

Extraction of *Sclerocarya birrea* samples for NMR and LC-MS

Methanol (d₄) (400 µl) was added to 15 mg of the dried leaf material, and this was vortexed for 1 min and sonicated for 20 min. The mixture was centrifuged at 10,000 rpm for 10 min and the supernatant (200 µl) was transferred into 3-mm NMR tubes. After the NMR analysis, the methanol-d₄ *S. birrea* extracts were diluted to 25% (v/v) ultrapure water. The diluted extracts were sonicated, centrifuged and filtrated before LC-MS analysis.

NMR analysis

¹HNMR spectra were acquired on a Bruker 600 MHz spectrometer (Wageningen University, Netherlands) operating at a proton NMR frequency of 600.13 MHz. A PRESAT pulse sequence was applied to suppress the water signal. For each sample, 64 K data points for the FID and 64 data points after Fourier transformation were used, 128 scans were acquired with 60 degree pulses at an acquisition time of 1.817 s. The spectral width was 18028 Hz with a relaxation delay of 4 s, and 0.2 Hz line broadening was used before Fourier transformation.

Data analysis

All NMR spectra were phase- and baseline corrected using MestReNova 10.01 (Mestrelab Research). Normalization was also done with MestReNova to ensure that all the spectra have the uniform norm. The regions containing the water and methanol peaks (4.60-5.00 ppm and 3.28-3.36 ppm) were excluded and the remaining spectral regions were divided into 0.04-ppm bins, converted to ASCII format, and imported to Microsoft Excel 2010.

Multivariate statistical analysis for NMR

The resulting Microsoft Excel file was imported to SIMCA version 13.0 (Umetrics) for multivariate data analysis. All imported data were Pareto-scaled for multivariate data analysis. Pareto scaling, in which each variable is divided by the square root of the standard deviation, gives greater weight to NMR data variables with higher intensities but is not as extreme as the use of no scaling [20]. Principal Component Analysis (PCA) was initially performed to determine the basic variation, the presence of extreme outliers and to obtain an overview of variation among the groups. The quality of the model was described by R² and Q² values. R² is defined as the proportion of variance in the data explained by the model, and indicates the goodness of fit and Q² is defined as the proportion of variance in the data predictable by the model and indicates predictability [21].

Untargeted metabolite profiling

Quantification of metabolites was achieved by 600 MHz library of Chenomx NMR suite version 8.2 which uses the concentrations of the known reference signals to determine the concentrations of individual compounds. The library contains 338 metabolites. Other metabolites were identified using published literature on compounds isolated from the leaves of *S. birrea*.

LC-MS analysis

The extracts were analysed by reverse phase LC-MS for their metabolite contents. The LC-QTOF MS analysis was carried out in Electrospray (ESI) negative mode. The mass rate and scan rate were set to record m/z 100-1500/s.

Multivariate statistical analysis for LC-MS

Quantification of *S. birrea* metabolite profiles of LC-MS analyses

was performed using XCMS online data analysis software, which can be accessed freely from the Scripps Centre for Metabolomics (<https://xcmsonline.scripps.edu>). This software approach employs peak alignment, matching, and comparison. The raw LC-MS files were first converted to MZxML files using MS convert. Peaks were subsequently extracted using XCMS. The final statistical analysis was performed by employing MetaboAnalyst 3.0-a comprehensive tool suite for metabolomic data analysis, which also can be accessed freely from <http://www.metaboanalyst.ca/MetaboAnalyst>. This software employs data normalization, multivariate analysis methods (PCA and PLS-DA), clustering analysis and many more analysis approaches. All the samples were auto scaled prior to the analysis.

Results

¹H NMR analysis

Figure 1 shows representative one-dimension ¹H NMR spectra of *Sclerocarya birrea* from Limpopo, Gauteng, North West, KwaZulu-Natal and Mpumalanga. No clear visual differences were observed in the overall spectroscopic fingerprint among *S. birrea* samples from five provinces. However, close inspection of the spectra revealed that the chemical compositions of metabolites between *S. birrea* samples obtained from five provinces were distinctly different. Visual inspection of the five spectra also revealed more intense sugar relative to aliphatic and aromatic compounds. Very few metabolites were annotated putatively using 600 MHz library from Chenomx NMR Suite version 8.2 which contains 338 metabolites, together with published literature [16,22] on compounds isolated from plant leaves. Intense signals in the sugar region from Gauteng (Figure 1) were however, clearly observed, indicating the presence of glycosylated metabolites and free sugars.

Metabolomic analysis of *Sclerocarya birrea* extracts from different provinces

To provide comparative interpretation and visualization of the fundamental variations and metabolite differences in the samples from different origins, Principal Component Analysis (PCA) and Orthogonal Partial Least Square-Discriminant Analysis (OPLS-DA) were applied to the ¹H NMR spectrum dataset. The PCA model for distinguishing *S. birrea* samples obtained from five provinces revealed R²X and Q² values of 0.85 and 0.78 respectively (Figure 2a). The PCA scores were used to determine whether metabolic fingerprints of *S. birrea* extracts were sufficiently different to distinguish geographical origin and to identify specific markers of each origin. To improve the clustering and to identify the metabolites responsible for the differences between the provinces, OPLS-DA model was constructed (Figure 2b), and revealed R²Y and Q² values of 0.73 and 0.62 respectively. The samples from five provinces were clearly separated in the OPLS-DA score plot (Figure 2b), which means that the non-correlated variation in X metabolites to Y metabolites was removed, resulting in maximum separation [23]. This maximum separation also indicates that the levels of *S. birrea* metabolites were strongly influenced by the various growing conditions in each region [24].

To further understand, the underlying variables contributing to the differentiation, the contribution plots were constructed. Figures 3a-3e shows the metabolites responsible for the separation observed in the OPLS-DA. All provinces showed a unique pattern of compounds important in their composition. KwaZulu-Natal samples were characterized by the higher levels of phenolics; quercetin 3-O-β-D glucopyranoside (bin at 4.19, 4.23, 5.41, 6.17, 6.36, 6.71, 7.42 and 7.57 ppm), gallic acid (bin at 7.04 ppm), quercetin

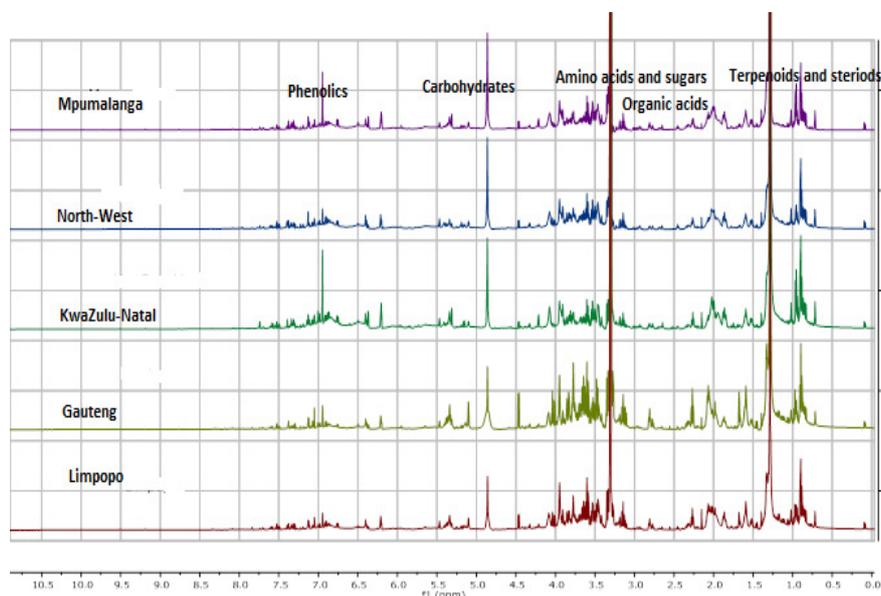


Figure 1: Average NMR spectra of *Sclerocarya birrea* from five provinces.

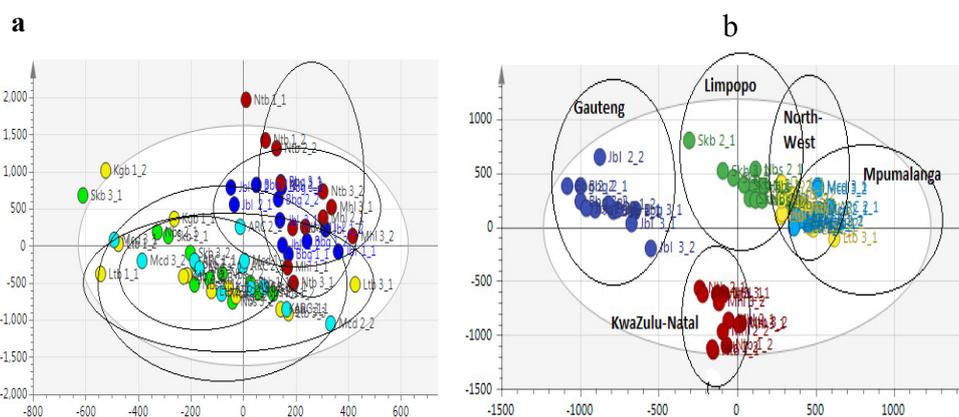
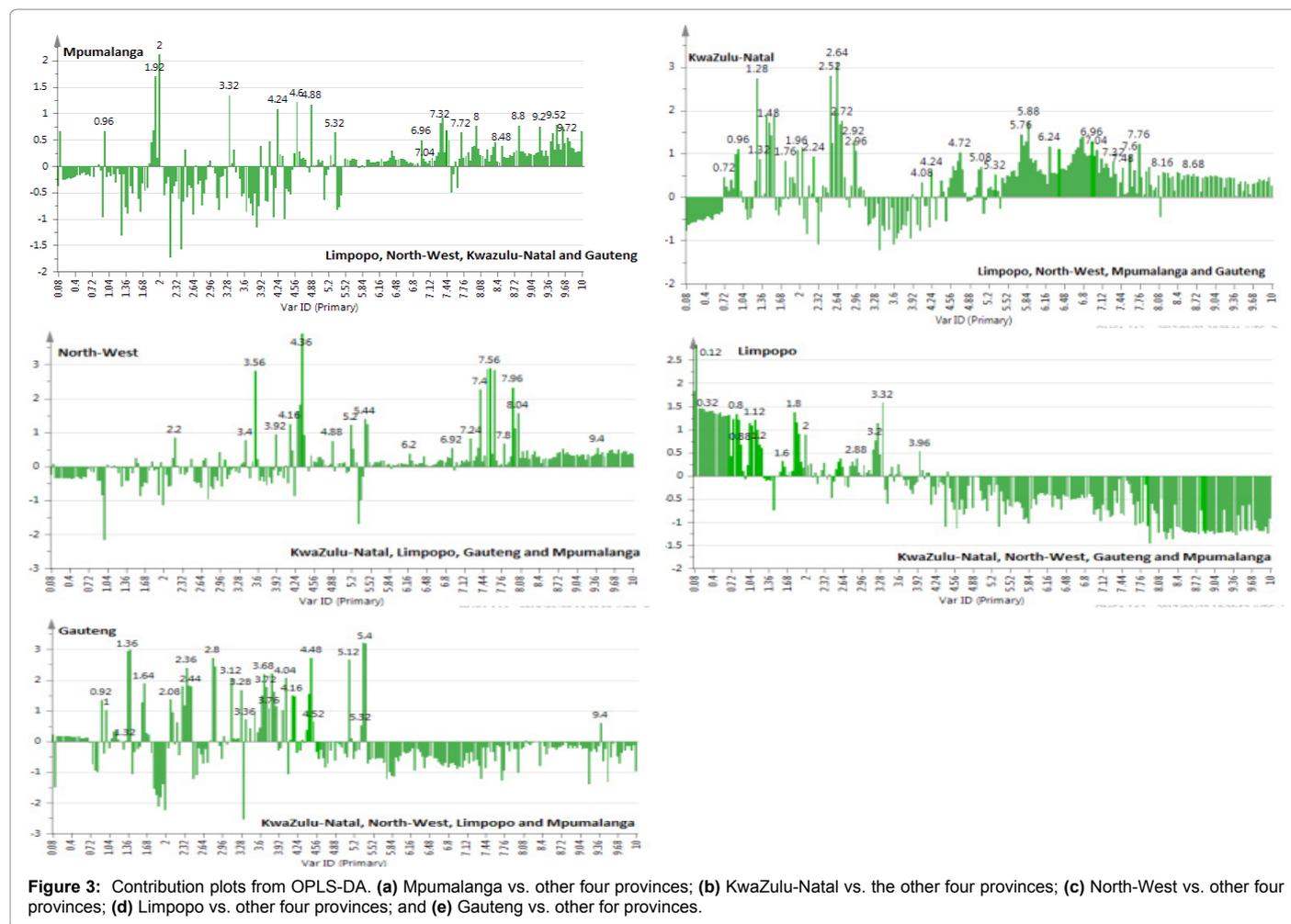


Figure 2: Representative PCA (a) and OPLS-DA (b) scores from ¹H NMR Spectra of *Sclerocarya birrea* from five provinces.

3-0-β-Dgalactopyranoside (bin at 4.04, 4.08, 5.36, 6.18, 6.39, 6.81, 7.51 and 7.65 ppm), quercetin 3-0-α-L-rhamnopyranoside (bin at 5.24, 6.19, 6.38, 6.85, 7.24 and 7.28 ppm) and threonine (bin at 1.32 ppm) compared to other provinces. Gauteng samples showed no phenolic compounds as metabolites that contributed to it being different to the other provinces, however the sugar; sucrose (bin at 3.44, 3.54, 3.68, 3.76, 3.78-3.90, 4.04, 4.17 and 5.39 ppm), unknown sugars (bins at 5.12, 5.32 and 5.4 ppm), lactate (bin at 1.32 and 4.04 ppm) and an unknown amino acid (bin at 0.92) appeared to be the important metabolites in this province. In addition to the profile observed in the samples from Gauteng, Limpopo also showed sucrose (bin at 3.44, 3.54, 3.68, 3.76, 3.78-3.90, 4.04, 4.17 and 5.39 ppm) as an important metabolite, fatty acid (bin at 0.88 ppm) was also observed in the samples from this province. Mpumalanga showed a combination of sugars, amino acids, and phenolics as the important metabolites. Higher levels of leucine (bin at 0.96 ppm), unknown sugars (bins at 4.24, 4.88 and 5.32 ppm), gallic acid (bin at 7.04 ppm), quercetin 3-0-β-D galactopyranoside (bin

at 4.19, 4.23, 5.41, 6.17, 6.36, 6.71, 7.42 and 7.57 ppm), quercetin 3-0-β-D galactopyranoside (bin at 4.04, 4.08, 5.36, 6.18, 6.39, 6.81, 7.51 and 7.65 ppm), kaempferol-3-0-α-L-rhamnopyranoside (bin at 5.43, 6.20, 6.38 ppm) and (-)-epigallocatechin 3-0-galloyl ester (bin at 2.92, 3.04, 5.65, 6.03 and 6.06 ppm), where observed as important metabolites in Mpumalanga samples. The profile observed in North West province is almost identical as the one from Mpumalanga. The samples from North-west presented quercetin 3-0-β-D galactopyranoside (bin at 4.19, 4.23, 5.41, 6.17, 6.36, 6.71, 7.42 and 7.57 ppm), quercetin 3-0-α-L-rhamnopyranoside (bin at 5.24, 6.19, 6.38, 6.85, 7.24 and 7.28 ppm), kaempferol-3-0-α-L-rhamnopyranoside (bin at 5.43, 6.20, 6.38) and quercetin 3-0-β-D galactopyranoside (bin at 4.04, 4.08, 5.36, 6.18, 6.39, 6.81, 7.51 and 7.65 ppm) as important metabolites.

The analysis of *S. birrea* leaf extracts from five provinces by ¹H NMR allowed the detection of essentially primary (polar) metabolites such as sugars, amino acids and organic acids and few secondary



metabolites. The abundance of the corresponding resonances indicated the higher natural concentrations of these metabolites in the leaves. The relatively low abundance of secondary metabolites and the large amount of resonance of primary metabolites in the spectrum made the detection of secondary metabolites, such as phenolic and flavonoids more difficult as compared to the detection of the primary metabolites. Liquid Chromatography Mass Spectroscopy was applied to further investigate the abundance of secondary metabolites. When Partial Least Square (PLS) was performed on the methanol extracts from LC-MS, the samples appeared to be grouped according to their places of origin (Figure 4). The samples from Mpumalanga showed some similarities to those from, North-West and KwaZulu-Natal. The VIP scores were generated and the assignment of the metabolites in the VIP scores were based on previously published literature (Figure 5).

The VIP scores (Table 1) confirms the NMR profile that indeed the samples from Limpopo and those from Gauteng contained low to no secondary metabolites as important in differentiating them to the other provinces, this is shown by green and light green colors in the annotated metabolites of the VIP scores. The VIP scores also showed that the metabolite levels of Quercetin 3-O-β-D-(6"-galloyl)-galactopyranoside were high in the samples from North-West followed by the ones from KwaZulu-Natal and Mpumalanga (red, orange, and

Sampling place	Latitude	Longitude
Limpopo province;		
-Mookgophong	24.5037°S	28.7140°E
-Ga-Sekgoboko	24.0894°S	28.9299°E
Gauteng province;		
-Jubilee	25.4057°S	28.2691°E
-Babelegi	25.3554°S	28.2790°E
North-West Province;		
-Lethlabile	25.4903°S	27.8415°E
-Kgabaatsane	25.5361°S	27.9571°E
Mpumalanga;		
-Machanodorp	25.6673°S	30.2455°E
-ARC-Nelspruit	25.4518°S	30.9697°E
KwaZulu-Natal;		
-Ntibane bushveld hideaway	27.7062°S	31.49032°S
-Mahlalela game range	27.7627°S	29.9344°S

Table 1: Leaf samples collected from different study sites.

yellow respectively). The annotated metabolites in Table 1, confirmed the highest levels of the metabolites Quercetin 3-O-β-D-(6"-galloyl)-galactopyranoside, Kaempferol-3-O-α-L-rhamnopyranoside and (-)-Epigallocatechin-3-O-galloyl ester in the samples from North-West, Mpumalanga and KwaZulu-Natal.

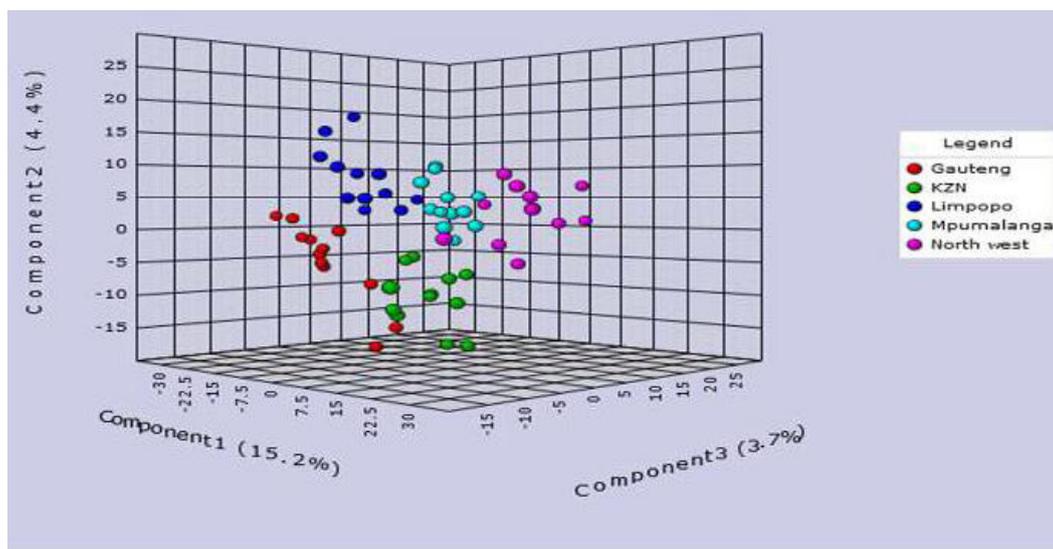


Figure 4: PLS scores of *Sclerocarya birrea* LC-MS from five provinces.

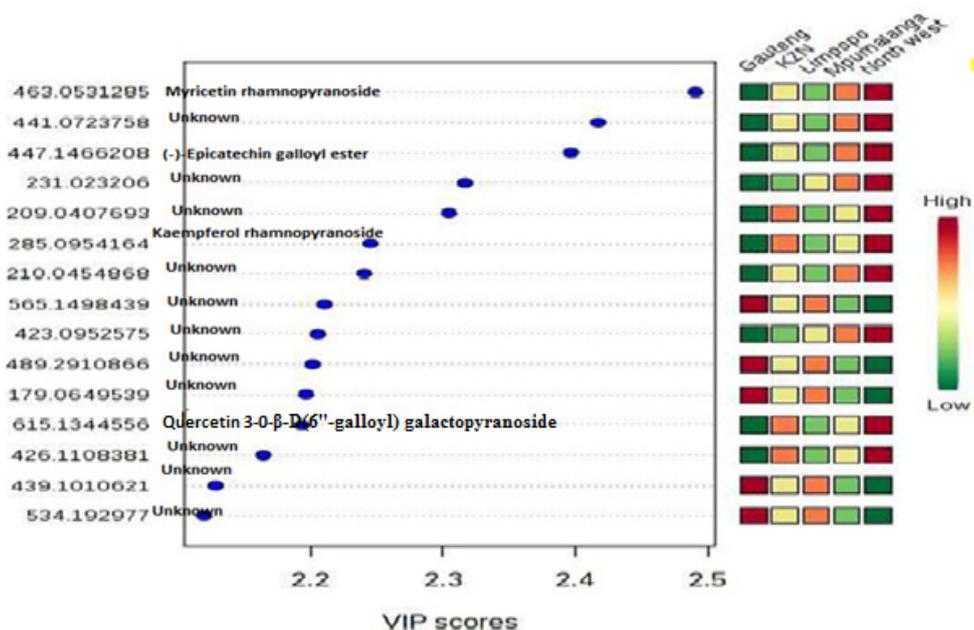


Figure 5: VIP scores from PLS scores of *Sclerocarya birrea* from five provinces.

Discussion

Virtual inspection of the ¹H NMR spectra showed no clear differences between samples from different provinces, this indicated the similarities within the species. Upon multivariate data analysis, five groups were observed according to the provinces, indicating the geographical influence on the chemical composition. The samples from Gauteng and KwaZulu-Natal was however clustered closer together and more separated from the other provinces. When the contribution plots were constructed on the samples from Gauteng, it was clear that no secondary metabolites contributed to the differences between Gauteng and other provinces. The putative annotation presented in Table 2 was

based on previously published metabolites isolated from the leaves of *S. birrea* [16]. The samples from Gauteng presented sugars, and some amino acids as the metabolites of importance. On the other side, the samples from KwaZulu-Natal showed very few sugars, few aliphatic and more aromatics to be important in the clustering of the samples from this province (Figure 3b). Visual inspection of the contribution plots from both KwaZulu-Natal and Gauteng confirms that the samples from Gauteng are rich with sugars while those from KwaZulu-Natal are rich in aromatic metabolites. It can therefore be concluded based on ¹HNMR data analysis that the separation of samples from Gauteng to those from KwaZulu-Natal, Limpopo, Mpumalanga and North West is influenced by the combination of sugars and some

Province	Metabolite	Chemical shift
KwaZulu-Natal	1. Quercetin 3- <i>O</i> -β- <i>D</i> - (6"-galloyl) glucopyranoside	4.19 (dd) 4.23 (dd) 5.41 (dd) 6.17 (d) 6.36 (d) 6.71 (d) 7.42 (d) 7.57 (dd)
	2. Quercetin 3- <i>O</i> -β- <i>D</i> - (6"-galloyl) galactopyranoside.	4.04 (dd) 4.08 (dd) 5.36 (d) 6.18 (d) 6.39 (d) 6.81 (d) 7.51 (d) 7.65 (dd)
	3. Quercetin 3- <i>O</i> -α- <i>L</i> - rhamnopyranoside	5.24 (d) 6.19 (d) 6.38 (d) 6.85 (d) 7.24 (dd) 7.28 (d)
	4. Gallic acid	7.04 (s)
	5. Threonine	1.32 (d)
Mpumalanga	1. Quercetin 3- <i>O</i> -β- <i>D</i> - (6"-galloyl) glucopyranoside	4.19 (dd) 4.23 (dd) 5.41 (dd) 6.17 (d) 6.36 (d) 6.71 (d) 7.42 (d) 7.57 (dd)
	2. Quercetin 3- <i>O</i> -β- <i>D</i> - (6"-galloyl) galactopyranoside.	4.04 (dd) 4.08 (dd) 5.36 (d) 6.18 (d) 6.39 (d) 6.81 (d) 7.51 (d) 7.65 (dd)
	3. Kaempferol-3- <i>O</i> -α- <i>L</i> - rhamnopyranoside	5.43 (d) 6.20 (d) 6.38 (d)
	4. Gallic acid	7.04 (s)
	5. (-)-Epigallocatechin-3- <i>O</i> -galloyl ester	2.92 (dd) 3.04 (dd) 5.65 (s) 6.03 (d) 6.06 (d) 7.04 (s) 2.92 (dd) 3.04 (dd) 5.65 (s) 6.03 (d) 6.06 (d)
	6. Leucine	0.96 (d) 1.70 (m) 3.73 (m)
Gauteng	1. Sucrose	3.44 (t) 3.54 (dd) 3.68 (s) 3.76 (t) 3.78-3.90 (m) 4.04 (t) 4.17 (d) 5.39 (d)
	3. Lactate	1.32 (d) 4.03 (q)
North-West	1. Quercetin 3- <i>O</i> -β- <i>D</i> - (6"-galloyl) Glucopyranoside	4.19 (dd) 4.23 (dd) 5.41 (dd) 6.17 (d) 6.36 (d) 6.71 (d) 7.42 (d) 7.57 (dd)
	2. Quercetin 3- <i>O</i> -α- <i>L</i> - rhamnopyranoside	5.24 (d) 6.19 (d) 6.38 (d) 6.85 (d) 7.24 (dd) 7.28 (d)
	3. Kaempferol-3- <i>O</i> -α- <i>L</i> -	5.43 (d) 6.20 (d) 6.38 (d)
	4. Quercetin 3- <i>O</i> -β- <i>D</i> - (6"-galloyl) galactopyranoside.	4.04 (dd) 4.08 (dd) 5.36 (d) 6.18 (d) 6.39 (d) 6.81 (d) 7.51 (d) 7.65 (dd)
	5. (-)-Epigallocatechin-3- <i>O</i> -galloyl ester	2.92 (dd) 3.04 (dd) 5.65 (s) 6.03 (d) 6.06 (d)
Limpopo	1. Sucrose	3.44 (t) 3.54 (dd) 6.68 (s) 3.76 (t) 3.78-3.90 (m) 4.04 (t) 4.17 (d) 5.39 (d)
	2. Fatty acids	0.88 (t)

Table 2: Putative assignment of *S. birrea* metabolites from NMR buckets [CD₃OD].

amino acids, while the separation of samples from KwaZulu-Natal to those from the other four provinces is mostly influenced by aromatic compounds. Complementarily to NMR data, the use of C18 reversed phase LC-MS detected metabolites such as flavonoids with the PLS that is analogous to the one obtained in the OPLS-DA from NMR. Liquid Chromatography Mass Spectroscopy detected mostly the secondary metabolites, similar to those partially detected by NMR (Table 2).

Very few flavonoids annotated from *Sclerocarya birrea* from different provinces supports previously published literature that concluded the use of *S. birrea* to treat various human ailments. The anticancer potential of this plant is evidenced by the presence of the flavonoids; quercetin and kaempferol. These two metabolites are believed to act synergistically in the reduction of cell proliferation of cancer cells; this was concluded in an *in-vivo* study [25]. Moreover, in the study conducted by Alkhalidy et al. [26], the metabolite kaempferol showed potential in the treatment of diabetes. Alkhalidy et al. [26] concluded that kaempferol treatment reversed the high fat diet impaired glucose transport-4 and AMP-dependent protein kinase expression in both muscles and adipose tissues from obese mice. This study provided the baseline information on the healing effect of *S. birrea*, by providing the snapshot of the metabolites found in each province. This study will attract researchers to use samples from different geographical origins when screening for activity since the geographical origin affects the abundance of the metabolites.

Conclusion

In this study, *Sclerocarya birrea* leaf extracts from different provinces were compared and characterized based on multivariate statistical

analysis of ¹H NMR and LC-MS based metabolomics data. Both the primary metabolites and secondary metabolites were annotated. *S. birrea* from different provinces exhibited significant differences in their metabolite profiles confirming that several metabolites can be used as markers for origin. The samples also showed some close similarities, in that some metabolites, Quercetin 3-*O*-β-*D*- (6"-galloyl) galactopyranoside, Kaempferol-3-*O*-α-*L*-rhamnopyranoside, (-)-Epigallocatechin-3-*O*-galloyl ester and gallic acid were common between samples from North-West, Mpumalanga and KwaZulu-Natal. The most significant difference was with regard to the sugars in the samples from Gauteng and Limpopo regions (Table 2) influenced by glycosylation. In this study, we demonstrated for the first time that ¹H NMR and LC-MS based profiling provides a fast and efficient method to fingerprint metabolic differences between same species of *S. birrea* grown in different regions.

Acknowledgement

The authors thank the department of Science and Technology for financial assistance.

References

1. Leakey R, Shackleton S, Plessis PD (2005) Domestication potential of Marula (*Sclerocarya birrea* subsp *caffra*) in South Africa and Namibia: 1. Phenotypic variation in fruit traits. *Agroforestry Systems* 64: 25-35.
2. Eloff J (2001) Antibacterial activity of Marula (*Sclerocarya birrea* (A. rich.) Hochst. subsp. *caffra*(Sond.) Kokwaro)(Anacardiaceae) bark and leaves. *J Ethnopharmacol* 76: 305-308.
3. Tanih NF, Ndip RN (2013) The acetone extract of *Sclerocarya birrea* (Anacardiaceae) possesses antiproliferative and apoptotic potential against human breast cancer cell lines (MCF-7). *Sci World J* 2013: 1-7.

4. Ojewole JA (2004) Evaluation of the analgesic, anti-inflammatory and anti-diabetic properties of *Sclerocarya birrea* (A. Rich.) Hochst. stem-bark aqueous extract in mice and rats. *Phytother Res* 18: 601-608.
5. Kim HK, Choi YH, Verpoorte R (2010) NMR-based metabolomic analysis of plants. *Nat Protocol* 5: 536-549.
6. Dunn WB, Ellis DI (2005) Metabolomics: current analytical platforms and methodologies. *TrAC Trends Analyt Chem* 24: 285-294.
7. Tikunov Y, Lommen A, De Vos CH, Verhoeven HA, Bino RJ, et al. (2005) A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. *Plant Physiol* 139: 1125-1137.
8. Moco S, Vervoort J, Bino RJ, De Vos RC, Bino R (2007a) Metabolomics technologies and metabolite identification. *Trends Analyt Chem* 26: 855-866.
9. Le Gall G, Colquhoun IJ, Davis AL, Collins GJ, Verhoeven ME (2003) Metabolite profiling of tomato (*Lycopersicon esculentum*) using ¹H NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. *J Agri Food Chem* 51: 2447-2456.
10. Moco S, Vervoort J, Bino RJ, De Vos RC, Bino R (2007b) Metabolomics technologies and metabolite identification. *Trends Analyt Chem* 26: 855-866.
11. Sawada Y, Hirai MY (2013) Integrated LC-MS/MS system for plant metabolomics. *Comput Struct Biotechnol J* 4: 1-6.
12. Moco S, Bino RJ, Vorst O, Verhoeven HA, De Groot J, et al. (2006) A liquid chromatography-mass spectrometry-based metabolome database for tomato. *Plant Physiol* 141: 1205-1218.
13. Allwood JW, Ellis DI, Goodacre R (2008) Metabolomic technologies and their application to the study of plants and plant-host interactions. *Physiol Plant* 132: 117-135.
14. Ward JL, Baker JM, Beale MH (2007) Recent applications of NMR spectroscopy in plant metabolomics. *Febs J* 274: 1126-1131.
15. Wolfender J, Ndjoko K, Hostettmann K (2001) The potential of LC-NMR in phytochemical analysis. *Phytochem Anal* 12: 2-22.
16. Braca A, Politi M, Sanogo R, Sanou H, Morelli I, et al. (2003) Chemical composition and antioxidant activity of phenolic compounds from wild and cultivated *Sclerocarya birrea* (Anacardiaceae) leaves. *J Agri Food Chem* 51: 6689-6695.
17. Krishnan P, Kruger NJ, Ratcliffe RG (2005) Metabolite fingerprinting and profiling in plants using NMR. *J Exp Botany* 56: 255-265.
18. Kim HK, Choi YH, Verpoorte R (2011) NMR-based plant metabolomics: where do we stand, where do we go? *Trends Biotechnol* 29: 267-275.
19. Yang S, Tschaplinski TJ, Engle NL, Carroll SL, Martin SL, et al. (2009) Transcriptomic and metabolomic profiling of *Zymomonas mobilis* during aerobic and anaerobic fermentations. *BMC Genom* 10: 1-34.
20. Robert AVB, Hoefsloot HC, Westerhuis JA, Smilde AK, Mariët JvW (2006) Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 7: 142.
21. Kim J, Jung Y, Bong Y, Lee K, Hwang G (2012) Determination of the geographical origin of kimchi by ¹H NMR-based metabolite profiling. *Biosci Biotechnol Biochem* 76: 1752-1757.
22. Verpoorte R, Choi Y, Kim H (2007) NMR-based metabolomics at work in phytochemistry. *Phytochem Rev* 6: 3-14.
23. Umetrics A (1992) SIMCA-P, SIMCA-P.
24. Wurtele ES, Chappell J, Jones AD, Celiz MD, Ransom N, et al. (2012) Medicinal plants: a public resource for metabolomics and hypothesis development. *Metabolites* 2: 1031-1059.
25. Ackland ML, Van De Waarsenburg S, Jones R (2005) Synergistic antiproliferative action of the flavonols quercetin and kaempferol in cultured human cancer cell lines. *In vivo* (Athens, Greece) 19: 69-76.
26. Alkhalidy H, Moore W, Zhang Y, Mcmillan R, Wang A, et al. (2015) Small Molecule Kaempferol Promotes Insulin Sensitivity and Preserved Pancreatic beta -Cell Mass in Middle-Aged Obese Diabetic Mice. *J Diabetes Res* 2015: 1-14.