

# SCIENTIFIC REPORTS



OPEN

## Metabolite profiling of *Dioscorea* (yam) species reveals underutilised biodiversity and renewable sources for high-value compounds

Elliott J. Price<sup>1,2</sup>, Paul Wilkin<sup>2</sup>, Viswambharan Sarasan<sup>2</sup> & Paul D. Fraser<sup>1</sup>

Received: 09 March 2016

Accepted: 15 June 2016

Published: 07 July 2016

Yams (*Dioscorea spp.*) are a multispecies crop with production in over 50 countries generating ~50 MT of edible tubers annually. The long-term storage potential of these tubers is vital for food security in developing countries. Furthermore, many species are important sources of pharmaceutical precursors. Despite these attributes as staple food crops and sources of high-value chemicals, *Dioscorea spp.* remain largely neglected in comparison to other staple tuber crops of tropical agricultural systems such as cassava (*Manihot esculenta*) and sweet potato (*Ipomoea batatas*). To date, studies have focussed on the tubers or rhizomes of *Dioscorea*, neglecting the foliage as waste. In the present study metabolite profiling procedures, using GC-MS approaches, have been established to assess biochemical diversity across species. The robustness of the procedures was shown using material from the phylogenetic clades. The resultant data allowed separation of the genotypes into clades, species and morphological traits with a putative geographical origin. Additionally, we show the potential of foliage material as a renewable source of high-value compounds.

*Dioscorea* (yam) species comprise a genus of 613 tuberous climbing plants<sup>1</sup>. Of these, seven to ten are cultivated on a large scale<sup>2,3</sup> and two (*D. alata*, *D. cayennensis* Lam. subsp. *cayennensis* and *D. cayennensis* Lam. subsp. *rotundata* (Poir.) J. Miège [referred to *D. rotundata* throughout]) are of primary importance as a staple crop, predominately in Western Africa<sup>4</sup>, for over 100 million people<sup>5</sup>. Approximately a further 50 species are eaten as wild-harvested staples or famine food and the genus holds importance for global food security<sup>6</sup>. In addition, *Dioscorea* species have been widely used in traditional medicines<sup>7,8</sup> and as a source of steroidal precursors<sup>9,10</sup>. Nevertheless yams are categorised as understudied and underutilised<sup>11,12</sup>.

Yam production is relatively expensive compared to other root and tuber crops due to high planting and labour costs, a long growing season and low yield per hectare<sup>3,13</sup>. The cost per 1000 calories from yam is estimated at 4 times that for cassava (*Manihot esculenta Crantz*)<sup>14</sup>. Yams hold cultural and social importance<sup>11</sup> and have preferred organoleptic properties compared with other carbohydrate sources, including cassava, potatoes (within *Solanum spp.*) and sweet potato (*Ipomoea batatas* (L.) Lam.)<sup>15</sup>. Sensorial preference, coupled with better storage qualities compared to crops such as cassava and plantain (*Musa spp.*)<sup>16</sup>, have led to high-demand as a cash crop<sup>13</sup>.

The abundance of steroidal C<sub>27</sub> saponins in the rhizome or tubers of some *Dioscorea* has long been known<sup>8,17–19</sup>. Diosgenin, the aglycone portion of the abundant saponin dioscin, has been industrially exploited as the starting material for the synthesis of pregnenolone-derived steroids<sup>10</sup>. Around 15 species of *Dioscorea* are used as a source of diosgenin<sup>20</sup>, with an estimated market value of \$500 million<sup>21</sup>. Over-harvesting of rare yams has threatened populations of several *Dioscorea* species worldwide<sup>18,22–24</sup>.

Use of *Dioscorea* species in traditional medicines is documented to at least 2000 BC<sup>25</sup>, including leaf material<sup>26</sup> as a constituent. Species such as *D. dumetorum* and *D. hispida* Dennst. and their relatives have been those primarily used for their poisonous properties, due to the presence of polar alkaloids<sup>8</sup>. Despite this, phytochemical analysis using metabolite profiling has not previously been conducted on the foliage of *Dioscorea* species in comparative biodiversity research.

<sup>1</sup>School of Biological Sciences, Royal Holloway University of London, Egham, Surrey, TW20 0EX, UK. <sup>2</sup>Royal Botanic Gardens, Kew, Richmond, Surrey, TW20 3AB, UK. Correspondence and requests for materials should be addressed to P.D.F. (email: p.fraser@rhul.ac.uk)

The foliage material of yam is an agricultural waste product yet some species generate substantial above-ground biomass each annual growth cycle. Utilising *Dioscorea* foliage has potential to provide a renewable source of high-value natural products whilst sustaining conservation of species.

Lebot *et al.* applied a metabolite profiling approach to tubers of some tropical root crops including selected *Dioscorea* species, where results showed inter- and intra- species diversity in *Dioscorea*<sup>27,28</sup>, emphasising the potential of a more global metabolomic investigation.

In this study, we analysed the polar extracts from leaf and petiole material of a diverse collection of *Dioscorea* via Gas Chromatography-Mass Spectrometry (GC-MS). The metabolite profiles obtained enabled investigation of species discrimination and comparison with phylogenetic relationships and morphological characteristics. The profiles provide insight into biochemically-related species and highlight *Dioscorea* species as potential sources of valuable compounds.

## Results

**The yam metabolome.** The devised platform for analysis of *Dioscorea* material (Supplementary Fig. S1) provided broad coverage of primary metabolism. Metabolic pathways which can be analysed by this platform include carbohydrate metabolism and amino acid biosynthesis, along with some nucleotides, secondary metabolites, monoamines and derivatives. A strong linear response ( $R^2 > 0.98$ ) was shown for 41 compounds measured (Supplementary Table S2).

Recovery rates for many compounds however, were low (<70% of the total extracted via three extractions, Supplementary Table S3). There were no qualitative changes (new metabolites extracted) following multiple extractions on samples, and thus one extraction is robust for fast screening. Overloading of the MS was necessary to achieve acceptable coverage of the metabolome and enhance the dynamic range.

A diagrammatic workflow of methods is provided in Supplementary Fig. S1.

**Species demarcation.** Profiling of the polar extracts from foliage material of 28 *Dioscorea* accessions, covering 19 species (Table 1), allowed consistent measurement of 151 features (Supplementary Table S4). Clustering of replicates through GPA (Fig. 1a) highlighted the robustness of analysis and the consensus arrangement described 85% of total variation. Univariate analysis to identify the most discriminatory variables allowed reduction of the dataset to 41 variables, which enabled a comparable degree of species demarcation as using all 151 features (Fig. 1b, Supplementary Fig. S5).

Species tended to group on the basis of phylogenetic and/or morphological traits. Species of the African (Afr) clade (comprising *D. elephantipes* and *D. sylvatica*) formed a distinct group characterised by abundance of shikimic acid and pyrogallol. The majority of compound leaved (CL) species (*D. pentaphylla*, *D. cochleari-apiculata* & *D. dumetorum*) migrated towards sucrose, citric acid; ascorbic acid and its degradation product erythronic acid to form a cluster. Exceptions to this were *D. bulbifera* and *D. antaly*.

*D. bulbifera* showed a profile more similar to the cultivated species of Enantiophyllum e.g. *D. alata* and *D. rotundata* (in the same plane on F1, Fig. 1b & F2, Fig. 1a) whereas *D. antaly* clustered with crop wild relatives of *D. rotundata* (*D. praehensilis* and *D. minutiflora*) and Stenophora lineages (*D. membranacea*, *D. rockii*) around the origin of the GPA. Species from the New World (*D. composita* and *D. altissima*) also clustered around the origin of the plot in the reduced dataset (Fig. 1b), yet *D. altissima* clustered with CL species on the total GPA (Fig. 1a). Species at the origin all presented higher levels of amino acids and monosaccharides (Fig. 1c).

*D. alata* and *D. preussii* (both Enantiophyllum) migrate from the origin, primarily due to the influence of scyllo-inositol (Fig. 1b,c). Glucose, fructose and xylulose are the predominant variables distinguishing *D. rotundata* from its crop wild relatives: *D. praehensilis* & *D. minutiflora* (all three Enantiophyllum). Samples of *D. sansibarensis* (Malagasy) were distinguishable on F2, yet not F1 (Fig. 1b) with higher sugar content in one sample driving the separation.

Cluster analysis highlighted the affinity of *D. rotundata* and its crop wild relatives (*D. praehensilis* and *D. minutiflora*) with the most basal lineages of *Dioscorea* (Stenophora and New World clades) (Fig. 2). Additionally, the Afr clade formed a tight cluster flanked by groupings of CL species, with some outliers, e.g. the South American species *D. altissima*. Cluster analysis on the metabolites showed that biochemically-related compounds tended to group (Supplementary Fig. S6), supporting the reliability of the approach. *D. pentaphylla* could be distinguished from other CL species due to relatively high abundance of dopamine and the derivative norepinephrine and *D. antaly* was distinguishable due to relatively high levels of catechins (catechin, epicatechin and gallocatechin; Supplementary Fig. S6).

The demarcation of accessions is robust even with the limited tissue availability and cultivation conditions used.

**Compound atlas.** The polar extracts from leaf material of many *Dioscorea* species showed numerous unknown compounds and many of these were abundant. Using *D. elephantipes* as an example (Fig. 3a) a more detailed investigation of the metabolome was conducted. Both polar and non-polar phases were analysed and all features, including unknowns, were measured on different structures of the plant (stem, leaf, root, inner and outer of parts of caudiciform tuber). Comprehensive coverage (totalling 535 features) showed similar discrimination as when using the 206 known and putative metabolites and solely the 121 identified features (Supplementary Fig. S7). Further reduction of this dataset by choosing the most discriminatory variables (5 or more groups following Kruskal-Wallis' and Conover-Iman post hoc) gave 38 features which can be used to discriminate regions of a single plant (Supplementary Fig. S7).

Shikimic acid was present in all regions in descending order: inner tuber, stem, leaf, outer tuber, roots. Sections were characterised by high abundance of particular metabolites. The inner caudex with amino acids,

<i>Dioscorea</i> species [in text]	Accession	Date collected	Glasshouse	Native habitat <sup>a</sup>	Clade <sup>b</sup>	Perennating organ type(s) <sup>c</sup>
<i>elephantipes</i> (L'Hér.) Engl. [ <i>D. elephantipes</i> (1)]	2007–447	21/11/2013	Tropical Nursery	South Africa [SA]	Africa- <i>Testudinaria</i>	Caudiciform perennial tuber
<i>membranacea</i> Pierre ex Prain & Burkill [ <i>D. membranacea</i> ]	1998–4292	21/11/2013	Tropical Nursery	Asia [Asia]	Stenophora	Rhizome
<i>elephantipes</i> (L'Hér.) Engl. [ <i>D. elephantipes</i> (2)]	2012–54	21/11/2013	Tropical Nursery	South Africa [SA]	Africa- <i>Testudinaria</i>	Caudiciform perennial tuber
<i>sylvatica</i> Eckl. [ <i>D. sylvatica</i> (1)]	1963–26704	21/11/2013	Tropical Nursery	South Africa [SA]	Africa- <i>Testudinaria</i>	Perennial tuber
<i>sylvatica</i> Eckl. [ <i>D. sylvatica</i> (2)]	1963–26705	21/11/2013	Tropical Nursery	South Africa [SA]	Africa- <i>Testudinaria</i>	Perennial tuber
<i>dumetorum</i> (Kunth) Pax [ <i>D. dumetorum</i> ]	1984–8405	21/11/2013	Tropical Nursery	Tropical Africa [TA]	Compound leaved-Lasiophyton (Lasiophyton)	Annual tuber & aerial bulbils
<i>antaly</i> Jum. & H.Perrier [ <i>D. antaly</i> ]	1998–523	21/11/2013	Tropical Nursery	Madagascar [Mad]	Compound leaved	Annual tuber
<i>bulbifera</i> L. [ <i>D. bulbifera</i> (1)]	1998–533	21/11/2013	Tropical Nursery	Asia [Asia]	Compound leaved (Opsophyton)	Annual tuber & aerial bulbils
<i>cochleariopiculata</i> De Wild [ <i>D. cochleari-apiculata</i> (1)]	1998–2987	21/11/2013	Tropical Nursery	Tropical Africa [TA]	Compound leaved-Lasiophyton	Annual tuber
<i>pentaphylla</i> L. [ <i>D. pentaphylla</i> ]	1996–4313	21/11/2013	Tropical Nursery	Asia [Asia]	Compound leaved-Botryosicyos	Annual tuber & aerial bulbils
<i>sansibarensis</i> Pax [ <i>D. sansibarensis</i> (1)]	1998–525	21/11/2013	Tropical Nursery	Madagascar [Mad]	Malagasy	Perennial tuber & aerial bulbils
<i>altissima</i> Lam. [ <i>D. altissima</i> ]	2005–1233	21/11/2013	Tropical Nursery	New World [NW]	N.A. <sup>d</sup>	Annual tuber
<i>cayennensis</i> Lam. subsp. <i>rotundata</i> (Poir.) Miège [ <i>D. rotundata</i> ]	1976–1475	25/11/2013	Palm House	Tropical Africa [TA]	(Enantiophyllum)	Annual tuber
<i>praehensilis</i> Benth. [ <i>D. praehensilis</i> ]	1960–1002	25/11/2013	Palm House	Tropical Africa [TA]	(Enantiophyllum)	Annual tuber
<i>preussii</i> Pax subsp. <i>preussii</i> [ <i>D. preussii</i> ]	1968–57006	25/11/2013	Palm House	Tropical Africa [TA]	Enantiophyllum (Macrocarpaea)	Annual tuber
<i>minutiflora</i> Engl. [ <i>D. minutiflora</i> ]	1960–1001	25/11/2013	Palm House	Tropical Africa [TA]	Enantiophyllum (Enantiophyllum)	Perennial tuber
<i>bulbifera</i> L. [ <i>D. bulbifera</i> (2)]	1987–1993	25/11/2013	Palm House	Asia [Asia]	Compound leaved (Opsophyton)	Annual tuber & aerial bulbils
<i>sansibarensis</i> Pax [ <i>D. sansibarensis</i> (2)]	1598–543	25/11/2013	Palm House	Madagascar [Mad]	Malagasy	Perennial tuber & aerial bulbils
<i>composita</i> Hemsl. [ <i>D. composita</i> (1)]	1969–11715	25/11/2013	Palm House	New World [NW]	New World I	Perennial tuber
<i>composita</i> Hemsl. [ <i>D. composita</i> (2)]	1978–1830	25/11/2013	Palm House	New World [NW]	New World I	Perennial tuber
<i>alata</i> L. [ <i>D. alata</i> ]	1982–1316	25/11/2013	Palm House	Asia (Asia)	Enantiophyllum (Enantiophyllum)	Annual tuber & aerial bulbils
<i>elephantipes</i> (L'Hér.) Engl. [ <i>D. elephantipes</i> (3)]	2012–54	25/11/2013	Princess of Wales	South Africa [SA]	Africa- <i>Testudinaria</i>	Caudiciform perennial tuber
<i>elephantipes</i> (L'Hér.) Engl. [ <i>D. elephantipes</i> (4)]	2001–2252	25/11/2013	Princess of Wales	South Africa [SA]	Africa- <i>Testudinaria</i>	Caudiciform perennial tuber
<i>cochleariopiculata</i> De Wild [ <i>D. cochleari-apiculata</i> (2)]	1998–2987	25/11/2013	Princess of Wales	Tropical Africa [TA]	Compound leaved-Lasiophyton	Annual tuber
<i>rockii</i> Prain & Burkill [ <i>D. rockii</i> ]	1996–4307	25/11/2013	Jodrell	Asia [Asia]	Stenophora	Rhizome
<i>cochleariopiculata</i> De Wild [ <i>D. cochleari-apiculata</i> (3)]	1998–515	25/11/2013	Jodrell	Tropical Africa [TA]	Compound leaved-Lasiophyton	Annual tuber
<i>glabra</i> Roxb. [ <i>D. glabra</i> ]	1996–4312	25/11/2013	Jodrell	Asia (Asia)	Enantiophyllum	Annual tuber
<i>cochleariopiculata</i> De Wild [ <i>D. cochleari-apiculata</i> (4)]	1995–1459	25/11/2013	Jodrell	Tropical Africa [TA]	Compound leaved-Lasiophyton	Annual tuber

**Table 1.** *Dioscorea* accessions collected from The Living Collections held at the Royal Botanic Gardens, Kew (<http://epic.kew.org/index.htm>). <sup>a</sup>Habitat used for geographical visualisation of AHC clustering (Fig. 4). <sup>b</sup>Clades from<sup>38</sup> and/or<sup>(51)</sup>. <sup>c</sup>Organ types from<sup>4</sup> and field experience. <sup>d</sup>N.A. – Not applicable, where data is unavailable.

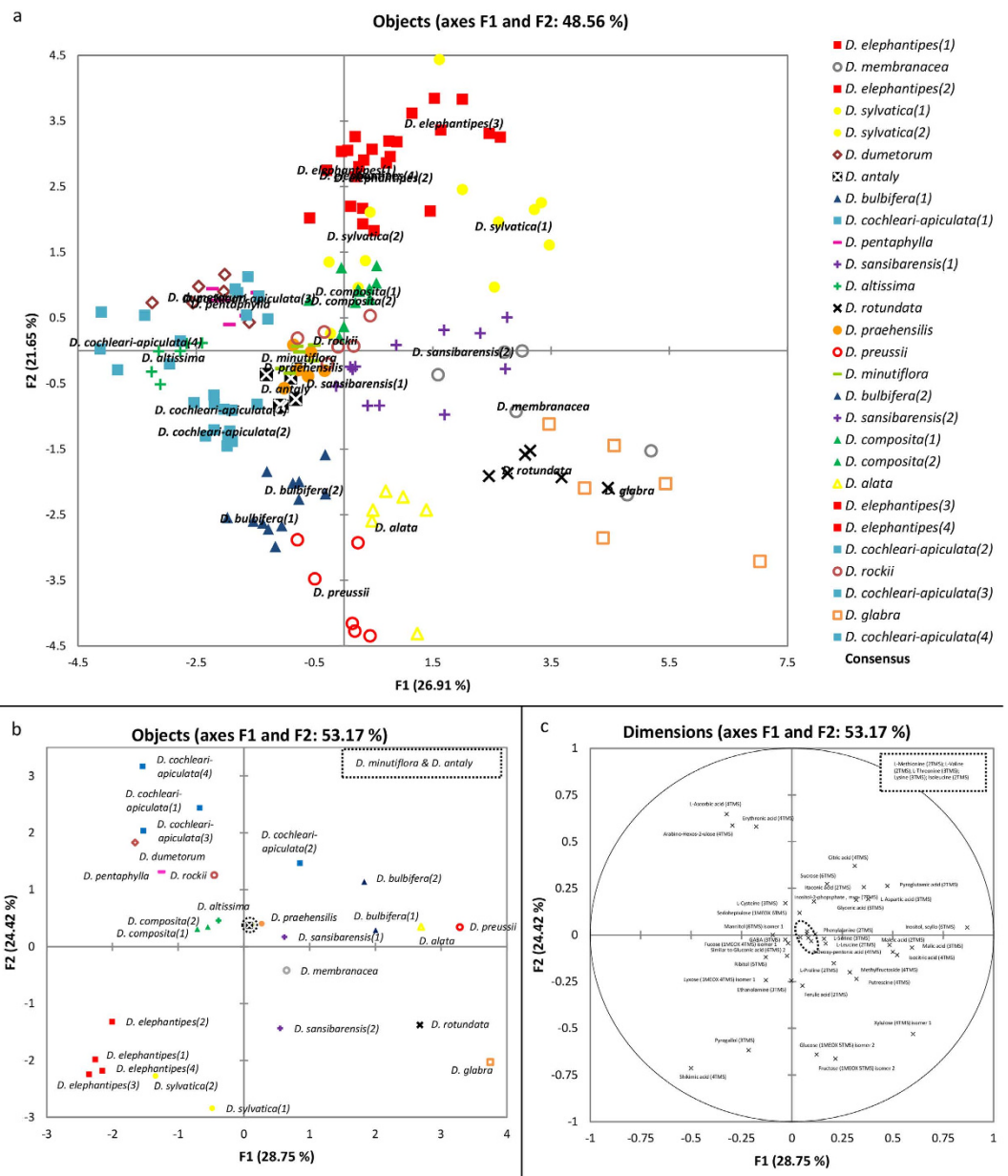
roots and outer caudex with trehalose and mannitol respectively and the leaf and stem with fructose, melibiose, GABA and erythronic acid.

**Shikimic acid quantification.** To validate the high level of shikimic acid in foliage of *D. elephantipes* & *D. sylvatica*, further material was sourced, along with another caudiciform species: *D. mexicana*. Samples of *D. elephantipes* and *D. sylvatica* showed similarly abundant shikimic acid as previously, however only trace amounts were present in *D. mexicana*.

An authentic standard showed a good linear response ( $R^2 = 0.9822$ ,  $y = 0.1923x$ ), relative to the internal standard, over the range 10–200 µg, with values in that of *D. elephantipes* and *D. sylvatica* approaching 8% (Fig. 3b).

## Discussion

Metabolite profiling via GC-MS is considered the gold standard technique for metabolite analysis<sup>29</sup> providing good resolution, high reproducibility<sup>30</sup> and can achieve broad compound coverage<sup>31</sup>. Equipment is relatively

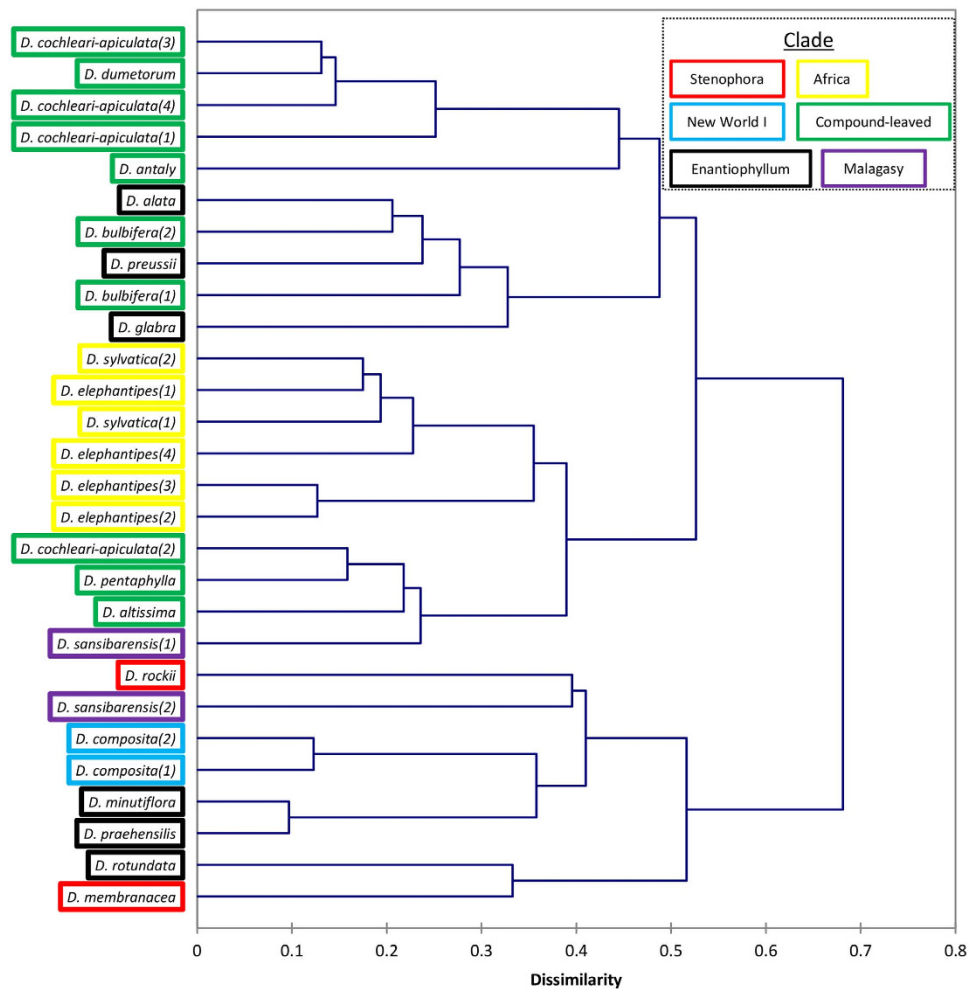


**Figure 1.** Generalised Procrustes Analysis on the polar fraction of metabolite extracts from leaf of *Dioscorea*, analysed by Gas Chromatography-Mass Spectrometry, provides (a) a transformed configuration from all six replicate analyses [ $R_c = 0.847$  (100<sup>th</sup> percentile);  $F_1 = 78.998$ ,  $F_2 = 72.105$  ( $p < 0.0001$ )] and (b) the consensus configuration from a reduced dataset of the 41 most discriminatory variables, with loadings shown in (c); which shows the same trends [ $R_c = 0.898$  (100<sup>th</sup> percentile);  $F_1 = 85.499$ ,  $F_2 = 64.471$  ( $p < 0.0001$ )].

affordable and widely used ensuring a wealth of available resources such as established protocols<sup>32</sup>, large compound libraries<sup>33</sup> and analysis software<sup>31</sup>. In comparison to other common analytical platforms GC-MS is more targeted towards intermediary metabolism with reliable compound identification. Therefore, a GC-MS based approach was favoured for initial biodiversity study of the *Dioscorea* metabolome.

A robust method (Supplementary Fig. S1) has been developed for metabolite analysis of *Dioscorea* species and applied to a diverse set of species and different plant organs. The approach provides representative coverage of the polar intermediary metabolome and can be easily extended to incorporate non-polar profiling. A low recovery rate was obtained for many compounds which may be due to the high starch (and sugar) content of material compromising extraction efficiency. Despite this, repeatable relative quantification can be achieved, evidenced by clustering of replicate samples analysed two months apart.

Sub-selections of metabolites can be identified which allows screening of small amounts of material (Fig. 1b, Supplementary Fig. S5) whilst remaining representative of total intermediary polar metabolism, as measured via this GC-MS methodology (Supplementary Fig. S7). Thus, a core set of small molecules could be defined to allow simplified rapid screening of *Dioscorea*.



**Figure 2.** Hierarchical tree of *Dioscorea* accessions based on mean ( $n = 6$ ) metabolite compositions shows relationship of chemotaxonomy with phylogenetic clades. Notably *D. rotundata* and crop-wild relatives (*D. praezensilis* and *D. minutiflora*) cluster with basal lineages of the Stenophora and New World I clades. Metabolite clustering is provided in a heat-map format in Supplementary Fig. S6.

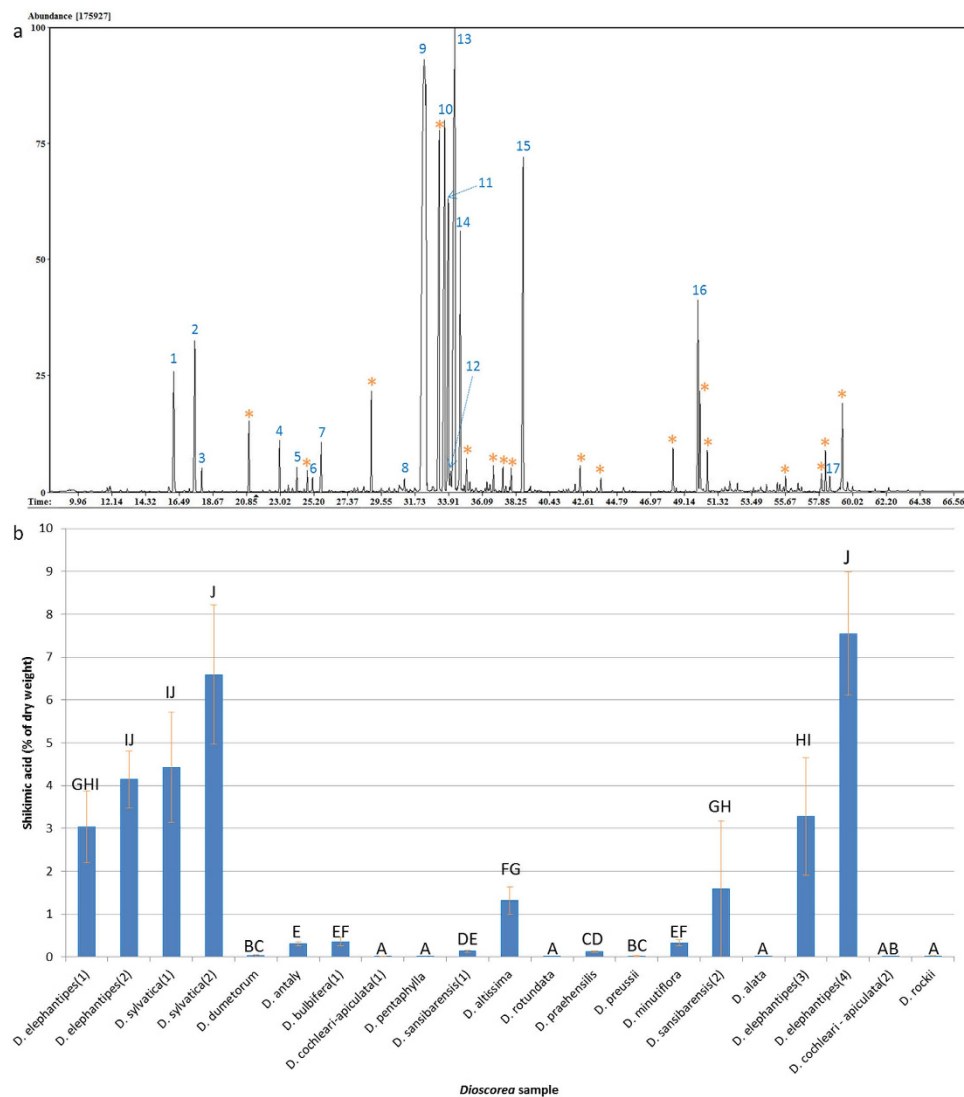
The verification of *Dioscorea* species is often noted to be problematic<sup>34,35</sup>. Furthermore, a recent genotyping by sequencing (GBS) analysis could not discriminate Guinea yam species unless combined with ploidy analysis<sup>36</sup>. Metabolomics can aid both identification and also assess biochemical diversity concurrently, which is extremely beneficial to ongoing breeding programs. Data generated from the developed platform showed that related species cluster (e.g. *D. elephantipes* with *D. sylvatica*; *D. dumetorum* with *D. cochleari-apiculata*) and that the little studied species *D. altissima* clustered with *D. pentaphylla* and the other CL species (Fig. 2). Interestingly, this matches the topology of an unpublished phylogenetic tree based on plastid marker data (personal communication, P. Wilkin), supporting the idea that a metabolomics approach can be used in place of or alongside conventional morphological descriptors used for characterisation<sup>37</sup>.

Well-sampled phylogenetic study suggest that Stenophora and New World clades are the most basal<sup>35,38</sup>. Within this work, species of these clades appear to be a centre of biochemical origin (mostly central in the GPA, Fig. 1). Relationships of cluster analysis indicate, as previously hypothesised, that *Dioscorea* originated in Asia with early transfer to the New World<sup>38</sup> (Fig. 4). Additionally, profiles obtained for CL and Afr clades are largely distinct. Species of the Enantiophyllum clades form a larger cluster overlapping other clades, which is not a surprise given that they are the youngest evolutionary lineages and inhabit a large geographical area.

Notably, *D. rotundata* and its crop wild relatives (*D. minutiflora* and *D. praezensilis*) have similar biochemical profiles to species of the Stenophora clade (Supplementary Fig. S6) and thus suggest the occurrence of convergent evolution. Around 90% of basal Stenophora species are distributed in Asia<sup>39</sup>, yet none in Africa where the central breeding programs of *Dioscorea* are based. Therefore, international co-operation will be greatly important for future breeding of these crops and capturing of traits from basal lineages.

A large proportion of unidentified metabolites have been measured in this study many of which are abundant and highlights the understudied nature of the genus. However, the genus may have potential for bioprospecting based on the finding of abundance of shikimic acid in stem and leaf material of *D. elephantipes* and *D. sylvatica* in this study.

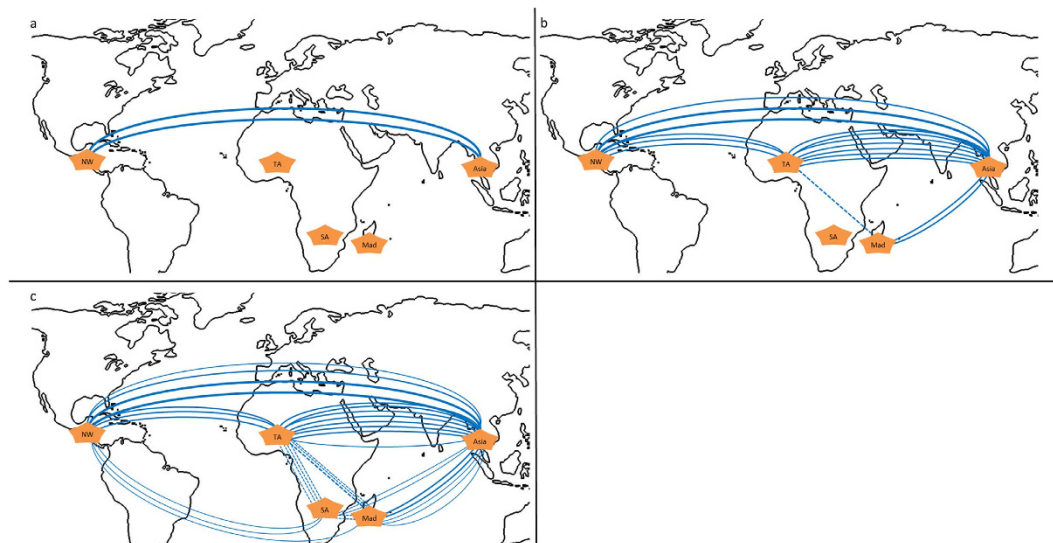




**Figure 3.** GC-MS analysis of *D. elephantipes* leaf material shows that (a) shikimic acid (4TMS) is often the most abundant peak recorded and (b) is significantly more abundant in species of the African clade. (a) Abundant peaks are: 1: Phosphate (3TMS), 2: Succinic-D<sub>4</sub> acid (internal standard), 3: MSTFA, 4: Malic acid (3TMS), 5: GABA (3TMS), 6: Threonic acid (4TMS) 7: Xylulose (4TMS) isomer 1, 8: Methylfructofuranoside (4TMS), 9: Shikimic acid (4TMS), 10: Fructose (1MEOX 5TMS) isomer 1, 11: Fructose (1MEOX 5TMS) isomer 2, 12: Galactose (1MEOX 5TMS) isomer 1, 13: Glucose (1MEOX 5TMS) isomer 1, 14: Glucose (1MEOX 5TMS) isomer 2, 15: Inositol, myo (5TMS), 16: Sucrose (8TMS), 17: Melibiose (8TMS). Many major unknowns (\*) are also present. (b) Groups from Bonferroni-corrected Conover-Iman post hoc following Kruskal-Wallis'. Error bars show 1 standard deviation (n = 6).

Shikimic acid plays a central role in biosynthesis of aromatic amino acids in plants, fungi and bacteria and as such is measured in many metabolomics studies. However, high abundance of shikimic acid within the leaf material of *Dioscorea* accessions has not previously been reported. A precursor used in the production of the anti-viral oseltamivir (Tamiflu®), shikimic acid is typically sourced from the fruits of Chinese star anise (*Illicium verum*). Levels found in foliage of *D. elephantipes* and *D. sylvatica* are approaching those for Chinese star anise<sup>40</sup> (Fig. 3b). Both species are under threat due to harvesting of tuber for traditional medicines<sup>41</sup>. The annual foliage is deemed waste material and can be used as an alternative source of shikimic acid which may promote the conservation and sustainable utilisation of these threatened wild species. The same could also be applied to other species of the genus offering a sustainable alternative to the destructive harvesting of tubers. Breeding for hybrids that produce better quality aerial biomass for less managed agroecosystems or as cover crops can improve livelihoods of farmers. Likewise selected clones can be mass propagated, using micropropagation, for industrial use and could potentially provide a profitable by-product for African yam production.

Initially, it was hypothesised that the abundance of shikimic acid may be related to tuber morphology yet only trace amounts were detected in *D. mexicana*, a caudiciform species of the New World. However, the hypothesis cannot be discarded as only one independent biological replicate could be sourced.



**Figure 4.** World map showing relationships between *Dioscorea* species from Asia (Asia) to the New World (NW), Tropical Africa (TA), South Africa (SA) and Madagascar (Mad) based on (a) first-degree, (b) second-degree and (c) third-degree linkages of samples following clustering on species-averaged metabolite compositions. Inter-continental transport is shown by dotted lines. World maps (from <https://openclipart.org/>) were modified in Microsoft Powerpoint.

Detailed study on *D. elephantipes* showed that different plant organs are easily definable by very few highly accumulated metabolites. The roots and outer caudex were abundant with trehalose and mannitol, both of which have high water retention capacities and may contribute to the high drought tolerance of this species and GABA in above ground foliage may represent a response to such stress.

Additionally, *D. pentaphylla* was shown to contain abundant dopamine yet due to a lack of availability for further material these findings could not be verified. However, the species may be a potential source for catecholamine-derived pharmaceuticals and thus a good candidate for bioprospecting.

GPA has been shown to be an effective approach for analysis of metabolomics data, especially useful in studies when representative QC material is not available for scaling of data (Fig. 1), such as when sample material is limited or not available to be pooled prior to screening, often the case in large-scale studies. Also, when comparisons across sample sets are taken at different times, e.g. over multiple growth seasons, GPA removes the need for timely scaling of samples to QC's. Additionally, different measurements, e.g. from improved compound libraries, can be integrated with previous data without the need to reanalyse all previous sample sets.

The platform represents progress for *Dioscorea* with potential to aid other studies, re-interpretation of historic data and implementation in breeding programs. Use of this GC-MS platform could be widely applied as cost is not prohibitive for developing countries growing yams (when compared to liquid chromatography (LC)-MS and other approaches). The ease of use and transferability of *Dioscorea*-specific compound libraries can provide the basis for metabolomics platform within breeding programs and allows the identification of diverse lineages.

Within this study robust results were shown from species grown under different glasshouse conditions. Future studies could use this metabolomics platform to investigate the influence of environmental factors on the metabolome and if the genetic determinates can overcome this influence.

Additionally, the many unknown abundant compounds in species (exampled Fig. 3a) highlight the further work required but provide potential leads for bioprospecting of this crop. The platform has been designed to allow extended analysis of non-polar<sup>42</sup> and secondary metabolites on other platforms<sup>43</sup> (e.g. LC -Photodiode Array (PDA)/-MS) from the same sample. This may prove useful, especially for species of the *Stenophora* and *Compound-leaved* clades which are widely utilised due to their high sterol and alkaloidal contents respectively. Therefore, the platform provides a basis for more holistic biochemical understanding of the economically, nutritionally and medicinally important yet understudied genus *Dioscorea*.

## Methods

**Reagents.** All reagents were of analytical grade.

**Plant material.** *Dioscorea* material was sourced from the Royal Botanic Gardens, Kew (Kew) Living Collections (<http://epic.kew.org/index.htm>) (Table 1). Youngest mature leaf and petiole material was sampled. Materials were cut from the vine and quenched in liquid nitrogen immediately before samples were lyophilised, homogenised and stored at  $-80^{\circ}\text{C}$  until further processing.

Additional dried leaf material of *D. elephantipes* (accessions 19900643A & 19280228C) and *D. sylvatica* (accession 19803437A) were sourced from the Royal Botanic Garden Edinburgh (RBGE) Living Collections (<http://elmer.rbge.org.uk/bgbase/livcol/bgbase/livcol.php>). Leaf material of *D. mexicana* Scheidw. (accession 8813370) was received from the Sukkulanten-Sammlung Zürich (<https://www.stadt-zuerich.ch/sukkulanten>).

For the compound atlas, a single commercially acquired *D. elephantipes* was used. The plant was halved and samples taken from leaf, stem, root, inner and outer parts of caudiciform tuber and a pooled quality control (QC) sample was prepared. Samples were quenched and processed as per the Kew Living Collections.

**Metabolite extraction.** Methanol (400  $\mu$ L) and water (400  $\mu$ L) were sequentially added to 10 mg aliquots of each sample in 2 mL plastic micro-centrifuge tubes, vortexed and rotated for 1 h at room temperature (22 °C). Chloroform (800  $\mu$ L) was added and the samples vortexed and centrifuged (3 min, 20,000 RCF) to partition extracts into upper (polar) and lower (organic) phases. A 100  $\mu$ L aliquot of the polar phase was taken into glass vials and succinic- $D_4$  acid added as internal standard (10  $\mu$ L of 1 mg/mL solution). Additionally, a 400  $\mu$ L aliquot of the non-polar phase, with myristic- $D_{27}$  acid as internal standard, was analysed for the compound atlas. Phases were dried under centrifugal evaporation and stored at  $-80$  °C until analyses. The complete Kew Living collections sample set was extracted on six independent occasions. *D. mexicana* and *D. elephantipes* for the compound atlas were extracted on three independent occasions. Samples received from RBGE were extracted once, due to limited availability of material.

**GC-MS based metabolite profiling.** Samples were re-dried for 30 min under centrifugal evaporation before methoxylation and silylation derivatisation via addition of methoxyamine hydrochloride (MeOx; 30  $\mu$ L, 20 g/L in pyridine) followed by *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA; 70  $\mu$ L); incubated (40 °C, 2 h) after addition of each<sup>33</sup>.

Samples (1  $\mu$ L) were injected into the GC-MS with a split/splitless injector at 290 °C. The injection of samples was made in splitless mode with polar samples of the Kew Living Collections also repeated on a 1:10 split. Metabolites were separated on a DB-5MS + DG 30 m (plus 10 m Duraguard)  $\times$  250  $\mu$ m  $\times$  0.25  $\mu$ m column (J&W Scientific, Folsom, California, US). The GC oven was held for 3 min at 70 °C before ramping at 4 °C/min to 325 °C and held for a min. Helium was the carrier gas at a flowrate of 1.3 mL/min. The interface with the MS was set at 280 °C and MS performed in full scan mode using 70 eV EI + and scanned from 50 to 1000 m/z. Retention time locking to ribitol was used (modified from<sup>44</sup>). A mixture of n-alkanes, ranging from 8 to 32 carbons, was used for retention index external calibration. Kew Living Collections sample sets (6) were run in two batches of three randomised-blocks, two months apart. This approach was used to assess robustness due to the lack of quality control samples. Samples for *D. elephantipes* of the compound atlas were analysed in three blocks within a single batch.

To identify chromatogram components found in the *Dioscorea* profiles, a mass spectral library was constructed from in-house standards, the NIST '11 MS library (National Institute of Standards and Technology, USA) and the Golm Metabolome Database (GMD)<sup>45</sup>, with additional manual searches of MassBank<sup>46</sup>, Human Metabolome Database (HMDB)<sup>47</sup> and the Yeast Metabolome Database (YMDB)<sup>48</sup>. Component peak identification and spectral deconvolution was performed using the Automated Mass Spectral Deconvolution and Identification System (AMDIS v2.71, NIST); using Kovat's retention indices (RI) and MS for identification using the metabolomics reporting guidelines<sup>49,50</sup>. Each compound was assigned a representative ion and response areas were integrated and expressed relative to internal standard.

Linearity was assessed by following the standard extraction procedure on different amounts of material (5, 10 and 20 mg) and the range further enhanced by taking a range of aliquots (10, 20, 50, 100 and 200  $\mu$ L). Recovery was assessed by expressing measurements from each extract of sample (10 mg) as a percentage of the total following three sequential extractions. All measurements for method development were conducted in triplicate and relative to internal standard. A shikimic acid standard series at 10, 20, 50, 100 and 200  $\mu$ g from methanolic stock was made on three occasions and analysed as per samples. Ion 255 was chosen to be representative.

**Statistical analysis and visualisation of the data.** All data analyses were performed using XLSTAT add-ins (Addinsoft) within Microsoft Excel. Generalised Procrustes Analysis (GPA) was performed using the Commandeur algorithm with 300 simulations. Agglomerative Hierarchical Clustering (AHC) was performed via Spearman dissimilarity with complete linkage. Principal Component Analysis (PCA) was conducted on the Spearman correlation matrix.

Kruskal-Wallis' one way analysis of variance was performed following normality and variance testing. Monte Carlo permutations (10,000) were used for p-value calculation. Conover-Iman post hoc tests ( $\alpha = 0.05$ ) were Bonferroni-corrected and selection of most discriminatory metabolites based on the number of groups generated. All univariate tests were two-tailed. More detailed explanations on the choice of statistical tests are provided in Supplementary Methods.

## References

- Govaerts, R., Wilkin, P. & Saunders, R. M. K. *World Checklist of Dioscoreales: Yams and Their Allies*. (Kew Publishing, 2007). Available at: <http://apps.kew.org/wcsp/home.do>. (Accessed: 23<sup>rd</sup> Sept. 2013).
- Lebot, V. Section III. *Yams In Tropical Root and Tuber Crops: Cassava, Sweet Potato, Yams and Aroids* 181–275, doi: 10.1079/9781845934248.0000 (CABI, 2008).
- Asiedu, R. & Sartie, A. Crops that feed the world 1. Yams. *Food Secur.* **2**, 305–315, doi: 10.1007/s12571-010-0085-0 (2010).
- Food and Agriculture Organization of the United Nations. FAOSTAT Database on Agriculture. (2013) Available at: <http://faostat3.fao.org/>. (Accessed: 28th June 2014).
- Mignouna, H. D., Abang, M. M. & Asiedu, R. Harnessing modern biotechnology for tropical tuber crop improvement: yam (*Dioscorea* spp.) molecular breeding. *African J. Biotechnol.* **2**, 478–485, doi: 10.5897/AJB2003.000-1097 (2003).
- Martin, F. W. & Degras, L. Tropical yams and their potential: part 6. Minor cultivated *Dioscorea* species. *Agricultural Handbook Number 538* (Science and Education Administration, United States Department of Agriculture in cooperation with Agency for International Development, 1978). Available at: <http://naldc.nal.usda.gov/naldc/download.xhtml?id=CAT87209435&content=PDF>. (Accessed: 23<sup>rd</sup> Sept. 2013).
- Shah, N. C. My experiences with the herbal plants & drugs as I knew Part XVI: *Dioscorea* & *Costus*. *Herb. Tech Ind.* 21–30 (2010).



8. Coursey, D. G. The civilizations of the yam: interrelationships of man and yams in Africa and the Indo-Pacific region. *Archaeol. Phys. Anthropol. Ocean.* **7**, 215–233, doi: 10.1002/j.1834-4453.1972.tb00158.x (1972).
9. Ramawat, K. G., Dass, S. & Mathur, M. In *Herbal Drugs: Ethnomedicine to Modern Medicine* (ed. Ramawat, K. G.) Ch. 2, 7–32, doi: 10.1007/978-3-540-79116-4\_2 (Springer, 2009).
10. Crabbé, P. Some aspects of steroid research based on natural products from plant origin. *Bull. des Sociétés Chim. Belges* **88**, 345–358, doi: 10.1002/bscb.19790880511 (1979).
11. Securing livelihoods through yams in *Proceedings of a Technical Workshop on Progress in Yam Research for Development in West and Central Africa held in Accra, Ghana, 11–13 September 2007* (eds. Nkamleu, B., Annang, D. & Bacco, N.M.) 329pp. (IFAD TAG 704, International Institute of Tropical Agriculture (IITA), 2009). Available at: [http://www.iita.org/c/document\\_library/get\\_file?uuid=727e968c-a9d3-4057-b984-9f3cc79a1432&groupId=25357](http://www.iita.org/c/document_library/get_file?uuid=727e968c-a9d3-4057-b984-9f3cc79a1432&groupId=25357). (Accessed: 23<sup>rd</sup> Sept. 2013).
12. Mignouna, H. D., Abang, M. M., Asiedu, R. & Geeta, R. True yams (Dioscorea): a biological and evolutionary link between eudicots and grasses. *Cold Spring Harb. Protoc.* **4**, 1–8, doi: 10.1101/pdb.em0136 (2009).
13. Osunde, Z. D. Minimising postharvest losses in yam (Dioscorea spp.): treatments and techniques in *Using Food Science and Technology to Improve Nutrition and Promote National Development* (eds. Robertson, G. L. & Lupien, J. R.) Ch. 12, 12pp. (International Union of Food Science & Technology, 2008). Available at: <http://www.iufost.org/publications/books/documents/Osundechapter12.pdf>. (Accessed: 26th Nov. 2013).
14. Diop, A. *Storage and Processing of Roots and Tubers in the Tropics*. (Food and Agriculture Organization of the United Nations, 1998). Available at: <http://www.fao.org/docrep/X5415E/X5415E00.htm>. (Accessed: 26th Nov. 2013).
15. Bhattacharjee, R. et al. In *Wild Crop Relatives: Genomic and Breeding Resources, Industrial Crops* (ed. Kole, C.) Ch. 4, 71–96, doi: 10.1007/978-3-642-21102-7\_4 (Springer - Verlag, 2011).
16. Sahoré, D. A., Nemlin, G. J. & Kamenan, A. Changes in nutritional properties of yam (Dioscorea spp.), plantain (Musa spp.) and cassava (Manihot esculenta) during storage. *Trop. Sci.* **47**, 81–88, doi: 10.1002/ts.200 (2007).
17. Heftmann, E., Bennett, R. D. & Bonner, J. Biosynthesis of diosgenin in Dioscorea tubers. *Arch. Biochem. Biophys.* **2**, 13–16, doi: 10.1016/0003-9861(61)90211-9 (1962).
18. Coursey, D. G. *Yams: An Account of the Nature, Origins, Cultivation and Utilisation of the Useful Members of the Dioscoreaceae*. (Longmans, Green and Co. Ltd., 1967).
19. Martin, F. W. The species of Dioscorea containing sapogenin. *Econ. Bot.* **23**, 373–379, doi: 10.1007/BF02860683 (1969).
20. NIIR Board of Consultants and Engineers. In *Cultivation and Processing of Selected Medicinal Plants*, Ch. 23, 270–292 (Asia Pacific Business Press Inc. 2006).
21. Singh, K. N. & Kaushal, R. Comprehensive notes on commercial utilization, characteristics and status of steroid yielding plants in India. *Ethnobot. Leaflet* **11**, 45–51 (2007).
22. Wilkin, P., Burrows, J., Burrows, S., Muasya, A. M. & Wyk, E. van. A critically endangered new species of yam (Dioscorea strydomiana Wilkin, Dioscoreaceae) from Mpumalanga, South Africa. *Kew Bull.* **65**, 421–433, doi: 10.1007/s12225-010-9227-y (2010).
23. Singh, Gopichand R. D., Meena, R. L., Kaul, V. K. & Singh, B. Influence of manure and plant spacing on growth and yield of Dioscorea deltoidea Wall: an endangered species. *J. Med. Plant Stud.* **1**, 184–190 (2013).
24. Part II: Articles on regional aspects of medicinal plants use in *Medicinal Plants for Forest Conservation and Health Care* (eds. Bodeker, G., Bhat, K. K. S., Burley, J. & Vantomme, P.) 78–158 (Food and Agriculture Organization of the United Nations, 1997). Available at: <http://www.fao.org/3/a-w7261e.pdf>. (Accessed: 22nd Aug. 2014).
25. Craufurd, P. Q., Summerfield, R. J., Asiedu, R. & Vara Prasad, P. V. Dormancy in yams. *Exp. Agric.* **37**, 147–181, doi: 10.1017/S001447970100206X (2001).
26. Sheikh, N., Kumar, Y., Misra, A. K. & Pfoze, L. Phytochemical screening to validate the ethnobotanical importance of root tubers of Dioscorea species of Meghalaya, North East India. *J. Med. Plants Stud.* **1**, 62–69 (2013).
27. Champagne, A. et al. Carotenoid profiling of tropical root crop chemotypes from Vanuatu, South Pacific. *J. Food Compos. Anal.* **23**, 763–771, doi: 10.1016/j.jfca.2010.03.021 (2010).
28. Champagne, A., Hilbert, G., Legendre, L. & Lebot, V. Diversity of anthocyanins and other phenolic compounds among tropical root crops from Vanuatu, South Pacific. *J. Food Compos. Anal.* **24**, 315–325, doi: 10.1016/j.jfca.2010.12.004 (2011).
29. Dunn, W. B. & Ellis, D. I. Metabolomics: current analytical platforms and methodologies. *TrAC Trends Anal. Chem.* **24**, 285–294, doi: 10.1016/j.trac.2004.11.021 (2005).
30. Du, X. & Zeisel, S. H. Spectral deconvolution for gas chromatography mass spectrometry-based metabolomics: current status and future perspectives. *Comput. Struct. Biotechnol. J.* **4**, doi: 10.5936/CSBJ.201301013 (2013).
31. Lu, H., Dunn, W. B., Shen, H., Kell, D. B. & Liang, Y. Comparative evaluation of software for deconvolution of metabolomics data based on GC-TOF-MS. *Trends Anal. Chem.* **27**, 215–227, doi: 10.1016/j.trac.2007.11.004 (2008).
32. Lisek, J., Schauer, N., Kopka, J., Willmitzer, L. & Fernie, A. R. Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.* **1**, 387–96, doi: 10.1038/nprot.2006.59 (2006).
33. Halket, J. M. et al. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J. Exp. Bot.* **56**, 219–43, doi: 10.1093/jxb/eri069 (2005).
34. Raz, L. Yam family in *Flora of North America* (eds. Flora of North America Editorial Committee) 26, 479–480 (Oxford University Press, 2003). Available at: [http://www.efloras.org/florataxon.aspx?flora\\_id=1&taxon\\_id=10280](http://www.efloras.org/florataxon.aspx?flora_id=1&taxon_id=10280). (Accessed: 19<sup>th</sup> June 2015).
35. Wilkin, P. et al. A plastid gene phylogeny of the yam genus, Dioscorea: Roots, fruits and Madagascar. *Syst. Bot.* **30**, 736–749, doi: 10.1600/036364405775097879 (2005).
36. Girma, G. et al. Next-generation sequencing based genotyping, cytometry and phenotyping for understanding diversity and evolution of guinea yams. *Theor. Appl. Genet.* **127**, 1783–1794, doi: 10.1007/s00122-014-2339-2 (2014).
37. IPGRI & IITA. *Descriptors for yam (Dioscorea spp.)*. (International Institute of Tropical Agriculture, Ibadan, Nigeria/International Plant Genetic Resources Institute, Rome, Italy, 1997). Available at: [http://www.bioversityinternational.org/uploads/tx\\_news/Descriptors\\_for\\_Yam\\_Dioscorea\\_spp.\\_310.pdf](http://www.bioversityinternational.org/uploads/tx_news/Descriptors_for_Yam_Dioscorea_spp._310.pdf). (Accessed: 23<sup>rd</sup> Sept. 2013).
38. Viruel, J. et al. Late Cretaceous – Early Eocene origin of yams (Dioscorea, Dioscoreaceae) in the Laurasian Palaeartic and their subsequent Oligocene – Miocene diversification. *J. Biogeogr.* **43**, 750–762, doi: 10.1111/jbi.12678 (2015).
39. Gao, X. et al. Phylogeny of Dioscorea sect. Stenophora based on chloroplast matK, rbcL and trnL-F sequences. *J. Syst. Evol.* **46**, 315–321, doi: 10.3724/SPJ.1002.2008.08007 (2008).
40. Ohira, H., Torii, N., Aida, T. M., Watanabe, M. & Smith, R. L. Rapid separation of shikimic acid from Chinese star anise (Illicium verum Hook. f.) with hot water extraction. *Sep. Purif. Technol.* **69**, 102–108, doi: 10.1016/j.seppur.2009.07.005 (2009).
41. Victor, J. E. & Dold, A. Dioscorea elephantipes (L'Hér.) Engl. National Assessment: Red List of South African Plants (2004). Available at: <http://redlist.sanbi.org/species.php?species=1777-12>. (Accessed: 8th Nov. 2013).
42. Lytovchenko, A. et al. Application of GC-MS for the detection of lipophilic compounds in diverse plant tissues. *Plant Methods* **5**, 4, doi: 10.1186/1746-4811-5-4 (2009).
43. t'Kindt, R., Morreel, K., Deforce, D., Boerjan, W. & Van Boclaer, J. Joint GC-MS and LC-MS platforms for comprehensive plant metabolomics: repeatability and sample pre-treatment. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* **877**, 3572–80, doi: 10.1016/j.jchromb.2009.08.041 (2009).
44. Enfissi, E. M. a. et al. Integrative transcript and metabolite analysis of nutritionally enhanced DE-ETIOLATED1 downregulated tomato fruit. *Plant Cell* **22**, 1190–1215, doi: 10.1105/tpc.110.073866 (2010).

45. Kopka, J. *et al.* GMD@CSB.DB: The Golm metabolome database. *Bioinformatics* **21**, 1635–1638, doi: 10.1093/bioinformatics/bti236 (2005).
46. Horai, H. *et al.* MassBank: A public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **45**, 703–714, doi: 10.1002/jms.1777 (2010).
47. Wishart, D. S. *et al.* HMDB 3.0-The Human Metabolome Database in 2013. *Nucleic Acids Res.* **41**, 801–807, doi: 10.1093/nar/gks1065 (2013).
48. Jewison, T. *et al.* YMDB: The yeast metabolome database. *Nucleic Acids Res.* **40**, 815–820, doi: 10.1093/nar/gkr916 (2012).
49. Fernie, A. R. *et al.* Recommendations for reporting metabolite data. *Plant Cell* **23**, 2477–82, doi: 10.1105/tpc.111.086272 (2011).
50. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis. *Metabolomics* **3**, 211–221, doi: 10.1007/s11306-007-0082-2. (2007)
51. Chaïr, H. *et al.* Use of cpSSRs for the characterisation of yam phylogeny in Benin. *Genome* **48**, 674–84, doi: 10.1139/g05-018 (2005).

## Acknowledgements

This work was supported by BBSRC-DTP No. 1638827 and in part the CGIAR Research Program on Roots, Tubers and Bananas (RTB). The authors are also grateful for the help provided by Paul Rees, Sam Phillips, Silke Roch, Anna Bowell, Lara Jewett and John Sitch with sampling at Kew. The authors kindly thank Dr. Urs Eggli for provision of *D. mexicana* material, Peter Brownless for provision of samples from RBGE and Mark Levy for donation of *D. elephantipes* used for compound atlas. P.D.F. and E.J.P. would also like to thank Dr Luis Augusto Becerra Lopez-Lavalle for his introduction and support to the RTB programme.

## Author Contributions

RTB project funds were obtained by P.D.F. BBSRC – DTP studentship was granted to P.D.F., V.S and P.W. Experiments were developed by P.D.F., E.J.P., V.S. and P.W. All experimental work was performed by E.J.P. All authors contributed to the interpretation of the results and preparation of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Price, E. J. *et al.* Metabolite profiling of *Dioscorea* (yam) species reveals underutilised biodiversity and renewable sources for high-value compounds. *Sci. Rep.* **6**, 29136; doi: 10.1038/srep29136 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>