v. 31, pp. 2527-2528, 1992.

MANSANARES, M. E.; FRONI-MARTINS, E. R.; SEMIR, J. Chromosome numbers in the genus *Lychnophora* Mart. (Lychnophorinae, Vernonieae, Asteraceae). Caryologia, v. 55, pp. 367-374, 2002.

MARCHANT, Y. Y.; GANDERS, F. R.; WAT, C.; TOWERS, G. H. N. Polyacetylenes in Hawaiian Bidens. Biochemical Systematics and Ecology, v. 12, pp. 167-178, 1984.

MARTINS, A. M.; CAMACHO, D.; SHUMAN, J.; SHA, W.; MENDES, P.; SHULAEV, V. A systems biology study of two distinct growth phases of Saccharomyces cerevisiae cultures. Current Genomics, v. 5, pp. 649-663, 2004.

MARTIUS C. F. P. Tabula geographica Brasilie et terrarum adjacentium. Tabula geographica quinque provicias florae Brasiliensis illustrans. In Flora Brasiliensis, v. 1, Fasc. 21, Monac-

chi et Lipsiae, edited by MARTIUS C. F. P. , 1824.

MARTIUS, C. F. P. Novum plantarum genus Lychnophora. Denkschr. K. Bayer. Botanische Gesellschaft Regensburg, v. 2, pp.148-159, 1822.

MASSART, D.L.; VANDEGINSTE, B. G. M.; DEMING, S. N.; MICHOTTE, Y.; KAUFMANN, L. Chemometrics: A Textbook. Elsevier, Amsterdam, 1988.

MATTFELD, J. Compositae. In Plantae Lützelburgianae brasilienses I, edited by PILGER, R. Notiblock Botanischer Garten Berlin-Dahlem, v. 8, pp. 428-451, 1923.

McCOMBIE, G. and KNOCHENMUSS, R. Small-Molecule MALDI Using the Matrix Suppression Effect to Reduce or Eliminate Matrix Background Interferences. Analytical Chemistry, v. 76, pp. 4990-4997, 2004.

MCLAFFERTY, F. W. Tandem mass-spectrometry. Science, v. 214, pp. 280-287, 1981.

MEHROTRA, B.; MENDES, P. Bioinformatics Approaches to Integrate Metabolomics and other Systems Biology Data. In Biotechnology in Agriculture and Forestry, v. 57, Plant Metabolomics, edited by SAITO, K.; DIXON, R. A.; WILLMITZER, L. Springer Verlag, Berlin Heidelberg, pp. 105-115, 2006.

MERCHANT, A.; RICHTER, A.; POPP, M.; ADAMS, M. Targeted metabolite profiling provides a functional link among eucalpyt taxonomy, physiology and evolution. Phytochemistry, v. 67, pp. 402-408, 2006.

MIRKIN, B. Clustering for data mining - a data recovery approach. Taylor & Francis Group, Boca Raton, 2005.

MOOLENAAR, S. H.; ENGELKE, U. F.; WEVERS, R. A. Proton nulcear magnetic resonance spectroscopy of body fluids in the field of inborn errors of metabolism. Annals of Clinical Biochemisry, v. 40, pp. 16-24, 2003.

MORALES-ESCOBAR, L.; BRACA, A.; PIZZA, C.; TOMMASI, N. New phenolic derivatives from *Vernonia mapirensis* Gleason. ARKIVOC, v. 7, pp. 349-358, 2007.

MORGENTHAL, K.; WIENKOOP, S.; WOLSCHIN, F.; WECKWERTH, W. Integrative Profiling of Metabolites and Proteins. In Methods in Molecular Biology, v. 358, Metabolomics: Methods and Protocols, edited by WECKWERTH, W. Totowa, pp. 57-75, 2007.

MORITZ, T.; JOHANSSON, A.I. Plant Metabolomics. In Metabolomics, Metabonomics and Metabolite Profiling, edited by GRIFFITHS, W. J. RSC Publishing, Cambridge, pp. 254-272, 2008.

MURRAY, K. K.; BOYD, R. K.; EBERLIN, M. N.; LANGLEY, G. J.; LI, L.; NAITO, Y. Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013). Pure and Applied Chemistry, ASAP article, 06.06.2013.

MYERS, N.; MITTERMEIER, R. A.; MITTERMEIER, C. G.; DA FONSECA, G. A. B.; KENT, J. Biodiversity hotspots for conservation priorities. Nature, v. 403, pp. 853-858, 2000.

NAKAJIMA, J.; LOEUILLE, B.; HEIDEN, G.; DEMATTEIS, M.; HATTORI, E.K.O.; MAGENTA, M.; RITTER, M.R.; MONDIN, C.A.; ROQUE, N.; FERREIRA, S.C.; TELES, A.M.; BORGES, R.A.X.; MONGE, M.; BRINGEL, Jr., J.B. A.; OLIVEIRA, C.T.; SOARES, P.N.; ALMEIDA, G.; SCHNEIDER, A.; SANCHO, G.; SAAVENDRA, M.M.; LIRO, R.M.; SOUZA-BUTURI, F.O.; PEREIRA, A.C.M.; MORAES, M.D.; SILVA, G.A.R. Asteraceae in Lista de Espécies da Flora do Brasil. Jardim Botânico do Rio de Janeiro. (`http://floradobrasil.jbrj.gov.br/2012/FB000055`). 2012

NAYAK, R.; KNAPP, D. R. Effects of thin-film structural parameters on laser desorption/ionization from porous aluminia. Analytical Chemistry, v. 79, pp. 4950-4956, 2007.

NICHOLSON, J. K.; WILSON, I. D. Understanding 'global' systems biology: metabonomics and the continuum of metabolism. Nature Reviews Drug Discovery, v.2, pp. 668-676, 2003.

NICHOLSON, J. K.; CONNELLY, J.; LINDON, J. C.; HOLMES, E. Metabonomics: a platform for studying drug toxicity and gene function. Nature Reviews Drug Discovery, v. 1, pp. 153-161, 2002.

NICHOLSON, J. K; LINDON, J. C.; HOLMES, E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica, v. 29, pp. 1181-1189, 1999.

NIELSEN, J.; JEWETT, M. C. The role of metabolomics in systems biology. In Metabolomics A Powerful Tool in Systems Biology, edited by NIELSEN, J.; JEWETT, M. C. Springer-Verlag, Heidelberg, pp. 1-10, 2007.

NORDSTRÖM, A. Data mining for metabolomics. In Metabolomics, Metabonomics and Metabolite Profiling, edited by GRIFFITHS, W. J. RSC Publishing, Cambridge, pp. 273-294, 2008.

OBER, A. G.; FRONCZEK, F. R.; FISCHER, N. H. Three benzofurans and a 1,4-dioxin derivative from *Calea* species: the molecular structures of calebertin and caleteucrin. Journal of Natural Products, v. 48, pp. 242-248, 1985.

OBER, A. G.; QUIJANO, L.; URBATSCH, L. E.; FISCHER, N. H. Guaianolides from *Calea subcordata*. Phytochemistry, v. 23, pp. 1289-1292, 1984a.

OBER, A. G.; FRONCZEK, F. R.; FISCHER, N. H. The molecular structure of a new eudesmanolide, subcordatolide B, from *Calea subcordata*. Journal of Natural Products, v. 47, pp. 626-629, 1984b.

OBER, A. G.; QUIJIANO, L.; URBATSCH, L. E.; FISCHER, N. H. Eudesmanolides from *Calea trichomata*. Phytochemistry, v. 23, pp. 910-912, 1984c.

OGAWA, K.; SASHIDA, Y. Caffeoyl derivatives of a sugar lactone and hydroxy acid from the leaves of *Bidens pilosa*. Phytochemistry, v. 31, p. 3657-3658, 1992.

OHIGASHI, H.; JISAKA, M.; TAKAGAKI, T.; NOZAKI, H.; TADA, T.; HUFFMAN, M. A.; NISHIDA, T.; KAJI, M.; KOSHIMIZU, K. Bitter Principle and a Related Steroid Glucoside from *Vernonia amygdalina*, a Possible Medicinal Plant for Wild Chimpanzees. Agricultural and Biological Chemistry, v.55, pp. 1201-1203, 1991.

OKSMAN-CALDENTEY, K.; SAITO, K. Integrating genomics and metabolomics for engineering plant metabolic pathways. Current Opinion in Biotechnology, v. 16, pp. 174-179, 2005.

OKUNADE, A. L. *Ageratum conyzoides* L. (Asteraceae). Fitoterapia, v. 73, pp. 1-16, 2002.

OLDIGES, M.; LÜTZ, S.; PFLUG, S.; SCHROER, K.; STEIN, N.; WIENDAHL, C. Metabolomics: current state and evolving methodologies and tools. Applied Microbiology and Biotechnology, v. 76, pp. 495-511, 2007.

OLIVER, S. G.; WINSON, M. K.; KELL, D. B.; BAGANZ, F. Systematic functional analysis of the yeast genome. Trends in Biotechnology, v. 16, pp. 373-378, 1998.

OOSTENDORP, M., U. F. ENGELKE, M. A. WILLEMSEN, R. A. WEVERS. Diagnosing inborn errors of lipid metabolism with proton nuclear magnetic resonance spectroscopy. Clinical Chemistry, v. 52, pp. 1395-1405, 2006.

ORTEGA, C. A.; MARÍA, A. O. M.; GIANELLO, J. C. Chemical components and biological activity of *Bidens subalternans*, *B. aurea* (Asteraceae) and *Zuccagnia puntacta* (Fabaceae). Molecules, v. 5, pp. 465-467, 2000.

PANERO, J. L. and FUNK, A. The value of sampling anamalous taxa in phylogenetic studies: Major clades of the Asteraceae revealed. Molecular Phylogenetics and Evolution, v. 47, pp. 757-782, 2008.

PANERO, J. L. and FUNK, A. Toward a phylogenetic subfamilial classification for the Compositae (Asteraceae). Proceedings of the Biological Society of Washington, v. 115, pp. 909-922, 2002.

PEDRIOLI, P. G. A.; ENG, J. K.; HUBLEY, R.; VOGELZANG, M.; DEUTSCH, E. W.; RAUGHT, B.; PRATT, B.; NILSSON, E.; ANGELETTI, R. H.; APWEILER, R.; CHEUNG, K.; COSTELLO, C. E.; HERMJAKOB, H.; HUANG, S.; JULIAN, R. K.; KAPP, JR. E.; MCCOMB, M. E.; OLIVER, S. G.; OMENN, G.; PATON, N. W.; SIMPSON, R.; SMITH, R.; TAYLOR, C. F.; ZHU, W.; AEBERSOLD, R. A common open representation of mass spectrometry data and its application to proteomics research. Nature Biotechnology, v. 22, pp. 1459-1466, 2004.

PERDUE JR., R. E.; TSICHRITZIS, F.; JAKUPOVIC, J. Prevernocistifolides from *Vernonia galamensis*. Phytochemistry, v. 34, pp. 1075-1077, 1993.

PEREIRA, R. L. C.; IBRAHIM, T.; LUCCHETTI, L.; SILVA, S. J. R.; MORAES, V. L. G. Immunosuppressive and anti-inflammatory effects of methanolic extract and the polyacetylene isolated from *Bidens pilosa* L. Immunopharmacology, v. 43, pp. 31-37, 1999.

PLUSKAL, T.; CASTILLO, S.; VILLAR-BRIONES, A.; ORESC, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics, v. 11, pp. 395-406, 2010.

POPE, G. A.; MACKENZIE, D. A.; DEFERNEZ, M.; AROSO, M. A. M. M.; FULLER, L. J.; MELLON, F. A.; DUNN, W. B.; BROWN, M.; GOODACRE, R.; KELL, D. B.; MARVIN, M. E.; LOUIS, E. J.; ROBERTS, I. N. Metabolic footprinting as a tool for discriminating between brewing yeasts. Yeast, v. 24, pp. 667-679, 2007.

PUEBLA, P.; ARANGUREN, N.; RINCÓN, J.; ROJAS, M.; GUERRERO, M.; SAN FELICIANO, A. Polar compounds isolated from the leaves of *Calea prunifolia* H. B. K. and their anti-adrenergic related vasodilator activity. Journal of the Brazilian Chemical Society, v. 22, pp. 2281-2285, 2011.

QUIJANO, L.; CALDERON, J. S.; GÓMEZ, F. G.; SORIA, I. E.; RÍOS, T. Highly oxygenated flavonoids from *Ageratum corymbosum*. Phytochemistry, v. 19, pp. 2439-2442, 1980.

QUIJANO, L.; DE VIVAR, R.; RIOS, R. Revision of the structures of caleine A and B, germacranolide sesquiterpenes from *Calea zacatechichi*. Phytochemistry, v. 18, pp. 1745-1747, 1979.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http:`

//www.R-project.org/.

RAAMSDONK, L. M.; TEUSINK, B.; BROADHURST, D.; ZHANG, N.; HAYES, A.; WALSH, M. C.; BERDEN, J. A.; BRINDLE, K. M.; KELL, D. B.; ROWLAND, J. J.; WESTERHOFF, H. V.; VAN DAM, K.; OLIVER, S. G. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nature Biotechnology, v.19, pp. 45-50, 2001.

REDL, K.; BREU, W.; DAVIS, B.; BAUER, R. Anti-inflammatory active polyacetylenes from *Bidens campylotheca*. Planta Medica, v. 60, pp. 58-62, 1994.

REDL, K.; DAVIS, B.; BAUER, R. Chalcone glycosides from *Bidens campylotheca*. Phytochemistry, v. 32, pp. 218-220, 1993.

REED, G. F.; LYNN, F.; MEADE, B. D. Use of coefficient of Variation in Assessing Variability of Quantitativa Assays. Clinical and Diagnostic Laboratory Immunology, v. 9, pp. 1235-1239, 2002.

RENNER, S. Phylogeny and classification of the Melastomataceae and Memecylaceae. Nordic Journal of Botany, v. 13, pp. 519-540, 1993.

RICHETT, H. W. (1966). In: GUILLET, G.; BÉLANGER, A.; ARNASON, T. Volatile monoterpenes in *Porophyllum gracile* and *P. ruderale* (Asteraceae): Identification, localization and insecticidal synergism with α-terthienyl. Phytochemistry, v. 49, pp. 423-429, 1998.

RIZHSKY, L.; LIANG, H.; SHUMAN, J.; SHUALEV, V.; DAVLETOVA, S.; MITTLER, R. When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress. Plant Physiology, v. 134, pp. 1683-1696, 2004.

ROBERTSON, D. G., REILY, M. D.; BAKER, J. D. Metabonomics in preclinical drug development. Expert Opin. Drug Metabolism and Toxicology, v. 1, pp. 363-376, 2005.

ROBINSON, H., SCHILLING, E. E.; PANERO, J. L. Eupatorieae. In Systematics, Evolution, and Biogeography of Compositae, edited by FUNK, V. A.; STUESSY, S. A.; BAYER, T. F.; RANDALL, J. IAPT, Vienna, pp.731-744, 2009.

ROBINSON, H. Generic and Subtribal Classification of American Vernonieae. Smithsonian Contributions to Botany, v. 89, pp.1-116, 1999.

ROBINSON, H. Five new species of *Lychnophora* from Bahia, Brasil (Vernonieae: Asteraceae). Phytologia, v. 53, pp. 369-384, 1983.

ROBINSON, H. A revision of the tribal and subtribal limits of the Heliantheae (Asteraceae). Smithsonian Contributions to Botany nr. 51, Smithsonian Institution Press, City of Washington, 1981.

ROBINSON, H. New species of Vernonieae (Asteraceae) VI. *Lychnophora souzae*. Phytologia, v. 46, pp. 104-106, 1980.

ROCHFORT, S. Metabolomics reviewed: A new "omics" platform technology for systems biology and implications for natural products research. Journal of Natural Produtcts, v. 68, pp. 1813-1820, 2005.

ROESSNER, U. Plant Metabolomics. In Metabolome Analysis: An Introduction, edited by VILLAS-BÔAS, S.G; ROESSNER, U.; HANSEN, M. A. E.; SMEDSGAARD, J.; NIELSEN, J. John Wiley & Sons, New Jersey, pp. 215-238, 2007.

ROESSNER, U.; PETTOLINO, F. The importance of anatomy and physiology in plant metabolomics. In Metabolomics A Powerful Tool in Systems Biology, edited by NIELSEN, J. and JEWETT, M. C. Springer Verlag, Berlin Heidelberg, pp. 253-278, 2007.

ROESSNER, U.; WILLMITZER, L.; FERNIE, A. Metabolic profiling and biochemical phenotyping of plant systems. Plant Cell Reports, v. 21, pp. 189-196, 2002.

ROESSNER, U.; LUEDEMANN, A.; BRUST, D.; FIEHN, O.; LINKE, T.; WILLMITZER, L.; FERNIE, A. R. Metabolic profiling allows comprehensive Phenotyping of genetically or environmentally modified plant systems. The Plant Cell, v. 13, pp. 11-29, 2001.

ROESSNER, U.; WAGNER, C.; KOPKA, J.; TRETHEWEY, R. N.; WILLMITZER, L. Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. The Plant Journal, v. 23, pp. 131-142, 2000.

RYAN, D. and ROBARDS, K. Metabolomics: the greatest omics of them all? Analytical Chemistry, v. 78, pp. 7954-7958, 2006.

SAITO, K.; MATSUDA, F. Metabolomics for Functional Genomics, Systems Biology, and Biotechnology. Annual Review of Plant Biology, v. 61, pp. 463-89, 2010.

SARG, T. M.; ATEYA, A. M.; ABBAS, F. A. Constituents and biological activity of *Bidens pilosa* L. grown in Egypt. Acta Pharmaceutica Hungarica, v. 61, pp. 317-323, 1991.

SARKER, S. D.; BARTHOLOMEW, B.; NASH, R. J.; ROBINSON, N. 5-*O*-methylhoslundin: an unusual flavonoid from *Bidens pilosa* (Asteraceae). Biochemical Systematics and Ecology, v. 28, pp. 591-593, 2000.

SASHIDA, Y.; OGAWA, K.; KITADA, M.; KARIKOME, H.; MIMAKI, Y.; SHIMOMURA, H. New aurone glucosides and new phenylpropanoid glucosides from *Bidens pilosa*. Chemical and Pharmaceutical Bulletin, v. 39, pp. 709-711, 1991.

SCHAD, M.; MUNGUR, R.; FIEHN, O.; KEHR, J. Metabolic profiling of laser microdissected vascular bundles of *Arabidopsis thaliana*. Plant Methods, v. 1, 2005.

SCHULTZ-BIPONTINUS, C. H. *Lychnophora* Martius und einige benachbarte Gattungen. Jahresbericht der Pollichia, v. 20-21, pp. 321-439, 1863.

SCHULTZ-BIPONTINUS, C. H. Cassiniaceae uniflorae. Jahresbericht der Pollichia, v. 18-19, pp. 157-190, 1861.

SEAMAN, F. C. Sesquiterpene lactones as taxonomic characters in the Asteraceae. The Botanical Review, v. 48, pp. 121-594, 1982.

SEGER, C.; STURM, S. Analytical aspects of plant metabolite profiling platforms: current standings and future aims. Journal of Proteome Research, v. 6, pp. 480-497, 2007.

SEMIR, J.; REZENDE, A. R.; MONGE, M.; LOPES, N. P. As Arnicas Endêmicas das Serras do Brasil – Uma Visão sobre a Biologia e a Química das Espécies de *Lychnophora* (Asteraceae). UFOP. Ouro Preto MG, 2011.

SEMIR, J. Revisão Taxonômica de *Lychnophora* Mart. (Vernonieae: Compositae). Doctoral thesis, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, 1991.

SHARMA, O. P. Plant taxonomy. Tata McGraw-Hill Publishing Company Limited, New Delhi, 7. ed., pp. 95, 2007.

SHROFF, R.; RULÍSEK, L.; DOUBSKÝ, J.; SVATOS, A. Acid-base-driven matrix-assisted mass spectrometry for targeted metabolomics. PNAS, v. 106, pp. 10092-10096, 2009.

SHULAEV, V. Metabolomics technology and bioinformatics. Briefings in Bioinformatics. v.7, pp. 128-139, 2006.

SIMPSON, M. G. Plant Systematics. Elsevier Academic Press, Burlington MA, pp. 9-10, 2006.

SMITH, C. A.; WANT, E. J.; O'MAILLE, G.; ABAGYAN, R.; SIUZDAK, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. Analytical Chemistry, v. 78, pp. 779-787, 2006.

SOGA, T.; Capillary Electrophoresis-Mass Spectrometry for Metabolomics. In Methods in Molecular Biology, v. 358, Metabolomics: Methods and Protocols, edited by WECKWERTH, W. Humana Press, Totowa, 2007, pp. 129-137, 2007.

SPRENGEL, C. Caroli Linnaei Systema Vegetabilium. Editio Decima Sexta, Gottingae, vol. 4, 1827.

SPRENGEL, C. Caroli Linnaei Systema Vegetabilium. Editio Decima Sexta, Gottingae, vol. 3, 1826.

STACKLIES, W.; REDESTIG, H.; SCHOLZ, M.; WALTHER, D.; SELBIG, J. pcaMethods - a Bioconductor package providing PCA methods for incomplete data. Bioinformatics, v. 23, pp. 1164-1167, 2007.

STEINBECK, C.; SPITZER, V.; STAROSTA, M.; VON POSER, G. Identification of two chromenes from *Calea serrata* by semiautomatic structure elucidation. Journal of Natural Products, v. 60, pp. 627-628, 1997.

STEUER, R.; MORGENTHAL, K.; WECKWERTH, W.; SELBIG, J. A Gentle Guide to the Analysis of Metabolomic Data. In Methods in Molecular Biology, v. 358, Metabolomics: Methods and Protocols, edited by WECKWERTH, W. Humana Press, Totowa, pp. 105-126, 2007.

STEUER, R.; KURTHS, J.; FIEHN, O.; WECKWERTH, W. Observing and interpreting correlations in metabolomic networks. Bioinformatics, v. 19, pp. 1019-1026, 2003.

STEVENS, P. F. (2001 onwards). Angiosperm Phylogeny Website. Version 12, July 2012. (`http://www.mobot.org/MOBOT/research/APweb/`). 2012.

STUESSY, T. F. Plant taxonomy - the systematic evaluation of comparative data. Columbia University Press, New York, 2. ed., 2009.

SUMNER, L. W.; MENDES, P.; DIXON, R. A. Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry, v. 62, pp. 817-836, 2003.

TABACHNICK, B.; FIDELL, L. S. Using Multivariate Statistics. Allyn and Bacon, Boston, 5. ed., 2007.

TAGUCHI, R. Application of electrospray ionization mass spectrometry for metabolomics. In Metabolomics: The Frontier of Systems Biology, edited by TOMITA, M. and NISHIOKA, T. Springer Verlag, Tokyo, pp. 25-35, 2005.

TANAKA, K. The Origin of Macromolecule Ionization by Laser Irradiation (Nobel Lecture). Angewandte Chemie International Edition, v. 42, pp. 3861-3870, 2003.

TANAKA, K.; WAKI, H.; IDO, Y.; AKITA, S.; YOSHIDA, Y.; YOSHIDA, T. Protein and Polymer Analyses up to m/z 100000 by Laser Ionization Time-of-Flight Mass Spectrometry.

Rapid Communications in Mass Spectrometry, v. 2, pp. 151-153, 1988.

TAYLOR, J.; KING, R. D.; ALTMANN, T.; FIEHN, O. Application of metabolomics to plant genotype discrimination using statistics and machine learning. Bioinformatics, v. 18, pp. S241-S248, 2002.

TCHINDA, A. T.; TANE, P.; AYAFOR, J. F.; CONNOLLY, J. D. Stigmastane derivatives and isovaleryl sucrose esters from *Vernonia guineensis* (Asteraceae). Phytochemistry, v.63, pp. 841-846, 2003.

TCHINDA, A. T.; TSOPMO, A.; TANE, P.; AYAFOR, J. F.; CONNOLLY, J. D.; STERNER, O. Vernoguinosterol and vernoguinoside, trypanocidal stigmastane derivatives from *Vernonia guineensis* (Asteraceae). Phytochemistry, v.59, pp. 371-374, 2002.

THOLEY, A.; HEINZLE, E. Ionic (liquid) matrices for matrix-assisted laser desorption/ionization mass spectrometry - applications and perspectives. Analytial and Bioanalytical Chemistry, v. 386, pp. 24-37, 2006.

TOMMASI, N.; PIZZA, C. Flavonol and chalcone ester glycosides from *Bidens leucantha*. Journal of Natural Products, v. 60, pp. 270-273, 1997.

TOMITA, M. Overview. In Metabolomics: The Frontier of Systems Biology, edited by TOMITA, M. and NISHIOKA, T. Springer Verlag, Tokyo, pp. 1-6, 2005.

TOUDAHL, A. B.; FILHO, S. A. V.; SOUZA, G. H. B.; MORAIS, L. D.; DOS SANTOS, O. D. H.; JÄGER, A. K. Chemical composition of the essential oil from *Microlicia graveolens* growing wild in Minas Gerais. Brazilian Journal of Pharmacognosy, v. 22, pp. 680-681, 2012.

TOYANG, N. J.; VERPOORTE, R. A review of the medicinal potentials of plants of the genus *Vernonia* (Asteraceae). Journal of Ethnopharmacology, v. 146, pp. 681-723, 2013.

TRUJILLO, E.; DAVIS, C.; MILNER, J. Nutrigenomics, proteomics, metabolomics, and the practice of dietetics. Journal of American Dietetic Association, v. 106, pp. 403-413, 2006.

TRYGG, J.; HOLMES, E.; LUNDSTEDT, T. Chemometrics in Metabonomics. Journal of Proteome Research, v. 6, pp. 469-479, 2007.

TRYGG, J.; GULLBERG, J.; JOHANSSON, A. I.; JONSSON, P.; MORITZ, T. Chemometrics in Metabolomics - An Introduction. In Biotechnology in Agriculture and Forestry, v. 57, Plant Metabolomics, edited by SAITO, K.; DIXON, R. A.; WILLMITZER, L. Springer Verlag, Berlin Heidelberg, pp. 117-128, 2006.

TWEEDDALE, H.; NOTLEY-MCROBB, L.; FERENCI, T. Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ("metabolome") analysis. Journal of Bacteriology, v. 180, pp. 5109-5116, 1998.

UBILLAS, R. P.; MENDEZ, C. D.; JOLAD, S. D.; LUO, J.; KING, S. R.; CARLSON, T. J.; FORT, D. M. Antihyperglycemic acetylenic glucosides from *Bidens pilosa*. Planta Medica, v. 66, pp. 82-83, 2000.

VAN BAREN, C. M.; JUAREZ, B.; FERRARO, G. E.; COUSSIO, J. D. Phenolic constituents of *Porophyllum obscurum*. Biochemical Systematics and Ecology, v. 22, p. 543, 1994.

VAN DEN BERG, R. A.; RUBINGH, C. M.; WESTERHUIS, J. A.; VAN DER WERF, M. J.; SMILDE, A. K. Metabolomics data exploration guided by prior knowledge. Analytica Chimica Acta, v. 651, pp. 173-181, 2009.

VAN DEN BERG, R. A.; HOEFSLOOT, H. JC.; 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, v. 7, pp. 142-156, 2006.

VAN KAMPEN, J. J. A.; BURGERS, P. C.; DE GROOT, R.; GRUTERS, R. A.; LUIDER, T. M. Biomedical application of MALDI mass spectrometry for small-molecule analysis. Mass Spectrometry Reviews, v. 30, pp. 101-120, 2011.

VARNEK, A.; BASKIN, I. I. Chemoinformatics as a Theoretical Chemistry Discipline. Molecular Informatics, v. 30, pp. 20-32, 2011.

VELOSO, H. P.; FILHO, A. L. R. R.; LIMA, J. C. A. Classificação da vegetação brasileira, adaptada a um sistema universal. IBGE, Departamento de Recursos Naturais e Estudos Ambientais, Rio de Janeiro, 1991.

VERPOORTE, R.; CHOI, Y.H; MUSTAFA, N. R.; KIM, H. K. Metabolomics: back to basics. Phytochemistry Reviews, v. 7, pp. 525-537, 2008.

VERPOORTE, R., CHOI, Y. H.; KIM, H. K. NMR-based metabolomics at work in phytochemistry. Phytochemistry Reviews, v. 6, pp. 3-14, 2007.

VIANT, M. R.; LUDWIG, C.; GÜNTHER, U. L. 1D and 2D NMR spectroscopy: from metabolic fingerprinting to profiling. In Metabolomics, Metabonomics and Metabolite Profiling, edited by GRIFFITHS, W. J. RSC Publishing, Cambridge, pp. 44-70, 2008.

VILLAS-BÔAS, S. G.; KOULMAN, A.; LANE, G. A. Analytical methods from the perspective of method standardization. In Metabolomics A Powerful Tool in Systems Biology, edited by NIELSEN, J. and JEWETT, M. C. Springer-Verlag, Berlin Heidelberg, pp. 11-52, 2007.

WANG, H. Y.; CHU, X.; ZHAO, Z. X.; HE, X. S.; GUO, Y. L. Analysis of low molecular weight compounds by MALDI-FTICR-MS. Journal of Chromatography B, v. 879, pp. 1166-1179, 2011.

WANG, J.; WANG, N.; YAO, X.; KITANAKA, S. Structures and anti-histamine activity of chalcones e aurones compounds from *Bidens parviflora* Willd. Asian Journal of Traditional Medicines, v. 2, pp. 23-29, 2007a.

WANG, N.; WANG, J.; YAO, X.; KITANAKA, S. Two new monoterpene glycosides and a new (+)-jasmololone glucoside from *Bidens parviflora* Willd. Journal of Asian Natural Products Research, v. 9, p. 449-455, 2007b.

WANG, N.; WANG, J.; YAO, X.; KITANAKA, S. Two neolignan glucosides and anti- histamine release activities from *Bidens parviflora* Willd. Chemical and Pharmaceutical Bulletin, v. 54, pp. 1190-1192, 2006.

WANG, N.; YAO, X.; ISHII, R.; KITANAKA, S. Bioactive sucrose esters from *Bidens parviflora*. Phytochemistry, v. 62, pp. 741-746, 2003.

WANG, N.; YAO, X.; ISHII, R.; KITANAKA, S. Antiallergic agents from natural sources. 3. Structures and Inhibitory effects on nitric oxide production and histamine release of five novel polyacetylene glucosides from *Bidens parviflora* Willd. Chemical and Pharmaceutical Bulletin, v. 49, pp. 938-942, 2001.

WANG, J.; YANG, H.; LIN, Z.; SUN, H. Flavonoids from *Bidens pilosa* var. radiata. Phytochemistry, v. 46, pp. 1275-1278, 1997.

WATANABE, T.; KAWASAKI, H.; YONEZAWA, T.; ARAKAWA, R. Surface-assisted laser desorption/ionization mass spectrometry (SALDI-MS) of low molecular weight organic compounds and synthetic polymers using zinc oxide (ZnO) nanoparticles. Journal of Mass Spectrometry, v. 43, pp. 1063-1071, 2008.

WATKINS, S. M.; GERMAN, J. B. Towards the implementation of metabolomic assessments of human health and nutrition. Current Opinion in Biotechnology, v. 13, pp. 512-516, 2002a.

WATKINS, S. M. and GERMAN, J. B. Metabolomics and biochemical profiling in drug discovery and development. Current Opinion in Molecular Therapeutics, v. 4, pp. 224-228, 2002b.

WATKINS, S. M.; REIFSNYDER, P. R.; PAN, H. J.; GERMAN, J. B.; LEITER, E. H. Lipid metabolome-wide effects of the PPARgamma agonist rostiglitazone. The Journal of Lipid

Research, v. 43, pp. 1809-1817, 2002c.

WATKINS, S. M.; HAMMOCK, B. D.; NEWMAN, J. W.; GERMAN, J. B. Individual metabolism should guide agriculture toward foods for improved health and nutrition. The American Journal of Clinical Nutrition, v. 74, pp. 283-286, 2001.

WAWRA, H. R. Itinera Principum S. Coburgi. Wien, v. 2, pp. 1-205, 1888.

WEBB, A. R. Statistical pattern recognition. John Wiley & Sons Ltd., Chichester, 2nd edition, 2002.

WEI, X.; SONG, W.; SHI, X.; KOO, I.; WANG, B.; ZHANG, J.; YIN, X.; TANG, Y.; BOGDANOV, B.; KIM, S.; ZHOU, Z.; MCCLAIN, C.; ZHANG, X. MetSign: A Computational Platform for High-Resolution Mass Spectrometry-Based Metabolomics. Analytical Chemistry, v. 83, pp. 7668-7675, 2011.

WEI, J.; BURIAK, J. M.; SIUZDAK, G. Desorption-ionization mass spectrometry on porous silicon. Nature, v. 399, pp. 243-246, 1999.

WHEELER, Q. D.; RAVEN, P. H.; WILSON, E. O. Taxonomy: Impediment or Expedient. Science, v. 303, p. 285, 2004.

WICKHAM, H. ggplot2: elegant graphics for data analysis. Springer New York, 2009.

WIESEMÜLLER, B., ROTHE, H.; HENKE, W. Phylogenetische Systematik. Springer-Verlag, Berlin Heidelberg, p. 1, 2003.

WILSON, I.D.; NICHOLSON, J. K. Metabonomics and global systems biology. In Metabolomics, Metabonomics and Metabolite Profiling, edited by GRIFFITHS, W. J. RSC Publishing, Cambridge, pp. 295-316, 2008.

WINK, M.; BOTSCHEN, F.; GOSMANN, C.; SCHÄFER, H.; WATERMAN, P. G. Chemotaxonomy seen from a phylogenetic perspective and evolution of secondary metabolism. Annual Plant Reviews, v. 40, pp. 364-433, 2010.

WISHART, D. S. Current progress in computational metabolomics. Briefings in Bioinformatics, v. 8, pp. 279-293, 2007.

WISHART, D. S.; QUERENGESSER, L. M. M.; LEFEBVRE, B. A.; EPSTEIN, N. A.; GREINER, R.; NEWTON, J. B. Magnetic resonance diagnostics: a new technology for high-throughput clinical diagnostics. Clinical Chemistry, v. 47, pp. 1918-1921, 2001.

WOLD, S. and SJÖSTRÖM, M. Chemometrics, present and future success. Chemometrics and Intelligent Laboratory Systems, v. 44, pp. 3-14, 1998.

WOLD, S. Chemometrics; what do we mean with it, and what do we want from it? Chemometrics and Intelligent Laboratory Systems, v. 30, pp. 109-115, 1995.

WOLFENDER, J. L.; RUDAZ, S.; CHOI, Y. H.; KIM, H. K. Plant Metabolomics: From holistic data to relevant biomarkers. Current Medicinal Chemistry, v. 20, pp. 1056-1090, 2013.

WOLNIAK, M.; TOMCZYKOWA, M.; TOMCZYK, M.; GUDEJ, J.; WAWER, I. Antioxidant activity of extracts and flavonoids from *Bidens tripartita*. Acta Poloniae Pharmaceutica - Drug Research, v. 63, pp. 441-447, 2007.

WU, H.; FRONCZEK, F. R.; BURANDT JR., C. L.; ZJAWIONY, J. K. Antileishmanial Germacranolides from *Calea zacatechichi*. Planta Medica, v. 77, pp. 749-753, 2011.

WU, L.; CHIANG, Y.; CHUANG, H.; LO, C.; YANG, K.; WANG, S.; SHYUR, L. A novel polyacetylene significantly inhibits angiogenesis and promotes apoptosis in human endothelial cells through activation of the CDK inhibitors and caspase-7. Planta Medica, v. 73, pp. 655-661, 2007.

WU, L.; CHIANG, H. C.; WANG, S.; YANG, G.; CHEN, Y.; LAI, L.; SHYUR, L. Polyacetylenes function as anti-angiogenic agents. Pharmaceutical Research, v. 21, pp. 2112-2119, 2004.

XIANG, Z.; WANG, X. Q.; CAI, X. J.; ZENG, S. Metabolomics study on quality control and discrimination of three *Curcuma* Species based on gas chromatograph-mass spectrometry. Phytochemical Analysis, v. 22, pp. 411-418, 2011.

XU, S.; LI, Y.; ZOU, H.; QIU, J.; GUO, Z.; GUO, B. Carbon Nanotubes as Assisted Matrix for Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. Analytical Chemistry, v. 75, pp. 6191-6195, 2003.

YANG, Y.; ZHANG, S.; HOWE, K.; WILSON, D. B.; MOSER, F.; IRWIN, D.; THANNHAUSER, T. W. A Comparison of nLC-ESI-MS/MS and nLC-MALDI-MS/MS for GeLC-Based Protein Identification and iTRAQ-Based Shotgun Quantitative Proteomics. Journal of Biomolecular Techniques, v. 18, pp. 226-237, 2007.

ZDERO, C.; BOHLMANN, F.; MUNGAI, G. M. Glaucolides and hirsutinolides from African *Vernonia* species. Phytochemistry, v. 30, pp. 2653-2654, 1991.

ZENOBI, R.; KNOCHENMUSS, R. Ion formation in MALDI mass spectrometry. Mass Spectrometry Reviews, v.17, pp.337-366, 1998.

ZHU, N.; LI, X.; LIU, G.; SHI, X.; GUI, M.; SUN, C.; JIN, Y. Constituents from aerial parts of *Bidens ceruna* L. and their DPPH radical scavenging activity. Chemical Research in Chinese

Universities, v. 25, pp. 328-331, 2009.

ZOLLO, P. H. A.; KUIATE, J. R.; MENUT, C.; LAMATY, G.; BESSIERE, J. M.; CHALCHAT, J. C.; GARRY, R. P. Aromatic plants of tropical central Africa. Part XX. The occurrence of 1-phenylhepta-1,3,5-triyne in the essential oil of *Bidens pilosa* L. from Cameroon. Flavour and Fragance Journal, v. 10, pp. 97-100, 1995.

ZULUETA, M. C. A.; TADA, M.; RAGASA, C. Y. A diterpene from *Bidens pilosa*. Phytochemistry, v. 38, pp. 1449-1450, 1995.