Australian Government
Department of Defence
Defence Science and
Technology Organisation

Chemical Investigations of the Castor Bean Plant *Ricinus communis*

Simon P. B. Ovenden, Christina K. Bagas, David J. Bourne, Eloise J. Pigott and Warren Roberts

Human Protection and Performance Division
Defence Science and Technology Organisation

DSTO-TR-2786

ABSTRACT

In 2009 a National Security Science and Technology grant was awarded to the Human Protection and Performance Division for the investigation of several forensic aspects of the castor bean plant *Ricinus communis*. A major focus of this grant was to understand the chemical composition of the seeds, and to ascertain if these differences could be used for provenance classification. This technical report will discuss progress made during these investigations.

RELEASE LIMITATION

Approved for public release

Published by

Human Protection and Performance Division
DSTO Defence Science and Technology Organisation
506 Lorimer St
Fishermans Bend, Victoria 3207 Australia

Telephone: (03) 9626 7000
Fax: (03) 9626 7999

AR-015-479
December 2012

APPROVED FOR PUBLIC RELEASE

Chemical Investigations of the Castor Bean Plant *Ricinus communis*

Executive Summary

Ricinus communis (commonly known as the castor bean plant) is an introduced species that now grows wild in Australia. There are approximately 250 cultivars known. In addition to castor oil, the seeds also produce the toxic lectin ricin. Ricin is declared by the Chemical Weapons Convention as a Schedule 1 agent. These are chemicals that are highly toxic and have no legitimate uses. Consequently, ricin is of interest to state and national law enforcement agencies. Given the above information, strategies that are able to determine cultivar and provenance of an extract from *R. communis* seeds are of interest to these agencies.

In 2009, Human Protection and Performance Division (HPPD) was awarded a Prime Minister and Cabinet (PM&C) National Security Science and Technology (NSST) grant to study *R. communis* and establish forensic methods for dealing with potential ricin white powder incidents. A particular focus of this work was to investigate if there are any chemical signatures in the seed extracts that would allow for provenance classification. In particular, the following aims were proposed:

- to gain an understanding of the different cultivars present throughout Australia via an extensive national collection program;
- to establish analytical methods to provenance extracts of *R. communis* through the understanding of both the inorganic [Inductively Coupled Plasma Mass Spectrometry (ICPMS)] and organic [via Liquid Chromatography Mass Spectrometry (LCMS) and proton Nuclear Magnetic Resonance (^{1}H NMR) spectroscopy] chemical fingerprints; and
- to interrogate the collected data using multivariate statistical analysis for the identification of inorganic and organic markers of provenance.

During the collection program, a great morphological diversity in specimens of *R. communis* was observed in Victoria, New South Wales and South Australia. In particular, many specimens were sighted and collected that had variations in leaf size, shape and colour, stem and inflorescence colour, as well as seed pod colour, seed size and seed shape. Conversely, it appeared from our field observations that Queensland and Western Australia have virtually no diversity in their *R. communis* populations. It was also noted that during these collection efforts no specimens of *R. communis* were sighted in Darwin, Northern Territory.

The chemical analysis of the extracts of *R. communis* yielded some interesting results. Firstly it was found that analysing the 2% acidic *R. communis* extracts was not readily applicable to IRMS and ICPMS techniques due to interference from residual acetic acid. However, Laser Ablation-Inductively Coupled Plasma Mass Spectrometry (LA-ICPMS) of the whole seed allowed for provenance determination. The 2% acidic *R. communis* extracts were able to be analysed by LCMS with no subsequent loss in sensitivity. However, only cultivar of *R. communis* extracts analysed was determined using this method.

^{1}H NMR is a non-destructive, non-selective analysis which is able to detect every compound in a mixture containing protons. In the field of metabolomics, it has been identified as a prudent starting point for any metabolomic investigation. NMR also has the advantage of being an inherently quantitative technique. An NMR spectrum therefore allows for an estimation of the relative amounts of compounds present in a mixture. NMR also allows for compound structural information to be ascertained to at least a functional group level. When applied in conjunction with LCMS, a greater understanding of the chemical composition of the mixture is achieved. This combination of ^{1}H NMR and LCMS, when applied to the analysis of the 2% acidic *R. communis* extracts, allowed for cultivar and provenance determinations to be made with a high degree of certainty.

This technical report documents the progress made against the chemistry milestones contained in the NSST grant. This report will inform the clients of this work program (AFP, Chemical Warfare Agent Laboratory Network (CWALN) members, other national security clients) of some of the capability that HPPD has for handling these extracts, and the type of information that is able to be extracted from them.

Authors

Simon P. B. Ovenden
Human Protection and Performance Division

Simon graduated from a BEd-Sci in 1994 and from a BSc(Hons) in 1995 from The University of Melbourne. In 1999 he completed a PhD in marine natural products chemistry from the same institution. He then completed two years post doctoral research in Singapore at the Centre for Natural Product Research isolating and elucidating novel natural products as potential drug leads. Following this Simon spent three years at Cerylid Bioscience in Melbourne, then approximately one year at the Australian Institute of Marine Science, in both cases as a Senior Research Scientist researching novel natural products as potential drug leads from Australian biota. He joined DSTO in 2006 as a Defence Scientist, and is currently a memebr of the Biomolecules Analysis group in the Chemical Defence Branch. Here, Simon applys his background in NMR spectroscopy and LCMS in the analysis of highly toxic mixtures.

Christine K. Bagas
Human Protection and Performance Division

Christina Bagas graduated from a BBiotech and Cell Biol in 2005 and a BSc(Hons) in 2006 from La Trobe University in collaboration with DSTO. The thesis was titled a Study Towards the Mass Spectrometry Characterisation of Sulfur-Mustard Oligonucleotides Adducts. In 2007 she joined DSTO as a Defence Scientist. In 2008 she completed a DipEd from La Trobe University. Curremtly she is a member of the Biomolecules Analysis group in the Chemical Defemce Branch, where she is working in the metabolomics research program on R. communis.

David J. Bourne
Human Protection and Performance Division

David Bourne graduated with a BAppSc from the University of South Queensland in 1974. He then worked as a quality assurance manager at Abbott Australasia for two years before joining the School of Biochemistry at University of NSW. He graduated with a BSc (Hons) in 1983 followed by a PhD in 1991, both in biochemistry at UNSW. During latter stages of PhD studies David was employed at the Biomedical Mass Spectrometry Facility in the School of Phamacology UNSW. In 1991 he started a post doctoral fellowship at the Research School of Chemistry, Australian National University where he synthesised some novel phosphorazine calibrants for use in Selected Ion Mass Spectrometry (SIMS) and optimisation of flow-SIMS. David then moved to the Australian Institute of Marine Science as a research scientist on a drug discovery project. He then joined DSTO in 1999 as a chemist/mass spectrometrist. He became task manager of an LRR task in 2002 looking at chemical agent sensors and then task manager of a toxins project in 2005.

Eloise J. Pigott
Human Protection and Performance Division

Eloise Pigott graduated with Honours in a Forensic Science degree at Deakin University (Waurn Ponds) in 2006. She joined DSTO in 2007 where she works as a Defence Scientist in the Biomolecules Analysis Group (Human Protection Performance Division). Currently she is involved in the metabolomics research program investigating the use of human biofluids for early predictive diagnostics in human health.

Warren Roberts
Maritime Platforms Division

Warren graduated from Swinburne University of Technology with a B.App.Sci (Medical Biophysics and instrumentation, 1993), GradDip. (Social Statistics, 1997), and a M.App. Sci (Haemorheology, 1999). During both undergraduate and post-graduate studies he was associated with the Australian Institute of Sport (Canberra), working as a research assistant on the physiological and biochemical assessment of the elite athlete. Warren joined Hewlett Packard in 1997 in a role of systems administrator for an Asia-pacific Windows NT server network. In 1998 he commenced with DSTO as a Defence Scientist working in the broad area of exercise physiology and contributing to research programs in thermal physiology, ergonomics, biomarkers of physical and cognitive fatigue, physiological monitoring, Naval crew fatigue, and RAN Escape and Evacuation modelling. Recently he has been involved in providing bioinformatics expertise to the predictive diagnostics and CW programs within HPPD.

This page is intentionally blank

Contents

This page is intentionally blank

Abbreviations

ANOVA	Analysis of Variance
AQIS	Australian Quarantine Inspection Service
BPC	Base Peak Chromatogram
cvSE	Cross Validation Standard Error
DIMS	Direct Infusion Mass Spectrometry
DL	Detection Limit
DNA	Deoxyribonucleic acid
FTICRMS	Fourier Transform Ion Cyclotron Resonance Mass Spectrometry
HRMS	High Resolution Mass Spectrometry
ICP-AES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICPMS	Solution based Inductively Coupled Plasma Mass Spectrometry
IRMS	Isotope Ratio Mass Spectrometry
LA-ICPMS	Laser Ablation Inductively Coupled Plasma Mass Spectrometry
LCMS	Liquid Chromatography Mass Spectrometry
LD_{50}	Lethal Dose amount required to kill 50% of a given test population
LV	Latent Variables
MS	Mass Spectrometry
MWCO	Molecular Weight Cut Off
m/z	Mass to charge ratio
NMR	Nuclear Magnetic Resonance spectroscopy
NSST	National Security Science and Technology
OPLS-DA	Orthogonal Projection to Latent Structures-Discriminate Analysis
PC	Principal Component
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PLS-DA	Partial Least Squares Discriminate Analysis
PM&C	Prime Minister and Cabinet
PQN	Probabilistic Quotient Normalisation
Q^2X	Predictive strength
R^2X	Strongest cumulative variation
RCA	*Ricinus communis* agglutinin
RCB	*Ricinus communis* biomarkers
RNA	Ribonucleic acid
SSP	Seed Storage Proteins
TSP	3-(trimethylsilyl)-2,2,3,3-d_4-propionic acid
UV	Unit Variance
VIP	Variable Importance in the Projection

This page is intentionally blank

1. Introduction

The castor bean plant *Ricinus communis* was a popular garden ornamental in Australian gardens in the 1960s. Due to its nature of producing large amounts of fertile seeds which are dispersed effectively, the plant's progeny readily escaped the confines of domestic gardens. Consequently *R. communis* has become a significant environmental weed found in many and varied locations around Australia. In addition to the seed containing castor oil, it also contains the toxic protein ricin.

Ricin is a heterodimeric type II ribosome-inactivating protein that consists of two chains (an A chain and a B chain) linked by a disulfide bond.[1] The lectin B chain binds to glycoproteins and glycolipids expressed on cell surfaces, facilitating the entry of the protein into the cytosol.[1] The A chain then inhibits protein synthesis by irreversibly inactivating eukaryotic ribosomes from the 28S ribosomal RNA loop contained within the 60S subunit.[1] This process prevents chain elongation of polypeptides and leads to cell death.[1] Ricin has an LD_{50} by intravenous injection of approximately 5 mg/kg in standard mouse models[2] and is thought to have a human LD_{50} by injection of 5-10 mg/kg.[2]

Ricin is listed in Schedule 1 of the Chemical Weapons Convention,[3] with attempts to use ricin for assassinations previously reported.[4] Consequently there is interest within the defence and law enforcement communities to develop analytical methods to investigate the alleged use of ricin both in chemical weapons (which could be required under the provisions of the Chemical Weapons Convention) and forensic analysis of a crime scene.[5-8]

In 2009, the Human Protection and Performance Division (HPPD) was awarded a Prime Minister and Cabinet (PM&C) National Security Science and Technology (NSST) grant to study *R. communis* and establish forensic methods for dealing with potential ricin white powder incidents. In particular, the following milestones were proposed:

Milestone 1: To gain an understanding of the different cultivars present throughout Australia via an extensive national collection program.

Milestone 2: To establish analytical methods to provenance extracts of *R. communis*. This was performed using two methods:

> **Method 1:** Through the analysis of isotope ratios of certain stable isotopes in an extract of the seed and a corresponding soil sample ($^{12}C/^{13}C$, $^{1}H/^{2}H$, $^{14}N/^{16}N$) via Isotope Ratio Mass Spectrometry (IRMS), and the metal ion profile in an extract of the seed, the corresponding soil sample and the whole seed using Inductively Coupled Plasma Mass Spectrometry (ICPMS).
>
> **Method 2:** Through the chemical analysis of the seed metabolome using Nuclear Magnetic Resonance (NMR) spectroscopy and Liquid Chromatography Mass Spectrometry (LCMS), with further interrogation of the generated data via multivariate statistical analysis.

Milestone 3: To identify if and when the DNA signature is lost during the preparation of a ricin extract using methods available in terrorist handbooks and/or the Internet. Additionally, the most efficient DNA clean up method for the preparation of a sample obtained from a clandestine laboratory was determined.

This technical report aims to discuss the scientific progress made against the first two milestones. The progress against Milestone 3 has been described in two previously published technical reports, and will not be discussed in detail.[9,10]

2. Results

2.1 Field collections

During July and August 2009, collections of plants were made from distinct geographic locations around Australia. These concentrated on the West Coast, South Australia and Far North Queensland (Figure 1). Initially it was planned to collect specimens from Darwin. However, this was omitted due to no sightings of the plant on earlier visits.

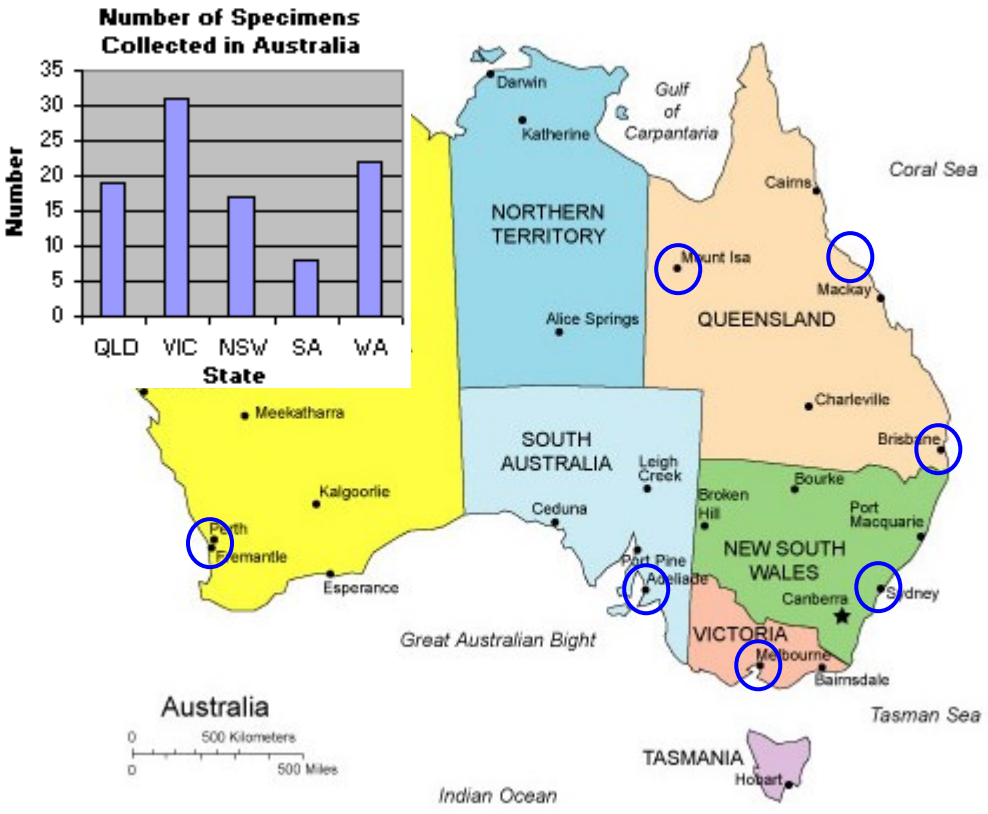


Figure 1 Map of Australia with blue circles indicating sites where specimens of R. communis and soil samples were collected. Inset: Graph showing the total number of collected specimens in the DSTO Australian mature seed library.

In total, 45 specimens were collected during these field trips, in addition to corresponding soil samples. After this collection effort, the DSTO Australian *R. communis* mature seed library contained 97 specimens (Figure 1 inset). This field work led to some interesting observations, in terms of cultivar population within the different states. The most diverse plant morphology was found in plants from New South Wales and Victoria. Queensland appeared to only have very limited diversity, with two specimen types observed in Brisbane. Genetic comparison of samples taken from Western Queensland (Cloncurry) and North Queensland (Killymoon Creek, near Townsville) indicated that two identical specimens were present in both locations, which were different to the specimens present in Brisbane. Also, there appeared to be no obvious stands of wild populations of *R. communis* in North Queensland north of the Herbert River at Ingham.

A subset of 25 specimens from these field collections were selected based on differences in location and morphology for further analysis (Appendix A). This selection formed the basis of ongoing studies of Australian specimens. The results from chemical analysis of these 25 specimens are discussed below in Section 2.2.2.3.

2.2 Cultivar and Provenance Determination

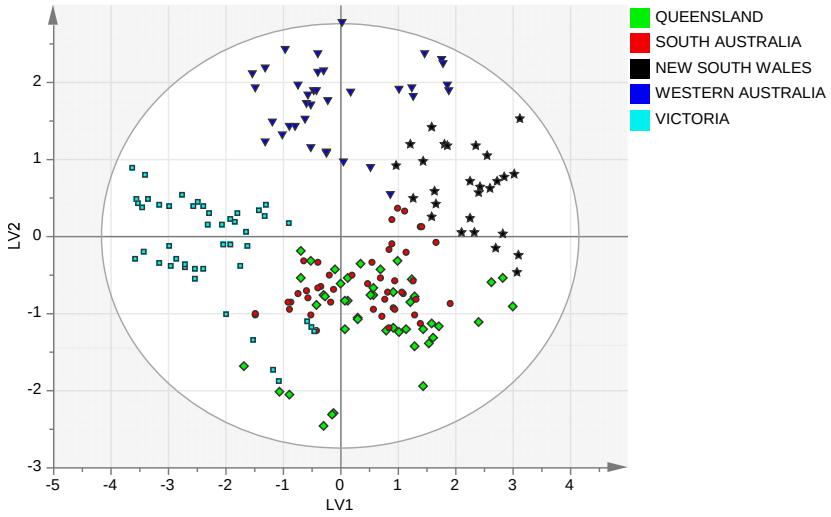
The importation of seeds of *R. communis* is restricted. Hence, the only available source is the progeny of garden specimens that grow around Australia. This limits investigators ability to trace an extract of *R. communis* to the geographic origin (provenance) due to the absence of a paper trail. Methods of analysis using routine analytical chemistry instrumentation for provenance determination would be useful to forensic and law enforcement agencies. To this end, investigations of *R. communis* extracts using mass spectrometry (ICPMS, IRMS LC-MS) and ^{1}H NMR were undertaken. The results obtained from these investigations are discussed in the following sections.

2.2.1 IRMS and ICPMS Analysis

The aim of using IRMS and ICPMS approaches was to determine if there was a stable isotope (^{1}H/^{2}H, ^{12}C/^{13}C, ^{14}N/^{16}N) and/or metal isotope composition link between a crude ricin extract and the location from which the seeds originated. For IRMS, data reproducibility was a significant restriction. Analysed independently of the ICPMS data, no significant trends were extracted from the IRMS data. Therefore only ICPMS data was analysed.

Analysis of Molecular Weight Cut Off (MWCO) and oil fractions, soil samples and whole seed via ICPMS was undertaken. There was significant intra-specimen variability in the data obtained for the MWCO and oil fractions from the two ICPMS techniques applied (LA-ICPMS and solution ICPMS). It was suspected that this was due to the residual acetic acid present in the solution, which was used during extraction. Furthermore, no correlations could be made between the composition of the seeds and the soil sampled from where the host plant resided. This was due to no soil being collected at multiple depths down to 3 m. Consequently, only data from the LA-ICPMS of the seed core could be used.

Following data pre-treatment, the LA-ICPMS data was subjected to OPLS-DA modelling. Samples were classified according to their state of origin ($R^2X = 0.83$, $Q^2X = 0.54$). It could be observed from the scores plot of LV1 vs. LV2 in Figure 2a that state specimens were clustering together. Other projections are shown in Appendix B. The loadings line plots are shown in Figure 2b.



(a)

(b)

Figure 2 OPLS-DA of the LA-ICPMS data. (a) Scores plot LV1vs. LV2; Vic (light blue squares), NSW (black stars), WA (dark blue triangles), Qld (green diamonds), SA (red circles); (b) Corresponding loadings line plot. Black: LV1; Blue: LV2; Red: LV3; Green: LV4.

The loadings line plot in Figure 2b allowed for each of the 15 isotopes to be interrogated for their ability to differentiate between the states. Each isotope was subjected to *t*-tests ($p \leq 0.007$) to confirm their validity. A summary of the results is shown in Table 1. Analysis of the data showed that ^{27}Al, ^{44}Ca, ^{55}Mn and ^{98}Mo did not contribute significantly to the observed clustering, highlighted by the yellow cells. Red cells identify isotopes that are decreased in

specimens from that state relative to other specimens. Cells in blue identify isotopes that are increased in specimens from that state relative to other specimens.

A representative line plot of the normalised LA-ICPMS data for ^{202}Hg is shown in Figure 3a. What can be seen from this are increased levels of ^{202}Hg in the Victorian and Western Australian specimens compared to the remaining states. Furthermore, compared to the specimens collected from all other states, New South Wales had decreased levels of ^{202}Hg.

Table 1 Isotopes identified as being significant for classification. Isotopes highlight (a) with red cells have decreased counts; (b) with blue cells have increased counts. Yellow cells made no contribution.

Isotope	Vic	NSW	SA	WA	Qld
^{24}Mg					
^{27}Al					
^{44}Ca					
^{53}Cr					
^{55}Mn					
^{57}Fe					
^{60}Ni					
^{65}Cu					
^{66}Zn					
^{75}As					
^{85}Rb					
^{88}Sr					
^{98}Mo					
^{138}Ba					
^{202}Hg					

Further analysis of the data led to two interesting observations. Firstly, the levels of ^{75}As were increased in South Australian specimens. Closer interrogation of the data for the South Australian specimens revealed that two specimens in particular (09-32 and 09-33) had significantly increased of ^{75}As. These specimens were collected from Blair Athol and Sefton Park respectively, neighbouring north Adelaide suburbs. The remaining three samples were collected from the Waterfall Gully in the Adelaide Hills (09-31), Reynella (09-27) in the southern suburbs of Adelaide, and Carrickalinga (09-30) on the coast 75 km south of Adelaide. Shown in Figure 3b is a line plot of the normalised LA-ICPMS data for ^{75}As. This plot clearly shows the increased levels of ^{75}As in the specimens from northern Adelaide compared to the other specimens.

The second observation was that levels of ^{85}Rb in the Queensland specimens. The specimens collected from both Cloncurry (09-66) and Killymoon Creek (09-70) were significantly increased in ^{85}Rb compared to any other specimens analysed. Shown in Figure 3c is the line plot of the normalised LA-ICPMS data for ^{85}Rb. The Cloncurry site is in western Queensland, while Killymoon Creek is near Townsville. Curiously, the Killymoon Creek specimen was collected approximately 40 km west of where the Townsville (09-72) specimen was collected, however it did not show increased levels of ^{85}Rb. Both the Cloncurry and Killymoon Creek

specimens were sampled on creek beds and it may this reason why the plants accumulated ^{85}Rb. Currently this is a tentative conclusion with further experimental work required.

While some interesting trends have been observed in the data, a further in-depth analysis is required and is currently being undertaken.

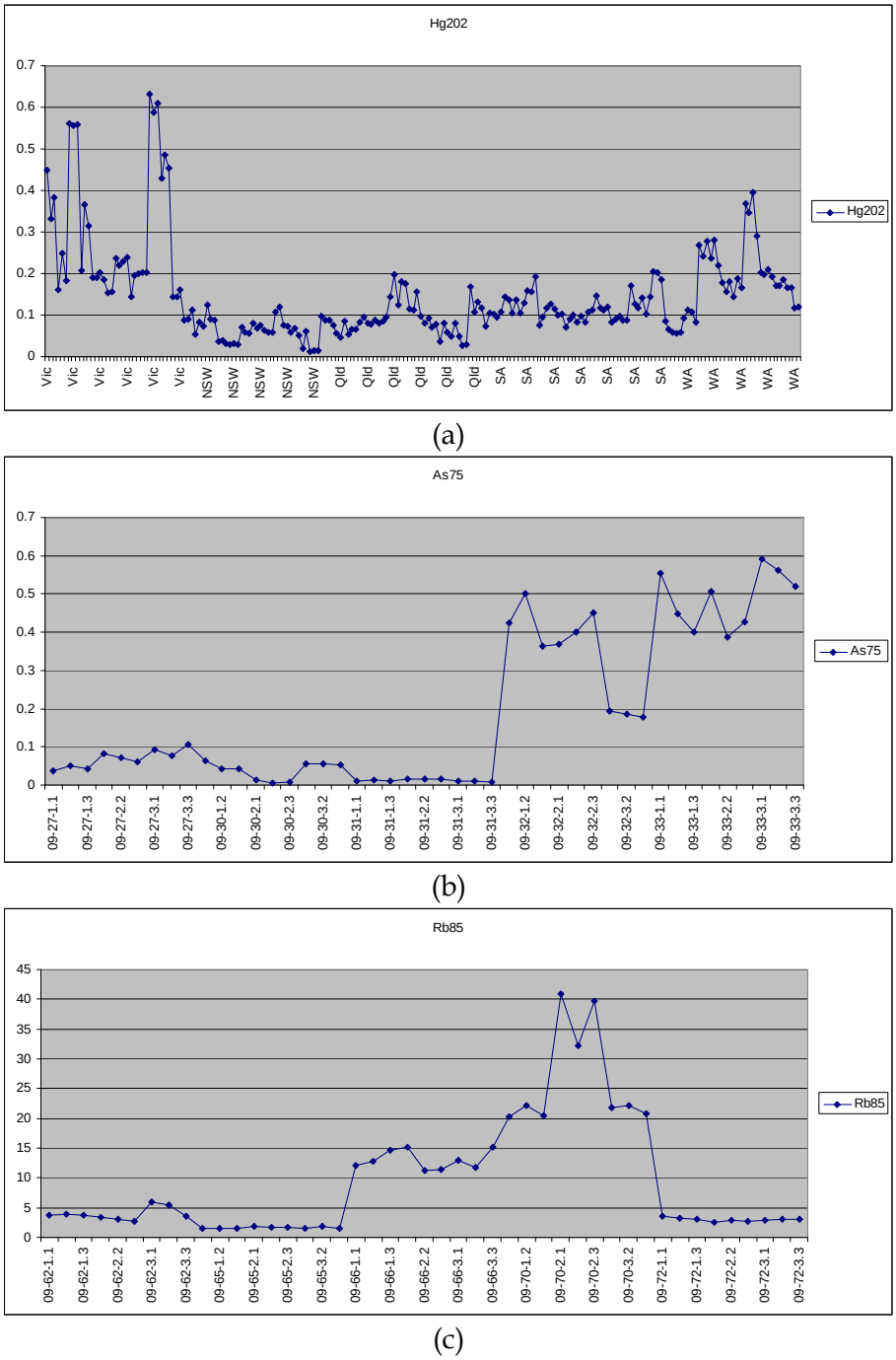


Figure 3 Line graphs for associated with isotopes from a particular state. (a) Line plot of the normalised LA-ICPMS data for ^{202}Hg (a) ^{75}As counts from SA specimens; (b) ^{85}Rb counts from Qld specimens.

2.2.2 NMR Based Metabolomics

Metabolomics is the study of the population of small molecules (metabolites) present at a particular time point within a biological system (plant, microbial or mammalian) and is referred to as the metabolome.[11,12] Through the study of the metabolome insights can be gained into the environment that the host biological system has been exposed too. Through the application of metabolomics to *R. communis* seeds, it was hypothesised that the environment in which the host plants were exposed to would be reflected in the metabolome. For this study, the environment is classified from a geographical stand point, as opposed to seasonal fluctuations. Given the disparate geography of Australia's state based capital cities, it is expected that the study of the metabolome would allow for provenance determination of the host plant to be made.

This study was divided into three sections. The first study was to analyse extracts of known cultivar and provenance from seed specimens supplied by Dstl. The second study analysed a larger population of seeds representing different cultivars collected from different countries and sourced from a seed supplier (Sandemann Seeds) in France. The third study concentrated on the analysis of seeds that were collected from various locations around Australia. Building models for provenance classifications with extracts of known cultivar allowed for genetic variations to be evaluated. If successful, this strategy could be applied to *R. communis* extracts for provenance determination of unknown cultivars.

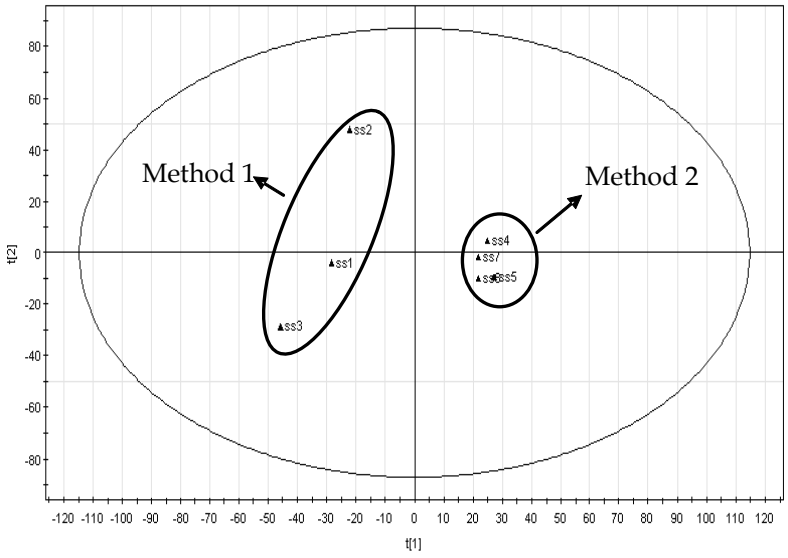
2.2.2.1 Study 1: Dstl Overseas Specimens

For this initial study, eight specimens of six cultivars ("*carmencita*" Tanzania, "*dehradun*" India, "*gibsonii*" Zimbabwe, "*impala*" Tanzania, "*sanguineus*" Spain and Tanzania, and "*zanzibariensis*" Kenya and Tanzania) were investigated. Following *R. communis* seed extraction and ^{1}H NMR analysis, the collected ^{1}H NMR was subjected to multivariate statistical analysis.

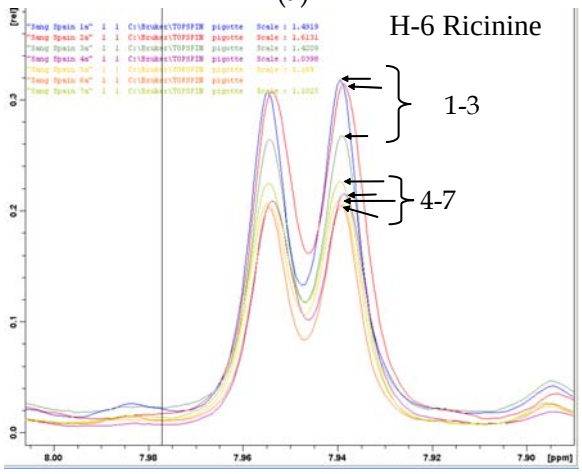
Initial OPLS-DA models indicated that whilst cultivar determination was possible, provenance determination of the *"zanzibariensis"* and *"sanguineus"* specimens was not. A principal components analysis (PCA) was conducted on the *"sanguineus"* Spain extracts. The PCA scores plot (Figure 4a) identified a difference between extraction method 1 (replicates 1-3) and extraction method 2 (replicates 4-7). On further analysis of all extracts from all cultivars, identical results were observed. On re-investigation of the ^{1}H NMR spectra for *"sanguineus"* Spain it was evident that the intensities of all resonances in the spectra differed between extraction method 1 and 2. This is clearly observed in the intensities of the H-6 ^{1}H NMR resonance for ricinine at δ 7.95 (Figure 4b). The three spectra with the highest intensities corresponded to extraction method 1. Conversely, the four spectra with the lowest correspond to extraction method 2.

After establishing that consistent separation was occurring based on extraction method across all of the collected spectra, PQN[13] was applied to remove the influence of extraction method. PQN calculates the most probable dilution factor from the distribution of quotients between the disparate spectra and the reference spectrum and then applies this to all affected spectra.[13] Separate OPLS-DA analysis conducted on spectra from replicates 4-7 resulted in a model that yielded good class separation between cultivar and provenance (data not shown). Hence,

these replicates were used as the standard set of spectra or reference spectra. A PQN adjusted data matrix was constructed, consisting of a combination of the original spectra from replicates 4-7 and the new PQN data set for replicates 1-3 of all cultivars.



(a)



(b)

Figure 4 (a) PCA scores plot of "sanguineus" Spain, highlighting the separation between extraction method 1 (replicates 1-3) and 2 (replicates 4-7); (b) Stacked ^{1}H NMR spectra of the H-6 resonance of ricinine (δ 7.95) of all "sanguineus" Spain replicates showing the varying intensities.

A seven-component OPLS-DA model of this adjusted data matrix identified class separation according to both cultivar and provenance (R^2X= 0.932, R^2Y= 0.886, Q^2Y= 0.758) with 50% of the variation (R^2X) explained by the first three latent variables. The scores plot (LV1 vs. LV2) in Figure 5 not only shows that each specimen occupies their own distinct regions, but also highlights the *"dehradun"* India specimen as markedly different from all other specimens based on LV1. This model also indicates that the "*zanzibariensis*" and "*sanguineus*" specimens cluster together according to their cultivar (negative loadings on LV2), yet still show separation based on provenance.

Examination of the loadings plot on LV1 (Figure 6a), revealed a strong positive contribution at δ 5.40, attributed to the anomeric ^{1}H NMR resonance of sucrose (Scheme 1). The strong contribution of these bins contributed to the distinct separation of *"dehradun"* India observed in the OPLS-DA model (Figure 5). Furthermore, the separation of *"impala"* Tanzania and *"zanzibariensis"* Kenya from the other specimens was also influenced by the relative amounts of sucrose. The average spectrum of each specimen was plotted to examine the relative amounts of sucrose present (Figure 6b). The *"dehradun"* was found to have significantly less sucrose that all other specimens (p<0.0001), while *"impala"* and *"zanzibariensis"* Kenya contained the highest relative amounts of sucrose (p<0.02). This observation supported the finding that the relative amounts of sucrose were responsible for explaining some of the observed class separation.

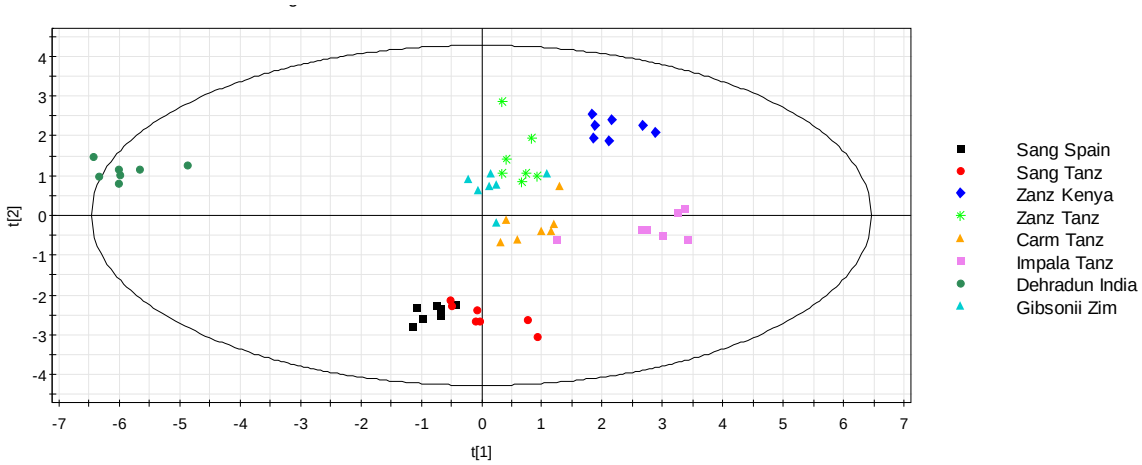
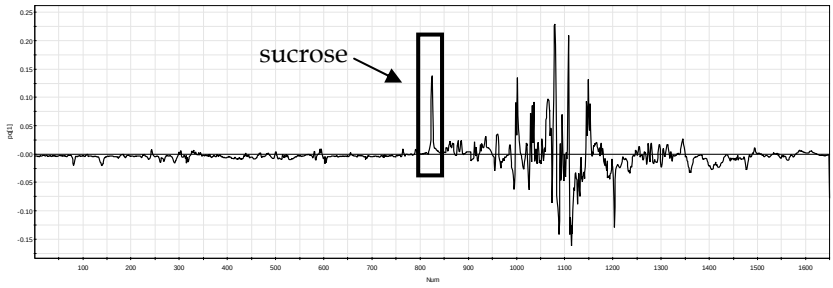


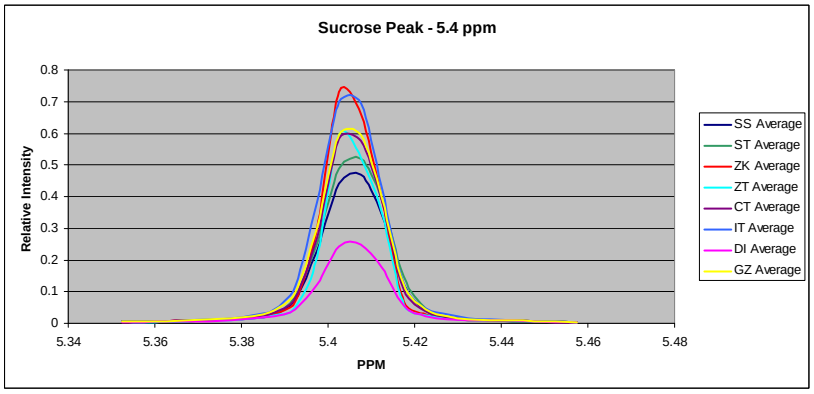
Figure 5 OPLS-DA model scores for LV1 and LV2 for all specimens assigned as their own cultivar/provenance.

The OPLS-DA scores plot (LV1 vs. LV3, Figure 6c) identified that LV3 was responsible for further specimen classification. The loadings plot of LV3 (Figure 6d) again identified bins 822-826, corresponding to the anomeric ^{1}H NMR resonance for sucrose, as responsible for positive loadings on LV3. Additionally, bins corresponding to the ^{1}H NMR resonances of H-5 (δ 6.5) and H-6 (δ 7.9) of ricinine,[14] *N*-demethyl[14] and *O*-demethyl ricinine[14] (identified by the boxes in Figure 6d, structures in Scheme 1) were equally responsible for negative loadings on LV3 (p<0.0001). The presence of sucrose, ricinine,[14] *N*-demethyl[14] and *O*-demethyl ricinine[14] was confirmed through isolation, 2D NMR and LC-MS.

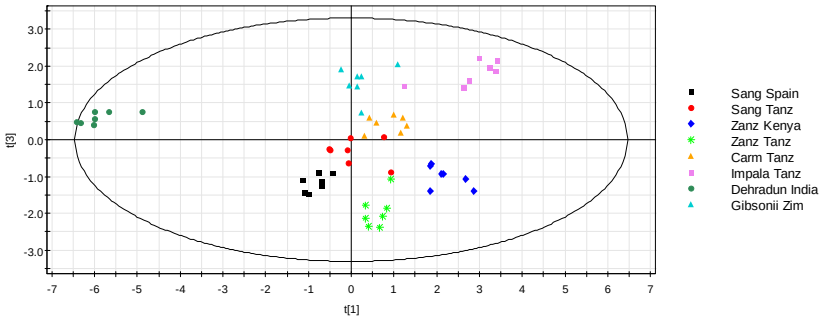
Further investigations were undertaken to establish an OPLS-DA model capable of classifying specimens according to provenance. This model explained 85% of the variation in the data (R^2X), with strong provenance separation (R^2Y= 0.884) and predictability (Q^2Y = 0.814). Of particular interest was that the two *"zanzibariensis"* specimens (both originating from Africa) clustered together (Figure (a), Appendix C), whereas the *"sanguineus"* specimens did not (originating from different continents). Consistent with previous observations, the *"dehradun"* specimen from India was found to again cluster in its own unique space, with negative loadings on LV1.



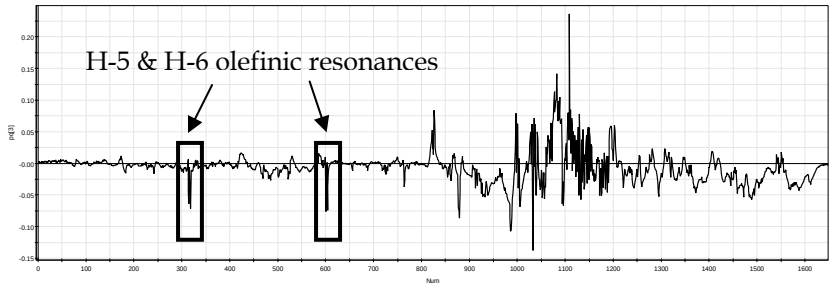
(a)



(b)

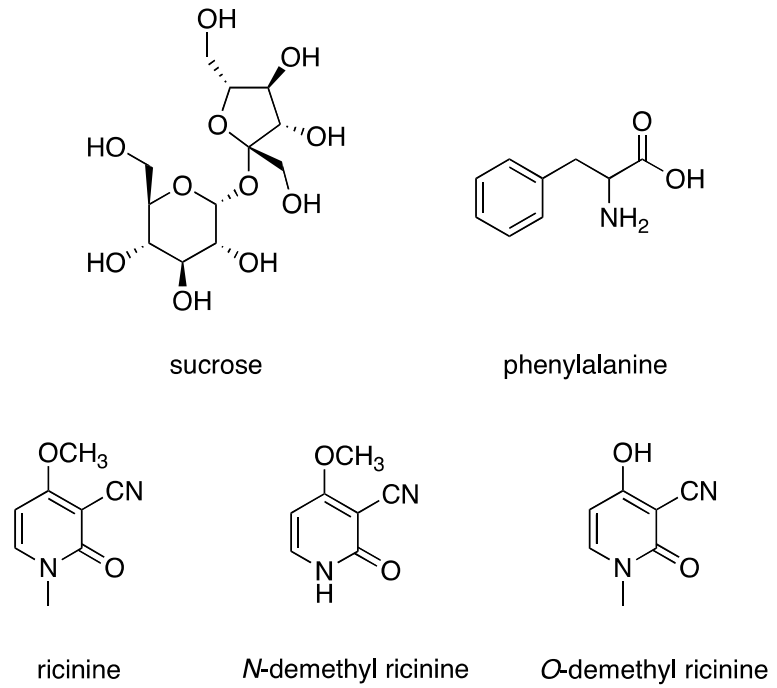


(c)



(d)

Figure 6 (a) Loadings plot of LV1. Box corresponds to the sucrose anomeric ^{1}H NMR resonance δ 5.40; (b) Comparison of the intensity of the anomeric ^{1}H NMR resonance of sucrose at δ 5.40 in the averaged spectrum across all specimens; (c) OPLS-DA model scores for LV1 and LV3 for all specimens assigned as their own cultivar/provenance; (d) Loadings plot of LV3. Boxes identify olefinic H-5 and H-6 resonances of ricinine, N-demethyl and O-demethyl ricinine as contributing to negative loadings.



Scheme 1 Structures of important compounds identified from OPLS-DA analysis.

To determine if further provenance separation could be achieved, the model was regenerated with only African specimens. Again, a strong two-component model (R^2X = 0.846) with excellent provenance separation (R^2Y = 0.913) and good predictability (Q^2Y = 0.742) was generated. Of particular note were the two "*zanzibariensis*" specimens (Tanzania and Kenya). Previously these clustered together according to their continent of origin (Figure (a), Appendix C). However, they were now separated according to their country of origin (Figure (d), Appendix C), despite the fact Tanzania and Kenya share a common border. Analysis of the loadings plots for these models (Figures (b), (c), (e), (f), Appendix C) again indicated that both sucrose and ricinine were contributing to the class separation.

Given the success of predicting provenance, an OPLS-DA model was generated to examine the possibility of cultivar determination amongst the individual African specimens. This model (Figures 8a and b) identified cultivar separation between all specimens (R^2X= 0.901, R^2Y= 0.893), with good predictability (Q^2Y = 0.753). The bins associated with the anomeric ^1H NMR resonance for sucrose, in addition to the olefinic ^1H NMR resonances for ricinine and analogues, again influenced the separation of specimens on LV1 and LV2 (Figure (g) and (h), Appendix C). The loadings plot of LV3 (Figure 7c) also showed that there were other unidentified compounds contributing to the model. In particular, some of the loadings associated with bins in the aromatic region of the data were contributing to negative loadings on LV3. Subsequent fractionation of the "*zanzibariensis*" Tanzania extract followed by 2D NMR and LC-MS identified phenylalanine (Scheme 1) that readily explained this observation. Also evident in the loadings plot for LV3 were bins most likely due to the anomeric protons of unresolved sugars. The compound responsible for these loadings requires further investigation to allow a positive identification.

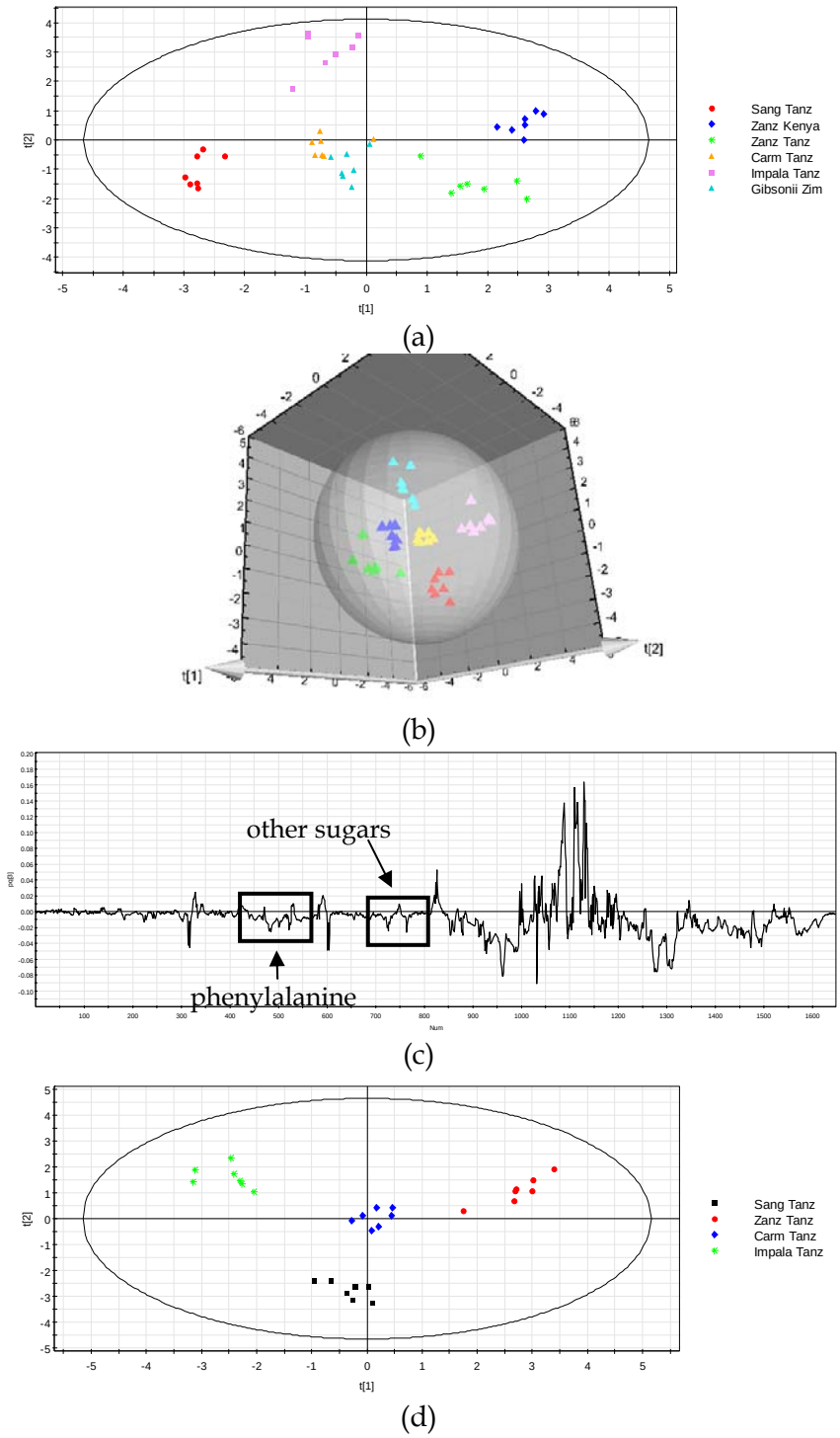


Figure 7 (a) OPLS-DA scores plot showing good separation of all the African specimens according to cultivar; (b) OPLS-DA scores plot using the same model as in (a), however looking at the first three LV to give a 3D plot; (c) Loadings plot of LV3 showing sucrose, ricinine, phenylalanine and other sugars (still to be identified) are contributing to the separation of the African specimens according to cultivar; (d) OPLS-DA scores plot showing good separation according to cultivars originating from Tanzania.

Furthermore, a similar cultivar model was generated from the four specimens originating from Tanzania. Strong separation was achieved between specimens ($R^2X = 0.849$, $R^2Y = 0.930$, $Q^2Y = 0.810$) as can be seen in Figure 7d. Again, this separation was again attributed to sucrose, ricinine, *N*-demethyl and *O*-demethyl ricinine.

To further explore the predictive strength of the OPLS-DA model, blind/validation extracts were introduced into the model described in Figure 5, with predicted values shown in Table 2. The three blinded *"gibsonii"* samples were correctly predicted, as were two of the three *"dehradun"* samples. The third *"dehradun"* sample (BS7) was predicted to be *'dehradun'*, *"carmencita"* or *"zanzibariensis"* Kenya, as no strong class classification was possible.

Tables 2 Prediction table of semi-blinded/validation samples according to all of the specimens. Strong prediction > 0.8 (green); 0.3 < weak prediction < 0.8 (orange); No prediction ≤ 0.3 (clear).

Obs ID	SS	ST	ZK	ZT	CT	IT	DI	GZ
BS1 (DI)	0.38	-0.16	0.09	0.02	-0.04	0.03	0.86	-0.17
BS2 (GZ)	-0.04	0.30	0.04	0.06	-0.14	-0.01	-0.08	0.88
BS4 (DI)	0.27	-0.07	-0.17	0.13	-0.12	0.04	0.97	-0.04
BS5 (GZ)	-0.35	-0.07	-0.20	0.27	0.18	0.21	0.03	0.91
BS7 (DI)	-0.40	0.01	0.47	-0.53	0.50	-0.10	1.38	-0.35
BS8 (GZ)	0.18	-0.13	-0.19	0.16	-0.02	0.03	0.03	0.94

SS: "sanguineus" Spain; ST: "sanguineus" Tanzania; ZK: "zanzibariensis" Kenya; ZT: "zanzibariensis" Tanzania; CT: "carmencita" Tanzania; IT: "impala" Tanzania; DI: "dehradun" India; GZ: "gibsonii" Zimbabwe

When the blinded samples were investigated for continent of origin (model in Figure 7a), every blinded sample was correctly predicted (Table 1a, Appendix C). Additionally, when the *"gibsonii"* Zimbabwe sample was predicted to be an African specimen (model in Figure 7b), all three blinded samples were correctly predicted (Table 1b, Appendix C). These three prediction tables indicate that the developed statistical models can be used as a tool to correctly identify blinded *R. communis* extracts according to cultivar or provenance or both. Additionally, all blinded samples could be correctly identified, despite three different extraction techniques being used. The results were further corroborated through the raw data matrix being analysed by and independent researcher, who generated a PLS-DA model (PLStoolbox), and correctly predicted the blinded samples.

These results demonstrate that for this initial study, cultivar and provenance were able to be determined for the eight specimens analysed. Utilising the loadings plots and 2D NMR, compounds were identified that contribute the observed class classifications in these models. While excellent results, to further strengthen the hypothesis, an expanded collection of overseas seeds was investigated.

These results have formed the basis of a manuscript recently published in the journal *Metabolomics*.[15]

2.2.2.2 *Study 2: Sandemann Seed Specimens*

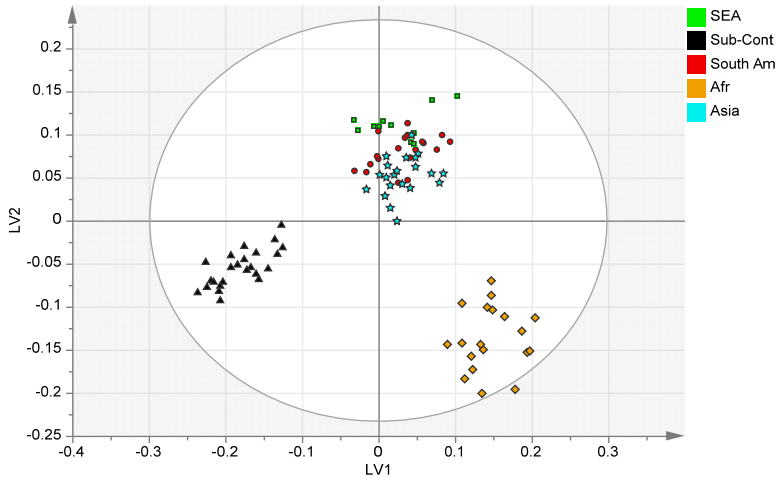
For the expanded study, a total of 18 specimens from 11 countries were analysed. These are tabulated Appendix A. Following data collection, pre-treatment and data reduction, specimens were class classified according to their continent of origin, and subjected to OPLS-DA ($R^2X = 0.89$, $Q^2X = 0.77$). The corresponding scores plots are shown in Figure 8.

As can be seen from these scores plots, depending on what LV combinations were compared, continent based clustering could be observed. In particular, Sub-Continent (black triangles) and African samples (yellow squares) in Figure 8a, South East Asian samples (green squares) in Figure 8b, South American (red circles), South East Asian samples (green squares) and Asian specimens (blue stars) in Figure 8c. The corresponding loadings plot for LV1 is shown in Figure 9a. The loadings plot indentified what resonances in the NMR spectra – and hence what compounds – were contributing to the observed class based clustering. In Figure 8a, African and Sub-Continent specimens were well separated. From the loadings plot in Figure 9a, ricinine (red box) and sucrose (green box) were identified as significant variables. Previous findings[15] identified that relative amounts of ricinine and sucrose were important discriminators for provenance. Furthermore, Figure 9a identified that resonances between δ 3.90 and δ 4.30 (blue box), and between δ 3.46 and δ 3.80 (black box) were important. Shown in Figure 9b are stack plots of the raw ^{1}H NMR data from two African (purple – "*zanzibariensis*" Kenya and green – "*impala*" Tanzania) and two Sub-Continent (blue – "*noori dehradun*" India and red – "*black diamond*" India) specimens for each of these regions.

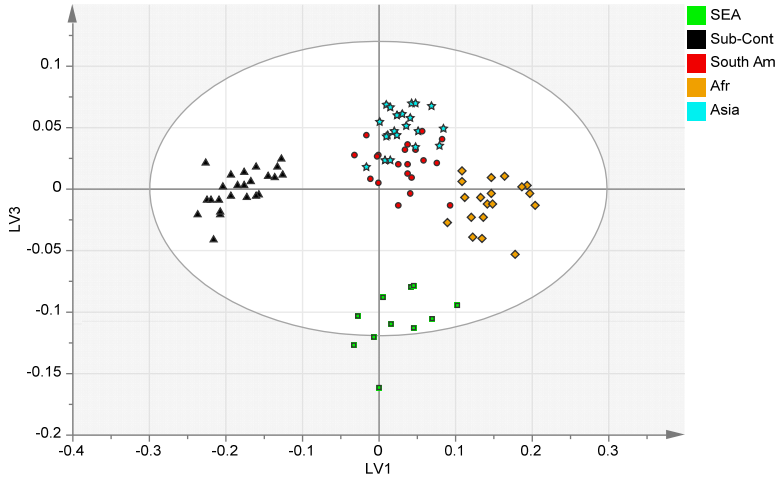
These ^{1}H NMR spectra stack plot show that more of the compounds responsible for the resonances between δ 3.90 and δ 4.30 are present in the African specimens compared to the Sub-Continent specimens (Figure 9b, top spectra). While for the region between δ 3.46 and δ 3.80, more of the compounds responsible for these resonances are present in the Sub-Continent specimens (Figure 9b, bottom spectra). Using this strategy, the remaining loadings plots (Figures 11a to c) were investigated.

Resonances identified by the boxes in Figure 10 were found to be significant. Of particular interest is the series of anomeric resonances identified by the red box (δ 5.05 to δ 5.30) in Figure 10a. There appears to be several different sugar species present in these extracts in differing amounts. These are important for the observed class clustering of South East Asian, South American and Asian specimens in Figure 8b and c.

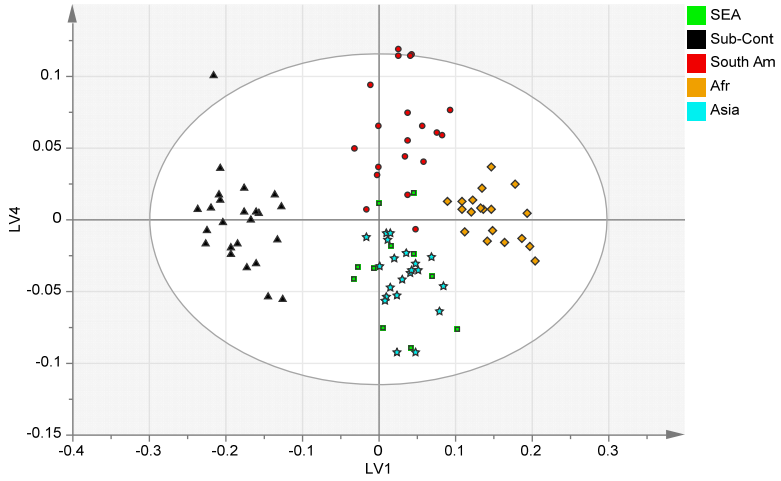
Currently the identity of the compounds associated with the coloured boxes in Figures 10a and 11 are being established. Once purified, their respective structures will be elucidated. Having identified structures in hand will allow for analytical method development to take place for a robust methodology for provenance determination.



(a)



(b)



(c)

Figure 8 OPLS-DA models of specimens of known cultivars. Specimens were classed according to continent of origin. (a) LV1 vs. LV2; (b) LV1 vs. LV3; (c) LV1 vs. LV4.

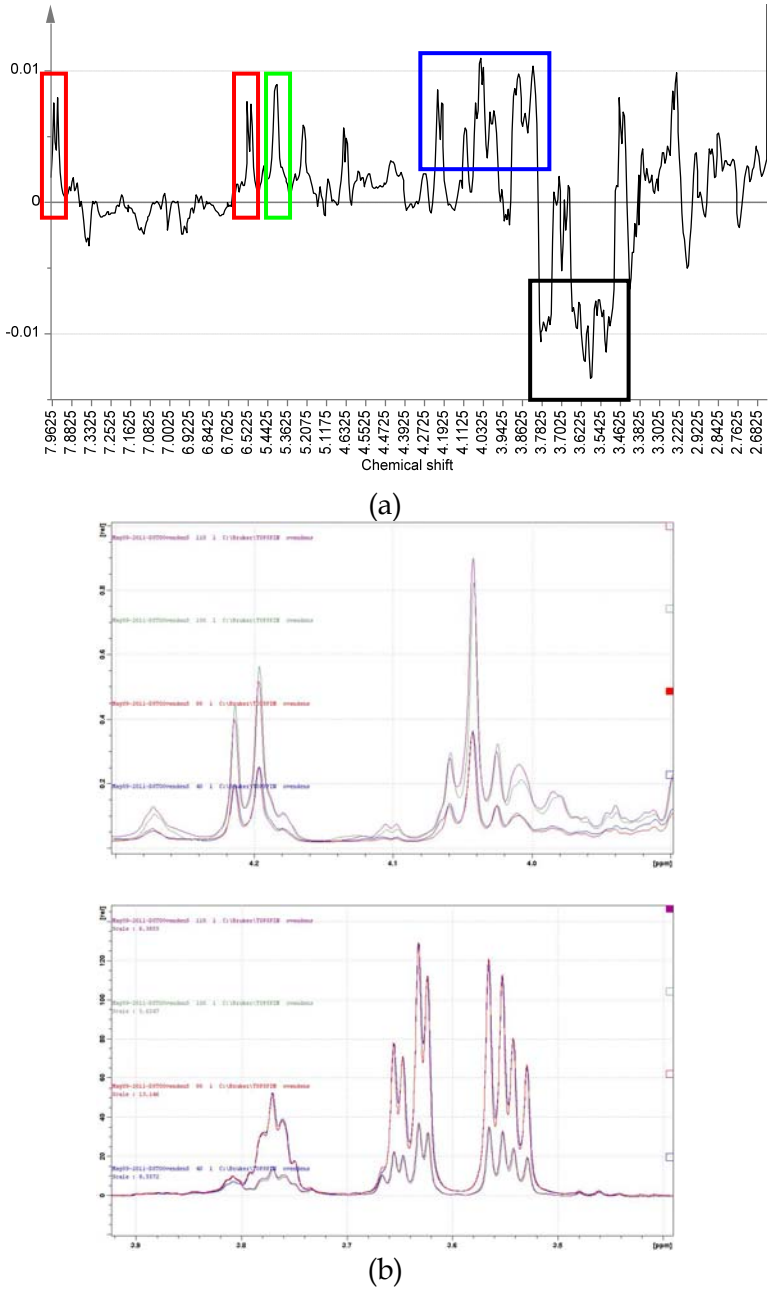
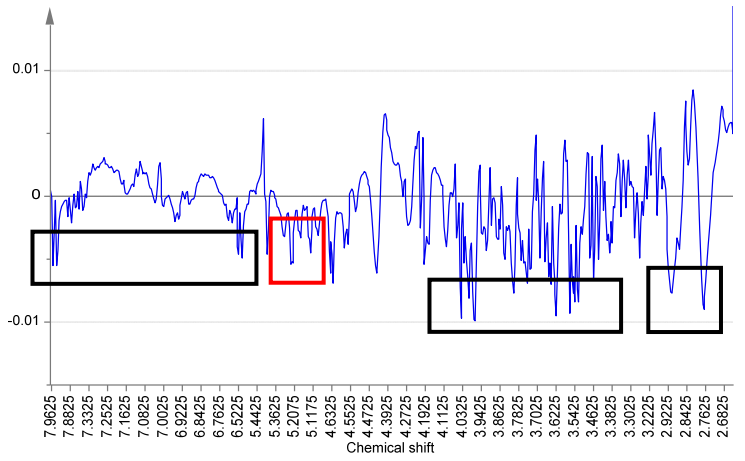


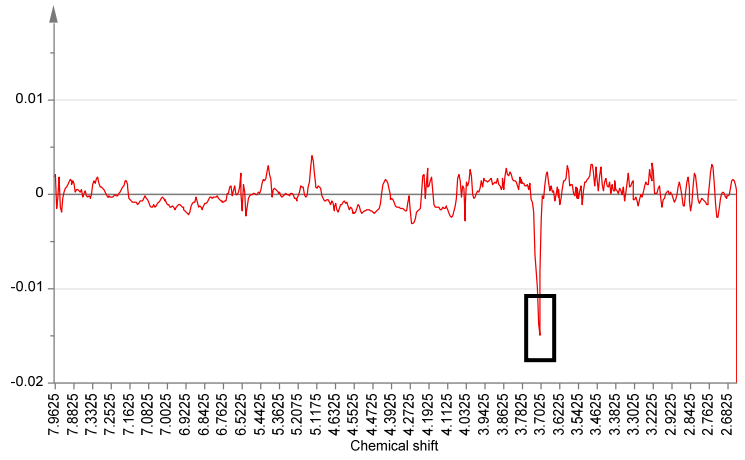
Figure 9 (a) Loadings line plot of LV1; (b) ^{1}H NMR stack plots of expanded regions between δ3.90 and δ4.30 (top) and between δ3.46 and δ3.80 (bottom).

2.2.2.3 *Study 3: Australian Specimens*

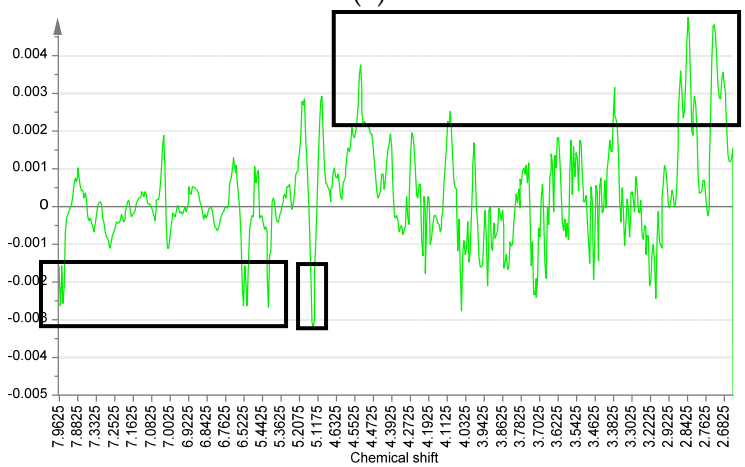
The previously discussed research on specimens of known cultivar and provenance was important in establishing proof of concept of the viability of the metabolomics approach. Further application of this methodology to Australian specimens was important to demonstrate its usefulness in an Australian context.



(a)



(b)



(c)

Figure 10 Associated loadings lines plots for the model scores plot in Figure 4. (a) LV2; (b) LV3; (c) LV4.

The 25 Australian specimens listed in Appendix A were extracted, subjected to ^{1}H NMR analysis, with data pre-treated as previously described. The collected data was classed according to state of origin, and subjected to OPLS-DA ($R^2X = 0.92$; $Q^2X = 0.68$). The scores plot for this model is shown in Figure 11a. Initial analysis readily identifies that the New South Wales specimens have clustered away from the Queensland specimens. The loadings plot for LV1 is shown in Figure 11b. For the New South Wales specimens the negative resonances were found to be important contributors to the observed class based clustering shown in Figure 11a. Figure 11c shows a stack plot of ^{1}H NMR spectra of a representative from each state for the region of the ^{1}H NMR spectra highlighted by the red box in Figure 11b. Immediately apparent is the New South Wales specimen (09-51 – blue) has more of the compounds represented by resonances at δ 9.12 and δ 8.83 as compared to the specimens from other states. The other region identified from Figure 13b was the area highlighted by the blue box. This area is complicated with many overlapping resonances, making it difficult to identify compounds responsible for the observed clustering. Isolation of these compounds will need to be undertaken to further understand the chemical composition.

The loadings line plot for LV2 is shown in Figure 12a. This identified a series of anomeric resonances (identified by the red box) at δ 5.10, δ 5.14, δ 5.19, δ 5.22, in addition to the sucrose anomeric resonance at δ 5.41, were important for the clustering of Victorian away from South Australian specimens in Figure 11a. Subsequent *t*-tests ($p \leq 0.004$) identified that all aside from the resonance at δ 5.22 were significant. Figure 12b shows a stack plot of all normalised ^{1}H NMR data for Victorian (red) and South Australian (black) specimens. Some general trends were able to be observed in this plot. In particular, the compound responsible for the anomeric resonances δ 5.19 and δ 5.10 were increased in the Victorian specimens. Furthermore, there appeared to be a general trend of increased amounts of sucrose, and another sugar with an anomeric resonance at δ 5.14, in the Victorian specimens.

Other scores plot projections are shown in Appendix D, along with the corresponding loadings line plots. These plots allowed for the identification of further resonances that contributed to the observed clustering. In particular, resonances at δ 7.32, consistent with the aromatic resonances of phenylalanine, contributed the clustering of South Australian specimens in Figure a, Appendix D. The scores plot in Figure c, Appendix D identified the ricinine[14] resonances, in addition to *O*- and *N*-demethyl ricinine analogues[14] making a significant contribution to the clustering of Western Australian specimens away from the other specimens. A stack plot of the ^{1}H NMR resonances from a representative of each state is shown in Figure 13. What can be seen from this is that compared to the other specimens, there is decreased amounts of ricinine from the Western Australian specimen compared to the other states. Furthermore, it appears that that *O*- and *N*- demethyl ricinine analogues[14] are increased in the Western Australian specimens. Further analysis and quantification studies are required to confirm this.

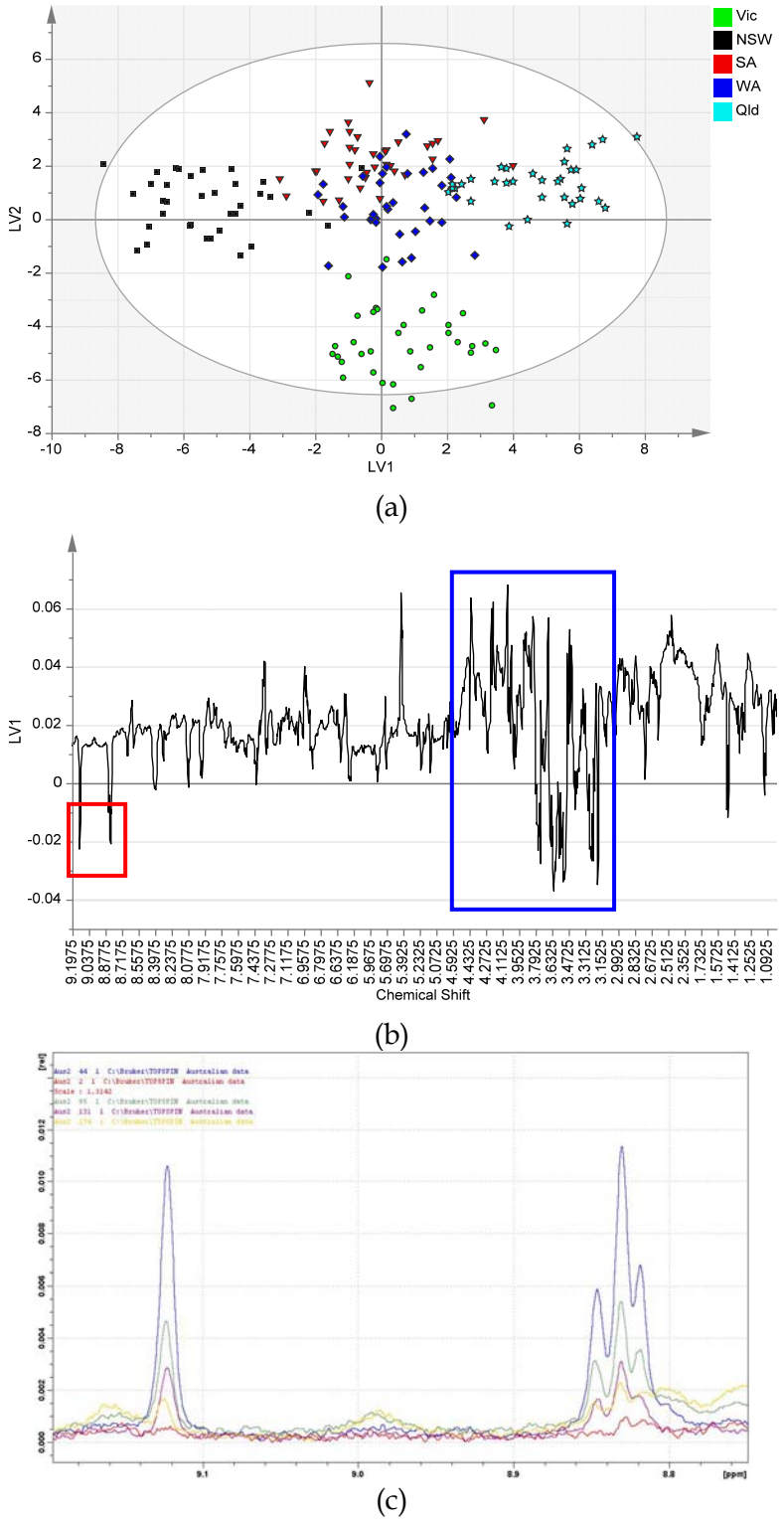


Figure 11 OPLS-DA analysis of Australia specimens. (a) Scores plot, LV1 vs. LV2 of Vic: green, NSW: black, SA: red, WA: dark blue and Qld: light blue; (b) Loadings line plot of LV1; (c) ^{1}H NMR stack plot of spectra from specimens from different states.

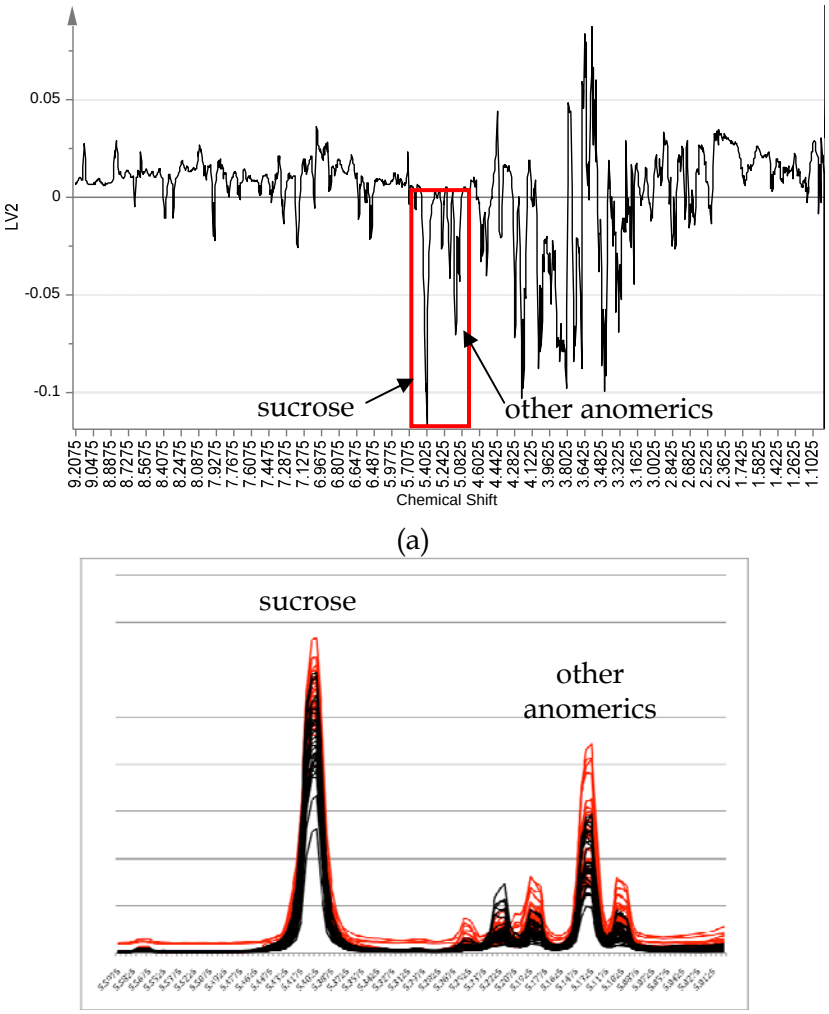


Figure 12 (a) Loadings line plot of LV2. Red box highlights the anomeric resonances; (b) stack plot of all normalised ^{1}H NMR data for Victorian (red) and South Australian (black) specimens.

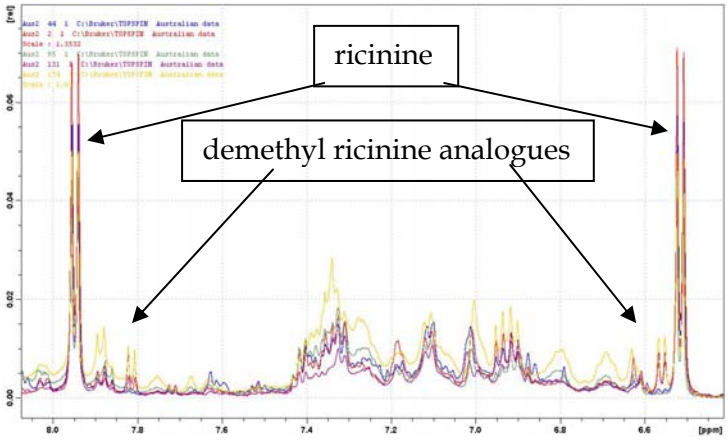


Figure 13 ^{1}H NMR stack plot (δ 8.01 – δ 6.40) of spectra from specimens from different states

2.2.2.3.1 Intra-state Comparisons

2.2.2.3.1.1 Queensland Specimens

While broad state based provenance classification was useful, for large states such Queensland and Western Australia, this would be of limited use. To this end, the Queensland data was investigated to further understand the ability to classify samples to a geographical region. In particular, the specimens collected from Cloncurry (09-66) and Killymoon Creek (09-70) were compared. These specimens were collected within two days of each other from different locations some 800 km apart, with Cloncurry situated in the arid North West of Queensland, and Killymoon Creek situated on the near Townsville. Morphologically, these two specimens looked identical, while PCR analysis[10] confirmed that genetically they were very closely related, if not identical.

The PCA ($R^2X = 0.95$, $Q^2X = 0.85$) scores plot of PC1 vs. PC2 (Figure 14a) indicated that there was a difference between these two specimens. The loadings plot shown in Figure 14b indicated that one of the main compounds responsible for the observed separation was ricinine (highlighted by the red boxes. A stack plot of the normalised ^{1}H NMR data for Cloncurry (09-66) and Killymoon Creek (09-70) is shown in Figure 14c. What can be seen in this plot is a general trend of more ricinine being present in the Killymoon Creek specimens as compared to the Cloncurry specimens. This finding is consistent with previous results[15] that have identified amounts of ricinine being sensitive to the local environment of the plant. Considering the genetic similarity of the specimens, these results would appear to be further evidence that the identification of differing chemistries due to the differing climates the host plants were exposed to.

2.2.2.3.1.2 Footscray Specimens

While the Queensland specimens were collected across the state from disparate geographical regions, all the Victorian specimens were collected within a 15 km radius of the CBD. However, the plants sampled were morphologically quite different from each other. In particular, the two Footscray specimens were morphologically very different and were growing approximately 20 m from of each other across a rail bridge. Specimens 09-05 had smooth seed pods that were are grey/green colour. Specimens 09-06 produced a bright red spiky seed pod. Considering this, as well as both plants being grown in the same soil type and exposed to identical micro-climates, they were excellent specimens to compare and to interrogate their respective metabolomes for differences.

Subsequently, the PCA ($R^2X = 0.84$; $Q^2X = 0.54$) scores plot of PC1 vs. PC2 is shown in Figure 15a, with the corresponding loadings line plot of PC1 shown in Figure 15b. For these two Footscray specimens, the sucrose anomeric proton resonance at δ 5.41 is a strong contributor to the separation. However, the anomeric proton resonance at δ 5.14 associated with an unknown sugar is the strongest contributor. A stack plot of the normalised ^{1}H NMR data for Footscray "red" (09-06) and Footscray "smooth" (09-05) is shown in Figure 15c. It can be seen here that there generally appears to be a greater amount of anomeric proton resonance at δ 5.14 present in the Footscray "red" (09-06) specimens compared to the Footscray "smooth" (09-05) specimens.

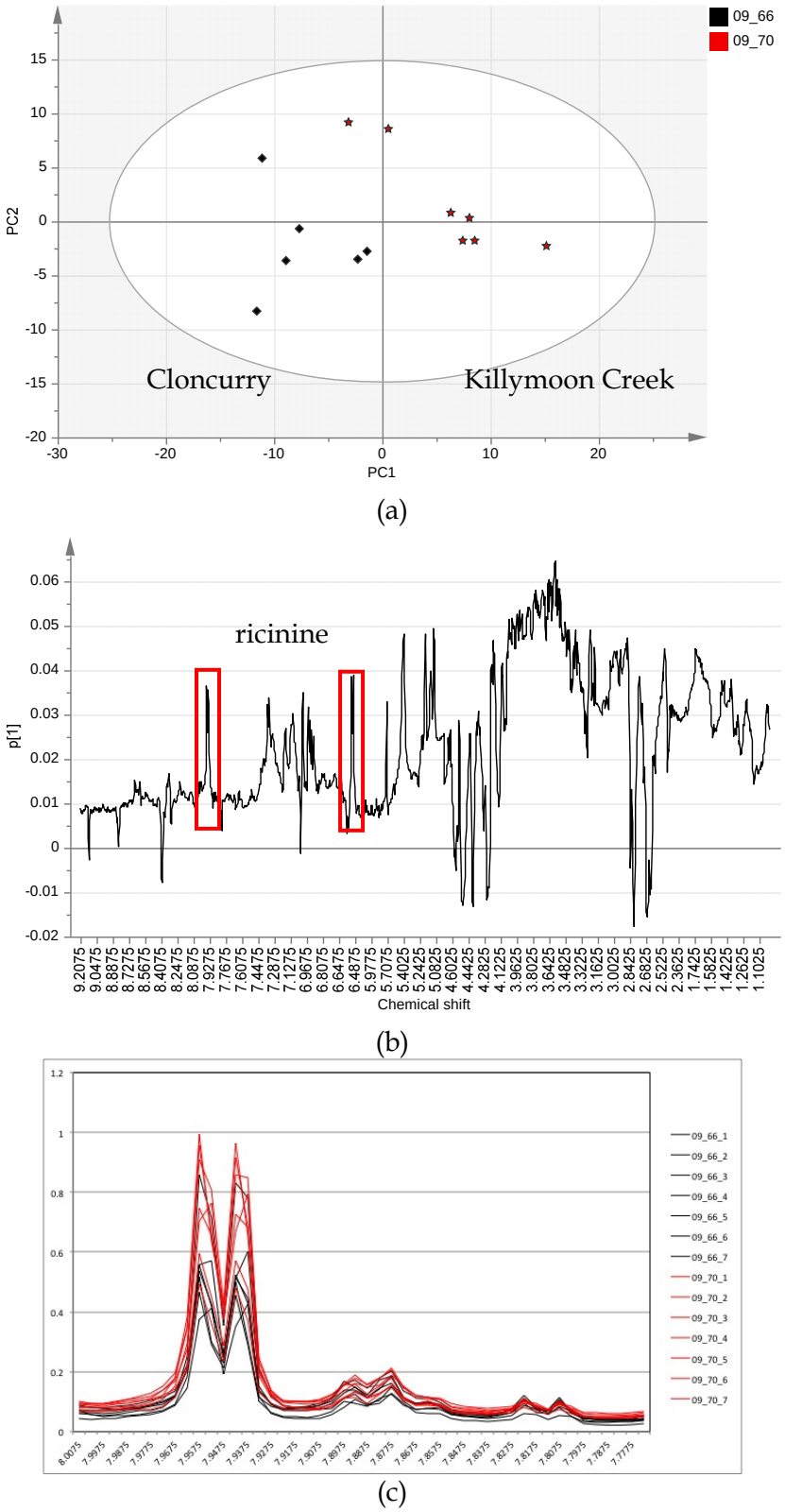


Figure 14 (a) PCA score plot (PC1 Vs. PC2) of Killymoon Creek and Cloncurry specimens; (b) Loadings line plot of PC1; (c) Stack plot of normalised ^{1}NMR data from Killymoon Creek and Cloncurry specimens.

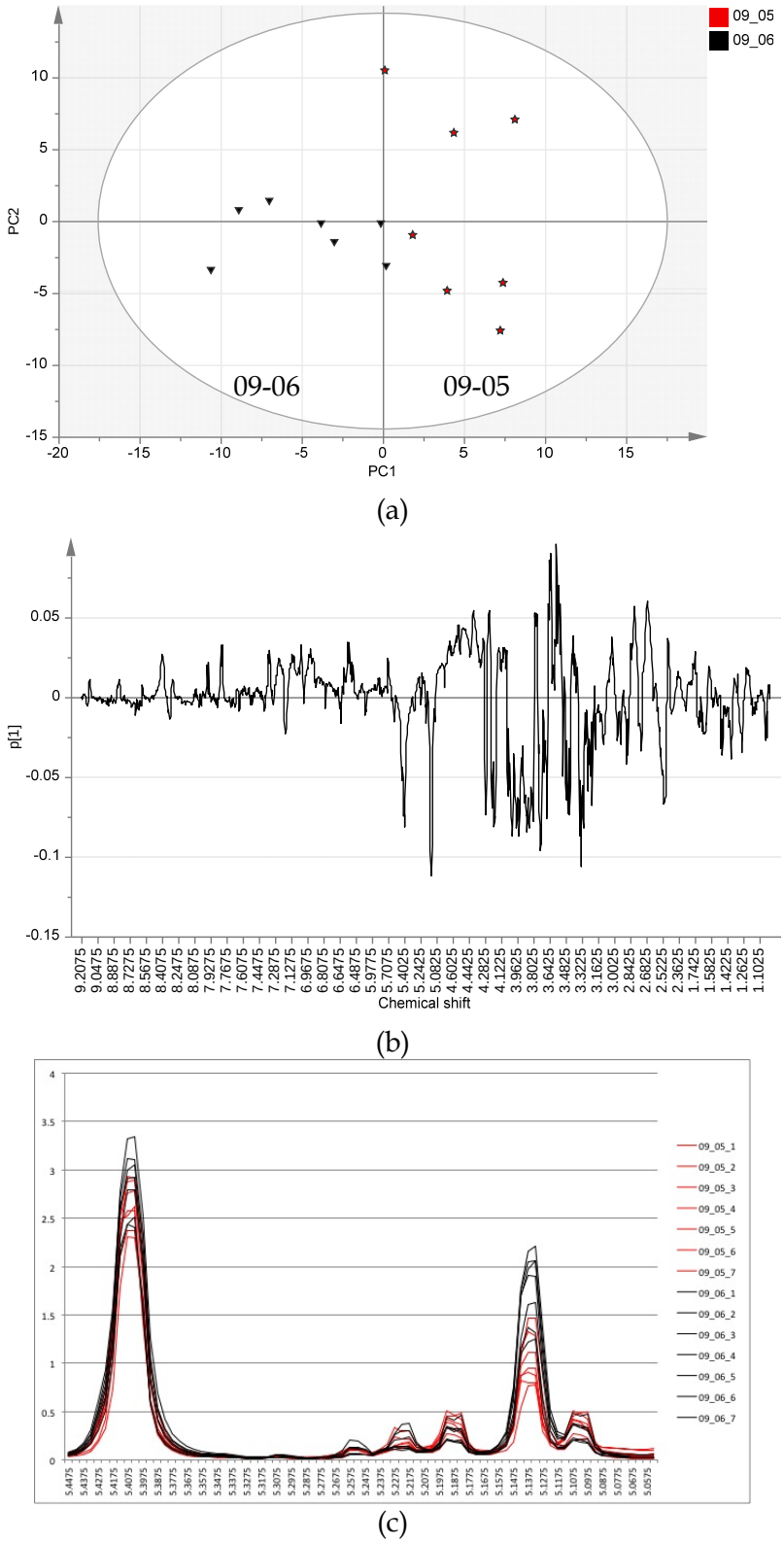


Figure 15 (a) PCA score plot (PC1 Vs. PC2) of Footscray "red" (09-06) and Footscray "smooth" (09-05); (b) Loadings line plot of PC1; (c) Stack plot of normalised ^{1}NMR data from Footscray "red" (09-06) and Footscray "smooth" (09-05) specimens.

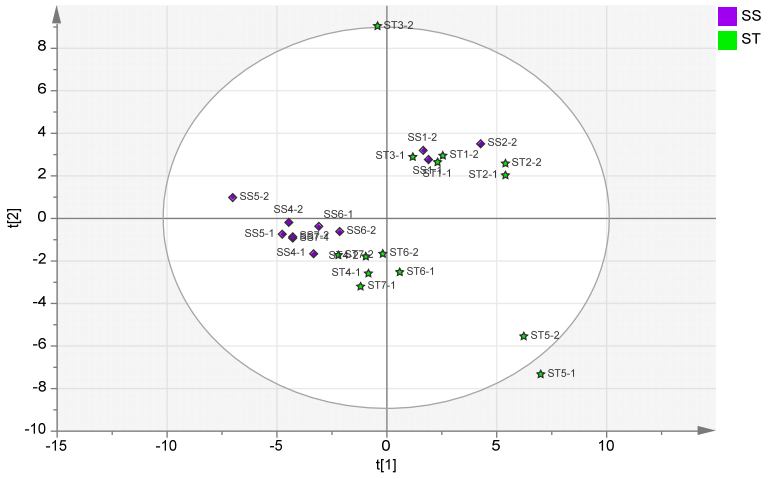
These analyses of the Queensland and Victorian specimens have identified fluctuations in the metabolome that may be explained by either the environment that the host plant was exposed to (Cloncurry vs. Killymoon Creek) or the inherent differences in the genome (Footscray "red" vs. Footscray "smooth"). Further work is required to completely understand the factors influencing these metabolomic differences, including a close study of the greenhouse progeny seed and comparison with the seed collected from the host plants, in addition to further PCR studies of the host plants to gain a greater understanding of how different the Australia population is.

2.3 Liquid Chromatography Mass Spectrometry (LCMS) based Metabolomics

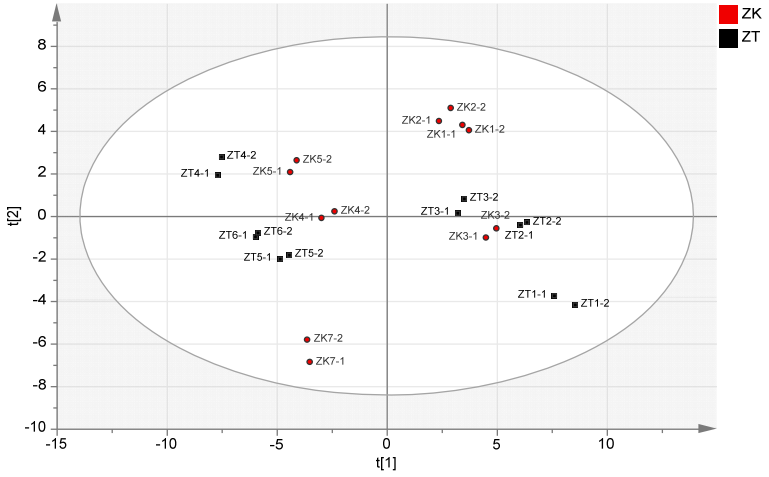
In collaboration with colleagues at the Swedish Defence Research Agency, FOI CBRN Defence and Security, it was demonstrated that Direct Infusion Mass Spectrometry (DIMS) analysis of the *R. communis* Biomarkers (RCB) and Seed Storage Protein populations of various *R. communis* extracts allowed for cultivar of an extract to be determined.[16] However, no provenance based classification could be made. To further investigate if it was possible for both cultivar and provenance to be determined using MS, LCMS analysis of eight specimens previously analysed via ^{1}H NMR was conducted.[15]

The two specimens each of "*sanguineus*" and "*zanzibariensis*" cultivars were analysed independently of the other specimens to understand what impact the local environment had on the metabolome. The PCA scores plots for the "*sanguineus*" ($R^2X = 0.45$, $Q^2X = 0.17$) and "*zanzibariensis*" ($R^2X = 0.48$, $Q^2X = 0.36$) specimens are shown in Figure 16a and 17b respectively. It was apparent from these scores plots that no provenance classification was observed. Furthermore, when each specimen was classified according to country of origin and subjected to OPLS-DA modelling, weak models with low predictive strength, and poor class classification were created. There are inherent difficulties in the LC-MS analysis of sugars and amino acids. Consequently, it could be expected that environment would have had no measureable impact when *R. communis* extracts when analysed by positive ion ESI LC-MS.

Considering these results, further analysis of the data was undertaken with each specimen classed according to cultivar. Subsequent OPLS-DA, variable selection using a combination of loadings scores of an individual variable, variable importance to projection (VIP) plot scores, and Cross Validation Standard Error (cvSE) were used to select variables of significance. This process removed variables that were not contributing significantly to the observed class classification. Applying the constraints that an individual variable needed to have a VIP score > 1, a cvSE < 1, and a loading score either > 0.05 or < -0.05, the data matrix was reduced to 65 variables. Outliers were removed using Hotelling T2 and DModX plots, and the reduced data matrix was again subjected to OPLS-DA ($R^2X = 0.84$, $Q^2X = 0.85$). The newly generated model had a significant increase in both the amount of variance explained and the predictive strength. The scores plot of LV1 vs. LV2 is shown in Figure 17a.



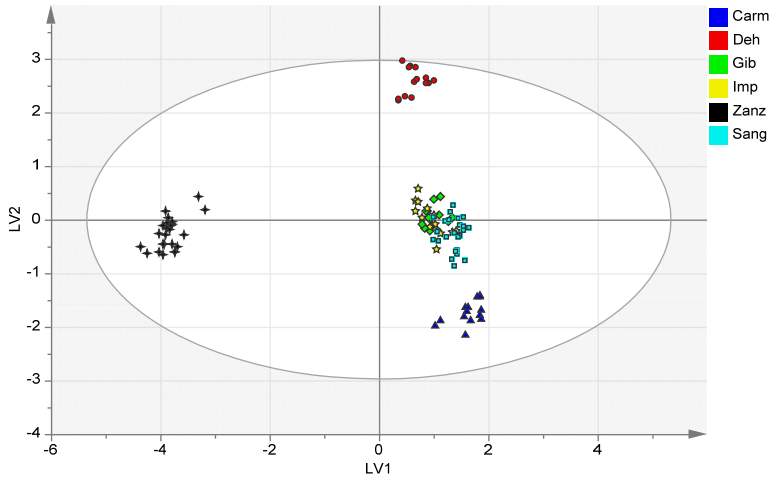
(a)



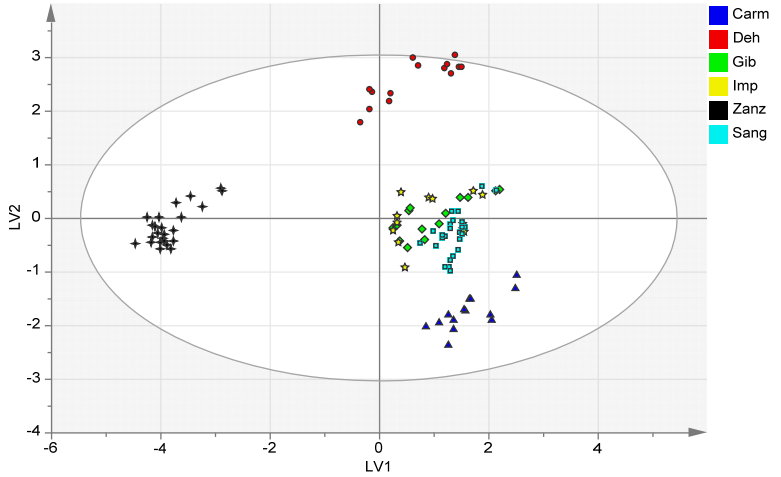
(b)

Figure 16 PCA scores plot of (a) "sanguineus" and (b) "zanzibariensis" specimens.

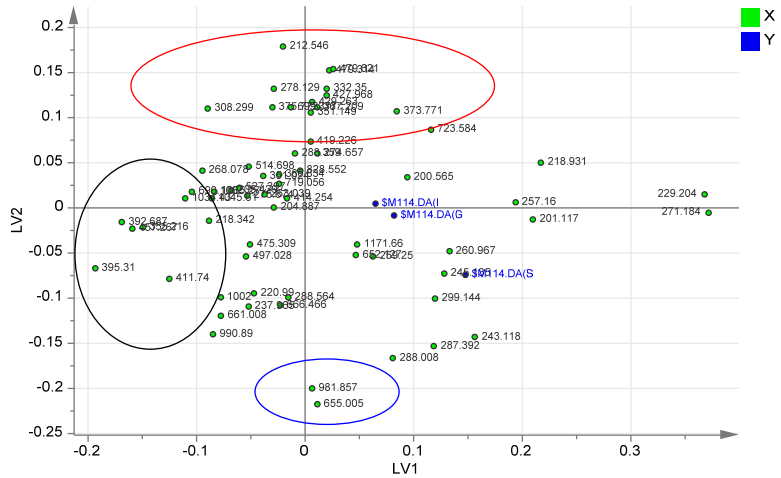
From this scores plot it was clearly identified that extracts from the "*zanzibariensis*" and "*dehradun*" cultivars clustered away from the other specimens. Additionally, there was a strengthening of the clustering of the "*carmencita*" cultivar away from other cultivars. To confirm the robustness of the model, a PLS-DA model ($R^2X = 0.88$, $Q^2X = 0.85$) was generated so permutation tests (100 rounds) could be conducted. The scores plot of LV1 vs. LV2 is shown in Figure 17b. What was initially noted was the similarity between the OPLS-DA scores plot shown in Figure 17a, and the PLS-DA scores plot shown in Figure 17b. The results of class based permutation testing (Figure a to f, Appendix E) confirmed that models based on the reduced data matrix were not over fitted for any class analysed. All permutations resulted in R^2X and Q^2X values significantly less that for the original model.



(a)



(b)



(c)

Figure 17 Results from OPLS-DA on the reduced data matrix. (a) Scores plot of LV1 vs. LV2; (b) PLS-DA scores plot of LV1 vs. LV2; (c) corresponding loadings scatter plot. Variables with significant loadings highlighted with coloured ellipses.

The loadings scatter plot corresponding to Figure 17a is shown in Figure 17c. Analysis of the loadings scatter plot allowed for the identification of variables that contributed to the observed clustering. For "*zanzibariensis*", ions at *m/z* 355.2, *m/z* 392.7, *m/z* 395.3, *m/z* 411.7, *m/z* 457.3, *m/z* 690.1 and *m/z* 1034.4 were identified (black ellipse). For "*dehradun*" ions, several ions of significance were identified (red ellipse), while for "*carmencita*", ions at *m/z* 655.0 and *m/z* 981.9 (blue ellipse) were identified. Other scores plots and their corresponding loadings scatter plots are shown in Figures g to i, Appendix D. These plots, in combination with those in Figure 17, allowed for the identification of a series of ions that could be used to discriminate between certain cultivars. In total, 24 ions were found to be significant contributors to the observed variance. Subsequent *t*-tests ($p \leq 0.001$) on these ions confirmed their validity.

High resolution mass spectrometry (HRMS) mass measurements were able to be made on 18 ions and molecular formulae proposed. These are summarised in Table 3. Six ions were readily identified as molecular ions of peptides. In particular, four ions were associated with RCB-1 (triply charged: *m/z* 689.9805^{3+}; doubly charged: *m/z* 1033.9788^{2+}) and RCB-3 (triply charged: *m/z* 654.6594^{3+}; doubly charged: *m/z* 981.4852^{2+}),[7] while two (*m/z* 718.6558^{3+} and *m/z* 828.0323^{3+}) were related to RCB-1.[7] These latter two ions were only present in extracts of "*impala*". Further investigations identified amino acid extensions of RCB-1.[7] The difference between RCB-1 and RCB-4 was the addition of Ser at the C-terminal. From Fourier Transform Ion Cyclotron Resonance Mass spectrometry (FTICRMS), it appears that the difference between RCB-1 and RCB-5 is the addition of Glu/Gln/Asp/Ser at the C-terminal. The proposed sequences for RCB-4 and RCB-5 are shown in Figure 18. Further MS/MS work is required to confirm these sequences.

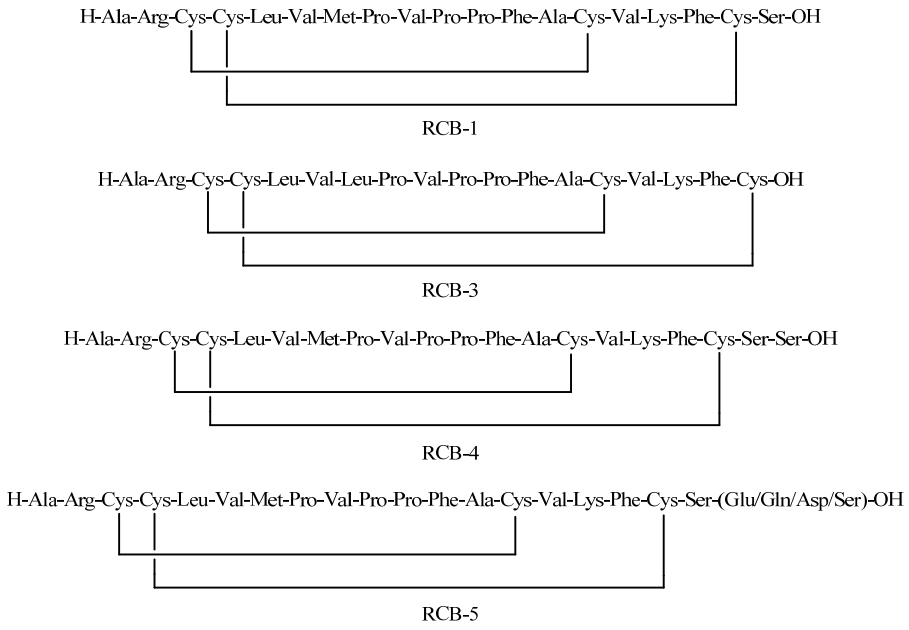


Figure 18 Sequences of the known RCB-1, -3, -4 and -5.

Of the remaining 12 ions, the molecular formulae of eight were confirmed through HRLC-MS/MS. Further interpretation of the MS/MS data for these eight ions allowed for some structural information to be elucidated. These MS/MS fragmentations are shown in Figure 19.

Table 3 Cultivar of R. communis with the corresponding identified ion of importance ($p \leq 0.001$) and proposed molecular formulae.

cultivar	ions (*m/z* $[M+H]^+$)[a]	Molecular Formula[b]
carmencita	229.2022	$C_{11}H_{25}N_4O$
	243.1818	$C_{11}H_{23}N_4O_2$
	261.0 @ 4.4 min	unknown
	271.2143	$C_{13}H_{27}N_4O_2$
	287.2083	$C_{13}H_{27}N_4O_3$
	654.6594^{3+}	RCB-3
	981.4852^{2+}	RCB-3
dehradun	278.1 @ 13.4 min	unknown
	332.1836	$C_{14}H_{26}N_3O_6$
	428.3 @ 7.6 min	unknown
	479.2887	$C_{24}H_{39}N_4O_6$
gibsonii	205.4 @ 5.1 min	unknown
	220.9 @ 2.0 min	unknown
	229.2022	$C_{11}H_{25}N_4O$
	238.0824	$C_{10}H_{12}N_3O_4$
	243.1818	$C_{11}H_{23}N_4O_2$
	259.1782	$C_{11}N_{23}N_4O_3$
	261.0 @ 4.4 min	unknown
	271.2143	$C_{13}H_{27}N_4O_2$
	497.0 @ 2.0 min	unknown
impala	287.2083	$C_{13}H_{27}N_4O_5$
	718.6558^{3+}	RCB-4[c]
	828.0323^{3+}	RCB-5[c]
sanguineus	229.2022	$C_{11}H_{25}N_4O$
	243.1818	$C_{11}H_{23}N_4O_2$
	259.1782	$C_{11}N_{23}N_4O_3$
	261.0 @ 4.4 min	unknown
	271.2143	$C_{13}H_{27}N_4O_2$
	287.2083	$C_{13}H_{27}N_4O_3$
	1033.9788^{2+}	RCB-1
zanzibariensis	355.1953	*$C_{11}H_{27}N_6O_7$*
	392.6974^{2+}	unknown peptide
	411.2741	*$C_{21}H_{37}N_3O_5$*
	457.1580	*$C_{19}H_{21}N_8O_6$*
	689.6551^{3+}	RCB-1
	1033.9788^{2+}	RCB-1

[a] Multiply charged ions identified
[b] Molecular Formula in italics are tentative
[c] sequence determined through HRMS and BLAST searches

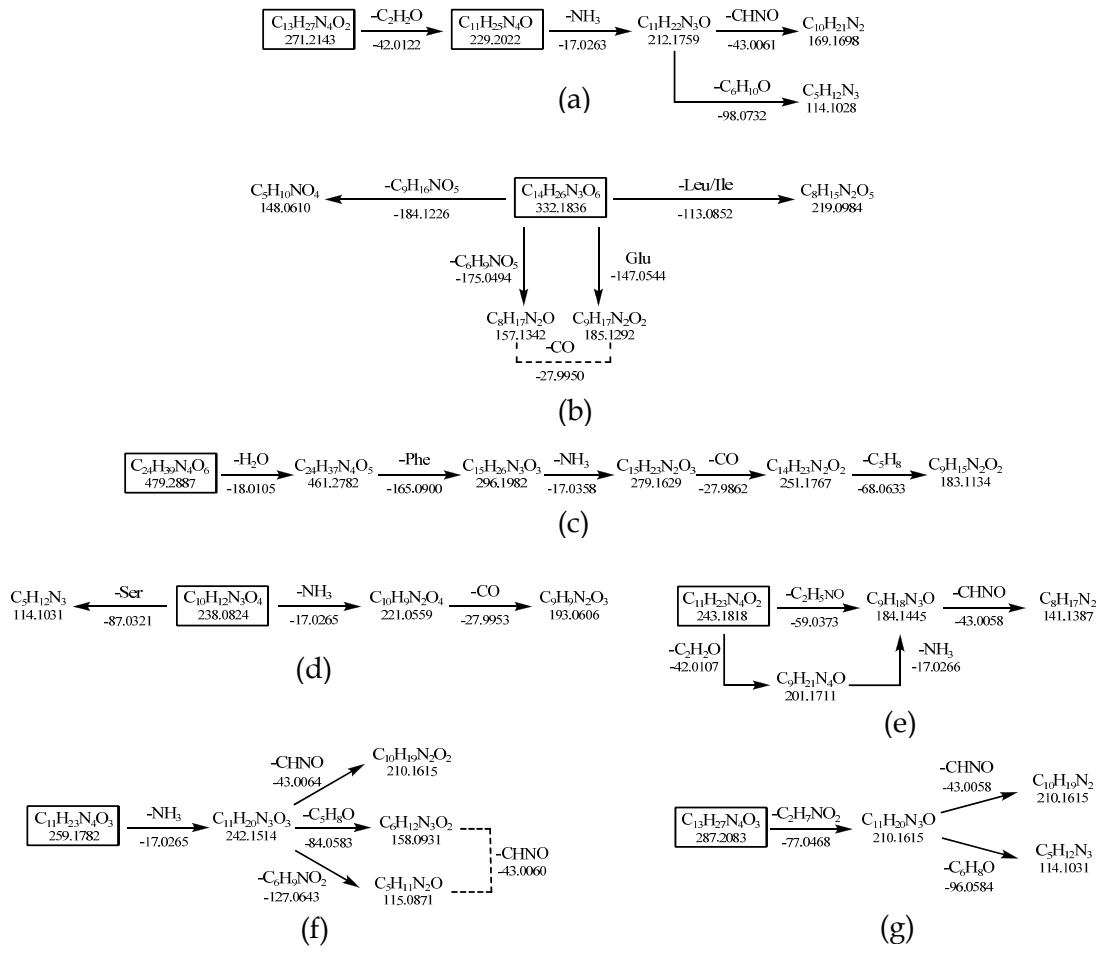


Figure 19 MS/MS fragmentations. (a) m/z 271.2143 and m/z 229.2022; diagnostic ions for "dehradun" at (b) m/z 332.1836; and (c) m/z 479.2887; and "gibsonii" at (d) m/z 238.0824; (e) m/z 243.1813; (f) m/z 259.1782; and (g) m/z 287.2083. Parent ions are highlighted in boxes.

An interesting observation from the data presented in Table 3 is that the ions at *m/z* 271.2143$^+$ and *m/z* 229.2022$^+$ were always present together. Analysis of the MS/MS (Figure 19a) spectra for these ions showed that these two compounds are related to each other, differing only by an acetate moiety. Considering that the extractions are performed in 2% aqueous acetic acid, it is possible that the ion at *m/z* 271.2143$^+$ is an artefact of the isolation process. Due to a lack of material, this currently remains unresolved. All other cultivars had unique ions identified that could be used for cultivar identification.

Analysis of the ions summarised in Table 3 established that extracts of "*sanguineus*" did not contain ions that were unique to this cultivar, with these ions present in extracts from one or more of "*carmencita*", "*impala*" and "*gibsonii*". However, only "*sanguineus*" extracts had all these ions present. It should also be noted that RCB-1[7] is present in all extracts. However, it is present in increased amounts in both "*sanguineus*" and "*zanzibariensis*" extracts relative to extracts of other cultivars.

All other extracts of cultivars had ions identified that were unique to that particular cultivar. Extracts of "*carmencita*" had four of the six ions present in extracts from other cultivars. However, the triply (*m/z* 654.6594^{3+}) and doubly (*m/z* 981.4852^{2+}) charged ions associated with the known peptide metabolite RCB-3 (Figure 18) were unique only to this cultivar.[7]

All the identified ions of importance for the "*dehradun*" extracts were unique to this cultivar. However, only the ions at *m/z* 332.1836^{+} and *m/z* 479.2887^{+} were abundant enough for HRMS/MS. Interpretation of the MS/MS data was suggestive of these ions being small peptides. For the ion at *m/z* 332.1836^{+} (Figure 19b), the sequence Leu/Ile-Ala-Glu was determined, with the loss of Glu and Leu/Ile residues from the C- and N-terminal respectively identified. For the ion at *m/z* 479.2887^{+} (Figure 19c), two Leu/Ile residues, and both a Phe and a Ser residue were identified. From the observed fragmentation in Figure 19c, it was apparent the Phe and Leu/Ile were positioned at the C- and N-terminal respectively. The positioning in the sequence of the remaining Leu/Ile and Ser residues was not able to be determined.

Of the nine ions identified in the "*gibsonii*" extracts, four ions were unique to this cultivar (Table 3). Accurate mass measurement could only be performed on one ion (*m/z* 238.0824^{+}), with formula validation achieved through HRMS/MS (Figure 19d). While the total structure was not able to be identified, loss of Ser residue from the N-terminal was identified. It appears that this molecule is a dipeptide, with some modification to the remaining amino acid.

In addition to the change in amounts of RCB-1 relative to other cultivars analysed, four additional ions were identified in the "*zanzibariensis*" extracts. While accurate mass measurements were performed on these, no MS/MS was possible due to the low ion abundance. Hence, the proposed molecular formulae for these ions are tentative. The presence of a doubly charged ion at *m/z* 392.6974^{2+} was also observed. Considering what has been identified in these extracts, it is expected that this to is likely to be a peptide.

The three ions remaining at *m/z* 243.1818 (Figure 19e), *m/z* 259.1782 (Figure 19f) and *m/z* 287.2083 (Figure 19g) were present in at least two of "*carmencita*", "*gibsonii*", "*impala*" and "*sanguineus*". While no amino acid residues were identified, there was homology between some the observed neutral losses shown in Figure 19. This includes the observation of losses of amino and amide functionalities, in addition to the loss of an acetate moiety. Again, this acetate moiety may be an artefact of the isolation process. Due to a lack of material, this currently remains unresolved. Considering the similarity in these losses, and what was previously identified, it is expected that these unresolved compounds are all modified peptides. A manuscript outlining this work has been submitted to the journal *Phytochemistry*.[17]

2.4 Environmental Considerations

2.4.1 Greenhouse Studies

An important consideration in these studies was to measure the impact the environment was having on the metabolome of the seed. To this end, all seeds that were investigated in these studies (listed in Appendix A) were grown in a greenhouse using the same potting mix,

humidity, temperature and water regimes. Through investigation of the metabolome of progeny seed collected from these greenhouse specimens, a comparison with original specimens could be made. This would then allow for an investigation of the impact of environment versus genetics on the metabolome. Plants were grown either in duplicate at Melbourne University, or triplicate at Australian Quarantine Inspection Service (AQIS).

Growing multiple specimens of each plant allowed for observations to be made between specimens of the same cultivar. While the duplicate specimens of the Australian seeds produced morphologically homogeneous plants, this was not the circumstance for some of the overseas specimens. In particular, triplicate "*zanzibariensis*" Tanzania specimens yielded two morphologically different plants. Similarly, the three "*zibo 108*" China specimens yielded plants with diverse plant and seed morphology. This was a concern as the overseas seeds were sourced from a single supplier. Hence, it was anticipated that they would be of consistent morphology. To further understand these observed differences, seeds from the three replicate specimens of "*zibo 108*" were analysed by ^{1}H NMR and subjected to both OPLS-DA (R^2X = 0.76, Q^2X = 0.26) and PLS-DA (R^2X = 0.99, Q^2X = 0.78) modelling. The corresponding scores plots are shown in Figures 21a and b respectively. No strong class classification of plants was observed. Furthermore, although a reasonable Q^2X value was obtained for the PLS-DA model, permutation testing indicated that the model was not robust. Permutation tests are based on scrambling sample labels, while the variables remain constant, and rebuilding the model. If the model is being over-fitted (i.e. classifications based on noise), then the ratios of R^2(new)/R^2(model) and Q^2(new)/Q^2(model) would approach one. This result was observed, with the corresponding plot shown in Appendix F. These data indicated that from a ^{1}H NMR perspective, no difference in the metabolome of the three "*zibo 108*" plants could be detected, despite the observed differences in plant morphology.

To validate the application of metabolomics for provenance and cultivar determination, verification that the chemical shift regions identified previously in Figures 10 and 11 as being critical for the observed class classification in Figure 8 was required. To this end, a comparison of the greenhouse seed progeny and supplied seed was undertaken. This was performed to confirm that these observations were as a consequence of the environment the host plants were exposed to, as opposed to the genetics unique to the cultivar of the host plant. If validated, this would be evidence for the impact the environment has on the plant's metabolome.

Firstly, the greenhouse data was scrutinised to ascertain if cultivar information could be discriminated for progeny seed. OPLS-DA modelling (R^2X = 0.94, Q^2X = 0.60) was performed, with the Hierarchical cluster analysis (HCA) dendrogram shown in Figure 21. What is observed is good class classification, with only the "*black diamond*" and "*Bangkok brown*" having multiple samples wrongly grouped. One specimen of "*lamoa red*" was incorrectly classified. Interestingly, the misclassified specimens were grouped together with specimens collected from the same country. It is not understood at this time why this would be the case. It could be that there is not a great deal of genetic difference between these specimens. What was apparent from this data is that the plants grown in the greenhouse generally retained their cultivar specificity.

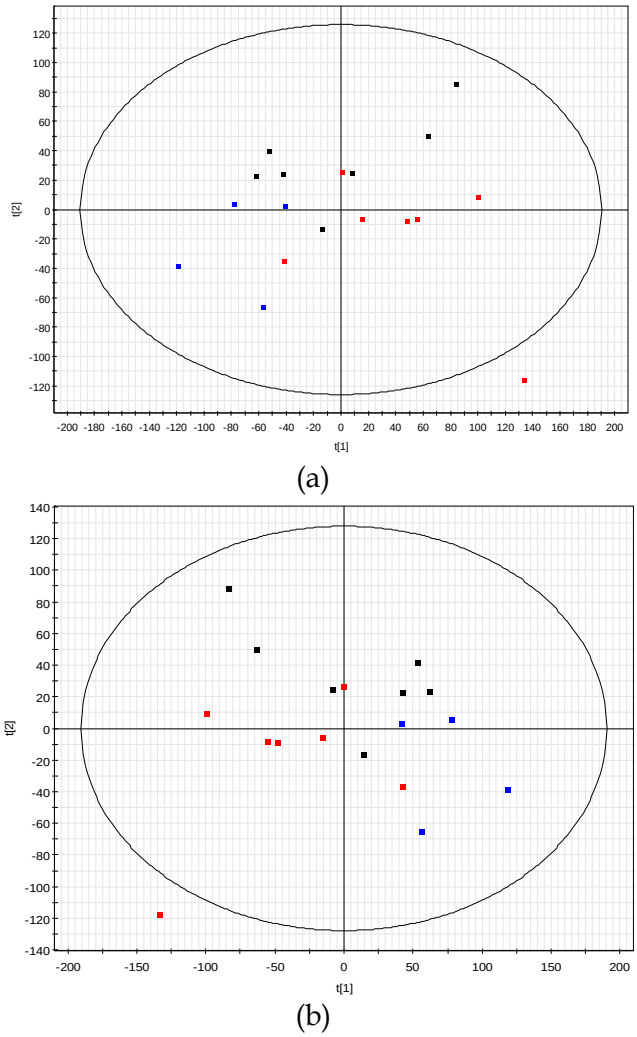


Figure 20 Analysis of the three "zibo 108" plants to assess for metabolome differences in the three morphologically different plants (black: tree 1; red: tree 2; blue: tree 3). (a) OPLS-DA scores plot (LV1 vs. LV2); (b) PLS-DA scores plot (LV1 vs. LV2).

Following this, data generated from the greenhouse plant progeny were compared with the data from the seed supplied specimens. This was done to understand if there was a significant difference between the greenhouse progeny seed, and the seed supplied by the seed supplier. PCA ($R^2X = 0.99$, $Q^2X = 0.98$) modelling was undertaken, with the subsequent scores plots (PC1 vs. PC2) are shown in Figure 22a. What is immediately apparent from this analysis is that there is a clear delineation in the scores plots between wild and greenhouse seeds. The corresponding loadings plot for PC1 is shown in Figure 22b. Interestingly, ricinine[14] and demethyl analogues[14] (blue box), sucrose (red box) and phenylalanine (green box) were found to be significant contributors to the observed separation on PC1. This is consistent with findings made in the initial study.[9] Primary metabolites such as sugars are required for basic function. It stands to reason that fluctuations in the primary metabolism could indeed be good indicators of environment. Plants exposed to harsh environments would potentially have lower levels of primary metabolites compared to those that are not.

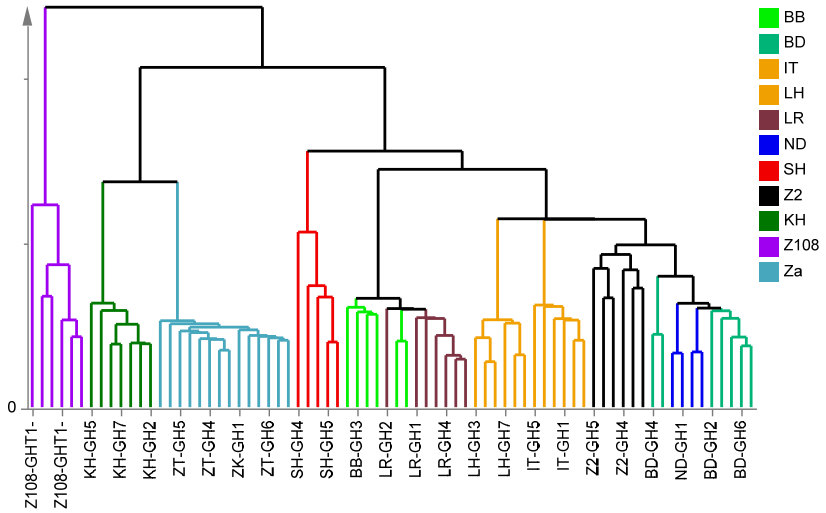


Figure 21 HCA dendrogram of the greenhouse specimens classed according to cultivar.

When comparing the loadings plots in Figure 22b with Figure 9a, it was apparent there are significant differences. It is expected that these were a manifestation of changes in the host plants secondary metabolism. Production of secondary metabolites in plants are influenced by the environment the plant is exposed to.[18] Consequently, any compounds that were shown to be present as a direct response to this will be strong candidates for provenance based biomarkers.

2.4.2 Seasonal Fluctuations

One of the more critical aspects of this project was to determine the effect the different seasons had on the observed metabolome. The basic premise of this research was to use the difference in environmental conditions across Australia to be able to identify gross geographic location. Ideally, however, this technique needs to be resistant to the more subtle seasonal environmental fluctuations that a wild plant is exposed to. To this end, a longitudinal study of three plants across 12 months was conducted.

The three plants were located within a twelve-kilometre radius from Melbourne central business district (Avondale Heights, Footscray and Richmond). The summary of climate observations in Melbourne during 2010 is shown in Figure 23. The city of Melbourne provided a good model to explore seasonal variation as the climate conditions evidently differentiate between seasons. The difference in mean temperature at 3 pm from February compared to July was 12.7 °C. There was a 123.4 millimetre difference in total precipitation from April compared to October.

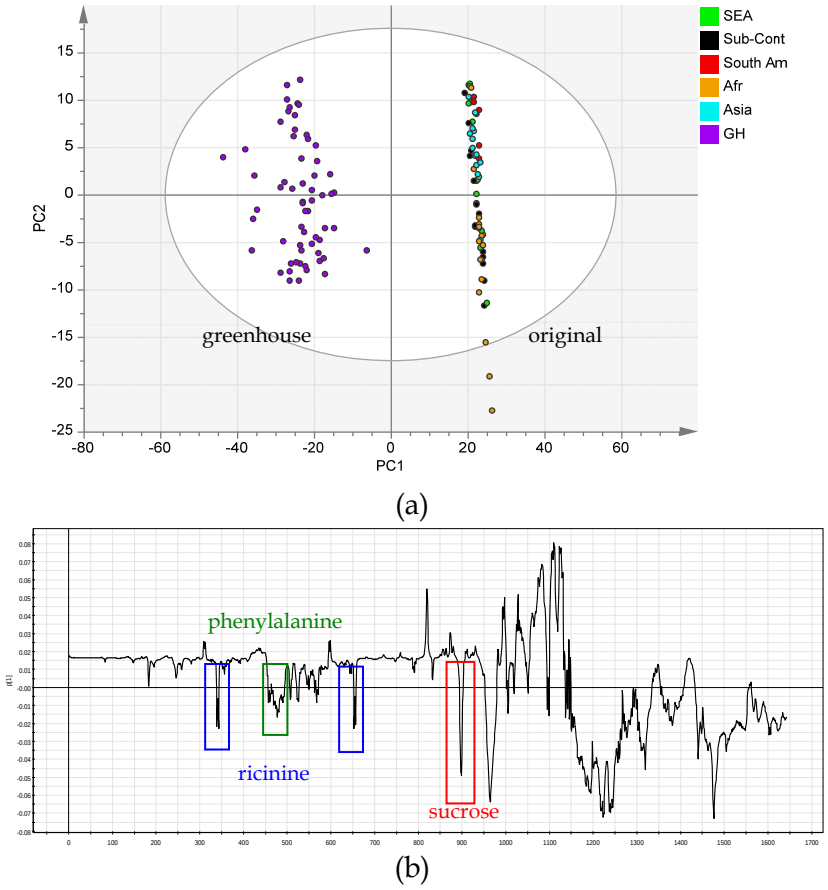


Figure 22 Comparison of wild specimens, and the progeny of identical specimens grown on a greenhouse. (a) PCA scores plot of PC1 vs. PC2 wild (multi colours) and greenhouse (purple) specimens; (b) The corresponding loadings plot for PC1.

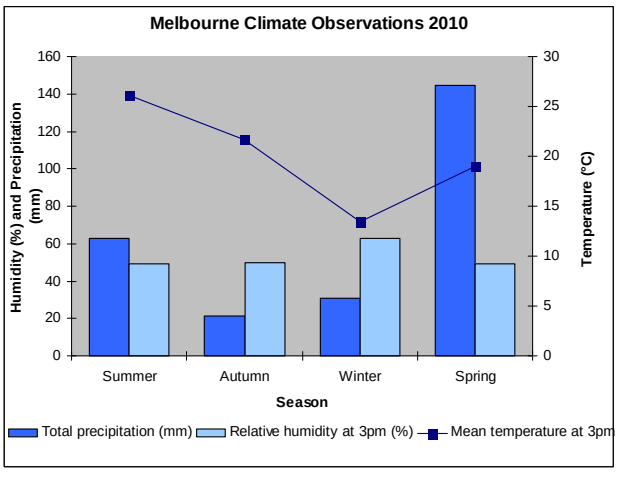
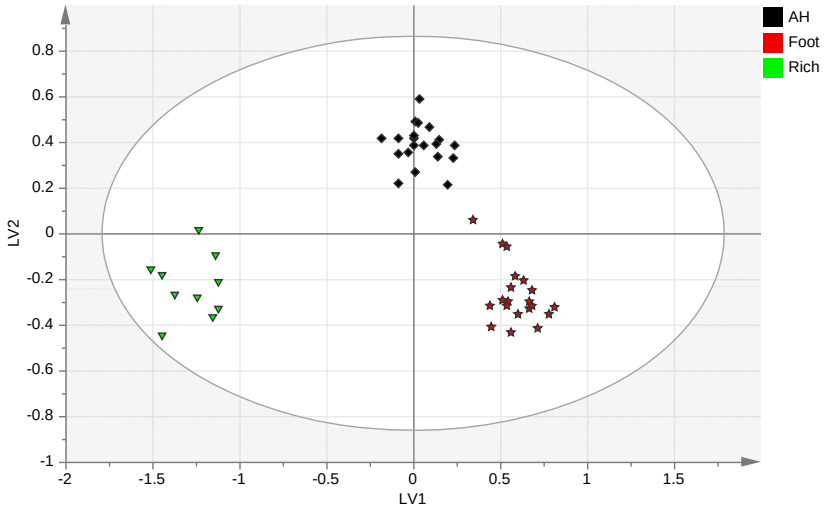


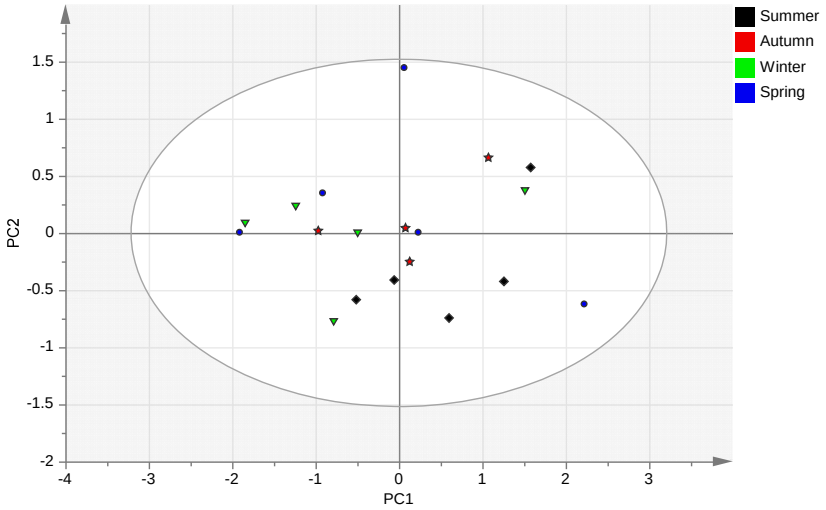
Figure 23 Climate observations for Melbourne during 2010.

As previously demonstrated, an OPLS-DA (R^2X= 0.88; Q^2X = 0.75) model was able to classify plants from disparate locations (Figure 24a). In contrast, an OPLS-DA (R^2X = 0.62; Q^2X = 0.06) model of seeds collected during different season from the same plant in Footscray resulted in

a significantly weaker model being generated (Figure 24b). Whilst a model was generated, the model statistics indicate that it was particularly weak. These observations lead to the conclusion that if there was a seasonal variation in the metabolome, it is minimal, and not impacting on the metabolome in a measurable way. This finding allows us to say with some confidence that any classification of an extract to a location is independent of seasonal climatic perturbations.



(a)



(b)

Figure 24 a) OPLS-DA scores plot (LV1 vs. LV2) of three separate specimens collected for seasonal variation analysis classed to specimen, Footscray B (black), Richmond (blue) and Avondale Heights (blue); (b) OPLS-DA scores plot of four separate specimens collected for seasonal variation analysis from Footscray classed to season. Summer (black), Autumn (red), Winter (blue) and Spring (green).

2.5 Milestone 3: DNA signature studies

"Terrorist cookbook" methods of ricin production are relatively crude, with the final products likely to contain residual DNA from the initial seed material. This residual DNA can be used for detection and identification of ricin by methods such as PCR, from which very small amounts of initial DNA can be detected with high specificity. However, the 'terrorist cookbook' methods use high quantities of chemicals such as salt, acetone, and acetic acid in the extraction process. The presence of these chemicals in the crude ricin preparations is likely to inhibit PCR enzymes, leading to false negative results. Additionally, plant substances such as oil and protein can also inhibit the PCR. The aim of this project was to determine a method for DNA purification from the crude ricin preparations that would remove PCR-inhibitory chemicals. Additionally, this project aimed to assess at what stage, if any, in the extraction procedures the DNA signature was lost such that detection by PCR was not possible.

Three published crude ricin extraction methods were used to generate a total of 14 ricin preparations, consisting of intermediate and final products. For each ricin sample, eight DNA purification techniques were used, and the results were compared for DNA yield and PCR efficiency. The Roche High Pure PCR Template Preparation kit was found to be the best technique for the extraction of DNA from the ricin preparations and was the only technique to give positive results for all samples in all PCR assays.

The ricin extraction methods were then used on seeds from three *R. communis* cultivars. In addition to the initial cultivar, two other cultivars were included to assess the applicability of the Roche High Pure PCR Template Preparation kit purification method on different bean phenotypes. In general, the PCR results obtained from the three-cultivar samples were similar to each other and to the initial result, indicating good reproducibility.

In summary, these results clearly demonstrate that sufficient DNA was present in the crude ricin preparations for detection using PCR methods, however purification of the DNA from the crude ricin extracts was necessary to remove PCR inhibition. Comparison of eight DNA purification methods indicated that some were superior in terms of the yield and purity obtained. This has positive implications for intelligence and forensic investigations, and therefore for the possible prosecution of individuals suspected to be extracting ricin for illegal, harmful use. A report has been generated for this work and has been be circulated.

3. Summary

The awarding of the NSST grant has allowed for the chemistry of *R. communis* to be investigated for cultivar and provenance determination, and to investigate the longevity of *R. communis* DNA signature in both crude and reasonably pure ricin preparations. Through these investigations several milestones were proposed. These, along with progress, are detailed below.

Milestone 1:
Through IRMS and ICPMS analysis of seed extracts, it was found that IRMS had limited application to provenance determination with the isotope ratios that were investigated. LA-ICPMS of the constituent parts of individual seeds yielded results that allowed a prediction of provenance to be made. However, this prediction could not be made on the supplied MWCO fractions in 2% acetic acid. This was due to the interference of the organic acid leading to ion suppression. A further limitation to the laser ablation technique is that metal ions can undergo polyvalent interactions, potentially leading to false positives. Therefore, expert interpretation of the generated data is required. At this stage LA-ICPMS on the whole seed is the technique best suited to provenance determination. Solution based ICPMS analysis of suspected dried powders of *R. communis* extracts requires significant method development, with investigations using ICP-AES underway.

Milestone 2:
Through the complete ^{1}H NMR and mass spectral analysis of extracts of known cultivar and provenance it has been demonstrated that there is significant potential for this methodology to be applied for both provenance and cultivar determinations. In particular, it was found that ^{1}H NMR based metabolomics of seed extracts, followed by supervised multivariate statistical analysis allowed for both continent and country to be identified. Within a country, specimens were able to be further distinguished into cultivar. Furthermore, physical quantities of sucrose, ricinine[14] and the demethyl analogues,[14] and phenylalanine were contributing to the observed classification. The results from this study have been published.[15] When comparing the statistical results of the ^{1}H NMR data of extracts from the seed supplier with progeny collected from the greenhouse, a significant difference was observed. This observation suggested that there is a marked difference in the metabolome for a seed grown in differing environmental conditions.

Conversely, LCMS based metabolomics was a satisfactory technique for cultivar determination. This is most likely as a consequence of the major discriminator compounds not being amenable to positive ESIMS. The results from this work have been submitted for peer-reviewed publication in *Phytochemistry*.[17]

When applied to extracts of Australian specimens, ^{1}H NMR based metabolomics analysis allowed for state based classification to be achieved. Greenhouse specimens are currently being extracted, and will be analysed and compared against collected data from the Australian specimens. Again, the aim here is to identify genetic vs. environmental marker compounds. Once completed, this work will be submitted for peer-reviewed publication.

It needs to be highlighted here that to further confirm that classification is due to environmental effects, PCR analysis needs to be conducted on both original and greenhouse seed progeny. This is to confirm genetic purity.

Milestone 3:
The analysis of the longevity of the DNA signature identified that *R. communis* DNA is significantly longer lived in an extract than first thought. This observation therefore makes PCR based methodologies to determine the potential presence of ricin in highly purified white powder a critical technique. An additional insight into this research was that for PCR to be

successful, an initial sample clean up is required before commencing PCR. Two technical reports have been published on the findings from this work.[9,10]

For completion, there are several investigations that need resolution. These are documented below:

- Isolation of discriminate provenance compounds needs to be conducted, and the corresponding structures elucidated.
- Complete the analysis of wild and greenhouse progeny seed extracts so environmental marker compounds can be confirmed.
- Complete the mass spectral analysis of seed extracts, including compound identification.
- Completions of the LA-ICPMS analysis, in addition to the evaluation of ICP-AES as a valid technique for provenance determination of aqueous acidic *R. communis* extracts.

Once this has been completed, these results will be written up for publication.

4. Experimental

4.1 Chemicals

All solvents used were analytical grade. Water and acetone were purchased from Merck. Acetic acid was purchased from Sigma-Aldrich. Deuterated NMR solvents (D_2O, d_4-acetic acid and TSP) were supplied by Cambridge Isotopes. MWCO filters (30 kDa) were obtained from Millipore Corporation (USA).

4.2 Collection of *R. communis* seed specimens

Collections of environmental samples of seed and soil specimens were made from various locations in Victoria, New South Wales, Queensland, South Australia and Western Australia during a three-week period in 2009. A total of twenty-five seed specimens (five from each State) were selected for metabolome analysis (Appendix A). Three plants from Victoria were selected as seasonal variation specimens. To this end, seeds were collected from each plant during each season within the same calendar year. The following plant characteristics and location details were recorded at the time of collection:

- GPS coordinates and general description and photographs of location and plant
- health and height of plant
- stem colour
- type of internode and length and number of nodes on main stem

- leaf and central vein colour, leaf shape, presence of waxy bloom, type of lancination on the third leaf from the top and number of lobes on leaf
- spike shape and compactness
- inflorescence colour and length of pedicel
- seed capsule colour, density, length and colour of spine
- seed size.

4.3 Extraction of *R. communis* seed specimens

Caution: *Ricin is a highly toxic protein, and extractions of R. communis need to be conducted with extreme care. All extraction work performed for this investigation was conducted in a laboratory within a fume cupboard. Laboratory coats, glasses and gloves were worn during all extraction work.*

For each specimen of *R. communis* analysed in this study, three mature seeds were selected randomly and extracted together to form a biological replicate. For Study 1, seven biological replicates were extracted from each of the eight specimens, resulting in 56 crude extracts. For this study two extraction methods were used which varied only in the procedure for the initial crushing of the seeds: Extraction method 1: Three biological replicates from each specimen were crushed with a mortar and pestle and transferred into a 50 mL Falcon tube containing 10 mL acetone. The mixture was sonicated for 20 min, and then centrifuged (room temperature, 3000 rpm for 30 min). Extraction method 2: Four biological replicates from each specimen were crushed using an Ultra-Turrax Tube Disperser containing the seeds, six glass mixing balls and 10 mL acetone. The seeds were blended for 8 min at maximum speed. The mixture was then transferred to a 50 mL Falcon tube and centrifuged (room temperature, 3000 rpm for 30 min). All further steps remained the same for all biological replicates. The acetone was decanted, and the seed mash again extracted with a 10 mL aliquot of acetone (room temperature, 20 min sonication, 30 min centrifugation at 3000 rpm). On removal of the acetone, the seed mash was extracted twice with 7.5 mL of 2% aqueous acetic acid solution (room temperature, 20 min sonication, 30 min centrifugation at 3000 rpm). The combined acetic acid extract was filtered twice through 30 kDa Molecular Weight Cut Off (MWCO) filters to remove both *R. communis* agglutinin and ricin. The aqueous extracts were stored at -30 ^{o}C until required for chemical analysis.

For Studies 2 and 3, extraction method 2 was used. In total, 7 biological replicated from 18 specimens (126 crude extracts) were analysed for study 2, with 7 biological replicated from 25 specimens (175 crude extracts) analysed for study 3.

For the blinded samples used in Study 1, two different cultivars (*"gibsonii"* Zimbabwe and *"dehradun"* India) were extracted using three different extraction techniques to give a total of six validation samples. In addition to extraction methods 1 and 2 used above, a third extraction method was also employed. This method involved crushing the seeds in the tube disperser with six glass mixing balls and 10 mL of 2% acetic acid. The seeds were crushed for 8 min at maximum speed. The mixture was then transferred to a 50 mL Falcon tube. Dichloromethane (20 mL) was then added to the Falcon tube and mixed gently. The solution

was centrifuged (4 °C, 1 h at 3000 rpm), then the acetic acid removed and twice filtered through a 30 kDa MWCO filters prior to analysis. These blinded extracts were then given to a third person for data collection and multivariate statistical analysis.

4.4 ICPMS multivariate statistical analysis

Three biological replicates from the 25 Australian specimens were selected for analysis. Of these, seeds from 3 specimens (9 seeds in total) were found to be of poor quality and could not be analysed. In total 22 specimens (for a total of 66 seeds) were subjected to LA-ICPMS. Each biological replicate was subjected to three LA-ICPMS analyses (technical replicates) in different locations on the core. In total 60 isotopes were analysed for, and any isotope with counts less than 100 is approaching the detection limit (DL) of the instrumentation. Consequently, any isotopes that had values less than 100 counts were removed from the data set. In total, the data set was composed of 15 isotopes (^{24}Mg, ^{27}Al, ^{44}Ca, ^{53}Cr, ^{55}Mn, ^{57}Fe, ^{60}Ni, ^{65}Cu, ^{66}Zn, ^{75}As, ^{85}Rb, ^{88}Sr, ^{98}Mo, ^{138}Ba, ^{202}Hg). The data matrix was normalised to the sum of the signal area, log transformed, scaled to Unit Variance (UV), and subjected to OPLS-DA. To further confirm the strength of the models, randomly selected samples were removed from the generated data matrix. The OPLS-DA models were rebuilt, and the withheld specimens used as a prediction set.

4.5 NMR sample preparation and data collection

^{1}H NMR data was collected on a Bruker Avance-500 NMR spectrometer (Bremen, Germany) operating at a ^{1}H NMR frequency of 500.13 MHz running Bruker Topspin 2.1 NMR software. The spectrometer was equipped with a standard geometry 5 mm diameter BBI (Broad Band Inverse) probe head. Each sample was freeze dried and resuspended in D_2O [with 0.01% (trimethylsilyl)-2,2,3,3-d_4-propionic acid (TSP) and 2% d_4-acetic acid] at a concentration of 25 mg/mL. A 600 µL aliquot of each extract was transferred to a 5 mm NMR tube immediately prior to analysis. All ^{1}H NMR data was collected using the noesypresat solvent suppression pulse sequence over a δ 20.00 sweep width with 64 scans and 64k data points. The total acquisition time was 8.17 s, the recycle delay time set to 5 s, and the pulse width (90^{o}) was manually calculated for each extract. The probe temperature was set to 298 K. Processing of the Free Induction Decay was performed with line broadening set to 1.0 Hz. All ^{1}H NMR spectra were referenced to TSP (δ 0.00 ppm) and manually phased and baseline corrected.

4.6 NMR multivariate statistical analysis

All collected ^{1}H NMR data was manually phased and baseline corrected, then binned into δ 0.005 bin widths from δ 0.50 to δ 9.50 (residual D_2O and acetic acid regions removed) using the Prometab v.3.3[19] script in Matlab 2009b (The Mathworks, USA). Binned spectra was then normalised to the area of the TSP peak, with a generalised log functions (overseas specimens λ = 1.2044 x 10^{-7}; Australian specimens λ= 4.3898 e-007)[20] applied to the data. The generated matrix was exported into SIMCA 13 (Umetrics AB, Umëa, Sweden) and subjected to Pareto

scaling. Data matrices were subjected to both OPLS-DA and PCA analysis. To further confirm the strength of the models, randomly selected samples were removed from the generated data matrix. The OPLS-DA models were rebuilt, and the withheld specimens used as a prediction set.

4.7 LCMS sample preparation and data collection

1 mL aliquots of extract were freeze dried, and resuspended at a concentration of 20 mg/mL in 2% aqueous acetic acid. Extracts were filtered through a 0.45 μm filter, then centrifuged at 10000 rpm for 5 min. Following this, a 20 μL injection of each extract was made onto an Agilent LC/MSD Trap XCT mass spectrometer connected to an Agilent 1100 series LC system comprising of an in-line degasser, binary pump, auto-injector, column heater and diode array detector. Data was collected via Agilent ChemStation LC for 3D software (Rev.A.09.03). Samples were eluted at 0.4 mL/min through a Phenomenex Luna 5 μm 50x2.0 mm C18 HPLC column, using gradient elution from H_2O (+ 0.05% formic acid) to 7:3 MeCN:H_2O (+ 0.05% formic acid) over 30 min. The order of the extracts was randomised to reduce the effect of any systematic errors. Furthermore, each extract was injected non-sequentially in duplicate (technical replicate). This provided a QC set to measure the robustness of the instrument, and also a predictive set to confirm the strength of the generated models.

4.8 LCMS multivariate statistical analysis

All Base Peak Chromatograms (BPC) were converted to mzXML format and imported into mzMine.[21] All BPC were aligned, resulting in a 2200 x 112 matrix. The data matrix was normalised to the sum of ion intensity,[22] then imported into SIMCA. Within SIMCA 13 the data was log transformed and Pareto scaled. All variables associated with the duplicate injections were removed, with the residual variables associated with the initial injections subjected to supervised OPLS modelling. Important variables were selected within SIMCA 13 using VIP and seCV scores. The resultant matrix (103 x 56) was again subjected to PLS-DA, OPLS-DA and PLS-DA modelling. To further confirm the strength of the models, randomly selected samples were removed from the generated data matrix. Subsequent PLS-DA and OPLS-DA models were rebuilt, and the withheld specimens used as a prediction set.

5. Acknowledgements

We would like to acknowledge the NSST unit with in PM&C for the funding of this work, and in particular Dr John Colton for the generous support he has given us during the lifetime of the grant. We would like to thank the AFP for their financial support, and specifically Dr Eric Wenger for his valuable contributions and suggestions. Also thanks to our collaborators from TSW Analytical at The University of Western Australia (Dr Rachel Scadding and Prof John Watling) for the ICPMS analysis, Prof Kliti Grice from Curtin University for IRMS analysis, and The University of Melbourne for the greenhouse for growing Australian specimens. We would also like to thank the staff at AQIS Knoxfield, in particular Kimberly Thomas, for growing overseas specimens in the PC2 greenhouse. Dr Simone Rochfort from The Victorian Department of Primary Industry gave valuable input and useful discussions. We would also like to acknowledge Ben Gordon and Elissa Caines for their contributions.

6. References

1. Audi, J.; Belson, M.; Patel, M.; Schier, J.; Osterloh, J. *JAMA – J. Am. Med. Assoc.* **2005**, *294*, 2342-2351.

2. Bradberry, S. M.; Dickers, K. J.; Rice, P.; Griffiths, G. D.; Vale, J. A. *Toxicological Reviews*, **2003**, *22*, 65-70.

3. Convention on the Prohibition of the Development, P., Stockpiling and use of Chemcial Weapons and on their Destruction,, Summary of CWC-Schedules and their Relevance to Chemical Warfare. In http://www.dfat.gov.au/cwco/download/index.htm, Weapons, O. f. t. P. o. C., Ed., last accessed May 17 2012.

4. Musshoff, F.; Madea, B. *Drug Test. Anal.* **2009**, *1*, 184-191.

5. Fredriksson, S. A.; Hulst, A. G.; Artursson, E.; de Jong, A. L.; Nilsson, C.; van Baar, B. L. M. *Anal. Chem.*, **2005**, *77*, 1545-1555.

6. Brinkworth, C. S.; Pigott, E. J.; Bourne, D. J. *Anal. Chem.* **2009**, *81*, 1529-1535.

7. Ovenden, S. P. B.; Fredriksson, S. A.; Bagas, C. K.; Bergstrom, T.; Thomson, S. A.; Nilsson, C.; Bourne, D. J. *Anal. Chem.* **2009**, *81*, 3986-3996.

8. Brinkworth, C. S. *Anal. Chem.* **2010**, *82*, 5246-5252.

9. Taylor, K.; Peeler, J.; Blackman, L.; Shahin, S. "Detection of residual DNA in crude ricin extracts by PCR" *DSTO Technical Report*, **2010,** DSTO-TR-2496.

10. Taylor, K.; Thirkettle-Watts, D. "*Ricinus communis* cultivar typing using SSR" *DSTO Technical Report*, **2010**, DSTO-TR-2637.

11. Krishnan, P.; Kruger, N. J.; Ratcliffe, R. G. *J. Exp. Bot.* **2005**, *56*, 255-265.

12. Dunn, W. B.; Bailey, N. J. C.; Johnson, H. E. *Analyst* **2005**, *130*, 606-625.

13. Dieterle, F.; Ross, A.; Schlotterbeck, G. A. *Anal. Chem.* **2006**, *78*, 4281-4290.

14. Darby, S. M.; Miller, M. L.; Allen, R. O.; *J. Forensic Sci.*, **2001**, *46*, 1033-1042.

15. Pigott, E. J.; Roberts, W.; Ovenden, S. P. B.; Rochfort, S.; Bourne, D. J. *Metabolomics*, **2012**, *8*, 634-642.

16. Fredriksson, S. A.; Ovenden, S. P. B. unpublished results.

17. Ovenden, S. P. B.; Pigott, E. J.; Rochfort, S.; Bourne, D. J. *Phytochemistry*, **2012**, submitted.

18. Sun, X.; Chen, P.; Cook, S. L.; Jackson, G. P.; Harnly, J. M.; Harrington, P. B. *Anal. Chem.*, **2012**, *84*, 3628-3634**.**

19. Viant, M. R. *Biochem. Biophys. Res. Commun.* **2003**, *310*, 943-948.

20. Parsons, H. M.; Ludwig, C.; Güther, U. L.; Viant, M. R. *BMC Bioinformatics*, **2007**, *8*, 234.

21. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. *BMC Bioinformatics*, **2010**, *11*, 395.

22. Wagner, M.; Naik, D.; Pothen, A. *Proteomics*, **2003**, *3*, 1692-1698.

Appendix A: Australian specimen collection sites

The 25 Australian specimens selected for study, including date and site of collection (with GPS co-ordinates).

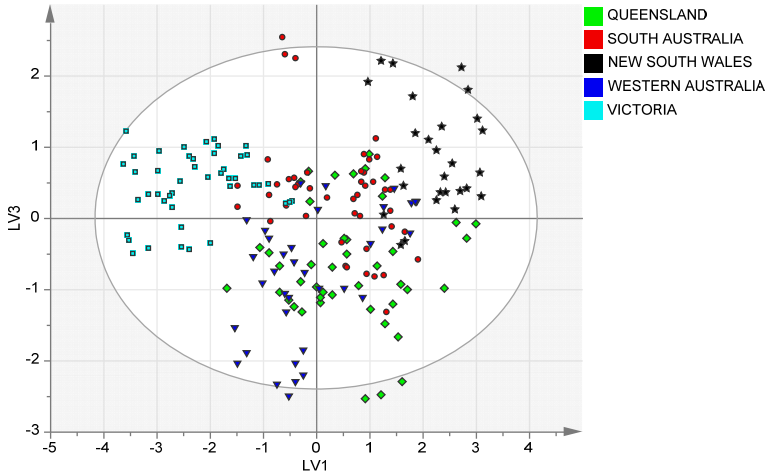
Specimen	Collection date	GPS	Location
07-19	09.07.2009	S 37° 45.604 E 144°51.678	Avondale Heights, Vic
08-02	29.02.2008	S 37°49.505 E 144° 59.344	Richmond, Vic
09-05	03.02.2009	S 37° 48.152 E 144°53.832	Footscray, Vic
09-06	03.02.2009	S 37° 48.167 E 144°53.863	Footscray, Vic
09-13	02.03.2009	S 37°44.407 E 144° 57.384	Coburg, Vic
09-03	28.01.2009	S 33°94.769 E 151° 16.490	Armcliffe, NSW
09-51	27.07.2009	S 37°02.17 E 150° 57.35	Holsworthy, NSW
09-54	28.07.2009	S 33°49.40 E 151° 01.09	Harris Park, NSW
09-59	29.07.2009	S 32°52.33 E 151° 44.46	Kooragang, NSW
09-60	30.07.2009	S 32°54.104 E 151° 45.13	Tighes Hill, NSW
09-27	13.07.2009	S 35°05.573 E 138° 32.210	Reynella, SA
09-30	13.07.2009	S 35°25.982 E 138° 19.397	Carrickalinga, SA
09-31	13.07.2009	S 34°57.255 E 138° 40.362	Waterfall Gully, SA
09-32	13.07.2009	S 34°52.05 E 138° 36.02	Blair Athol, SA
09-33	14.07.2009	S 34°52.40 E 138° 36.09	Sefton Park, SA
09-35	15.07.2009	S 31°53.746 E 115° 48.313	Osbourne Park, WA
09-37	15.07.2009	S 31°53.37 E 115° 46.13	Scarborough, WA
09-40	15.07.2009	S 32°02.02 E 115° 45.16	North Fremantle, WA
09-41	15.07.2009	S 32°02.02 E 115° 45.16	North Fremantle, WA
09-46	29.04.2009	S 31°57.315 E 115° 54.565	Ascot, WA
09-62	06.08.2009	S 27°24.628 E 152° 58.449	Mitchelton, Qld
09-65	07.08.2009		Brisbane Airport, Qld
09-66	10.08.2009		Cloncurry, Qld
09-70	11.08.2009		Killymoon Creek, Qld
09-72	11.08.2009	S 33°49.40 E 151° 01.09	Townsville, Qld

Specimens of known cultivar and provenance to be investigated using both NMR and LC-MS based metabolomics. No seasonal information was available for these seeds.

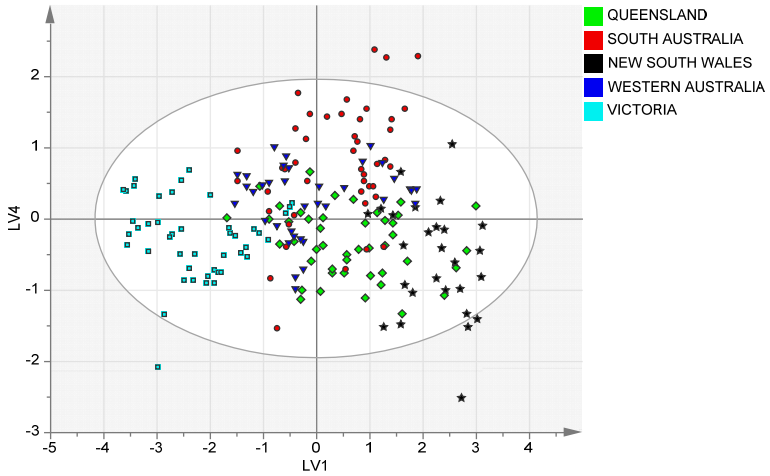
Specimen	Location
"*noori dehradun*"	India
"*black diamond 1*"	India
"*zibo 108*"	China
"*zibo 2*"	China
"*kranti quetta*"	Pakistan
"*sarahHybrid*"	Paraguay
"*lyra hybrid*"	Brazil
"*bangkok brown*"	Phillipines
"*lamao red*"	Phillipines
"*impala*"	Tanzania
"*zanzibariensis*"	Kenya
"*zanzibariensis*"	Tanzania
"*dehradun*"	India
"*gibsonii*"	Zimbabwe
"*impala*"	Tanzania
"*sanguineus*"	Spain
"*sanguineus*"	Tanzania
"*carmencita*"	Tanzania

Appendix B: LA-ICPMS scores plots

Other projections of the ICPMS data. (a) Scores plots of LV1 vs. LV3; (b) LV1 vs. LV4.



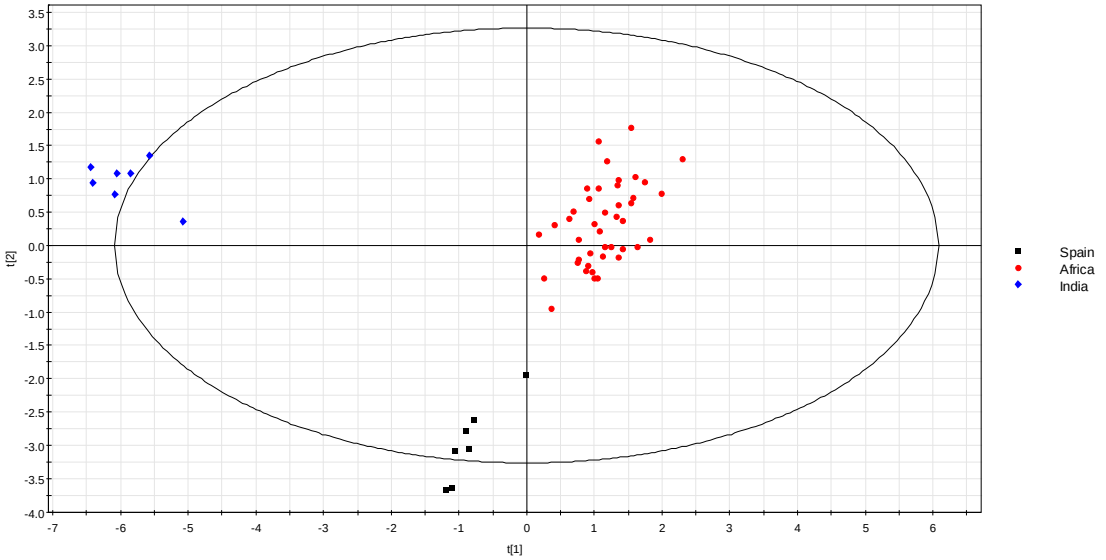
(a)



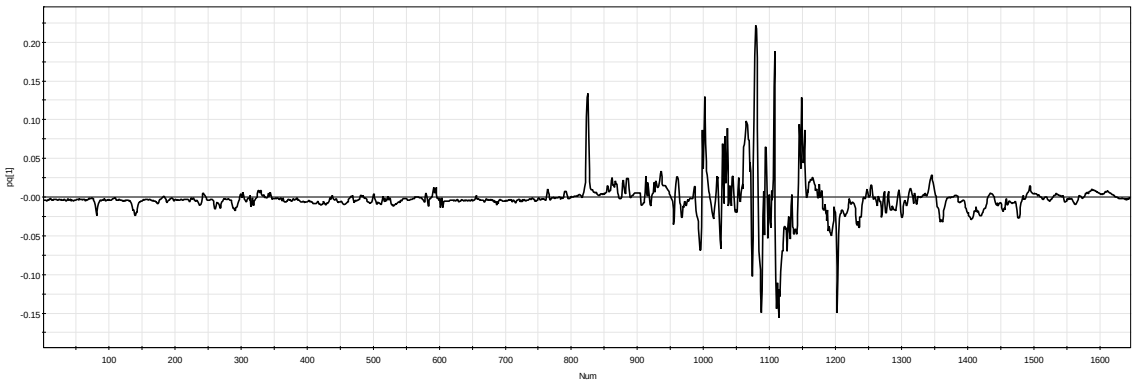
(b)

Appendix C: Supporting data for Study 1

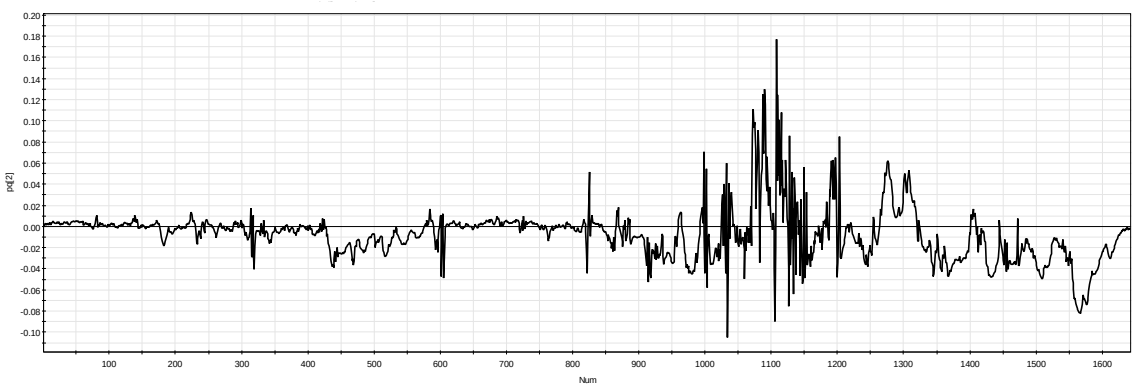
(a) OPLS-DA model scores plot classifying specimens according to their continent of origin; (NB: Spain and India are labelled by their country name rather than continent as there was only one specimen from their corresponding continents); (b) Loadings plot of LV1; (c) Loadings plot of LV2; (d) OPLS-DA scores plot of the first three latent variables (3D) separating African specimens according to their country of origin; (e) Loadings plot of LV1; (f) Loadings plot of LV2.



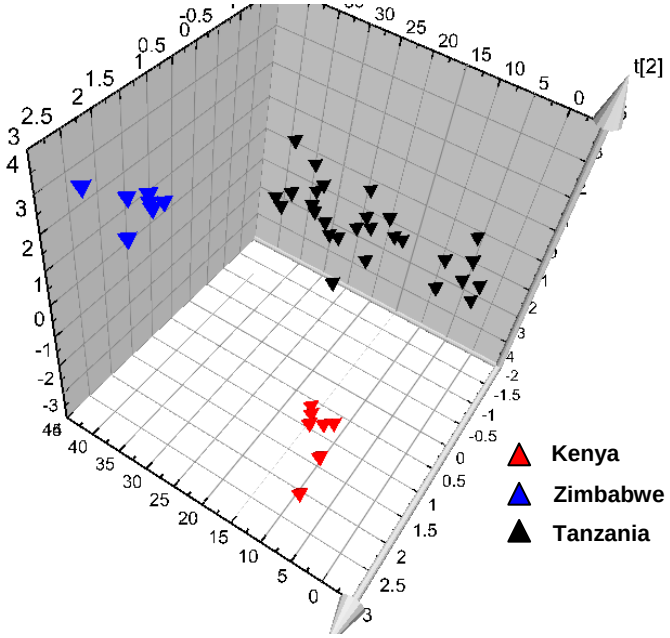
(a)



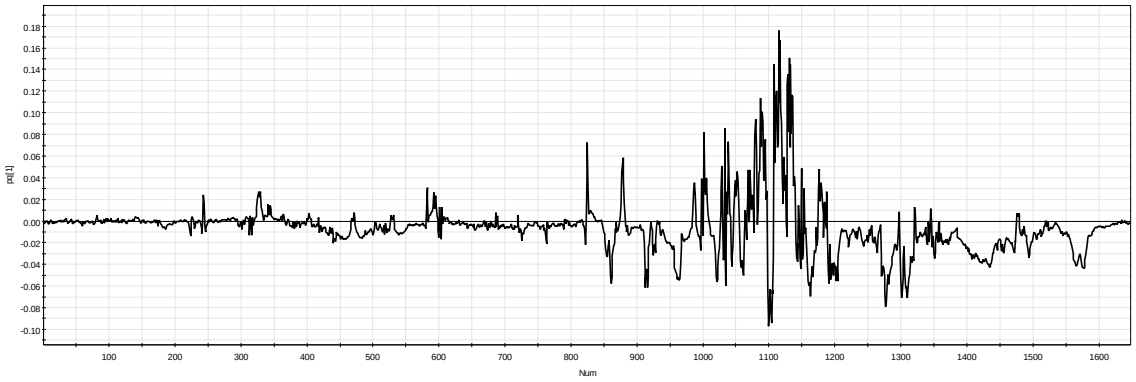
(b)



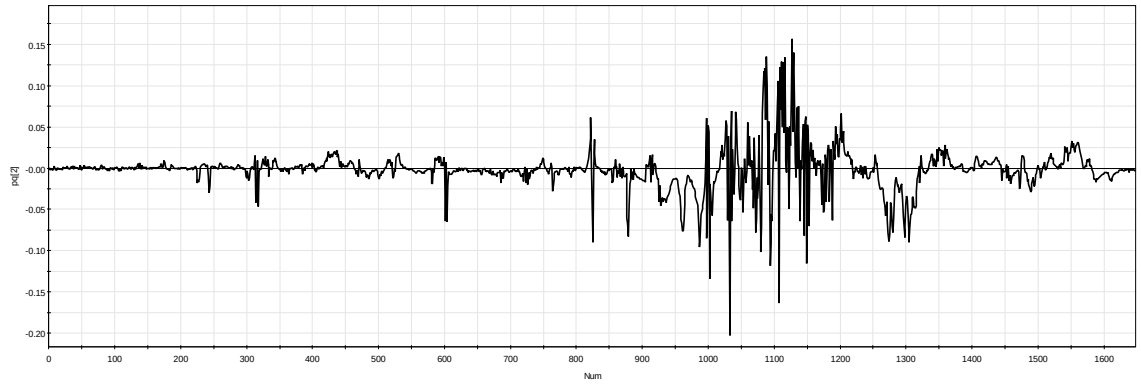
(c)



(d)

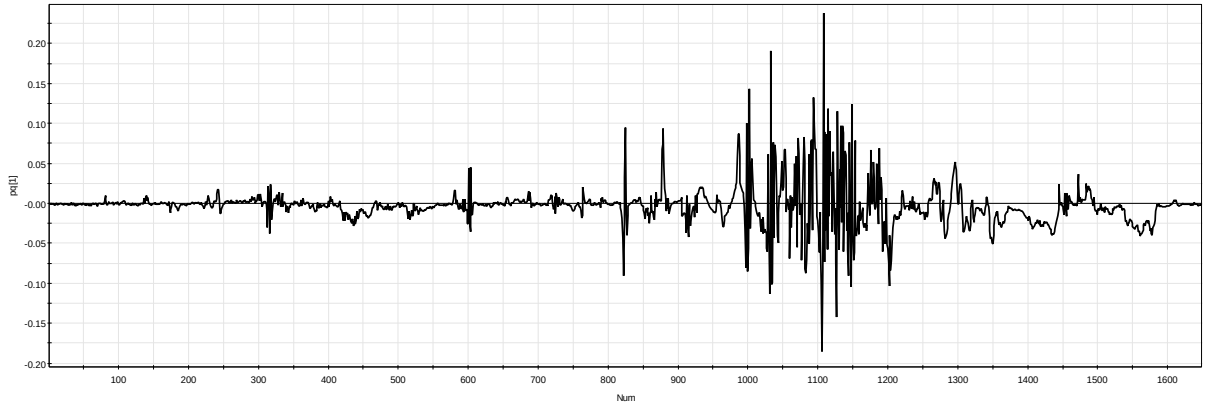


(e)

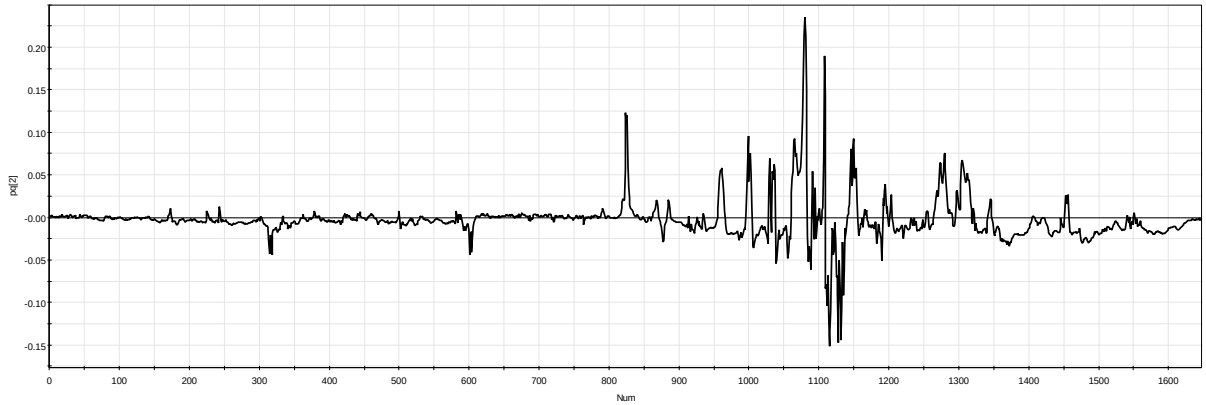


(f)

Loadings plots for cultivar classification OPLS-DA model generated in Figure 7a and 8b. (g) Loadings on LV1; (h) Loadings on LV2.



(g)



(h)

Table 1: (a) Prediction table according to continent of origin correctly identifying all blinded samples; (b) Prediction table according to African specimens correctly identifying the "*gibsonii*" Zimbabwe blinded samples.

Obs ID	Spain	Africa	India
BS1 (DI)	0.33	-0.17	0.84
BS2 (GZ)	0.05	1.02	-0.07
BS4 (DI)	0.24	-0.23	0.99
BS5 (GZ)	-0.20	1.13	0.07
BS7 (DI)	-0.31	-0.03	1.34
BS8 (GZ)	0.25	0.72	0.03

(a)

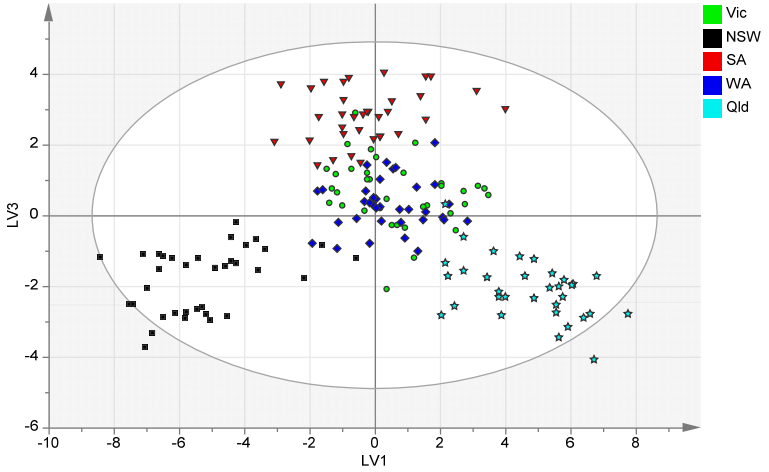
Obs ID	ST	ZK	ZT	CT	IT	GZ
BS2 (GZ)	0.21	0.07	-0.01	-0.12	0.01	0.84
BS5 (GZ)	-0.23	-0.19	0.26	0.18	0.21	0.76
BS8 (GZ)	-0.05	-0.20	0.22	0.02	0.05	0.97

(b)

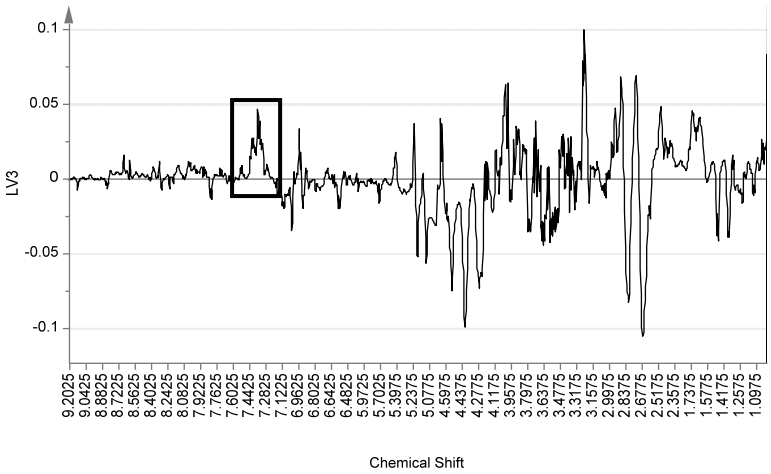
SS: "*sanguineus*" Spain; ST: "*sanguineus*" Tanzania; ZK: "*zanzibariensis*" Kenya; ZT: "*zanzibariensis*" Tanzania; CT: "*carmencita*" Tanzania; IT: "*impala*" Tanzania; DI: "*dehradun*" India; GZ: "*gibsonii*" Zimbabwe

Appendix D: Supporting data for Study 3

(a) Scores plot of LV1 vs. LV3; (b) Corresponding loadings line plot of LV3. Black box highlights aromatic ^{1}H NMR resonances of phenylalanine.

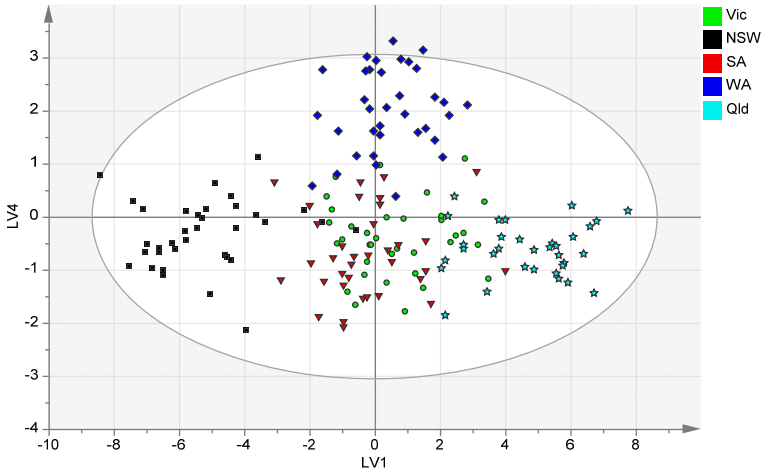


(a)

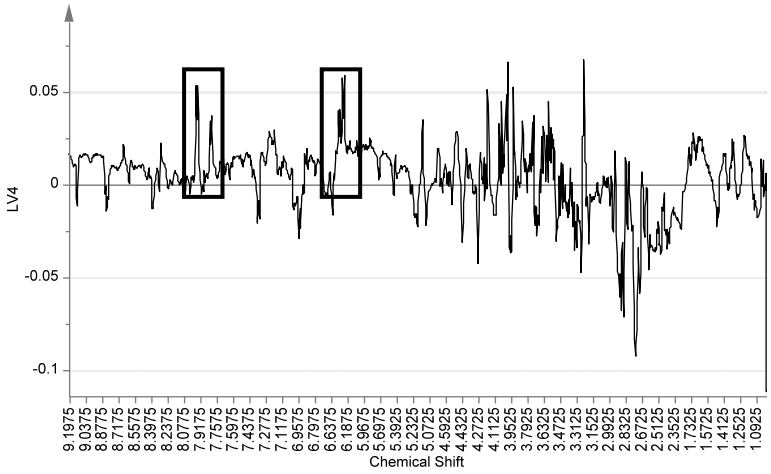


(b)

(c) Scores plot of LV1 vs. LV4; (d) Corresponding loadings line plot of LV4. Black box highlights ^{1}H NMR resonances of ricinine and *O*- and *N*- demethyl ricinine.



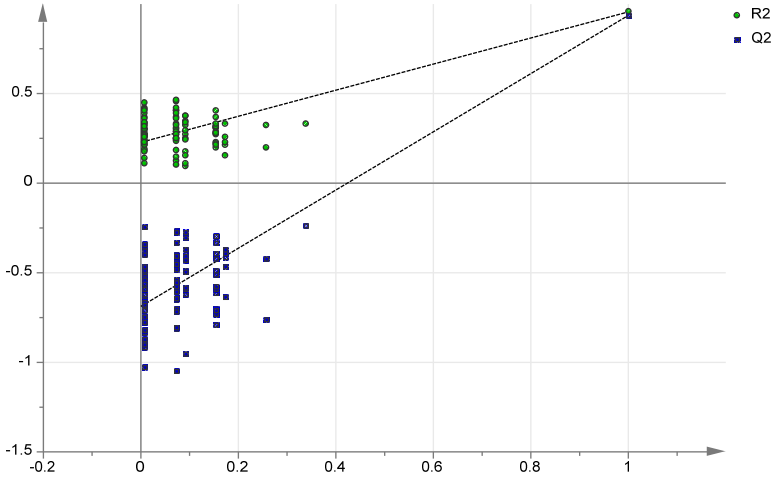
(c)



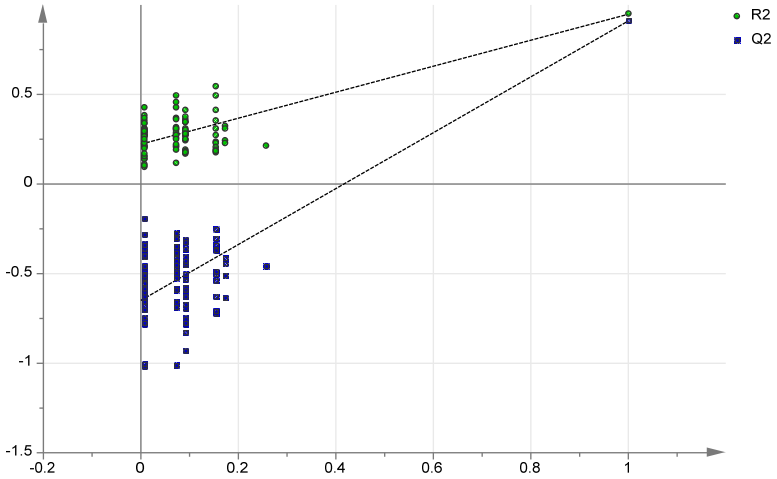
(d)

Appendix E: Supplementary PCA, PLS-DA & OPLS-DA analysis of LCMS data

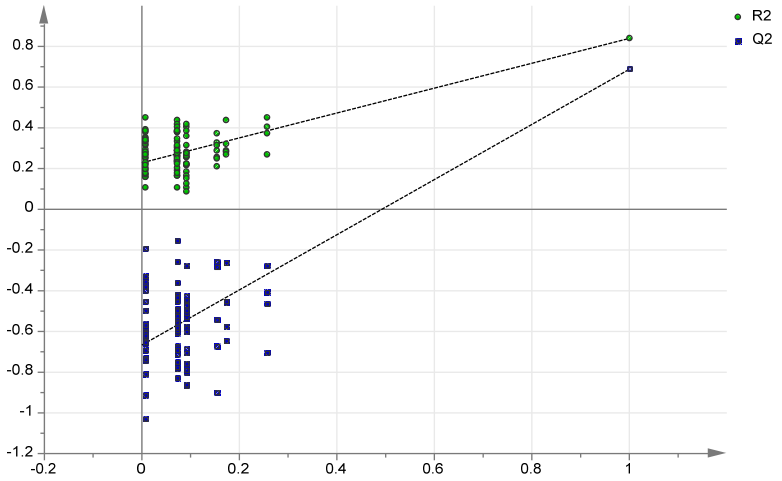
Permutation testing (100 rounds) on the PLS-DA model of the reduced data matrix. (a) "*carmencita*"; (b) "*dehradun*"; (c) "*gibsonii*"; (d) "*impala*"; (e) "*sanguineus*"; (f) "*zanzibariensis*".



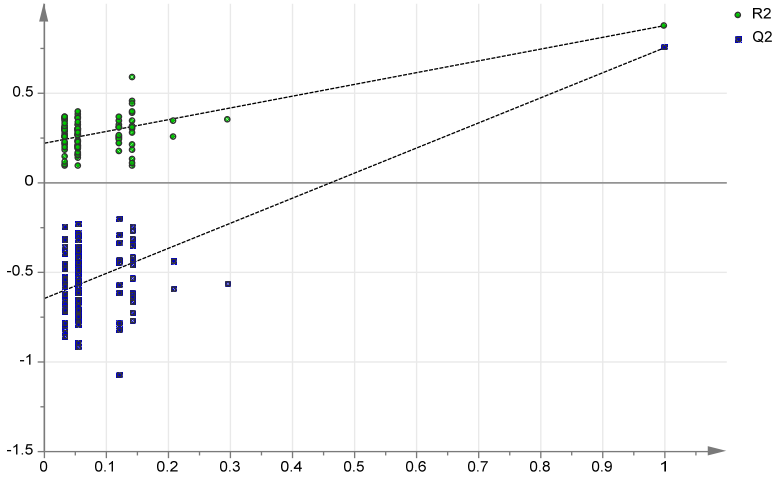
(a)



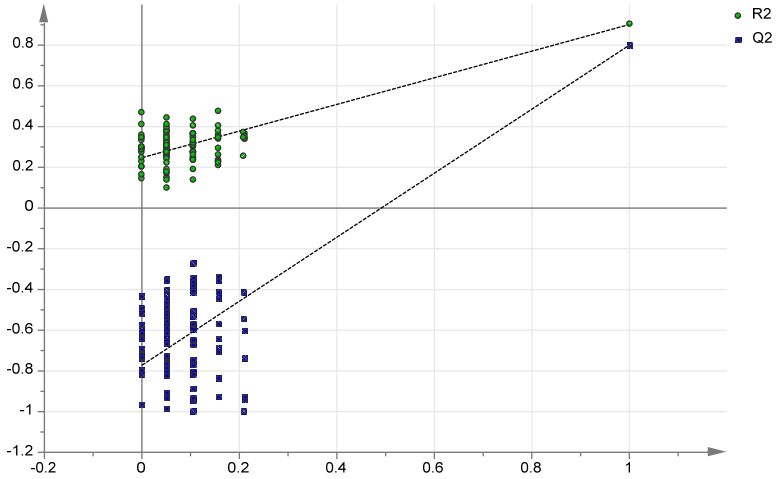
(b)


R2
Q2
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-0.2
0
0.2
0.4
0.6
0.8
1

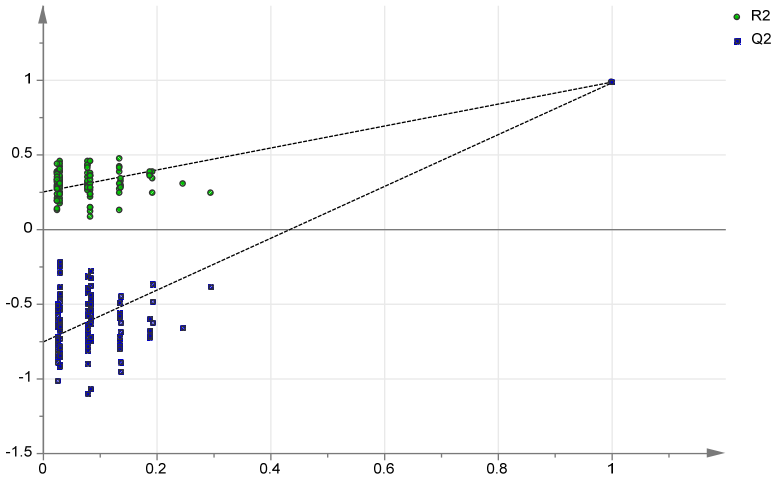
(c)


R2
Q2
0.5
0
-0.5
-1
-1.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

(d)

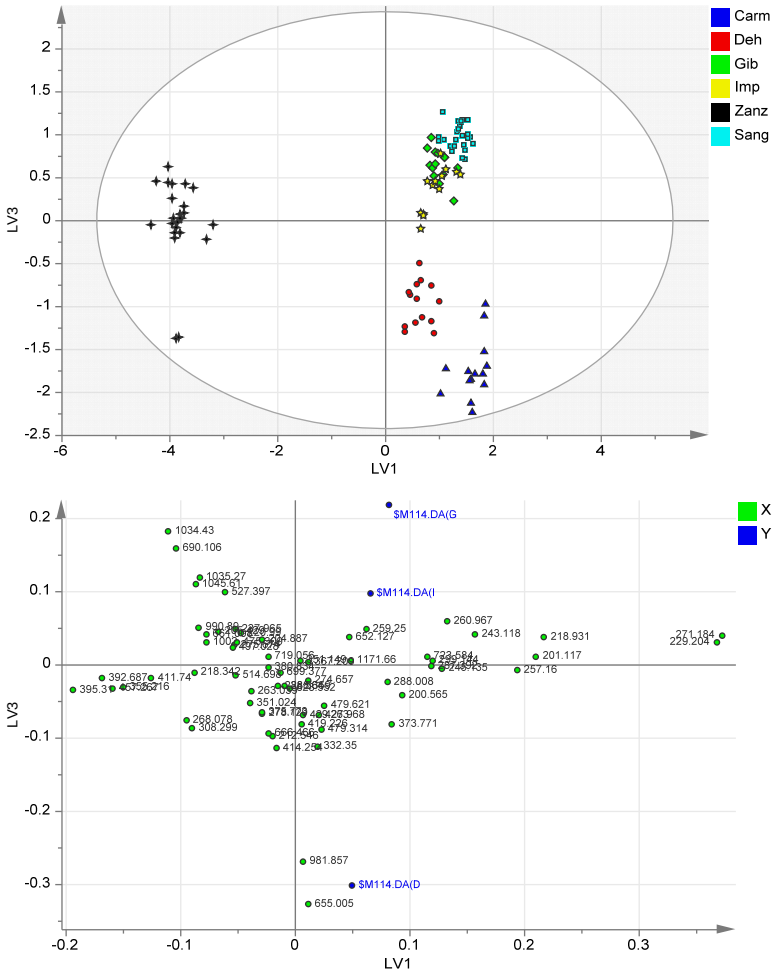

R2
Q2
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-0.2
0
0.2
0.4
0.6
0.8
1

(e)

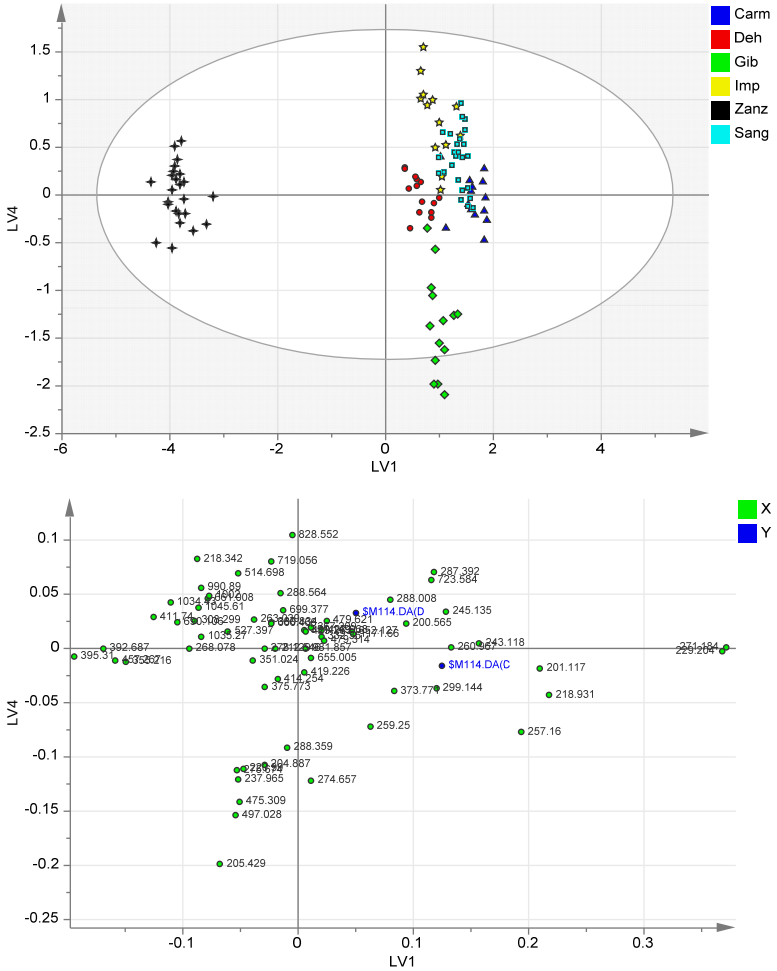


(f)

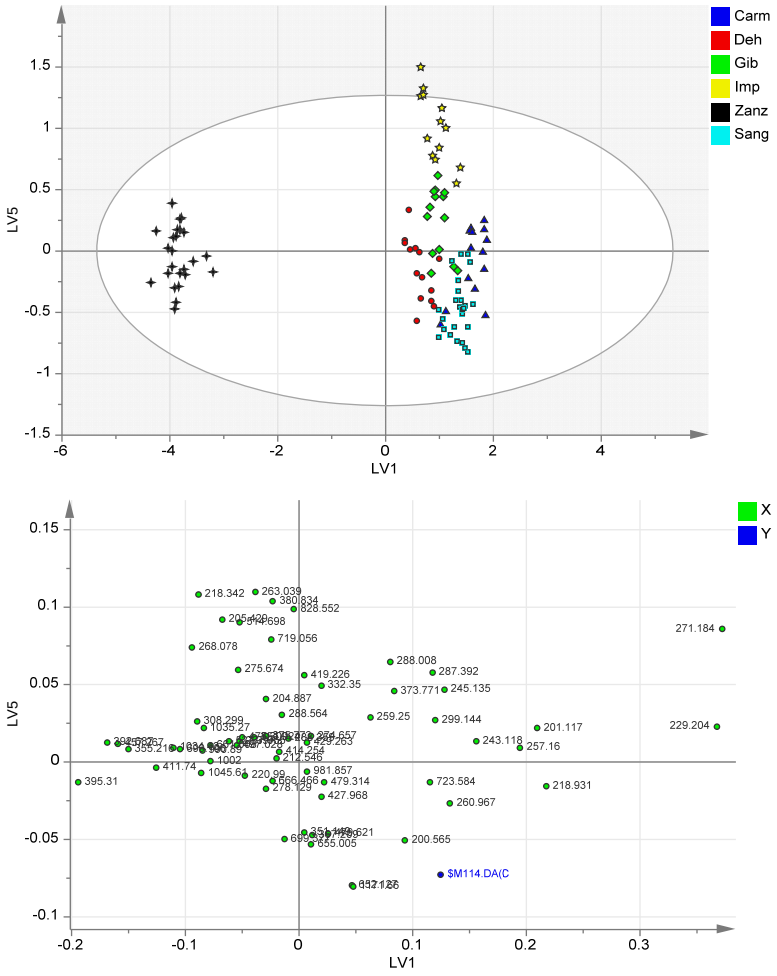
Scores and loadings scatter plots from the reduced data matrix. (g) LV1 vs. LV3; (h) LV1 vs. LV4; (i) LV1 vs. LV5.



(g)


Carm
Deh
Gib
Imp
Zanz
Sang
LV4
LV1
X
Y
828.552
218.342
719.056
514.698
287.392
723.584
288.564
288.008
$M114.DA(D
245.135
200.565
699.377
$M114.DA(C
201.117
218.931
299.144
259.25
257.16
288.359
274.657
475.309
497.028
205.429

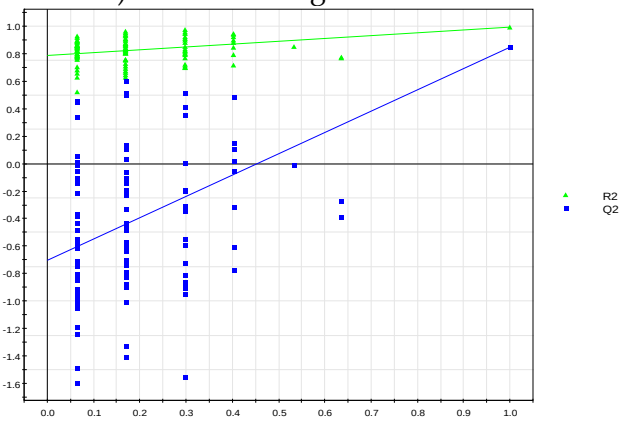
(h)


Carm
Deh
Gib
Imp
Zanz
Sang
LV5
LV1
1.5
1
0.5
0
-0.5
-1
-1.5
-6
-4
-2
0
2
4
X
Y
0.15
0.1
0.05
0
-0.05
-0.1
-0.2
-0.1
0
0.1
0.2
0.3
218.342
263.039
380.834
828.552
719.056
268.078
275.674
419.226
332.35
288.008
287.392
373.771
245.135
204.887
288.564
259.25
299.144
201.117
271.184
229.204
243.118
257.16
1002
411.74
981.857
395.31
723.584
218.931
479.314
427.968
260.967
200.565
655.005
$M114.DA(C

(i)

Appendix F: "zibo 108" greenhouse ^{1}H NMR permutation test

Permutation testing (100 times) results of the greenhouse "*zibo 108*" PLS-DA model.



DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION
DOCUMENT CONTROL DATA

1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)

2. TITLE

Chemical Investigations of the Castor Bean Plant *Ricinus communis*

3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION)

Document	(U)
Title	(U)
Abstract	(U)

4. AUTHOR(S)

Simon P. B. Ovenden, Christina K. Bagas, David J. Bourne, Eloise J. Pigott and Warren Roberts

5. CORPORATE AUTHOR

DSTO Defence Science and Technology Organisation
506 Lorimer St
Fishermans Bend Victoria 3207 Australia

6a. DSTO NUMBER	6b. AR NUMBER	6c. TYPE OF REPORT	7. DOCUMENT DATE
DSTO-TR-2786	AR-015-479	Technical Report	December 2012

8. FILE NUMBER	9. TASK NUMBER	10. TASK SPONSOR	11. NO. OF PAGES	12. NO. OF REFERENCES
2012/1093030	NS 08/0003	AFP, PM&C	58	19

13. DSTO Publications Repository

http://dspace.dsto.defence.gov.au/dspace/

14. RELEASE AUTHORITY

Chief, Human Protection and Performance Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111

16. DELIBERATE ANNOUNCEMENT

No limitations

17. CITATION IN OTHER DOCUMENTS Yes

18. DSTO RESEARCH LIBRARY THESAURUS

Chemical analysis, Chromatography, Research, Spectrometry, Spectroscopy

19. ABSTRACT

In 2009 a National Security Science and Technology grant was awarded to the Human Protection and Performance Division for the investigation of several forensic aspects of the castor bean plant *Ricinus communis*. A major focus of this grant was to understand the chemical composition of the seeds, and to ascertain if these differences could be used for provenance classification. This technical report will discuss progress made during these investigations.