CrossMark

ORIGINAL ARTICLE

Metabolite profiling of yam (*Dioscorea spp.*) accessions for use in crop improvement programmes

Elliott J. Price[1,2] · **Ranjana Bhattacharjee**[3] · **Antonio Lopez-Montes**[3] · **Paul D. Fraser**[1]

Abstract

Introduction Ninety-seven percent of yam (*Dioscorea* spp.) production takes place in low income food deficit countries (LIFDCs) and the crop provides 200 calories a day to approximately 300 million people. Therefore, yams are vital for food security. Yams have high-yield potential and high market value potential yet current breeding of yam is hindered by a lack of genomic information and genetic resources. New tools are needed to modernise breeding strategies and unlock the potential of yam to improve livelihood in LIFDCs.

Objectives Metabolomic screening has been undertaken on a diverse panel of *Dioscorea* accessions to assess the utility of the approach for advancing breeding strategies in this understudied crop.

Methods Polar and lipophilic extracts from tubers of accessions from the global yam breeding program have been comprehensively profiled via gas chromatography-mass spectrometry.

Results A visual pathway representation of the measured yam tuber metabolome has been delivered as a resource for biochemical evaluation of yam germplasm. Over 200 compounds were routinely measured in tubers, providing a major advance for the chemo-typing of this crop. Core biochemical redundancy concealed trends that were only elucidated following detailed mining of global metabolomics data. Combined analysis on leaf and tuber material identified a subset of metabolites which allow accurate species classification and highlighted the potential of predicting tuber composition from leaf profiles. Metabolic variation was accession-specific and often localised to compound classes, which will aid trait-targeting for metabolite markers.

Conclusions Metabolomics provides a standalone platform with potential to deliver near-future crop gains for yam. The approach compliments the genetic advancements currently underway and integration with other '–omics' studies will deliver a significant advancement to yam breeding strategies.

✉ Paul D. Fraser
p.fraser@rhul.ac.uk

Elliott J. Price
elliott.james.price@gmail.com

Ranjana Bhattacharjee
r.bhattacharjee@cgiar.org

Antonio Lopez-Montes
a.lopez-montes@cgiar.org

[1] Royal Holloway University of London, Egham, Surrey TW20 0EX, UK

[2] Royal Botanic Gardens, Kew, Richmond, Surrey TW20 3AB, UK

[3] International Institute of Tropical Agriculture, Oyo Road, PMB 5320 Ibadan, Nigeria

1 Significance statement

The tubers of numerous *Dioscorea* (yam) species are an important staple of many tropical countries yet; numerous constraints hinder breeding, including a lack of genetic resources. Biochemical phenotyping accessions of the breeding program can be rapidly implemented to advance yam breeding strategies. The present study illustrates utility of

metabolomics as a useful tool and resource to augment the breeding of new yam varieties with improved consumer and agronomic traits.

2 Introduction

Over the last two decades global yam production has doubled, reaching a total production of 60 million tonnes and having a net value internationally of ~$12 billion. Approximately 97% of production is from low income food deficit countries (LIFDCs) (Food and Agriculture Organization of the United Nations 2015), with Western Africa accounting for over 90% of total production each year in this period. In spite of high total production, no Western African country is in the top five countries for yield of this crop per area farmed (Food and Agriculture Organization of the United Nations n.d.).

Compared to cassava, yam is four times more expensive per calorie to produce (Diop 1998), with increased labour requirements (Nweke and Ezumah 1992) and lower yields (Kenyon et al. 2006). Despite these comparative parameters, in regions of low-technology farming, yams produce the highest amount of food per hectare and per season (Kumar 2014). Unlike other root and tuber crops, yams may be stored for a period of 4–6 months at ambient temperature (Bricas and Attaie 1997; Knoth 1993) and as such, yams are vital for the food security of growing regions. To date yams as a food source and cash crop have been understudied and underutilised (Mignouna et al. 2009; Nkamleu et al. 2009). This is surprising considering their high yield potential, high food and market values and numerous post-harvest options for storage and utilisation (Asiedu and Sartie 2010). These are all properties suited to the combating of food insecurity, achieving the United Nations Sustainable Development Goals and the sustainable intensification of agriculture.

Large species diversity confers a wide range of ecological and agronomic traits within the genus, providing intrinsic opportunity for crop improvements through the exploitation of natural variation (Fernie et al. 2006). Conventional breeding strategies have had significant successes (Mignouna et al. 2003) however, to create the step-change required to combat our ever increasing global population and diminishing resources, modern tools and resources must be developed and implemented. A lack of availability of next generation '–omics' resources and information has hindered application of molecular breeding in yam (Bhattacharjee et al. 2011).

In order to fully exploit the potential of this crop species, resources and tools are required that add value to existing and new discovery pipelines. Adding value to these pipelines will result in better competitive varieties with improved and new traits, as well as increasing the rapidity with which the varieties can be generated. Recent advances in the application of next generation sequencing has enabled capture of genomes from non-model species, including yams, with the development of reference genomes for *D. rotundata* (Tamiru et al. 2017), *D. alata & D. dumetorum* and large-scale screening of genetic diversity via genotyping-by-sequencing (GBS) and whole genome re-sequencing (WGRS) of these species and crop wild relatives to allow trait mapping and identification of single nucleotide polymorphism (SNP) variants (Fukushima et al. 2009).

Genomics represents one level of cellular regulation when coupled with other '–omics' technologies such as transcriptomics, proteomics, metabolomics and phenomics the power of trait characterisation is greatly improved (Fukushima et al. 2009). This is particularly true of metabolomics where traits of interest can, in many cases be directly associated with metabolite composition (Bino et al. 2004).

The quality traits of yams are understudied with regards to breeding, largely due to the complicated genetic architecture of *Dioscorea* hindering association mapping. Progression has been slow, even when implementing next-generation technologies, due to the lack of fundamental molecular knowledge. For example, the ploidy status of many accessions in current breeding programs is unknown. Only recently have the chromosome numbers of major edible species been clarified and definitive studies assessing ploidy have required inference through marker segregation (Arnau et al. 2009; Bousalem et al. 2006; Scarcelli et al. 2005). Furthermore, dioecy, intra and inter-species hybridisation and the large phenotypic plasticity of yams has made the scoring and association of numerous interdependent traits a convoluted process.

Metabolomics platforms produce large-scale biochemical phenotypes that can be representative of quality traits. Association of phenotype to metabolite profiles can provide measureable markers comparable to genetic quantitative trait loci (QTL) e.g. metabolite quantitative trait loci (mQTL) which can be recorded through breeding program both independent of underlying genetic mechanisms and combined with typical genomic analysis to enhance linkage of genotype with phenotype, such metabolite-based genome-wide association analysis (mGWAS) for functional genomics (Adamski and Suhre 2013; Luo 2015).

A pilot study has been undertaken to investigate utility of metabolomics within yam breeding programs. Metabolite profiling a set of *Dioscorea* accessions routinely used in yam breeding programmes has been illustrated. Such approaches will facilitate the capture of biochemical diversity within this crop species, enable comparisons across other tuber food crops and contribute to the deciphering of underlying biochemical and molecular mechanisms associated with agronomic and consumer traits.

A GC-MS platform was opted for as costs are not as prohibitive for developing countries when compared to LC-MS,

NMR and other approaches. The relatively reduced complexity of data analysis, ease of operation and higher reproducibility would hopefully facilitate skill transferability, education and capacity building enabling incorporation of metabolomics in yam breeding programmes at national and regional levels.

3 Results

Using a GC-MS based platform, tuber material from 49 accessions across four *Dioscorea* species from the global breeding program was analysed (Table 1). Collectively, 152 metabolites (Table S1) were identified based on the guidelines produced by the international metabolomics society (Fernie et al. 2011; Sumner et al. 2007). A customised *Dioscorea* GC-MS metabolite library, previously built using extracts from leaf tissue (Price et al. 2016), was extended to include non-polar metabolites providing comprehensive coverage of key sectors of *Dioscorea* metabolome found in tubers. Additionally, 89 unknown molecular features have also been measured (Table S2) to allow comprehensive platform coverage. Using these data, a model of the tuber biochemical pathways has been created (Fig. 1).

Of the 152 identified compounds, 41 were not detected in one or more species, with 9 being unique to an individual species and a further 7 uniquely absent in only one species. However, these differences were accession specific and it was not possible to classify species based on presence/absence profiles. Twenty-eight of the identified compounds were detectable in both polar and non-polar extracts (Table S1), which is not unexpected given the crude nature of separation between phases. Multivariate statistical analysis was conducted on both polar and non-polar extracts independently along with the combined data set.

Clustering on the combined dataset showed that total metabolite profiles and the composition of identified metabolites provide the same separation of material (Figure S1). Species were largely distinguishable, though variants of *D. rotundata* and *D. cayennensis* showed overlap and grouped together (Fig. 2). Discrimination among species was also apparent when profiling polar extracts (Fig. 3a), with separation along principal component (PC) 1 driven by sugars and that on PC2 via organic acids, yet not possible following analysis on the non-polar dataset (Fig. 3b). Notably, accession *TDd 3774* (*D. dumetorum*) largely diverged from the sample set due to an abundance of fatty acids (Figure S2) however, presented a polar profile indiscriminate from many other accessions of the *D. dumetorum* (Fig. 3a). Excluding *TDd 3774* showed that within non-polar profiles two groups were apparent: those possessing an abundance of sterols and fatty acids versus those with low amounts (Figure S3).

Representation of the abundance of the most discriminatory variables in the most diverse accessions highlighted complex distributions of metabolites across accessions (Figure S4). As per qualitative differences, quantitative differences were largely accession specific. Metabolite–metabolite correlation on *D. dumetorum* accessions showed that many biochemically-related compounds correlate (Fig. 4, Figure S5), e.g. amino acids positively correlated within their own class and with nucleotides and cofactors and secondary metabolites and fatty acids positively correlate with their own compound class whilst fatty acid derivatives negatively correlate with other fatty acids.

The platform can provide robust analysis relevant to underlying metabolic pathways and thus has potential to investigate metabolic perturbation due to different biotic and abiotic stresses or investigate postharvest physiological deterioration. For example, the accessions of *D. dumetorum* could be segregated into three lineages (Figure S6). Class 1 (red branch) showed increased levels of amino acids and phosphate, Class 2 (green branch) were distinguished by glycerol mono-pentadecanoate and homogentisic acid whilst Class 3 (blue branch) showed increased abundances of sugars and fatty acids.

3.1 Analysis of metabolite gradients over tuber sub-sections

Different sections of yam tuber show different enzymatic activities (Oluoha 1988; Wellington and Ahmad 1993; Wheatley et al. 2002) and phenotypic differences reportedly observed when mini-sett sections are used as planting material. To investigate the impact on metabolic functions a representative accession was selected per each species and analyses conducted on the head, middle, tail and skin sections of tuber independently.

In total, 162 metabolite features were detected and 93 identified (Table S3). The skin from all tubers displayed a similar profile, distinguished from other sections of tuber largely due to reduced amino acid content (Figure S7). Only a limited number of metabolites showed significant differences across different tuber sections; however these were species-specific with no single metabolite showing significant differences across sections for all species (Figure S8). All tubers, except those of *D. rotundata*, showed a gradient of malic acid concentration from highest abundance in tail to lowest in head. *Dioscorea bulbifera* showed a low-to-high gradient for many amino acids from head-to-tail and the same gradient was evident in *D. dumetorum* but for sugars. Sections of *D. rotundata* showed the highest number of compounds which were significantly different in Sections (16) however; there were not clear distribution patterns or metabolite gradients evident for many compounds. Despite different enzymatic activity having been shown

Table 1 *Dioscorea* accessions sourced from the IITA Yam Breeding Unit of the International Institute of Tropical Agriculture

Dioscorea species [in text]	Abbreviation	Accession Numbers
alata L. [*D. alata*]	TDa	98/01176
		297
		98/001166
		291
		00/00194
bulbifera L. [*D. bulbifera*][a]	TDb	3059
		3079
		3072
		3688
		3048
cayennensis Lam. subsp. *cayennensis* [*D. cayennensis*]	TDc	04-71-2
		03–5
		95 – 17
		04-97-4
dumetorum (Kunth) Pax [*D. dumetorum*]	TDd	08-38-8
		08-36-14
		3108
		08-14-42
		1315
		3112
		4118
		4088
		08-37-12
		3774
		08-37-27
		08-37-16
		3104
		08-38-57
		3947
		05–6
		3100
		08-36-12
		08-38-18
		3109
		3648
		08-14-6
		08-36-88
		08-13-1
		08-3879
cayennensis Lam. subsp. *rotundata* (Poir.) J. Miège [*D. rotundata*]	TDr	EHObia
		99/02607
		omi-Efun
		97/00917
		95/01932
		97/00793
		97/00777
		04-219
		EHuRu
		Ponna

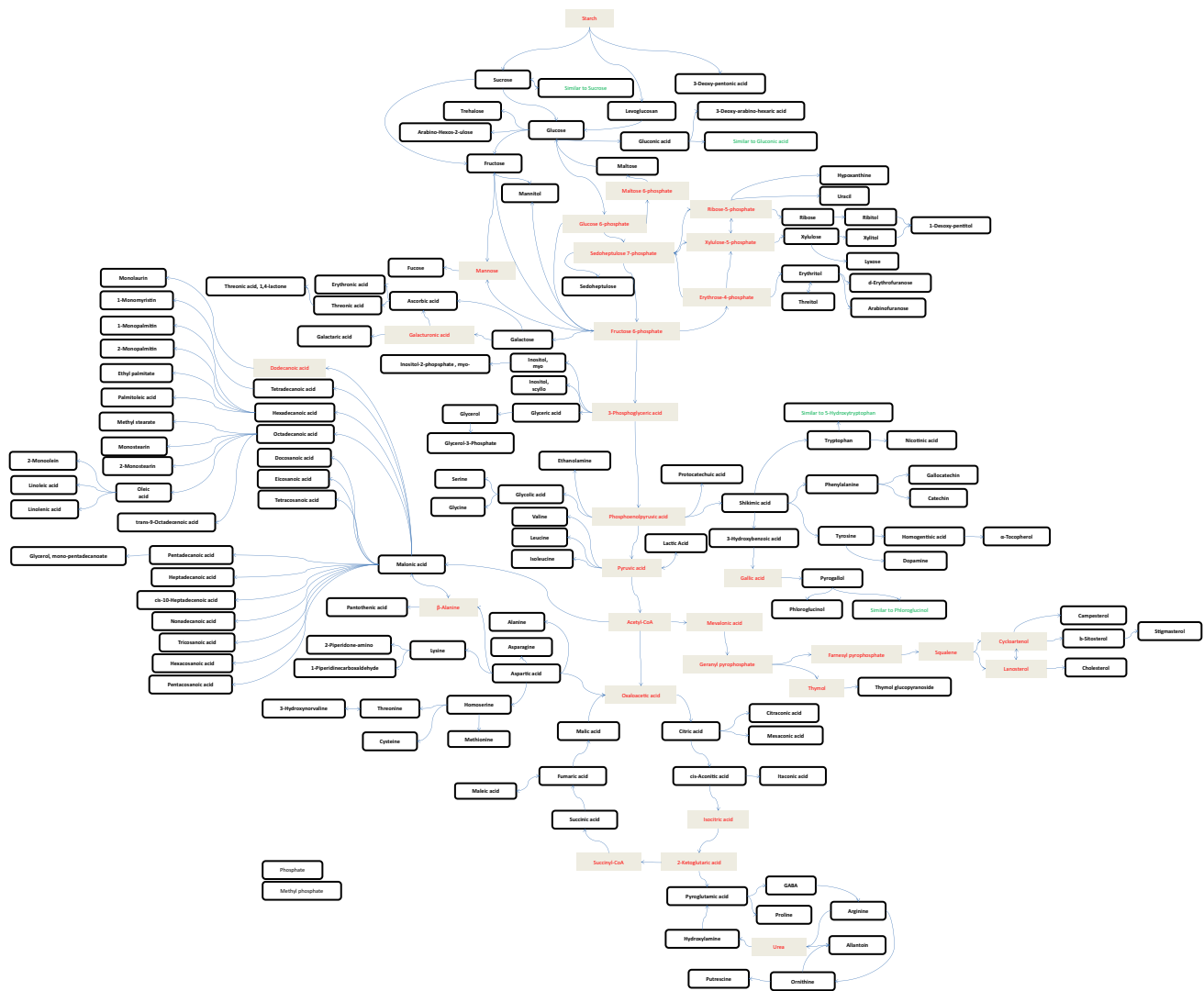[a]Aerial tubers (bulbils) were provided for accessions of *D. bulbifera*

Fig. 1 Map of measured primary metabolome of *Dioscorea* tubers. Visual representation of biochemical pathways of compound recorded in tuber extracts by the GC-MS profiling platform. Metabolites with preliminary identification are shown in green. Metabolites not detected by the platform are shown in red

from different portions of yam, these previous studies could not define patterns of biochemical properties across head, middle and tail sections. Additionally, the results were not unexpected as work across sections of potato noted that large differences between individual tubers, plants and growing season limited interpretation (Merlo et al. 1993).

3.2 Tuber and leaf comparison

Using the method developed, a second crop at a different geographical location was produced to provide a comparison between tuber and leaf material. In this case the material was generated under polytunnel conditions in the UK. In total 337 metabolite features were measured via

the platform, 258 in tuber and 312 in leaf respectively (Table S4). Preliminary analysis showed that similar pattern of species separation was evident for tuber and leaf material (Figure S9) and so a PLS-DA model using species as classifier was created (Figure S10). The model allowed species classification (90% accurate on validation set) with only a sample of *D. cayennensis* wrongly predicted as being *D. rotundata*. The top 50 variables of importance in the model were selected and used in a new model to distinguish metabolites driving species delineation (Fig. 5). *Dioscorea dumetorum* was defined by fatty acids, *D. rotundata* and *D. cayennensis* by TCA cycle intermediates and phosphate, and *D. alata* and *D. bublifera* both largely by sugars.
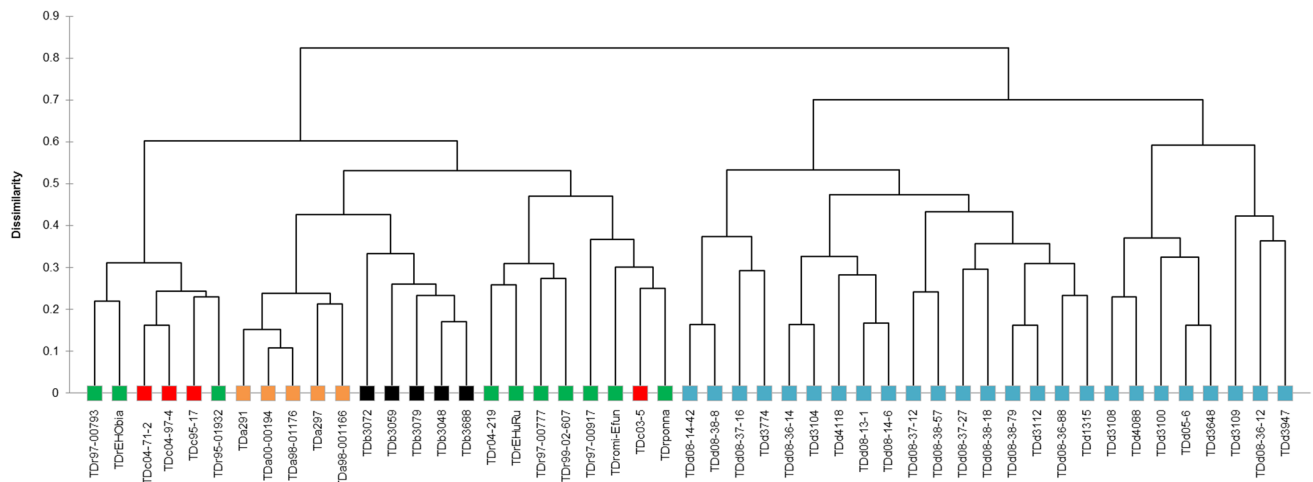
Fig. 2 Dendrogram of *Dioscorea* accessions using identified compounds measured by the GC-MS platform. Complete-linkage agglomerative hierarchical clustering (AHC) on the Spearman dissimilarity matrix of mean (n = 3) metabolite abundances recorded in tuber extracts. *Dioscorea dumetorum* (blue); *D. rotundata* (green); *D. alata* (yellow); *D. cayennensis* (red) and *D. bulbifera* (black) recorded in triplicate

4 Conclusions

A diverse panel of accessions from four *Dioscorea* species routinely used in yam breeding programmes was used to create a semi-automated GC-MS library for chemo-typing. The library provides comprehensive coverage of intermediary metabolism and abundant specialised metabolites. A pathway map of the measured yam tuber metabolome was created (Fig. 1). *Dioscorea* tuber presents many challenges for metabolomics analysis due to high starch content yet, simple modifications to conventional methods (such as an extended centrifugation time and chilling of extracts prior to evaporation) allowed robust and repeatable metabolite profiling.

Exploratory analysis of phenolic compounds (Champagne et al. 2011) and carotenoids (Champagne et al. 2010) in some *Dioscorea* species showed both inter- and intra- species diversity and emphasised the potential of global metabolomic investigation. Previous study evidenced that species could largely be discriminated by metabolite profiling polar extracts of leaf (Price et al. 2016), which has further been shown true for tuber material in this work (Fig. 3).

The chemotaxonomic relationship of species (Fig. 2) also closely matches that of phylogenetic studies, with *D. dumetorum* being most divergent and monophyletic support for *D. bulbifera* and *D. alata*, whilst *D. rotundata* and *D. cayennensis* are more genetically similar (Chaïr et al. 2005; Sartie et al. 2012). Other studies distinguished *D. rotundata* from *D. cayennensis* using molecular markers (Ramser et al. 1997), morphological characteristics and isozyme markers (Dansi et al. 2000; Hamon and Touré 1990) with discriminatory enzymatic descriptors related to central carbon metabolism of leaf. Whilst the measured tuber metabolome did not discriminate between the two cultigens (Fig. 2), inclusion of leaf material may offer increased resolution, as evidenced in the exploratory study conducted (Fig. 3).

Following global metabolomics approaches, redundancy within the biochemical profiles of these breeding materials concealed trends (Figure S1) and limited data interpretation. This is a common bottleneck for analysis of metabolomics investigations (Kuehne et al. 2015). However, detailed analyses reveal variation across broad pathways in an accession-specific manner (Figure S4). Metabolite profiling of both tuber and leaf material revealed that the same trends could be observed in leaf as tuber material (Figure S9) and though a small set of accessions was studied, a subselection of metabolites could be used to accurately predict species (Fig. 5, Figure S10). Similar trends have been observed in other tuberous crops such as cassava, potato and sweet potato. Future analysis using bigger datasets will be needed to validate if the leaf profiles can be representative or even predictive of tuber metabolite composition. If so, initial phenotypic screening of breeding programs could be undertaken on leaf material and profiling of root and tuber crops would be significantly more rapid; conducted prior to tuber formation; requiring less labour for material harvesting and analyses being cheaper and easier.

Identifying biochemical signatures of phenotypic traits would allow metabolite-marker based breeding (Fernie and Schauer 2009). Given the lack of genomic information and genetic resources that currently hinders the applicability of gene based marker-assisted breeding (Bhattacharjee et al. 2011; Hana; Chaïr et al. 2016; White et al. 2016), implementing metabolomics offers great potential for

Fig. 3 GPA of **a** polar and **b** non-polar extracts from tuber material of *Dioscorea* accessions. Consensus (n = 3) configurations following Generalised Procrustes Analysis conducted on all metabolite features recorded in the respective phases. *Dioscorea dumetorum* (blue); *D. rotundata* (green); *D. alata* (yellow); *D. cayennensis* (red) and *D. bulbifera* (black) recorded in triplicate (n = 3)
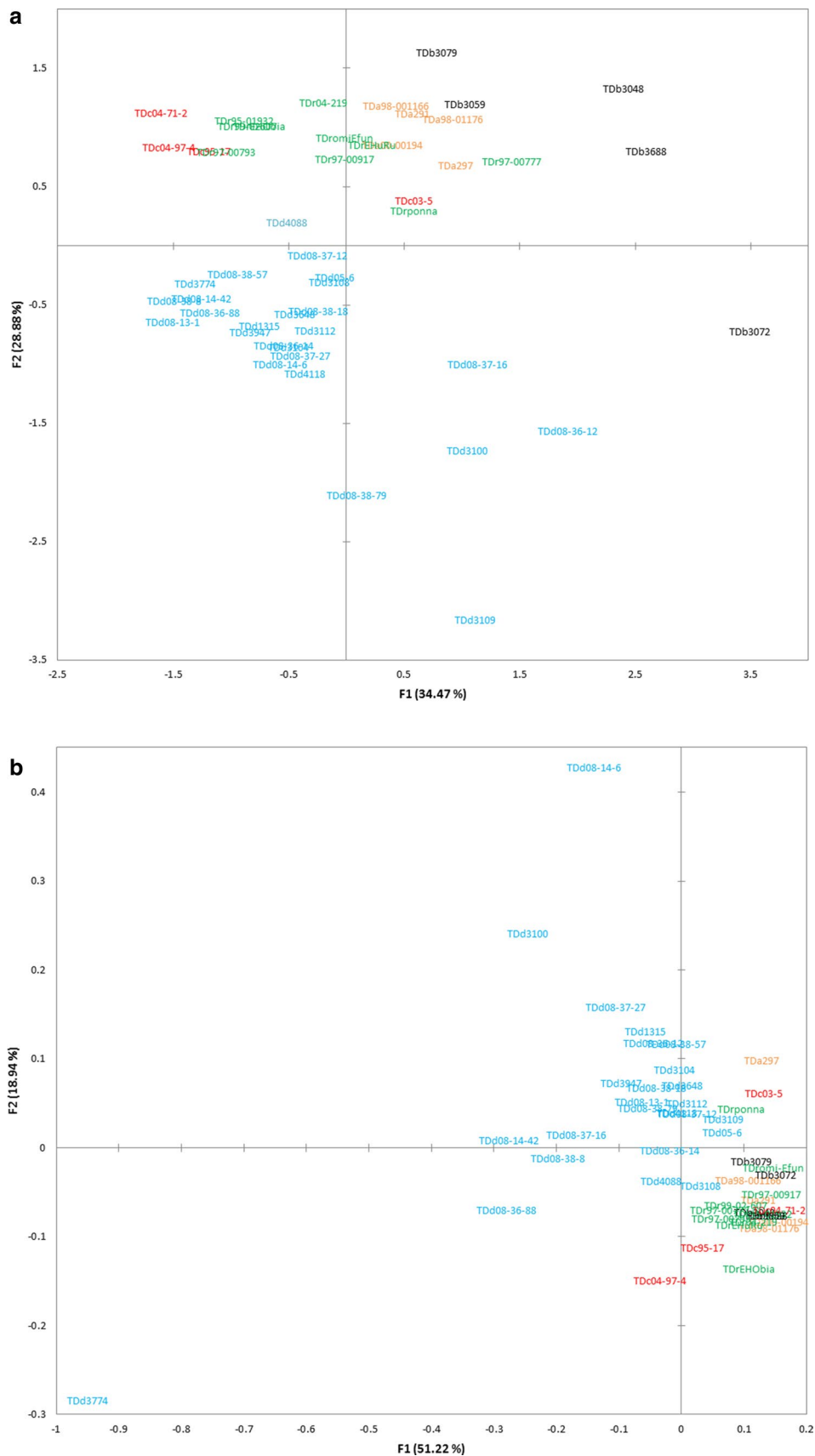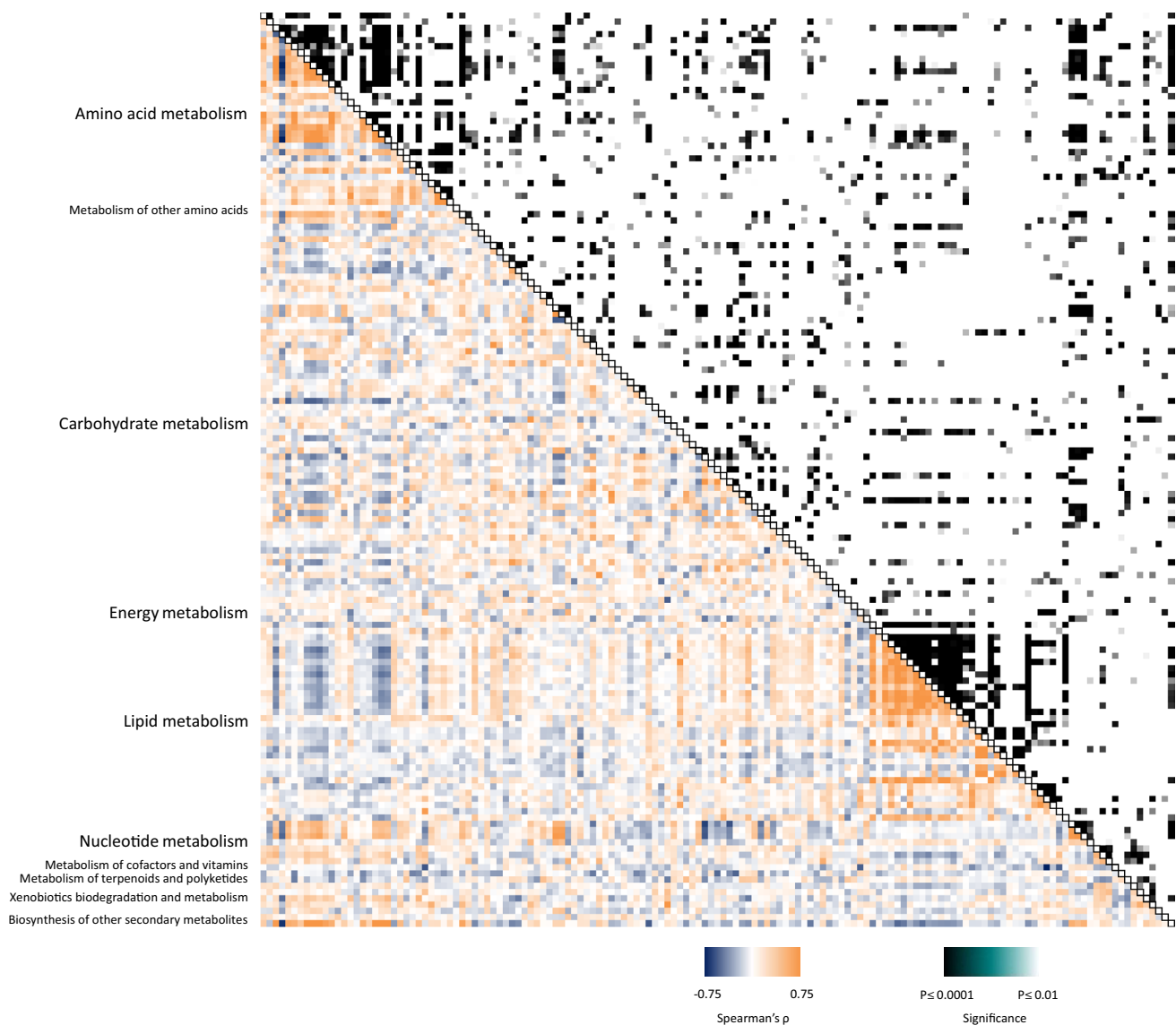
Fig. 4 Heat-map of metabolite–metabolite correlation for *D. dumeto-rum*. Spearman correlation between metabolites across all replicates of *D. dumetorum* (25 accessions, n=3) shows that compounds typically have significant correlations within compound class and among biochemically-related pathways. In the coloured area rectangles represent Spearman's rho and in the black and white area, rectangles represent the respective p-value. Individual metabolites shown in Figure S4

near-future crop gains. Metabolite profiles provide a direct biochemical measurement of quality traits which can be utilised as markers to investigate trait inheritance independent of association with genetic mechanisms. Additionally, information derived from integrating biochemical profiles with the morphological descriptors traditionally recorded throughout breeding programs can subsequently be integrated with genomic data when it becomes available, making the characterisation of genes much more rapid (Tohge and Fernie 2010). The genome sequence of *D. rotundata* has recently been published and sex-linked markers determined (Tamiru et al. 2017). However, transferability of genetic markers to other species is dependent on taxonomic closeness (Tamiru et al. 2015) and differing sex-determination mechanisms between species and environmental plasticity may still hinder breeding. In comparison, metabolite markers may be easier to implement in multi-species crop breeding programmes, such as for yam and are simpler to associate with environmental factors. Additionally, in wheat metabolite markers have been shown to be as equally predictive of complex traits as genetic markers (Riedelsheimer et al. 2012) and more effective when use of both is combined (Ward et al. 2015). Thus, integrating a metabolomics component into current genetics-led yam breeding would hopefully remove some of the constraints realised for typical breeding approaches.

Fig. 5 A reduced PLS-DA model classifies species based on leaf and tuber metabolite profiles. Loading plot of metabolites recorded in leaf and tuber extracts used as variables to predict species. The top 50 variables of importance (VIPs) were selected from an initial PLS-DA model created using all identified metabolites and validated using a random subset of > 25% of variables recorded (Figure S9)
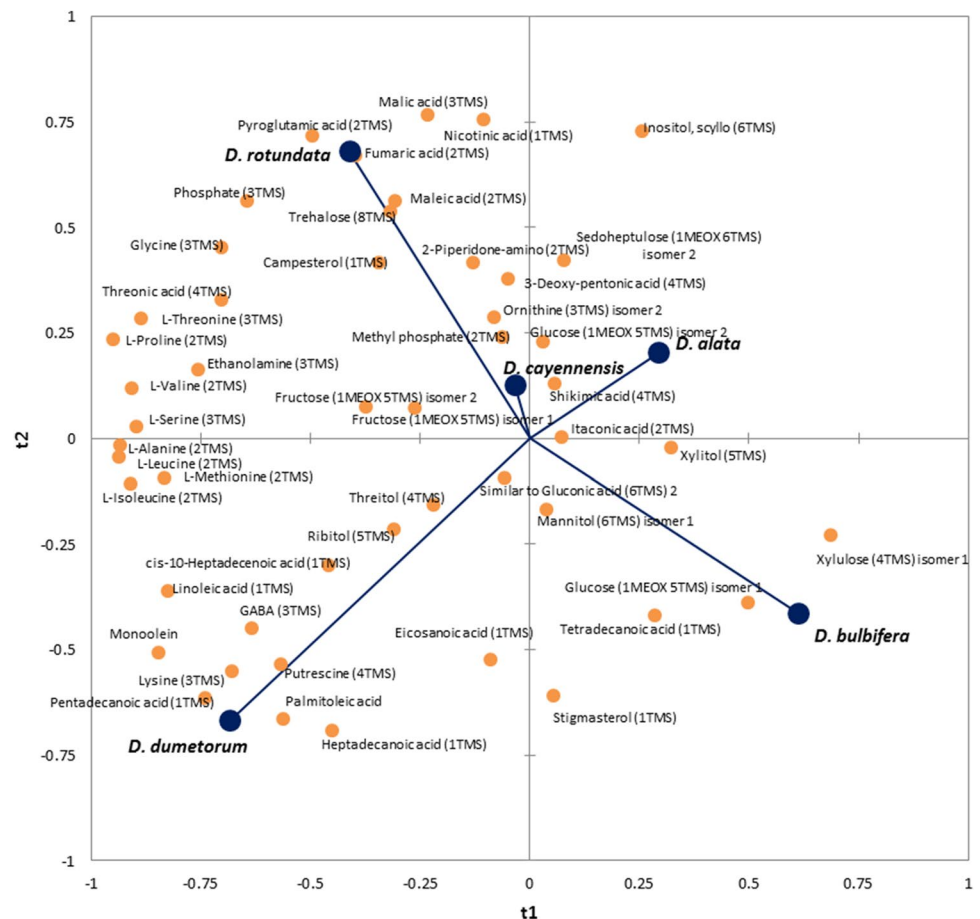
In 2014, Adaramola et al. (2014) integrated ploidy, qualitative phytochemical content and morphological traits to design *D. dumetorum* breeding strategy; however, phytochemical analyses were crude and only a weak association between terpenoid detection and ploidy was found. The relationship between gene dosage and quality traits is complex, with many levels of regulation (transcriptional, translational, enzymatic etc.) and genotype to phenotype relationships indirect, especially in mixed polyploid crops such as yam. Biochemical markers are more closely associated to phenotypic traits and dynamic measurements of metabolome composition are able to represent phenotypic changes. Therefore, metabolome to phenotype relationships can provide a simplified system with which to assess quality traits and provide a stepping stone for interpretation of genotype-phenotype interaction.

Within the limits of the platforms dynamic range it was clear that the total number of metabolites detected in tuber was less than those found in the leaf. Typically the differences were both quantitative and qualitative, with fewer pathway intermediates such as sugar derivatives and a reduced number of phenylpropanoid pathway derived compounds. This is to be expected, given the abundance of sucrose reduces the dynamic range measurable and lower

expression of phenylpropanoid metabolism in tuber. Information on post-harvest storage of root and tuber crops is important for these types of study, with far more metabolomics signatures recorded when the crop is immediately harvested, allowing easier differentiation among species etc. Time-series studies on each species could be undertaken in future studies to investigate crucial aspect of post-harvest deterioration.

Moreover, alternative platforms such as LC- or CE-MS and NMR could be used and would provide enhanced coverage of the metabolome and increase accession demarcation. In particular, enhanced measurement of secondary metabolites would aid more detailed analysis and is advocated when investigating complex quality traits such as disease susceptibility, tolerance and resistance and sensory-related parameters.

This study showcases comprehensive practical analysis of the yam metabolome. Application of this technology to a larger sample set of the breeding program is needed with integration of data from multiple '–omics' platforms to provide a broader systems level knowledge base. Since metabolomics only provides a temporal snapshot of biochemical phenotype, analysis over numerous growing seasons will be required to identify robust quality trait markers. However,

being able to correlate biochemical signatures with numerous agronomic and sensorial traits and investigate inheritance independent of the underlying genetic mechanisms provides potential to speed breeding increasing the rapidity of selection. Interpretation of data from the platform is not compromised by the complicated genetic architecture which slows genomics approaches and in turn provides a segue for future elucidation and association of genotype to phenotype.

This work increases the phytochemical resources available for yam and will facilitate the deciphering of underlying molecular and biochemical mechanisms associated with traits of interest; add value to the discovery pipeline and provide a resource to aid rational design of breeding towards better varieties that are suitable for future sustainable intensification.

5 Methods

5.1 Reagents

All reagents were of analytical grade.

5.2 Plant material

Forty-nine accessions covering four species of *Dioscorea* were sourced from the IITA Yam Breeding Unit of the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria (Table 1) and shipped to the Royal Holloway University of London (RHUL), United Kingdom for further analysis.

For standard extractions tuber from three biological replicates per accession were sectioned laterally and longitudinally into 12; and six representative sections per tuber frozen in liquid nitrogen (Figure S10). Sections were freeze-dried; skin peeled and ground (via a cryogenic mill) to a homogenous powder prior to extraction. All samples were stored at $-80\,°C$ prior to further processing.

Five accessions (*TDa 98-01176, TDb 3059, TDc 04-71-2, TDd 08-14-42, TDr EHuRu*) underwent spatial metabolomics analyses whereby tuber from a fourth biological replicate was sectioned into individual head, middle and tail portions (Figure S10). Each section was then processed as above and analysed with triplicate technical repetition. The peeled skin was also analysed in duplicate.

Additionally, tuber of a fifth replicate for these accessions was divided into mini-setts planted in polytunnel at RHUL, UK. Tuber and leaf material from three plants per accession were harvested following 9 months growth (May 2014–January 2015). Tuber material was processed as per standard extracts, yet on the same day as harvest. Leaf material was processed as previously described (Price et al. 2016).

5.3 Gas Chromatography-Mass Spectrometry (GC-MS) Profiling

Material was extracted, derivatised and analysed via GC-MS as previously described (Price et al. 2016) with modifications: complete phase separation of $MeOH:H_2O/CHCl_3$ required centrifugation (14,000 rpm) for 10 min. Polar phase aliquots were stored overnight at $-20\,°C$ and then dried under inert nitrogen. Samples were subsequently stored at $-80\,°C$. Both polar and non-polar aliquots were analysed; except portions of head, middle and tail, where only polar aliquots were analysed. Only splitless injections onto the GC were undertaken. All matches to NIST database were of a probability greater than 80%. Peaks were only included in analyses if identified in all three replicates of any sample for each study.

5.4 Data processing and statistics

Polar and non-polar datasets were combined; removing duplicates by selecting the maximal response recorded per each compounds present in both phases.

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. Samples which were more visually divergent on the GPA plots were selected for representation in scattergrams. Agglomerative hierarchical clustering (AHC) was performed using complete linkage on the spearman dissimilarity matrix. Partial Least Squares Discriminant Analysis (PLS-DA) was performed on centred and reduced variables and validated by a random subset of ≥ 25% of variables. Metabolite–metabolite correlations were performed via spearman correlation coefficients and heatmaps arranged via assignment of compounds into KEGG pathways. Pathway displays were created in Microsoft Powerpoint. Kruskal–Wallis' one way analysis of variance were 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 the selection of most discriminatory metabolites based on the number of groups generated i.e. the more groups a metabolite segregated the accessions into, the more discriminatory the metabolite. All univariate tests were two-tailed.

5.5 Accession numbers

Provided in Table 1.

Compliance with ethical standards

References

Adamski, J., & Suhre, K. (2013). Metabolomics platforms for genome wide association studies-linking the genome to the metabolome. *Current Opinion in Biotechnology, 24*(1), 39–47. doi:10.1016/j.copbio.2012.10.003.

Adaramola, T. F., Sonibare, M. a., Sartie, A., Lopez-Montes, A., Franco, J., & Albach, D. C. (2014). Integration of ploidy level, secondary metabolite profile and morphological traits analyses to define a breeding strategy for trifoliate yam (*Dioscorea dumetorum* (Kunth) Pax). *Plant Genetic Resources, 14*, 1–10. doi:10.1017/S1479262114000975.

Arnau, G., Nemorin, A., Maledon, E., & Abraham, K. (2009). Revision of ploidy status of *Dioscorea alata* L. (Dioscoreaceae) by cytogenetic and microsatellite segregation analysis. *Theoretical and Applied Genetics, 118*(7), 1239–1249. doi:10.1007/s00122-009-0977-6.

Asiedu, R., & Sartie, A. (2010). Crops that feed the World 1. Yams. *Food Security, 2*(4), 305–315. doi:10.1007/s12571-010-0085-0.

Bhattacharjee, R., Gedil, M., Sartie, A., Otoo, E., Dumet, D., Kikuno, H., et al. (2011). Dioscorea. In C. Kole (Ed.), *Wild crop relatives: Genomic and breeding resources-industrial crops* (pp. 71–96). Berlin: Springer.

Bino, R. J., Hall, R. D., Fiehn, O., Kopka, J., Saito, K., Draper, J., et al. (2004). Potential of metabolomics as a functional genomics tool. *Trends in Plant Science, 9*(9). doi:10.1016/j.tplants.2004.07.004.

Bousalem, M., Arnau, G., Hochu, I., Arnolin, R., Viader, V., Santoni, S., & David, J. (2006). Microsatellite segregation analysis and cytogenetic evidence for tetrasomic inheritance in the American yam *Dioscorea trifida* and a new basic chromosome number in the Dioscoreae. *Theoretical and Applied Genetics, 113*(3), 439–451. doi:10.1007/s00122-006-0309-z.

Bricas, N., & Attaie, H. (1997). La consommation alimentaire des ignames Synthèse des connaissances et enjeux pour la recherche. L'igname, plante séculaire et culture d'avenir, pp. 21–30.

Chaïr, H., Perrier, X., Agbangla, C., Marchand, J. L., Dainou, O., & Noyer, J. L. (2005). Use of cpSSRs for the characterisation of yam phylogeny in Benin. *Genome, 48*(4), 674–684. doi:10.1139/g05-018.

Chaïr, H., Scarcelli, N., Arnau, G., Pavis, C., Akakpo, R., Dansi, A., et al. (2016). CIRAD, IRD and INRA yam genomic initiatives: Unlocking genetic diversity and accelerating yam. Plant & Animal Genome Conference XXIV. San Diego.

Champagne, A., Bernillon, S., Moing, A., Rolin, D., Legendre, L., & Lebot, V. (2010). Carotenoid profiling of tropical root crop chemotypes from Vanuatu, South Pacific. *Journal of Food Composition and Analysis, 23*(8), 763–771. doi:10.1016/j.jfca.2010.03.021.

Champagne, A., Hilbert, G., Legendre, L., & Lebot, V. (2011). Diversity of anthocyanins and other phenolic compounds among tropical root crops from Vanuatu, South Pacific. *Journal of Food Composition and Analysis, 24*(3), 315–325. doi:10.1016/j.jfca.2010.12.004.

Dansi, A., Mignouna, H. D., Zoundjihékpon, J., Sangare, A., Asiedu, R., & Ahoussou, N. (2000). Using isozyme polymorphism to assess genetic variation within cultivated yams (Dioscorea cayenensis/Dioscorea rotundata complex) of the Republic of Benin. *Genetic Resources and Crop Evolution, 47*(4), 371–383. doi:10.1023/A:1008718411962.

Diop, A. (1998). Storage and processing of roots and tubers in the tropics. D. J. B. Calverley (Ed.). Food and Agriculture Organization of the United Nations. http://www.fao.org/docrep/X5415E/X5415E00.htm.

Fernie, A. R., Aharoni, A., Willmitzer, L., Stitt, M., Tohge, T., Kopka, J., et al. (2011). Recommendations for reporting metabolite data. *The Plant Cell, 23*(7), 2477–2482. doi:10.1105/tpc.111.086272.

Fernie, A. R., & Schauer, N. (2009). Metabolomics-assisted breeding: A viable option for crop improvement? *Trends in Genetics, 25*(1), 39–48. doi:10.1016/j.tig.2008.10.010.

Fernie, A. R., Tadmor, Y., & Zamir, D. (2006). Natural genetic variation for improving crop quality. *Current Opinion in Plant Biology, 9*(2), 196–202. doi:10.1016/j.pbi.2006.01.010.

Food and Agriculture Organization of the United Nations. (2015). Low-Income Food-Deficit Countries (LIFDC)—List for 2015. http://www.fao.org/countryprofiles/lifdc/en/.

Food and Agriculture Organization of the United Nations. (n.d.). FAOSTAT Database on Agriculture. Rome, Italy: Food and Agriculture Organization of the United Nations. http://faostat3.fao.org/. Accessed 28 June 2014.

Fukushima, A., Kusano, M., Redestig, H., Arita, M., & Saito, K. (2009). Integrated omics approaches in plant systems biology. *Current Opinion in Chemical Biology, 13*, 532–538. doi:10.1016/j.cbpa.2009.09.022.

Hamon, P., & Touré, B. (1990). The classification of the cultivated yams (Dioscorea cayenensis-rotundata complex) of West Africa. *Euphytica, 47*(3), 179–187. doi:10.1007/BF00024240.

Kenyon, L., Anandajayasekeram, P., & Ochieng, C. (2006). *A synthesis/lesson-learning study of the research carried out on root and tuber crops commissioned through the DFID RNRRS research programmes between 1995 and 2005*. London: Crop Protection Programme (CPP) of the UK Department For International Development (DFID)

Knoth, J. (1993). *Traditional storage of Yams and Cassava and its improvement*. Tamale: GTZ-Postharvest Project.

Kuehne, A., Zamboni, N., Sauer, U., Sévin, D. C., Kuehne, A., Zamboni, N., & Sauer, U. (2015). Biological insights through nontargeted metabolomics. *Current Opinion in Biotechnology, 34*, 1–8. doi:10.1016/j.copbio.2014.10.001.

Kumar, M. (2014). *Genetics and Breeding of Vegetables*. Oxford: Oxford Book Company.

Luo, J. (2015). Metabolite-based genome-wide association studies in plants. *Current Opinion in Plant Biology, 24*, 31–38. doi:10.1016/j.pbi.2015.01.006.

Merlo, L., Geigenberger, P., Hajirezaei, M., & Stitt, M. (1993). Changes of carbohydrates, metabolites and enzyme activities in potato tubers during development, and within a single tuber along astolon-apexgradient. *Journal of Plant Physiology, 142*(4), 392–402. doi:10.1016/S0176-1617(11)81243-5.

Mignouna, H. D., Abang, M. M., & Asiedu, R. (2003). Harnessing modern biotechnology for tropical tuber crop improvement: Yam (*Dioscorea* spp.) molecular breeding. *African Journal of Biotechnology, 2*(12), 478–485. doi:10.5897/AJB2003.000-1097.

Mignouna, H. D., Abang, M. M., Asiedu, R., & Geeta, R. (2009). True yams (*Dioscorea*): A biological and evolutionary link between eudicots and grasses. *Cold Spring Harbor Protocols, 4*(11), 1–8. doi:10.1101/pdb.emo136.

Nkamleu, B., Annang, D., & Bacco, N. M. (2009). Securing Livelihoods through Yams. In B. Nkamleu, D. Annang, & N. Bacco (Eds.), *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* (p. 329). Nigeria: International Institute of Tropical Agriculture (IITA).

Nweke, F. I., & Ezumah, H. C. (1992). Cassava in African farming and food systems: Implications for use in livestock feeds. In S. K. Hahn, L. Reynolds, & G. N. Egbunike (Eds.), *Cassava as livestock feed in Africa: Proceedings of the IITA/ILCA/University of Ibadan Workshop on the Potential Utilization of Cassava as Livestock Feeds in Africa. 14–18 November 1988, Ibadan, Nigeria* (p. 9). Ibadan, Nigeria: International Institute of Tropical Agricukture and International Livestock Centre for Africa. http://www.fao.org/wairdocs/ilri/x5458e/x5458e04.htm#cassava in food crop and livestock production systems.

Oluoha, U. (1988). Delimitation of physiological regions in yam tubers (Dioscorea sp.) and distribution pattern of saccharide degrading enzymes, cell sap pH and protein in these regions. *Biologica Plantarum, 30*(3), 210–218.

Price, E. J., Wilkin, P., Sarasan, V., & Fraser, P. D. (2016). Metabolite profiling of Dioscorea (yam) species reveals underutilised biodiversity and renewable sources for high-value compounds. *Scientific Reports, 6*(July), 29136. doi:10.1038/srep29136.

Ramser, J., Weising, K., Terauchi, R., Kahl, G., Lopez-Peralta, C., & Terhalle, W. (1997). Molecular marker based taxonomy and phylogeny of Guinea yam (*Dioscorea rotundata–D. cayenensis*). Genome/National Research Council Canada = Génome/Conseil national de recherches Canada, *40*(6), 903–15. http://www.ncbi.nlm.nih.gov/pubmed/18464876.

Riedelsheimer, C., Czedik-eysenberg, A., Grieder, C., Lisec, J., Technow, F., Sulpice, R., et al. (2012). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature Genetics, 44*(2), 217–220. doi:10.1038/ng.1033.

Sartie, A., Asiedu, R., & Franco, J. (2012). Genetic and phenotypic diversity in a germplasm working collection of cultivated tropical yams (Dioscorea spp.). *Genetic Resources and Crop Evolution, 59*(8), 1753–1765. doi:10.1007/s10722-012-9797-7.

Scarcelli, N., Daïnou, O., Agbangla, C., Tostain, S., & Pham, J. L. (2005). Segregation patterns of isozyme loci and microsatellite markers show the diploidy of African yam Dioscorea rotundata (2n = 40). *Theoretical and Applied Genetics, 111*(2), 226–232. doi:10.1007/s00122-005-2003-y.

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. a., et al. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics, 3*(3), 211–221. doi:10.1007/s11306-007-0082-2.

Tamiru, M., Natsume, S., Takagi, H., White, B., Yaegashi, H., Shimizu, M., et al. (2017). Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. *BMC Biology, 15*(1), 86. doi:10.1186/s12915-017-0419-x.

Tamiru, M., Yamanaka, S., Mitsuoka, C., Babil, P., Takagi, H., Lopez-Montes, A., et al. (2015). Development of genomic simple sequence repeat markers for yam. *Crop Science, 55*(5), 2191–2200. doi:10.2135/cropsci2014.10.0725.

Tohge, T., & Fernie, A. R. (2010). Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function. *Nature Protocols, 5*(6), 1210–1227. doi:10.1038/nprot.2010.82.

Ward, J., Rakszegi, M., Bedő, Z., Shewry, P. R., & Mackay, I. (2015). Differentially penalized regression to predict agronomic traits from metabolites and markers in wheat. *BMC Genetics, 16*(1), 19. doi:10.1186/s12863-015-0169-0.

Wellington, M. A., & Ahmad, M. H. (1993). Glutathione and beta-amylase, peroxidase and o-diphenolase activities during sprouting of minisett yarns (Dioscorea sp). *Journal of the Science of Food and Agriculture, 62*, 225–228.

Wheatley, A. O., Iyare, O. A., & Asemota, H. N. (2002). Effect of section of yam (Dioscorea cayenensis) tuber used in minisett on the biochemical properties of the resultant tuber. *Journal of the Science of Food and Agriculture, 82*(13), 1579–1583. doi:10.1002/jsfa.1225.

White, B. H., Bhattacharjee, R., Verweij, W., Lopez-Montes, A., Kumar, P. L., Kamoun, S., & Corpas, M. (2016). Progress on the genome characterisation of *D. alata*. Plant & Animal Genome Conference XXIV.