toxins

MDPI

Article

Metabolome-Based Classification of Snake Venoms by Bioinformatic Tools

Luis L. Alonso [1,2], Julien Slagboom [1,2], Nicholas R. Casewell [3], Saer Samanipour [4,*] and Jeroen Kool [1,2,*]

1 Division of BioAnalytical Chemistry, Amsterdam Institute of Molecular and Life Sciences, Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands
2 Centre for Analytical Sciences Amsterdam (CASA), 1012 WX Amsterdam, The Netherlands
3 Centre for Snakebite Research and Interventions, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK
4 Van 't Hof Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
* Correspondence: s.samanipour@uva.nl (S.S.); j.kool@vu.nl (J.K.)

Abstract: Snakebite is considered a neglected tropical disease, and it is one of the most intricate ones. The variability found in snake venom is what makes it immensely complex to study. These variations are present both in the big and the small molecules found in snake venom. This study focused on examining the variability found in the venom's small molecules (i.e., mass range of 100–1000 Da) between two main families of venomous snakes—Elapidae and Viperidae—managing to create a model able to classify unknown samples by means of specific features, which can be extracted from their LC–MS data and output in a comprehensive list. The developed model also allowed further insight into the composition of snake venom by highlighting the most relevant metabolites of each group by clustering similarly composed venoms. The model was created by means of support vector machines and used 20 features, which were merged into 10 principal components. All samples from the first and second validation data subsets were correctly classified. Biological hypotheses relevant to the variation regarding the metabolites that were identified are also given.

Keywords: venom variation; metabolomics; data analysis; script-controlled peak integration

Key Contribution: A pipeline was developed to automatically extract MS1 features from metabolomics LC–MS measurements of venoms. It was applied in this research to investigate inter- and intra-family differences in venom composition.

check for updates

Citation: Alonso, L.L.; Slagboom, J.; Casewell, N.R.; Samanipour, S.; Kool, J. Metabolome-Based Classification of Snake Venoms by Bioinformatic Tools. *Toxins* **2023**, *15*, 161. https://doi.org/10.3390/toxins15020161

Received: 15 December 2022
Revised: 31 January 2023
Accepted: 10 February 2023
Published: 15 February 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Neglected tropical diseases (NTDs)—such as dengue and leishmaniasis—are a group of diseases that mostly afflict populations found in low- and middle-income countries of the tropics and thus have not received as much attention as other diseases [1]. Within NTDs, snakebite stands out for its high mortality rate, although it was only recently accepted as a Category A neglected tropical disease by the WHO in 2017 [2]. Snakebite results in more than 1.8 million envenomings and 125,000 deaths annually [3]. Although snakebite envenoming can be found in all inhabited continents, it is mostly a public health concern in tropical and sub-tropical areas in developing countries, and in part, this is due to the lack of health system resources found within these countries.

Snake venoms are an intricate mixture of bioactive compounds, and their composition varies extensively between and within snake species. They primarily comprise small molecules—such as neurotransmitters or polyamines—salts, metals and proteins, the latter being the most abundant type of biomolecule found. Although multiple types of proteins can be found within snake venom [4–6], four stand out as generally being the most important ones based on toxicity and relative abundance [7,8]: three-finger toxins (3FTxs),

phospholipases (PLA_2s), snake venom metalloproteases (SVMPs) and snake venom serine proteases (SVSPs).

In addition to these proteinaceous toxins, earlier studies have also demonstrated the presence of other smaller molecules (<1000 Da) in snake venoms [9–11]. For a number of these molecules—or metabolites—such as peptides, amines or hydrocarbons, it was shown that they potentially play a role in aiding [8] or inhibiting [12] protein toxins within the venom. However, these small molecules have not been thoroughly studied due to proteins being the main and most recognizable cause of toxicity of the venom. Nevertheless, they could be of significant help when trying to study the inter- and intra-group variability of venoms and when aiming at gaining better holistic understanding of venom functioning.

Independent of their size, most of the toxins within snake venoms can perturb homeostasis of the human body, and the pathologies associated with snake venoms can be summarized in three main groups: hemotoxic, cytotoxic or neurotoxic [13]. Variability in the composition of venoms can be found both between different taxa [4] and within the same taxonomic unit [14–16].

This variability is not only found in snake venom proteinaceous toxins. Small molecules can also be found within snake venoms at different concentrations depending on taxonomic differences of the snakes and thus could also be used for classifying snake venoms. The purpose of these molecules within the venoms is not fully understood yet, partially because of the broad variety of different metabolites found, each of them potentially having one—or several—different functions [8]. Additionally, a generalized classification of the said metabolites has not been reached yet—probably due to the lack of information regarding these molecules. One of the purposes of this research will be looking at classification of small molecules in venoms depending on the source of the said metabolites, which can be useful for the classification of snake venoms but also for understanding the relevance of small molecules in this biological matrix by considering how they might be involved in, for example, human envenoming, maintaining venom gland integrity and ecological relationships between snake and prey. Three groups stand out: endogenous metabolites produced by the venom gland, relevant to the venom activity; endogenous metabolites unrelated to venom, generated through metabolic pathways separate from those associated with venom; and exogenous metabolites coming from the snake's environment.

The classification of snakes that this study aims for must be based on species, which are somewhat unique to the snake's taxonomic group. This would mean that, by creating the said model, we are not only developing a way of classifying snake venoms—and as such, the snakes from which the venoms were derived—but also a workflow for untargeted studies of the chemical species able to define each taxonomic group. Because the said species must be found in only that group for us to be able to classify the venom based on the said molecules, the biological and evolutionary reasons for those species to be found in the venoms could be inferred and further studied. For this purpose, a comprehensive list containing the m/z-values and retention times for all features found in the samples is automatically generated and included in "Table S1: Level of each metabolite" in the Supplementary Materials.

Within the order Squamata, two families stand out as the most dangerous ones: Elapidae and Viperidae [5]. The Viperidae are mainly known for their hemotoxicity [6], while Elapidae snakebite manifests through systemic neurotoxic and cytotoxic illnesses [17]. As can be seen in Figure 1, *Dendroaspis* and *Naja* are part of the Elapidae family, whereas *Crotalus* and *Bothrops* are part of the Viperidae family. Within the *Naja* genus, spitting and non-spitting cobras can be recognized, the first one being able to spit out venom at attackers. Thus, their venom is specialized in causing more pain, which is achieved by upregulating the PLA_2s toxins [18]. This taxonomic distribution can be found in Figure 1. No subclade was presented further than Asian *Naja* due to the lack of representation from non-spitting African cobras in the model.

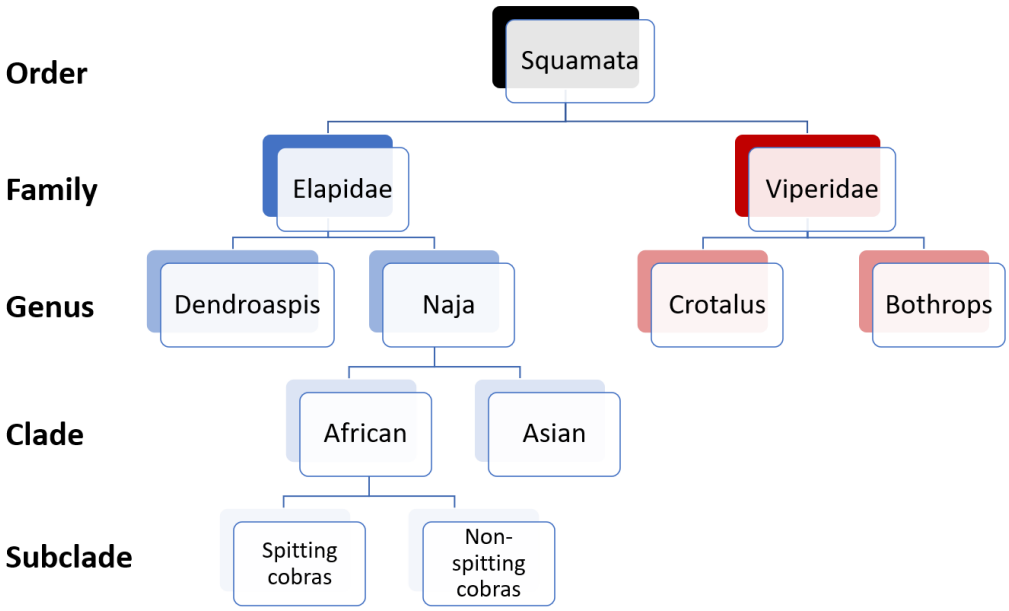


Figure 1. Order, family, genus, clade and subclade of the individuals that produced the samples.

To analyze the venoms of different species and focus on their variability, various statistical tools can be used. Principal component analysis (PCA) is one of the most important techniques used in exploratory data analysis. It was developed in 1901 by Karl Pearson [19], and it involves the projection of a dataset into a series of vectors that hold information about the variance found within the said dataset.

After PCA, other statistical techniques can be used to group the data and create a model able to classify samples based on the said groups. If a proper model is built, the family prediction of a snake based solely on the components of its venom can be produced.

This model could be applied in a high-throughput fashion for recognition and classification of snake venoms using data from liquid chromatography–mass spectrometry (LC–MS) analyses of venoms as input, followed by automated feature extraction from the mass spectra to, finally, introduction of the class-defining features into the model. This would eventually output—within certain confidence intervals—the taxon of the snake.

The automated extraction of data from the liquid chromatography–mass spectrometry (LC–MS) analyses, which is always performed when following the analytical workflow, also results in the creation of a comprehensive list containing all m/z-values and retention times of all compounds detected in each venom within the mass range used for detection (i.e., 100 to 1000 in our case). Thus, by following the analytical workflow, two main results are obtained: a classifier able to predict the family of the snake (in this case either Elapidae or Viperidae) and a comprehensive list containing the processed analytical data on all small molecules found for each venom.

The bioinformatic tools in combination with the analytical approach presented in this study can help better explain why some metabolites are found more commonly in some taxa and could indicate the evolutionary origin and function of the said small molecules.

2. Results

The overall workflow started with the LC–MS (Liquid Chromatography—Mass spectrometry) analysis of the small molecules in 50 different venoms—in duplicate—followed by extraction and alignment of the features found (i.e., all detected peaks in MS and their corresponding properties, such as m/z-value and retention time) for all the analyses. By doing this, a coherent matrix containing all samples (i.e., collection of features per venom and their intensity) is generated; the columns of the matrix are the features found, and the rows are the samples. This comprehensive list can be found in "Table S1: Level of each metabolite" in the Supplementary Materials.

Afterward, the measured samples were divided in three subsets in order to perform cross-validation of the model. After the validity of the model was proven, the model was looked at to check what features were recognized as relevant, and their concentrations in each venom were then investigated. This step was performed in conjunction with the identification of several metabolites associated with these relevant features for the model by means of the acquired MS/MS spectra of these metabolites or by standard addition when a metabolite was commercially available.

A graphic depiction of this workflow is included in Chart 1.

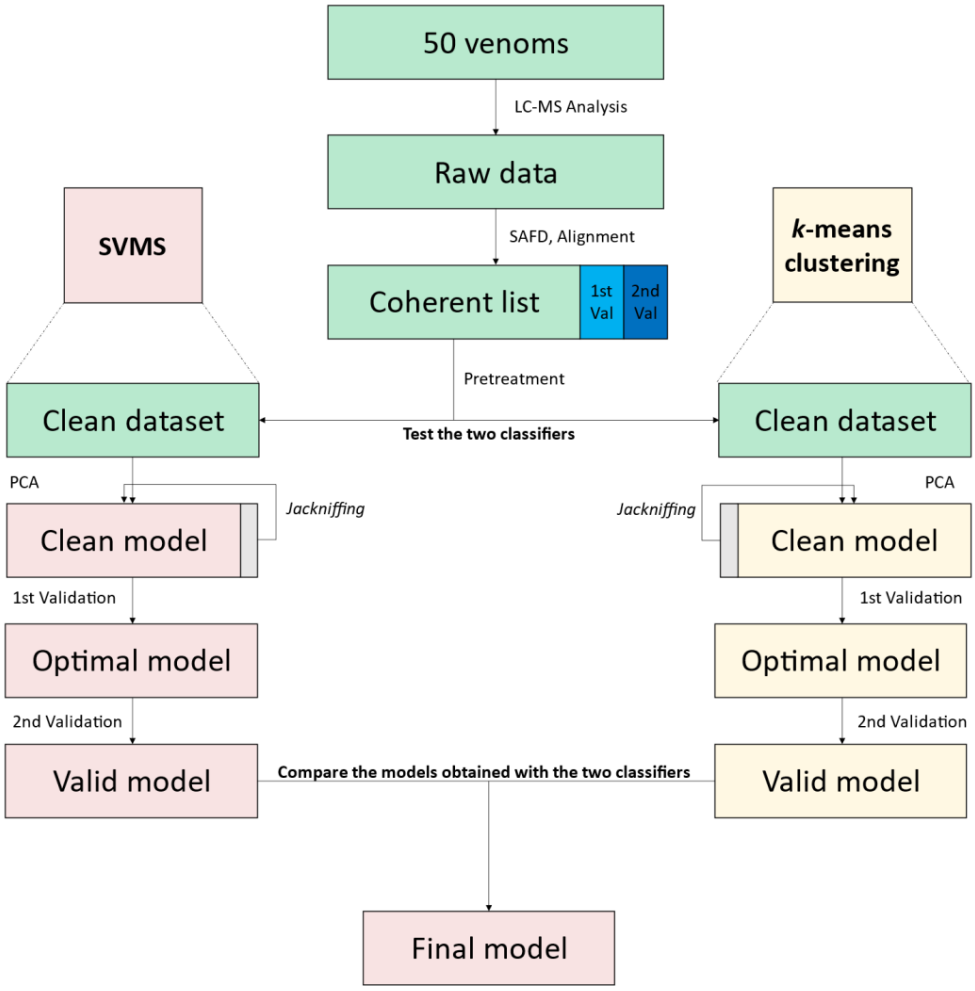


Chart 1. Graphic depiction of the workflow followed to develop the classifier model. After the development of the coherent list with all the metabolites and their levels in each venom, two classifiers were used to optimize the dimensionality reduction of this list.

2.1. Feature Extraction

To check for repeatability of the analyses, the venom of *Naja siamensis* was spiked with a mixture of standards—specified in Section S5—and analyzed several times before the analysis of all venoms was performed. This venom was chosen arbitrarily, as the same function could be attained by any other venom. As can be seen in "Figure S1—Analysis of repeatability" in the Supplementary Materials, the retention times of the features were sufficiently repeatable for the bioinformatic tool to extract the features and to recognize the same ones among the different repetitions—even if the retention times shifted slightly—meaning that the method was repeatable. The mixture of standards added to this venom was used to investigate whether normalization of the intensity of all features found by an internal standard would render more reproducible results. Thus, the intensities of the

signals of these standards were used to normalize the intensities extracted from the features found by means of recalculating all the intensities as a percentage relative to the chosen standard's signal intensity for all standards included separately. The highest repeatability was found when the intensities extracted were not normalized. This is probably because all intensities were normalized using a single standard at a time, and internal standards should preferably be similar to the compounds that are being standardized, which was impossible for this study, since none of the compounds measured were known at the point of measurement and of data processing. Additionally, taking into account the complexity of snake venom, no analyte can be similar to all of the metabolites found, meaning that no single internal standard could be used. Thus, no standardization was performed on the data. However, one of the standards (d9-Caffeine) was included in all the measurements to check for any issues during the analyses. By monitoring this standard in all measurements, the stability of the analyses could be investigated, from which it was found that the measurements were performed robustly.

After testing the repeatability of the analyses, the 50 different snake venoms (5 *Dendroaspis*, 15 *Naja*—divided into 5 African spitting cobras, 5 African non-spitting cobras, 4 Asian non-spitting cobras and 1 Asian spitting cobra—27 *Crotalus* and 3 *Bothrops*) were analyzed in duplicate by LC–MS in the order specified in "Table S2: Order of the injections" in the Supplementary Materials. The MS1 data from each analysis were exported and converted into .mzXML by MSConvert and run through the SAFD bioanalytical tool, which extracts the features and their intensities from the .mzXML files, in conjunction with "Computer code S1" of the Supplementary Materials, which aligns the features between samples, allowing us to obtain a coherent list containing all features and their retention times, m/z-values and intensities for each sample.

Afterward, all samples were divided in three subsets: the model subset, the first validation subset and the second validation subset. The model subset was used to generate a model that would undergo the different validation tests. The first validation subset was used to test the optimal number of PCs used to project the samples onto. The second validation subset was used to test the validity of the model created after the number of PCs had been chosen.

2.2. Model Building

The model we developed to classify snakes based on the metabolome of their venoms was built in four steps: pretreatment of the data, PCA, jackknifing and validation.

The data were pretreated to enhance the reproducibility of the method by only considering the features that would appear in both repetitions of a venom when their signals were at least three times the intensity of those features when they were found in Mili-Q Water, which was used as a (blank) sample. Additionally, a variable called "Rep" was defined, which would allow a feature to be taken into account only when it was found across a "Rep" number of samples. Finally, the data were mean centered and autoscaled. By doing this, a clean dataset focused in the variability of each feature was obtained.

PCA is an explorative matrix decomposition method, which allows for dimensionality reduction [20]. It also generates new variables (principal components, PCs), which are composed of the observed variables. The importance of each observed variable in each PC is defined by the loadings matrix, which, when the observed variables are projected onto it, leads to the scores matrix. The scores matrix is thus the projection of the original dataset onto the principal components, allowing for easy multivariate representation of the data within two or three dimensions. More information on this is given in "Section S1—Further information regarding Principal Component Analysis" in Document S1 in the Supplementary Materials.

Leave one out—or *jackknifing*—is a resampling-without-replacement method introduced by Maurice Quenouille in 1949 [21]. It consists of extracting part of the samples with which the model is built and rebuilding the model without the said samples. If performed several times, it allows for the creation of numerous models, depending on how many

samples are extracted and what samples they are. In this case, it was used to optimize the Rep variable. For each possible value of Rep, four samples were extracted from each family—Elapidae and Viperidae—and the models were built. Then, the eight samples extracted were projected onto the loadings to obtain their scores and to be classified. This was repeated 1000 times per Rep value to check the robustness of the model. This way of validating a model is called cross-validation, and it is useful when trying to optimize certain variables within the same dataset, as we do not need to obtain copious amounts of data. Additionally, it avoids bias in the creation of the model due to virtually using all the data as samples to be classified.

As mentioned, two validations were performed with data subsets, which were not part of the model-building data subset. The first validation subset was used to optimize the number of PCs to be considered within the clean PCA model (built with the clean dataset), thus creating the optimal model. The second validation subset was used to validate the final model and check its ability to classify unknown samples.

2.3. Classification

Some statistical techniques can be used to group the data and create a model able to classify samples based on the said groups. There are several tools that can be used for this goal. For this case study, two will be tested: one probabilistic and one binary linear.

Proximity classification after *k*-means clustering. Introduced by MacQueen in 1967 [22], the clustering method consists of analyzing the difference between the individual scores' sample vectors within the PCA model and the mean vector of each group or cluster. Thus, the Euclidean distance between each nth dimensional score and each cluster center is calculated.

To perform the classification, these two values are summed, and an algorithm is written, so that the distance to each center is divided by the said sum and multiplied by 100. This value is recorded as the first percentage, and the other percentage is calculated by subtracting the said number from 100. Thus, two percentages are the outcome. Because *k*-means clustering is based on proximity to the cluster center, the highest percentage was chosen as the indicator of the class of the snake family. These percentages are classified in five groups, which represent how certain the algorithm is about the samples coming from the specified cluster based on the indicator mentioned—the highest percentage. This classification is specified in Table 1.

Table 1. Classification of the classifying ability of the *k*-means models.

Qualitative Classification	Thresholds
Hard hits	(80, 100] %
Hits	(55, 80] %
Unsure	(45, 55] %
Miss	(20, 45] %
Hard miss	[0, 20] %

Support vector machines (SVMs). Introduced in 1992 by Boser, Guyon and Vapnik [23], SVMs work by generating a hyperplane, which separates the two classes of data. The samples will be classified as either of the groups depending on which side of the hyperplane they lay on. The SVM chosen during research was a linear Kernel with a hard margin.

2.4. Workflow

2.4.1. Model That Used *k*-Means Clustering

After applying PCA, the *jackknifing* step revealed that the appropriate value for the Rep variable was 36, because in 72 of the 100 iterations, all the samples that had been previously extracted were correctly classified. This was the highest number of correctly assigned sub-models. This led to the model being created with seven variables.

The first validation revealed that the number of PCs that could be used to correctly assign the first validation subset was every number in between three and seven. However, if the number of PCs is the same as the number of variables, there is an overfitting issue. Thus, according to the workflow explained in "Section S2—Further information regarding *k*-means clustering modelling" in Document S1 in the Supplementary Materials, the number of PCs chosen was four, as adding more PCs would lead to overfitting of the data [24].

For the second step of the validation, the second validation subset was projected onto the principal components and classified based on Table 1.

2.4.2. Model That Used Support Vector Machines

The lines of code that run the model building can be found in the "Computer Code S2" document of the Supplementary Materials.

After applying PCA, the jackknifing step revealed that the Rep value that generated the most robust model (98.1% of the *jackknifing* models correctly classified all the *jackknifing* samples) was 20, and the model contained 20 variables. Therefore, 20 features were chosen as relevant to the classification; the m/z-values of those features were: 150.0, 152.0, 175.0, 193.0, 203.2, 205.1, 215.0, 216.0, 237.0, 345.2, 385.2, 399.2, 413.1, 430.2, 431.2, 444.2, 445.2, 859.3 and 971.2 m/z.

The first validation step revealed that the number of PCs that could be used to correctly assign the first validation subset was every number in between 3 and 20. However, if the number of PCs is the same as the number of variables, there is a high chance of overfitting. Thus, according to the workflow explained in "Section S3—First Validation of the *k*-means clustering model" in Document S1 in the Supplementary Materials, the number of PCs chosen was 10, as adding more PCs would lead to overfitting of the data.

For the second validation, the data matrix was multiplied by the loadings and then classified by the SVMs.

2.5. Results of PCA of the Small Molecules Found in Snake Venom

2.5.1. *k*-Means Clustering

The results of the analysis performed on this model can be found in "Section S4—Results of the *k*-means clustering model" in Document S1 in the Supplementary Materials.

Although the optimized model was able to correctly classify most of the samples, only 50% were classified as hard hits. This lack of certainty could be explained by the overall trend of the Viperidae family not being defined by a single direction in any of the PCs—in contrast to what we found for Elapidae. This distribution over two PCs can result in lower classification power, as more extreme values for the seven features—the ones defined as relevant by the model—within the Viperidae family lead to a scattering of the samples all over the subspace created by the PCs (instead of close clustering), rather than an increase in density of the Viperidae cluster.

2.5.2. Support Vector Machines

As mentioned, a classifier based on SVMs with Rep = 20 and n° PCs = 10 was built. A 2D representation of this model, including the scores projected onto the two first PCs, can be found in Figure 2. Two PCs were chosen for this representation, as it is easier to visualize, but the actual subspace is contained in 10 dimensions, as they were the ones chosen to describe most of the variance of the model.

A table containing the values for all the loading values is included in "Table S3: Relevance of each feature" in the Supplementary Materials.

The clustering in Figure 2 already looks more organized than the one we could see for the model obtained when Rep was set to 36 (Figure S3 of "Section S4—Results of the *k*-means clustering model" in Document S1 in the Supplementary Materials, where the representation of the optimized *k*-means model can be found). Whereas Elapidae scores seem to distribute most of their variance in one direction—almost solely along the second PC—Viperidae scores are distributed along both PCs, although mostly along the first one.

Figure 2 also shows how non-spitting cobras seem to have high values for PC1 and low values for PC2 compared to the rest of the Elapidae family, which indicates that further classification could probably be performed when a bigger dataset is used.

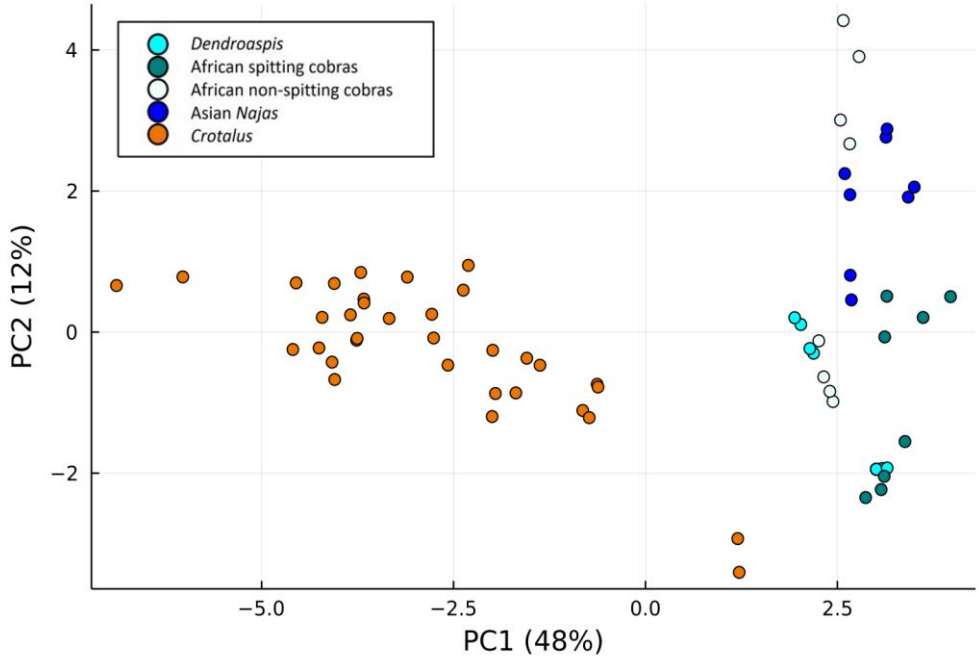


Figure 2. Two-dimensional representation of the scores generated by the model for Rep = 20 and nº PCs = 10 of PCs 1 and 2. Orange colors represent Viperidae venoms; dark blue colors represent Asian *Naja* venoms; white colors represent African non-spitting cobra venoms; dark cyan colors represent African spitting cobra venoms; and light cyan colors represent *Dendroaspis* venoms.

To understand the information held in the PCs, their loadings are represented in "Section S5—Representation and analysis of the Loadings in the SVM model" in Document S1 of the Supplementary Materials. The first principal component gives similar relevance to each of the features, using all of them to define most of the variance in the model. However, while features 1, 2 and 20 (the ones with weights higher than 0) are the only ones to be considered when classifying a sample as a member of the Elapidae family, features 6, 7 and 8 (the ones with the lowest weights) are essential when classifying a sample as Viperidae and explaining most of their variance.

It is interesting to look at all the features and samples at the same time to better understand the patterns within the data; therefore, the autoscaled values of those 20 features were summarized in a heatmap depicted in Figure 3.

Similar to what we found when Rep = 36 was used to create the model (Figure S6 of "Section S4—Results of the *k*-means clustering model" in Document S1 in the Supplementary Materials), the first features and the last one (150.0, 152.0, 175.0, 193.0 and 971.2 m/z) seem to be consistently more abundant in Elapidae than in Viperidae. Because the last feature (971.2 m/z) also holds some extreme variability within the Elapidae family, but its intensity can be similar to that of Viperidae in some cases, this feature is pushed back to the fifth component. Nonetheless, the first four variables—with quite different values between the two families—are found in the second principal component and thus explain much more variance. As can also be seen in the heatmap, features 7 and 9 seem to have somewhat of an influence when defining the Elapidae. These variables are held inside the second, third and fifth components. If we project the scores of those second and third PCs—only the most important two out of the three are used for clarity in the representation—the resulting graph (Figure 4) should explain most of the variance found within Elapidae. Only the said

family is represented for clarity, as introducing the Crotalus family would make the image too crowded, and it would add no additional value.

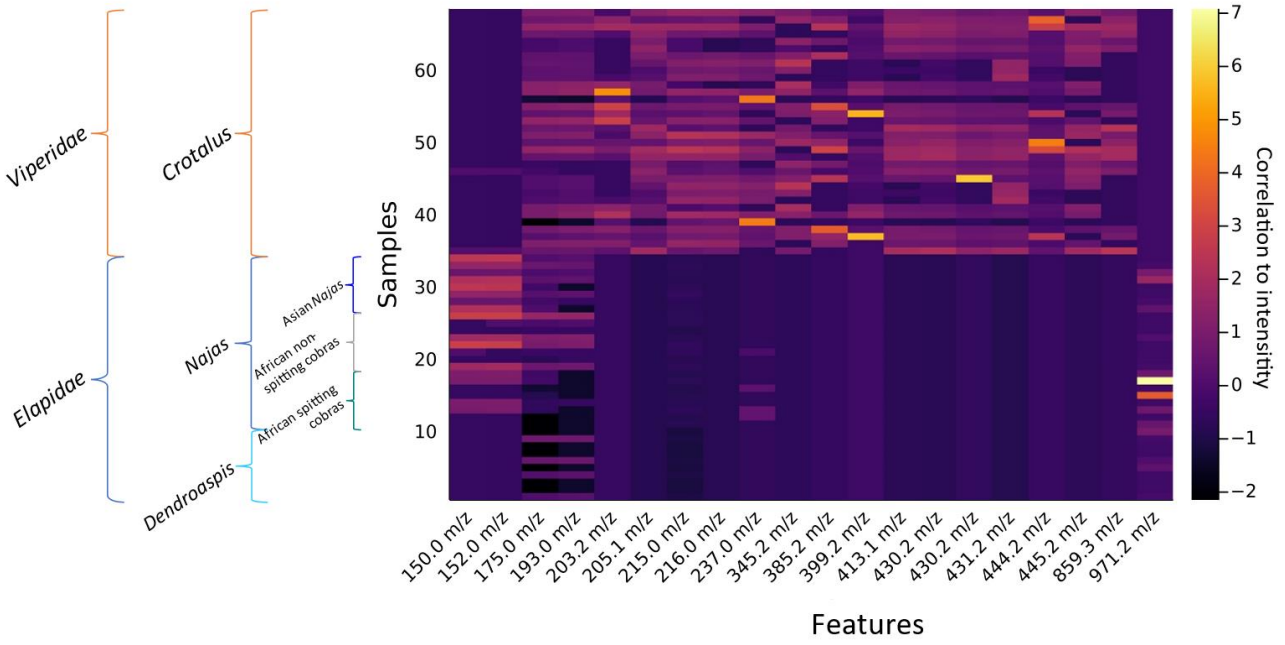


Figure 3. Heatmap of the autoscaled matrix for PCA with Rep = 20. Twenty features were considered relevant, and, in the figure, a correlation between the color and the original data can be seen (lighter colors signify more signal, and darker colors signify less signal). The values represented are the ones from the autoscaled matrix. The subclades were also classified for biological relevance based on their family and genus.

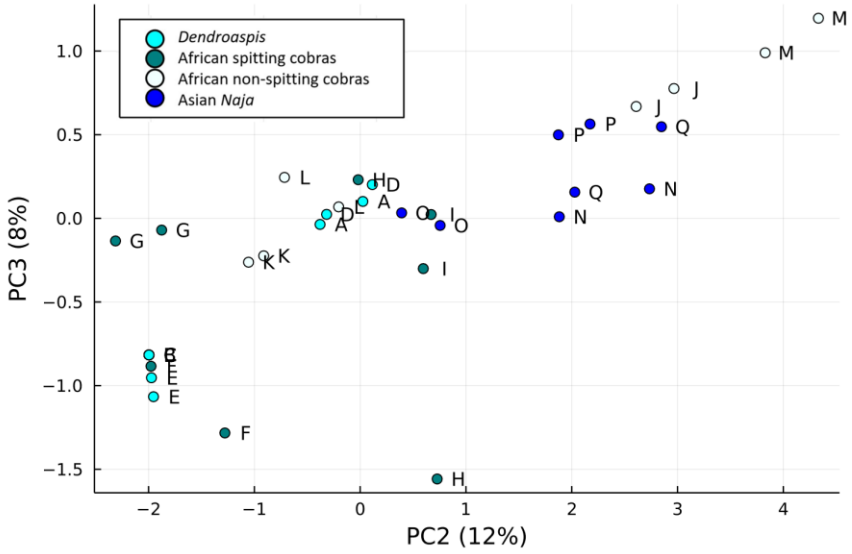


Figure 4. Two-dimensional representation of the Elapidae scores generated by the model for Rep = 20 and nº PCs = 10 of PCs 2 and 3. Subspecies were also indicated. Dark blue colors represent Asian *Naja* venoms; white colors represent African non-spitting cobra venoms; dark cyan colors represent African spitting cobra venoms; and light cyan colors represent *Dendroaspis* venoms. Legend: A: *Dendroaspis angusticepsis*, B: *Dendroaspis jamesoni jamesoni*, C: *Dendroaspis jamesoni kaimosae*, D: *Dendroaspis polylepsis*, E: *Dendroaspis viridis*, F: *Naja mossambica*, G: *Naja nigricinta*, H: *Naja nigricollis*, I: *Naja nubiae*, J: *Naja haje*, K: *Naja legionis*, L: *Naja nivea*, M: *Naja subfulva*, N: *Naja atra*, O: *Naja kaouthia*, P: *Naja naja*, Q: *Naja oxiana*.

It is noticeable to mention that the feature with 215.0 m/z seems to be able to clearly differentiate between *Dendroaspis* and *Naja* while also differentiating *Dendroaspis* from *Crotalus*, as occurred in the *k*-means model. However, no PC relies on this variable to generate such clustering, probably due to the low number of *Dendroaspis* samples, which biases the model generation. PC 3 does take this variable into account, but, because this principal component also considers the feature with 237.0 m/z, the ability of the model to cluster *Dendroaspis* separately is strongly diminished, as can be seen in Figure 4.

Most of the variability found in Elapidae useful for classification can be essentially summarized in features 1, 2, 3, 4, 7, 8 and 9, which—by looking at Figure 4—means that these two PCs should be sufficient for explaining most of the relevant variability. However, the feature with 215.0 m/z is not relevant enough in these PCs for us to see the clustering of *Dendroaspis*, probably due to the aforementioned bias. However, by adding more samples to the model, this variable should stand out as a paramount feature to define and cluster the said group. As will be described in the Section 3, this feature is thought to derive from the Deoxyribose 5-monophosphate metabolite, commonly known for partaking in the pentose phosphate pathway, which is used to anabolize precursors for the synthesis of nucleotides and to maintain the redox state of the cell [25].

The PCs in the model that hold the highest variance and that contain information concerning the features that define the variability found within Viperidae are the first PC (as it holds information concerning variables 5 to 19 and explains most of the variability within the model) and the second one (as it contains information concerning the variability of features 3 and 4, which also define Viperidae). These PCs can be found in Figure S8 of "Section S5—Representation and analysis of the Loadings in the SVM model" in Document S1 of the Supplementary Materials.

After analyzing the model built for Rep = 20 and nº PCs = 10 (Figure 2), the samples within the second validation subset were projected onto this model, thereby obtaining their scores. In Figure **??**, the scores of both validation subsets were included with those already in the model. Only snakes from the Viperidae family were chosen for the second validation model to test its ability to predict the family of the snake from which a specific venom originates. It is relevant to mention that some of the said samples derive from a genus that was not included in the model—*Bothrops*—which was done to test the ability of the model to predict the family of a genus, which had not been presented to the model. In brief, this means that if the model was able to correctly classify those samples, the features found—and their concentrations—could be assumed to be homogeneous throughout the different families. However, this statement should be explored in further studies, as the prediction of one genus, which was not included in the model, cannot be extrapolated to all other genera.

The results of the validation demonstrated that 100% of the second validation subset samples were correctly classified, thus showing that the SVM model rendered excellent results, whereas the *k*-means approach did not. Even samples coming from a genus unseen by the SVM model (*Bothrops*) were correctly identified within the Viperidae family. As can be seen, the representation of the second validation shows how the clustering defined by the hyperplane works, as the distribution of the validation samples does not seem to follow a random pattern, but rather, the same trend, which could be seen in the model itself, can be found for the validation subsets.

To understand the reason for this clustering, the features present in the model were looked into more thoroughly in order to recognize the metabolites that were associated with the said features, and thus, the metabolites that were able to discriminate between different taxa.

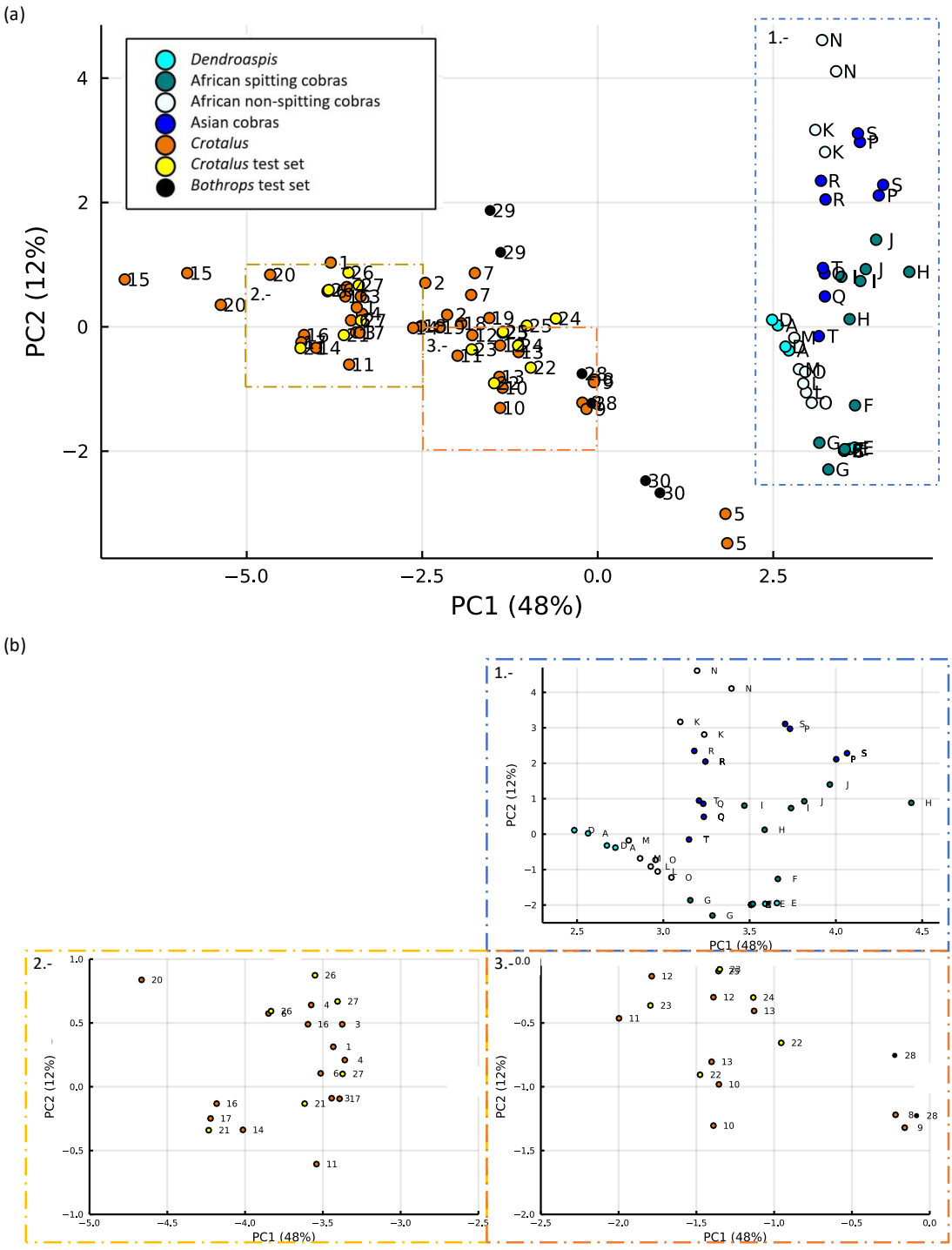


Figure 5. (**a**) Two-dimensional representation of PCs 1 and 2 holding the scores generated by the totality of the samples when using the PCs obtained for the model set with Rep = 20 and nº PCs = 10. Red colors represent Viperidae snakes, while blue and white colors represent Elapidae ones. All the genera, species, subspecies and clades are also shown. (**b**) Three zoom-in views (i.e., 1, 2 and 3) of the densest parts of a) to further clarify the relative position of each of the samples. Black dots represent *Bothrops* venoms; yellow dots represent *Crotalus* venoms in the second validation set; orange dots represent *Crotalus* venoms in the model dataset; dark blue dots represent Asian *Naja* venoms; white dots represent African non-spitting cobra venoms; dark cyan dots represent African spitting cobra venoms; and light cyan dots represent *Dendroaspis* venoms. Legend: A: *Dendroaspis angusticepsis*, B: *Dendroaspis jamesoni jamesoni*, C: *Dendroaspis jamesoni kaimosae*, D: *Dendroaspis polylepsis*, E: *Dendroaspis*

viridis, F: *Naja mossambica*, G: *Naja nigricinta*, H: *Naja nigricollis*, I: *Naja nubiae*, J: *Naja pallida*, K: *Naja annulifera*, L: *Naja haje*, M: *Naja legionis*, N: *Naja nivea*, O: *Naja subfulva*, P: *Naja atra*, Q: *Naja kaouthia*, R: *Naja naja*, S: *Naja Naja siamensis*, T: *Naja* oxiana, 1: *Crotalus adamanteus*, 2: *Crotalus culminatus*, 3: *Crotalus durissus cumanensis*, 4: *Crotalus durissus terrificus*, 5: *Crotalus vegrandis (1)*, 6: *Crotalus vegrandis (2)*, 7: *Crotalus enyo*, 8: *Crotalus horridus*, 9: *Crotalus horridus atricaudatus (1)*, 10: *Crotalus horridus atricaudatus (2)*, 11: *Crotalus lepidus kluberri*, 12: *Crotalus lepidus maculosus*, 13: *Crotalus morulus*, 14: *Crotalus molossus molossus*, 15: *Crotalus molossus nigricans*, 16: *Crotalus molossus oaxacus*, 17: *Crotalus phyrrus (1)*, 18: *Crotalus phyrrus (2)*, 19: *Crotalus cerberus*, 20: *Crotalus ravus ravus*, 21: *Crotalus lorenzoensis*, 22: *Crotalus ruber ruber*, 23: *Crotalus scutellatus salvini*, 24: *Crotalus scutellatus scutellatus*, 25: *Crotalus tortuguensis*, 26: *Crotalus triseratus triseratus*, 27: *Crotalus viridis viridis*, 28: *Bothrops asper*, 29: *Bothrops alternatus*, 30: *Bothrops atrox*.

In the first place, the identification of the metabolites was performed via comparison of the m/z-values of the features with two databases of metabolites; one included the ones found in snake venom and the other one metabolites only found in humans [8,26]. As a second step, the metabolites from which we had stock solutions were spiked onto the venom samples, which contained features to check whether their intensity increased after the addition. Because we did not have stock from all expected metabolites, LC–MS/MS was performed on a pooled sample containing four of the venoms (*Naja Nigricinta*, *Crotalus vegrandis (2)*, *Crotalus culminatus* and *Crotalus lorenzoensis*), which, in conjunction, contained all the metabolites relevant to the model. With the acquired data, the MS/MS spectra resulting from fragmentation of accurate masses that correlated to features of the specific metabolites under investigation, could be compared with MS/MS fragmentation spectra retrieved from the Human Metabolome Database (HMDB) [26], which contains all human metabolites found and relevant information about them, such as their mass and fragmentation spectra. This database was chosen, as it is the only one that includes MS/MS fragmentation spectra of a high number of metabolites (220,945). When a match was found for a feature's accurate mass and its MS/MS spectrum, that feature was designated as the metabolite listed in the database.

In Table 2, the names, retention times and m/z-values of each metabolite relevant to the model can be found. The second-to-last column accounts for the confidence level at which the metabolites were determined, and the last column specifies the genera for which that feature was found in more than 50% of the genera samples. The confidence level was assigned by means of standards within the Metabolomic Standards Initiative (MSI) [27]. Level 1 confidence corresponds to validated identification—confirmation by use of a standard. Level 2 confidence signifies that both the m/z-value and the MS/MS pattern correspond to those of the mentioned metabolite. The level of confidence "- " corresponds to lack of confidence level assignment, as only m/z-values were cross-referenced. Because Villar-Briones and Aird [8] (referenced as (1) in Table 2) analyzed metabolites in venoms, whereas Wishart et al. (HMDB) [26] (referenced as (2) in Table 2) took into account metabolites only found in humans, it is notable to mention that the metabolites in accordance with Villar Briones and Aird's data are expected to be identified more reliably due to the fact that we are dealing with snake venom metabolites and not human metabolites.

All relevant metabolites for the model eluted from LC within the first 20 min of the LC–MS analyses, which means that the analysis time could probably be adapted and shortened after this time threshold without losing any crucial information when the analyses are performed for snake venom family classification by means of the defined SVM model.

More confidence in the identification of metabolites with no MSI confidence level assignment would be needed to robustly elucidate the specific venom components that underpin the differences between different snake genera.

For further understanding of the patterns found in snakes' venom metabolites, the relevant features of all analyzed samples were explored in the heatmap found in Figure 6.

Table 2. Summary of the name, *m*/*z*-values, retention time, identification confidence level and the main clade they are found in of each relevant metabolite.

Feature	Metabolite	*m*/*z*-Value	Average Retention Time (min)	Confidence Level (MSI)	Mostly Seen in the Clade:
1	Methionine (1)	149.95	4.2	1	Asian cobras
2	Guanine (1)	151.95	4.4	2	Asian cobras
3	Aconitic acid (1)	175.02	5.3	2	All but African spitting cobras
4	Citric acid (1)	193.03	5.3	1	All but African spitting cobras
5	4-Ethylphenilsulphate (2)	203.22	4.3	2	*Bothrops* and *Crotalus*
6	Tryptophan (1)	204.10	17.5	1	*Crotalus*
7	Deoxyribose 5-monophosphate (2)	215.02	5.1	-	All but *Dendroaspis*
8	O-Phospho-4-hydroxy-L-threonine (2)	216.02	5.1	-	*Bothrops* and *Crotalus*
9	[Unknown]Na+	237.00	4.9	-	*Bothrops* and *Crotalus*
10	pEKS (tripeptide) (1)	345.18	5.1	2	*Crotalus*
11	TPPA (tetrapeptide) (1)	385.21	5.2	2	*Crotalus*
12	pERI (tripeptide) (1)	399.24	13.2	-	*Crotalus*
13	Unknown	413.15	16.2	-	*Crotalus*
14	pENW (tripeptide) (1)	430.17	16.3	2	*Bothrops* and *Crotalus*
15	Unknown	430.27	16.3	-	*Crotalus*
16	Unknown	431.17	16.0	-	*Bothrops* and *Crotalus*
17	Unknown	444.20	16.2	-	*Bothrops* and *Crotalus*
18	Unknown	445.19	16.4	-	*Bothrops* and *Crotalus*
19	Unknown	859.34	16.1	-	*Crotalus*
20	Unknown	971.17	17.8	-	African spitting cobras

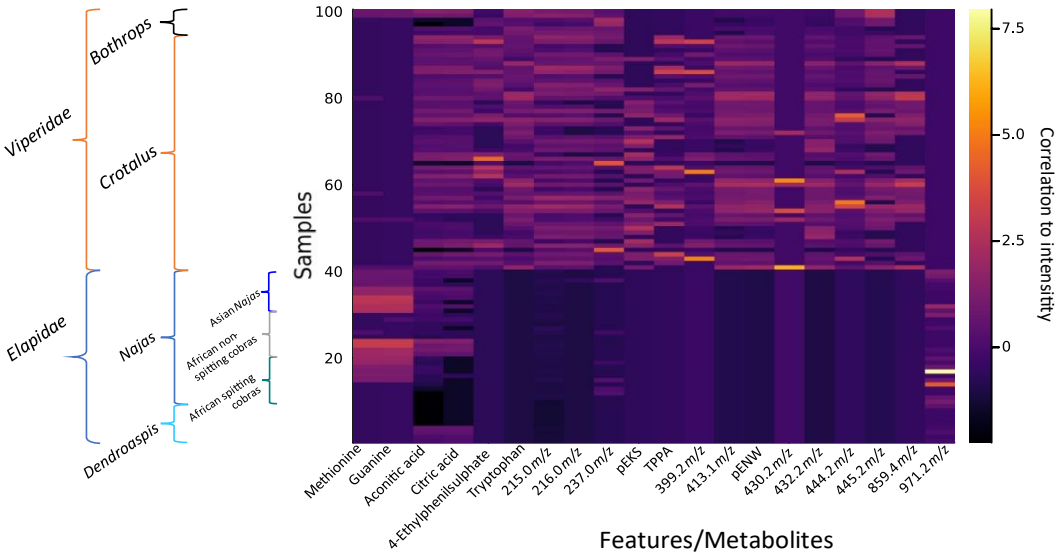


Figure 6. Heatmap of the autoscaled complete dataset for PCA with Rep = 20. These 20 features are relevant to the model, and, in the figure, a correlation between the color and the autoscaled data is given (lighter colors signify more signal, and darker colors signify less signal). The family, genus and clade were also classified for biological relevance based on their family and genus.

3. Discussion

Several patterns could be found regarding the intensities of these features, which the model found relevant. These patterns were able to differentiate between families and even genera and clades, meaning that there is probably a biological reason as to why the scores of these samples are distributed in such manner. It is important to mention here how the model is able to group both *Bothrops* and *Crotalus* venoms under the same family—Viperidae—even though the intensity of several features does not comply with the general trend of Viperidae.

When analyzing the patterns found in the intensity of each feature, the fact that citrate was mostly seen in Viperidae venoms, whereas it was not found in *Dendroaspis*, stands out. Citrate is known for its protease inhibition activity, due to which it is thought to be present in snake venom to prevent degradation of the venom by its own metalloproteases [28]. Additionally, this study also reported that citrate was able to inhibit a PLA_2, which was investigated. As can be seen in Figure 6, citric acid is present, on average, in higher concentrations in the Viperidae family, which is known to have venom compositions that contain, on average, much higher concentrations of proteases than Elapidae venoms [29]. This is in accordance with other studies, which analyzed citrate in snake venom [8,28]. This metabolite is known to work by chelating metals, such as zinc, thus being able to inactivate metalloproteases residing in the venom gland. Additionally, it lowers the pH of the matrix, thereby lowering the enzymatic activity of metalloproteases. Once a venom enters the prey's body, citrate can possibly also assist in disrupting the normal coagulation processes by chelating Ca^{2+}, the most important metal ion in the coagulation cascade [30]. Citrate is therefore suggested to be present in metalloprotease-rich venoms to prevent self-harm while potentially aiding in their hemotoxicity once the venom enters the prey's bloodstream. Spitting cobras can use their venom as a defense mechanism, thus, their venom aims toward cytotoxicity [31]. These venoms have low amounts of metalloproteases [32], and, with the hypothesis postulated above, it makes sense that almost no citrate was found in these venoms. Endogenous peptides, such as pENW, are also known to be able to inhibit metalloproteases [12,33–35], which could be a reason as to why they were mostly found in hemotoxic Viperidae venoms. The presence of these glutamate-containing oligopeptides in the venom is known to contribute to reducing proteolytic activity of the SVMPs, thus preventing venom degradation by its own toxins [35], which could explain the correlation between the concentration of these oligopeptides and SVMPs found in Viperidae venom.

Naja venoms—especially those from African non-spitting cobras—on average contained higher methionine concentrations than the other taxa. This metabolite is not found in the rest of the genera, meaning that methionine is not a trait of the Elapidae venoms but rather specifically of the *Naja* venoms.

Dendroaspis is the genus, which contained the least amount of metabolite with a feature that had a m/z-value of 215.02. As stated, this feature is suspected to derive from deoxyribose 5-monophosphate, which is known to take part in the pentose phosphate pathway (PPP). PPP is a metabolic pathway involved in the generation of NADPH and pentoses, and it has been demonstrated as a major regulator of cellular redox homeostasis and nucleotides biosynthesis [36].

Thirteen features and metabolites were found in most of the *Crotalus* genera, and metabolite identification of those features—with MS/MS and/or standards—would be needed to draw further conclusions. Tryptophan, an uncommon metabolite in other taxa, was frequently observed in the *Crotalus* venoms, for which we currently do not have a hypothesis.

There does not seem to be any specific metabolite able to fully define only the *Bothrops* genus. This might be due to the fact that only three *Bothrops* samples were analyzed and that these samples were not used in the model. However, even if the defined SVM model was not able to indicate which metabolites are unique for the said genera, it was indeed able to classify them as Viperidae due to the features these samples contained and their respective intensities.

Considering classification regarding the origin of the metabolites previously stated in the Introduction (endogenous metabolites produced by the venom gland, endogenous metabolites unrelated to venom and exogenous metabolites), it is interesting to note that endogenous metabolites, which could be related to the snake's metabolism rather than involvement in venom functioning—such as methionine and tryptophan—had different abundance levels between Elapidae and Viperidae. The only endogenous metabolite identified probably related to venom activity and/or venom gland protection (citric acid) followed the expected pattern of taxa differentiation based on the biological relevance of the said analyte within the venom. No exogenous metabolites were identified in this study within the set of features in the model.

In addition to creating a model able to discriminate between Elapidae and Viperidae based on their metabolites, this research can also be considered as an untargeted metabolomic study, as a list with all the features found (i.e., peaks detected in MS and their corresponding m/z-values and intensities) in all the venoms measured was created and included in "Table S1: Level of each metabolite" in the Supplementary Materials. These types of studies are helpful when identifying core metabolites in the venoms but also for elucidating the theories explaining why those species are biologically relevant. This study has certain limitations that allow for bias to potentially modify the results. For example, factors such as sex, age and/or geographical location were not considered in the model building in this study, as these data were not available for a large part of the venoms in our venom collection, and/or some venom samples were derived from pooled milkings of different individuals. Additionally, some of the analytes were found at the beginning of the chromatographic analyses—eluting almost with dead volume—which is known to be detrimental for their ionization in the MS due to the high number of species being ionized during this time frame. Because no traceback of those factors—and many others—could be performed, the samples must be assumed to represent the totality of their family, genus, species and subspecies. However, this study is the first attempt to initiate research that focuses on the metabolites found in venom, as they can also be useful for understanding these complex mixtures.

In order to be able to investigate specific metabolites and/or features, a dashboard tool can be used to easily visualize the variables of choice. For this, "Table S1: Level of each metabolite" in the Supplementary Materials is used and imported into the dashboard tool. We performed this for our dataset, and all the relevant representations of the data can be found in https://public.tableau.com/app/profile/luisce99/viz/Metabolome-basedclassificationofsnakevenomsbybioinformatictools/Dashboard1, (accessed on 11 February 2023). Due to the intrinsically interactive properties of the dashboard, the information selected can be sorted and presented in different representative ways, such as bar graphs and pie charts. This implies that users can choose the metabolites and/or features they want to investigate and from there be able to immediately obtain information regarding their abundance level in each venom, or in a clade or taxa. Examples of visualizations presented using the dashboard can be found in "Section S7—Examples of data visualization with the dashboard" in Document S1 of the Supplementary Materials.

This software is able to automatically sort the input data into different variables, clean and interpret the data, and understand SQL-type queries and implement them into one's dataset. It can also sort the data in different tables and create new variables with unions or intersects between different tables. The main reason this software was also used in this study is due to its ability to easily create quick representations of the input data and merge them into a single dashboard with all the required information represented in clearly structured overviews.

4. Conclusions

A high-throughput analytical LC–MS-based pipeline for analysis and automated extraction of all masses corresponding to small molecules in the range of 100 to 1000 m/z-values in snake venoms was set. Within this workflow, a comprehensive list containing all

the features found and their characteristics—m/z-values, retention times and intensities—in the samples was obtained. This list can be found in "Table S1: Level of each metabolite" in the Supplementary Materials.

After investigating two ways of classifying snake venoms by their metabolites (*k*-means and SVMs), the SVM approach showed not only much more robustness but also much more reliability when attempting to determine whether a venom originated from an Elapidae or a Viperidae snake. This is probably because, as we transform the sample matrix into a score matrix, the two groups are naturally separated within the subspace of the loadings. Thus, the SVM—being a binary linear method—recognizes this separation and takes advantage of it.

From the results of the validation of the model, it can be inferred that metabolites are consistently present in the snake venoms and that their levels are sufficiently similar between samples from the same family (and sufficiently different between families) to be used for family classification of snake venoms. This differentiation can also be found—with less confidence—within other taxa (genera, clades, etc.). MS/MS analysis of the said metabolites would lead to improvement in terms of their identification level. Some of the metabolites that could be identified presented distinctly different concentration levels in venoms of different families due to potentially explainable biological relevance. Hypotheses regarding the said levels are proposed in this study.

To further develop the project, analyses of additional and different venoms would be needed. By analyzing more venoms, the classification could reach and recognize deeper patterns in the data, thus theoretically being able to also fully classify venoms into subfamily, genus and eventually maybe even the diet of a specific snake. In order to achieve this, the said venoms should also include biological and geographical information, such as age, gender, location, etc., from the snake they came from, so that the result is more representative. Follow-up targeted studies could also be performed on the metabolites identified, as understanding the biological reasons for the presence of some of these small molecules in venoms could provide insightful information concerning snake venom.

This study mainly focused on setting up and demonstrating the analytical methodology to obtain a coherent list of metabolites and the venoms they are found in by means of developing a family classifier. Future studies should focus on the metabolite variation related to morphological and ecological parameters.

5. Materials and Methods

5.1. Chemical and Biological Reagents

Water was purified with a Milli-Q Plus system (Millipore, Amsterdam, The Netherlands). DMSO was supplied by Riedel-de-Haen (Zwijndrecht, The Netherlands). Acetonitrile (ACN; ULC/MS grade), trifluoroacetic acid (TFA) and formic acid (FA) were obtained from Biosolve (Valkenswaard, The Netherlands). All salts used for buffer preparation were of analytical grade and bought from Merck (Kenilworth, USA), Fluka (Bucharest, Romania) or Sigma-Aldrich (Darmstadt, Germany). Micro-90® concentrated cleaning solution was supplied by Sigma-Aldrich. Lyophilized snake venoms (Worksheet 1 of Workbook 1 of SI) were provided by the Centre for Snakebite Research & Interventions (Liverpool School of Tropical Medicine (LSTM), UK) and the historical collections of Freek J. Vonk (FV) and Manjunatha Kini (K) and stored on a long-term basis at −80 °C. Stock solutions of crude venoms (5.0 mg/mL) were prepared in water prior to analysis and stored at −80 °C. A total of 50 venoms were studied; 15 of those venoms originated from the genus *Naja* and 5 from the genus *Dendroaspis*, both within the Elapidae family. A total of 27 of those venoms originated from the genus *Crotalus* and 3 from the genus *Bothrops*, both within the Viperidae family. Volumes of 50 µL of 1 mg/mL solutions of each venom were injected into the HPLC-MS. Solutions of Phenylalanine, Caffeine, d9-Caffeine, Nortriptyline and Metoprolol (10 mM) were prepared in Milli-Q (except for Nortriptyline, which was dissolved in DMSO—Riedel-de Haen) and stored at −80 °C. They were used as internal standards for this study

due to their small size, the ability to absorb at the wavelengths studied and disparate retention times. All these chemicals were bought from Sigma Aldrich.

The origin of the lyophilized venoms, the order in which they were injected and the model the samples were used on are included in "Table S2: Order of the injections" in the Supplementary Materials.

5.2. Analysis of Small Molecules

5.2.1. Separation

Reverse-phase high-performance liquid chromatography (RP-HPLC, or LC for short) was carried out using a Shimadzu HPLC system managed by Shimadzu LabSolution software (Shimadzu, s-Hertogenbosch, The Netherlands). Four types of samples were analyzed in the order specified in Table 1 of the SI: 50 µL water injections acting as blanks, 50 µL of each of the mentioned venoms spiked with d9Caffeine (2.5 µM; used as a standard), a mixture of standards (Phenylalanine 2.5 µM, Caffeine 250 nM, d9-Caffeine 250 nM, Nortriptyline 250 nM and Metoprolol 250 nM) and 50 µL of Naja siamensis venom spiked with the said mixture to test the repeatability of the method and the capability of those standards to normalize the intensities of the features found in *Naja siamensis*. All samples were injected using a SIL-30AC autosampler, and the column used was a Waters Xbridge Peptide BEH300 C18 analytical column (100 × 4.6 mm, 3.5 µm particle size and 300Å pore size with a flow rate of 0.5 mL/min). Mobile phase A consisted of 97.9% H_2O, 2% ACN, 0.1% FA, and mobile phase B consisted of 97.9% ACN, 2% H_2O, 0.1% FA. The gradient program was set as follows: linear increase to 50% B in 30 min, followed by linear increase to 90% B in 4 min, isocratic elution for 5 min at 90% B, down to 5% B in 1 min and then equilibration for 10 min.

5.2.2. Detection

For all LC–MS analyses, a flow splitter was added after separation. 90% of the flow was directed toward an SPD-20A Prominence UV Detector to record UV spectra at both 220 and 254 nm. The remaining 10% of the flow was directed toward a MaXis QTOF mass spectrometer (Bruker Daltonics, Bremen, Germany) hyphenated to the HPLC via an ESI source operating in positive ion mode. The following parameters were used: temperature 220°C, capillary voltage 4.5 kV, gas flow 8.0 L/min, Nebulizer pressure 1.8 Bar. The spectra were stored at a rate of 2 Hz in the range of 100 to 1500 m/z-values. otofControl software was used for instrument control (Version 5.2-Build 0.9; Bruker, MA, US).

To acquire MS/MS data on the relevant metabolites found in the model, an Impact II mass spectrometer (Bruker Daltonics, Bremen, Germany) in data-dependent mode was hyphenated to the HPLC via an ESI source operating in positive ion mode. The following parameters were used: temperature 380 °C, capillary voltage 4.5 kV, gas flow 8 L/min, Nebulizer pressure 1.8 Bar, CI gradient from 25 to 55 kV.

5.2.3. Data Processing

Data processing consisted of several steps. First, the output data from the MS were converted to mzXML. This was performed by the MSConvert software with the thresholds specified in "Section S6—Thresholds applied on MSConvert" in Document S1 of the Supplementary Materials. Once the mzXML files were obtained, both SAFD—an algorithm able to recognize features by fitting a 3D Gaussian able to align the features within one sample by means of their m/z-values, retention times and intensities [37]—and the algorithm given in "Computer Code S1" in the Supplementary Materials were used to identify the features in each sample and align them into a coherent matrix. The latter based the said alignment on similarities between retention time windows and m/z-values. With this information, it was possible to recognize the same feature in different venoms, thus being able to align the same feature in all venoms, thus being able to create the mentioned coherent matrix with all venoms and features. After the coherent matrix was developed, the measured samples were divided in three subsets in order to perform cross-validation

of the model. After the validity of the model was proven, the model was looked at to check what features were recognized as relevant, and their concentrations in each venom were then investigated. This step was performed in conjunction with the identification of several metabolites associated with those features relevant to the model by means of the acquired MS/MS spectra of these metabolites or by standard addition when a metabolite was commercially available.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/toxins15020161/s1, Computer code S1: Feature Alignment, Computer code S2: Model, Documents S1, Figure S1: Analysis of repeatability; Table S1: Level of each metabolite; Table S2: Order of the injections; Table S3: Relevance of each feature.

Author Contributions: Conceptualization, J.K. and L.L.A.; methodology, L.L.A. and S.S.; software: L.L.A. and S.S.; investigation, J.S. and L.L.A.; resources, J.K. and S.S.; data curation, L.L.A.; writing—original draft preparation, L.L.A.; writing—review and editing, J.S., N.R.C., S.S. and J.K.; visualization, S.S. and L.L.A.; supervision, J.K. and S.S.; project administration, J.K.; funding acquisition, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the Wellcome Trust [221712/Z/20/Z]. For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in this article and supplementary materials.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hotez, P.J.; Aksoy, S.; Brindley, P.J.; Kamhawi, S. What constitutes a neglected tropical disease? *PLoS Negl. Trop. Dis.* **2020**, *14*, e0008001. [CrossRef] [PubMed]
2. Chippaux, J.-P. Snakebite envenomation turns again into a neglected tropical disease! *J. Venom. Anim. Toxins Incl. Trop. Dis.* **2017**, *23*, 38. [CrossRef] [PubMed]
3. Mohapatra, B.; Warrell, D.A.; Suraweera, W.; Bhatia, P.; Dhingra, N.; Jotkar, R.M.; Rodriguez, P.S.; Mishra, K.; Whitaker, R.; Jha, P.; et al. Snakebite Mortality in India: A Nationally Representative Mortality Survey. *PLoS Negl. Trop. Dis.* **2011**, *5*, e1018. [CrossRef]
4. Fry, B.G.; Scheib, H.; van der Weerd, L.; Young, B.; McNaughtan, J.; Ramjan, S.F.R.; Vidal, N.; Poelmann, R.E.; Norman, J.A. Evolution of an Arsenal. *Mol. Cell. Proteom.* **2008**, *7*, 215–246. [CrossRef]
5. Gutiérrez, J.M.; Calvete, J.J.; Habib, A.G.; Harrison, R.A.; Williams, D.J.; Warrell, D.A. Snakebite envenoming. *Nat. Rev. Dis. Prim.* **2017**, *3*, 17063. [CrossRef]
6. Slagboom, J.; Kool, J.; Harrison, R.A.; Casewell, N.R. Haemotoxic snake venoms: Their functional activity, impact on snakebite victims and pharmaceutical promise. *Br. J. Haematol.* **2017**, *177*, 947–959. [CrossRef]
7. Ferraz, C.R.; Arrahman, A.; Xie, C.; Casewell, N.R.; Lewis, R.J.; Kool, J.; Cardoso, F.C. Multifunctional Toxins in Snake Venoms and Therapeutic Implications: From Pain to Hemorrhage and Necrosis. *Front. Ecol. Evol.* **2019**, *7*, 218. [CrossRef]
8. Villar-Briones, A.; Aird, S. Organic and Peptidyl Constituents of Snake Venoms: The Picture Is Vastly More Complex Than We Imagined. *Toxins* **2018**, *10*, 392. [CrossRef] [PubMed]
9. Aird, S.; Villar Briones, A.; Roy, M.; Mikheyev, A. Polyamines as Snake Toxins and Their Probable Pharmacological Functions in Envenomation. *Toxins* **2016**, *8*, 279. [CrossRef]
10. Casewell, N.R.; Wüster, W.; Vonk, F.J.; Harrison, R.A.; Fry, B.G. Complex cocktails: The evolutionary novelty of venoms. *Trends Ecol. Evol.* **2013**, *28*, 219–229. [CrossRef]
11. Hutchinson, D.A.; Savitzky, A.H.; Burghardt, G.M.; Nguyen, C.; Meinwald, J.; Schroeder, F.C.; Mori, A. Chemical defense of an A sian snake reflects local availability of toxic prey and hatchling diet. *J. Zool.* **2013**, *289*, 270–278. [CrossRef] [PubMed]
12. Huang, K.-F.; Hung, C.-C.; Wu, S.-H.; Chiou, S.-H. Characterization of Three Endogenous Peptide Inhibitors for Multiple Metalloproteinases with Fibrinogenolytic Activity from the Venom of Taiwan Habu (*Trimeresurus mucrosquamatus*). *Biochem. Biophys. Res. Commun.* **1998**, *248*, 562–568. [CrossRef] [PubMed]
13. WHO. *Guidelines for the Prevention and Clinical Management of Snakebite in Africa*; WHO: Geneva, Switzerland, 2010.
14. Alape-Girón, A.; Sanz, L.; Escolano, J.; Flores-Díaz, M.; Madrigal, M.; Sasa, M.; Calvete, J.J. Snake Venomics of the Lancehead Pitviper Bothrops asper: Geographic, Individual, and Ontogenetic Variations. *J. Proteome Res.* **2008**, *7*, 3556–3571. [CrossRef]

15. Amorim, F.; Costa, T.; Baiwir, D.; De Pauw, E.; Quinton, L.; Sampaio, S. Proteopeptidomic, Functional and Immunoreactivity Characterization of Bothrops moojeni Snake Venom: Influence of Snake Gender on Venom Composition. *Toxins* **2018**, *10*, 177. [CrossRef] [PubMed]
16. De Oliveira, I.S.; Cardoso, I.A.; de Castro Figueiredo Bordon, K.; Carone, S.E.I.; Boldrini-França, J.; Pucca, M.B.; Zoccal, K.F.; Faccioli, L.H.; Sampaio, S.V.; Rosa, J.C.; et al. Global proteomic and functional analysis of Crotalus durissus collilineatus individual venom variation and its impact on envenoming. *J. Proteom.* **2019**, *191*, 153–165. [CrossRef] [PubMed]
17. Kalam, Y.; Isbister, G.K.; Mirtschin, P.; Hodgson, W.C.; Konstantakopoulos, N. Validation of a cell-based assay to differentiate between the cytotoxic effects of elapid snake venoms. *J. Pharmacol. Toxicol. Methods* **2011**, *63*, 137–142. [CrossRef]
18. Kazandjian, T.D.; Petras, D.; Robinson, S.D.; Van Thiel, J.; Greene, H.W.; Arbuckle, K.; Barlow, A.; Carter, D.A.; Wouters, R.M.; Whiteley, G.; et al. Convergent evolution of pain-inducing defensive venom components in spitting cobras. *Science* **2022**, *371*, 386–390. [CrossRef]
19. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [CrossRef]
20. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. A* **2016**, *374*, 20150202. [CrossRef]
21. Quenouille, M.H. Approximate Tests of Correlation in Time-Series. *J. R. Stat. Soc. Ser. B Methodol.* **1949**, *11*, 68–84. [CrossRef]
22. MacQueen, J. Some Methods for Classification and Analysis of Multivariate Observations. In *Multivariate Observations*; University of California Press: Oakland, CA, USA, 1967; p. 17.
23. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algortihm for optimal margin classifiers. In Proceedings of the COLT '92: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992; pp. 144–152. [CrossRef]
24. Catell, R.B. The scree test for the number of factors. *Multiva. Behav. Res.* **1966**, *1*, 140–161. [CrossRef] [PubMed]
25. Alfarouk, K.O.; Ahmed, S.B.M.; Elliott, R.L.; Benoit, A.; Alqahtani, S.S.; Ibrahim, M.E.; Orozco, J.D.P.; Cardone, R.A.; Reshkin, S.J.; Harguindey, S. The Pentose Phosphate Pathway Dynamics in Cancer and Its Dependency on Intracellular pH. *Metabolites* **2020**, *10*, 285. [CrossRef]
26. Wishart, D.S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B.L.; et al. HMDB 5.0: The Human Metabolome Database for 2022. *Nucl. Acids Res.* **2022**, *50*, D622–D631. [CrossRef] [PubMed]
27. Reisdorph, N.A.; Walmsley, S.; Reisdorph, R. A Perspective and Framework for Developing Sample Type Specific Databases for LC/MS-Based Clinical Metabolomics. *Metabolites* **2019**, *10*, 8. [CrossRef]
28. Odell, G.V.; Ferry, P.C.; Vick, L.M.; Fenton, A.W.; Decker, L.S.; Cowell, R.L.; Ownby, C.L.; Gutierrez, J.M. Citrate inhibition of snake venom proteases. *Toxicon* **1998**, *36*, 1801–1806. [CrossRef]
29. Tasoulis, T.; Isbister, G. A Review and Database of Snake Venom Proteomes. *Toxins* **2017**, *9*, 290. [CrossRef]
30. Bouchard, J.; Madore, F. Role of citrate and other methods of anticoagulation in patients with severe liver failure requiring continuous renal replacement therapy. *Clin. Kidney J.* **2009**, *2*, 11–19. [CrossRef]
31. Panagides, N.; Jackson, T.; Ikonomopoulou, M.; Arbuckle, K.; Pretzler, R.; Yang, D.; Ali, S.A.; Koludarov, I.; Dobson, J.; Sanker, B.; et al. How the Cobra Got Its Flesh-Eating Venom: Cytotoxicity as a Defensive Innovation and Its Co-Evolution with Hooding, Aposematic Marking, and Spitting. *Toxins* **2017**, *9*, 103. [CrossRef] [PubMed]
32. Dam, S.H.; Friis, R.U.W.; Petersen, S.D.; Martos-Esteban, A.; Lautsen, A.H. Snake Venomics Display: An online toolbox for visualization of snake venomics data. *Toxicon* **2018**, *152*, 60–64. [CrossRef]
33. Yee, K.; Pitts, M.; Tongyoo, P.; Rojnuckarin, P.; Wilkinson, M. Snake Venom Metalloproteinases and Their Peptide Inhibitors from Myanmar Russell's Viper Venom. *Toxins* **2016**, *9*, 15. [CrossRef]
34. Francis, B.; Kaiser, I.I. Inhibition of metalloproteinases in Bothrops asper venom by endogenous peptides. *Toxicon* **1993**, *31*, 889–899. [CrossRef] [PubMed]
35. Robeva, A.; Politi, V.; Shannon, J.D.; Bjarnason, J.B.; Fox, J.W. Synthetic and endogenous inhibitors of snake venom metalloproteinases. *Biomed. Biochim. Acta* **1991**, *50*, 769–773. [PubMed]
36. De Santis, M.C.; Porporato, P.E.; Martini, M.; Morandi, A. Signaling Pathways Regulating Redox Balance in Cancer Metabolism. *Front. Oncol.* **2018**, *8*, 126. [CrossRef] [PubMed]
37. Samanipour, S.; O'Brien, J.W.; Reid, M.J.; Thomas, K.V. Self Adjusting Algorithm for the Nontargeted Feature Detection of High Resolution Mass Spectrometry Coupled with Liquid Chromatography Profile Data. *Anal. Chem.* **2019**, *91*, 10800–10807. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.