Australian Government
Department of Defence
Defence Science and
Technology Organisation

Cultivar Determination of *Ricinus communis* via the Metabolome: a Proof of Concept Investigation

Simon P. B. Ovenden, Benjamin R. Gordon[1], Christina K. Bagas, Bob Muir, Simone Rochfort[2] and David J. Bourne

Human Protection and Performance Division
Defence Science and Technology Organisation

[1]La Trobe University

[2]Department of Primary Industries – Victoria

DSTO-TR-2338

ABSTRACT

Investigations were undertaken to ascertain the appropriateness of studying the metabolome of *Ricinus communis* for cultivar and provenance determination. Seeds from fourteen *R. communis* specimens (a total of 56 seeds) collected from the east coast of Australia were analysed by various analytical chemistry methods. The data collected from these investigations were then analysed using Principal Component Analysis. The outcomes from these investigations are discussed in this technical report.

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-014-609
August 2009

APPROVED FOR PUBLIC RELEASE

Cultivar Determination of *Ricinus communis* via the Metabolome: a Proof of Concept Investigation

Executive Summary

Ricinus communis (commonly known as the castor bean plant) is an introduced species of plant that now grows wild in Australia, with some 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 forensic agencies. There are many analytical strategies that are available to be applied. One such strategy worth consideration was metabolomics. Metabolomics is the study of the metabolome of an organism. The metabolome can be defined as the pool of extractable chemistry produced by an organism, through the interaction between the organisms' genome and the environment. To this end, a proof of concept study was undertaken to investigate the appropriateness of studying the metabolome of *R. communis* seeds for cultivar and provenance determination. Subsequently, fourteen *R. communis* specimens (a total of 56 seeds) collected from the east coast of Australia were analysed by High Pressure Liquid Chromatography-Ultra Violet (HPLC-UV), Liquid Chromatography-Mass Spectrometry (LC-MS) and ^{1}H Nuclear Magnetic Resonance (NMR) spectroscopy. The data collected from these analyses were then further analysed using Principal Component Analysis (PCA). For HPLC-UV analysis, the seed extract from seven *R. communis* specimens were unambiguously identified by PCA as belonging to separate classes relating to specimen. LC-MS data allowed unique ions to be identified for five specimens. Conversely ten specimens were unambiguously segregated in the PCA of the ^{1}H NMR data. Furthermore, the ratio between the known biomarker ricinine, and two demethylricinine analogues, was found to be important for specimen determination. These combined analyses suggested that a combination of HPLC-UV, LC-MS and ^{1}H NMR in conjunction with PCA could allow for specimen differentiation to be made. Arguments allowing for these conclusions to be made are discussed in detail in this technical report.

Authors

Simon P. B. Ovenden
Human Protection and Performance Division

Simon Ovenden 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 in Analysis and Verification, where he is applying his background in Nuclear Magnetic Resonance (NMR) spectroscopy and mass spectrometry. Simon currently leads the metabolomics research program on R. communis at DSTO.

Benjamin R. Gordon
La Trobe University

Ben Gordon graduated with a BChemSc(Hons) in 2008 in collaboration with DSTO. His honours thesis formed part of this tech report and was titled Metabolomic Studies of Ricinus communis For Forensic Applications. In 2009 he joined the Australian Institute of Marine Science (AIMS) as a chemist for the sustainable supply of bioresources. His work involves extraction and isolation chemistry to produce pure compounds from target organisms for external biodiscovery clients.

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 Analysis and Verification. In 2008 she completed a DipEd from La Trobe University. Currently she is working in the metabolomics research program on R. communis at DSTO.

Bob Muir
Human Protection and Performance Division

Bob Muir completed a BSc (Hons) in chemistry in 1995 and a PhD in environmental/analytical chemistry in 1998. Bob has engaged in research in the field of chemical warfare agent detection and analysis from 1998 to present, working for both the UK and Australian Governments.

Simone Rochfort
Department of Primary Industries – Victoria

Simone Rochfort completed her PhD in marine natural products chemistry at the University of Melbourne in 1996. After a postdoctoral fellowship with Dr Jeffrey Wright, National Research Council of Canada, she returned to Australia to take up a research position with AstraZeneca and Griffith University where she worked on the discovery of natural products for human pharmaceuticals. Dr Rochfort's research in the pharmaceutical and biotech industries continued until 2004 when she joined the Victorian Department of Primary Industries. She is currently employed as a Principal Research Scientist and applies her natural products research interests to metabolomics.

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. Currently David is the Task Leader of the divisional DIO program with additional Group Leader responsibilities for chemical analysis and verification

---

Contents

1. INTRODUCTION 1

2. RESULTS AND DISCUSSION 4
2.1 HPLC-UV Results and Discussion 4
2.1.1 *R. communis* seed selection and extraction 5
2.1.2 HPLC method development and instrument stability 6
2.1.3 HPLC-UV data analysis and PCA 8
2.1.3.1 PC1 vs. PC2 9
2.1.3.2 PC1 vs. PC3 12
2.1.3.3 PC3 vs. PC4 13
2.1.3.4 PC4 vs. PC5 14
2.1.4 Summary of outcomes from HPLC-UV and LC-MS analysis 16
2.2 ^{1}H NMR spectroscopy and PCA of *R. communis* seed extracts 17
2.2.1 Sample preparation and analysis 17
2.2.2 NMR statistical analysis 18
2.2.2.1 Model One 18
2.2.2.2 Model Two 19
2.2.2.3 Model Three 20
2.2.3 Summary of the ^{1}H NMR analysis 27

3. CONCLUSIONS 27

4. EXPERIMENTAL 29
4.1 Chemicals 29
4.2 General Experimental 29
4.3 Collection and extraction of *R. communis* seed specimens 29
4.4 HPLC-UV data collection and multivariate statistical analysis 30
4.5 NMR sample preparation and data collection 30
4.6 NMR multivariate statistical analysis 31

5. ACKNOWLEDGMENTS 31

6. REFERENCES 31

APPENDIX A: SEED IMAGE AND LOCATION 37

APPENDIX B: HPLC-UV BIN NUMBER AND PC 39

APPENDIX C: EXTRACTED MASS SPECTRA 41

APPENDIX D: TYPICAL ^{1}H NMR SPECTRUM 43

APPENDIX E: ^{1}H NMR STACK PLOTS 45

1. Introduction

Ricinus communis, more commonly known as the castor bean plant, is indigenous to Eastern Africa,[1] parts of east Asia and South America[2] and has been in cultivation for four thousand years.[3] The plant is grown for the seeds, which produce up to 60% castor oil by weight. Castor oil is a basic constituent in a variety of industries, including aviation, hydraulic fluid, engine lubricant and paint medium.[4] It is also used as a healing agent in many folk medicine remedies as a purgative.[4] More recently castor oil has been used in exclusive organic hair care treatments.[5] Annual world production of castor oil is in excess of one million tonnes, with the primary producers being China and India.[6,7]

R. communis has also been widely admired as a garden ornamental and was prevalent in Australian homes during the 1960's due in part to their striking spike colour and distinctive leaves (Figure 1). However, as a consequence of *R. communis* being a prolific producer of seeds, these garden specimens now grow wild in many geographic locations within Australia. Indeed in some states *R. communis* has been declared a noxious weed. The combined drivers of developing cultivars for industrial castor oil production and for garden ornamentals have led to some 250 cultivars being available.[8] Consequently, there are a striking array of differences (including height, leaf size, shape and colour, stem colour, seed size and colour) between cultivars and they often bear little resemblance to each other.[9]

Figure 1: Images of three different specimens of R. communis

The seeds from *R. communis* not only contain castor oil but also the toxic plant lectin ricin. Ricin is declared by the Chemical Weapons Convention (CWC) as a Schedule 1 agent,[10] which are chemicals that are highly toxic and have no legitimate uses.[11] Ricin is also listed as the second highest priority on the list for terrorism agents by the United States Centers for Disease Control.[12,13] Ricin has an LD_{50} of approximately 2 µg/kg in standard mouse models,[14] and is thought to have a human LD_{50} of 3 – 30 µg/kg.[15] Ricin is a heterodimeric type II ribosome-inactivating protein (RIP).[16] There are numerous naturally occurring RIP toxins found in both plants and microbes.[16] They are defined by their N-glycosidase activity, which selectively depurinates adenine within a highly conserved fourteen nucleotide region of the 28S rRNA subunit of the large 60S ribosome.[17-19] This results in the inhibition of protein manufacture within the cell, preventing chain elongation of polypeptides and leading to apoptosis.[15,20]

Ricin is made of two approximately 32 kDa proteins (the A-chain and B-chain, see Figure 2)[21] linked by a disulfide bond. The A-chain (also known as the N-glycosidase enzyme) is responsible for exerting the toxicity of ricin. The B-chain (also known as the lectin) facilitates the entry of ricin into the cytosol by attaching to glycolipids and glycoproteins on the surface of the cell.[3]

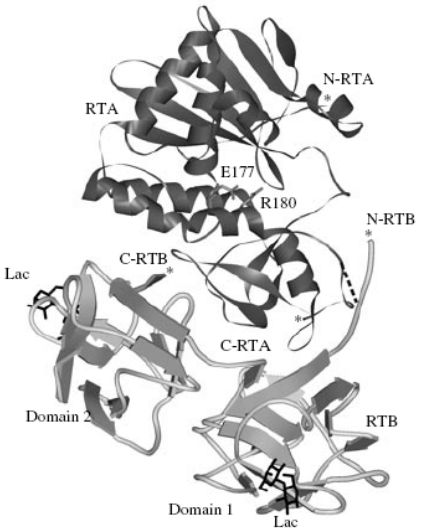


Figure 2: Scheme of the 3D of ricin (Protein Data Bank ID 2AAI). The disulfide bond between RTA (dark gray) and RTB is shown with a dashed line. Lactose molecules (Lac) bound in the galactose-binding sites of RTB are in black. Positions of the side chains of Glu177 (E177) and Arg180 (R180), which are important for the catalytic action of RTA, are indicated. The N and C termini of RTA and RTB (N/C-RTA, N/C-RTB) are asterisked.

Ricin was thought to have been used in the assassination of the Bulgarian dissident Georgi Markov.[22] A pellet allegedly impregnated with ricin was attached via a spring loaded mechanism to the tip of an umbrella. The action of poking the umbrella into Markov's thigh successfully injected him with approximately 0.5 mg of ricin, ultimately causing his death.[22,23] Ricin has also been implicated in several recent incidents, which continue to highlight the risk associated with its use. In 2003 a package containing ricin was discovered in a South Carolina postal centre with a note threatening to poison water supplies if certain demands were not met (Figure 3).[24,25] While in 2004 a "white powder" incident in the US Senate office in Washington subsequently prompted the Department of Health and Human Services (USA) to develop the need for a ricin-specific response protocol.[26]

In March 2008 a grand jury indicted Roger Von Bergendorff on ricin possession charges. A search of his room found castor beans in addition to a copy of the book "Anarchist's Cookbook" and a collection of instructions on poisons and other dangerous recipes. Authorities were alerted to his activities after he poisoned himself with ricin.[28-30] There have also been documented cases of ricin having been used in suicide and suicide attempts.[31]

Figure 3: Envelope and letter implicated with ricin in the 2003 South Carolina postal facility incident[27]

The prevalence of *R. communis* in the environment, the ease of seed collection, the toxicity of ricin, and its Schedule 1 status, necessitate that domestic and international law enforcement agencies have the ability to determine cultivar and provenance of a seed extract.[32] There are several analytical techniques available that can specifically determine provenance, however they require dedicated equipment and trained operators.[33-35] This makes their implementation as a general analytical method less appealing. An alternative approach is to study the total pool of extractable chemistry from a seed of *R. communis* (the metabolome), and analyse the results via multivariate statistical analysis (chemometrics). Commonly known as metabolomics (or metabonomics), it can be clearly defined as a method that seeks to identify and quantify the complete set of metabolites in a cell or tissue type quickly without bias.[36,37] The study of the metabolome found in an organism is the result of the interaction between the organisms' genome and the environment. Metabolomics has been applied to commercial (wheat, olive, wines) and forensic (cannabis) plant crops and has enabled either or both of cultivar and provenance to be determined.[38-41] Advantageously, studying the metabolome relies on the data generated from standard laboratory analytical equipment such as High Pressure Liquid Chromatography-Ultra Violet (HPLC-UV), Liquid Chromatography-Mass Spectrometry (LC-MS), and to a lesser extent Nuclear Magnetic Resonance (NMR) spectroscopy.

In an effort to prove the concept that metabolomic methodologies can be applied to *R. communis* seeds to determine cultivar, an investigation into the seed metabolome was undertaken. The aim was to use data from the standard analytical equipment documented above and analyse the results via chemometrics for cultivar identification. Discussed in this technical report are the results obtained from these initial analytical investigations into the metabolome of *R. communis* seeds, and future investigations that will be conducted.

2. Results and Discussion

2.1 HPLC-UV Results and Discussion

HPLC-UV data has been employed to study the metabolome of red wines for cultivar and provenance.[41] Generally though it has been avoided for many reasons, including complexities surrounding peak alignment of UV chromatograms.[42] HPLC-UV has in the main been restricted to "targeted applications", where the presence/absence of a specific compound, or a group of compounds has been monitored for.[43-50]

From a metabolomics perspective, it is desirable to analyse an extract of biological material in a "non-targeted" manner. This approach allows for the metabolome to be analysed without bias towards a particular structure class. While there are complexities surrounding peak alignment of UV chromatograms, it is crucial that they are aligned so that the apexes of common peaks match. Any error in peak alignment will directly impact on the outcomes generated from multivariate statistical analysis. There is literature precedence for performing alignment of UV data using techniques such as Correlation Optimised Warping (COW) and Dynamic Time Warping (DTW),[51] however they require dedicated software.

There are advantages in using HPLC-UV for metabolome analysis. Compared to LC-MS, there is no potential for signal suppression due to sample and/or eluant matrix effects. This makes it an excellent analytical technique for the analysis of complex mixtures. Additionally, compounds that have a UV chromophore but will not ionise in a mass spectrometer can be identified. Data collection via a Photo Diode Array (PDA) detector is very powerful. Used diligently, it can be utilised to distinguish between classes of compounds present in the metabolome.

Considering this, the application of HPLC-UV for the analysis of the metabolome of *R. communis* seeds for cultivar determination was evaluated. Given the nature of the data collected for each specimen extract (multiple variables of retention time and peak area per injection per sample), a statistical classification technique is ideally suited to find patterns in the resulting data. As such, principal component analysis (PCA) was selected for this study. PCA is a dimension reducing technique that is applied to large data matrices. The resulting derived principal components (PC) explain the variance of the original data matrix in simple linear combinations of descriptive variables. For successful PCA, the bulk of the data variance should be explained with values of 70% or greater generally considered fit for purpose.[52]

Complex chromatographic data sets on biological systems such as the case presented here, by their very nature will contain differences in chromatogram composition. Therefore, simply overlaying chromatograms from different specimens, while informative, will reveal gaps in one or more of the chromatograms. This makes accurate retention time alignment of peaks difficult and usually results in missing values in the generated data matrix. Various techniques have been used to reduce or eliminate such missing values. Approaches include substituting zeros with minimum non-zero values[53] and use of chromatographic warping.[51] Alternative variations on PCA have also been successfully applied such as non-linear PCA[54] and iterative replacement of missing values.[55]

For the analysis of determining cultivar of *R. communis* extracts, the novel approach of binning the HPLC-UV chromatographic data was investigated. Each bin had a defined retention time width, with the area under the chromatogram for each bin determined. Thus every bin contained some portion of the chromatogram that could be integrated, completely removing missing values from the data matrix. This approach makes use of the known biomarker ricinine as an internal standard, which is ubiquitous to all extracts of mature *R. communis* seeds (Figure 4).[56,57] Each chromatogram was referenced to the ricinine retention time (6.6 min), binned using a Microsoft Excel macro developed in-house,[58] then subjected to PCA. Bins were then identified as being critical for discrimination between specimens. The corresponding LC-MS data for these critical bins were then interrogated with the aim to identify ions within that bin that were unique to that particular specimen.

Discussed below is the rationale for seed selection and extraction, HPLC-UV method development, data analysis and PCA for UV chromatographic data, and identified ions of importance for specimen determination. Also, the observed seed-to-seed biological variation of extracts from seeds of a common *R. communis* specimen is discussed.

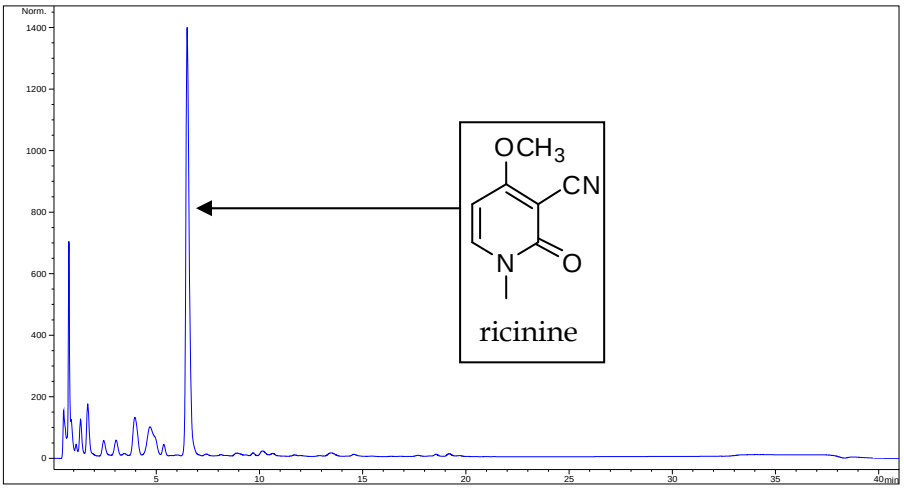


Figure 4: A standard chromatogram at 254 nm for a <30 kDa MWCO fraction of R. communis seed extract highlighting the ricinine peak. Ricinine has a strong UV absorbance at 254 nm, and a consistent retention time of 6.6 min.

2.1.1 *R. communis* seed selection and extraction

Analysing a samples metabolome provides a snapshot of an organisms metabolism at a particular point in time.[59] Consequently, for the preparation of a sample for metabolomics analysis, it is important that there is no bias towards certain compound classes. As such, the development of a reliable sample preparation method is essential to reduce analytical variability, increase the robustness of the data and allow for reliable measurement of biological variability.

Currently, the DSTO seed library of Australian specimens of *R. communis* is made up of collections from unknown cultivars. Consequently, specimens selected for this analysis were

made with respect to the morphology of the host plant. The rationale behind this was that plants of significantly different morphology were anticipated to be different cultivars. Four individual seeds from fourteen different specimens were analysed for a total of 56 seeds. This was so the seed-to-seed biological variation prevalent within seeds from the one specimen could be analysed. A table of seed images, collection code numbers and corresponding geographic location is shown in Appendix A.

Each seed was treated according to a "terrorist textbook" method of preparation. Seeds were crushed and the castor oil removed with acetone. The acetone extract was filtered, and the residual mash treated with 2% acetic acid to yield a crude ricin extract. This crude ricin extract was then treated with a 30 kDa Molecular Weight Cut Off (MWCO) filter to remove both the ricin and *Ricinus communis* agglutinin (RCA), making the extracts safer to handle. Extracts were then analysed using HPLC-UV, LC-MS and ^{1}H NMR, with the collected data analysed using PCA.

The seed extract of *R. communis* was a very complex matrix, with hundreds of compounds present. To obtain reliable HPLC-UV and LC-MS data, four 20 μL injections from a 20 mg/mL solution of each extract were made. In total, 224 separate injections were made and analysed.

2.1.2 HPLC method development and instrument stability

Previous metabolome analysis of *R. communis* seeds conducted at DSTO allowed for an HPLC-UV method to be developed. This method used a 50 mm x 2 mm Phenomenex Luna C18 5 μm HPLC column at a flow rate of 0.4 mL/min, with a linear gradient from 100% H_2O (+ 0.05% formic acid) to 70:30 MeOH:H_2O (+ 0.05% formic acid) over 30 min. The column was flushed with 100% MeOH (+ 0.05% formic acid) for four minutes then re-equilibrated for five minutes with the initial conditions before the next injection was made. This method was evaluated against combinations of other reversed phase methods (MeCN vs. MeOH as the organic phase, formic acid vs. trifluroacetic acid as an acid modifier, C18 vs. phenyl-hexyl HPLC column). However it was found that these conditions generated optimal HPLC-UV and LC-MS data. Column temperature was set at 25°C, and the UV chromatogram at 254 nm was recorded and analysed.

The results that were to be ultimately obtained from PCA of the collated data were clearly reliant on a stable HPLC-UV system. Length of time for system equilibration, injection volume repeatability, peak shape, and retention time stability were seen to be critical parameters. To critique these variables, six repeat injections of 0.5 mg/mL bradykinin (in H_2O) were made. This analysis showed that the injection volume was consistent, as the area under the bradykinin peak was identical (within experimental error), and a consistent peak shape was obtained. The retention time for the bradykinin peak became consistent after the fourth repeat injection. Hence, before daily HPLC-UV analysis eight blank injections were made to ensure a well equilibrated analytical system. To guarantee that no spurious peaks due to the MWCO filters would influence the statistical analysis, blank samples containing 2% acetic acid in water (the same solution used for the extractions) were passed through MWCO filters in the same way as the seed extracts after the removal of castor oil. These blank samples became the source material for the eight blank injections that were made at the beginning of daily analysis.

The bradykinin standard was also injected post equilibration and periodically during daily analysis. This was to ensure that HPLC performance was maintained. Shown in Figure 5a are four injections of bradykinin made during a day's analysis. This figure clearly shows that HPLC performance was maintained during the analyses. Identical observations were made whenever analysis was conducted.

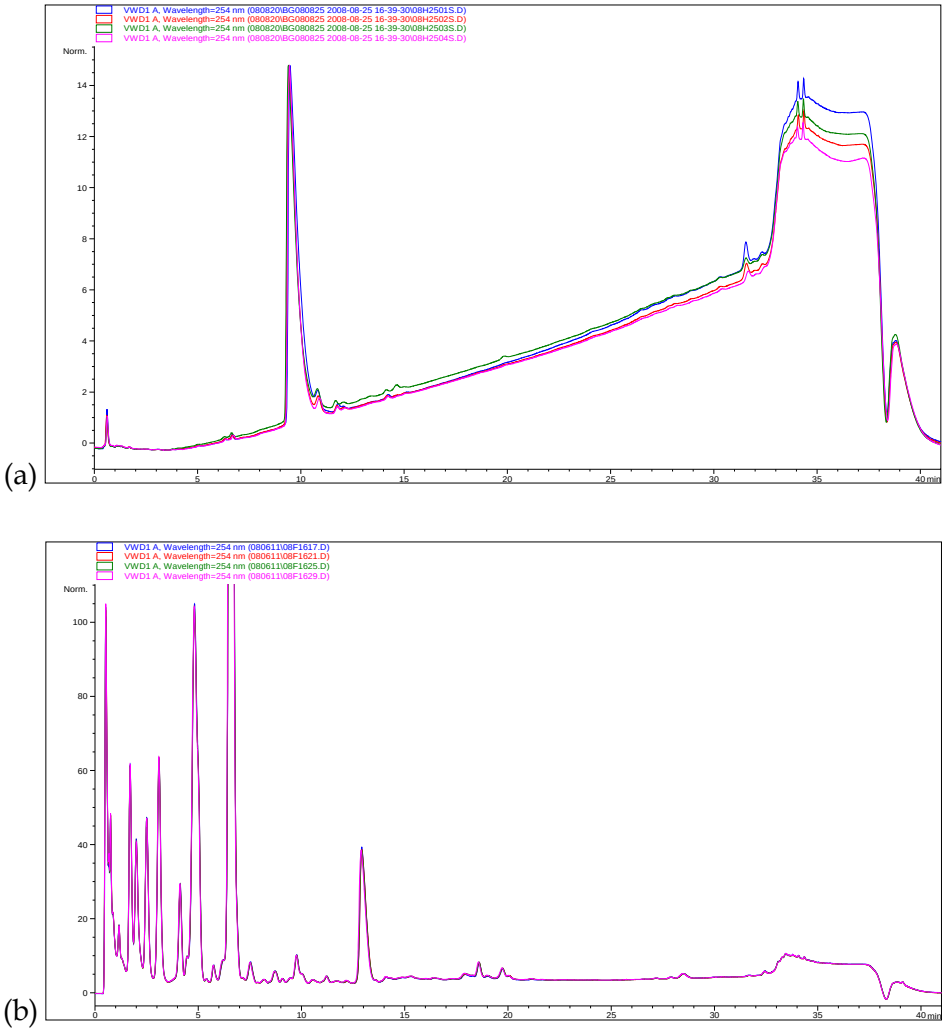


Figure 5: (a) UV chromatograms of four injections of the bradykinin standard made throughout one day of analysis. Minimal retention time drift was observed for bradykinin over the course of a day's analysis; (b) Four repeat injections from one individual seed of the Braybrook specimens. Excellent instrument stability and performance was highlighted through these two observations.

Another condition for HPLC-UV stability is the reproducibility of chromatograms for a complex mixture. For the analysis of *R. communis* seed extracts, instrument stability was excellent as expected. Shown in Figure 5b is a stack plot of four repeat injections of an extract from a Braybrook seed. Outside minor experimental error, the returned UV chromatogram had consistent peak intensity and retention times. Therefore, any observed perturbations in

the UV chromatogram between the metabolome of separate *R. communis* seed extracts were ascribed to inherent differences in the metabolite content.

2.1.3 HPLC-UV data analysis and PCA

Retention time correction to ricinine and binning of the HPLC-UV chromatogram offered some advantages over traditional integration methods. The main advantage was the removal of zero values from the data matrix subjected to PCA. Binning data reduces the statistical impact any minor perturbations in retention time for common compounds will have on PCA. It also allows for an easy way to handle chromatographic data with significant complexity through overlapping peaks. Hence, the 224 HPLC-UV chromatograms were retention time corrected to ricinine, arbitrarily binned into 114 bins of equal time width (approximately 20 s), and imported into Minitab. The data was standardised (involves mean centering, where the variable mean coincides with the origin of the PC), normalised (ensures data vectors are of equal length), and autoscaled (makes each variable possess a mean of zero and unit variance). Subsequent PCA identified six PC's (PC1 35.9%; PC2 16.9%; PC3 16.7%; PC4 8.7%; PC5 5.4%; PC6 4.1%) that accounted for 88% of the data variation of the original data matrix.

With PCs selected, visualisation of the results was performed. This involved plotting the scores for each PC against each other. In other words, a comparison was made of bins (retention times) that were responsible for explaining a particular PC against bins that were responsible for another PC. As each PC accounts for a certain percentage of data variability, each scores plot was able to identify differences and similarities between specimens.

In each scores plot a red dot represents one HPLC-UV analysis of an extract from a seed of a particular *R. communis* specimen. Four red dots grouped together by a black ellipse represents four repeat injections made from the extract of one seed of a particular *R. communis* specimen. A coloured ellipse grouping together four black ellipses indicates all HPLC-UV analyses performed from the four individual seeds of one *R. communis* specimen (sixteen in total). If a coloured ellipse (hence specimen) within a scores plot had on a particular PC axis a score less than -1, or greater than 1, then that PC was interpreted as being significant in explaining the variation of that specimen. The adoption of this value was arbitrary but did simplify the interpretation of the scores and loadings. These were then treated as definitive results for a particular specimen. The subsequent loadings plots were then investigated for each PC. This allowed for an identification of which bins had the most influence on the selected PC. Once identified, the mass spectral data corresponding to these bins were analysed. This was to identify potential specimen specific biomarker compounds.

Apparent in all the scores plots discussed were significant seed-to-seed biological variation for some of the specimens studied. Hence, in this initial study not all of the *R. communis* seed specimen extracts were identified from these scores plots. However, some meaningful results were obtained. Of the fourteen specimens analysed, seven specimens had all their seed extracts accounted for from four scores plots (PC1 vs. PC2; PC1 vs. PC3; PC3 vs. PC4; and PC4 vs. PC5). For the purposes of the following discussion, only these specimens will be discussed in detail, as it was felt that the *intra* specimen seed-to-seed metabolome variation was not as pronounced. Each of the scores plots are addressed and discussed individually in the following sections, with important results highlighted and discussed.

2.1.3.1 *PC1 vs. PC2*

Immediately apparent in the scores plot for PC1 vs. PC2 (Figure 6) was that the four repeat injections from individual seeds for a specific specimen were tightly grouped. This highlighted the excellent stability and performance of the instrument during the analysis.

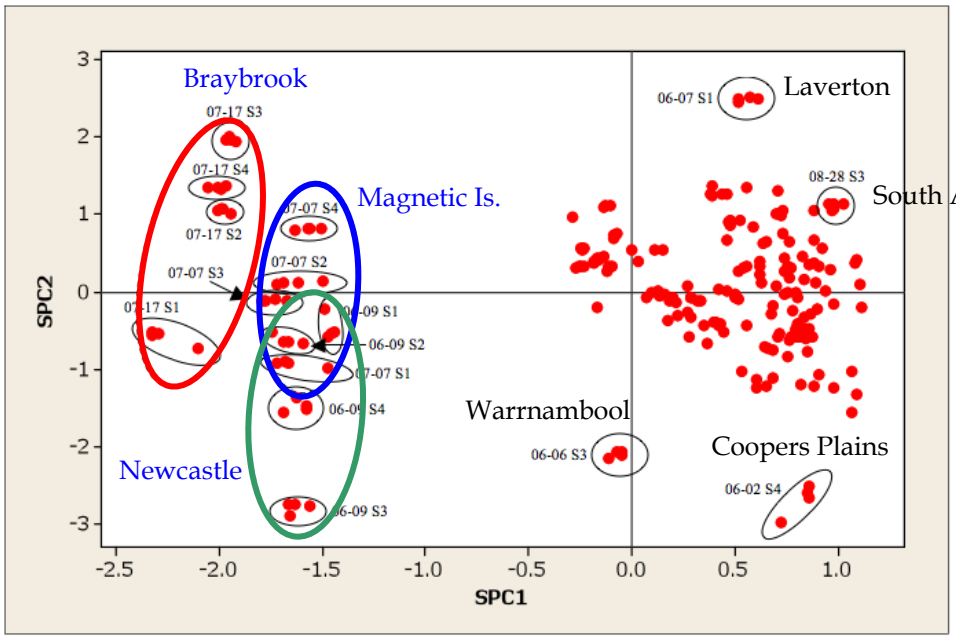


Figure 6: Scores plot of PC 1 vs. PC2. The Braybrook, Magnetic Island and Newcastle specimens were well accounted for by this PCA.

Evident from Figure 6 was that all four seeds from the Braybrook, Magnetic Island and Newcastle specimens populated similar space in the scores plot. Compared to the other specimens, these three specimens were heavily influenced by the negative loadings comprising PC1. This indicated that there were components described by these loadings within the metabolome of the Braybrook, Newcastle and Magnetic Island specimens that were not present in the other specimens. Additionally, three of the seeds from the Braybrook specimen showed positive separation on PC2. This could indicate that if a larger number of seeds were extracted (to average out seed-to-seed metabolome variation), Braybrook could contain components in PC2 that also account for the observed variation. Conversely for Magnetic Island and Newcastle specimens, PC2 allowed only some separation. What is evident from the analysis of Figure 6 is that there appeared to be some clear trends that could allow for discrimination between these specimens.

Also evident from Figure 6 was that only one seed each from the Warrnambool, Coopers Plains, South Arm and Laverton specimens were clearly separated. All remaining seeds from these four specimens, and all other specimens investigated, were not able to be identified. This is highlighted by the number of unassigned injections evident in Figure 6 (red dots without solid ellipses around them). This indicates that neither PC1 nor PC2 are responsible in accounting for the metabolome variation of these specimens. However, as four specimens had one seed explained by Figure 6, it could also be that seed-to-seed metabolome variation is

significant for these specimens. More seeds per specimen will need to be extracted in a single batch so that any seed-to-seed biological variability can be averaged, hence removing this issue.

The loadings plot for PC1 and PC2 allowed for an identification of which bins (and hence retention time range) characterise each PC, and is shown in Figure 7. Each vector represents one bin, and all vectors (by definition) were identical in length. The longer a vector appears in Figure 7, the more closely it was aligned to an axis. The closer that a vector was to an axis, the more it accounted for the variance observed in that PC. Outlined in Appendix B is the bin number (with corresponding retention time) with the PC that best explains the variance of that particular bin.

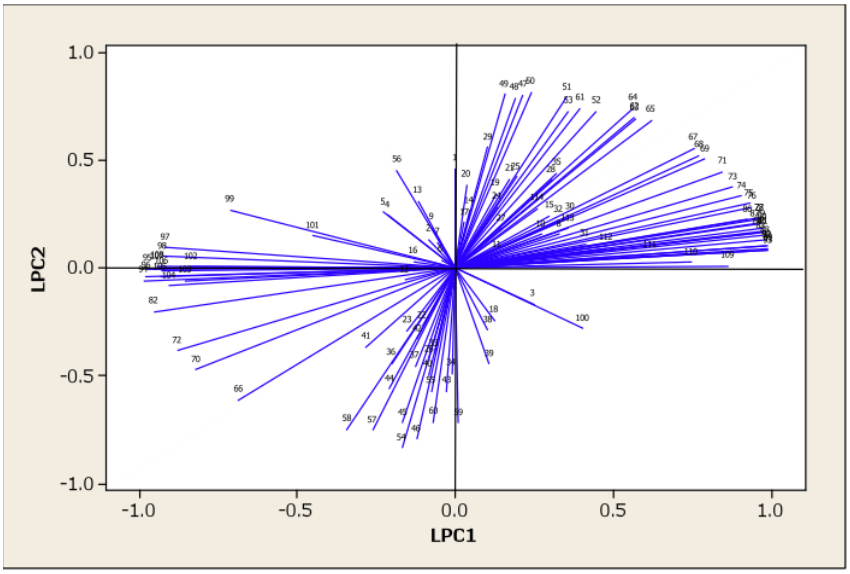


Figure 7: Loadings plot of PC1 and PC2. Vectors most closely aligned to the axis indicate the bins with the strongest influence.

With the bins identified that accounted for the derived PC, the associated mass spectral data was analysed. The aim was to identify the presence or absence of compounds which could allow for one *R. communis* specimen to be distinguished over another. Analysis of the HPLC-UV data for bins responsible for PC1 (Appendix B) showed that across a majority of the bins, all specimens had essentially identical bin composition. However, for the Braybrook specimen, there were some significant observed differences for bins 81 to 83 (28 to 29 min) as compared to the other specimens. The mass spectra for bins 81 to 83 for the Braybrook, Newcastle and Magnetic Island specimens are shown in Figure 8. As highlighted, the ion at *m/z* 461.6 was completely absent in the Braybrook specimen. Additionally, the ion at *m/z* 353.6 was significantly reduced in intensity. It should be noted that these compounds were present in varying amounts in all other specimens analysed.

Furthermore, the Braybrook specimen appeared to contain a very minor ion at *m/z* 661.6 that was not present in any other specimen extract (data not shown). Further work needs to be undertaken to validate these observations for the Braybrook specimen, but these initial results

were encouraging. It is also worth noting that the Magnetic Island specimen contained the peptide biomarker RCB-3[60] in bins 67 and 68 as a triply charged ion at *m/z* 655.3, as did the Coopers Plains and Clifton Hill specimens. This peptide was originally discovered in seeds from the cultivar "*Carmencita*" collected from Tanzania. Our analyses to date on all specimens of *R. communis* suggests that this peptide is not very common, with it only being present in four of some 22 specimens analysed. On its own, RCB-3 is not a definitive biomarker for the Magnetic Island specimen. However, considering the uncommon nature of RCB-3, it is expected that it will in future be exploited as a biomarker for a small subset of *R. communis* cultivars.

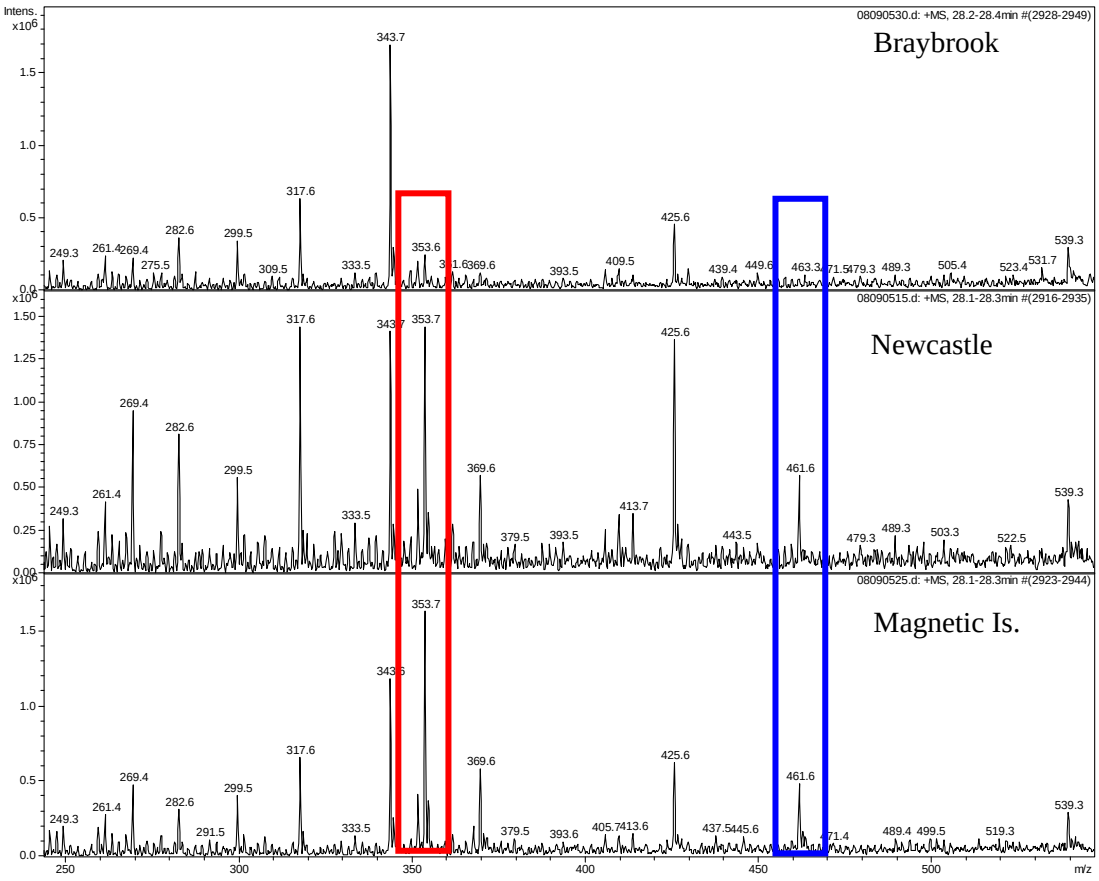


Figure 8: Mass spectral data for bins 81 to 83 for the Braybrook, Newcastle and Magnetic Island specimens. Red box: ion at m/z 353.6; Blue box: ion at m/z 461.6.

Shown in Figure 9 are actual images of seeds from the Braybrook, Magnetic Island and Newcastle specimens. The seeds from these three specimens are visually very similar. Many of the other specimens investigated during this study looked similar to these seed samples. These results therefore underlined the power of the chemometrics approach to studying *R. communis* seed metabolome for cultivar determination. For this example the metabolomics/chemometrics approach has allowed for specimen differentiation to be made between similar looking seeds.

Figure 9: Images of the seeds from Braybrook (left), Magnetic Island (centre) and Newcastle (right) specimens

2.1.3.2 PC1 vs. PC3

The scores plot of PC1 vs. PC3 is shown in Figure 10 and identified which specimens were accounted for by both PC1 and PC3. While more ambiguous than the scores plot of PC1 vs. PC2, some interesting results were garnished.

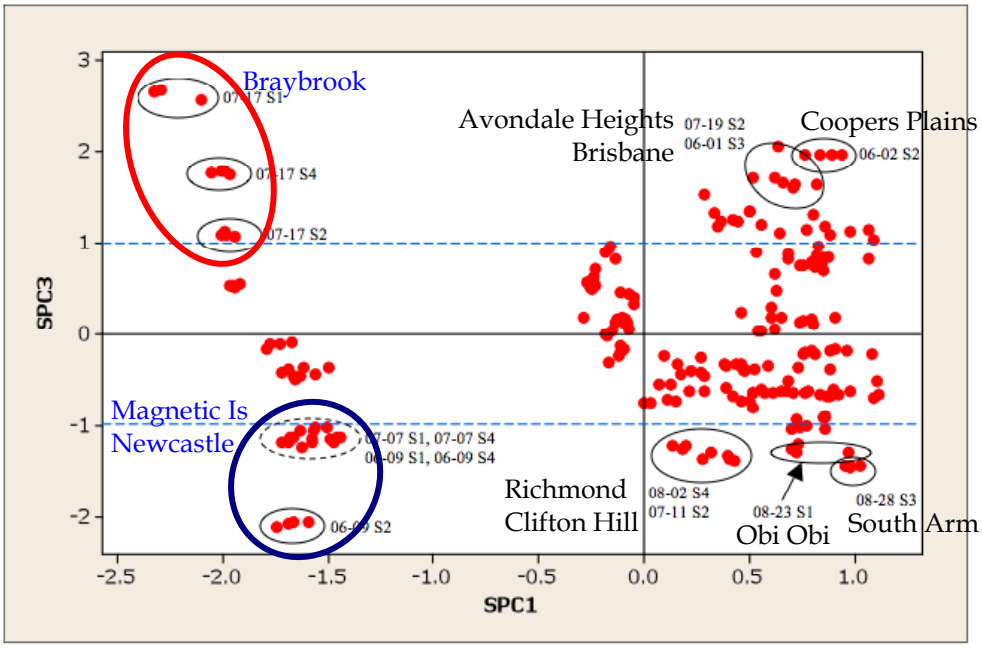


Figure 10: Scores plot of PC1 vs. PC3 showing data best accounted for by PC3. The Braybrook and Newcastle specimens had three seeds accounted for by this scores plot.

As for the scores plot of PC1 vs. PC2 (Figure 6), PC1 vs. PC3 best explained the same three specimens: Braybrook, Magnetic Island and Newcastle. PC3 clearly allows differentiation of Braybrook (+ve PC3) from Magnetic Island and Newcastle (-ve PC3) specimens. For the Magnetic Island specimen only two seeds were accounted for by PC3. It is unclear from this if the bins that made up PC3 were contributing significantly to the observed variation for the Magnetic Island specimen. It is suspected that this ambiguity was due to the seed-to-seed biological variation in the sampled Magnetic Island seed specimens.

Conversely, three of the four seed extracts for both Braybrook and Newcastle specimens were accounted for by PC3. This was thought to be significant, hence mass spectra for the bins that accounted for PC3 were analysed for potential biomarkers. These results are outlined in Table 1. Stack plots of mass spectra for those bins that yielded potential biomarkers are shown in Appendix C.

From this, several potential specimen biomarkers were identified. For the Braybrook specimen, a doubly charged ion at *m/z* 594.2 was identified for bins 10 to 13. An additional ion at *m/z* 644.6 for bins 41 to 45 was also identified. A significant reduction in the intensity of the ion at *m/z* 261.5 in bins 21 to 28 was also observed. Of particular interest was the doubly charged ion at *m/z* 594.2, which may be a small peptide. There is a propensity for *R. communis* seeds to produce small peptides.[60] Indeed, unpublished analysis of the metabolome of the "*Dehradun*" cultivar has allowed for the identification of several unique peptides.[61] For the Newcastle specimen, a compound was identified with a molecular ion at *m/z* 621.4 in bins 10 to 13, with the additional observation of the absence of an ion at *m/z* 583.7 in bins 16 to 18.

Table 1: Identified ions in PC3 for Braybrook and Newcastle

Bin responsible for PC3	Braybrook (*m/z*) (+ve PC3 axis)	Newcastle (*m/z*) (-ve PC3 axis)
1 to 5	-	-
10 to 13	594.2^{2+}	621.4
16 to 18	-	absence of 583.7
21 to 28	significantly reduced 261.5	-
31 to 36	-	-
41 to 45	644.6	-
56	-	-

2.1.3.3 PC3 vs. PC4

The scores plot of PC3 vs. PC4 (Figure 11) allowed for all seeds from the Avondale Heights specimen to be identified, with bins that accounted for PC3 responsible for the differentiation.

Analysis of the corresponding mass spectra for the Avondale Heights specimen identified no specific compounds. However, a potentially useful observation for bins 10 to 13 was observed. The Avondale Heights extract did not contain ions at either *m/z* 408.7 or *m/z* 559.5. Of the other six *R. communis* that had all four seed extracts identified, two had ions at *m/z* 559.5 present (Newcastle and Coopers Plains), while four had both ions at *m/z* 408.7 and *m/z* 559.5 (Footscray, Magnetic Island, Braybrook, and Warrnambool) present. To use the absence of compounds in a metabolome as an indicator of cultivar is a powerful observation. Further investigations are required to validate these observations.

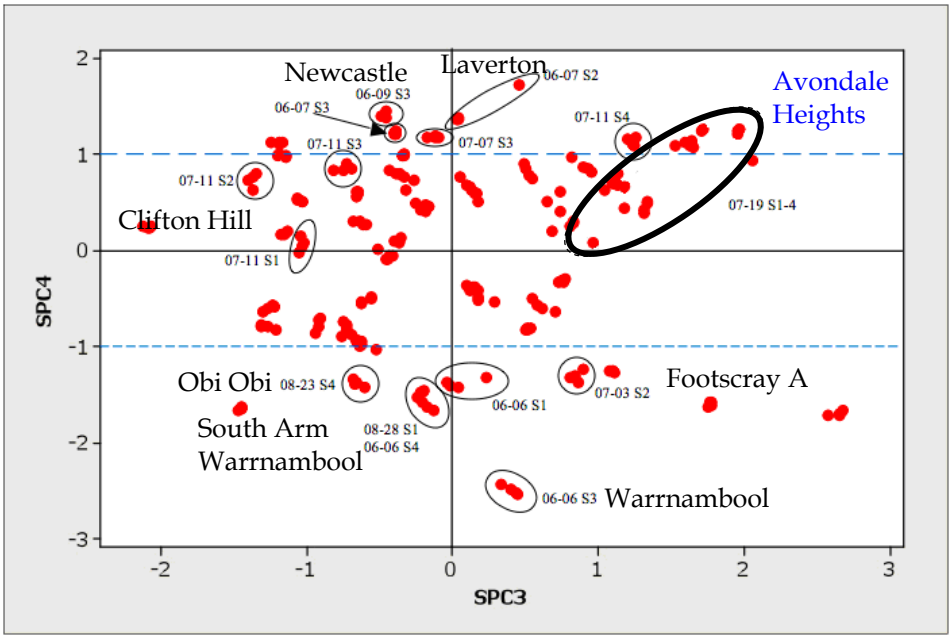


Figure 11: Scores plot of PC3 vs. PC4. Avondale Heights is the only specimen that is fully accounted for by this scores plot.

2.1.3.4 PC4 vs. PC5

The scores plot of PC4 vs. PC5 (Figure 12) yielded perhaps the most striking scores plot of all. This scores plot shows that the Warrnambool, Footscray and Coopers Plains specimens were highly associated with the bins comprising PC5. Additionally, three of the four Warrnambool specimens show a strong association with PC4.

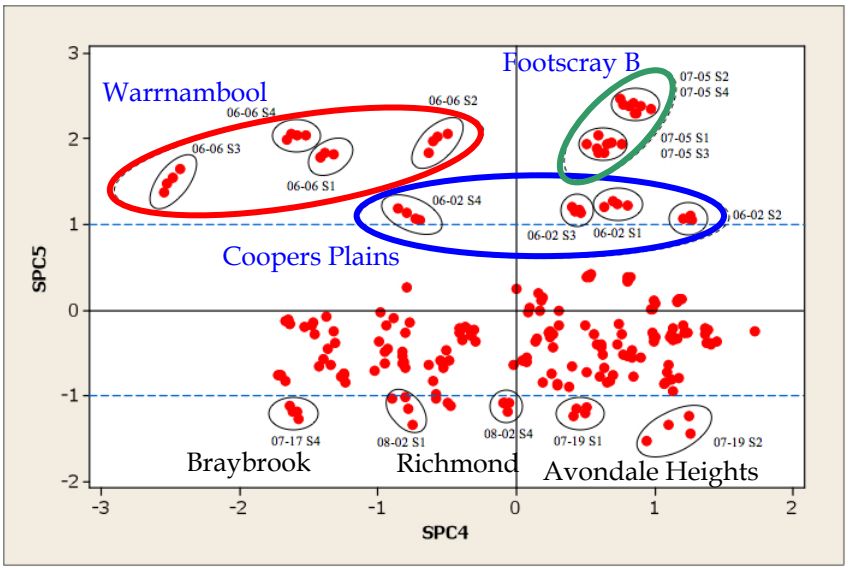


Figure 12: Scores plot of PC4 vs. PC5. Red ellipse: Warrnambool, Blue: Coopers Plains, Green: Footscray B

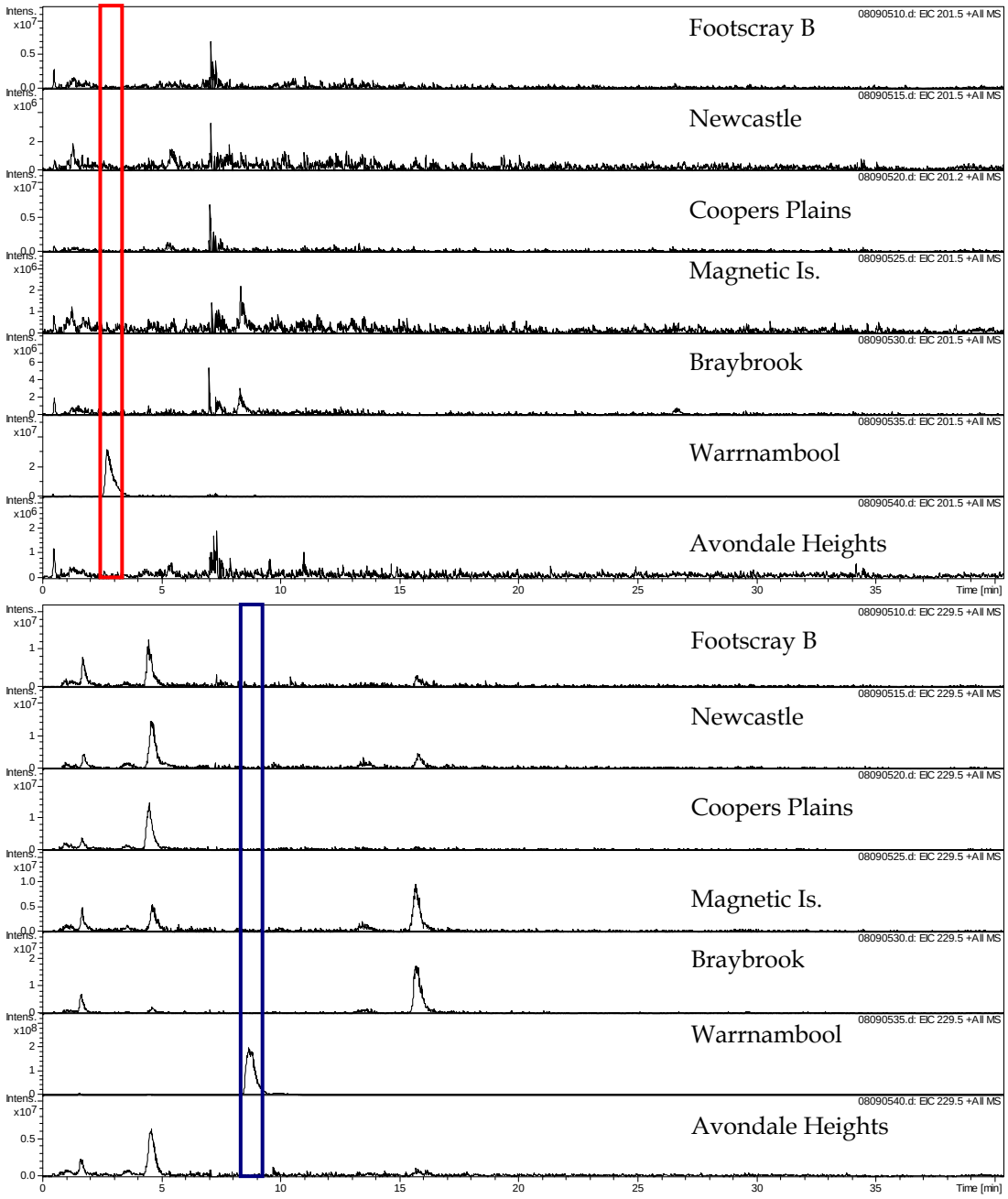


Figure 13: EIC for ions at m/z 201.5 (top spectra), and m/z 229.5 (bottom spectra). Red box highlights bin 9, blue box highlights bin 26. Only the Warrnambool specimen has both these ions present in these bins.

There were six bins that contributed to PC4. Of these, for the Warrnambool specimen (the only specimen accounted for by PC4), analysis of the mass spectra for four of the bins yielded no diagnostic compounds. However, analysis of the remaining two bins yielded some intriguing results. For bin 9, an ion was identified at *m/z* 201.5, while for bin 26 an ion was identified at *m/z* 229.5. Extracted Ion Chromatograms (EIC) for these two ions are shown in Figure 13, with bin 9 highlighted by the red box, and bin 26 highlighted by the blue. It is clear that only the

Warrnambool metabolome has these two ions present in these two bins. No ions of interest were identified from the mass spectra of bins that contribute to PC5 for the Warrnambool, Coopers Plains or Footscray B specimen. There are several reasons as to why no diagnostic ions were identified for bins explained by PC5. These include minor amounts of low intensity ions, and/or ion suppression due to matrix effects. Despite this, the identification of the two ions in bins 9 and 26 from PC4 suggested that LC-MS may have provided evidence for diagnostic indicators of Warrnambool seed specimens.

The scores plot of PC4 vs. PC5 suggested that the Warrnambool, Footscray and Coopers Plains specimens had metabolomes that were both distinct from themselves as well as from the other specimens analysed. Shown in Figure 14 are images of these three seeds specimens. As can be seen, the Warrnambool specimen was morphologically distinct from Coopers Plains and Footscray B. While this may have implied that the results generated from PC4 vs. PC5 scores plot were expected, the results also indicated that studying the metabolome of *R. communis* seeds has the potential to successfully establish cultivar, *a priori* of specimen knowledge.

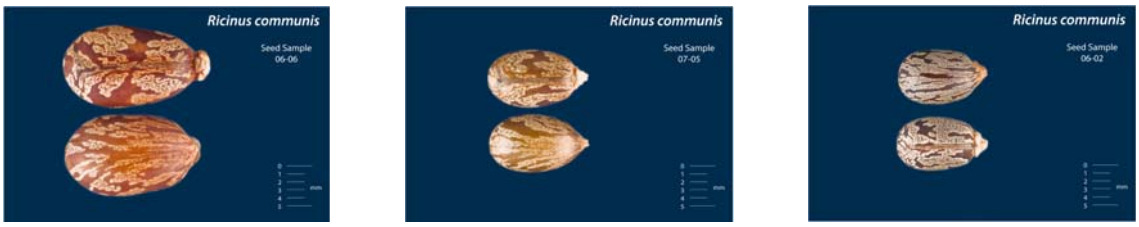


Figure 14: Seed pictures of specimens Warrnambool (left), Footscray B (centre) and Coopers Plains (right)

2.1.4 Summary of outcomes from HPLC-UV and LC-MS analysis

In total, four seeds each from fourteen specimens (56 seeds in total) of *R. communis* were investigated by HPLC-UV and chemometrics for cultivar determination. Of these fourteen specimens, seven had all four seeds accounted for by the various scores plots discussed. Five of the seven specimens had some observed differences in the mass spectra of the bins responsible for the PC that best explained that specimen. These results are outlined in Table 2.

Table 2: Bins and observed significant molecular ions for each specimen

Specimen	PC	Mass spec observations
Braybrook	1	Reduced intensity of *m/z* 353.7, no *m/z* 461.6, minor *m/z* 661.6
	3	*m/z* 594.2^{2+}, *m/z* 644.6, reduced intensity of *m/z* 261.5
Magnetic Island	1	RCB-3 at 655.4^{3+}
Newcastle	1	No definitive masses identified
	3	*m/z* 621.4, reduced intensity of *m/z* 583.7
Avondale Heights	3	Complete absence of *m/z* 408.7 and *m/z* 559.5
Warrnambool	4	*m/z* 201.5 and *m/z* 229.5
	5	No definitive masses identified
Coopers Plains	5	No definitive masses identified
Footscray B	5	No definitive masses identified

The results obtained during this part of the research work were encouraging, and suggestive that the HPLC-UV metabolomics approach for cultivar determination has merit. Clearly these results are yet to be validated, and hence could not be considered definitive. There are two issues that need to be addressed: the seed-to-seed biological variation; and the variation in the metabolome due to seasonal and local environment. Addressing these two critical points will allow for the validation of identified compounds above as specimen specific biomarkers.

2.2 ^{1}H NMR spectroscopy and PCA of *R. communis* seed extracts

Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful analytical technique that is widely used in a variety of plant based metabolomic applications.[42,46,62-70] One dimensional ^{1}H NMR has been applied to the metabolomic determination of provenance of Italian olive oils[71,72] and propolis samples,[73] cultivar and provenance determination of French[74,75] and Italian[76] wines, the variety of apples used for apple juice,[69] and ecotype determinations of *Arabidopsis thaliana*.[77] These applications suggest that ^{1}H NMR can be applied to determine *R. communis* cultivar from seed extracts. While NMR analysis is an insensitive technique compared to HPLC-UV and LC-MS, it does have several advantages including the non-selective nature of the analysis (all metabolite structure classes can be analysed using NMR); the ease of sample preparation; the non-destructive nature of the analysis (sample can be recovered); and the ability to both derive qualitative structural information, and to quantitate metabolites relatively easily.

To this end ^{1}H NMR was investigated as an analytical technique to study the seed metabolome of *R. communis*. There were two main aims of this investigation: To determine if there was measurable seed-to-seed biological variation in the metabolome when analysed by ^{1}H NMR (as observed for HPLC-UV analysis); and to investigate the metabolome of *R. communis* seeds to identify specimen.

2.2.1 Sample preparation and analysis

Given that NMR is less sensitive than other analytical techniques, unused extracts from the HPLC-UV analysis of the four seeds were combined, hence forming fourteen extracts. These combined extracts were then freeze dried, and resuspended in a solution of 2% d_4-acetic acid in D_2O. Added to each sample for analysis was an internal standard of 0.1% 3-(trimethylsilyl)-2,2,3,3,-d_4-propionic acid (TSP, referenced to δ 0.00 ppm). The addition of acetic acid allowed for all extracts to have their ^{1}H NMR data collected at a common pH, minimising chemical shift perturbations due to differing pH. Each sample was made up to a concentration of 20 mg/mL, sonicated for 30 s, and centrifuged to remove insoluble material.

To determine if there was any observed seed-to-seed variation in the ^{1}H NMR data on individual seed extracts, and to measure the influence this had on the outcomes of the PCA, the four seed extracts from the Richmond and South Arm specimens were individually analysed before being combined to form two samples. Therefore a total of 22 ^{1}H NMR spectra were collected and submitted to PCA. To suppress residual HDO signal in the sample and increase the sensitivity of the experiment, the "Watergate" solvent suppression pulse sequence was employed.

2.2.2 NMR statistical analysis

2.2.2.1 Model One

Manually phased and baseline corrected ^{1}H NMR data was binned within the statistical program Matlab™ using the PLStoolbox with the ProMetab script.[78] The chemical shift range for binning was between δ 0.5 ppm and δ 10.0 ppm. Bin widths were set at δ 0.002 ppm, for a total of 4750 bins. The binned data was normalised and subjected to PCA. The scores plot of PC1 vs. PC2 shown in Figure 15. Overall 57% of the observed variability in the data was accounted for by PC1 and PC2. It was expected from this analysis that the individual and combined extracts from Richmond and South Arm would occupy similar space in the scores plot. As can be seen in Figure 15, this was not the case, with only two Richmond, and three South Arm extracts occupying similar space in the scores plot. More importantly, the combined extracts from these two specimens did not group with the individual extracts.

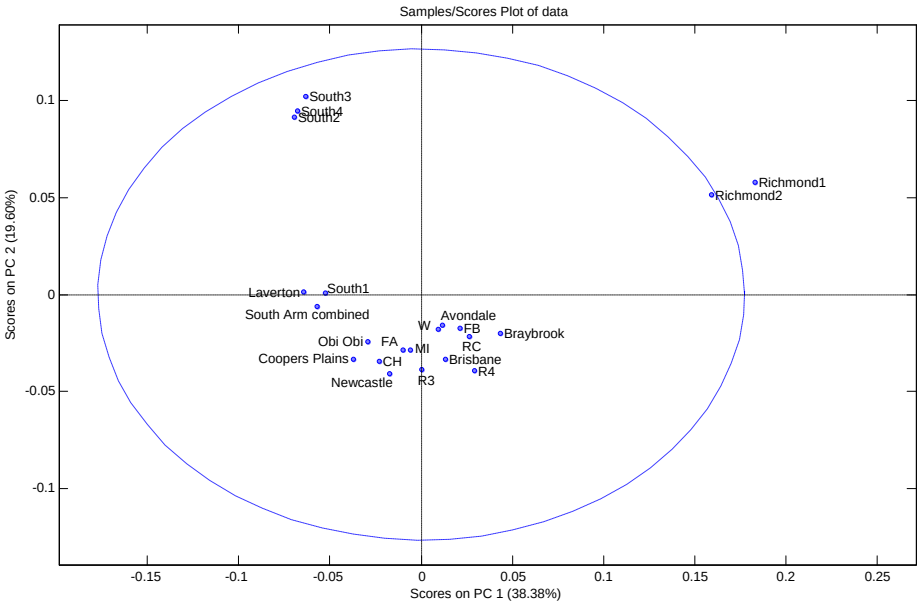


Figure 15: Model one scores plot of PC1 vs. PC2. A lack of correlations between extracts of common seed specimens was observed. South Arm and Richmond extracts annotated with numbers indicate individual seed extracts analysed. Abbreviations for some specimens have been used to declutter scores plot: CH – Clifton Hill; FA – Footscray A; FB – Footscray B; MI – Magnetic Island; R3 – Richmond sample 3; R4 – Richmond sample 4; RC – Richmond combined; W – Warrnambool.

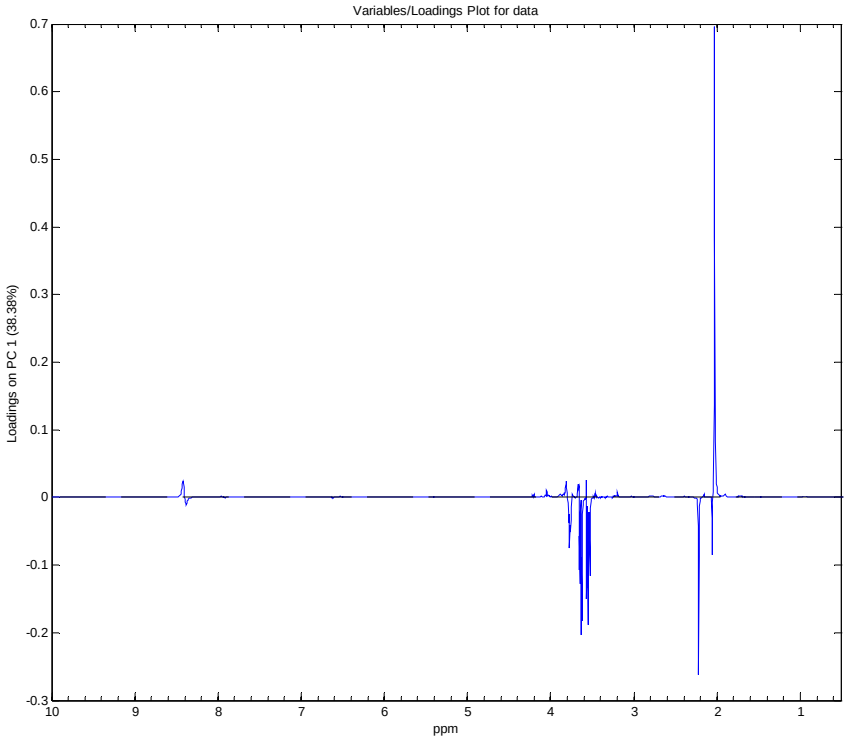


Figure 16: Loadings plot of PC1 clearly showing the influence of residual acetic acid on positive PC1, and quantity of sugar in an extract on negative PC1

Subsequent analysis of the loadings plots (Figure 16) for PC1 clearly showed that residual undeuterated acetic acid from the NMR solvent, in addition to sugars and inositol naturally occurring in the *R. communis* seed extracts, had the greatest influence on the scores plot in Figure 15. A ^{1}H NMR spectrum highlighting these specific regions is shown in Appendix D.

2.2.2.2 Model Two

To reduce the effect these residual resonances had on loading, a log function was applied to the binned data. Subsequent PCA showed that 76.2% of the variance was accounted for by PC1, PC2 and PC3. The scores plot of PC1 vs. PC2, and PC1 vs. PC3 are shown in Figure 17. Corresponding loadings plots are shown in Figure 18. Applying the log function increased the robustness of the model, with PC1 accounting for 51.2% of the variability. This is reflected in the scores plots shown in Figure 17. In the plot of PC1 vs. PC2 (Figure 17a), the Richmond and South Arm extracts are starting to correlate into the same space. The scores plot of PC1 vs. PC3 (Figure 17b) showed strong grouping for all South Arm extracts, and a further tightening of the Richmond extracts.

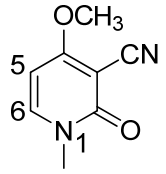
The loadings in Figure 18 suggest that applying a log function to the data has significantly reduced the influence that acetic acid and sugar/inositol had on PC2 (Figure 18a). For PC3 (Figure 18b), in addition to the residual acetic acid and sugar/inositol resonances, residual formic acid at δ 8.5 ppm from the HPLC solutions had a significant influence (Appendix D). Interesting, applying this log function highlighted a relationship between the amount of

ricinine (**1**) (positive influence on PC2 and PC3) with the amount of sugar/inositol in these extracts (negative influence on PC2 and PC3). Due to the other interferences in the extracts this relationship could not be further investigated. However, this could be a critical observation and one that will be further investigated.

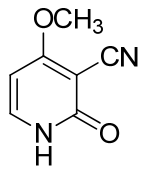
As a result of the influences the resonances due to formic acid, acetic acid and sugar/inositol had on the PCA, a model was built using a reduced chemical shift range between δ 5.6 and δ 8.2 ppm. The same bucket width of δ 0.002 ppm was used, for a total of 1300 buckets. A full analysis of the data was made using this model, including an analysis of the seed-to-seed variation that was observed. These results are discussed in the next section.

2.2.2.3 Model Three

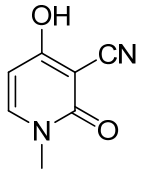
Firstly, the Richmond and South Arm extracts were analysed to ascertain if the seed-to-seed metabolome variation was significant. A visual analysis of the ^{1}H NMR data immediately identified an interesting observation. There was a significant difference in the quantities of compounds responsible for downfield resonances in these extracts. In particular, the Richmond seeds contained significantly more compounds with downfield resonances compared to the South Arm seeds. This is best shown by the intensity of the ricinine (**1**) doublets for the two specimens at δ 6.52 ppm (H-5) and δ 7.95 ppm (H-6, Figure 19). [56,57]



ricinine (**1**)

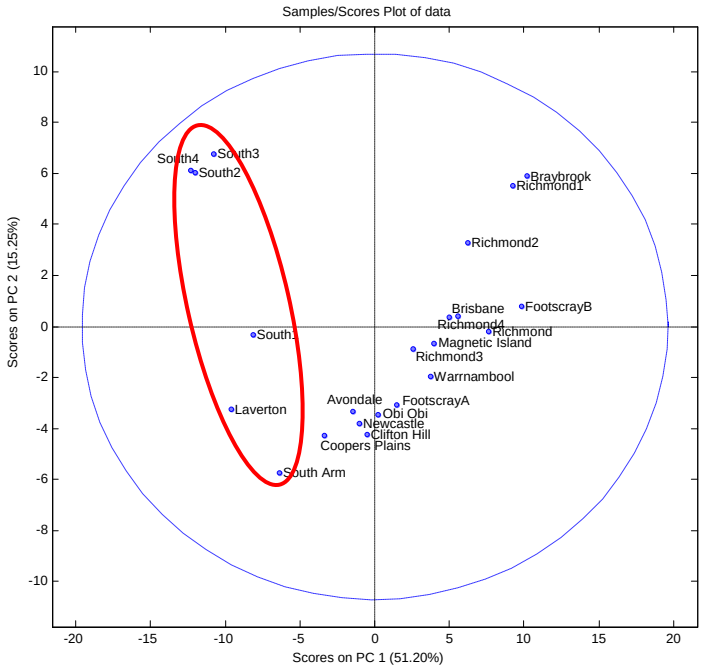


N-demethylricinine (**2**)

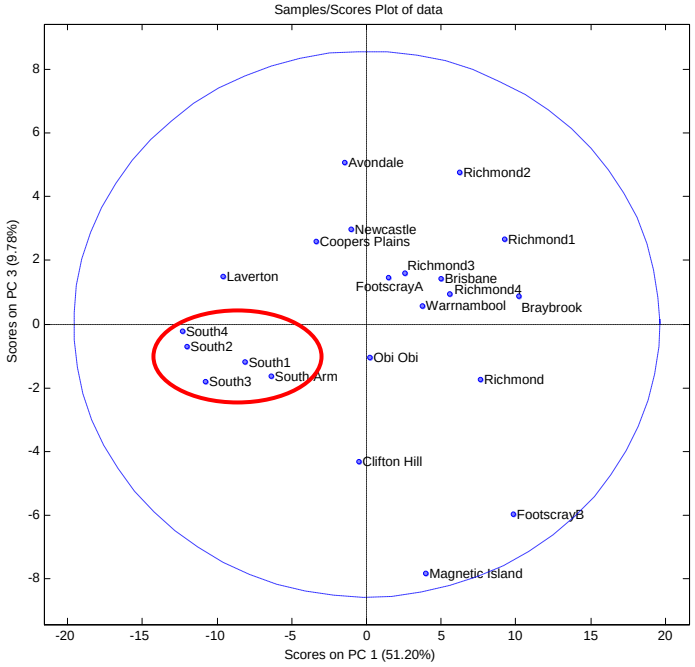


O-demethylricinine (**3**)

Also evident in the ^{1}H NMR for these two specimens were doublet resonances at δ 6.62 ppm and δ 7.92 ppm, which could be explained by either or both of *N*-demethylricinine (**2**) and *O*-demethylricinine (**3**).[57] There is evidence in the LC-MS data that either one or both of **2** and **3** are present in these extracts. At this point in time no discrimination between these two compounds can be made. This will be the subject of further follow up investigations via isolation and NMR analysis. It is also worth noting the over integration of the H-5 protons for both **1** and **2**/**3**. The chemical shift values for **1** have been verified with an authentic standard. It appeared that for these extracts there is another molecule with a doublet at δ 6.52 ppm for **1**, and δ 6.62 ppm for **2**/**3**. Further investigations to elucidate the structure of this compound are in progress.

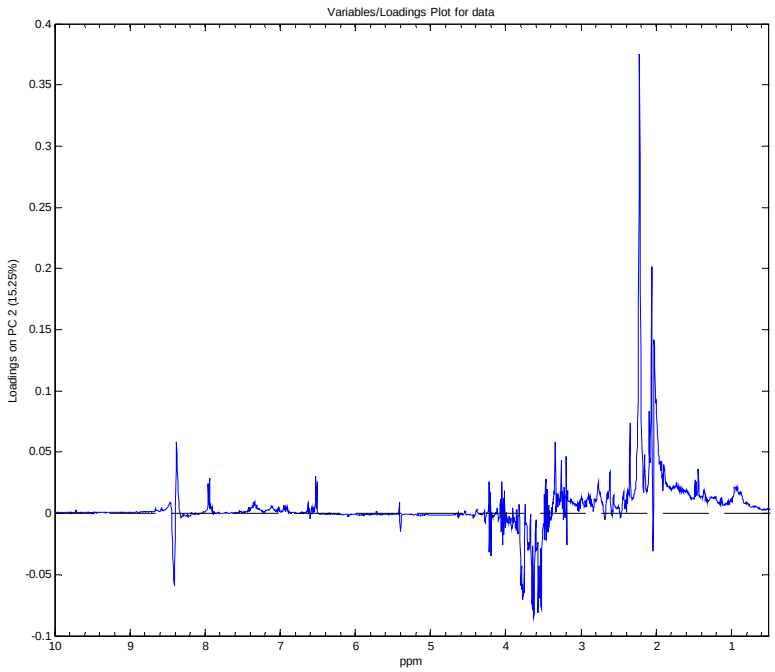


(a)

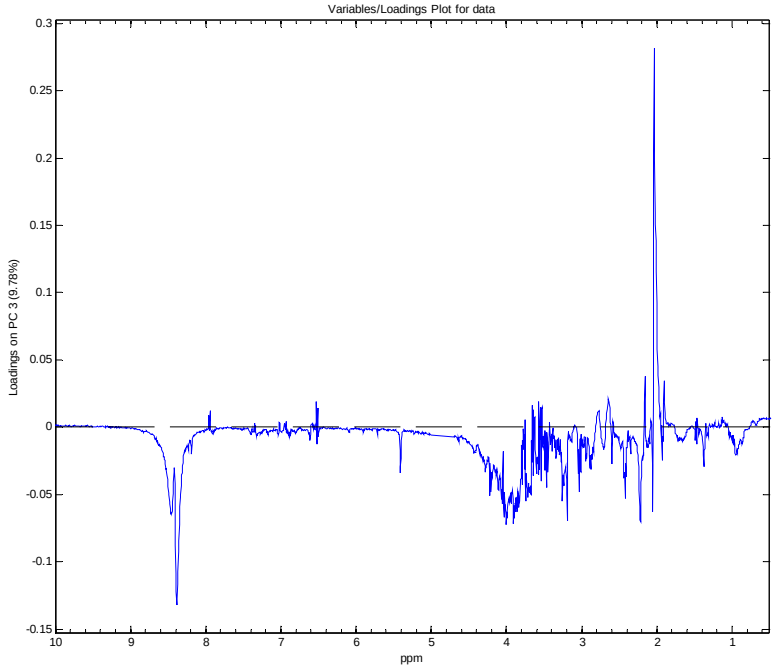


(b)

Figure 17: PCA scores plots of binned NMR data after applying a log function. (a) PC1 vs. PC2; (b) PC1 vs. PC3. Note the tight grouping of South Arm extracts in the scores plot of PC1 vs. PC3.



(a)



(b)

Figure 18: Loadings plots on (a) PC2; and (b) PC3. The influence due to residual acetic acid and sugar is still significant even after a log function is applied to the data. Residual formic acid in the extracts is influencing loadings on PC3.

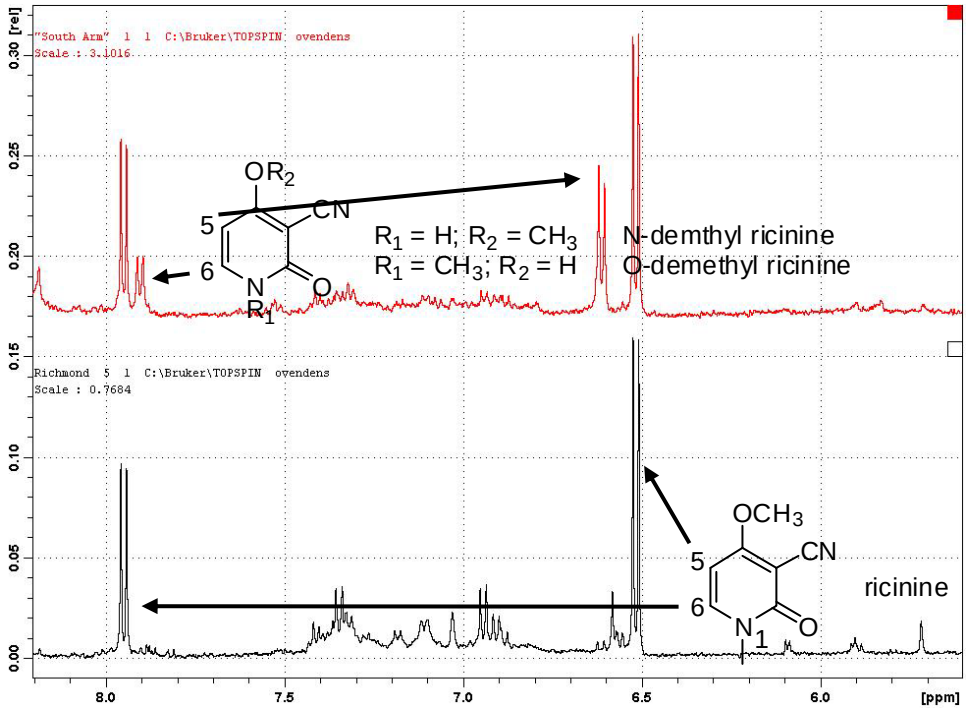
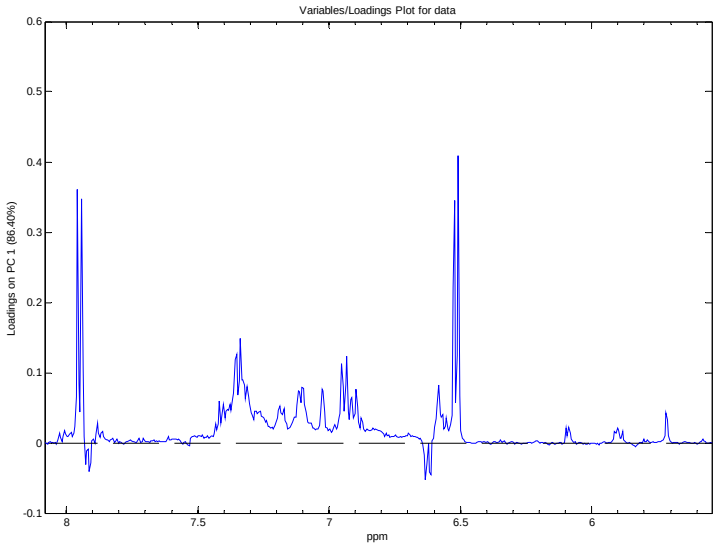
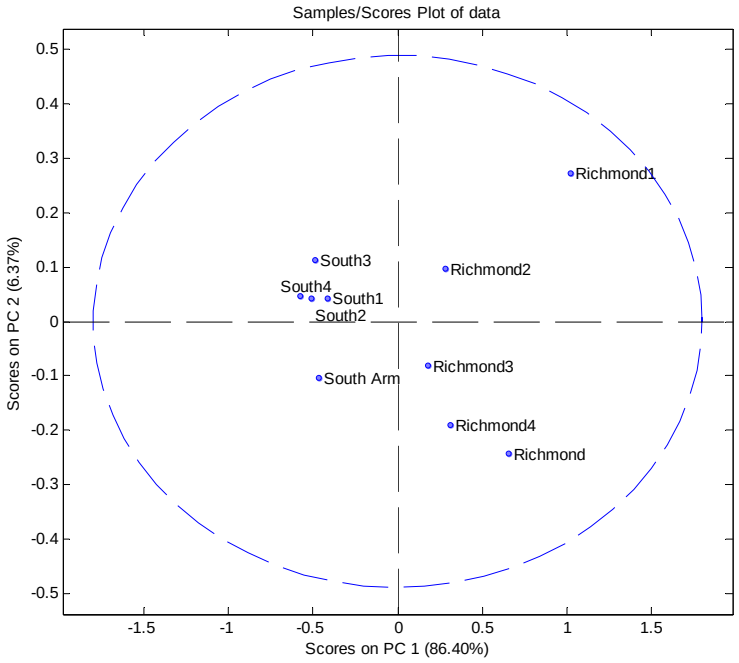


Figure 19: ^{1}H NMR spectra of the South Arm and Richmond combined extracts. Ricinine and demthylricinine analogues dominate the spectrum for South Arm. Conversely, ricinine and other unidentified compounds dominate the Richmond spectrum.

Examination of the loadings plot for PC1 (Figure 20a) shows that **2**/**3** was weighted negatively for PC1, with the South Arm specimens seemingly containing more **2**/**3**. Conversely, **1** was weighted positively for PC1 along with several unidentified metabolites, with the Richmond specimens containing more of these metabolites. A total of 90.97% of the variance was explained by PC1 (86.4%) and PC2 (6.37%). The scores plot of PC1 vs. PC2 (Figure 20b) showed excellent grouping of the South Arm extracts, while the Richmond extracts were not so well grouped. It is unclear at this stage why the combined Richmond extract was not grouped with the individual extracts. It may be that there was an issue with sample handling prior to ^{1}H NMR analysis of the combined extract, leading to some compound decomposition. What was clear from this was that there was little seed-to-seed variation in *R. communis* seed metabolome described by this narrow region of the ^{1}H NMR spectrum. This contrasted with the HPLC-UV analysis, where significant seed-to-seed variation was observed.



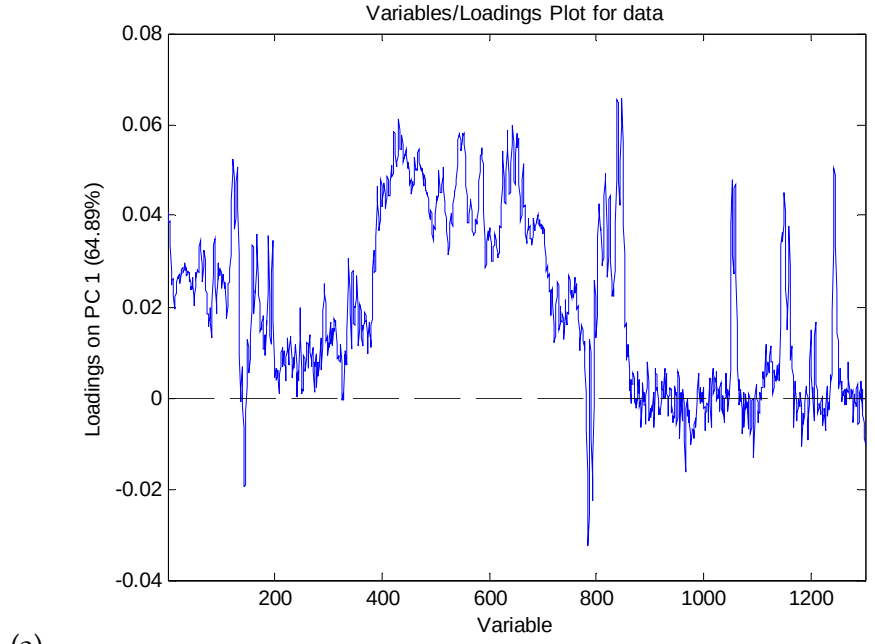
(a)



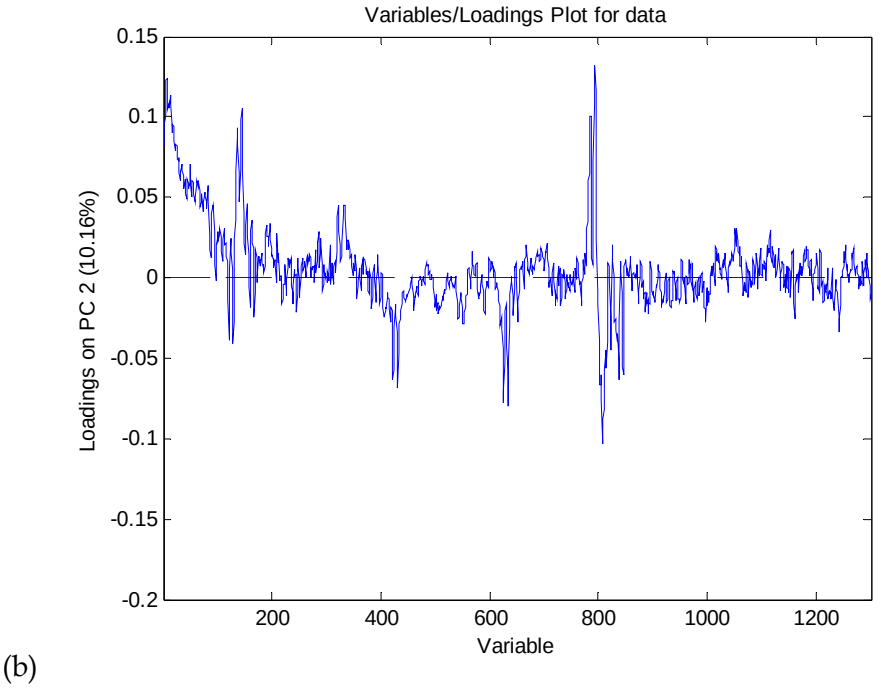
(b)

Figure 20: (a) Loading plot for PC1 from the ^{1}H NMR of downfield region of South Arm and Richmond individual and combined extracts. (b) Scores plot of PC1 vs. PC2.

The remaining twelve combined *R. communis* metabolome extracts were then added to the PCA model. Loadings plots for PC1 (Figure 21a) shows that **1** had a positive effect on PC1, while **2**/**3** had a negative effect. For PC2 (Figure 21b) the loadings plot clearly showed that **2**/**3** had a positive effect. There are additional aromatic and olefinic resonances that had strong positive and negative influences on the loadings plots. In total 75.05% of the variance was explained by PC1 and PC2.



(a)



(b)

Figure 21: Loadings plot for model three: (a) PC1 loadings; (b) PC2 loadings

The subsequent scores plot of PC1 vs. PC2 shown in Figure 22 clearly shows clustering of the individual and combined seed extracts from the South Arm specimen. All Richmond analyses cluster as compared to the scores plot shown in Figure 20. It is suspected that this was a consequence of the amounts of **1** and **2/3** in the Richmond extracts being comparatively similar when compared to the other extracts analysed. This observation seems to suggest that

the relative ratio of **1** to **2/3** produced in seeds fluctuates between specimens. Therefore, this variation may indeed be a powerful observation in the determination of specimen.

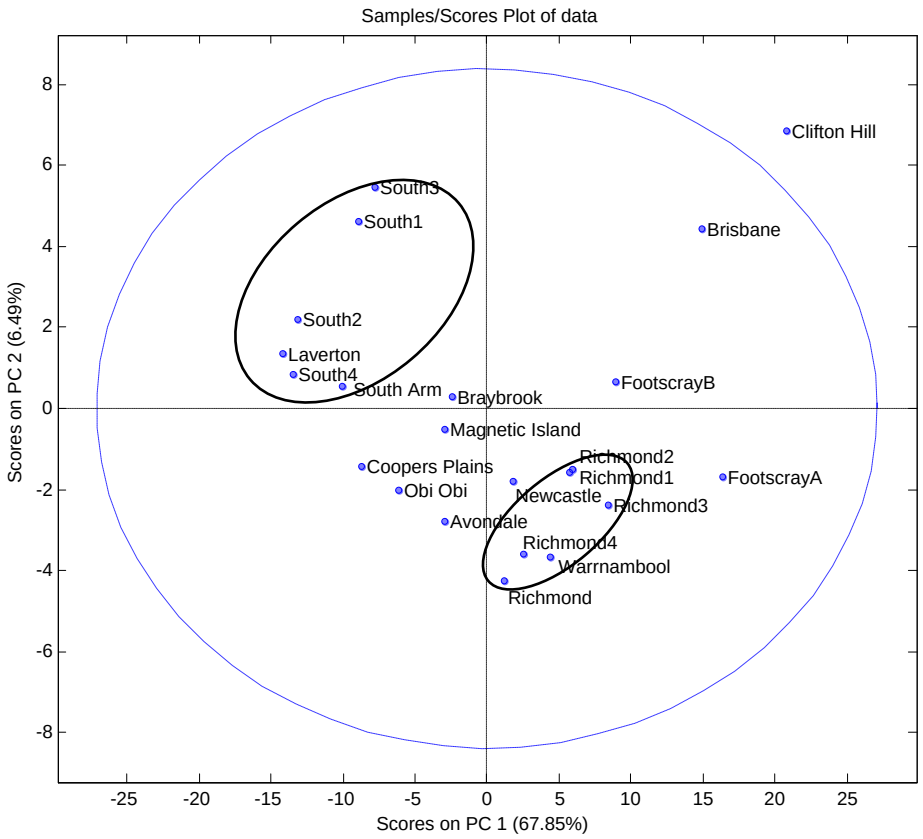


Figure 22: Scores plot of PC1 vs. PC2 for the 22 ^{1}H NMR analyses. All Richmond extracts clustered, as did the Warrnambool extract. Similarly all South Arm extracts clustered, as did the Laverton extract.

Other significant observations include the Brisbane and Clifton Hill specimens being explained by PC1 and PC2. The Newcastle and Avondale Heights specimens were explained by negative loadings on PC2. Both Footscray A and B specimens were explained by positive loadings on PC1, with negative PC2 loadings having some influence on Footscray A. Alternatively, Coopers Plains and Obi Obi specimens were explained by negative loadings on PC1 and PC2. Finally the Braybrook and Magnetic Island specimens were not strongly influenced by loadings on either PC1 or PC2. While it may be the case that the ratio of **1** to **2/3** is influencing the scores plot for these specimens, it is unclear what influence the other compounds were having. Significant qualitative/quantitative work needs to be performed on both environmental and controlled growth specimens. This will allow for a greater understanding into the chemistry involved.

A further observation from Figure 22 was that the Richmond extracts cluster with the Warrnambool extract, while the South Arm extracts cluster with the Laverton extract. There was a significant difference in the amounts of **1** and **2/3** present in the extracts, as noted in their respective ^{1}H NMR spectra (Appendix E). The amount of **2/3** present in the South Arm and Laverton extracts was significantly increased compared to **1**. This observation was reversed for the Richmond and Warrnambool extracts, with virtually no **2/3** present in the ^{1}H NMR spectra. This observation gives further credence to the ratio of **1** to **2/3** being a potential indicator of specimen.

2.2.3 Summary of the ^{1}H NMR analysis

These investigations into the ^{1}H NMR of the seed metabolome of *R. communis* again yielded promising results. The third model evaluated where the data was log transformed, binned between δ 5.6 and δ 8.2 ppm, and subjected to PCA yielded the most definitive results. This analysis suggested that the amounts of **1** and **2/3** could be utilised as identifiers of specimen. While this was a very limited study on the applicability of applying NMR to analyse *R. communis* extracts for cultivar determination, it is evident from the analysis presented that further investigations are warranted.

3. Conclusions

This research aimed to prove the concept that specimen determination of *R. communis* could be elucidated through the study of the seed metabolome and chemometrics. A total of 56 seeds from fourteen specimens were analysed by HPLC-UV, LC-MS and ^{1}H NMR, with the data subjected to PCA. The results discussed in this technical report show that there is merit in further pursuing the metabolome of seed extracts of *R. communis* for cultivar determination.

Analysis of the metabolome via HPLC-UV and LC-MS identified all four seed extracts from seven of the fourteen specimens studied. Furthermore, a number of unique molecular ions in addition to the absence of several molecular ions were identified for five of the seven specimens (summarised in Table 2). Of particular interest is the Warrnambool extract, where unique molecular ions were identified at *m/z* 201.5 and *m/z* 229.5. For the seven specimens that did not have all their seeds unambiguously identified, it is currently thought that there was high intra specimen seed-to-seed biological variability.

For the results of the HPLC-UV, chemometrics, and LC-MS analyses to be developed into a forensic methodology, there are several critical issues that need addressing, including:

- The amount of seed that needs to be extracted to average out biological variation;
- The validation of the presence/absence of identified molecular ions as specimen biomarkers;
- Improving the resolution of the HPLC-UV chromatography for improved results from PCA; and
- Reanalysing the data using a supervised method such as Partial Least Squares – Discriminant Analysis (PLS-DA) as more data is collected.

Each of these issues are currently being addressed and form the basis of an ongoing work program. This will involve performing extractions on larger quantities of seeds; growing specimens in controlled greenhouse conditions to validate identified biomarkers; and moving LC-MS platforms to an Ultra High Pressure Liquid Chromatography triple quadrupole mass spectrometer utilising different solid phases (such as Hydrophobic Interaction Liquid Chromatography) to obtain improved UV signal resolution. This expanded work program will involve investigating many more specimens from many different locales, including extracts from eight specimens of the six known cultivars "*Carmencita*" (Zimbabwe), "*Dehradun*" (unknown African locality), "*Gibsonii*" (Zimbabwe), "*Impala*" (Tanzania), "*Sanguineus*" (Spain and Tanzania) and "*Zanzibariensis*" (Kenya and Tanzania). The increased data would then allow for the supervised PLS-DA method to be applied to the collected data.

The PCA on the ^{1}H NMR data of extracts of fourteen combined *R. communis* specimens was very encouraging. Given that for model three PC1 and PC2 accounted for 75% of the variance this also suggested that the final model constructed was quite strong. This analysis suggested that the relative amounts of **1** and **2/3** may well be important in determining specimen. Further research investigations are currently focused on expanding the number of specimens analysed by NMR and chemometrics. These will be performed in conjunction with the above HPLC-UV and LC-MS investigations. Collected NMR data will be analysed using both unsupervised (PCA) and supervised (PLS-DA) multivariate statistical analysis methods. It is also anticipated that one dimensional ^{13}C NMR data, and two dimensional heteronuclear based NMR pulse sequences such ^{1}H-^{13}C gHMQC will be employed to investigate the metabolome of seed extracts. Collection of NMR data on these extracts will be done in conjunction with the NMR facility at Bio21 Institute at The University of Melbourne. Analysis will be performed on an 800MHz NMR spectrometer with cryoprobe that will decrease acquisition times, and increase the sensitivity and resolution of the experiments. This will increase the amount of information collected during NMR analysis, hopefully leading to further refining and strengthening of the statistical models. Finally, using a high field strength NMR spectrometer may allow for an enhanced ability to detect biomarker compounds specific to a particular cultivar through sensitivity and resolution gains.

Funding has been secured (through a National Security Science and Technology grant) to continue investigating all aspects of the metabolome for cultivar determination. It has also allowed for an expansion of the program to include provenance determination. Some of the questions raised in the conclusions will be the primary focus of this NSST grant. It needs to be highlighted that as data is collected across multiple analysis platforms (NMR, HPLC-UV and LC-MS), it will be compared, contrasted and linked via multivariate statistical analysis methods. Combining data from multiple platforms is rare in the metabolomics literature.[79] Conducting the analysis this way will provide national and international forensic agencies a novel and unique way to analyse *R. communis* extracts for cultivar, and potentially provenance, determination.

4. Experimental

4.1 Chemicals

All solvents used were analytical grade. MeOH, acetone, MeCN and H_2O were purchased from Merck. Trifluoroacetic acid, acetic acid and formic acid were analytical grade and purchased from Sigma-Aldrich. Deuterated NMR solvents (D_2O, d_4-acetic acid, TSP, d_4-MeOH, d_6-DMSO) were supplied by Cambridge Isotopes.

4.2 General Experimental

HPLC-UV and LC-MS data were collected on 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, equipped with Agilent ChemStation LC for 3D software (Rev.A.09.03). The capillary was operated in positive-ion mode at a constant temperature of 350°C. The electrospray needle was held at +3500 V, the skimmer at +40 V and cap exit at +136 V. Octopole 1 and 2 were set at +12 V and +1.74 V respectively. The rf was set at 200 Vpp. Lenses 1 and 2 were set at -5 V and -60 V respectively. Nitrogen was used as the high-flow nebuliser gas at a pressure of 50 psi and the nitrogen drying gas was set at a temperature of 350°C with a flow rate of 12 L/min. Data was acquired in the range of *m/z* 100 – 1800.

NMR data was collected on a Bruker Avance (Bremen, Germany) NMR spectrometer operating at a ^{1}H NMR frequency of 500.13 MHz. The spectrometer was running Bruker Biospin Topspin 2.0 NMR software. The spectrometer was equipped with a standard geometry 5 mm diameter BBI probe head. Each sample was referenced to the internal standard TSP at δ 0.00 ppm.

4.3 Collection and 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 these investigations were conducted in a PC2 designated laboratory within a laminar flow cytotoxic drug cabinet. Staff performing extractions wore gowns, safety glasses and gloves during all extraction work.*

Collections of environmental samples of seed specimens of *R. communis* were made from various locations in Melbourne and eastern Australia. Plant morphologies (seed size and colour, leaf and stem colour, plant height and pod features), date of collection and GPS co-ordinates were recorded at the time of collection and details entered into the in-house database "Castorbase". From this library fourteen specimens were selected for metabolome analysis. The cultivars of the fourteen specimens of *R. communis* have not been established, hence selections were based on differences in plant morphology, and are assumed at this time to indicate unique cultivars.

For each specimen of *R. communis*, four mature seeds were selected and separately ground with a mortar and pestle. The individually ground seeds were agitated separately in 20 mL of acetone for 1 hour to remove the castor oil from the seed pulp. The acetone was removed via filtration (filter paper), and the seed mash washed twice with 25 mL aliquots of clean acetone to remove residual castor oil. The residual seed mash was allowed to air dry. Subsequently, 20 mL of 2% aqueous acetic acid solution added, and the resultant solution agitated for 2 hours. The aqueous acid solution was again filtered (filter paper) and the residual seed mash washed a further two times with approximately 5 mL of 2% acetic acid solution. The combined filtrate was then twice passed through a 30 kDa MWCO filter to remove both *R. communis* Agglutinin (RCA) and the ricin toxin from the aqueous acid extract. The combined <30 kDa MWCO fractions were then stored at -30°C until required for chemical analysis.

4.4 HPLC-UV data collection and multivariate statistical analysis

HPLC-UV data was collected on 20 µL aliquots of 20 mg/mL solutions of all fourteen specimen extracts. Aliquots were injected onto a Phenomonex Luna 5 µm, 50 x 2 mm, C18 reversed phase HPLC column at 25°C with gradient elution from 100% H_2O + 0.05% formic acid to 70:30 MeOH:H_2O (+ 0.05% formic acid) over 30 min, then to 100% MeOH + 0.05% formic acid over 1 min and held at this for 4 min. Blank injections and injections of the bradykinin standard (20 µL of a 5 mg/mL solution) were made using the same gradient conditions.

Collected HPLC-UV data at 254 nm was retention time corrected to the internal standard ricinine (retention time = 6.6 min) and binned via an in-house Microsoft Excel macro.[58] Each UV chromatogram was divided into 114 bins with a width of approximately 21 s/bin. A 224 x 114 data matrix was formed, which was standardised, normalised, autoscaled, and subjected to PCA using Minitab™.

4.5 NMR sample preparation and data collection

The four individual seed extracts from each specimen were combined, forming fourteen combined extracts for analysis. In addition, before they were combined for analysis, the Richmond and South Arm extracts were analysed individually to ascertain seed-to-seed biological variation. In total, 22 samples were subjected to ^{1}H NMR analysis. Each sample subjected to ^{1}H NMR analysis was made up to a concentration of 20 mg/mL in D_2O (with 0.1% TSP and 2% d_4-acetic acid). Solutions were vortexed for 30 s then centrifuged for 3 min. A 600 µL aliquot of each extract was transferred to a 5 mm NMR tube immediately prior to analysis. One dimensional "Watergate" ^{1}H NMR spectra were collected over an 11 ppm sweep width with 128 scans and 16k data points. The recycle delay time was set to 5 s, and the pulse width was 8.5 µs (90^o). Probe temperature was set to 298 K. Processing of the Free Induction Decay (FID) was performed with line broadening (LB) set to 1.0 Hz, and linear prediction (SI) set to 32k. Each spectrum was phased and baseline corrected.

4.6 NMR multivariate statistical analysis

Phased and baseline corrected ^{1}H NMR data was binned within the statistical program Matlab™ using the PLStoolbox with the ProMetab script.[78] Bin widths were set at δ 0.002 ppm, with three models built. Model one and two were constructed over a chemical shift range 0.5 ppm to 10.0 ppm. Model one data was normalised and mean-centered, while for model two a log function was also applied. Model three was constructed over the chemical shift range 5.6 ppm to 8.2 ppm. The binned data was normalised, mean-centered and a log function applied.

5. Acknowledgments

We would like to thank Alex Theo and Justin Doward for their collective efforts toward writing the Microsoft Excel macros used during these studies. We also would like to thank Stuart Thomson and Eloise Pigott for their forthright analysis on aspects of the work, their clear thinking and valuable input into analysis. We would also like to thank DSTO for the financial support given to Ben Gordon during his Honours year.

6. References

1. Harvey, J. H., in *John Evelyn's "Elysium Britannicum" and European gardening*, (Ed. O'Malley, T., Wolschke-Bulmahn, J.) **1947**, Vol 17, pp. 1787-1949.

2. Knight, J., in *Guides to growers*, (Ed. Department of Agriculture, Victoria) **1894**, No 13, pp. 1-8.

3. Olsnes, S. *Toxicon.* **2004**, *44*, 361-370.

4. Millard, C. B., LeClaire, R. D., in *Chemical warfare agents: chemistry, pharmacology, toxicology, and therapeutics*, (Ed. Romano, J. A., Lukey, B. J., Salem, H.) **2007**, Ch. 17, pp. 432-452.

5. Halliday, C. Cutting edge hair, *The Age* 30th September **2007.**

6. Internet database: Oil World, Oil world annual 2006. *Oil World* http://www.oilworld.biz/app.php?fid=1060&fpar=0&isSSL=0&aps=0&blub=65707b39e9020b063148896592bf6f2e **2006**, last accessed April 30 2009.

7. Internet database: Food and Agriculture Organization of the United Nations, FAOSTAT Agricultural data. In http://faostat.fao.org/site/567/default.aspx#ancor, last accessed April 30 2009.

8. Internet database: Royal Botanic Gardens World checklist series. http://www.kew.org/wcsp/qsearch.do, last accessed April 30 2009.

9. Weiss, E. A., *Castor sesame & safflower*. Leonard Hill Books: London, **1971**.

10. 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 April 30 2009.

11. Department of Foreign Affairs and Trade, Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction (Chemical Weapons Convention) - Annex on Chemicals.

12. Khan, A. S.; Morse, S.; Lillibridge, S. *The Lancet* **2000**, *356*, 1179-1182.

13. Rotz, L. D.; Khan, A. S.; Lillibridge, S. R.; Ostroff, S. M.; Hughes, J. M. *Emerg. Infect. Dis.* **2002**, *8*, 225-230.

14. Fu, T.; Burbage, C.; Tagge, E. P.; Brothers, T.; Willingham, M. C.; Frankel, A. E. *Int. J. Immunopharmaco.* **1996**, *18*, 685-692.

15. Doan, L. G. *Clin. Toxicol.* **2005**, *42*, 201-208.

16. Endo, Y.; Tsurugi, K.; Yutsudo, T.; Takeda, Y.; Ogasawara, T.; Igarashi, K. *Eur. J. Biochem.* **1988**, *171*, 45-50.

17. Chen, C.; Jiang, L.; Michalczyk, R.; Russu, I. M. *Biochem.* **2006**, *45*, 13606-13613.

18. Endo, Y.; Tsurugi, K. *J. Biol. Chem.* **1987**, *262*, 8128-8130.

19. Endo, Y.; Mitsui, K.; Motizuki, M.; Tsurugi, K. *J. Biol. Chem.* **1987**, *262*, 5908-5912.

20. Frigerio, L.; Roberts, L. *J. Exp. Bot.* **1998**, *49*, 1473-1480.

21. Koszlov, J. V.; Sudarkina, O. J.; Kurmanova, A. G. *Mol. Biol.* **2006,** *40*, 635-646.

22. Crompton, R.; Gall, D. *Med. Leg. J.* **1980**, *48*, 51-62.

23. Knight, B. *Br. Med. J.* **1979**, *1*, 350-351.

24. Gibson, J.; Drociuk, D.; Fabian, T.; Brundage, S. *Morbid Mortal Wkly. rep.* **2003**, *52*, 1129-1131.

25. Schier, J. G.; Patel, M. M.; Belson, M. G.; Patel, A.; Schwartz, M.; Fitzpatrick, N.; Drociuk, D.; Deitchman, S.; Meyer, R.; Litovitz, T.; Watson, W. A.; Rubin, C. H.; Kiefer, M. *Am. J. Public Health* **2007**, *97*, (Supplement 1), S152.

26. Hotlermann, K.; Hughes, C. A.; Asamoah, S.; Bell, L.; Giovachino, M.; Morrow, R.; Trabert, E., Response to a ricin incident; Guidelines for Federal, State, and Local Public Health and Medical Officials In Services, D. o. H. a. H., Ed. USA, 2006.

27. www.anthraxinvestigation.com/ricin.html, last accessed April 30 2009.

28. CNN, Grand jury indicts man for ricin possession. In CNN.com international **2008**.

29. From Correspondents in Las Vegas, Man critical after deadly ricin poisoning. *Herald Sun* **2008**.

30. Ritter, K., Guns, 'anarchist book' found with ricin In Press, T. A., Ed. ABC news **2008**.

31. Associated Press, Ricin scare closes Phoenix hospital In Freedom communications, INC. : **2008**.

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

33. Benson, S.; Lennard, C.; Maynard, P.; Roux, C. *Forensic Sci. Int.* **2006**, *157*, 1-22.

34. Hays, P. A.; Remaud, G. S.; Jamin, E.; Martin, Y. L. *J. Forensic Sci.* **2000**, *45*, 552-562.

35. Christoph, N.; Rossmann, A.; Schlicht, C.; Voerkelius, S. *A.C.S. symposium series* **2007**, *952*, 166-179.

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

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

38. Choi, Y. H.; Kim, H. K.; Hazekamp, A.; Erkelens, C.; Lefeber, A. W. M.; Verpoorte, R. *J. Nat. Prod.* **2004**, *67*, 953-957.

39. Brescia, M.; Alviti, G.; Liuzzi, V.; Sacco, A. *J. Am. Oil Chem. Soc.* **2003**, *80*, 945-950.

40. Hamzehzarghani, H.; Kushalappa, A. C.; Dion, Y.; Rioux, S.; Comeau, A.; Yaylayan, V.; Marshall, W. D.; Mather, D. E. *Physiol. Mol. Plant Pathol.* **2005**, *66*, 119-133.

41. Makris, D. P.; Kallithraka, S.; Mamalos, A. *Talanta* **2006**, *70*, 1143-1152.

42. Sato, S.; Soga, T.; Nishioka, T.; Tomita, M., *Plant J.* **2004**, 40, 151-163.

43. Yamazaki, M.; Nakajima, J.; Yamanashi, M.; Sugiyama, M.; Makita, Y.; Springob, K.; Awazuhara, M.; Saito, K., *Phytochemistry* **2003**, 62, 987-995.

44. Van Nederkassel, A. M.; Xu, C. J.; Lancelin, P.; Sarraf, M.; MacKenzie, D. A.; Walton, N. J.; Bensaid, F.; Lees, M.; Martin, G. J.; Desmurs, J. R.; Massart, D. L.; Smeyers-Verbeke, J.; Vander Heyden, Y. *J. Chromatogr. A* **2006**, *1120*, 297-298.

45. Burns, J.; Fraser, P. D.; Bramley, P. M. *Phytochemistry* **2003**, *62*, 939-947.

46. Defernez, M.; Gunning, Y. M.; Parr, A. J.; Shepherd, L. V.; Davies, H. V.; Colquhoun, I. J. *J. Agric. Food Chem.* **2004**, *52*, 6075-6085.

47. Cichelli, A.; Pertesana, G. P. *J. Chromatogr. A* **2004**, *1046*, 141-146.

48. Mohana, M.; Reddy, K.; Jayshanker, G.; Suresh, V.; Sarin, R. K.; Sashidhar, R. B. *J. Sep. Sci.* **2005**, *28*, 1558-1565.

49. Reddy, M. M.; Ghosh, P.; Rasool, S. N.; Sarin, R. K.; Sashidhar, R. B. *J. Chromatogr. A* **2005**, *1088*, 158-168.

50. Schaneberg, B. T.; Crockett, S.; Bedir, E.; Khan, I. A. *Phytochemistry* **2003**, *62*, 911-918.

51. Nielsen, N.-P. V.; Carstensen, J. M.; Smedsgaard, J. *J. Chromatogr. A* **1998**, *805*, 17-35.

52. Lindon, J. C.; Holmes, E.; Nicholson, J. K. *Prog. NMR Spectro.* **2001**, *39*, 1-40.

53. Smilde, A. K.; van der Werf, M. J.; Bijlsma, S.; van der Werff-van der Vat, B. J.; Jellema, R. H. *Anal. Chem.* **2005**, *77*, 6729-6736.

54. Scholz, M.; Kaplan, F.; Guy, C. L.; Kopka, J.; Selbig, J. *Bioinformatics* **2005**, *21*, 3887-3895.

55. Grung, B.; Manne, R. *Chemometr. Intell. Lab. Syst.* **1998**, *42*, 125-139.

56. (a) Kang, S. S.; Cordell, G. A.; Soejarto, D. D.; Fong, H. H. S. *J. Nat. Prod* **1985**, *48*, 155-156; (b) L. Skursky, D. Burleson, G. R. Waller, *J. Biol. Chem.* **1969**, *244*, 3238-3242.

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

58. Theo, A. Microsoft™ Excel Macro, **2008**.

59. Dunn, W. B.; Ellis, D. I., *TrAC* **2005**, *24*, 285-294.

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

61. Bourne, D. J. personal communication.

62. Maraschin, R. P.; Ianssen, C.; Arsego, J. L.; Capel, L. S.; Dias, P. F.; Cimadon, A. M. A.; Zanus, C.; Caro, M. S. B.; Maraschin, M. *Spec. Publ. - R. Soc. Chem. FIELD Full Journal* **2002**, *286*, 255-260.

63. Defernez, M.; Colquhoun, I. J. *Phytochemistry* **2003**, *62*, 1009-1017.

64. Seger, C.; Sturm, S. *J. Proteome Res.* **2007**, *6*, 480-497.

65. Ward, J. L.; Baker, J. M.; Beale, M. H. *Febs J.* **2007**, *274*, 1126-1131.

66. Van der Kooy, F.; Verpoorte, R.; Marion Meyer, J. J. *S. Afr. J. Bot.* **2008**, *74*, 186-189

67. Bailey, N. J.; Oven, M.; Holmes, E.; Nicholson, J. K.; Zenk, M. H. *Phytochemistry* **2003**, *62*, 851-858.

68. Bailey, N. J.; Sampson, J.; Hylands, P. J.; Nicholson, J. K.; Holmes, E. *Planta Med.* **2002**, *68*, 734-738.

69. Belton, P. S.; Colquhoun, I. J.; Kemsley, E. K.; Delgadillo, I.; Roma, P.; Dennis, M. J.; Sharman, M.; Holmes, E.; Nicholson, J. K.; Spraul, M. *Food Chem.* **1998**, *61*, 207-213.

70. Ott, K. H.; Aranibar, N.; Singh, B.; Stockton, G. W. *Phytochemistry* **2003**, *62*, 971-985.

71. Mannina, L.; Patumi, M.; Proietti, N.; Bassi, D.; Segre, A. L. *J. Agric. Food Chem.* **2001**, *49*, 2687-2696.

72. D'Imperio, M.; Mannina, L.; Capitani, D.; Bidet, O.; Rossi, E.; Bucarelli, F. M.; Quaglia, G. B.; Segre, A. *Food Chem.* **2007**, *105*, 1256-1267.

73. Watson, D. G.; Peyfoon, E.; Zheng, L.; Lu, D.; Seidel, V.; Johnston, B.; Parkinson, J. A.; Fearnley, J. *Phytochem. Anal.* **2006**, *17*, 323-331.

74. Pereira, G. E.; Gaudillere, J. P.; VanLeeuwen, C.; Hilbert, G.; Lavialle, O.; Maucourt, M.; Deborde, C.; Moing, A.; Rolin, D. *J. Agric. Food Chem.* **2005**, *53*, 6382-6389.

75. Periera, G. E.; Gaudillere, J. P.; Leeuwen, C. V.; Hilbert, G.; Maucourt, M.; Deborde, C.; Moing, A.; Rolin, D., *Anal. Chim. Acta* **2006**, 346-352.

76. Brescia, M. A.; Caldarola, V.; De Giglio, A.; Benedetti, D.; Fanizzi, F. P.; Sacco, A. *Anal. Chim. Acta* **2002**, *458*, 177-186.

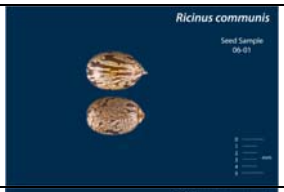
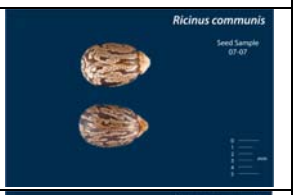
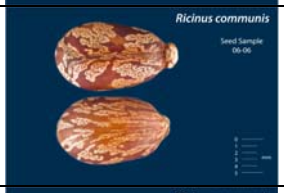
77. Ward, J. L.; Harris, C.; Lewis, J.; Beale, M. H. *Phytochemistry* **2003**, *62*, 949-957.

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

79. Rochfort, S. *J. Nat. Prod.* **2005**, *68*, 1813-1820

Appendix A: Seed Image and Location

Below is a table containing details of specimens of *R. communis* that were studied, including DSTO code numbers, geographic location and seed image.

Collection Code	Geographic Location	Seed Image	Collection Code	Geographic Location	Seed Image
06-01	Brisbane, Qld		07-07	Magnetic Island, Qld	
06-02	Coopers Plains Qld		07-11	Clifton Hill, Vic	
06-06	Warrnambool, Vic		07-17	Braybrook, Vic	
06-07	Laverton, Vic		07-19	Avondale Heights, Vic	
06-09	Newcastle, NSW		08-02	Richmond, Vic	
07-03	Footscray A, Vic		08-23	Obi Obi, Qld	
07-05	Footscray B, Vic		08-28	South Arm, NSW	

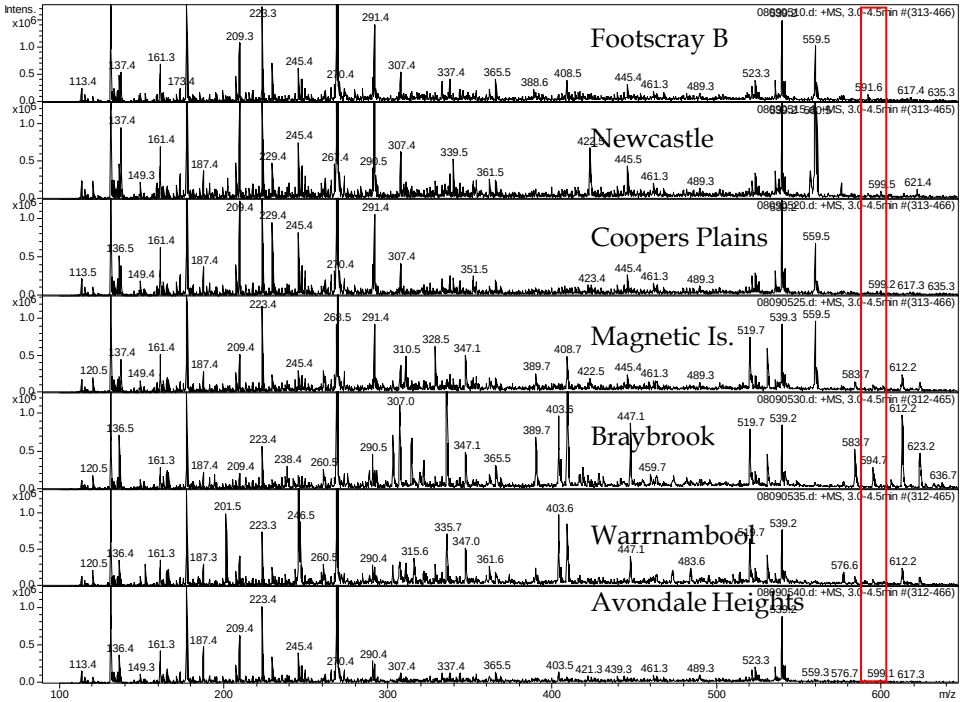
Appendix B: HPLC-UV Bin number and PC

Bin number, corresponding retention time, and the principal component that best explain each bin are listed in the Table below.

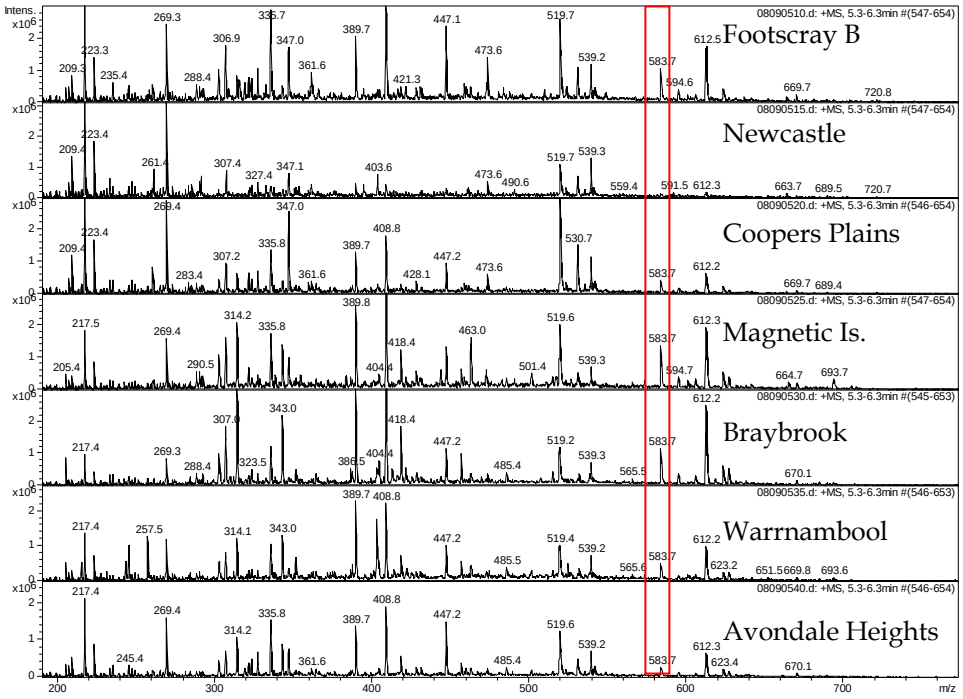
Bin retention times and Principal Components

Bin Number	start time (mm:ss)	end time (mm:ss)	Bin centre (mm:ss)	PC
1	00:00	00:21	00:11	3
2	00:21	00:42	00:32	3
3	00:42	01:03	00:53	3
4	01:03	01:24	01:14	3
5	01:24	01:45	01:35	3
6	01:45	02:06	01:56	5
7	02:06	02:27	02:17	5
8	02:27	02:48	02:38	5
9	02:48	03:09	02:59	4
10	03:09	03:31	03:20	3,5
11	03:31	03:52	03:41	3
12	03:52	04:13	04:02	3
13	04:13	04:34	04:23	3
14	04:34	04:55	04:44	4
15	04:55	05:16	05:05	4
16	05:16	05:37	05:26	3
17	05:37	05:58	05:47	3
18	05:58	06:19	06:08	3,4
19	06:19	06:40	06:29	4
20	06:40	07:01	06:51	4
21	07:01	07:22	07:12	3,4
22	07:22	07:43	07:33	3
23	07:43	08:04	07:54	3
24	08:04	08:25	08:15	3
25	08:25	08:46	08:36	3
26	08:46	09:07	08:57	3,4
27	09:07	09:28	09:18	3
28	09:28	09:49	09:39	3
29	09:49	10:11	10:00	2,4
30	10:11	10:32	10:21	4
31	10:32	10:53	10:42	3
32	10:53	11:14	11:03	3
33	11:14	11:35	11:24	3
34	11:35	11:56	11:45	3
35	11:56	12:17	12:06	3
36	12:17	12:38	12:27	3
37	12:38	12:59	12:48	5
38	12:59	13:20	13:09	5
39	13:20	13:41	13:31	4
40	13:41	14:02	13:52	2,5
41	14:02	14:23	14:13	3,5
42	14:23	14:44	14:34	3
43	14:44	15:05	14:55	2,3
44	15:05	15:26	15:16	2,3
45	15:26	15:47	15:37	2,3
46	15:47	16:08	15:58	2
47	16:08	16:29	16:19	2
48	16:29	16:51	16:40	2
49	16:51	17:12	17:01	2
50	17:12	17:33	17:22	2
51	17:33	17:54	17:43	2
52	17:54	18:15	18:04	2
53	18:15	18:36	18:25	2
54	18:36	18:57	18:46	2
55	18:57	19:18	19:07	2
56	19:18	19:39	19:28	3
57	19:39	20:00	19:49	2
58	20:00	20:21	20:11	2
59	20:21	20:42	20:32	2
60	20:42	21:03	20:53	2
61	21:03	21:24	21:14	2
62	21:24	21:45	21:35	1,2
63	21:45	22:06	21:56	1,2
64	22:06	22:27	22:17	1,2
65	22:27	22:48	22:38	1,2
66	22:48	23:09	22:59	1,2
67	23:09	23:31	23:20	1,2
68	23:31	23:52	23:41	1,2
69	23:52	24:13	24:02	1,2
70	24:13	24:34	24:23	1
71	24:34	24:55	24:44	1
72	24:55	25:16	25:05	1
73	25:16	25:37	25:26	1
74	25:37	25:58	25:47	1
75	25:58	26:19	26:08	1
76	26:19	26:40	26:29	1
77	26:40	27:01	26:51	1
78	27:01	27:22	27:12	1
79	27:22	27:43	27:33	1
80	27:43	28:04	27:54	1
81	28:04	28:25	28:15	1
82	28:25	28:46	28:36	1
83	28:46	29:07	28:57	1
84	29:07	29:28	29:18	1
85	29:28	29:49	29:39	1
86	29:49	30:11	30:00	1
87	30:11	30:32	30:21	1
88	30:32	30:53	30:42	1
89	30:53	31:14	31:03	1
90	31:14	31:35	31:24	1
91	31:35	31:56	31:45	1
92	31:56	32:17	32:06	1
93	32:17	32:38	32:27	1
94	32:38	32:59	32:48	1
95	32:59	33:20	33:09	1
96	33:20	33:41	33:31	1
97	33:41	34:02	33:52	1
98	34:02	34:23	34:13	1
99	34:23	34:44	34:34	1
100	34:44	35:05	34:55	5
101	35:05	35:26	35:16	5
102	35:26	35:47	35:37	1
103	35:47	36:08	35:58	1
104	36:08	36:29	36:19	1
105	36:29	36:51	36:40	1
106	36:51	37:12	37:01	1
107	37:12	37:33	37:22	1
108	37:33	37:54	37:43	1
109	37:54	38:15	38:04	1
110	38:15	38:36	38:25	1
111	38:36	38:57	38:46	1,6
112	38:57	39:18	39:07	6
113	39:18	39:39	39:28	6
114	39:39	40:00	39:49	6

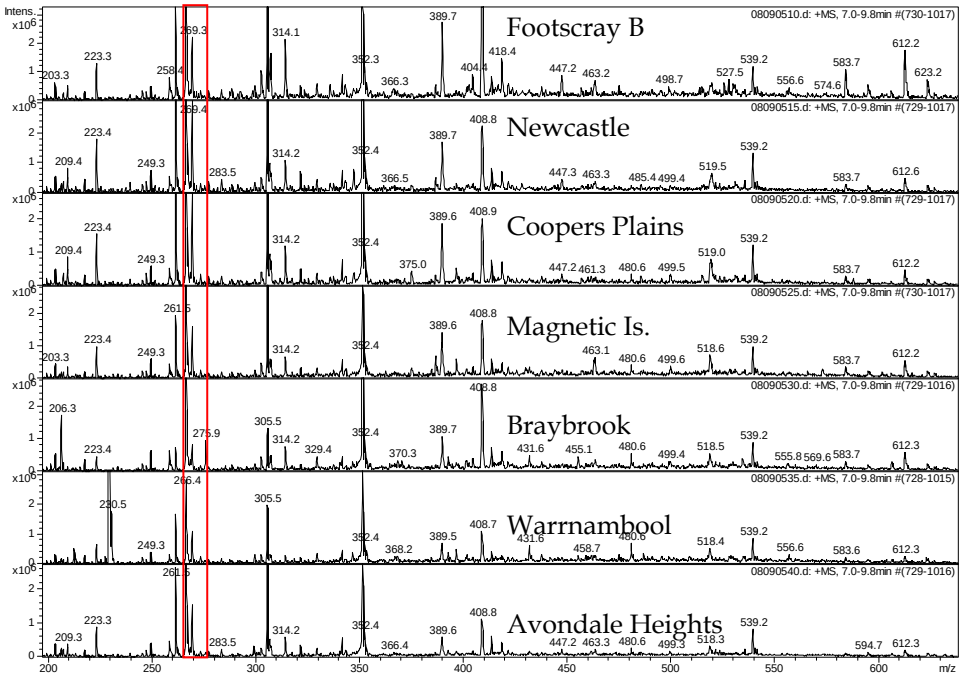
Appendix C: Extracted Mass Spectra



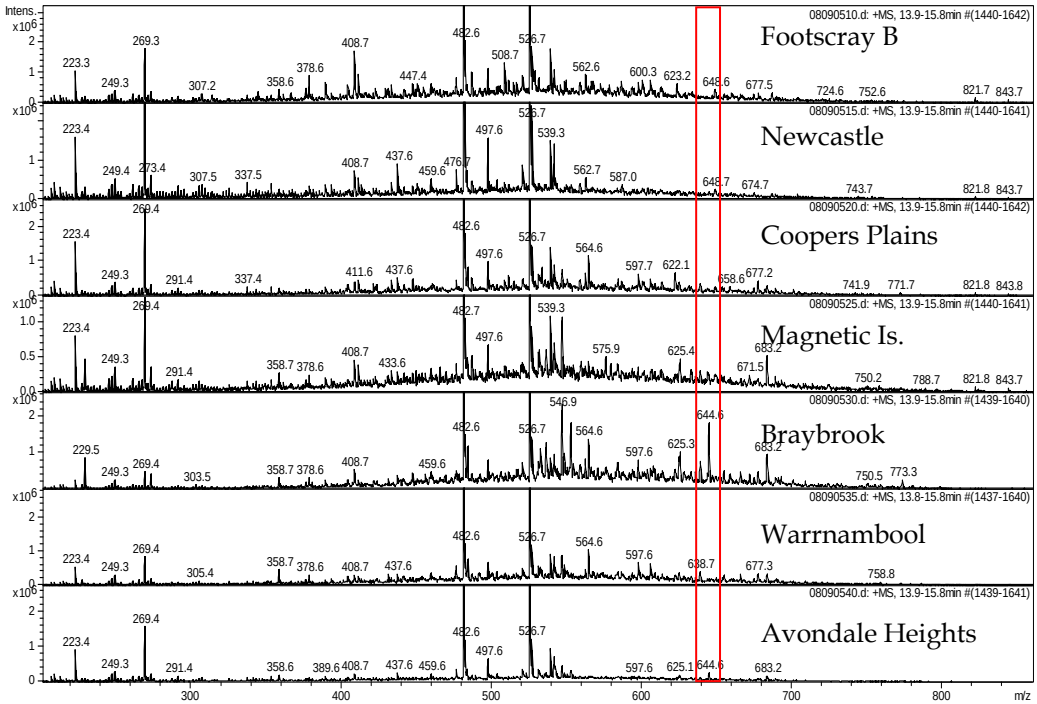
Extracted mass spectra for bins 10 to 13 for the seven analysed *R. communis* specimens. Highlighted is the observed compound in the Braybrook extract at *m/z* 594.6^{2+}.



Extracted mass spectra for bins 16 to 18 for the seven analysed *R. communis* specimens. Highlighted is the compound at *m/z* 583.7, which is absent in the Newcastle metabolome.



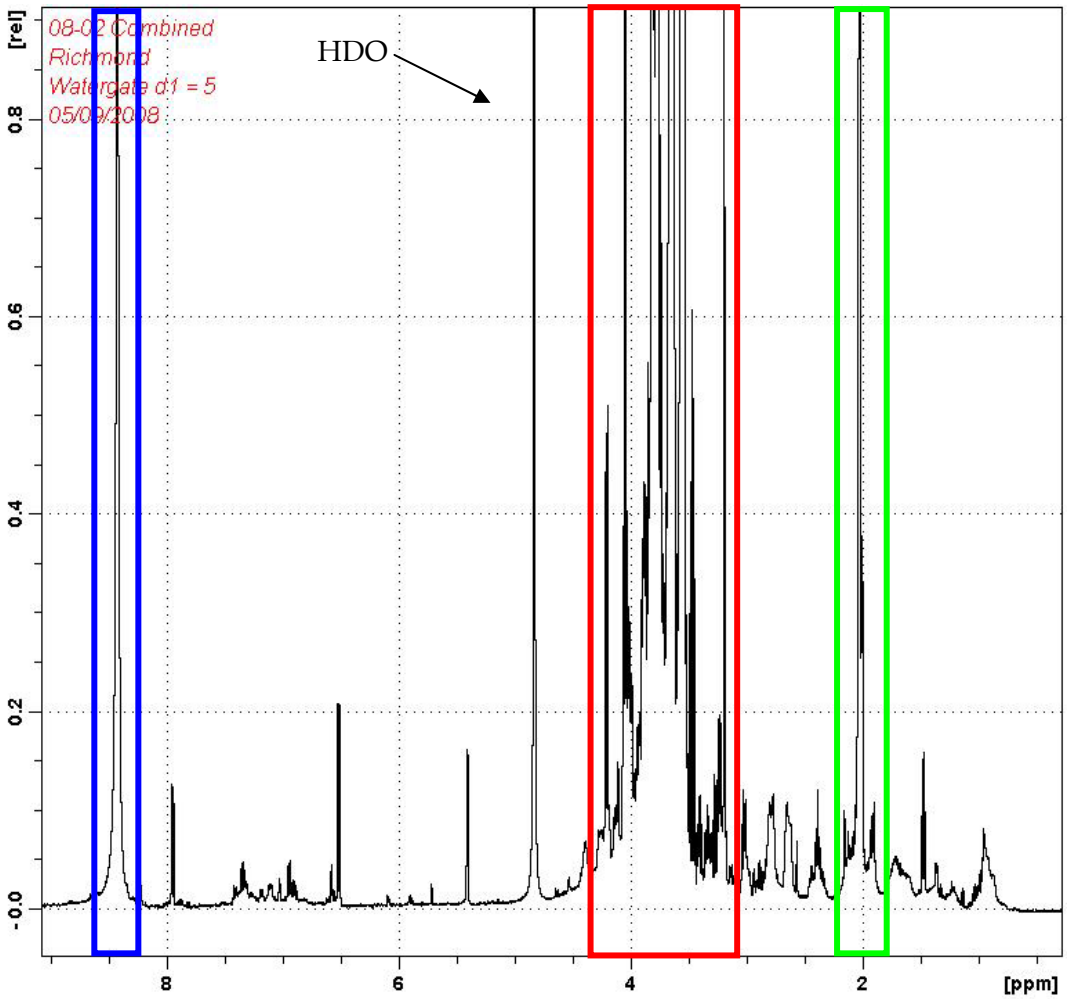
Extracted mass spectra for bins 21 to 28 for the seven analysed *R. communis* specimens. Highlighted is the compound at *m/z* 261.5. This compound is significantly reduced in the Braybrook metabolome.



Extracted mass spectra for bins 41 to 45 for the seven analysed *R. communis* specimens. Highlighted is the compound at *m/z* 644.6. This compound has a significant presence in the Braybrook metabolome, a very minor presence in the Avondale Heights metabolome, and is not present in any other metabolome.

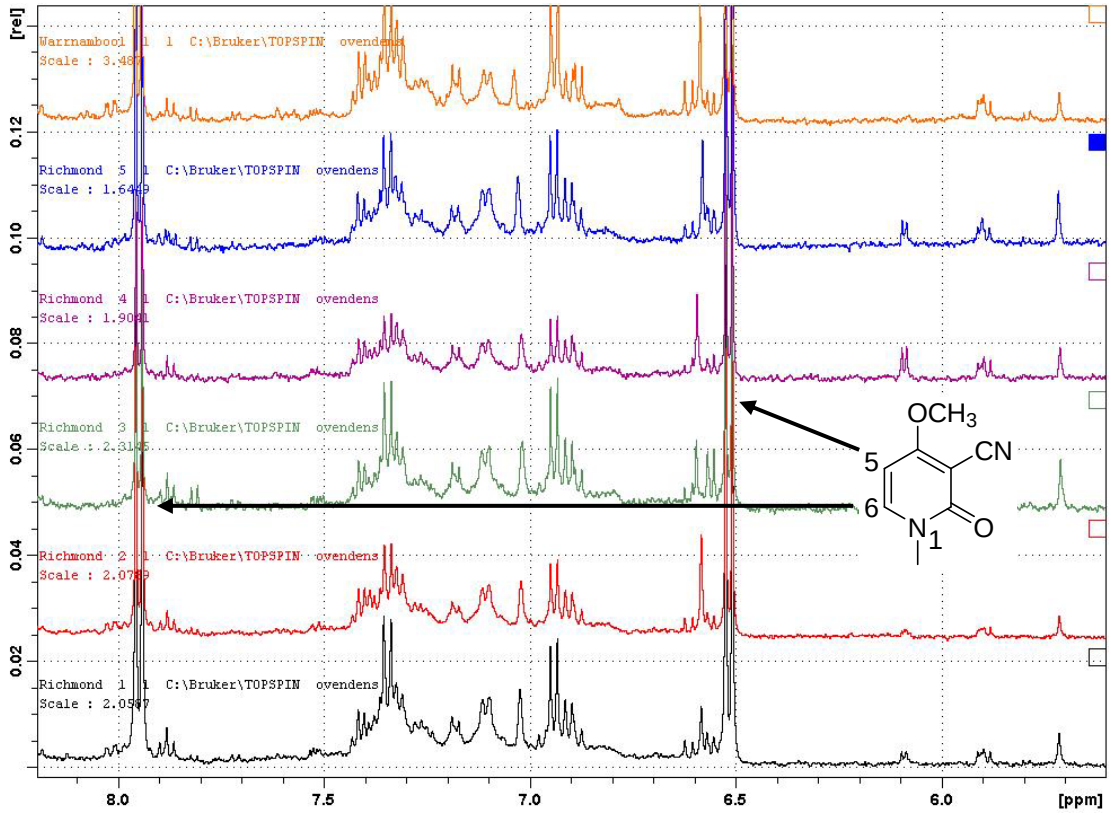
Appendix D: Typical ^{1}H NMR spectrum

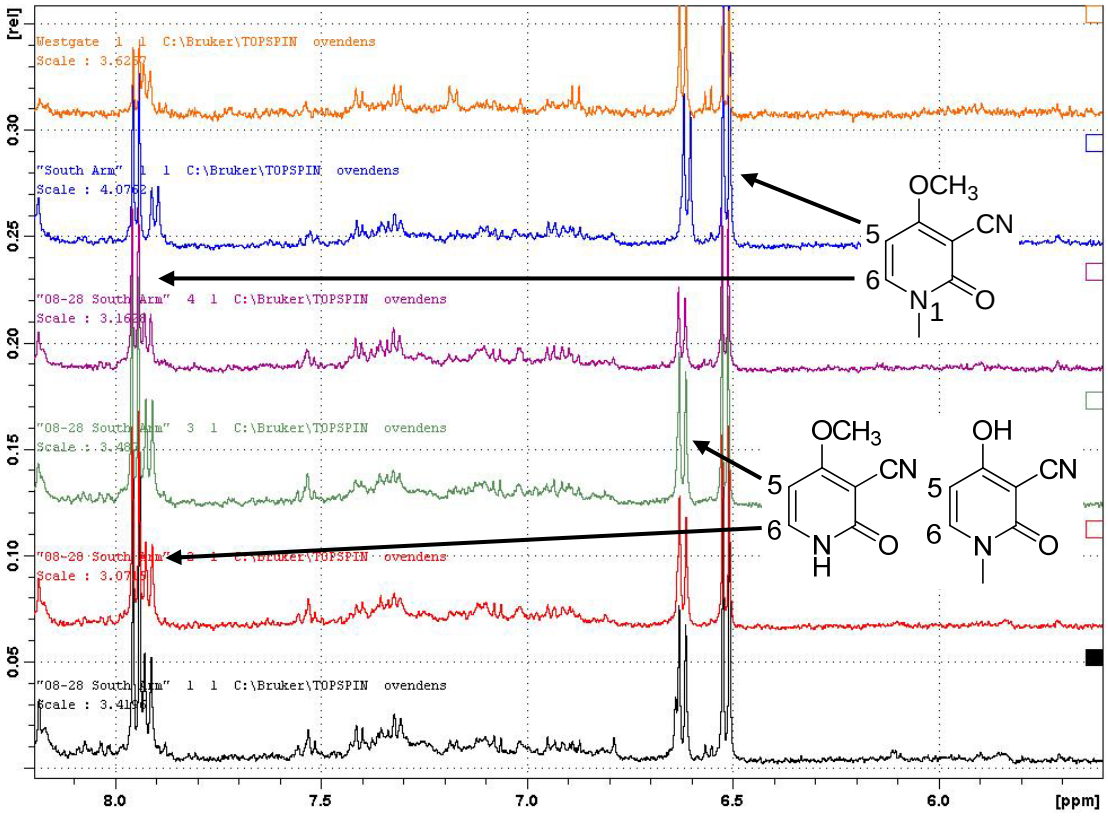
Below highlights the areas of the ^{1}H NMR spectrum that contained compounds that influenced the PCA. The blue box indicates the presence of residual formic acid, the red box indicates resonances due to the presence of sugar and inositol moieties in the extract, while the green box indicates the presence of residual acetic acid.



Appendix E: ^{1}H NMR stack plots

Below is an expansion and stack plot of the ^{1}H NMR spectrum between δ 5.6 and δ 8.2 that was the focus of model 3. The first is a comparison of the individual and combined Richmond extracts with the combined Warrnambool extract. Highlighted are the resonances for H-5 and H-6 of ricinine. The second is a comparison of the individual and combined South Arm extracts with the combined Laverton extract, highlighting H-5 and H-6 for the demethyl analogues. What is clearly evident in these spectra is the significant reduction in the amount of demethyl analogues of ricinine in the Richmond and Warrnambool extracts.




[rel]
0.30
0.25
0.20
0.15
0.10
0.05
8.0
7.5
7.0
6.5
6.0
[ppm]
OCH3
CN
OH
5
6
N
H
O

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION
DOCUMENT CONTROL DATA

1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)

2. TITLE

Cultivar Determination of *Ricinus communis* Via the Metabolome: a Proof of Concept Investigation

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, Benjamin R. Gordon, Christina K. Bagas, Bob Muir, Simone Rochfort and David J. Bourne

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-2338	AR-014-609	Technical Report	August 2009

8. FILE NUMBER	9. TASK NUMBER	10. TASK SPONSOR	11. NO. OF PAGES	12. NO. OF REFERENCES
2009/1037961	07/086	CTSTC	46	79

13. DOWNGRADING/DELIMITING INSTRUCTIONS

To be reviewed three years after date of publication.

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 http://web-vic.dsto.defence.gov.au/workareas/library/resources/dsto_thesaurus.shtml

Chemical analysis, Chromatography, Research, Spectrometry, Spectroscopy

19. ABSTRACT

Investigations were undertaken to ascertain the appropriateness of studying the metabolome of *Ricinus communis* for cultivar and provenance determination. Seeds from fourteen *R. communis* specimens (a total of 56 seeds) collected from the east coast of Australia were analysed by various analytical chemistry methods. The data collected from these analyses were then analysed using Principal Component Analysis. The outcomes from these investigations are discussed in this technical report.