

Molecular breeding bases of trifoliate yam (*Dioscorea dumetorum* (Kunth) Pax) post- harvest hardening phenomenon of tubers

Von der Fakultät für Mathematik und Naturwissenschaften der Carl von Ossietzky Universität
Oldenburg zur Erlangung des Grades und Titels eines
Doktors der Naturwissenschaften (Dr. rer. nat.)

Angenommene Dissertation von

Herr Christian Siadjeu

Geboren am 10.09.1986 in Yaounde

Gutachter: Prof. Dr. Dirk C. Albach

Zweitgutachter: Prof. Dr. Sascha Laubinger

Tag der Disputation: 19.03.2021

Printed with the support of the German Academic Exchange Service

ERKLÄRUNG

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst und nur die angegebenen Hilfsmittel verwendet habe. Ausserdem versichere ich, dass ich die allgemeinen Prinzipien wissenschaftlicher Arbeit und Veröffentlichung, wie sie in den Leitlinien guter wissenschaftlicher Praxis der Carl von Ossietzky Universität Oldenburg festgelegt sind, befolgt habe.

Oldenburg, den _____

ACKNOWLEDGEMENTS

I would first like to thank my supervisor Prof. Dirk Albach for believing in me and giving me the opportunity to join his research group. Your consistent advice, motivation, encouragement, guidance, support, and welcoming atmosphere were critical and help me to successfully achieve my PhD thesis.

I would like to thank Prof. Bernd Weisshaar for a great support and supervision, and Dr. Boas Pucker for guidance and support during my time in Bielefeld. Our discussions and your meaningful advices were meaningful for my research. Besides, I would like to thank Prisca Viehöver and Dr. Daniela Holtgräwe for the welcoming atmosphere and advices during my stay in Bielefeld.

I wish to express my deepest gratitude to the German Academic Exchange Service (DAAD) for funding my doctoral stay in Germany, the Alexander von Humboldt foundation, and the German Society of Plant Sciences (DBG) for their practical and financial supports.

I wish to show my gratitude to Prof. Sascha Laubinger, for a great support and advice including insightful discussions and suggestions.

I am indebted to Eike Mayland-Quellhorst for great and continuous assistance. Your support and advice were important in the achievement of this PhD thesis. I also thank Silvia Kempen and Holger Ihler for their support.

I would like to highlight all my colleagues from research groups “Biodiversity and Evolution of Plants” and “Functional Ecology” and especially my friends Yonatan Aguilar Cruz and Jessica Ying Ling Tay for their support and the lovely, enjoyable, and relaxing atmosphere during my PhD journey.

Last, but not least, I would like to thank my family for the indefeatable support and all people from far or near who helped me to successfully accomplish this PhD thesis.

TABLE OF CONTENTS

ERKLÄRUNG	i
ACKNOWLEDGEMENTS	ii
TABLE OF CONTENTS	iii
LIST OF MAIN ABBREVIATIONS.....	v
ABSTRACT	1
ZUSAMMENFASSUNG.....	3
GENERAL INTRODUCTION	5
Yams: an overview.....	5
Trifoliate yam: <i>D. dumetorum</i>	7
Taxonomy, origin and distribution	7
Morphology and biology	8
Hardening phenomenon in plants	9
Hardening in <i>D. dumetorum</i>	10
Importance of <i>D. dumetorum</i> and yam production	10
Yams breeding	11
Molecular yam breeding.....	12
Yam ploidy	13
Yam: post-harvest storage practices	13
Research rationale	14
ARTICLES.....	16
List of articles	16
ARTICLE I.....	17
GENETIC DIVERSITY AND POPULATION STRUCTURE OF TRIFOLIATE YAM (<i>DIOSCOREA DUMETORUM</i> KUNTH) IN CAMEROON REVEALED BY GENOTYPING-BY-SEQUENCING (GBS).....	17
ARTICLE II.....	39
HIGH CONTIGUITY DE NOVO GENOME SEQUENCE ASSEMBLY OF TRIFOLIATE YAM (<i>DIOSCOREA DUMETORUM</i>) USING LONG READ SEQUENCING	39
ARTICLE III	66
TRANSCRIPTOME SEQUENCING REVEALS GENES CONTROLLING THE POST- HARVEST HARDENING OF TRIFOLIATE YAM <i>DIOSCOREA DUMETORUM</i>	66
OVERALL CONCLUSION AND FUTURE PERSPECTIVES	120
REFERENCES	122

CURRICULUM VITAE	134
AUTHOR CONTRIBUTIONS	139

LIST OF MAIN ABBREVIATIONS

AFLP:	Amplified Fragment Length Polymorphism
AH:	After harvest
CIRAD:	Agricultural Research Centre for International Development
DAH:	Day after Harvest
DEG:	Differential expressed gene
GBS:	Genotyping-By-Sequencing
GO:	Gene Ontology
MAE:	Month after emergence
MYB:	MYoloBlastosis
PCR:	Polymerase Chain Reaction
PHH:	Post-Harvest Hardening
QTLs:	Quantitative Trait Tocus
RAPD:	Random Amplified Fragment DNA
RFLPs:	Restriction Fragment Length Polymorphisms
RNA:	Ribonucleic Acid
SNP:	Single Nucleotide Polymorphism
SSRs:	Simple Sequence Repeats

ABSTRACT

Yams (*Dioscorea* spp.) are important and economical food for millions of people in the tropics and subtropics. *Dioscorea dumetorum* has the highest nutritional value compared to the other cultivated yam species with important medicinal properties. Despite these advantages, its cultivation is jeopardized by the post-harvest hardening (PHH) phenomenon characterized by the inability of tubers to soften during cooking. The PHH begins within 24 h after harvest and causes tubers to become inedible. Although several techniques have been deployed to solve this issue, they failed to overcome the PHH. The main objective of this thesis was to provide the basis of genetic breeding for avoidance of the PHH in *D. dumetorum*. For molecular breeding either the genetic basis of hardening needs to be known and/or variation within the species must be present. Firstly, we asked are there genetic variability between *D. dumetorum* accessions in Cameroon? Are *D. dumetorum* accessions in Cameroon grouped into populations? What is the diversity of *D. dumetorum* accessions with respect to the ploidy level and genome size? Using genotyping-by-sequencing, we showed that *D. dumetorum* accessions in Cameroon are diverse. Structure analysis revealed that *D. dumetorum* in Cameroon is structured into four populations, among which three populations were clearly determined. We also showed that there are diploid and triploid levels in *D. dumetorum*. Most of *D. dumetorum* accessions were diploid, but triploid level occurs predominantly in male accessions. The 1C genome was found to be small. Secondly, owing to heterozygous nature of yam, we hypothesized that releasing a genome resource of *D. dumetorum* is needed to improve for non-hardening accessions as well as for other desired characters. Based on long reads sequencing technologies, we released the first and highly contiguity *D. dumetorum* genome sequence. Overall, 35,269 protein-encoding gene models were predicted, and functional annotations were assigned. Finally, we assumed that the PHH is controlled by genes that are differentially expressed and up regulated after harvest. Harnessing the *D. dumetorum* genome sequencing and using transcriptome analysis based on RNA-Seq, we demonstrated that genes are differential and up regulated after harvest. Gene ontology (GO) analysis showed that the PHH in *D. dumetorum* is a protective reaction against stress due to sun light and water losses. Functional classification revealed that genes associated with light-harvesting proteins, cell wall strengthening, lignification, and transcription factors from (MYoloBlastosis) MYB family genes were involved in the PHH.

The significance of this study is that it defines the framework for the improvement of PHH avoidance and for other significant characters in yam as well as for yam conservation by

providing valuable information regarding the diversity of *D. dumetorum* in Cameroon, and the genetic basis of the PHH, by releasing a genome sequence.

ZUSAMMENFASSUNG

Yamswurzeln (*Dioscorea* spp.) sind wichtige und wirtschaftlich relevante Nahrungsmittel für Millionen von Menschen in den Tropen und subtropen. *Dioscorea dumetorum* hat den höchsten Nährwert im Vergleich zu den anderen kultivierten Yamswurzelnarten. Trotz dieser verbrauchsbedingten Vorteile wird der Anbau durch das Nachernte-Härtung Phänomen (PHH) beschränkt, bei dem sich die Knollen während des Kochens nicht weich werden können. Das PHH beginnt innerhalb von 24 Stunden nach der Ernte und führt dazu, dass Knollen ungenießbar werden. Für die molekulare Züchtung muss entweder die genetische Grundlage der Verhärtung bekannt sein und/oder Variationen innerhalb der Spezies vorhanden sein. Obwohl mehrere Techniken eingesetzt wurden, um dieses Problem zu lösen, konnten sie das PHH nicht überwinden. Hauptziel der Doktorarbeit war es, die Grundlagen der genetischen Technik zur Vermeidung der Nachernte-Härtung bei *D. dumetorum* zu schaffen. Erstens fragten wir, ob es genetische Variabilität zwischen *D. dumetorum* Sorte in Kamerun gibt? Sind *D. dumetorum* Sorte in Kamerun nach Populationen gruppiert? Wie groß ist die Vielfalt der *D. dumetorum* Sorte in Bezug auf Ploidie und Genomgröße? Mithilfe von Genotyping-By-Sequenzierung konnten wir zeigen, dass *D. dumetorum* Sorte in Kamerun vielfältig sind. Die Strukturanalyse ergab, dass *D. dumetorum* in Kamerun in vier Populationen unterteilt ist, von denen drei Populationen eindeutig bestimmt wurden. Wir haben auch gezeigt, dass es dioploid und triploide Werte bei *D. dumetorum* gibt. Die meisten *D. dumetorum* Sorte waren diploid, aber triploid tritt vor allem bei männlichen Sorte auf. Bei dem 1C-Genom wurde festgestellt, dass es klein ist. Zweitens, aufgrund der heterozygoten Natur der Yamswurzel, haben wir die Hypothese aufgestellt, dass die Freisetzung einer Genomressource von *D. dumetorum* notwendig ist, um nicht-härtende Akzessionen sowie andere erwünschte Merkmale zu verbessern. Basierend auf Long-Lese-Sequenzierungstechnologien haben wir die erste und sehr zusammenhängende Genomsequenz von *D. dumetorum* veröffentlicht. Insgesamt wurden 35 269 protein-kodierende Genmodelle vorhergesagt und funktionelle Annotationen zugewiesen. Schliesslich gingen wir davon aus, dass das PHH durch Gene kontrolliert wird, die nach der Ernte differenziert exprimiert und hochreguliert werden. Mithilfe der *D. dumetorum* Genomsequenzierung und der Transkriptomanalyse basierend auf RNA-Seq konnten wir zeigen, dass Gene nach der Ernte differenzial und hochreguliert sind. Geneontologische Analysen (GO) zeigten, dass das PHH in *D. dumetorum* eine Schutzreaktion Gene Stress durch Sonnenlicht und Wasserverluste ist. Die funktionelle Klassifizierung ergab, dass Gene, die mit

Lichtsammelproteinen, Zellwandstärkung, Lignifikation und Transkriptionsfaktoren aus der MYB-Familie assoziiert sind, an den PHH beteiligt waren.

Die Bedeutung der Studie liegt darin, dass sie durch die Bereitstellung wertvoller Informationen über die Diversität von *D. dumetorum* in Kamerun und die genetischen Grundlagen der PHH sowie durch die Freigabe einer Genomsequenz die Rahmenbedingungen für die Verbesserung der PHH-Vermeidung und anderer signifikanter Merkmale in der Yamswurzel sowie für die Yamswuchterhaltung definiert.

GENERAL INTRODUCTION

Yams: an overview

Yams (*Dioscorea* spp., *Dioscoreaceae*, *Monocotyledons*) constitute an important staple food for over hundreds of millions of people in the tropics. Yams are likely to originate from tropical areas of three separated continents (Africa, Southeast Asia and South America. The genus *Dioscorea* comprises over 600 species of which a few are cultivated and consumed: *Dioscorea alata*, *D. cayenensis*, *D. rotundata*, *D. bulbifera*, *D. esculenta*, *D. dumetorum*, *D. opposita*, *D. japonica*, *D. nummularia*, *D. pentaphylla*, *D. transversa*, *D. trifida* (Lebot, 2009; Siadjeu et al., 2015). This genus is further divided into several sections, of which the five most important are *Enantiophyllum*, *Lasiophyton*, *Combium*, *Opsophyton*, and *Macrogynodium* (Bai and Ekanayake, 1998). In Cameroon, there are twenty-three yam species, of which, sixteen are wild and seven are cultivated (Dumont et al., 1994). The most important are *D. dumetorum*, *D. alata*, *D. bulbifera*, *D. rotundata*, *D. cayenensis* (Westphal et al., 1985). In Cameroon, yams occur in all regions of Cameroon and they vary with respect to abundancy (Table 1).

Table 1. Abundance of yam species in the ecological regions of Cameroon (Dumont *et al.*, 1994). (+++) abundant, (++) fairly common, (+) occasional

Ecological region	<i>D. cayenensis</i>	<i>D. rotundata</i>	<i>D. dumetorum</i>	<i>D. alata</i>	<i>D. bulbifera</i>
West, North-West		+++	+++	+	++
Littoral, South-West	++	++	++	++	+
Centre, East, South	++	++	+	+	
Adamaoua		++		+	
North		++	+	+	
Far North		+			

Yams possess an aerial and underground structures. The aerial structure consists of a voluble stem on which leaves, and inflorescences are formed. The underground structure consists of a fibrous root system and the tubers which are storage organs. Optimum stem development is usually achieved though staking and varies in size depending on the species (Bai and Ekanayake, 1998).

Yams produce an annual underground storage organ called tubers, which wither when regrowth starts, and a new tuber can be formed concomitantly. Unlike the Irish potato, the yam tuber has no buds or eyes, no scale leaves and not terminal bud at the distal end of the tuber. Some species form perennial tubers, which become larger and more lignified as the plant grows. There are enormous variations in size, shape and number of tubers per plant within and between species. Many yam species (*D. alata*, *D. bulbifera*, *D. pentaphylla*) also produce edible aerial tubers called bulbils in the axils of the leaves (Lebot, 2009).

The root system is shallow. Yam plants possess three different types of adventitious roots (Bai and Ekanayake, 1998). Firstly, there are adventitious roots arising from the corm-like structure at the base of the stem from which originated tuber. These feed the plant by absorbing water and nutrients. Then, there are adventitious roots originating from the body of the yam tuber, which are generally thin and short. These roots are alive and may feed the plant for a short period of time during the tuber growth. Finally, adventitious roots emerging from the exposed lower nodes of the growing plants which do not reach the ground and can feed the plant.

The vine-like stem of yam can reach several meters in length depending on the variety and growing conditions. Twining orientation is a characteristic of each taxonomic section in the genus *Dioscorea* (Bai and Ekanayake, 1998). Species belonging to the section *Enantiophyllum* twine to the right or clockwise. Whereas, species belonging to *Opsophyton*, *Lasophyton*, *Combilium*, and *Macrogynodium* twine to the left or anti-clockwise. Depending on the species, the stem can be winged, spiny or not, hairy or not, and circular, rectangular or polygonal in cross-section (Coursey, 1967).

Yam leaves are generally simple, cordate and can be also lobed or forming of three leaflets. They are always carried on long, not sheathing petioles (Lebot, 2009). The leaves differ in size between and within species, and between cultivars and different parts of a single plant with an average area range of 50-200 cm² (Lebot, 2009). Leaves organization on the stem of phyllotaxis is spiral although it appears to be opposite or alternating depending on the species, and very often it can be alternating at the lower part of the stem and opposite at the upper part (Lebot, 2009).

Yam plants are dioecious, however monoecious plants with staminate and pistillate flowers or complete flowers on the same plant are sometimes observed (Bai and Ekanayake, 1998). Yam inflorescences are simple or compound. Commonly known as spikes (racemes), they are predominantly verticillate. The male flower consists of a calyx of three sepals and a

corolla of three petals, organized regularly and almost identical in size and appearance, with three or six stamens (Owueme 1978). The female flower possesses a trilocular ovary located below the corolla with three stigmas. Each lodge of the ovary contains two ovules (Govaerts et al. 2007). Pollination is suggested to be entomophilous. Fruits are trilocular, dry and dehiscent capsules of 1-3 cm long, and can produce six seeds (Lebot, 2009).

Trifoliate yam: *D. dumetorum*

Taxonomy, origin and distribution

Dioscorea dumetorum also known as “Bitter yam” or “trifoliate yam” or sweet yam (Cameroon) belongs to the subdivision *spermatophytina*, the superorder *Liliana*, the family *Dioscoreaceae*, and the section *Lasiophyton*. The *Lasiophyton* section contains 51 edible species, the most important of which are *D. pentaphylla*, *D. hispida* and *D. dumetorum* (Degras, 1993).

Trifoliate yam *D. dumetorum* is native to Africa, specifically in the region known as yam belt comprising Nigeria, Benin, Togo, Ghana, Cameroon and Côte d'Ivoire (Figure 1) (Bhattacharjee et al., 2011). However, cultivation of this species is limited in West and Central Africa, mainly in Cameroon and Nigeria (Adaramola et al., 2014). In Cameroon, it is mainly cultivated in the western part of the country.

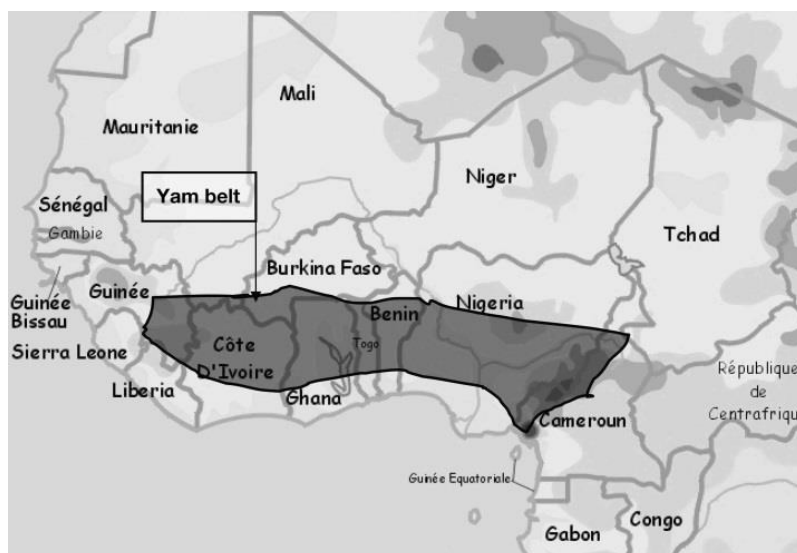


Figure 1. Yam belt in Africa. Source CIRAD

Morphology and biology

Trifoliate yam is a vigorous non-winged, herbaceous, climbing plant that has an underground structure and an aerial structure (Bai and Ekanayake, 1998). The underground structure consists on the one hand of the fibrous root system and on the other hand of the tuber or storage organ. The aerial structure is made up of a vine-like stem twining in clockwise direction on which the leaves and inflorescences are formed.

According to Degras (1993), *D. dumetorum* as other yam species have two root types. The first type consists of coronary roots 3-6 mm thick, and 1-3 m long. The number of roots varies from one to a few dozens. These roots grow at the base of the nodal region and extend across the upper layer of the ground. Its growth is essentially plagiotrophic. They rarely branched and emit very few radicals. The second type consists of the roots on the tuber which are rarely longer than a few centimetres and are 1-2 mm thick in diameter.

A trifoliate yam plant produces 3 to 12 large, rough and obconic tubers located in the upper soil layer (Figure 2a). Tuber skins are thin, and flesh colour varies from white to saffron yellow. It also can produce bulbils that vary in size and covered with rootlets (Figure 2b). Its stems are hairy, sometimes slightly spiny, covered with strong, reddish hairs and twine to the left (clockwise) (Figure 2c). The leaves are thick and trifoliate with three to five ascending and hairy nerves more or less dense on the lower part (Figure 2d). Their petiole is long, hairy and can be spiny (Hamon et al, 1995).

The inflorescences are cyme of two to five flowers and form highly branched groups (Degras, 1993). They are complex in male plant and form condensed and composite axis. The fruits are glossy, elongated, slightly hairy, and are reflexed upwards at maturity (Hamon et al, 1995).

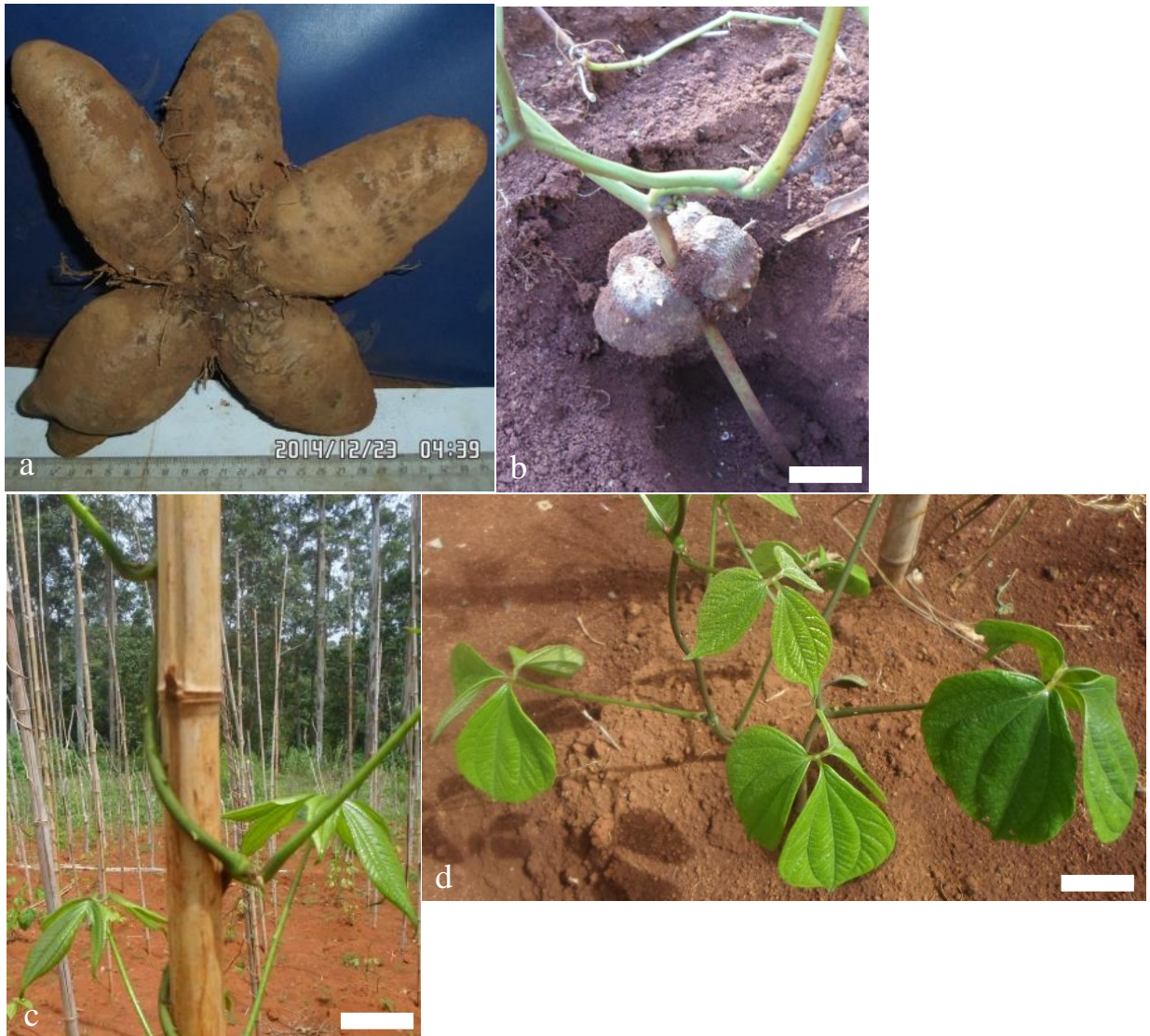


Figure 2. Morphology of *D. dumetorum* plant: a: tubers, b: bulbils, c: stem, d: leaves. Scale bar = 5cm

Hardening phenomenon in plants

Hardening phenomenon in plants also known as hard-to-cook defect is characterized by failure of sufficient softening during cooking. This phenomenon is well known in legumes like beans and chickpeas (Shehata, 1992; Reyes-Moreno et al., 2001). Several hypotheses have been proposed to explain this phenomenon. The first mechanism supposes a decrease in intracellular phytic acid, which is a calcium chelator (Mattson, 1946). This decrease leads to cations release, which during water soaking and cooking, are able to displace the monovalent (Na^+) cations bound to the pectin, thus causing pectin insolubilization and strengthening cell adhesion (Medoua, 2005). However, this mechanism does not fully explain hard-to cook defect in plants. Indeed, non-significant correlations between phytate content and hardening have been reported (Crean and Haisman, 1963; Henderson and Ankrah, 1985). Other mechanisms proposed

involve enzymatic or non-enzymatic lignification reactions in which phenolic compounds are precursors and would create bridges between pectins (Garcia et al., 1998).

Hardening in *D. dumetorum*

Diverse biochemical and structural changes have been associated with the post-harvest hardening phenomenon (PHH). It was characterized by thickened cell wall and middle lamella in the cells of hardening tubers (Figure 3) (Treche and Delpuch, 1982; Sefa-Dedeh and Afoakwa, 2002). This thickening was associated with an increase of cell wall components (lignin, cellulose, hemicellulose) (Sefa-Dedeh and Afoakwa, 2002). Medoua and Mbofung, (2006) divided the PHH in reversible and irreversible components connected with the decrease of phytate and the increase of total polyphenol respectively. The PHH was attributed to the defence reactions of tubers which refuse to loss water (Treche and Delpuch, 1982).

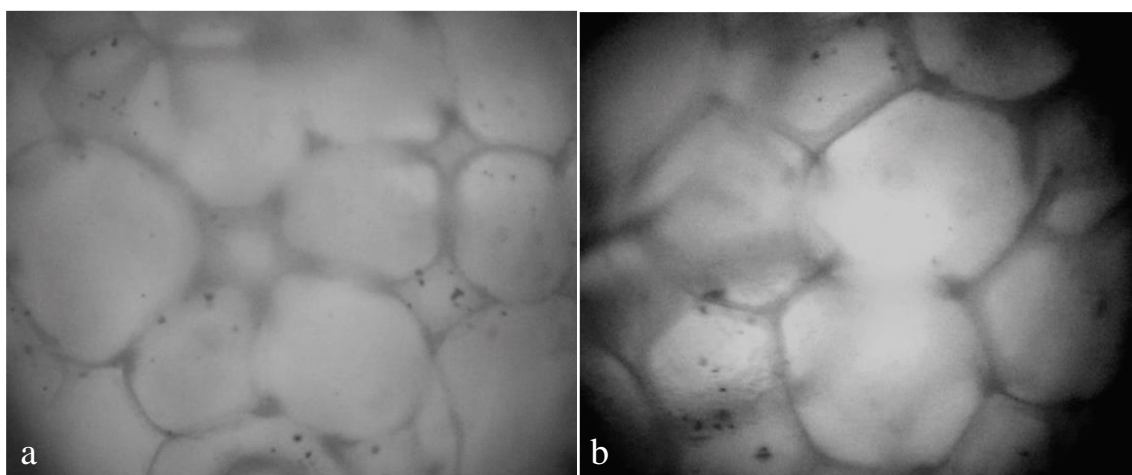


Figure 3. Parenchyma microstructure of fresh (a) and hardening (b) tubers x400 (Siadjeu et al., 2016)

Importance of *D. dumetorum* and yam production

Compared to other yam species, tubers of *D. dumetorum* possess the highest nutritional value. It differs from other yams with high protein contents (9.6%) and having proportionally balanced essential amino acids (Mbome Lape and Treche, 1994; Agbor-Egbe and Treche, 1995). Its starch has a similar structure to cereal starch, and easy digestibility and absorption of nitrogen and other nutrients (Mbome Lape and Treche, 1994). Tubers of yellow intense variety possess similar provitamin A and carotenoids contents to yellow corn maize lines selected for accumulated high provitamin A concentration (Ferede et al., 2010). In the pharmaceutical field, a bioactive compound (dioscoretine) has been identified (Iwu et al., 1990). This compound

possesses hypoglycaemic activities and used in antidiabetic medication in Nigeria (Iwu et al., 1990; Nimenibo-Uadia and Oriakhi, 2017).

Trifoliate yam is a high-yielding species, with yields of 10 and 40t/ha produced under traditional farming systems and in agronomic stations respectively (Treche, 1989; Agbor-Egbe and Treche, 1995). Trifoliate yam possesses high level of resistance to two major nematode parasites of yam (Kwoseh, 2000). Although yams are cultivated throughout the tropics, 97% of its production is confined in Africa (FAOSTAT, 2018). Nigeria alone produces 67 % of Africa production (Table 2). However, Cameroon has the highest yield per ha and therefore presents a great potential for yam production in Africa.

Table 2. Yam production around the world (FAOSTAT, 2018)

	Area harvested (ha)	Yield (hg/ha)	Production (T)
Cameroon	58716	114922	674776
Nigeria	5990184	79351	47532615
Africa	8500055	82924	70485850
World	8690716	83515	72580851

Yams breeding

The primary objectives for yam breeding include high and stable tuber yield, resistance to pests (e.g., nematodes) and diseases (e.g., viruses, anthracnose), tuber characteristics (shape, food quality) as required by consumers (Asiedu et al., 2003). However, new characters have been gradually added that differ in species, and depending on consumers' preferences, economical purposes, and environmental conditions. In *D. dumetorum*, the principal objectives for breeding focus on reducing the root system in the proximal area of the tuber and to eliminate the post-harvest hardening (Degras, 1993).

Traditionally, yam breeding employs sexual and asexual reproduction. The asexual selection is made using vegetative propagation on traditional material by farmers according to their objectives and based on the type of tubers produced (Hamon et al., 1997). The sexual hybridization in yams is constrained by several barriers making yam breeding challenging. They include scarcity of flowering, poor synchronization of the male and female flowering phases, dioecy, polyploidy, and high heterozygosity (Asiedu et al., 1998). However, advances have been made in studies of the floral biology of yams mainly of *D. alata* and *D. cayenensis*-

rotundata (Lebot, 2009). Thus, many parental genotypes (mainly of *D. rotundata*) have been identified over the years combining good agronomic qualities, reliable flowering and high fertility (Asiedu et al., 2003). In addition, real seeds have been used in yam breeding programs and natural hybridizations have been reported (Girma et al., 2014).

Molecular yam breeding

Application of molecular breeding or marker-assisted breeding in yam have been impeded by a lack of next generation omics resources and information (Bhattacharjee et al., 2011). Molecular markers are meaningful tools for application such as estimating genetic diversity and phylogenetic relationships, cultivar identification, mapping of major effect genes and quantitative trait locus (QTLs), assessing population structure, selection of interesting genotypes in breeding programs, and for validation of progenies obtained from genetic hybridizations (Tamiru et al., 2015). A clearer understanding of the available genetic diversity and the breeding potential of specific cultivars is relevant to select parents use in breeding programs (Mignouna et al., 2003a).

The early effort in yam molecular genetics was dedicated to the development of polymorphic DNA markers and estimation of their potential utilization in yams (Mignouna et al., 2003b). The first molecular markers were Restriction Fragment Length Polymorphisms (RFLPs), which were used to assess the origin and phylogeny of the Guinea yams (Terauchi et al., 1992). The success of this markers was limited and there was the need to extract a set of yam gene sequences as a priority for anchoring yam genetic maps (Mignouna et al., 2003b). In addition, RFLP was not cost efficient. In this regard, it has been abandoned in favour of PCR-based markers such as Random Amplified Fragment DNA (RAPD) and Amplified Fragment Length Polymorphism (AFLP).

Unlike RFLP, RAPD marker is faster, easier and do not require prior sequence information (Mignouna et al., 2003a). This marker was used to assess genetic diversity of yam in Benin and to identify molecular markers linked to yam mosaic virus resistance (Mignouna et al., 2002a; Zannou et al., 2009). However, the level of polymorphism detected in the mapping populations was low, and therefore RAPDs were not considered to be good targeting system in mapping (Mignouna et al., 2003b).

In yams, compared to RAPD, AFLP markers have a high level of polymorphism, and are more robust and reliable (Mignouna et al., 2003a). These markers have been applied for

genetic diversity in *D. dumetorum* (Sonibare et al., 2010), to estimate the diversity of *D. alata* cultivars and their relationship to *D. nummularia* Lam (Malapa et al., 2005). Furthermore, a genetic linkage map was constructed for *D. rotundata* and *D. alata* based AFLP markers (Mignouna et al., 2002b; (Petro et al., 2011).

Simple Sequence Repeats markers (SSRs) or microsatellites, owing to their high abundance and wide distribution across genome, high level of polymorphism was needful during yam genomics progress (Darkwa et al., 2020). Thus, SSRs markers were developed in many yam species: *D. tokoro*, *D. rotundata*, *D. japonica*, *D. alata*, *D. abyssinica*, *D. praehensilis*, *D. trifida*, *D. cayenensis* (Terauchi et al., 1994; Mignouna et al., 2003a; Mizuki et al., 2005; Tostain et al., 2006; Hochu et al., 2006; Siqueira et al., 2011; Silva et al., 2014).

Yam ploidy

Poor sexual reproduction of yams is often related to the polyploidy nature of yam (Egesi et al., 2002). It is important in the genus *Dioscorea* and most cultivated species are polyploidy with ploidy levels ranging from 2x to more than 10x (Nemorin, 2012). Several studies have been conducted to determine the level of ploidy of yams by chromosome counting or flow cytometry (Miège, 1954; Zoundjihekpon et al., 1990; Hamon et al., 1992; Gamiette et al., 1999; Dansi et al., 2001). However, due to the small size of yam chromosomes (0,2 - 2 μ m) (Ramachandran, 1968), which tend to agglutinate, these studies are not always consistent. *Dioscorea alata* has long been considered to have six levels of ploidy (2n=30, 40, 50, 60, 70 et 80), but recent studies demonstrated that it has rather three levels (2n=40, 60 et 80) (Abraham and Nair, 1990; Gamiette et al., 1999; Malapa et al., 2005; Arnau et al., 2009).

Yam : post-harvest storage practices

According to Opara (2003), three conditions are necessary for successful storage in yams: aeration, temperature reduction and regular inspection of tubers. Ventilation prevents condensation of moisture on the surface of the tuber and eliminates respiration heat. The low temperature reduces losses due to respiration, emergence and rotting. However, cold storage must be maintained at around 12-15 °C below which irreversible physiological deterioration occurs (chilling injury). Regular tuber inspection enables to remove sprouted and rotten tubers, and to check for the presence of rodents and insects.

In yams, there are traditional storage and improved practices. The traditional yam storage methods differ among yam producing countries and even within the country a range of

practices may be employed (Ravi et al., 1996). Yams can simply be stored by keeping them in the ground until food is needed or for sale. Although this practice exposes tubers to attacks by pests (e.g. termites and rodents), it is still used in West Africa (Ravi et al., 1996). Another simple storage practice, utilized in some regions of Asia and Africa, comprises purely of stacking the tubers into small heaps in places (e.g. rock caves or between the buttresses of large trees), where they are protected from the sun and flooding (Ravi et al., 1996). Moreover, Yam tubers can be stored under a conical protective roof made of maize (or millet stalks) or in trench silos or in sheds.

Other methods using modern techniques have been developed to address the shortcomings of traditional methods. Curing is a method of self-healing of tubers by exposure to high temperature and humidity immediately after harvest. High temperature and humidity promote cork cells formation in the cambium, which then move towards the wound areas which they close with multi-layers of wound periderm (Bautista, 1990). Treatment of yam tubers with pesticides (e.g. deltamethrin and thiabendazole) have been shown to control insect damage and reduce rots during storage (Girardin, 1996; Sauphanor et al., 1987). Manual de-sprouting significantly reduces the relative mass losses of tubers (Girardin, 1996).

Research rationale

The use of *D. dumetorum* tubers as food is hampered by a post-harvest hardening phenomenon that limits their storage time and results in significant post-harvest losses, with negative economic consequences for farmers (Medoua, 2005). It begins within 24 h after harvest and renders tubers unsuitable for human consumption (Sefa-Dedeh and Afoakwa, 2002). Tubers harden and become resistant to cooking and chewing, even after prolonged cooking (Treche and Delpeuch, 1982; Sealy et al., 1985). The usual storage conditions in West Africa (in an aerated shed, away from sun) lead to 100% losses after 4 months of storage (Treche and Delpeuch, 1979). Apart from this hardening phenomenon; no changes have been reported that could impede the use of hardening tuber (Medoua, 2005). Moreover, some studies demonstrated that even hardened, *D. dumetorum* tubers have a nutritional composition greater than other yams and comparable to cereals (Treche, 1989; Agbor-Egbe and Treche, 1995; Treche and Agbor-Egbe, 1996). In order to compensate for the hardening phenomenon, and not loss the value of tubers as a source of food and energy, the hardening tubers have been processed into flour (Treche et al., 1984) and then in porridge (Mbome Lape, 1991) for better storage. However, all processes developed to date for processing hardening tubers into flour result in

products with poor sensory characteristics (Medoua et al., 2005). In this regards, hardened tubers were soaked in different salt solutions (Medoua et al., 2007) and in water for fermentation (Medoua et al., 2008), but the hardening phenomenon has not been overcome. Thus, genetic breeding of *D. dumetorum* appears to be the most appropriate alternative method to address the hardening phenomenon. The key to success for any breeding program is linked to the availability and assessment of genetic diversity of the desired traits. Indeed, Treche (1989) noticed that 3% of *D. dumetorum* tubers are not subject to hardening after one month of storage, and then Sefa-Dedeh and Afoakwa (2002) reported that hardening in *D. dumetorum* varies between yellow and white cultivars. Finally, Siadjeu et al. (2016) identified a cultivar that does not harden and could provide useful information for breeding program of *D. dumetorum* against the hardening. This suggests that the hardening in *D. dumetorum* is likely controlled by molecular factors or genes. Therefore, the main objective of this study was to provide the basis of genetic breeding for the avoidance of the post-harvest hardening phenomenon in *D. dumetorum*. To achieve that goal, the study was divided in three chapters corresponding to three articles.

The first chapter (ARTICLES I) assessed the genetic diversity and population structure of *D. dumetorum* in Cameroon using genotyping-by-sequencing (GBS). Forty-four *D. dumetorum* samples were collected in the main yam growing regions. Many of these accessions were utilized for the PHH estimation. We assumed that *D. dumetorum* accessions are diverse and structured into population in Cameroon. In addition, ploidy level and genome size were determined, which are relevant for breeding programs.

The second chapter (ARTICLES II) aimed to de novo assembled and annotated *D. dumetorum* genome using long read sequencing. Given the highly heterozygous nature of *Dioscorea*, the availability of a *D. dumetorum* genome sequence of *D. dumetorum* is important for breeding program against the PHH and other characters of interest. Thus, a yam accession, which does not express the hardening was selected for the assembly.

The third chapter (ARTICLES III) identified candidate genes involved in the post-harvest hardening phenomenon in *D. dumetorum*. Four *D. dumetorum* accessions among which one is not subject to the hardening were collected at different period of times during the tuber development and after harvest to capture genes which are up regulated and involved in the hardening.

ARTICLES

List of articles

Siadjeu, C.*, Mayland-Quellhorst, E., and Albach, D. C. (2018). Genetic diversity and population structure of trifoliate yam (*Dioscorea dumetorum* Kunth) in Cameroon revealed by genotyping-by-sequencing (GBS). BMC Plant Biology, 18:359.

Siadjeu, C., Pucker, B., Viehöver, P., Albach, D.C., Weisshaar, B*. (2020). High contiguity *de novo* genome sequence assembly of trifoliate yam (*Dioscorea dumetorum*) using long read sequencing. Genes, 11: 274.

Siadjeu, C.*, Mayland-Quellhorst, E., Laubinger, S., and Albach, D. C. (2021). Transcriptome sequencing reveals genes controlling the post-harvest hardening of trifoliate yam *Dioscorea dumetorum* (submitted to Plants, manuscript ID: plants-1130765).

ARTICLE I:

GENETIC DIVERSITY AND POPULATION STRUCTURE OF
TRIFOLIATE YAM (*DIOSCOREA DUMETORUM* KUNTH) IN
CAMEROON REVEALED BY GENOTYPING-BY-SEQUENCING
(GBS)

Siadjeu et al. *BMC Plant Biology* (2018) 18:359
<https://doi.org/10.1186/s12870-018-1593-x>

BMC Plant Biology

RESEARCH ARTICLE

Open Access



Genetic diversity and population structure of trifoliate yam (*Dioscorea dumetorum* Kunth) in Cameroon revealed by genotyping-by-sequencing (GBS)

Christian Siadjeu^{*} , Eike Mayland-Quellhorst and Dirk C. Albach

Abstract

Background: Yams (*Dioscorea* spp.) are economically important food for millions of people in the humid and sub-humid tropics. *Dioscorea dumetorum* (Kunth) is the most nutritious among the eight-yam species, commonly grown and consumed in West and Central Africa. Despite these qualities, the storage ability of *D. dumetorum* is restricted by severe postharvest hardening of the tubers that can be addressed through concerted breeding efforts. The first step of any breeding program is bound to the study of genetic diversity. In this study, we used the Genotyping-By-Sequencing of Single Nucleotide Polymorphism (GBS-SNP) to investigate the genetic diversity and population structure of 44 accessions of *D. dumetorum* in Cameroon. Ploidy was inferred using flow cytometry and gbs2ploidy.

Results: We obtained on average 6371 loci having at least information for 75% accessions. Based on 6457 unlinked SNPs, our results demonstrate that *D. dumetorum* is structured into four populations. We clearly identified, a western/north-western, a western, and south-western populations, suggesting that altitude and farmers-consumers preference are the decisive factors for differential adaptation and separation of these populations. Bayesian and neighbor-joining clustering detected the highest genetic variability in *D. dumetorum* accessions from the south-western region. This variation is likely due to larger breeding efforts in the region as shown by gene flow between *D. dumetorum* accessions from the south-western region inferred by maximum likelihood. Ploidy analysis revealed diploid and triploid levels in *D. dumetorum* accessions with mostly diploid accessions (77%). Male and female accessions were mostly triploid (75%) and diploid (69%), respectively. The 1C genome size values of *D. dumetorum* accessions were on average 0.333 ± 0.009 pg and 0.519 ± 0.004 pg for diploids and triploids, respectively.

Conclusions: Germplasm characterization, population structure and ploidy are an essential basic information in a breeding program as well as for conservation of intraspecific diversity. Thus, results obtained in this study provide valuable information for the improvement and conservation of *D. dumetorum*. Moreover, GBS appears as an efficient powerful tool to detect intraspecific variation.

Keywords: Cameroon, *D. dumetorum*, Yam, Genetic diversity, Population structure, GBS, Ploidy

* Correspondence: christian.siadjeu@uni-oldenburg.de
Institute for Biology and Environmental Sciences, Biodiversity and Evolution of Plants, Carl-von-Ossietzky University Oldenburg, Carl-von-Ossietzky Str. 9-11, 26111 Oldenburg, Germany



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Background

Yams (*Dioscorea* spp.) constitute a staple food for over 300 million people in the humid and sub-humid tropics. About 600 species are described and are widely distributed throughout the tropics [1]. *Dioscorea dumetorum* has the highest nutrient value among eight other yam species commonly grown and consumed in West and Central Africa [2]. The species originated in tropical Africa and occurs in both wild and cultivated forms. Its cultivation is restricted to West and Central Africa [3], and widespread in western Cameroon. Tubers of *D. dumetorum* are protein-rich (9.6%) with fairly balanced essential amino acids and its starch is easily digestible [4–6]. Agronomically, *D. dumetorum* is high-yielding, with yield of 40 tons/hectare recorded in agricultural stations [7]. *Dioscorea dumetorum* is also recognized for its pharmaceutical properties. A novel bioactive compound dioscoretine has been identified in *D. dumetorum* [8], which can be used advantageously as a hypoglycemic agent in anti-diabetic medications [9].

Despite these qualities, the storage ability of *D. dumetorum* is restricted by severe postharvest hardening of the tubers, which begins within 24 h after harvest and renders them unsuitable for human consumption [2]. According to Treche and Delpeuch [10], the usual storage conditions in West Africa (under airy warehouse, shelter from sunlight) induce 100% losses after 4 months of storage. It is manifested by the loss of culinary quality due to a combination of factors resulting from normal but inadvertently deleterious reactions leading to textural changes [11]. Therefore, *D. dumetorum* is consumed exclusively during its limited harvest period and only freshly harvested tubers are cooked and sold to consumers. To add more value to *D. dumetorum* as an important source of food and energy, hardened tubers are transformed into instant flour [12]. However, flour obtained directly from hardened tubers has poor organoleptic qualities, such as coarseness in the mouth [4]. Thus, other techniques have been used, such as salt soaking treatment [13] and fermentation [14], but the hardening phenomenon has not been surmounted. Consequently, molecular breeding of *D. dumetorum* appears as the appropriate method to overcome this phenomenon.

The study of genetic diversity is an important, early step in plant breeding. Highlighting this variability is part of the characterization of germplasm under investigation. In our recent study on the phenotypic diversity of *D. dumetorum* we found relatively high diversity of morphological characters suggesting high underlying genetic diversity [15]. Indeed, the expression of morphological characters are subject to agro-climatic variations and thus provides limited genetic information. Therefore, molecular markers that are not subject to

environmental variations are necessary for estimation of genetic diversity. The development of molecular markers over the last 30 years has enabled the study of diversity and evolution as well as germplasm characterization [16]. Among these markers, Single Nucleotide Polymorphisms (SNPs) have emerged as the most widely used genotyping markers due to their abundance in the genome allowing not only germplasm characterization but also quantification of relative proportions of ancestry derived from various founder genotypes of currently grown cultivars [16]. Moreover, the development of traditional markers like SSRs, RFLPs and AFLPs was a costly, iterative process that involved either time-consuming cloning and enzyme testing or primer design steps that could not easily be parallelized [17].

Genotyping-By-Sequencing (GBS) has emerged as a new approach to mitigate these constraints. The method has been demonstrated to be suitable for population studies, germplasm characterization, genetic improvement, trait mapping in a variety of diverse organisms and thereby, SNP discovery and genotyping of multiple individuals are performed cost-effectively and efficiently [18]. GBS is performed by an initial digest of sample DNA with restriction enzymes reducing genome complexity followed by a round of PCR to generate a high-throughput sequencing library [19]. Reducing genome complexity with restriction enzymes is quick, extremely specific and highly reproducible [19]. Unlike other similar approaches using restriction enzymes, GBS is technically simple [20]. Besides, bioinformatic pipelines are publicly available [21] and GBS can be easily applied to non-model species with limited genomic information [20]. This method has been successfully used on Cassava (*Manihot esculenta* Crantz) [22], guinea yam [23] and water yam [24], which demonstrated the power of GBS-SNP genotyping as a suitable technology for high-throughput genotyping in yam.

Genetics of yams is least understood and remains largely neglected among the major staple food crops due to several biological constraints and research neglect [25]. Some progress has been made in germplasm characterization and the development of molecular markers for genome analysis. Various dominant molecular markers (AFLP, RAPD) have been used on yam with little success (e.g., [9]). Additionally, genomic microsatellite markers have been developed for yam species [24–32]. However, no markers have been developed for *D. dumetorum* and its genetics is the least known in spite of its qualities among the cultivated yam. Until now, no information is available using SNP genotyping to assess population structure, genetic diversity and the relationship among *D. dumetorum* cultivars.

A possible additional factor that influences population structure and genetic diversity is polyploidy. Polyploidy

has several advantages for plant breeding such as the increment in plant organs ("gigas" effect), buffering of deleterious mutations, increased heterozygosity, and heterosis (hybrid vigor) [33]. In yam, ploidy increase is correlated with growth vigor, higher and more stable tuber yield and increased tolerance to abiotic and biotic stress [33, 34]. Recent studies using flow cytometry revealed diploid and triploid levels in *D. dumetorum* with predominance of the diploid cytotype [35, 36]. Therefore, the objective of this study is to understand the genetic diversity and the population structure of *D. dumetorum* using the genotyping-by-sequencing (GBS) in relation to ploidy information.

Methods

Plants materials

Overall, 44 accessions of *D. dumetorum* were used in this study (Table 1). All these accessions were collected from different localities in the major yam growing regions (western, south-western, and north-western) of Cameroon, with an additional three accessions of *D. dumetorum* from Nigeria complementing the dataset (Fig. 1). Western and north-western regions belong to agro-ecological zone (AEZ) 3 and the south-western region to AEZ 4 of Cameroon [38]. Most of these accessions were previously used for morphological characterization [15] and hardness assessment [39]. Here, we selected some characters related to tubers (Fig. 2). The yam tubers of these accessions were planted in April 2015 at the "Ferme Ecole de Bokué" in the western region of Cameroon (latitude 05°20.040' N and longitude 010°22.572' E). Silica-dried young leaves were transported to Oldenburg (Germany) for molecular analyses. Genomic DNA was extracted using an innuPREP Plant DNA kit (Analytik Jena, Jena, Germany).

Preparation of libraries for next-generation sequencing

A total of 200 ng of genomic DNA for each sample were digested with 1 Unit MspI (New England Biolabs, NEB) in 1x NEB4 buffer in 30 µl volume for 1 h at 37 °C. The restriction enzyme was heat inactivated by incubation at 80 °C for 20 min. Afterwards, 15 µl of digested DNA were transferred to a new 96well PCR plate, mixed and stored on ice first with 3 µl of one of the 192 L2 ligation adaptors (Ovation Rapid DR Multiplex System, Nugen Technologies, Leek, The Netherlands) and then with 12 µl master mix (combination of 4.6 µl D1 water/ 6 µl L1 ligation buffer mix/ 1.5 µl L3 ligation enzyme mix). Ligation reactions were incubated at 25 °C for 15 min followed by inactivation of the enzyme at 65 °C for 10 min. Then, 20 µl of the kits 'final repair' master mix were added to each tube and the reaction was incubated at 72 °C for 3 min. For library purification,

reactions were diluted with 50 µl TE 10/50 (10 mM Tris/HCl, 50 mM EDTA, pH:8.0) and mixed with 80 µl magnetic beads, incubated for 10 min at room temperature and placed for 5 min on a magnet to collect the beads. The supernatant was discarded and the beads were washed two times with 200 µl 80% Ethanol. Beads were air-dried for 10 min and libraries were eluted in 20 µl Tris Buffer (5 mM Tris/HCl, pH 9). Each of the 45 libraries (including one technical repeat) were amplified with 10 µl of the purified restriction product in 20 µl PCR reactions using 4 µl MyTaq (Bioline) 5x buffer, 0.2 µl polymerase and 1 µl (10 pmol/µl) of standard Illumina TrueSeq amplification primers. Cycle number was limited to ten cycles. Then, 5 µl from each of the 48 amplified libraries were pooled. PCR primers and small amplicons were removed by magnetic bead purification using 0.6 volume of beads. The PCR polymerase was removed by an additional purification on Qiagen MinElute Columns. The pooled library was eluted in a final volume of 20 µl Tris buffer (5 mM Tris/HCl, pH 9). The final library pool was sent to LGC genomics (Berlin, Germany) and sequenced on an Illumina NextSeq with 1.5 million 150 bp paired-end reads for each sample. Additional steps at LGC for the sequencing preparation were normalization, reamplification and size selection. Normalization was conducted using Trimmer Kit (Evrogen). For this 1 µg pooled GBS library in 12 µl was mixed with 4 µl 4x hybridization buffer, denatured for 3 min at 98 °C and incubated for 5 h at 68 °C to allow re-association of DNA fragments. 20 µl of 2 x DSN master buffer was added and the samples were incubated for 10 min at 68 °C. One unit of DSN enzyme (1 U/µl) was added and the reaction was incubated for another 30 min. The reaction was terminated by the addition of 20 µl DSN Stop Solution, purified on a Qiagen MinElute Column and eluted in 10 µl Tris Buffer (5 mM Tris/HCl pH 9). The normalized library pools were re-amplified in 100 µl PCR reactions using MyTaq (Bioline). Primer i5-Adaptors were used to include i5-indices into the libraries, allowing parallel sequencing of multiple libraries on the Illumina NextSeq 500 sequencer. Cycle number was limited to 14 cycles. The nGBS libraries were size selected using Blue Pippin, followed by a second size selection on a LMP-Agarose gel, removing fragments smaller than 300 bp and those larger than 400 bp. Libraries were sequenced on an Illumina NextSeq 500 using Illumina V2 Chemistry.

GBS data analysis

GBS data were analyzed using the custom software pipeline iPyrad (Versions: 0.7.19 and 0.7.28) developed by Eaton and Ree [21] for population genetic and phylogenetic studies. It includes seven steps to demultiplex and quality filtering, cluster loci with consensus

Table 1 Characteristics of *D. dumetorum* accessions used in this study. * Area belongs to agro-ecological zone 3, ** to agro-ecological zone 4

Code	Accession name	Geographic origin	Roots on the tuber surface	Tuber flesh color	Tuber hardening	Altitude (m)
A07W	Baigon 1	West*	Few	Yellow	Yes	1120
A08W	Fonkouankem 1	West*	Many	Yellow	Yes	1167
A09I	Ibo sweet 3	Nigeria	Few	Yellow	No	56
A10W	Bangangté 1	West*	Few	Yellow	Yes	1350
A11S	Muyuka 1	Southwest**	Few	White	Yes	554
A12S	Muyuka 5	Southwest**	Few	White	Yes	554
B07N	Bambui 1	Northwest*	Few	Yellow	Yes	1262
B08W	Bangou 1	West*	Many	Yellow	Yes	1350
B09W	Bamendjou 2	West*	Many	Yellow	Yes	1647
B10W	Bana 1	West*	Few	Yellow	Yes	1167
B11S	Bekora 1	Southwest**	Few	White	Yes	60
B12S	Country yam 2	Southwest**	Few	White	Yes	62
C07S	Banga bakundu sweet 1	Southwest**	Few	White	Yes	62
C08I	Ibo sweet 2	Nigeria	Many	White	Yes	56
C09W	Bamendjou 1	West*	Few	Yellow	Yes	1647
C10N	Nkwen 1	Northwest*	Few	Yellow	Yes	1251
C11S	Bekora 2	Southwest**	Few	White	Yes	60
C12S	Muyuka 3a	Southwest**	Few	White	Yes	62
D07S	Country yam 1	Southwest**	Few	White	Yes	62
D08W	Bamokonbou 1	West*	Many	Yellow	Yes	1414
D09S	Muyuka 3	Southwest**	Few	White	Yes	62
D10N	Bambalang 1	Northwest*	Few	Yellow	Yes	1185
D11S	Mabondji sweet yellow 1	Southwest**	Few	Yellow	Yes	80
D12S	Penda-boko sweet 1	Southwest**	Few	White	Yes	554
E07S	Lysoka sweet 1	Southwest**	Many	Yellow	Yes	60
E08I	Ibo sweet 1	Nigeria	Few	White	Yes	56
E09W	Dschang 1	West*	Few	Yellow	Yes	1337
E10S	Mbongue sweet 1	Southwest**	Few	White	Yes	62
E12W	Fong-tongo 1	West*	Many	Yellow	Yes	1460
F07 N	Mankon 1	Northwest*	Few	Yellow	Yes	1253
F08 W	Bayangam 1	West*	Few	Yellow	Yes	1560
F09S	Mabondji sweet white 1	Southwest**	Few	White	Yes	80
F10 N	Batibo 1	Northwest*	Few	Yellow	Yes	1127
F11S	Buea sweet white yam 1	Southwest**	Few	White	Yes	438
G07 N	Guzang 1	Northwest*	Many	Yellow	Yes	1233
G08 W	Bayangam 2	West*	Many	Yellow	Yes	1560
G09 W	Bangang 1	West*	Few	Yellow	Yes	1776
G10 N	Fundong 1	Northwest*	Few	Yellow	Yes	1554
H06N	Babungo 1	Northwest*	Few	Yellow	Yes	1182
H07S	Buea sweet 1	Southwest**	Many	Yellow	Yes	554
H08W	Bangang 2	West*	Many	Yellow	Yes	1776
H09W	Babadjou 1	West*	Few	Yellow	Yes	1395
H10N	Bafut 1	Northwest*	Few	Yellow	Yes	1123
H11S	Alou 1	Southwest**	Few	Yellow	Yes	1606

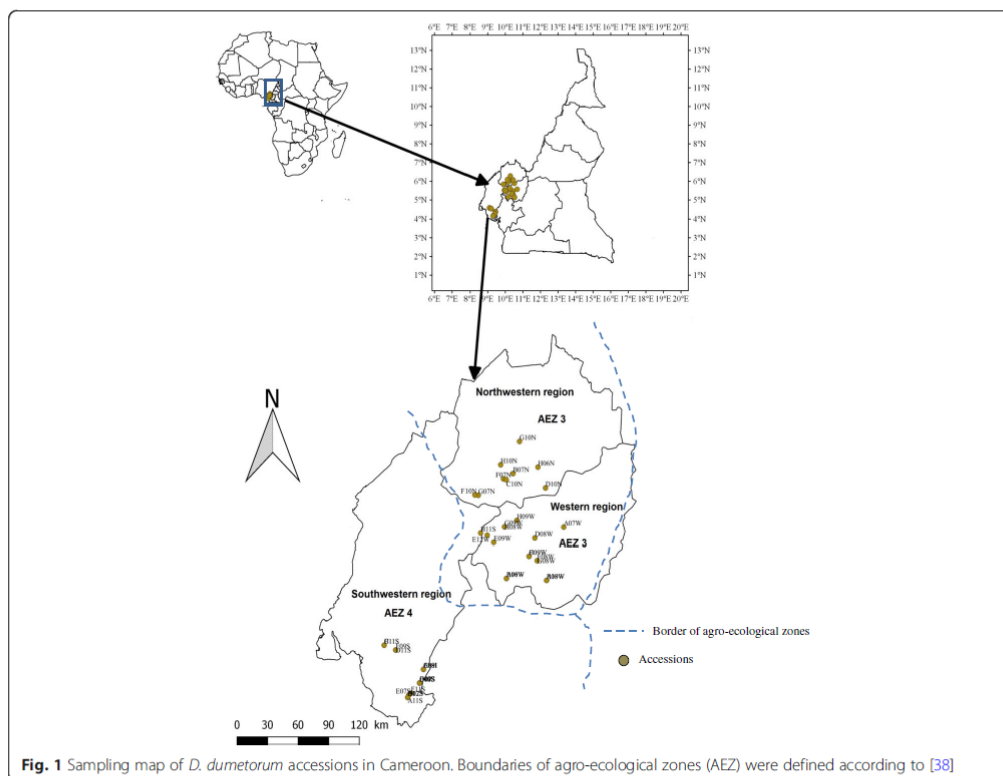


Fig. 1 Sampling map of *D. dumetorum* accessions in Cameroon. Boundaries of agro-ecological zones (AEZ) were defined according to [38]

alignments and SNP calling with SNP filtering to the final SNP matrix, which can be transferred into various output formats. We have conducted demultiplexing and QC separately to retrieve fastq sequences as input for iPyrad. The restriction sites and barcodes were trimmed for each sequence, bases with a quality score less than PHRED 20 were changed into N and sequences having more than 5% of N were discarded. Step 3 of iPyrad used in our de-novo SNP analysis VSEARCH [40] for dereplication and merging of paired reads and for clustering reads per sample into putative loci with 85% sequence similarity. Alignments of consensus sequences of the putative loci were built with MUSCLE [41]. After estimation of sequencing errors (Π) and heterozygosity (ϵ), consensus alleles were estimated with these estimated parameters and the number of alleles was recorded. Resulting consensus alleles were again clustered with VSEARCH and aligned with MUSCLE. Base SNPs were called when loci were observed in at least 75% of the samples, had not more than 20 SNPs and eight indels and heterozygous sites in 50% of the samples, but all samples were treated as diploid, thus allowing two haplotypes per polymorphic site.

Phylogenetic inference

An unrooted tree was generated using the neighbor-net method in SplitsTree (Version 4.14.6) [42] based on the concatenated GBS data. To control whether or not the introduction of triploid accessions affected our phylogenetic analysis, we constructed dendrograms with and without triploid accessions.

Historical relationship between accessions (TreeMix)

Historical relationships between *D. dumetorum* accessions including possible gene flow events was assessed through the maximum likelihood method implemented in TreeMix (version 1.13) [43]. TreeMix reconstructs the possible migrations between populations based on allele frequency of genomic data. It uses a method that allows for both population splits and gene flow. We defined the population parameter as 0, because we worked at the individual level. Of 25,541 SNPs loci investigated, 157 SNPs were filtered to get a gap free matrix and used to determine the relationships between the accessions. The tree was built with the confidence of 1000 bootstrap replicates and visualized with toytree (version 0.1.4) and toyplot (version 0.16.0).

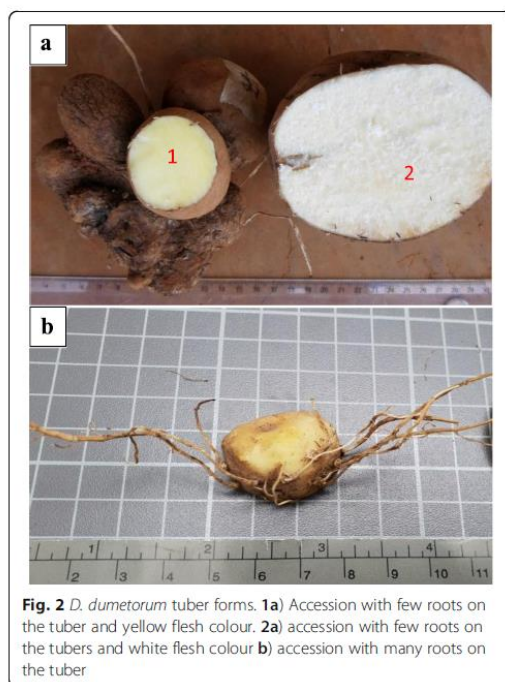


Fig. 2 *D. dumetorum* tuber forms. **1a)** Accession with few roots on the tuber and yellow flesh colour. **2a)** accession with few roots on the tubers and white flesh colour **b)** accession with many roots on the tuber

Population structure analysis

Analysis of population structure was performed using the software STRUCTURE [44] and Maverick [45]. Structure uses a Bayesian model-based clustering method with a heuristic approach for estimation whereas Maverick uses a computation technique called Thermodynamic Integration (TI). However, the mixture modeling framework is identical in both programs [45]. The analysis was carried out in STRUCTURE using the admixture model across 10 replicates (K of 2 to 5) of sampled unlinked SNPs (one randomly chosen SNP per ipyrad-cluster). A burn-in period of 10,000 iterations and 100,000 Markov Chain Monte Carlo (MCMC) replicates were run. The true number of clusters (K) was detected using Evanno's method [46] implemented in STRUCTURE HARVESTER [47]. The MCMC implementation of Maverick differs slightly, although the core model assumed is identical to that used in Structure [45]. Thus, the admixture model across five replicates (K of 2 to 5) was run with a burn-in period of 2000 iterations and 10,000 MCMC. The best value of K was detected in 25 TI rungs each for a range of K (2 to 5) with default settings.

Ploidy/genome size estimation

For each accession, about 1 cm² of young leaf was co-chopped with a standard using a razor blade in a

Petri dish containing 1.1 mL ice-cold Otto I buffer (0.1 M citric acid monohydrate and 5% Triton X-100). We used *Solanum lycopersicum* L. 'Stupicke' (1C = 0.98 pg; [48]) as the internal standard. The chopped material and buffer were then filtered through a Cell-Tric 30-μm filter into a plastic tube, and 50 μL RNase were added. After incubation in a water bath for 30 min at 37 °C, 450 μL of the solution were transferred to another tube, to which 2 mL Otto II (propidium iodide + Na₂HPO₄) were added. This solution was placed at 4 °C for 1 h. The samples were analyzed using a CyFlow flow cytometer (Partec GmbH, Münster, Germany). For each accession, three replicates comprising 5000 counts were measured. We measured the genome size of 17 out of 44 *D. dumetorum* accessions due to the loss of certain accessions, in which the sex has been identified. Ploidy level of the remaining accessions (27) was assessed using the R package gbs2ploidy [49]. This method infers cytotypes based on the allelic ratios of heterozygous SNPs identified during variant calling within each individual. Data was prepared by acquiring a *.vcf output file for all specimens from ipyrad using VCFConverter2.py (<https://github.com/dandewaters/VCF-File-Converter>) as in [50]. Cytotypes were estimated in two ways: 1) without reference to accessions of known ploidy and 2) with reference to accessions of known ploidy and 2) with reference of 17 accessions for which ploidy is known) from flow cytometry as set of triploids and diploids to the 27 remaining accessions.

Results

GBS data analysis summary

We generated an average of 2.2 million raw reads per *D. dumetorum* accessions by Illumina sequencing (Table 2). After filtering we obtained on average 1.3×10^4 reads clustered at 85%, with an average depth per accession of 53. The maximum likelihood average estimate of heterozygosity ($\epsilon = 1.1 \times 10^{-2}$) was greater than the sequence error rate ($\Pi = 6 \times 10^{-3}$). Consensus sequences were called for each cluster, yielding on average 32,532 reads per accession. We recorded on average 6371 loci recovered in at least 75% of accessions. Accession D09S had a markedly higher proportion of missing data.

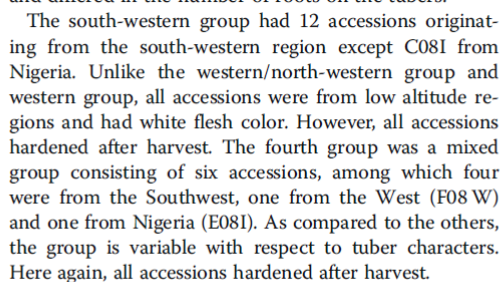
Phylogenetic inference

The unrooted neighbor-net clustered the 44 accessions of *D. dumetorum* into four groups: a western/north-western group, a western group, a southwestern group and a mixed group (Fig. 3). However, two accessions (E10S and H06N) were not clustered in these groups. Triploid accessions did not affect the topology of the network (Additional file 1: Figure S1).

The western/north-western group had 16 accessions, 88% were from the western and north-western regions (50% were from West and 50% from North-west).

Table 2 Summary statistics of filtering and clustering GBS data from *D. dumetorum*

Accessions code	Raw reads	clusters total at 85%	Heterozygosity (ε)	Sequence error rate (Π)	Consensus reads	Loci with 75% of samples	Mean depth
A07W	97,6092	85,513	0.01	0.006	23,641	6411	22.03
A08W	4,208,284	188,712	0.013	0.006	44,753	6926	93.15
A09I	6,850,560	230,744	0.016	0.006	50,290	6710	153.87
A10W	1,403,568	105,485	0.01	0.006	29,316	6806	32.24
A11S	2,355,324	153,699	0.012	0.007	36,254	6797	50.5
A12S	2,224,370	194,901	0.012	0.007	39,005	6599	38.43
B07N	6,315,512	270,497	0.015	0.006	50,836	6776	132.11
B08W	3,224,481	152,315	0.011	0.006	38,488	6938	75.73
B09W	4,821,443	227,297	0.013	0.006	49,469	6873	107.54
B10W	2,246,679	123,018	0.011	0.006	32,260	6978	54.09
B11S	1,675,059	114,487	0.011	0.007	31,882	6589	35.9
B12S	2,823,683	140,521	0.012	0.006	37,261	6843	64.72
C07S	10,482,599	328,036	0.016	0.006	65,451	6621	219.94
C08I	1,051,320	100,553	0.01	0.007	25,978	6340	23.27
C09W	3,945,117	180,341	0.013	0.006	42,150	6935	91.99
C10N	1,060,648	94,092	0.01	0.007	25,254	6592	24.77
C11S	1,937,689	125,788	0.012	0.006	31,880	6815	43.75
C12S	479,412	60,756	0.009	0.007	16,586	4535	10.32
D07S	3,137,681	162,346	0.012	0.006	38,116	6809	72.13
D08W	1,127,041	85,710	0.01	0.006	25,476	6491	26.71
D09S	24,555	8580	0.012	0.006	883	208	0.2
D10N	814,774	77,508	0.01	0.006	22,223	6298	19.69
D11S	2,860,962	166,597	0.012	0.006	41,004	6818	60.81
D12S	476,212	58,172	0.011	0.007	16,603	4777	11.11
E07S	310,575	49,544	0.01	0.007	11,810	3427	5.73
E08I	2,329,405	156,765	0.011	0.006	36,697	6783	52.83
E09W	1,182,619	119,258	0.01	0.006	27,799	6354	23.22
E10S	1361,644	110,821	0.012	0.007	28,732	6665	28.6
E12W	701,654	79,193	0.01	0.006	21,530	5765	15.23
F07 N	2,413,712	138,285	0.011	0.006	33,944	6969	56.7
F08 W	1,735,184	114,265	0.011	0.006	30,603	6665	40.36
F09S	1,643,257	110,155	0.011	0.007	30,166	6733	37.7
F10 N	739,588	68,577	0.01	0.006	20,579	5869	16.84
F11S	1,542,460	98,117	0.01	0.006	28,884	6651	36.72
G07 N	3,845,340	171,087	0.013	0.006	40,668	6924	88.43
G08 W	1,637,683	115,971	0.011	0.006	31,144	6917	38.44
G09 W	675,273	68,986	0.01	0.006	20,229	5793	15.56
G10 N	2,277,786	128,426	0.011	0.006	32,260	6955	54.21
H06N	4,854,717	207,191	0.014	0.006	49,949	6919	107.4
H07S	3,731,409	192,342	0.012	0.006	46,340	6790	76.41
H08W	3,057,515	146,899	0.012	0.006	35,916	6936	71.65
H09W	1,579,956	107,146	0.01	0.006	28,594	6951	39.04
H10N	1,920,829	109,146	0.011	0.006	31,018	6942	46.4
H11S	1,530,490	103,703	0.01	0.006	29,508	6831	34.52
Mean	2,399,867	132,535	0.011	0.006	32,532	6371	53.43



In population P1, accessions were from the western and north-west region except accessions A09I (Nigeria) and H11S (south-western region). Here, three accessions were assigned 100% to P1, twelve as admixture between P1 and P4 and one accession A09I as admixture of P1xP2xP3xP4.

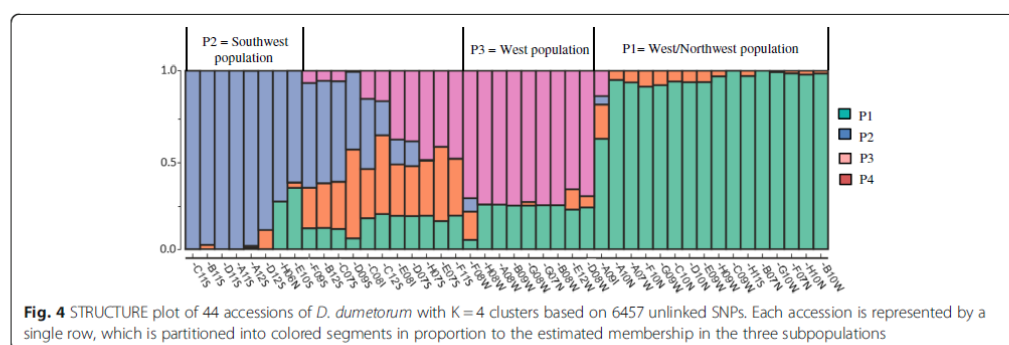


Fig. 4 STRUCTURE plot of 44 accessions of *D. dumetorum* with $K = 4$ clusters based on 6457 unlinked SNPs. Each accession is represented by a single row, which is partitioned into colored segments in proportion to the estimated membership in the three subpopulations

In contrast, all the accessions of population P2 were from the south-western region except H06N (North-west). Four accessions were assigned 100% to P2, two accessions as admixture P2xP4, while two each as admixture P1xP2xP4 and P1xP2. Regarding P3, almost all accessions (8) were from the western region except G07N from the north-western region. Conversely, no accession was assigned 100% to P3. Five were assigned as admixture P1xP3, three classified as P1xP2xP3 and one as P1xP2xP3xP4. Moreover, the population structure did not change with the increased values of $K = 5$ (Additional file 3: Figure S3). Comparing results from the STRUCTURE analysis with the neighbor-net, we found generally similar results. Thus, P1 corresponds to the west/north-western population, P2 to the south-western population, and P3 to the western population. No accessions belonging to P4 were identified.

Ploidy/genome size estimation

We found that 13 (76%) accessions of *D. dumetorum* were diploid (2x) and four (24%) were triploid (3x) (Table 3). The 1C genome size values for *D. dumetorum* measured here were on average 0.333 ± 0.009 pg and 0.519 ± 0.004 pg for diploids and triploids, respectively. The standard coefficient of variation (CV) of each measurement was < 5% for all runs (Additional file 4: Table S1). Comparing the data with sex, we found that diploid accessions, were 69% female and 31% were male. For triploid accessions, 75% were male and 25%

female. With respect to geographic origin, all triploid accessions come from the southwestern region.

Using the R package gbs2ploidy on accessions with known ploidy (17), we assessed the sensitivity of gbs2ploidy on our GBS data. The probability of concurrence between flow cytometry and gbs2ploidy was 35%, with 8 of 17 accessions assigned to the opposite cytotype and three (A09I, B09W, E08I) being inconclusive. The probability of correct diploid and triploid assignments was 38 and 25%, respectively. Training gbs2ploidy with reference accessions from flow cytometry on the remaining accessions (27), we found that 21 (78%) accessions were diploids and 6 (22%) triploids with the mean assignment probability of 74 and 73%, respectively. Regarding diploid accessions, seven, five and nine accessions originated from western, north-western and south-western regions, respectively. For triploids, three were from north-western, two from western and one from south-western regions. In summary, 34 accessions of *D. dumetorum* (77%) were diploid (2x) and 10 (23%) were triploid (3x). Triploid accessions originated mainly (70%) from the south-western region.

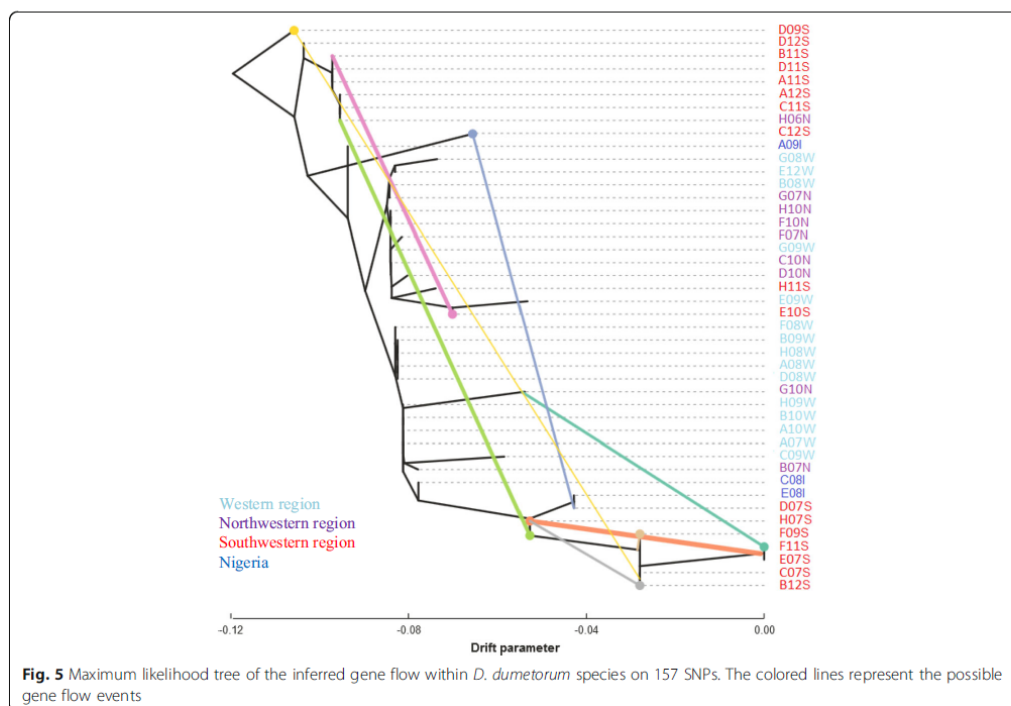
Historical relationship between accessions

We used TreeMix in order to determine splits and gene flow between *D. dumetorum* accessions. We constructed the tree allowing between no migration and ten migration events. We found eight gene flow events between *D. dumetorum* accessions (Fig. 5). Despite the

Table 3 Ploidy level/genome size, sex and origin of *D. dumetorum* accessions

	No. of accessions	Sex		Geographic origin				1C-values (pg)
		Male	Female	West	Northwest	Southwest	Nigeria	
Diploid (76%)	13	4 (31%)	9 (69%)	6 (46%)	1 (8%)	4 (31%)	2 (15%)	0.333 ± 0.009
Triploid (24%)	4	3 (75%)	1 (25%)	0 (0)	0 (0)	4 (100%)	0 (0)	0.519 ± 0.004
Diploid* (78%)	21	—	—	7 (33%)	5 (24%)	9 (43%)	0 (0)	—
Triploid* (22%)	6	—	—	2 (33%)	1 (17%)	3 (50%)	0 (0)	—

*Ploidy level estimated using the R package gbs2ploidy



likelihood for the tree with nine migration events being highest (but almost similar to eight migrations), we chose the one with eight events because the ninth migration was redundant (Additional file 5). The migration events involved eleven accessions from the south-western region and two (G10 N and H06N) from the north-western region. We did not find a migration event involving A08, which does not harden after harvest, as well accessions originating from the western region and Nigeria. C12S (2x, few root and white flesh) was possibly the result of gene flow between D07S (2x, female, few root and white flesh) and D09S (3x, male, few root and white flesh) or their ancestors; C07S (3x, male, few roots and white flesh) and E07S (2x, male, many root and yellow flesh) were possibly the result of a introgression between H06N (2x, few roots and yellow flesh) and H07S (2x, male, many roots and yellow flesh). Furthermore, allowing migrations altered the topology of the tree compared to the tree with no migrations events (Additional file 6: Figure S4).

Discussion

Genotyping-by-sequencing is an innovative, robust and cost-effective approach allowing multiplexing individuals in one library to generate thousands to millions of

SNPs across a wide range of species [51]. In our study, we identified on average 30,698 reads per accession. After filtering to avoid the effect of missing data, 5054 loci were kept for the analyses. In total, 26,325 SNPs were investigated. These numbers are similar to a previous study using the same pipeline in another non-model species [21].

The unrooted neighbor-net tree (Fig. 3) clustered *D. dumetorum* accessions into four groups: A western/north-western group, a western group, a south-western group and a mixed group. The West and North-west belong to agro-ecological zone III (Western Highlands) and the Southwest belongs to agro-ecological zone IV. This result disagrees with previous results using morphological characters [15], in which there was no clear separation of *D. dumetorum* accessions according to agro-ecological zone. However, morphological markers are subject to environmental conditions and thus provide limited genetic information. Moreover, Sonibare et al. [9] using AFLP on *D. dumetorum* accessions from three countries didn't find a clear separation according to the area of collection. However, SNP markers are the most abundant in a genome and suitable for analysis on a wide range of genomic scales [52, 53]. In combination with high-throughput sequencing, thousands to millions

of SNP generated using GBS [54] allow assessing more efficiently the genetic diversity compared to AFLP. This was already suggested by Saski et al. [24], who stated that GBS is a powerful tool for high-throughput genotyping in yam.

Our assignment test results based on STRUCTURE also separate *D. dumetorum* accessions into four populations in which three were clearly identified, the western/north-western population, the western population and the south-western population. On the contrary, MavricK revealed that *D. dumetorum* was structured into two populations in accordance with known agro-ecological zones (Additional file 2, Figure S2). However, the number of loci investigated was large (more than hundreds of loci). In this situation, the heuristic approximation implemented in STRUCTURE appears to be better [45]. Furthermore, tuber flesh color of all accessions in the western and north-western region was yellow whereas the majority of accessions from south-western have white tuber flesh. Our results suggest that altitude and farmers-consumers preference played a role as a barrier between *D. dumetorum* populations. Indeed, AEZ 3 corresponds to western highlands covering the western and northwestern region. It is characterized by high altitude (1000–2740 m), low temperature (annual mean 19 °C) and annual rainfall of 1500 to 2000 mm. In contrast, AEZ 4 comprises mainly humid forest covering the south-western and littoral regions. It is characterized by low altitude (< 700 m except a few mountains), with an annual rainfall of 2500 to 4000 mm and a mean temperature of 26 °C [38]. All three regions of Cameroon belong to the yam belt, where the species occurs in both wild and cultivated forms. Nevertheless, its center of origin remains unknown so far excluding an explanation for the origin of the separation of populations in Cameroon. Tuber quality is an important criterion of adoption of yam varieties by farmers and consumers [55]. Thus, the difference regarding tuber flesh color in the western/north-western and south-western regions could be explained by different preferences of consumers in these regions, which also depends on yam food form. In the western and north-western regions, yam tubers are almost exclusively consumed as boiled tubers contrary to the South-west where tubers are consumed either boiled or pounded. Consumers in Cameroon probably prefer yellow tubers in the boiled and white tubers in the pounded form. Indeed, Egesi et al. [56] demonstrated that flesh color determines a general preference for boiled or pounded yam in *D. alata*. Assuming white flesh as the ancestral character state based on its predominant occurrence in other yams species, we assumed that the yellow flesh color has evolved several times (probably four times) because it is present in our four groups inferred, although a single origin with subsequent intraspecific hybridization or losses cannot

be excluded. Yams with many roots have likely evolved once, in the western region probably due to environmental conditions of the highland with occasional scarcity of water. The root system has an important physiological function in nutrient and water absorption. It is well known, that several root system traits are considered to be important in maintaining plant productivity under drought stress [57]. The occurrence of mutations related to yellow flesh color and many roots on the tuber in the south-western region (mixed group) was probably caused by artificial crossing of genetically diverse accessions in the region.

The importance of gene flow within and between our four main groups in *D. dumetorum* can be seen in the high proportion of admixture. This observation could be explained by the efforts, which have been made in the past in Cameroon, especially in the South-west to improve *D. dumetorum* [7]. Indeed, genetic diversity can be increased by breeding activities [58]. Especially noteworthy is the fourth groups with all individuals assigned to it being admixed, suggesting the absence of genetically unambiguous accessions belonging to this group from Cameroon (Fig. 3). It is possibly that genetically unambiguous individuals of this group was not sampled within Cameroon or went extinct, but our preferred hypothesis is that such plants originated from Nigeria. This finding further corroborates a close relationship between *D. dumetorum* accessions from Nigeria and Cameroon. The South-west and North-west regions of Cameroon share a common border with Nigeria. Exchanges of *D. dumetorum* accessions between farmers on both side of the border are well-known, providing gene flow and interbreeding. Indeed, Sonibare et al. [9] reported that introduction of the *D. dumetorum* germplasm to Central African countries has been affected by the activities of farmers from Nigeria.

TreeMix results obtained in our study also indicate that there was more gene flow between accessions from the south-western region than in the western/north-western region. These findings support the admixture result of STRUCTURE discussed above and allow refinement of our understanding of genotypes crossed in the past. However, regarding the sample with non-postharvest hardening, we did not detect any gene flow. This suggests that the sample was not used, yet, in any breeding in Cameroon and that non-postharvest hardening appears yet to be restricted in *D. dumetorum* to Nigeria. Thus, a broader study on the genetic diversity involving samples across the distribution range of the species is needed to track the origin of this character and the ancestry of this sample.

Ploidy is another factor possibly relevant for population structure and breeding causing hybrid vigor (heterosis) and buffering of deleterious mutations. Our

analysis revealed that 77% of *D. dumetorum* accessions were diploid and 23% were triploid. This result is broadly consistent with previous findings, in which 83% were diploid and 17% triploid [36] and 60% diploid and 40% triploid [37]. However, the probability of concurrence between flow cytometry and gbs2ploidy was low (35%). In fact, a limitation of the gbs2ploidy method is low coverage, especially if possible ploidy levels for the species are unknown [49]. The authors reported that this problem could be resolved by including validated reference samples with known cytotypes in the analysis as done in our study.

The association between sex and ploidy showed a predominance of triploids for male accessions and diploids for female accessions. These findings partially contradict those of Adaramola et al. [37] in which a predominance of diploid for male accessions has been reported. However, Adaramola et al. [37] outlined that a more systematic sampling method that ensures an equal number of *D. dumetorum* accession may change their results, which was the case in our study. The 1C genome size values of *D. dumetorum* accessions ranged on average from 0.33 to 0.52 pg for diploids and triploids, respectively. This supports the results of Obidiegwu et al. [36], who found that the 1C genome of five diploid and one triploid *D. dumetorum* clones ranged from 0.35 to 0.53 pg, respectively. Thus, *D. dumetorum* appears to have a very small size genome (1C-value ≤ 1.4 pg) following the categories of [59]. TreeMix results suggested admixture of some accessions between different ploidy levels. Triploid accessions may either be the result of a possible admixture between triploid (3x) or diploid (2x) male with diploid (2x) females, although sex of accessions H06N and C12S has not been determined. Similar results were reported in *D. alata* [60]. This suggests that the occurrence of triploid accessions in *D. dumetorum* is most likely due to the involvement of unreduced (2n) gametes in the pollen rather than the egg cell. This was confirmed by artificial crossing of triploid (3x) male and diploid (2x) female we performed in the field (Siadjeu unpublished data, Additional file 7: Figure S5). Finally, the predominant occurrence of triploid accessions in the southwestern region coincides with the more intensive breeding program in the region and may be explained by it since it is known that hybridization between genetically diverse accessions of a species may increase the number of unreduced gametes [61].

Conclusions

In this study, we reported population structure, genetic diversity and ploidy/genome size of *D. dumetorum* in Cameroon using GBS. We demonstrated that *D. dumetorum* is structured into populations. There is a high genetic variability of *D. dumetorum* accessions in

Cameroon. We revealed intraspecific hybridization and provided useful information regarding ploidy/genome size of *D. dumetorum*. All this information is relevant for conservation and a breeding program of *D. dumetorum*. However, we did not infer a firm relationship of the sample with postharvest hardening, the character most important for future breeding efforts, suggesting a broad study with respect to this character in West and Central Africa will be needed to elucidate its origin. Finally, GBS appears as an efficient powerful tool for phylogeographic studies in yams.

Additional files

Additional file 1: Figure S1. Phylogenetic relationships within *D. dumetorum* based on multilocus concatenated SNP sequences alignment from GBS data of 34 diploid accessions. (PDF 259 kb)

Additional file 2: Figure S2. Estimates of the model evidence for $K = 2, 5$ using Ti estimator **a)** log-evidence and **b)** the evidence and Structure estimator Delta K ΔK **c)** (PDF 95 kb)

Additional file 3: Figure S3. STRUCTURE plot of 44 accessions of *D. dumetorum* with $K = 2, 3, 5$ clusters based on 6457 unlinked SNPs. (PDF 133 kb)

Additional file 4: Table S1. Coefficient of variation of ploidy measurements using flow cytometric and ploidy level per accessions estimated by gbs2ploidy. * Ploidy level assessed by gbs2ploidy (PDF 47 kb)

Additional file 5: Edges, weight of migrations (8 and 9) and likelihood for migrations 0.9. (TREEOUT 1 kb)

Additional file 6: Figure S4. Maximum likelihood tree of the inferred gene flow within *D. dumetorum* species with no gene flow events. (PDF 125 kb)

Additional file 7: Figure S5. Flowering and fructification of *D. dumetorum*. **a)** male flower, **b)** female flower. Bar scale = 3 cm. **c)** fruits, **d)** seeds. Bar scale = 2 cm (PDF 288 kb)

Abbreviations

AEZ: Agro-ecological zone; AFLP: Amplified Fragment Length Polymorphism; CV: Coefficient of Variation; EDTA: Ethylenediaminetetra-acetic acid; GBS: Genotyping-By-Sequencing; MCMC: Markov Chain Monte Carlo; P: Population; PCR: Polymerase Chain Reaction; RAPD: Random Amplified Polymorphic DNA; RFLP: Restriction Fragment Length Polymorphism; SNP: Single Nucleotide Polymorphism; SSR: Single Sequence Repeats; Ti: Thermodynamic Integration

Acknowledgements

We thank Silvia Kempen, for her support in the analysis of genome size. We also thank Dr. Robert Verity for his support to deal with bad runs of Ti after analysis on Maverick and two anonymous reviewers who helped to improve this paper.

Funding

This study was financially supported by the German Academic Exchange Service (DAAD), and Appropriate Development for Africa foundation (ADAF) for the yam accessions collection and the field work in Cameroon. Moreover, the funders had no role in the design of the study and collection, analysis and interpretation of data, decision to publish, or the preparation of the manuscript.

Availability of data and materials

The GBS data was deposited in European Nucleotide Archive (ENA) (Access No. PRJEB27526; <https://www.ebi.ac.uk/ena/data/search?query=PRJEB27526>). All data generated during the current study are included in this published article and its supplementary information as Additional files 1, 2, 3, 4, 5, 6, and 7.

Authors' contributions

CS contributed to design of the experiments, collection of yam accessions, data analysis and interpretation and manuscript writing. EMQ: DNA extraction and sequencing, DCA: data interpretation and manuscript writing. All authors have read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 9 July 2018 Accepted: 6 December 2018

Published online: 18 December 2018

References

- Viruel J, Forest F, Paun O, Chase MW, Devvey D, Sousa Couto R, et al. A nuclear Xdh analysis of yams (*Dioscorea*, *Dioscoreaceae*) congruent with plastid trees reveals a new Neotropical lineage. *Bot J Linn Soc*. 2018; in press.
- Sefa-Dedeh S, Afoakwa EO. Biochemical and textural changes in trifoliate yam *Dioscorea dumetorum* tubers after harvest. *Food Chem*. 2002;79:27–40.
- Degras L. The yam: a tropical root crop. Yam a trop. Root crop. London: Macmillan Press Ltd; 1993.
- Medoua GN, Mbome IL, Agbor-Egbe T, Mbofung CMF. Physicochemical changes occurring during postharvest hardening of trifoliate yam (*Dioscorea dumetorum*) tubers. *Food Chem*. 2005;90:597–601.
- Mbome Lape I, Treche S. Nutritional quality of yam (*Dioscorea dumetorum* and *D. rotundata*) flours for growing rats. *J Sci Food Agric*. 1994;66:447–55.
- Afoakwa EO, Sefa-Dedeh S. Chemical composition and quality changes occurring in *Dioscorea dumetorum* pax tubers after harvest. *Food Chem*. 2001;75:85–91.
- Treche S. Potentialités nutritionnelles des ignames (*Dioscorea* spp.) cultivées au Cameroun. Editions de l'ORSTOM CE et T, editor. Paris; 1989.
- Iwu MM, Okunji C, Akah P, Corley D, Tempesta S. Hypoglycaemic Activity of Dioscoretine from Tubers of *Dioscorea dumetorum* in Normal and Alloxan Diabetic Rabbits. *Planta Med*. 1990;56:264–7.
- Sonibare MA, Asiedu R, Albach DC. Genetic diversity of *Dioscorea dumetorum* (Kunth) Pax using amplified fragment length polymorphisms (AFLP) and cpDNA. *Biochem Syst Ecol*. 2010;38:320–34.
- Treche S, Delpeuch F. Evidence for the development of membrane thickening in the parenchyma of tubers of *Dioscorea dumetorum* during storage. *C R Acad Sc Paris t*. 1979;288:67–70.
- Medoua GN. Potentiels nutritionnel et technologique des tubercules durcis de l'igname *Dioscorea dumetorum* (Kunth). Cameroon: Doctoral thesis, Ngaoundere University; 2005.
- Treche S, Mbome Lape I, Agbor-Egbe T. Variations de la valeur nutritionnelle au cours de la préparation des produits séchés à partir d'ignames cultivées. *Rev Sci Tech (Sci Santé)*. 1984;7–22.
- Medoua GN, Mbome IL, Egbe TA, Mbofung CMF. Salts soaking treatment for improving the textural and functional properties of trifoliate yam (*Dioscorea dumetorum*) hardened tubers. *J Food Sci*. 2007;72E:464–E469.
- Medoua GN, Mbome IL, Agbor-Egbe T, Mbofung CMF. Influence of fermentation on some quality characteristics of trifoliate yam (*Dioscorea dumetorum*) hardened tubers. *Food Chem*. 2008;107:1180–6.
- Siadjeu C, MahbouSomoToukam G, Bell JM, Nkwate S. Genetic diversity of sweet yam *Dioscorea dumetorum* (Kunth) Pax revealed by morphological traits in two agro-ecological zones of Cameroon. *African J Biotechnol*. 2015; 14:781–93.
- Deschamps S, Llaca V, May GD. Genotyping-by-sequencing in plants. *Biology (Basel)*. 2012;1:460–83.
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet*. 2011;12:499–510.
- De Donato M, Peters SO, Mitchell SE, Hussain T, Imumorin IG. Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *PLoS One*. 2013;8:1–8.
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 2011;6:1–10.
- Wickland DP, Battu G, Hudson KA, Diers BW, Hudson ME. A comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow, GB-easy. *BMC Bioinformatics*. 2017;18:586.
- Aetion DAR, Ree RH. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis: Orobanchaceae*). *PLoS One*. 2013;6:689–706.
- Rabbi NY, Kulakow PA, Manu-Aduening JA, Dankyi AA, Asibuo JY, Parkes EY, et al. Tracking crop varieties using genotyping-by-sequencing markers: a case study using cassava (*Manihot esculenta* Crantz). *BMC Genet*. 2015;16:1–11.
- Girma G, Hyma KE, Asiedu R, Mitchell SE, Gedil M, Spillane C. Next generation sequencing based genotyping, cytometry and phenotyping for understanding diversity and evolution of Guinea yams. *Theor Appl Genet*. 2014;127.
- Saski CA, Bhattacharjee R, Scheffler BE, Asiedu R. Genomic resources for water yam (*Dioscorea alata* L.): analyses of estsequences, de novo sequencing and GBS libraries. *PLoS One*. 2015;10:1–14.
- Mignouna HD, Abang MM, Asiedu R. Harnessing modern biotechnology for tropical tuber crop improvement: yam (*Dioscorea* spp.) molecular breeding. *African J Biotechnol*. 2003;2:475–85.
- Terauchi R, Konuma A. Microsatellite polymorphism in *Dioscorea tokoro*, a wild yam species - genome. *Genome*. 1994;37:794–801.
- Mizuki I, Tani N, Ishida K, Tsumura Y. Development and characterization of microsatellite markers in a clonal plant, *Dioscorea japonica* Thunb. *Mol Ecol Notes*. 2005;5:721–3.
- Tostain S, Scarcelli N, Brotter P, Marchand JL, Pham JL, Noyer JL. Development of DNA microsatellite markers in tropical yam (*Dioscorea* sp.). *Mol Ecol Notes*. 2006;6:173–5.
- Hochu I, Santoni S, Bousalem M. Isolation, characterization and cross-species amplification of microsatellite DNA loci in the tropical American yam *Dioscorea trifida*. *Mol Ecol Notes*. 2006;6:137–40.
- Siqueira MVB, Marconi TG, Bonatelli ML, Zucchi MI, Veasey EA. New microsatellite loci for water yam (*Dioscorea alata*, *Dioscoreaceae*) and cross-amplification for other *Dioscorea* species. *Am J Bot*. 2011;98:e144–e146.
- Silva LR, Bajay MM, Monteiro M, Mezette TF, Nascimento WF, Zucchi MI, Pinheiro JB, Veasey EA. Isolation and characterization of microsatellites for the yam *Dioscorea cayenensis* (*Dioscoreaceae*) and cross-amplification in *D. rotundata*. *Genet Mol Res*. 2014;13(2):2766–71.
- Tamiru M, Yamanaka S, Mitsuoka C, Babil P, Takagi H, Lopez-Montes A, et al. Development of genomic simple sequence repeat markers for yam. *Crop Sci*. 2015;55:2191–200.
- Sattler MC, Carvalho CR, Clarindo WR. The polyploidy and its key role in plant breeding. *Planta*. 2016;243:281–96.
- Malapa R, Arnau G, Noyer JL, Lebot V. Genetic diversity of the greater yam (*Dioscorea alata* L.) and relatedness to *D. nummularia* lam. and *D. transversa* Br. As revealed with AFLP markers. *Genet Resour Crop Evol*. 2005;52:919–29.
- Lebot V. Tropical root and tuber crops. In: Verheye WH, editor. oils, plant growth crop prod. Encyclopedi. Oxford: Eloss Publishers; 2010. p. 9.
- Obidiegwu JE, Rodriguez E, Loureiro J, Muoneke CO, Asiedu R. Estimation of the nuclear DNA content in some representative of genus *Dioscorea*. *Sci Res Essay*. 2009;4:448–52.
- Adaramola TF, Sonibare MA, Sartie A, Lopez-Montes A, Franco J, Albach DC. Integration of ploidy level, secondary metabolite profile and morphological traits analyses to define a breeding strategy for trifoliate yam (*Dioscorea dumetorum* (Kunth) Pax). *Plant Genet Resour*. 2014;14:1–10.
- IRAD. Second report on the state of plant genetic resources for food and agriculture in Cameroon. Yaoundé: Institute of Agricultural Research and Development (IRAD); 2008. Available from: <http://www.fao.org/docrep/013/i1500e/Cameroon.pdf>
- Siadjeu C, Panyoo EA, Mahbou Somo Toukam G, Bell M, Nono B, Medoua GN. Influence of cultivar on the postharvest hardening of trifoliate yam (*Dioscorea dumetorum*) tubers. *Adv Agric*. 2016;2016:1–18.
- Rognes T, Flouri T, Nichols B, Quince C, Mahé FVSEARCH: A versatile open source tool for metagenomics. *PeerJ*. 2016;4:e2584.
- Edgar RCMUSCLE. A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 2004;5:1–19.
- Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 2006;23:254–67.
- Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*. 2012;8:1–17.

44. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945–59.
45. Verity R, Nichols RA. Estimating the number of subpopulations (K) in structured populations. *Genetics*. 2016;203:1827–35.
46. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 2005;14:2611–20.
47. Earl DA, vonHoldt BM. Structure harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conserv Genet Resour*. 2012;4:359–61.
48. Dolezel J, Greilhuber J, Lucretti S, Meister A, Lysak MA, Nardi L, Obermayer R. Plant genome size estimation by flow cytometry: inter-laboratory comparison. *Ann Bot*. 1998;82:17–26.
49. Gompert Z, Mock KE. Detection of individual ploidy levels with genotyping-by-sequencing (GBS) analysis. *Mol Ecol Resour*. 2017;17:1156–67.
50. Burns M, Hedin M, Tsurusaki N. Population genomics and geographical parthenogenesis in Japanese harvestmen (*Opiliones, Sclerosomatidae, Leiobunum*). *Ecol Evol*. 2018;8:36–52.
51. Torkamaneh D, Laroche J, Belzile F. Genome-wide SNP calling from genotyping by sequencing (GBS) data: a comparison of seven pipelines and two sequencing technologies. *PLoS One*. 2016;11:1–14.
52. Rafalski A. Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol*. 2002;5:94–100.
53. Zhu YL, Song QJ, Hyten DL, Van Tassel CP, Matukumalli LK, Grimm DR, et al. Single-nucleotide polymorphisms in soybean. *Genetics*. 2003;163:1123–34.
54. DaCosta JM, Sorenson MD. DdRAD-seq phylogenetics based on nucleotide, indel, and presence-absence polymorphisms: analyses of two avian genera with contrasting histories. *Mol Phylogenet Evol*. 2016;94:122–35.
55. Asiedu R, Wanyera NM NS and NQ. Yams. In: Fuccillo, Dominic, Sears, Linda, Stapleton P, editor. *Biodivers trust Conserv use plant genet Resour CGIAR centres*. Cambridge: Cambridge University Press; 1997. p. 371.
56. Egesi CN, Asiedu R, Egunjobi JK, Bokanga M. Genetic diversity of organoleptic properties in water yam (*Dioscorea alata* L.). *J Sci Food Agric*. 2003;83:858–65.
57. Janiak A, Kwaśniewski M, Szarejko I. Gene expression regulation in roots under drought. *J Exp Bot*. 2016;67:1003–14.
58. Arnau G, Bhattacharjee R, Mn S, Malapa R, Lebot V, Abraham K, et al. Understanding the genetic diversity and population structure of yam (*Dioscorea alata* L.) using microsatellite. *Mark Theory*. 2017;12:1–17.
59. Leitch IJ, Chase MW, Bennett MD. Phylogenetic analysis of DNA C-values provides evidence for a small ancestral genome size in flowering plants. *Ann Bot*. 1998;82:85–94.
60. Nemorin A, David J, Maledon E, Nudol E, Dalon J, Arnau G. Microsatellite and flow cytometry analysis to help understand the origin of *Dioscorea alata* polyploids. *Ann Bot*. 2013;112:811–9.
61. Ramsey J, Schemske DW. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu Rev Ecol Syst*. 1998;29:467–501.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



Additional file 1. Siadjeu et al. (2018)

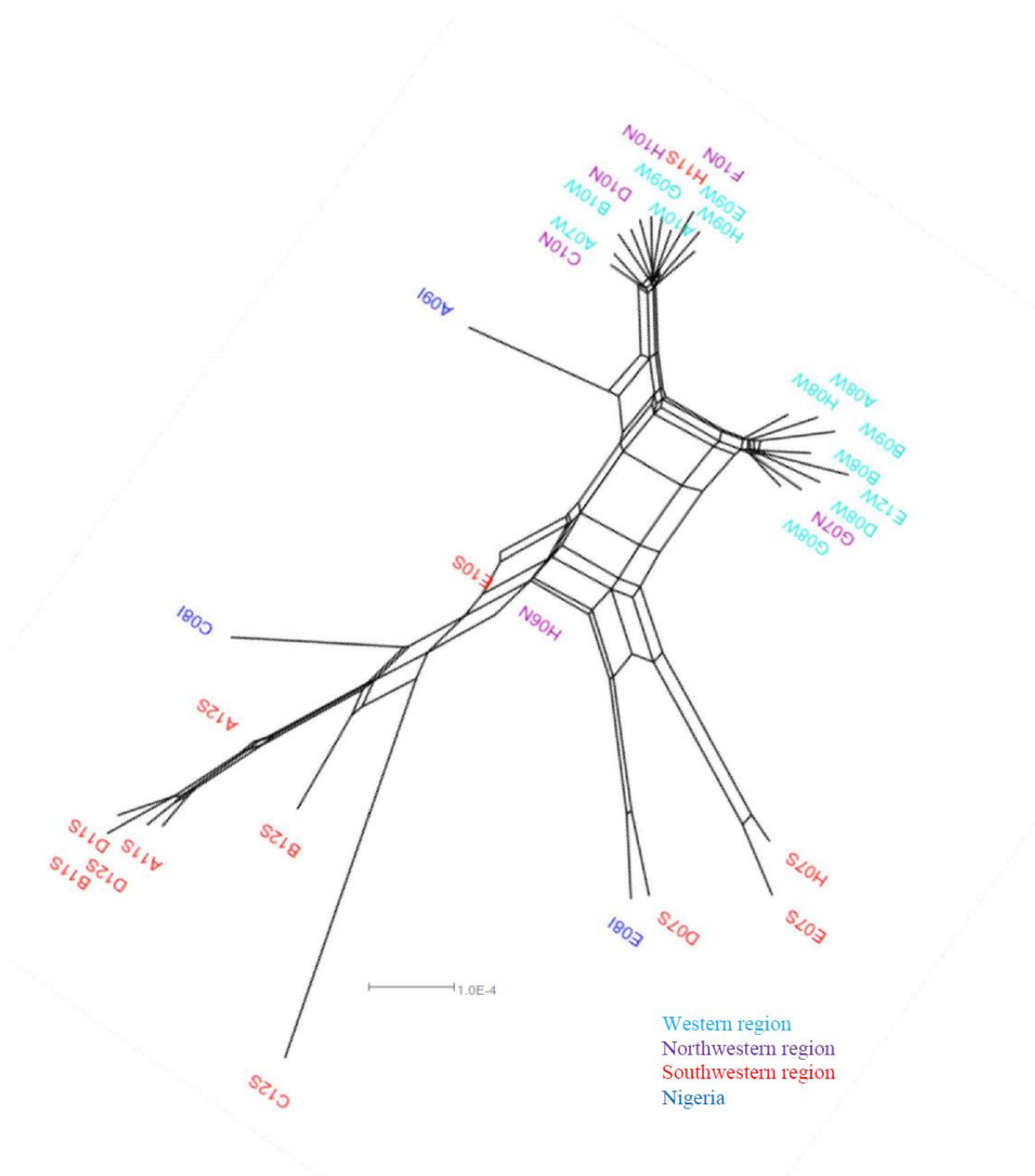


Figure S1. Phylogenetic relationships within *D. dumetorum* based on multilocus concatenated SNP sequences alignment from GBS data of 34 diploid accessions

Additional file 2. Siadjeu et al. (2018)

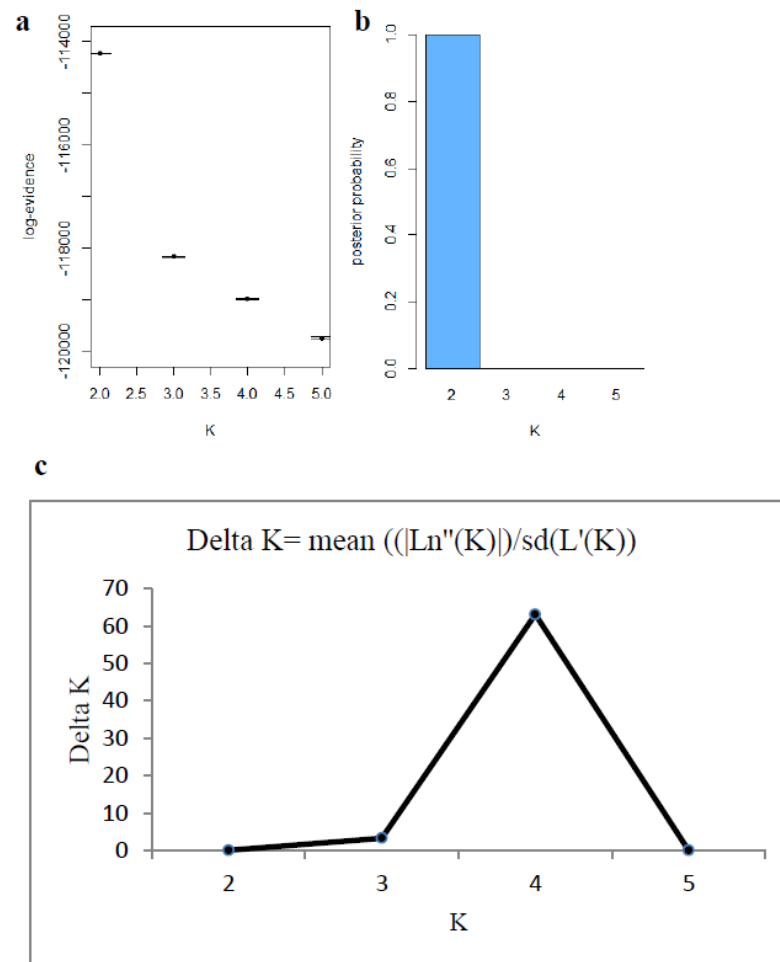


Figure S2. Estimates of the model evidence for K = 2:5 using TI estimator **a)** log-evidence and **b)** the evidence and Structure estimator Delta K ΔK **c)**

Additional file 3. Siadjeu et al. (2018)

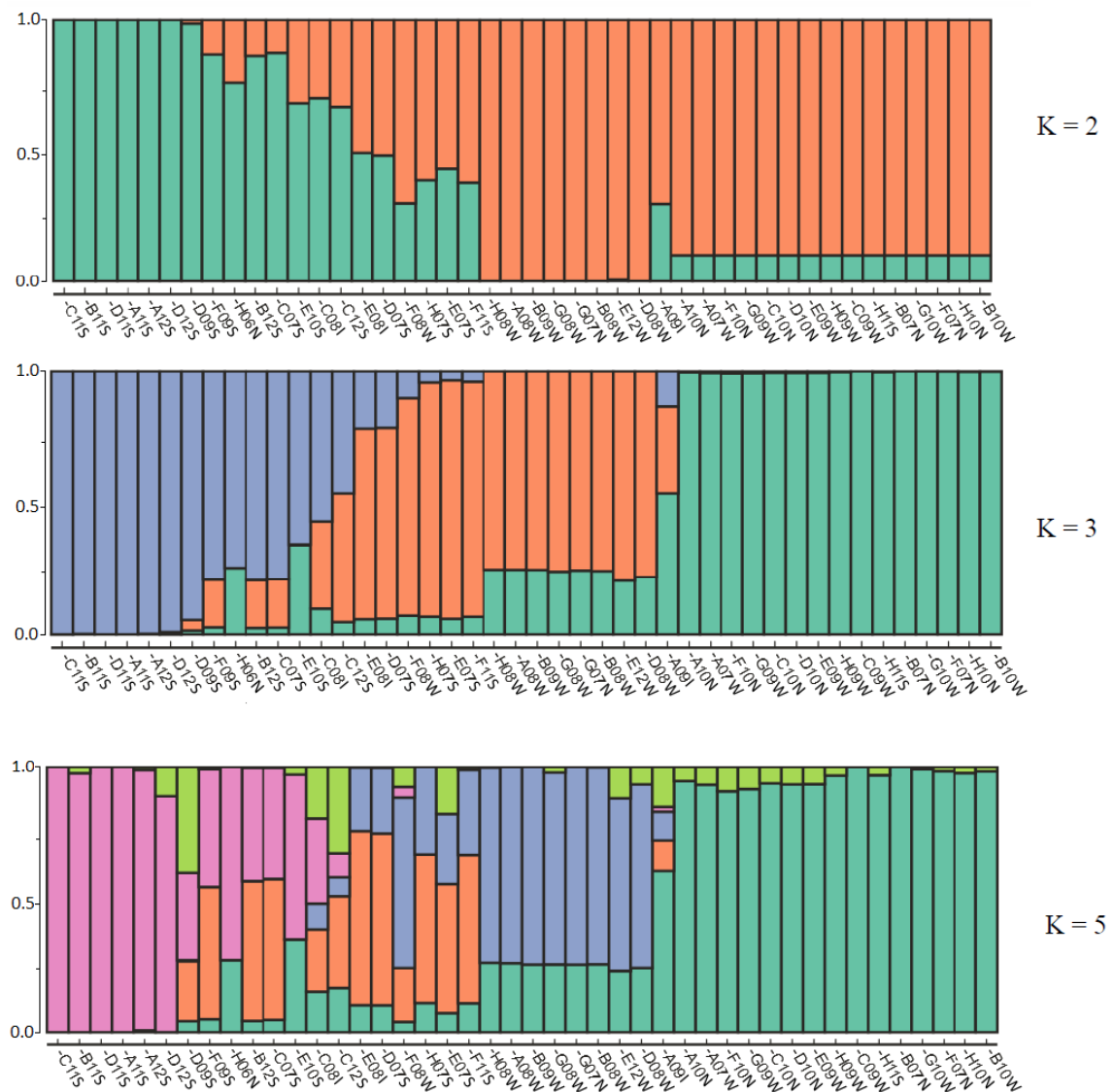


Figure S3. STRUCTURE plot of 44 accessions of *D. dumetorum* with K = 2, 3, 5 clusters based on 6457 unlinked SNPs.

Additional file 4. Siadjeu et al. (2018)

Table S1. Coefficient of variation of ploidy measurements using flow cytometric and ploidy level per accessions estimated by gbs2ploidy. * Ploidy level assessed by gbs2ploidy

Code	Origin	Sex	Ploidy level	1C-values	CV	Probability
A09I	Nigeria	Male	2x	0.329	4.52	
B08W	West	Female	2x	0.320	4.93	
B09W	West	Female	2x	0.345	4.40	
C08I	Nigeria	Male	2x	0.340	3.97	
D07S	Southwest	Female	2x	0.330	4.08	
E07S	Southwest	Male	2x	0.322	4.86	
E08I	Nigeria	Female	2x	0.336	3.89	
E12W	West	Female	2x	0.333	4.64	
G07N	Northwest	Female	2x	0.322	4.79	
G08W	West	Female	2x	0.338	4.54	
G09W	West	Female	2x	0.350	4.22	
H07S	Southwest	Male	2x	0.329	4.17	
H08W	West	Female	2x	0.333	4.50	
C07S	Southwest	Male	3x	0.514	4.94	
C11S	Southwest	Female	3x	0.520	3.08	
D09S	Southwest	Male	3x	0.520	4.60	
F09S	Southwest	Male	3x	0.523	3.51	
A07W*	West		2x			79.88
A08W*	West		2x			55.35
A10W*	West		2x			76.19
A11S*	Southwest		2x			77.65
A12S*	Southwest		2x			75.04
B10W*	West		2x			79.22
B11S*	Southwest		2x			60.45
B12S*	Southwest		2x			70.51
C10N*	Northwest		2x			73.86
C12S*	Southwest		2x			75.72
D08W*	West		2x			61.9
D10N*	Northwest		2x			83.35
D11S*	Southwest		2x			71.07
D12S*	Southwest		2x			51.05
E09W*	West		2x			74.7
E10S*	Southwest		2x			71.55
F10N*	Northwest		2x			84.42
H06N*	Northwest		2x			84.78
H09W*	West		2x			79.1
H10N*	Northwest		2x			83.66
H11S*	Southwest		2x			79.89
B07N*	Northwest		3x			88.51
C09W*	West		3x			83.71
F07N*	Northwest		3x			73.83
F08W*	West		3x			55.14
F11S*	Southwest		3x			58.61
G10N*	Northwest		3x			80.17

Additional file 5. Siadjeu et al. (2018)

Edges, weight of migrations (8 and 9) and likelihood for migrations 0:9.

Results of 8 Migrations

0.283022 NA NA NA G10N:0.0273162 F11S:0

0.480978 NA NA NA (F11S:0,E07:0):0.0279907 H07:0

0.245203 NA NA NA D07:0 C12S:0.0376164

0.390885 NA NA NA B11S:0 E10S:0

0.375997 NA NA NA H06N:0

((F09:0,((F11:0,E07:0):0.0279907,(C07S:0,B12S:0):0):0):0.024769,H07S:0):8.33947e-07

0.133431 NA NA NA (C07S:0,B12S:0):0 D09S:0.0138161

0.246653 NA NA NA (F09:0,((F11:0,E07:0):0.0279907,(C07S:0,B12S:0):0):0):0.024769 F09S:0

0.246845 NA NA NA H07S:0 B12S:0

Results of 9 Migrations

0.282806 NA NA NA G10:0.0272958 F11:0

0.490486 NA NA NA (F11:0,E07:0):0.0279641 H07:0

0.245381 NA NA NA D07:0 C12:0.0376335

0.390802 NA NA NA B11:0 E10:0

0.374222 NA NA NA H06:0

((F09:0,((F11:0,E07:0):0.0279641,(C07:0,B12:0):1.62225):0):0.0243862,H07:0):4.48654e-06

0.114021 NA NA NA (C07:0,B12:0):1.62225 D09:0

0.247588 NA NA NA (F09:0,((F11:0,E07:0):0.0279641,(C07:0,B12:0):1.62225):0):0.0243862 F09:0

0.265545 NA NA NA H07:0 B12:0

0.965544 NA NA NA F09:0 (C07:0,B12:0):1.62225

Likelihood of 9 migrations

Starting ln(likelihood) with 0 migration events: 4174.55

Exiting ln(likelihood) with 1 migration events: 4314.25

Exiting ln(likelihood) with 2 migration events: 4336.66

Exiting ln(likelihood) with 3 migration events: 4341.66

Exiting ln(likelihood) with 4 migration events: 4346.32

Exiting ln(likelihood) with 5 migration events: 4348.77

Exiting ln(likelihood) with 6 migration events: 4355.06

Exiting ln(likelihood) with 7 migration events: 4356.21

Exiting ln(likelihood) with 8 migration events: 4359.38

Exiting ln(likelihood) with 9 migration events: 4360.99

Additional file 6. Siadjeu et al. (2018)

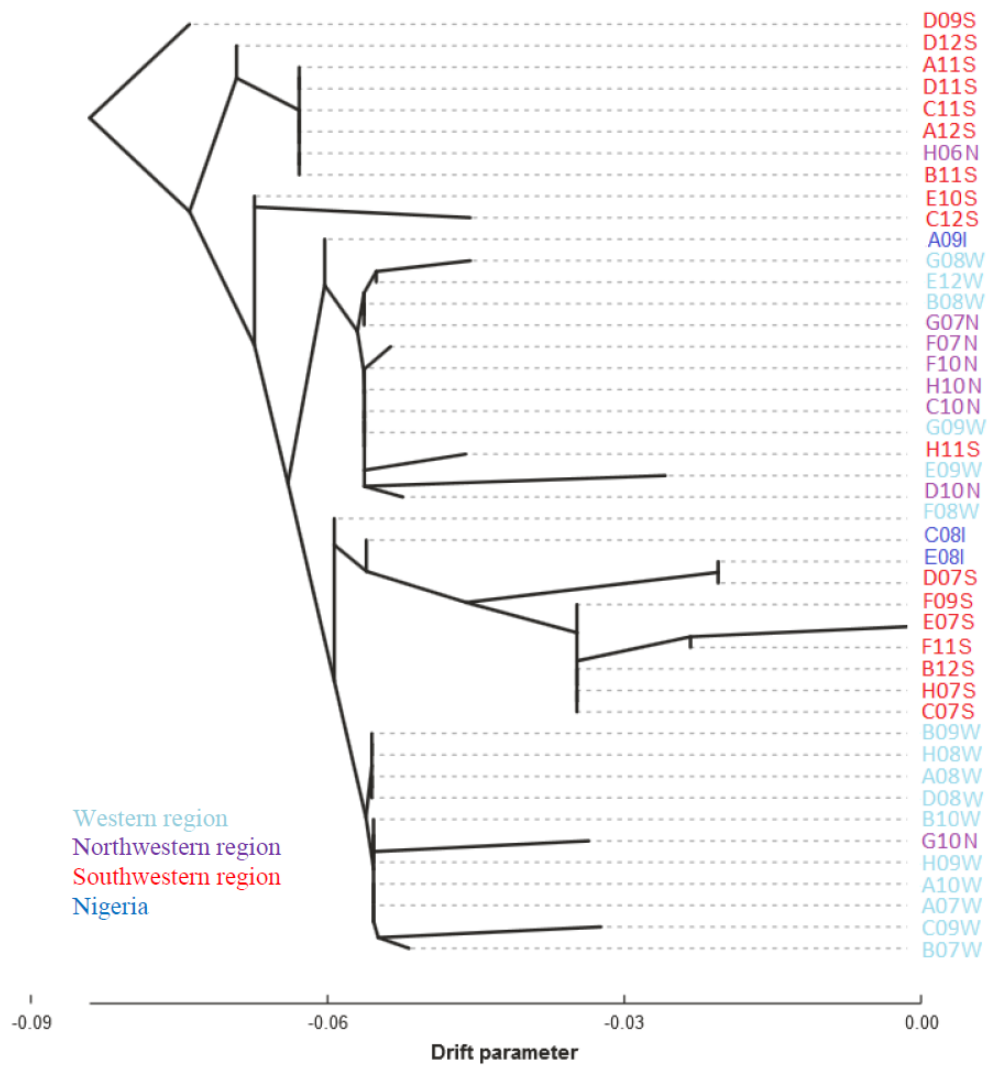


Figure S4. Maximum likelihood tree of the inferred gene flow within *D. dumetorum* species with no gene flow events. (PDF 125 kb)

Additional file 7. Siadjeu et al. (2018)

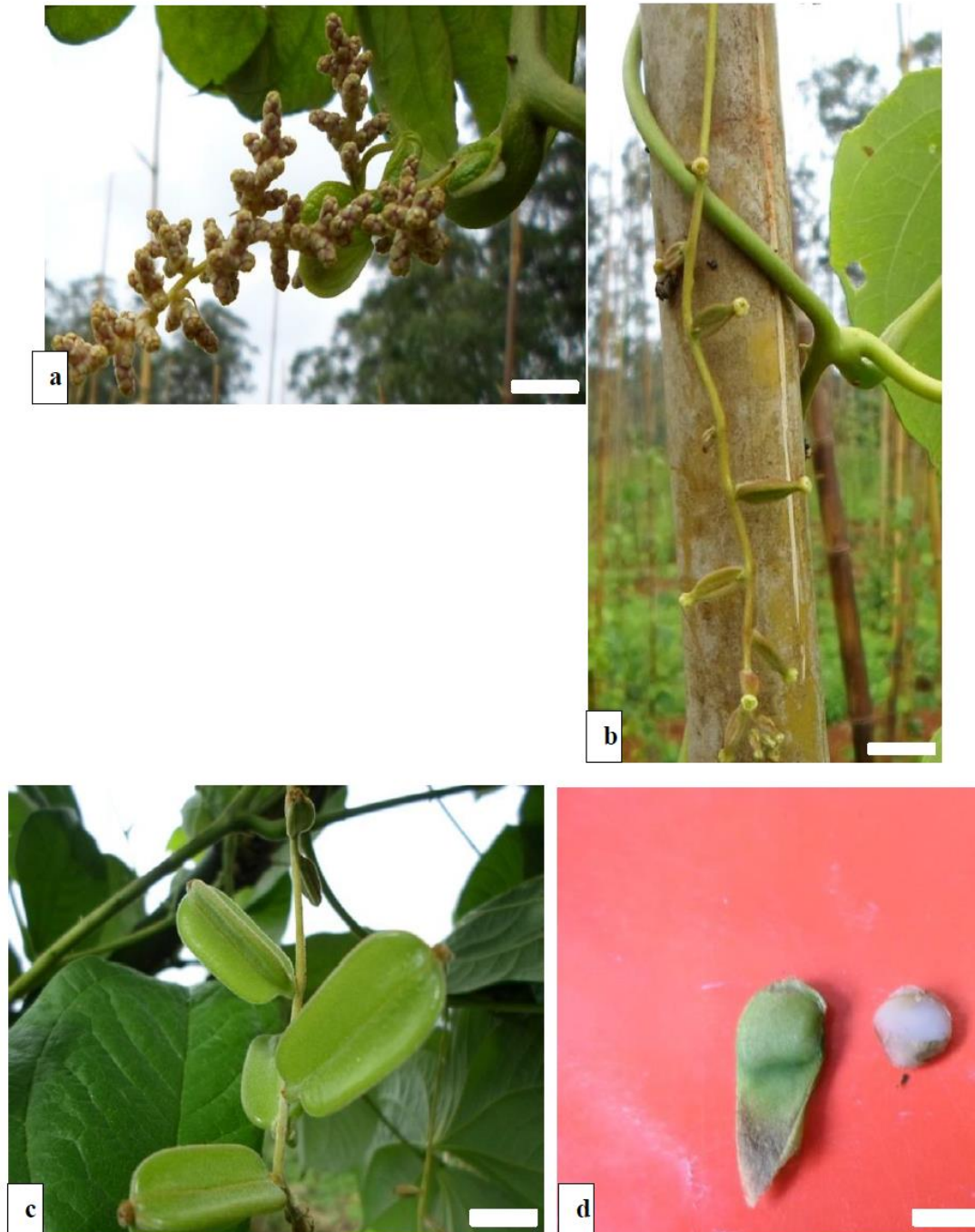


Figure S5. Flowering and fructification of *D. dumetorum*. a) male flower, b) female flower. Bar scale = 3 cm. c) fruits, d) seeds. Bar scale = 2 cm

ARTICLE II:

HIGH CONTIGUITY DE NOVO GENOME SEQUENCE
ASSEMBLY OF TRIFOLIATE YAM (*DIOSCOREA*
DUMETORUM) USING LONG READ SEQUENCING



Article

High Contiguity De Novo Genome Sequence Assembly of Trifoliolate Yam (*Dioscorea dumetorum*) Using Long Read Sequencing

Christian Siadjeu ^{1,2,†}, Boas Pucker ^{2,3,†}, Prisca Viehöver ², Dirk C. Albach ¹ and Bernd Weisshaar ^{2,*}

¹ Institute for Biology and Environmental Sciences, Biodiversity and Evolution of Plants, Carl-von-Ossietzky University Oldenburg, Carl-von-Ossietzky Str. 9-11, 26111 Oldenburg, Germany; christian.siadjeu@uol.de (C.S.); dirk.albach@uol.de (D.C.A.)

² Genetics and Genomics of Plants, Faculty of Biology, Center for Biotechnology (CeBiTec), Bielefeld University, Sequenz 1, 33615 Bielefeld, Germany; bpucker@cebitec.uni-bielefeld.de (B.P.); viehoeve@cebitec.uni-bielefeld.de (P.V.)

³ Molecular Genetics and Physiology of Plants, Faculty of Biology and Biotechnology, Ruhr-University Bochum, Universitätsstraße 150, 44801 Bochum, Germany

* Correspondence: bernd.weisshaar@uni-bielefeld.de; Tel.: +49-521-106-8720

† These authors contributed equally to this work.

Received: 31 January 2020; Accepted: 29 February 2020; Published: 4 March 2020

Abstract: Trifoliolate yam (*Dioscorea dumetorum*) is one example of an orphan crop, not traded internationally. Post-harvest hardening of the tubers of this species starts within 24 h after harvesting and renders the tubers inedible. Genomic resources are required for *D. dumetorum* to improve breeding for non-hardening varieties as well as for other traits. We sequenced the *D. dumetorum* genome and generated the corresponding annotation. The two haplophases of this highly heterozygous genome were separated to a large extent. The assembly represents 485 Mbp of the genome with an N₅₀ of over 3.2 Mbp. A total of 35,269 protein-encoding gene models as well as 9941 non-coding RNA genes were predicted, and functional annotations were assigned.

Keywords: yam; *D. dumetorum*; nanopore sequencing; genome assembly; comparative genomics; read depth

1. Introduction

The yam species *Dioscorea dumetorum* (Trifoliolate yam) belongs to the genus *Dioscorea* comprising about 600 described species. The genus is widely distributed throughout the tropics [1] and includes important root crops that offer staple food for over 300 million people. Eight *Dioscorea* species are commonly consumed in West and Central Africa, of which *D. dumetorum* has the highest nutritional value [2]. Tubers of *D. dumetorum* are protein-rich (9.6%) with a fairly balanced essential amino acids composition [3]. The provitamin A and carotenoid contents of the tubers of deep yellow genotypes are equivalent to those of yellow corn maize lines selected for increased concentrations of provitamin A [4]. The deep yellow yam tubers are used in antidiabetic treatments in Nigeria [5], probably due to the presence of dioscoretine, which is a bioactive compound with hypoglycaemic properties [6]. Yet, *D. dumetorum* constitutes an underutilized and neglected crop species despite its great potential for nutritional, agricultural, and pharmaceutical purposes.

Unlike other yam species, the agricultural value of *D. dumetorum* is limited by post-harvest hardening, which starts within 24 h after harvest and renders tubers inedible. Previous research showed that among 32 *D. dumetorum* cultivars tested, one cultivar was not affected by the hardening

phenomenon [7]. This discovery provides a starting point for a breeding program of *D. dumetorum* against the post-harvest hardening phenomenon. *Dioscorea* cultivars are obligate outcrossing plants that display highly heterozygous genomes. Thus, methods of genetic analysis routinely applied in inbreeding species such as linkage analysis using the segregation patterns of an F2 generation or recombinant inbred lines are inapplicable to yam [8]. Furthermore, the development of marker-assisted selection requires the establishment of marker assays and dense genetic linkage maps. Thus, access to a complete and well-annotated genome sequence is one essential step towards the implementation of comprehensive genetic, genomic, and population genomic approaches for *D. dumetorum* breeding. So far, a genome sequence assembly for *Dioscorea rotundata* (Guinea yam) [8] and a reference genetic map for *Dioscorea alata* (Greater yam) [9] have been released. However, these two species belong to the same section of *Dioscorea* (*D. sect. Enantiophyllum*) but are distant from *D. dumetorum* (*D. sect. Lasiophyton*) in phylogenetic analyses [10,11]. They also differ in chromosome number [8,12,13] making it unlikely that genetic maps can be directly transferred to *D. dumetorum*. Here, we report long read sequencing and de novo genome sequence assembly of the cultivar *D. dumetorum* Ibo sweet 3 that does not display post-harvest hardening.

2. Materials and Methods

2.1. Sampling and Sequencing

The *D. dumetorum* accession Ibo sweet 3 that does not display post-harvest hardening had been collected in the South-West region of Cameroon in 2013 [7]. Tubers of this accession were transferred to Oldenburg (Germany) and the corresponding plants were cultivated in a greenhouse at 25 °C. The haploid genome size of the Ibo sweet 3 genotype had been estimated to be 322 Mbp through flow cytometry [14].

High molecular weight DNA was extracted from 1 g of leaf tissue using a CTAB-based method modified from [15]. After grinding the sample in liquid nitrogen, the powder was suspended in 5 mL CTAB1 (100 mM Tris-HCl pH 8.0, 20 mM EDTA, 1.4 M NaCl, 2% CTAB, 0.25% PVP) buffer supplemented with 300 µL β-mercaptoethanol. The suspension was incubated at 75 °C for 30 min and inverted every 5 min. Next, 5 mL dichloromethane was added and the solutions were gently mixed by inverting. The sample was centrifuged at 12,000 g at 20 °C for 30 min. The clear supernatant was mixed with 10 mL CTAB2 (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% CTAB, 0.125% PVP) in a new reaction tube by inverting. Next, a centrifugation was performed at 12,000 g at 20 °C for 30 min. After discarding the supernatant, 1 mL NaCl (1 M) was added to re-suspend the sediment by gently flicking the tube. By adding an equivalent amount of 1 mL isopropanol and careful mixing, the DNA was precipitated again and the sample was centrifuged as described above. After washing the sediment with 1 mL of 70% ethanol, 200 µL TE buffer (10 mM Tris pH 8.0, 0.1 mM EDTA) containing 2 mg DNase-free RNaseA were added. Re-suspension and RNA degradation were achieved by incubation overnight at room temperature. DNA quality and quantity were assessed via NanoDrop2000 (Thermo Fisher Scientific, Waltham, MA, USA) measurement, agarose gel electrophoresis, and Qubit (Invitrogen, Carlsbad, CA, USA) measurement. The short read eliminator (SRE) kit (Circulomics, Baltimore, MD, USA) was used to enrich long DNA fragments following the suppliers' instructions. Results were validated via Qubit measurement.

Library preparation was performed with 1 µg of high molecular weight DNA following the SQK-LSK109 protocol (Oxford Nanopore Technologies, ONT, Oxford, UK). Sequencing was performed on four R9.4.1 flow cells on a GridION. Flow cells were treated with nuclease flush (20 µL DNaseI (New England Biolabs (NEB), Ipswich, MA, USA) and 380 µL nuclease flush buffer) once the number of active pores dropped below 200, to allow successive sequencing of multiple libraries on an individual flow cell. Live base calling was performed on the GridION by Guppy v3.0 (ONT).

A total of 200 ng high molecular weight gDNA was fragmented by sonication using a Bioruptor (Diagenode) and subsequently used for Illumina library preparation. End-repaired fragments were

size selected by AmpureXp Beads (Beckmann-Coulter) to an average size of 650 bp. After A-tailing and adaptor ligation fragments that carried adaptors on both ends were enriched by eight cycles of PCR (TruSeq Nano DNA Sample Kit; Illumina, San Diego, CA, USA). The final library was quantified using PicoGreen (Quant-iT, Invitrogen, CA, USA) on a FLUOstar plate reader (BMG labtech, Ortenberg, Germany) and quality checked by HS-Chips on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The PE library was sequenced in 2×250 nt mode on an Illumina HiSeq-1500 instrument.

2.2. Genome Assembly and Polishing

Genome size prediction was performed with GenomeScope [16], findGSE [17], and gce [18] based on k-mer histograms generated by JellyFish v2 [19] as previously described [20] for different k-mer size values. In addition, MGSE [20] was run on an Illumina read mapping with single copy BUSCOs as reference regions for the haploid coverage calculation. Smudgeplot [21] was run on the same k-mer histograms (also for different k-mer size values) as the genome size estimations to estimate the ploidy.

Canu v1.8 [22] was deployed for the genome assembly. Raw ONT reads were provided as input to Canu for correction and trimming. Subsequently, Canu assembled the genome sequence from the resulting polished reads. The following optimized parameters were used “genomeSize = 350 m’, ‘corOutCoverage = 200’ ‘correctedErrorRate = 0.12’ batOptions = -dg 3 -db 3 -dr 1 -ca 500 -cp 50’ ‘minReadLength = 10000’ ‘minOverlapLength = 5000’ ‘corMhapFilterThreshold = 0.0000000002’ ‘ovlMerThreshold = 500’ ‘corMhapOptions = -threshold 0.85 -num-hashes 512 -num-min-matches 3 -ordered-sketch-size 1000 -ordered-kmer-size 14 -min-olap-length 5000 -repeat-idf-scale 50’”. The selected parameters were optimized for the assembly of a heterozygous genome sequence and our data set. The value for the genome size, estimated to be 322 Mbp, was increased to 350 Mbp to increase the number of reads utilized for the assembly process. A total of 66.7 Gbp of ONT reads with an N_{50} of 23 kbp was used for assembly, correction, and trimming.

ONT reads were mapped back to the assembled sequence with minimap2 v2.17 [23], using the settings recommended for ONT reads. Next, the contigs were polished with racon v1.4.7 [24] with -m 8 -x -6 -g -8 as recommended prior to the polishing step with medaka. Two runs of medaka v0.10.0 (<https://github.com/nanoporetech/medaka>) polishing were performed with default parameters (-m r941_min_high) using ONT reads. Illumina short reads were aligned to the medaka consensus sequence using BWA-MEM v. 0.7.17 [25]. This alignment was subjected to Pilon v1.23 [26] for final polishing in three iterative rounds with default parameters for the correction of all variant types and -mindepth 4.

Downstream processing was based on a previously described workflow [27] and was performed by customized Python scripts for purging of contigs shorter than 100 kbp and calculation of assembly statistics (<https://github.com/bpucker/yam>). In general, sequences were kept if matching a white list (*D. rotundata*) and discarded if matching a black list (bacterial/fungal genome sequences). Sequences with perfect matches against the genome sequences of plants that were sequenced in the lab in parallel (*Arabidopsis thaliana*, *Beta vulgaris*, and *Vitis vinifera*) were discarded as well. Contigs with less than 3-fold average coverage in an Illumina short read mapping were compared against nt via BLASTn with an e-value cut-off at 10^{-10} to identify and remove additional bacterial and fungal sequences.

For sorting (“scaffolding” according to linkage groups) the *D. dumetorum* assembly we employed *D. rotundata* pseudochromosomes. *D. rotundata* pseudochromosome sequences longer than 100 kbp were split into chunks of 1000 bp and subject to a BLASTn search against the *D. dumetorum* assembly with a word size of 12. Hits were considered if the similarity was at least 70% and if at least 70% of the query length were covered by the alignment. To avoid ambiguous hits against close paralogs or between repeat units, BLAST hits were excluded if the second hit exceeded 90% of the score of the top hit. The known order of all chunks on the *D. rotundata* sequence was considered as a “pseudo genetic map” to arrange the *D. dumetorum* contigs via ALLMAPS v0.9.14 [28].

2.3. Genome Sequence Annotation

Hints for gene prediction were generated by aligning *D. rotundata* transcript sequences (TDr96 v1.0) [8] as previously described [29]. BUSCO v3 [30] was applied to generate a species-specific set of AUGUSTUS gene prediction parameter files. For comparison of annotation results, the *D. rotundata* genome assembly GCA_002260605.1 [8] was retrieved from NCBI. Gene prediction hints of *D. dumetorum* and dedicated parameters were subjected to AUGUSTUS v3.3 [31] for gene prediction with previously described settings [29]. Various approaches involving AUGUSTUS parameter files for rice and maize genome sequences provided by AUGUSTUS, as well as running the gene prediction on a sequence with repeats masked by RepeatMasker v4.0.8 [32] with default parameters, were evaluated. BUSCO was applied repeatedly to assess the completeness of the gene predictions. The best results for *D. dumetorum* genome sequence annotation were obtained by using an unmasked assembly sequence and by applying yam specific AUGUSTUS gene prediction parameter files generated via BUSCO as previously described [30,33]. Predicted genes were filtered based on sequence similarity to entries in several databases (UniProt/SwissProt, Araport11, *Brachypodium distachyon* v3.0, *Elaeis guineensis* v5.1, GCF_000005425.2, GCF_000413155.1, *Musa acuminata* Pahang v2). Predicted peptide sequences were compared to these databases via BLASTp [34] using an e-value cut-off of 10^{-5} . Scores of resulting BLASTp hits were normalized to the score when searched against the set of predicted peptides. Only predicted sequences with at least 0.25 score ratio and 0.25 query length covered by the best alignment were kept. Representative transcript and peptide sequences were identified per gene to encode the longest possible peptide as previously established [29,35]. GO terms were assigned via InterProScan5 [36]. Reciprocal best BLAST hits against Araport11 [35] were identified based on a previously developed script [27]. Remaining sequences were annotated via best BLAST hits against Araport11 with an e-value cut-off at 0.0001. The Araport11 annotation was transferred to predicted sequences.

Prediction of non-protein coding RNA genes like tRNA and rRNA genes was performed based on tRNAscan-SE v2.0.3 [37,38] and INFERNAL (cmscan) v1.1.2 [39] based on Rfam13 [40].

RepeatModeler v2 [41] was deployed with default settings for the identification of repeat family consensus sequences.

2.4. Assembly and Annotation Assessment

The percentage of phase-separated and merged regions in the genome assembly was assessed with the focus on predicted genes. Based on Illumina and ONT read mappings, the average coverage depth per gene was calculated. The distribution of these average values per gene allowed the classification of genes as phase-separated (haploid read depth) or merged (diploid read depth). As previous studies revealed that Illumina short reads have a higher resolution for such coverage analysis [42], we focused on the Illumina read data set for these analyses. Sequence variants were detected based on this read mapping as previously described [43]. The number of heterozygous variants per gene was calculated and compared between the groups of putatively phase-separated and merged genes. Predicted peptide sequences were compared against the annotation of other species including *A. thaliana* and *D. rotundata* via OrthoFinder v2 [44].

Sequence reads and assembled sequences are available at ENA under the project ID ERP118030 (see File S1 for details). The assembly described in this manuscript is available under GCA_902712375. Additional annotation files including the contigs assigned to organelle genomes are available as a data publication from the institutional repository of Bielefeld University at <https://doi.org/10.4119/unibi/2941469>.

Alleles covered by the fraction of phase-separated gene models were matched based on reciprocal best BLAST hits of the coding sequences (CDSs) following a previously described approach [27]. Alleles were considered a valid pair that represents a single gene if the second best match displayed 99% or less of the score of the best match. A customized Python script for this allele assignment is available on github (<https://github.com/bpucker/yam>).

3. Results

In total, we generated 66.7 Gbp of ONT reads data representing about 207× coverage of the estimated 322 Mbp haploid *D. dumetorum* genome. Read length N_{50} of the raw ONT data set was 23 kbp and increased to 38 kbp through correction, trimming, and filtering. Additionally, 30 Gbp of Illumina short read data (about 100× coverage) were generated. After polishing, the final assembly represents 485 Mbp of the highly heterozygous *D. dumetorum* genome with an N_{50} of 3.2 Mbp (Table 1). Substantial improvement of the initial assembly through various polishing steps was indicated by the increasing number of recovered BUSCOs (File S2). The final assembly displayed more BUSCOs (92.30% out of 1440 included in the embryophyta data set, see File S2 for details on the various BUSCO classes) compared to the publicly available genome sequence assembly of *D. rotundata* (v0.1) for that we detected 81.70% BUSCOs with identical parameters. Since there is no genetic map available for *D. dumetorum*, we transferred linkage group assignments from *D. rotundata* to our assembly. In total, 206 contigs comprising 330 Mbp were assigned to a linkage group, while 718 contigs remained unplaced with a total sequence of 155 Mbp (File S3). One plastid and six mitochondrial contigs were identified based on sequence similarity to *D. rotundata* organelle genome sequences (see <https://doi.org/10.4119/unibi/2941469>); the assignment was confirmed by very high coverage in the read mapping. Our *D. dumetorum* plastid sequence turned out to be almost identical to the data recently provided for the *D. dumetorum* plastome [11].

Table 1. Statistics of selected versions of the *Dioscorea dumetorum* genome assembly (see File S5 for a full table).

	Initial Assembly	Racon1	Medaka2	Pilon3	Final
Number of contigs	1172	1172	1215	1215	924
Max. contig length (bp)	20,187,448	20,424,333	17,910,017	17,878,854	17,878,854
Assembly size (bp)	501,985,705	508,061,170	507,215,754	506,184,192	485,115,345
Assembly size without N (bp)	501,985,705	508,061,170	507,215,754	506,184,192	485,115,345
GC content	37.74%	37.66%	37.87%	37.59%	37.57%
N_{50} (bp)	3,896,882	3,930,287	2,598,889	2,593,751	3,190,870
N_{90} (bp)	136,614	138,199	137,206	136,754	156,407
BUSCO (complete)	85.70%	89.80%	91.90%	92.30%	92.30%

Haploid genome size estimations based on k-mer distributions of the Illumina sequence reads ranged from 215 Mbp (gce) over 254 Mbp (GenomeScope) to 350 Mbp (findGSE, MGSE) (File S4). The differences between the estimates might be influenced by the repeat content of the *D. dumetorum* genome (see below).

Different gene prediction approaches were evaluated (File S6) leading to a final set of 35,269 protein-encoding gene models. The average gene model spans 4.3 kbp, comprises six exons and encodes 455 amino acids (see File S6 for details). The gene prediction dataset for *D. dumetorum* is further supported by the identification of 6475 single copy orthologs between *D. dumetorum* and *D. rotundata* as well as additional orthogroups (File S7). Based on these single copy orthologs, the similarity of *D. dumetorum* and *D. rotundata* sequences was determined to be mostly above 80% (File S8). If the phase-separated allelic gene models were considered (Figure 1), 3352 additional single copy orthologs were detected. Functional annotation was assigned to 23,835 gene models (File S9). Additionally, 9941 non-coding RNA gene models were predicted including 784 putative tRNA genes (see <https://doi.org/10.4119/unibi/2941469>). Finally, and in addition to gene models encoding proteins and various RNA types, we identified 1129 repeat consensus sequences with a combined length of 1.3 Mbp (File S10). The maximal repeat consensus length is 17.4 kbp, while the N_{50} is only 2.5 kbp.

Average read mapping depth per gene was analyzed to distinguish genes annotated in separated haplophases as well as merged sequences, respectively (Figure 1, File S11). About 64% of all predicted protein-encoding gene models were in the expected range of the haploid read mapping depth between 50-fold and 150-fold and about 27% are merged with a read depth between 150-fold and 250-fold. Only 6% of all genes show an average read depth below 50-fold and only 1% show an average coverage

higher than 250-fold. It should be noted that the gene models annotated in the phase separated part will cover in general two alleles per gene. A total of 22,885 gene models, representing the 64% in the range of the haploid read mapping depth, were sorted into allelic pairs which was successful for 8492 genes. The findings presented above can be explained by a diploid genome. An analysis with Smudgeplot indicated hints for a tetraploid genome from analysis with a k-mer size of 19, while the other three investigated k-mer sizes supported a diploid genome (File S12).

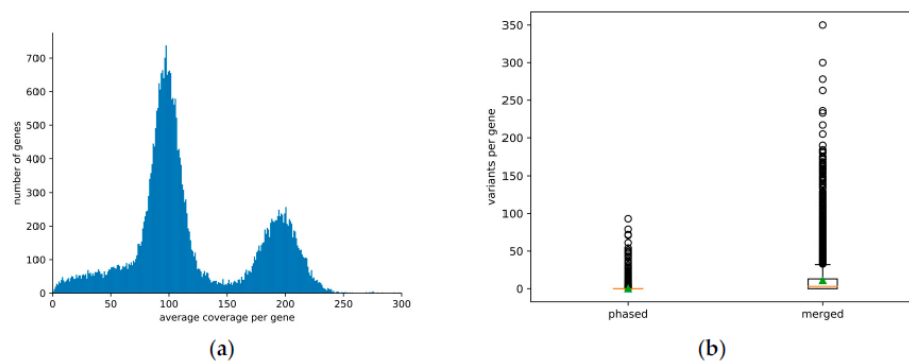


Figure 1. (a) Distribution of the average sequencing read depth per gene model. Predicted gene models were classified into phase-separated and merged based on the average read depth value deduced from the analysis presented here. The haploid read depth with Illumina short reads ranges from 50-fold to 150-fold. (b) Number of heterozygous sequence variants in phase separated and merged genes. The high proportion of heterozygous variants in merged gene models is due to the mapping of reads originating from two different alleles to the same region of the assembly.

4. Discussion

The release of genome sequences of many model and crop plants has provided new opportunities for gene identification and studies of genome evolution, both ultimately serving the process of plant breeding [45] by allowing discovery of genes responsible for important agronomic traits and the development of molecular markers associated with these traits. Here, we present the first genome sequence for *D. dumetorum*, an important crop for Central and Western Africa, and the second genome sequence for the genus. Our assembly offers a great opportunity to understand the evolution of yam and to elucidate some biological constraints inherent to yam including a long growth cycle, poor to non-flowering, polyploidy, vegetative propagation, and a heterozygous genetic background [46]. Yam improvement has been challenging due to these factors preventing the genetic study of important traits in yam [47].

Oxford Nanopore Technology sequencing has proven to be a reliable and affordable technology for sequencing genomes thus replacing Illumina technique for de novo genome sequencing due to substantially higher assembly continuity [42,48]. Large fractions of the genome sequence were separated into phases, while regions with lower heterozygosity are merged into one representative sequence. Coverage analysis with Illumina read mapping allowed to classify predicted gene models as ‘phase-separated’ or ‘merged’ based on an average coverage around 100-fold or around 200-fold, respectively. While this distinction is possible at the gene model level, whole contigs cannot be classified this way. Several Mbp long contigs comprise alternating phase-separated and merged regions. Therefore, it is likely that the contigs represent a mixture of both haplophases with the risk of switching between phases at each merged region. Since the haplophases cannot be resolved continuously through low heterozygosity regions, purging of contigs to reduce the assembly into a representation of the haploid genome might be advantageous for some applications in the future. The bimodal coverage distribution (Figure 1a) supports the assumption that *D. dumetorum* Ibo sweet 3 has a diploid genome. This is supported by Smudgeplot for three out of four k-mer sizes tested while

the shortest k-mer size used (19) finds indications for tetraploidy. Since a high ploidy would result in more distinct coverage peaks as observed for a genome with up to pentaploid parts [42], we assume that the genome is diploid. The weak hint for tetraploidy might be due to a whole genome duplication event early in the diversification of the genus. The N_{50} of 3.2 Mbp is in the expected range for a long read assembly of a highly heterozygous plant species which contains some quite repetitive sequences as others reported similar values before [49]. Due to regions of merged haplophases the total assembly size of 485 Mbp is smaller than expected for a fully phase-separated “diploid” genome sequence based on the haploid genome size estimation of 322 Mbp.

We noticed an increase in the number of BUSCOs through several polishing rounds. Initial assemblies of long reads can contain numerous short insertions and deletions as these are the major error type in ONT reads [50]. As a result, the identification of CDSs and deduced open reading frames is hindered through apparent disruptions of some CDS. Through the applied polishing steps, the number of such apparent frame shifts is reduced thus leading to an increase of detected BUSCOs.

D. dumetorum has 36 chromosomes [12], so with 924 contigs we are far from chromosome-level resolution but considerably better than the other genome assembly published in the genus, that of *D. rotundata* with 40 chromosomes [8]. Knuth [51] circumscribed *D. dumetorum* and *D. rotundata* in two distant sections *D. sect. Lasiophyton* and *D. sect. Enantiophyllum*, respectively. Additionally, phylogenetically the two species are quite distantly related with a last common ancestor about 30 million years ago [11,52]. Comparing our predicted peptides to the *D. rotundata* peptide set [8], we identified about 9800 single copy orthologs (6475 in the whole set of 35,269 gene models plus 3352 with a relation of one gene in *D. rotundata* and two phase-separated alleles in *D. dumetorum*) which could elucidate the evolutionary history of those species. The total number of predicted protein-encoding gene models was determined to be 35,269, but this number includes two copies of about 11,300 gene models (see Figure 1) as these are represented by two alleles each. The CDS-based pairing we performed detected about 8500 of the theoretical maximum of 11,300 cases which is a good success rate given the fact that close paralogs and also hemizygous genome regions contribute to the detected number of phase-separated gene models. If phase-separated gene models (alleles) are excluded, a number of about 24,000 genes would result for *D. dumetorum*. This fits to the range detected in other plant genomes [53,54]. The BUSCO results support this interpretation with about 40% of BUSCOs that occur with exactly two copies. Therefore, the true number of protein-encoding genes of a haploid *D. dumetorum* (Trifoliolate yam) genome could be around 25,000, also considering that the BUSCO analysis indicated by 5.8% missing BUSCOs that still a small fraction of the genome is not represented in the current assembly. This gene number fits well to gene numbers of higher plants based on all available annotations at NCBI/EBI [54]. The average length of genes and the number of encoded amino acids are in the same range as previously observed for other plant species from diverse taxonomic groups [33,55].

It should be noted that the assignment of *D. dumetorum* sequences to the *D. rotundata* pseudochromosomes and indirectly the respective linkage groups contain the risk of incorrect assignments. However, although *D. rotundata* and *D. dumetorum* are evolutionary separated, *D. rotundata* is the most closely related species with genetic and genomic resources.

Our draft genome has the potential to provide the basis for new ways to breed with *D. dumetorum*, for example avoiding the post-harvest hardening phenomenon, which begins within 24 h after harvest and makes it necessary to process the tubers within this time to allow consumption [2]. The family Dioscoreaceae consists of more than 800 species [56] and the post-harvest hardening phenomenon has only been reported from *D. dumetorum* [57], outlining the singularity of this species among yam species. We predicted a large number of genes, which will include putative genes controlling the post-harvest hardening on *D. dumetorum* and many useful bioactive compounds detected in this yam species, which is considered the most nutritious and valuable from a phytomedicine point of view [58]. Ongoing work will try to identify these genes and polymorphisms for making them available for subsequent breeding.

In summary, we present the first de novo nuclear genome sequence assembly of *D. dumetorum* with very good contiguity and partially separated phases. Our assembly has no ambiguous bases and

offers a well applicable gene model annotation. This assembly unraveled the genomic structure of *D. dumetorum* to a large extent and will serve as a reference genome sequence for yam breeding by helping to identify and develop molecular markers associated with relevant agronomic traits, and to understand the evolutionary history of *D. dumetorum* and yam species in general.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4425/11/3/274/s1>, File S1: Sequencing overview with ENA identifiers of runs. File S2: Results of BUSCO analysis of different assembly versions. File S3: AGP file describing contig assignment to *D. rotundata* pseudochromosomes. File S4: Genome size estimation overview using four different tools. File S5: General statistics of different assembly versions. File S6: Comparison of results from different gene prediction approaches. File S7: Orthogroups of predicted peptides of *D. rotundata* and *D. dumetorum*. File S8: Similarity of *D. dumetorum* and *D. rotundata* based on single copy orthologs. File S9: Functional annotation of predicted genes in the *D. dumetorum* genome sequence. File S10: Consensus sequences of repeat elements detected in the *D. dumetorum* genome sequence. File S11: Average short read mapping coverage of predicted genes in the *D. dumetorum* genome sequence. File S12: Results from Smudgeplot analyses.

Author Contributions: C.S., B.P., D.C.A., and B.W. designed the study. C.S. collected the sample. B.P. performed DNA extraction, ONT sequencing, and genome assembly. P.V. performed Illumina sequencing. C.S. and B.P. processed the assembly. B.P. performed gene prediction and evaluation. C.S. and B.P. wrote the initial draft. B.W. and D.C.A. revised the manuscript. All authors read and approved the final version of the manuscript.

Funding: This research was partly funded by the German Academic Exchange Service (DAAD, No. 57299294) with a fellowship to CS.

Acknowledgments: We thank the Appropriate Development for Africa foundation (ADAF) for the yam collection in Cameroon and for permission to study the plants in Germany in the framework of our mutual protocol agreement. We also thank the German Society of Botany (DBG) for supporting the research stay of CS in Bielefeld. We acknowledge support for the Article Processing Charge by the Deutsche Forschungsgemeinschaft and the Open Access Publication Fund of Bielefeld University.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Viruel, J.; Forest, F.; Paun, O.; Chase, M.W.; Devey, D.; Couto, R.S.; Segarra-Moragues, J.G.; Catalan, P.; Wilkin, P. A nuclear Xdh phylogenetic analysis of yams (*Dioscorea* Dioscoreaceae) congruent with plastid trees reveals a new Neotropical lineage. *Bot. J. Linn. Soc.* **2018**, *187*, 232–246. [\[CrossRef\]](#)
2. Sefa-Dedeh, S.; Afoakwa, E.O. Biochemical and textural changes in trifoliate yam *Dioscorea dumetorum* tubers after harvest. *Food Chem.* **2002**, *79*, 27–40. [\[CrossRef\]](#)
3. Alozie, Y.E.; Akpanabiatu, M.; Eyong, E.U.; Umoh, I.B.; Alozie, G. Amino Acid Composition of *Dioscorea dumetorum* Varieties. *Pak. J. Nutr.* **2009**, *8*, 103–105.
4. Ferede, R.; Maziya-Dixon, B.; Alamu, O.E.; Asiedu, R. Identification and quantification of major carotenoids of deep yellow-fleshed yam (tropical *Dioscorea dumetorum*). *J. Food Agric. Environ.* **2010**, *8*, 160–166.
5. Nimenibo-Uadia, R.; Oriakhi, A. Proximate, Mineral and Phytochemical Composition of *Dioscorea dumetorum* Pax. *J. Appl. Sci. Environ. Manag.* **2017**, *21*, 771–774.
6. Iwu, M.M.; Okunji, C.O.; Ohiaeri, G.O.; Akah, P.; Corley, D.; Tempesta, M.S. Hypoglycaemic activity of dioscoretine from tubers of *Dioscorea dumetorum* in normal and alloxan diabetic rabbits. *Plant. Med.* **1990**, *56*, 264–267. [\[CrossRef\]](#)
7. Siadjeu, C.; Panyoo, E.A.; Toukam, G.M.S.; Bell, J.M.; Nono, B.; Medoua, G.N. Influence of Cultivar on the Postharvest Hardening of Trifoliate Yam (*Dioscorea dumetorum*) Tubers. *Adv. Agric.* **2016**, 2658983. [\[CrossRef\]](#)
8. Tamiru, M.; Natsume, S.; Takagi, H.; White, B.; Yaegashi, H.; Shimizu, M.; Yoshida, K.; Uemura, A.; Oikawa, K.; Abe, A.; et al. Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. *BMC Biol.* **2017**, *15*, 86. [\[CrossRef\]](#)
9. Cormier, F.; Lawac, F.; Maledon, E.; Gravillon, M.C.; Nudol, E.; Mournet, P.; Vignes, H.; Chair, H.; Arnau, G. A reference high-density genetic map of greater yam (*Dioscorea alata* L.). *Theor. Appl. Genet.* **2019**, *132*, 1733–1744. [\[CrossRef\]](#)
10. Ngo Ngwe, M.F.; Omokolo, D.N.; Joly, S. Evolution and Phylogenetic Diversity of Yam Species (*Dioscorea* spp.): Implication for Conservation and Agricultural Practices. *PLoS ONE* **2015**, *10*, e0145364. [\[CrossRef\]](#)

11. Magwe-Tindo, J.; Wieringa, J.J.; Sonke, B.; Zapfack, L.; Vigouroux, Y.; Couvreur, T.L.P.; Scarcelli, N. Complete plastome sequences of 14 African yam species (*Dioscorea* spp.). *Mitochondrial DNA Part B-Resour.* **2019**, *4*, 74–76. [CrossRef]
12. Mieg, J. Nombres chromosomiques et répartition géographique de quelques plantes tropicales et équatoriales. *Revue de Cytologie et de Biologie Végétales* **1954**, *15*, 312–348.
13. Hui-Chen, C.; Mei-Chen, C.; Ping-Ping, L.; Chih-Tsun, T.; Fang-Ping, D. A cytotoxic study on Chinese *Dioscorea*, L.—The chromosome numbers and their relation to the origin and evolution of the genus. *J. Syst. Evol.* **1985**, *23*, 11–18.
14. Siadjeu, C.; Mayland-Quellhorst, E.; Albach, D.C. Genetic diversity and population structure of trifoliate yam (*Dioscorea dumetorum* Kunth) in Cameroon revealed by genotyping-by-sequencing (GBS). *BMC Plant Biol.* **2018**, *18*, 359. [CrossRef]
15. Rosso, M.G.; Li, Y.; Strizhov, N.; Reiss, B.; Dekker, K.; Weisshaar, B. An *Arabidopsis thaliana* T-DNA mutagenised population (GABI-Kat) for flanking sequence tag based reverse genetics. *Plant Mol. Biol.* **2003**, *53*, 247–259. [CrossRef]
16. Vurture, G.W.; Sedlazeck, F.J.; Nattestad, M.; Underwood, C.J.; Fang, H.; Gurtowski, J.; Schatz, M.C. GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* **2017**, *33*, 2202–2204. [CrossRef]
17. Sun, H.; Ding, J.; Piednoel, M.; Schneeberger, K. findGSE: Estimating genome size variation within human and *Arabidopsis* using k-mer frequencies. *Bioinformatics* **2018**, *34*, 550–557. [CrossRef]
18. Liu, B.; Shi, Y.; Yuan, J.; Hu, X.; Zhang, H.; Li, N.; Li, Z.; Chen, Y.; Mu, D.; Fan, W. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *arXiv* **2013**, arXiv:1308.2012.
19. Marcais, G.; Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **2011**, *27*, 764–770. [CrossRef]
20. Pucker, B. Mapping-based genome size estimation. *bioRxiv* **2019**. [CrossRef]
21. Ranallo-Benavidez, T.R.; Jaron, K.S.; Schatz, M.C. GenomeScope 2.0 and Smudgeplots: Reference-free profiling of polyploid genomes. *bioRxiv* **2019**, 747568. [CrossRef] [PubMed]
22. Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **2017**, *27*, 722–736. [CrossRef] [PubMed]
23. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [CrossRef]
24. Vaser, R.; Sović, I.; Nagarajan, N.; Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **2017**, *27*, 737–746. [CrossRef] [PubMed]
25. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.
26. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **2014**, *9*, e112963. [CrossRef]
27. Pucker, B.; Holtgräwe, D.; Rosleff Sörensen, T.; Stracke, R.; Viehöver, P.; Weisshaar, B. A De Novo Genome Sequence Assembly of the *Arabidopsis thaliana* Accession Niederzenz-1 Displays Presence/Absence Variation and Strong Synteny. *PLoS ONE* **2016**, *11*, e0164321. [CrossRef]
28. Tang, H.; Zhang, X.; Miao, C.; Zhang, J.; Ming, R.; Schnable, J.C.; Schnable, P.S.; Lyons, E.; Lu, J. ALLMAPS: Robust scaffold ordering based on multiple maps. *Genome Biol.* **2015**, *16*, 3. [CrossRef]
29. Pucker, B.; Holtgräwe, D.; Weisshaar, B. Consideration of non-canonical splice sites improves gene prediction on the *Arabidopsis thaliana* Niederzenz-1 genome sequence. *BMC Res. Notes* **2017**, *10*, 667. [CrossRef]
30. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [CrossRef]
31. Keller, O.; Kollmar, M.; Stanke, M.; Waack, S. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* **2011**, *27*, 757–763. [CrossRef] [PubMed]
32. Smit, A.F.A.; Hubley, R.; Green, P. RepeatMasker Open-4.0. Available online: <http://www.repeatmasker.org> (accessed on 2 December 2018).
33. Pucker, B.; Feng, T.; Brockington, S. Next generation sequencing to investigate genomic diversity in Caryophyllales. *bioRxiv* **2019**. [CrossRef]
34. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]

35. Cheng, C.Y.; Krishnakumar, V.; Chan, A.; Thibaud-Nissen, F.; Schobel, S.; Town, C.D. Araport11: A complete reannotation of the Arabidopsis thaliana reference genome. *Plant J.* **2017**, *89*, 789–804. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Lowe, T.M.; Eddy, S.R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **1997**, *25*, 955–964. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Chan, P.P.; Lowe, T.M. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. In *Gene Prediction: Methods and Protocols*; Kollmar, M., Ed.; Springer: Berlin/Heidelberg, Germany, 2019; Volume 1962, pp. 1–14.
39. Nawrocki, E.P.; Eddy, S.R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **2013**, *29*, 2933–2935. [\[CrossRef\]](#)
40. Kalvari, I.; Argasinska, J.; Quinones-Olvera, N.; Nawrocki, E.P.; Rivas, E.; Eddy, S.R.; Bateman, A.; Finn, R.D.; Petrov, A.I. Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* **2018**, *46*, D335–D342. [\[CrossRef\]](#)
41. Flynn, J.M.; Hubley, R.; Goubert, C.; Rosen, J.; Clark, A.G.; Feschotte, C.; Smit, A.F. RepeatModeler2: Automated genomic discovery of transposable element families. *bioRxiv* **2019**. [\[CrossRef\]](#)
42. Pucker, B.; Ruckert, C.; Stracke, R.; Viehover, P.; Kalinowski, J.; Weisshaar, B. Twenty-Five Years of Propagation in Suspension Cell Culture Results in Substantial Alterations of the Arabidopsis Thaliana Genome. *Genes* **2019**, *10*, 671. [\[CrossRef\]](#)
43. Baasner, J.S.; Howard, D.; Pucker, B. Influence of neighboring small sequence variants on functional impact prediction. *bioRxiv* **2019**. [\[CrossRef\]](#)
44. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*, 238. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Ruggieri, V.; Alexiou, K.G.; Morata, J.; Argyris, J.; Pujol, M.; Yano, R.; Nonaka, S.; Ezura, H.; Latrasse, D.; Boualem, A.; et al. An improved assembly and annotation of the melon (*Cucumis melo* L.) reference genome. *Sci. Rep.* **2018**, *8*, 8088. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Mignouna, H.D.; Abang, M.M.; Asiedu, R. Harnessing modern biotechnology for tropical tuber crop improvement: Yam (*Dioscorea* spp.) molecular breeding. *Afr. J. Biotechnol.* **2003**, *2*, 12.
47. Mignouna, H.D.; Abang, M.M.; Asiedu, R. Genomics of Yams, a Common Source of Food and Medicine in the Tropics. In *Genomics of Tropical Crop Plants*; Moore, P.H., Ming, R., Eds.; Springer: New York, NY, USA, 2008; pp. 549–570. [\[CrossRef\]](#)
48. Michael, T.P.; Jupe, F.; Bemm, F.; Motley, S.T.; Sandoval, J.P.; Lanz, C.; Loudet, O.; Weigel, D.; Ecker, J.R. High contiguity Arabidopsis thaliana genome assembly with a single nanopore flow cell. *Nat. Commun.* **2018**, *9*, 541. [\[CrossRef\]](#)
49. Paajanen, P.; Kettleborough, G.; Lopez-Girona, E.; Giolai, M.; Heavens, D.; Baker, D.; Lister, A.; Cugliandolo, F.; Wilde, G.; Hein, I.; et al. A critical comparison of technologies for a plant genome sequencing project. *Gigascience* **2019**, *8*, giv163. [\[CrossRef\]](#)
50. Salmela, L.; Walve, R.; Rivals, E.; Ukkonen, E. Accurate self-correction of errors in long reads using de Bruijn graphs. *Bioinformatics* **2017**, *33*, 799–806. [\[CrossRef\]](#)
51. Knuth, R. Dioscoreaceae. In *Das Pflanzenreich*; Engelm., A., Ed.; Engelmann, W.: Leipzig, Germany, 1924.
52. Viruel, J.; Segarra-Moragues, J.G.; Raz, L.; Forest, F.; Wilkin, P.; Sanmartin, I.; Catalan, P. Late Cretaceous-Early Eocene origin of yams (Dioscorea, Dioscoreaceae) in the Laurasian Palaeoartic and their subsequent Oligocene-Miocene diversification. *J. Biogeogr.* **2016**, *43*, 750–762. [\[CrossRef\]](#)
53. Wendel, J.F.; Jackson, S.A.; Meyers, B.C.; Wing, R.A. Evolution of plant genome architecture. *Genome Biol.* **2016**, *17*, 37. [\[CrossRef\]](#)
54. Pucker, B.; Brockington, S.F. Genome-wide analyses supported by RNA-Seq reveal non-canonical splice sites in plant genomes. *BMC Genom.* **2018**, *19*, 980. [\[CrossRef\]](#)
55. Pucker, B.; Holtgräwe, D.; Stadermann, K.B.; Frey, K.; Huettel, B.; Reinhardt, R.; Weisshaar, B. A chromosome-level sequence assembly reveals the structure of the Arabidopsis thaliana Nd-1 genome and its gene set. *PLoS ONE* **2019**, *14*, e0216233. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Barton, H. Yams: Origins and Development. *Encyclopaedia Glob. Archaeol.* **2014**, 7943–7947.

57. Treche, S.; Delpeuch, F. Physiologie Vegetale—Mise en evidence de l'apparition d'un epaississement membranaire dans le parenchyme des tubercules de *Dioscorea dumetorum* au cours de la conservation. *Comptes Rendus de l'Académie des Sciences. Série D: Sciences Naturelles* **1979**, *288*, 67–70.
58. Price, E.J.; Wilkin, P.; Sarasan, V.; Fraser, P.D. Metabolite profiling of *Dioscorea* (yam) species reveals underutilised biodiversity and renewable sources for high-value compounds. *Sci. Rep.* **2016**, *6*, 29136. [[CrossRef](#)]



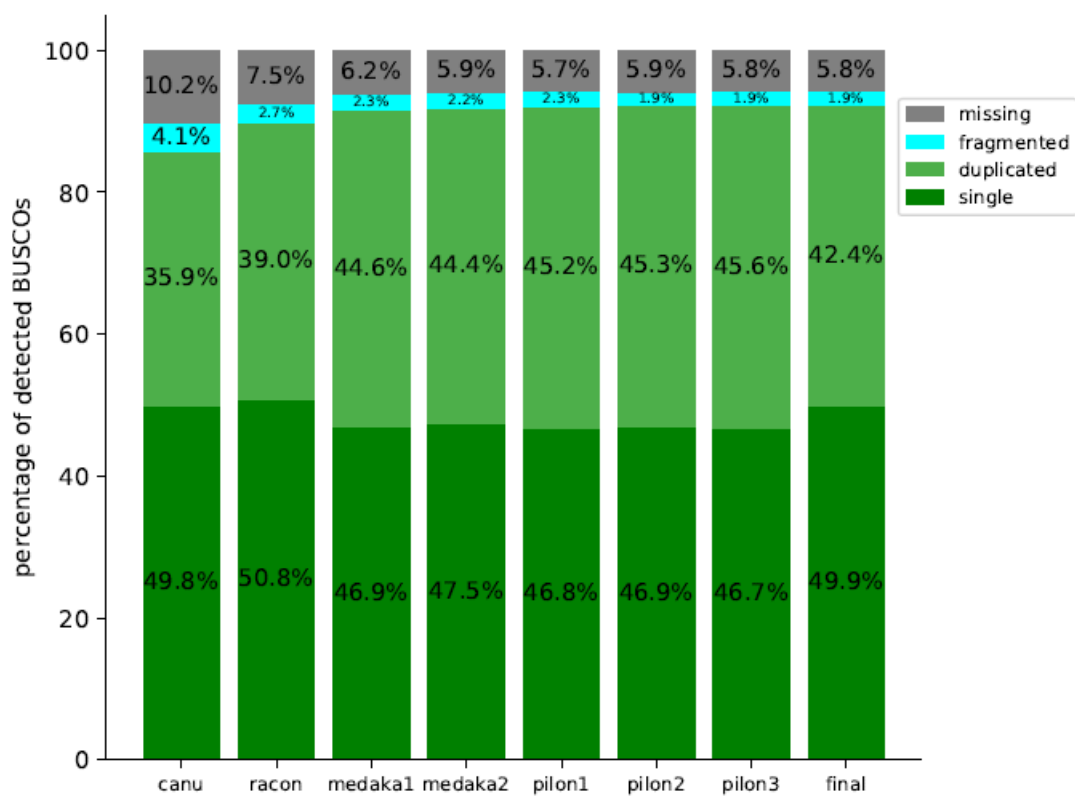
© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Supplementary material. Siadjeu et al. (2020)

File S1: Sequencing overview with ENA identifiers of runs

RunID	Technology	LibraryName
ERR3610915,ERR3863353	ONT R9.4.1	1095cfd4
ERR3610916,ERR3863645	ONT R9.4.1	292eae2c
ERR3611125,ERR3863644	ONT R9.4.1	2d3a938d
ERR3611126,ERR3863643	ONT R9.4.1	31913587
ERR3611127,ERR3866805	ONT R9.4.1	57b36f80
ERR3611128,ERR3863660	ONT R9.4.1	6dddb329
ERR3611129,ERR3863642	ONT R9.4.1	811fe0b8
ERR3611130,ERR3863647	ONT R9.4.1	8f2d9277
ERR3611131,ERR3863648	ONT R9.4.1	94c34238
ERR3611132,ERR3863354	ONT R9.4.1	96f7257c
ERR3611133,ERR3863646	ONT R9.4.1	bd07af06
ERR3611134,ERR3863659	ONT R9.4.1	c06ad413
ERR3611135,ERR3863641	ONT R9.4.1	e4399e7c
ERR3611136,ERR3870375	ONT R9.4.1	e5e2e37f
ERR3611137,ERR3863071	ONT R9.4.1	f7d8039f
ERR3611139	Illumina	
ERR3611140	Illumina	

Supplementary material. Siadjeu et al. (2020)



File S2: Results of BUSCO analysis of different assembly versions

Supplementary material. Siadjeu et al. (2020)

File S3: AGP file describing contig assignment to *D. rotundata* pseudochromosomes (cf. external additional file)

Supplementary material. Siadjeu et al. (2020)

File S4: Genome size estimation overview using four different tools

k-mer	GenomeScope	findGSE	Gce	MGSE
19	251	337	206	-
21	253	351	211	-
23	255	344	215	-
25	257	354	219	-
-	-	-	-	350

Supplementary material. Siadjeu et al. (2020)

File S5: General statistics of different assembly versions

	Initial assembly	Racon1	Medaka1	Medaka2	Pilon1	Pilon2	Pilon3	Final*
Number of contigs	1172	1172	1329	1361	1361	1361	1172	924
Max. contig length (bp)	20 187 448	20 424 333	17 902 291	17 910 017	17 883 451	17 882 417	20 276 997	17 878 854
Assembly size (bp)	501 985 705	508 061 170	508 002 724	507 922 122	507 180 447	507 014 617	503 597 432	485 115 345
GC content	37.74%	37.66%	37.85%	37.88%	37.60%	37.59%	37.58%	37.57%
N50	3 896 882	3 930 287	3 200 042	2 598 889	2 593 818	2 594 536	3 913 348	3 190 870
N90	136 614	138 199	136 918	136 202	136 150	136 144	137 065	156 407
BUSCO	85,70%	89,80%	91,50%	91,90%	92,00%	92,20%	92,30%	92,30%

Supplementary material. Siadjeu et al. (2020)

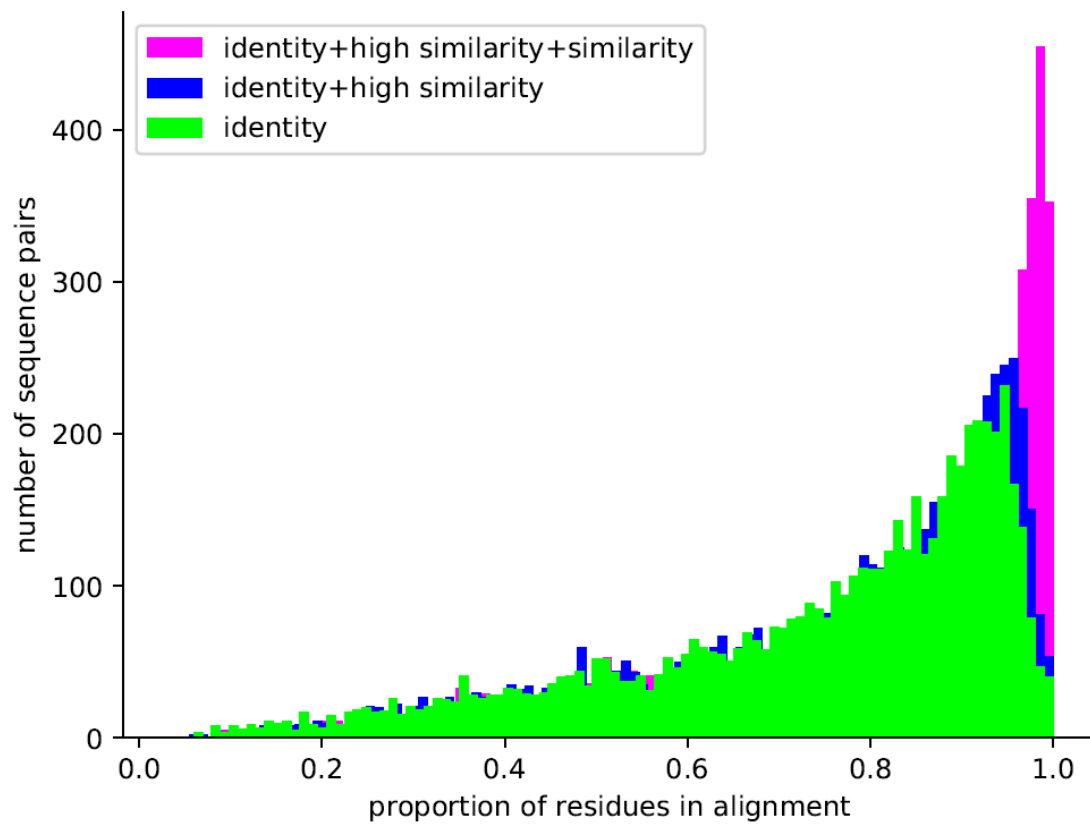
File S6: Comparison of results from different gene prediction approaches

ID	gene_number	gene_length	exons_per_gene	average_mRNA_length	average_peptide_length	BUSCO
final	35269	4282	6	1915	455	C:87.0%[S:47.8%,D:39.2%],F:4.9%,M:8.1%,n:1440
masked_maize	19304	9407	8	2625	703	C:70.2%[S:42.4%,D:27.8%],F:9.7%,M:20.1%,n:1440
masked_rice	29467	5699	6	1964	465	C:67.7%[S:41.7%,D:26.0%],F:11.9%,M:20.4%,n:1440
masked_yam	81826	3411	5	1895	351	C:88.8%[S:49.2%,D:39.6%],F:5.1%,M:6.1%,n:1440
unmasked_maize	25684	9636	8	2781	758	C:73.7%[S:44.0%,D:29.7%],F:9.0%,M:17.3%,n:1440
unmasked_rice	37911	5669	6	2041	494	C:70.7%[S:43.3%,D:27.4%],F:11.6%,M:17.7%,n:1440
unmasked_yam	92335	3540	5	1931	381	C:89.5%[S:48.9%,D:40.6%],F:5.1%,M:5.4%,n:1440

Supplementary material. Siadjeu et al. (2020)

File S7: Orthogroups of predicted peptides of *D. rotundata* and *D. dumetorum* (cf. external additional file)

Supplementary material. Siadjeu et al. (2020)



File S8: Similarity of *D. dumetorum* and *D. rotundata* based on single copy orthologs

Supplementary material. Siadjeu et al. (2020)

File S9: Functional annotation of predicted genes in the *D. dumetorum* genome sequence (cf. external additional file)

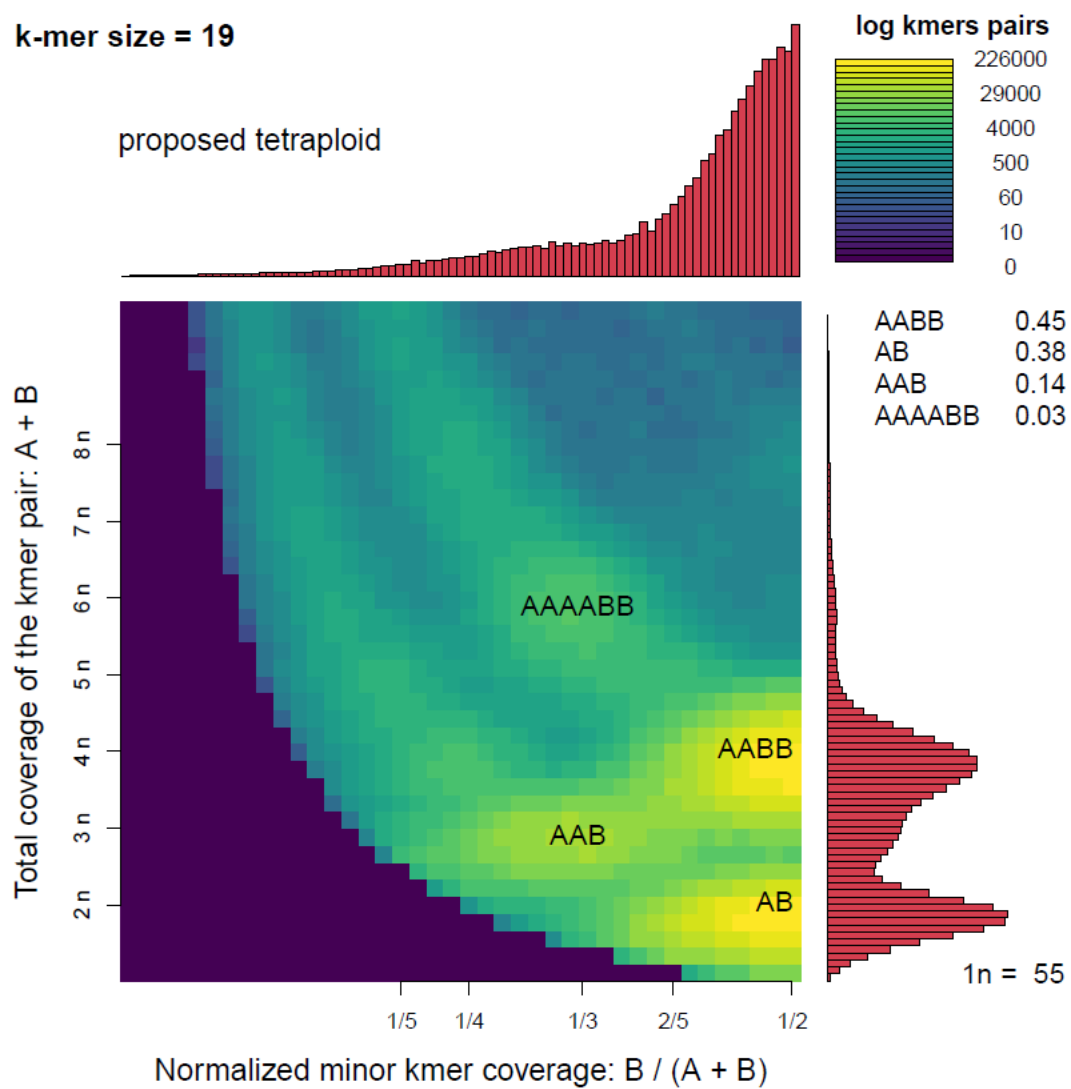
Supplementary material. Siadjeu et al. (2020)

File S10: Consensus sequences of repeat elements detected in the *D. dumetorum* genome sequence (cf. external additional file)

Supplementary material. Siadjeu et al. (2020)

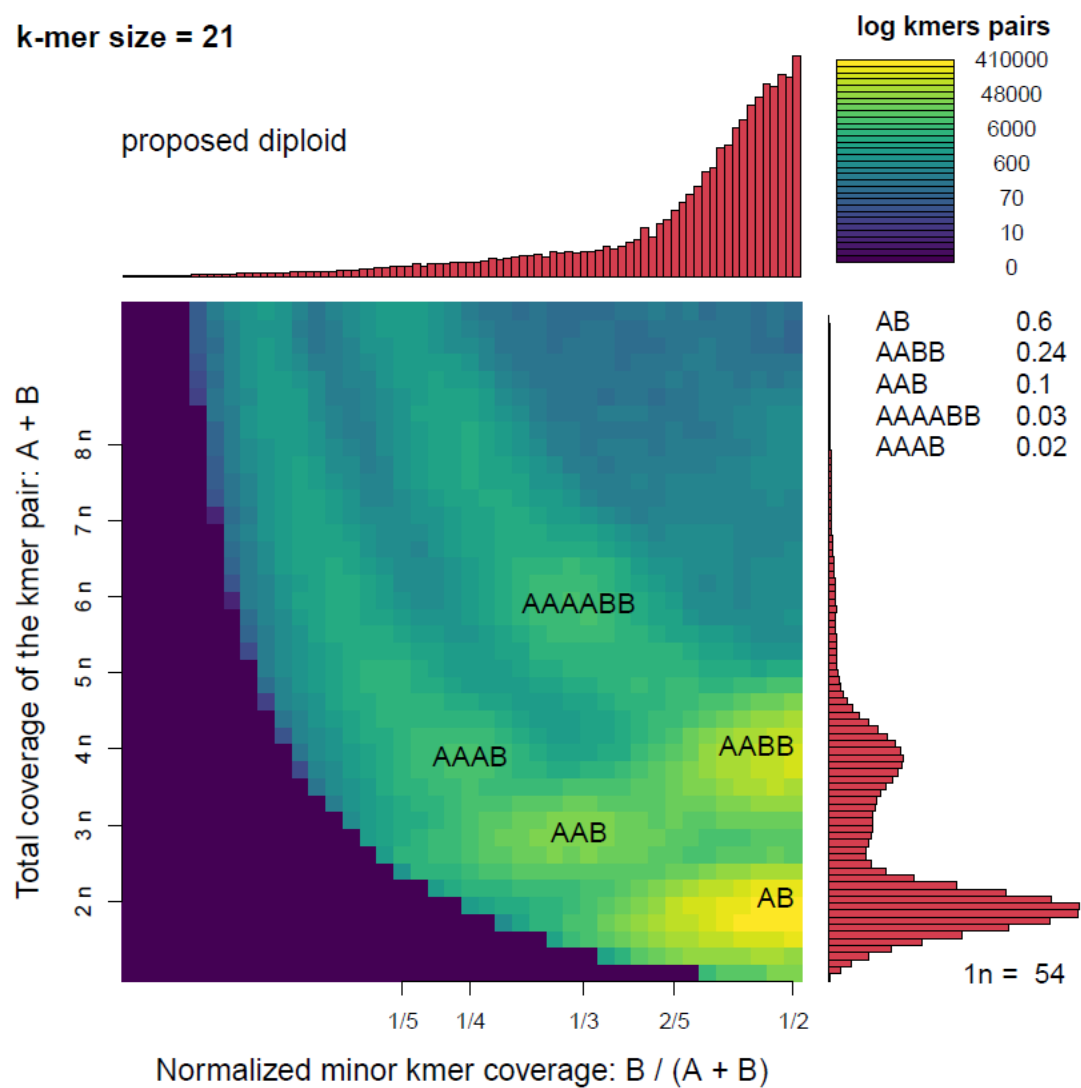
File S11: Average short read mapping coverage of predicted genes in the *D. dumetorum* genome sequence (cf. external additional file)

Supplementary material. Siadjeu et al. (2020)



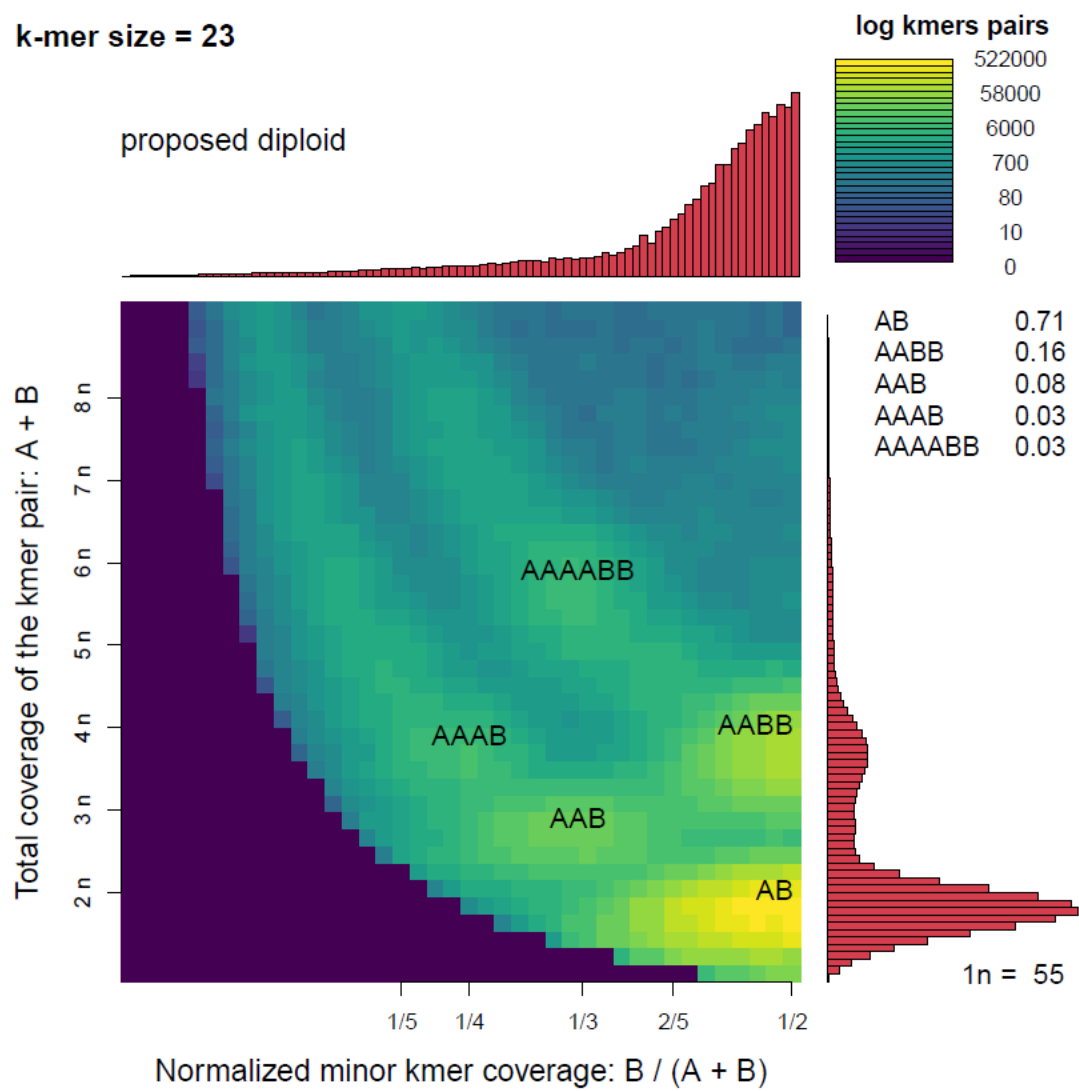
File S12: Results from Smudgeplot analyses

Supplementary material. Siadjeu et al. (2020)



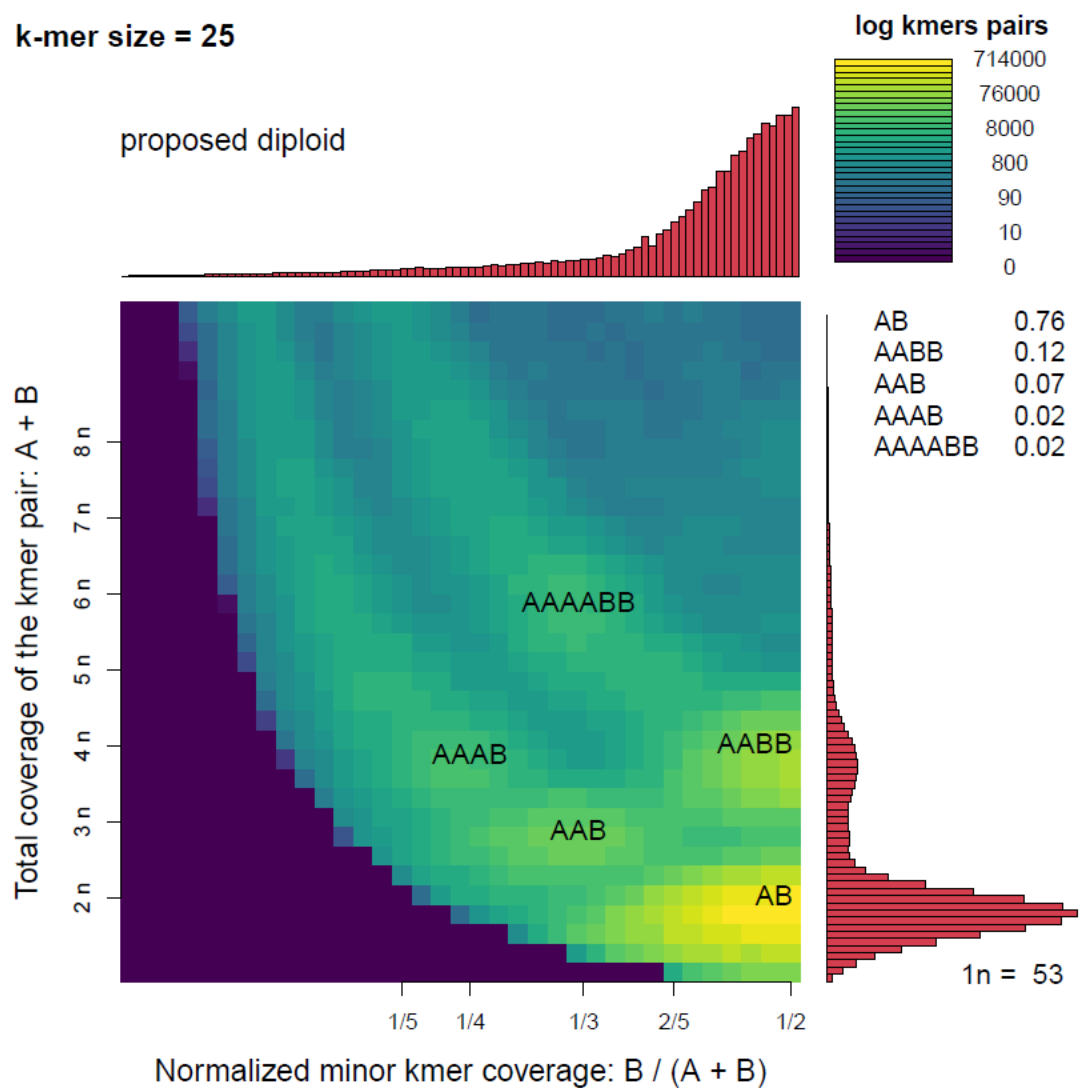
File S12: Results from Smudgeplot analyses continued

Supplementary material. Siadjeu et al. (2020)



File S12: Results from Smudgeplot analyses continued

Supplementary material. Siadjeu et al. (2020)



File S12: Results from Smudgeplot analyses continued

ARTICLE III:

TRANSCRIPTOME SEQUENCING REVEALS CANDIDATE GENES CONTROLLING THE POST-HARVEST HARDENING OF TRIFOLIATE YAM *DIOSCOREA DUMETORUM*

Transcriptome sequencing reveals candidate genes controlling the post-harvest hardening of trifoliate yam *Dioscorea dumetorum*

Christian Siadjeu^{1*}, Eike Mayland-Quellhorst¹, Sascha Laubinger, Dirk C. Albach¹

¹Institute for Biology and Environmental Sciences, Carl-von-Ossietzky University Oldenburg, Carl-von-Ossietzky Str. 9-11, 26111 Oldenburg, Germany

* Corresponding author: christian.siadjeu@uol.de

Abstract

Storage ability of trifoliate yam (*D. dumetorum*) is restricted by a severe post-harvest hardening (PHH) phenomenon, which starts within the first 24h after harvest and renders tubers inedible. Previous work has only focused on the biochemical changes affecting PHH in *D. dumetorum*. To the best of our knowledge, the candidate genes responsible for hardening of *D. dumetorum* have not been identified. Here, transcriptome analyses of *D. dumetorum* tubers were performed in yam tubers of four developmental stages: 4 months after emergence (4MAE), immediately after harvest (AH), 3 days after harvest (3DAH) and 14 days after harvest (14DAH) of four accessions (Bangou 1, Bayangam 2, Fonkouankem 1, and Ibo sweet 3) using RNA-Seq. In total, between AH and 3DAH, 165, 199, 128 and 61 differentially expressed genes (DEGs) were detected in Bayangam 2, Fonkouankem 1, Bangou 1 and Ibo sweet 3, respectively. Functional analysis of DEGs revealed that genes encoding for cellulose synthase A (CESA), xylan O-acetyltransferase (XOAT), chlorophyll a/b binding protein 1, 2, 3, 4 (LHCB1, LHCB, LHCB3, and LCH4) and a MYB transcription factor were predominantly and significantly up-regulated 3DAH. A hypothetical mechanism of this phenomenon and its regulation has been proposed. These findings provide the first comprehensive insights into gene expression in yam tubers after harvest and valuable information for molecular breeding against the PHH.

1. Introduction

Yams constitute an important food crop for over 300 million people in the humid and subhumid tropics. Among the eight yam species commonly grown and consumed in West and Central Africa, trifoliate yam (*Dioscorea dumetorum*) is the most nutritious [1]. Tubers of *D. dumetorum* are rich in protein (9.6%), well balanced in essential amino acids (chemical score of 0.94) and its starch is easily digestible [2-3]. *Dioscorea dumetorum* is not only used for human alimentation but also for pharmaceutical purposes. A bio-active compound, dioscoretine, has been identified in *D. Dumetorum* [4], which has been accepted pharmaceutically and which can be used advantageously as a hypoglycemic agent in situations of acute stress. The tubers are, therefore, commonly used in treating diabetes in Nigeria [5].

Despite of these qualities, the storage ability of this yam species is restricted by severe post-harvest hardening (PHH) of the tubers, which begins within 24 h after harvest and renders them unsuitable for human consumption [1]. The PHH of *D. dumetorum* is separated into a reversible component associated with the decrease of phytate and an irreversible component associated with the increase of total phenols [6]. The mechanism of PHH is supposed to start with enzymatic hydrolyzation of phytate and subsequent migration of the released divalent cations to the cell wall where they cross-react with demethoxylated pectins in the middle lamella. This starts the lignification process in which the aromatic compounds accumulate on the surface of the cellular wall reacting as precursors for the lignification [7].

Whereas physiological changes associated with hardening of yam tubers are now reasonably well understood, knowledge is still limited in terms of ways to overcome the hardening. Naturally, occurring genotypes lacking PHH have been identified [8], which offers a chance to understand the genetic basis of hardening. Therefore, the next step is to understand the genetic background of this genotype and its relationship to other genotypes, which has been conducted using GBS (Illumina-based genotyping-by-sequencing [9]. Further insights have been gained by sequencing and analyzing the genome of the non-hardening genotype Ibo sweet 3 [10].

Here, we analyze the transcriptome of the non-hardening accession Ibo sweet 3 and three hardening accessions to identify genes involved in the PHH phenomenon. The study of the transcriptome examines the presence of mRNAs in a given cell population and usually includes some information on the concentration of each RNA molecule, as a factor of the number of reads sequenced, in addition to the molecular identities. Unlike the genome, which is roughly fixed for a given cell line when neglecting mutations, the transcriptome varies from organ to organ, during development and based on external environmental conditions. In

particular, transcriptome analysis by RNA-seq enables identification of genes that have differential expression in response to environmental changes or developmental stage and mapping genomic diversity in non-model organisms [11]. Differential gene expression analysis under different conditions has, therefore, allowed an increased insight into the responses of plants to external and internal factors and into the regulation of different biological processes. High-throughput sequencing technologies allow an almost exhaustive survey of the transcriptome, even in species with no available genome sequence [12]. Indeed, transcriptome analysis based on high-throughput sequencing technology has been applied to investigate gene expression of hardening in carrot [13]. In yam, it helped elucidate flavonoid biosynthesis regulation of *D. alata* tubers [14].

A lack of availability of next generation ‘-omics’ resources and information had hindered application of molecular breeding in yam [15], which has recently been overcome by the publication of two genome sequences in the genus [10-16]. Here, we report the first transcriptomic study of *D. dumetorum* and the first to evaluate the influence of genes on the PHH phenomenon in a monocot tuber using transcriptomics. We aim to close this gap by identifying candidate genes involved in the PHH phenomenon of *D. dumetorum* to facilitate breeding non-hardening accessions of *D. dumetorum*.

2. Results

2.1. Descriptive Statistics of RNA-Seq Data

Transcriptome sequences of four *D. dumetorum* accessions (Bangou 1, Bayangam 2, Fonkouankem 1 and Ibo sweet 3) were analysed during four tuber developmental stages: 4 months after emergence (4MAE), 9 months after emergence (harvest time, AH), 3 days after harvest (3DAH) and 14 days after harvest (14DAH) to determine putative genes involved in the PHH phenomenon. After trimming, 943,323,048 paired-end raw reads (150-bp in length) were generated for 48 samples (Supplementary S1). Among these 242.7, 224.6, 233.9 and 242.1 million reads belonged to Bangou 1, Bayangam 2, Fonkouankem 1 and Ibo sweet 3. On average, 90% of all clean reads were aligned to the *D. dumeotrum* reference genome v1.0. Furthermore, an average of 56 % of those reads were uniquely mapped to the reference genome sequence. A PCA plot showed the normalized read counts of all samples (Figure 1). The first two principal components (PCs) explained 69% of the variability among samples. Samples four months after emergence were separated from those AH, 3DAH and 14DAH. No clear separation was observed between AH, 3DAH, and 14DAH. However, taking into account accession specificity, AH was separated from 3DAH and 14DAH. This finding indicated a difference in

transcriptome expressions of accessions before and after harvest. One biological replicate of each accession at a specific time point did not cluster with others likely due to individual variability between plants.

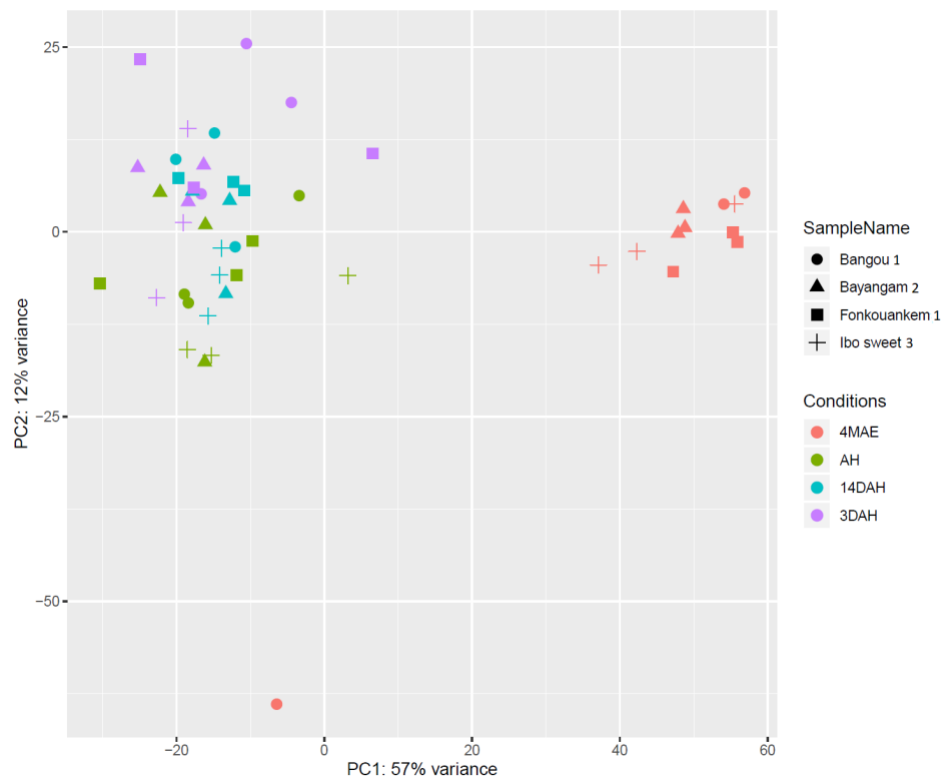


Figure 1. PCA plot of normalized counts using the variant stabilizing transformation (VST). Four *D. dumetorum* accessions (symbols) at four different sampling points (colors) are shown.

2.2. Differential Expression Analysis

Two well established statistical analysis methods (edgeR and DESeq2) to assess differentially expressed genes (DEGs) based on read counts were employed. We used two strategies to determine DEGs in *D. dumetorum* after harvest: STAR_DESeq2, and STAR_edgeR. The design model for DE analysis was \sim Accession + Condition + Accession:Condition. We carried out multiple comparison at the accession, condition and interaction accession*conditions levels. The STAR_DESeq2 approach yielded the highest number of DEGs (Figure 2) and the results were selected for downstream analysis (Supplementary S2). Pairwise comparisons (4MAE vs AH, 3DAH vs AH, 14DAH vs AH, 14DAH vs 3DAH) of gene expressions among the four accessions were performed (Figure 2).

However, since the PHH in *D. dumetorum* tubers occurs after harvest, we focused on gene expressions after harvest. A decrease of up-regulated DEGs and an increase of down-regulated DEGs were noticed among the three accessions that do harden from harvest to 14DAH (Figure 2). The accession that does not harden depicted a different pattern. Comparing 3DAH vs AH, significantly DEGs were detected in Bayangam 2 (165 DEGs), Fonkouankem 1 (199 DEGs), Bangou 1 (128 DEGs) and Ibo sweet 3 (61 DEGs). Amongst these, 120, 112, 83 and 16 were up-regulated in Bayangam 2, Fonkouankem 1, Bangou 1 and Ibo sweet 3, respectively. For 14DAH vs AH, significantly DEGs were obtained in Bayangam 2 (162 DEGs), Bangou 1 (201 DEGs), Fonkouankem 1 (161 DEGs) and Ibo sweet 3 (46 DEGs). Among these, 126, 83, 47, and 13 were up-regulated DEGs in Bayangam 2, Bangou 1, Fonkouankem 1 and Ibo sweet 3, respectively. In total, the highest number of significantly up-regulated DEGs were detected in Bayangam 2 and the lowest in Ibo sweet 3. A mixture analysis of the three accessions that do harden irrespective of accession was performed (Supplementary S3). Pairwise comparisons of gene expression among the three stages or conditions (AH, 3DAH and 14DAH) in the combined analysis of the three hardening accessions detected 59, 40 and 13 up-regulated DEGs between 3DAH vs AH, 14DAH vs AH and 14DAH vs 3DAH, respectively (Supplementary S3, Supplementary S4). Whereas, 14 (3DAH vs AH), 36 (14DAH vs AH), and 56 (14DAH vs 3DAH) were down-regulated.

In order to understand the difference between Ibo sweet 3 (the non-hardening accession) and the other accessions, a multiple pairwise comparison (Bayangam 2 vs Ibo sweet 3, Bangou 1 vs Ibo sweet 3, Fonkouankem 1 vs Ibo sweet 3) after harvest (3DAH vs AH, 14DAH vs AH) was carried out (Figure 3, Supplementary S5). After harvest to 3DAH (3DAH vs AH), significantly DEGs were acquired comparing Bayangam 2 vs Ibo sweet 3 (111 DEGs), Fonkouankem 1 vs Ibo sweet 3 (111 DEGs) and Bangou 1 vs Ibo sweet 3 (80 DEGs). Amongst these, 101, 80 and 62 were up-regulated DEGs in Bayangam 2 vs. Ibo sweet 3, Fonkouankem 1 vs Ibo sweet 3 and Bangou 1 vs Ibo sweet 3, respectively. For 14DAH vs AH, 88, 85 and 91 significantly DEGs were detected comparing Bayangam 2 vs Ibo sweet 3 (88 DEGs), Fonkouankem 1 vs Ibo sweet 3 (85 DEGs) and Bangou 1 vs Ibo sweet 3 (91 DEGs). Among these, 80, 30 and 22 were up-regulated in Bayangam 2 vs Ibo sweet 3, Fonkouankem 1 vs Ibo sweet 3 and Bangou 1 vs Ibo sweet 3, respectively.

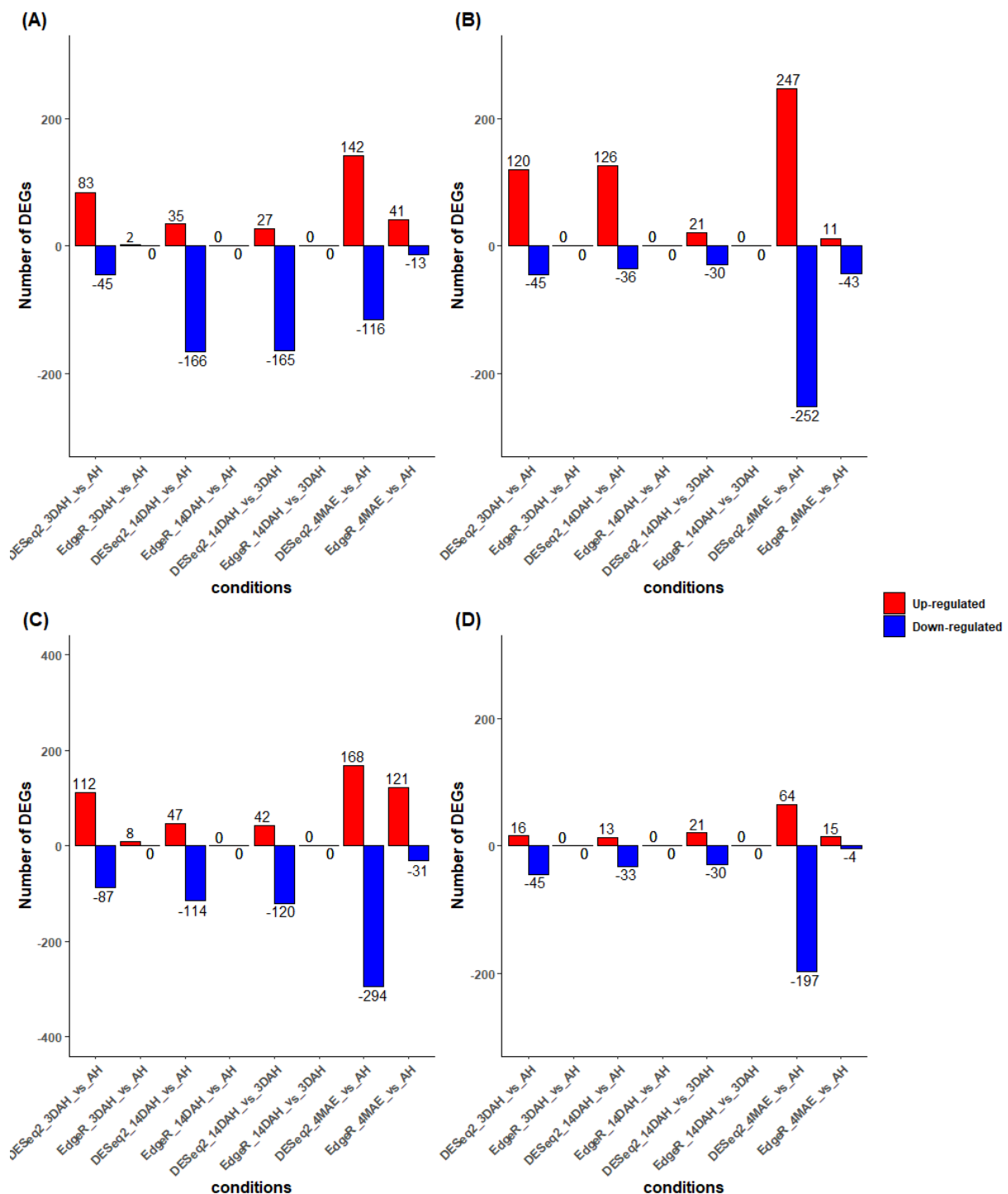


Figure 2. The number of DEGs based on the comparison of DESeq2 and EdgeR 4MAE and after harvest (AH, 3DAH, and 14DAH). (A) Bangou 1, (B) Bayangam 2, (C) Fonkouankem 1, (D) Ibo sweet 3 (non-hardening accession). Blue represents down-regulated transcripts and red represents up-regulated transcripts.

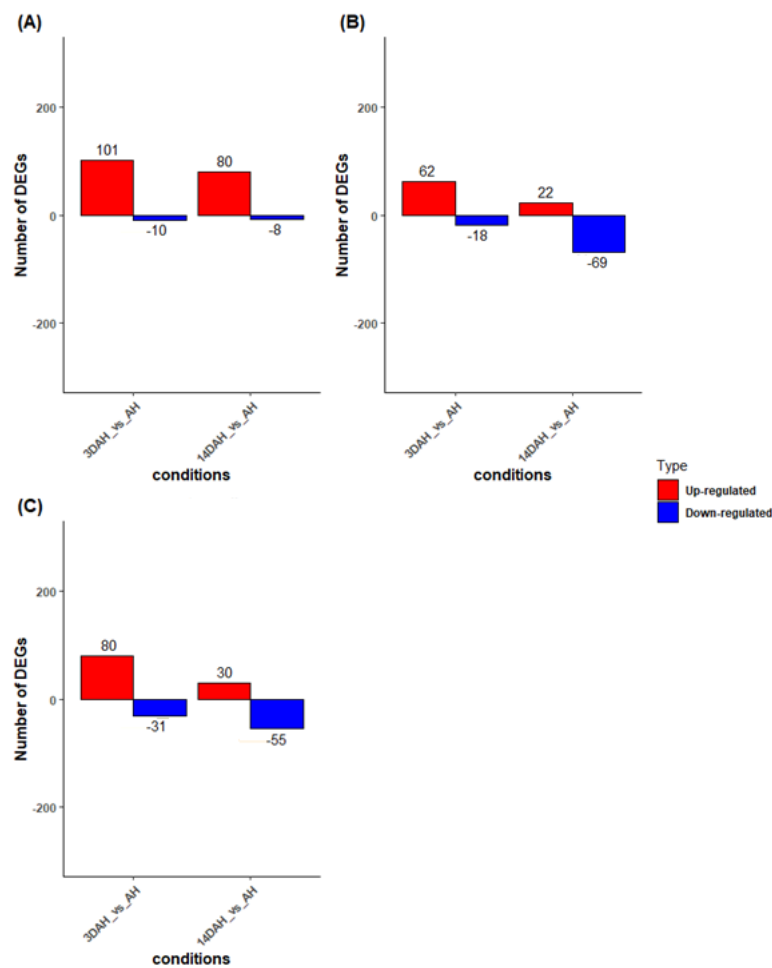


Figure 3. The number of DEGs based on the comparison between Ibo sweet 3 and other accessions after harvest (AH, 3DAH, and 14DAH). (A) Ibo sweet 3 vs Bayangam 2, (B) Ibo sweet 3 vs Bangou 1, (C) Ibo sweet 3 vs Fonkouankem 1. Blue represents down-regulated transcripts and red represents up-regulated transcripts.

2.3. GO Enrichment and Functional Classification of DEGs with KEGG and Mapman

For better comprehension of the PHH phenomenon, gene ontology (GO) term annotation and enrichment were performed on up-regulated DEGs resulting from pairwise comparisons (3DAH vs AH, 14DAH vs AH) of all three accessions that do harden (Figure 4 A). Compared with 3DAH and AH, out of the 59 up-regulated DEGs, 38 were significantly annotated with 43 GO terms, most of which were involved in biological processes related to cellular processes, response to stimuli, and metabolic processes. Likewise, for 14 DAH vs AH, 23 up-regulated genes (out of 40) were significantly enriched regarding biological processes in

relation to cellular processes, response to stimuli, and metabolic processes (Figure 4 B). GO term analysis of each hardening accession separately revealed that cellular processes, metabolic processes, response to stimuli and response to stress were in the top 10 of the most common enriched GO terms 3DAH and 14DAH (Figure 4 C, D).

Pathway-based analysis with Kyoto encyclopedia of genes and genomes (KEGG) revealed that metabolic pathway (Ko01100) was the most enriched with seven and six up-regulated transcripts followed by biosynthesis of secondary metabolites pathway (Ko01110) with three and one up-regulated transcripts 3DAH and 14DAH, respectively (Figure 5 A, B). Based on MapMan, photosynthesis (Bin 1, 23 genes) and RNA biosynthesis pathways (Bin 15, eight genes) were the most enriched 3DAH. Likewise, 14DAH, photosynthesis (six genes) and RNA biosynthesis (six genes) were the most enriched pathways (Figure 5 A, B).

Analysis of each hardening accession separately showed a similar pattern for KEGG and MapMan pathway-based annotations (Figure 5 C, D). Metabolic pathway was the most enriched followed by biosynthesis of secondary metabolites pathway 3DAH and 14 DAH. Fourteen (in Bayangam 2), twelve (in Fonkouankem 1), and nine (in Bangou 1) up-regulated transcripts related to metabolic pathway were recovered 3DAH. Conversely, fifteen, five, and three up-regulated transcripts were identified in Bayangam 2, Fonkouankem 1, and Bangou 1 respectively 14DAH. The MapMan pathway enrichment revealed that photosynthesis, RNA biosynthesis and cell wall organization pathways were in the top 5 of the most common enriched pathway across the hardening accessions 3DAH. However, protein homeostasis, phytohormone action, and protein modification were among the top 5 of the most common enriched pathways 14DAH. These results highlight that the PHH likely occurs predominantly in *D. dumetorum* 3DAH.

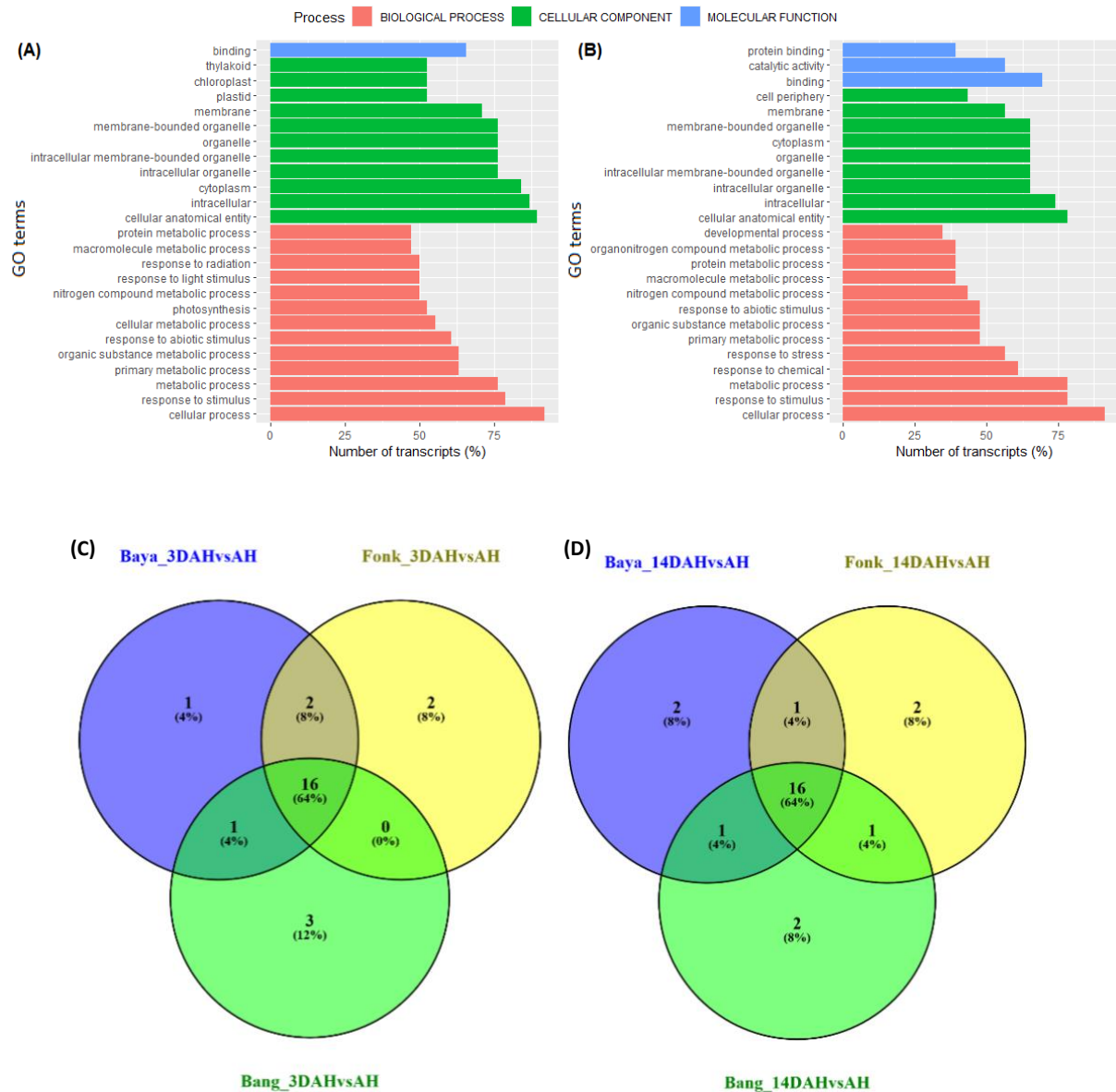


Figure 4. Functional annotation of the top up-regulated enriched GO terms of *D. dumetorum* tubers after harvest (AH, 3DAH, 14DAH). (A) and (B) the top 25 up-regulated enriched GO terms of the combined analysis of three hardening accessions 3DAH and 14DAH, respectively. (C) and (D) Venn diagram of the enrichment of the top 20 GO terms of each hardened accessions 3DAH and 14DAH, respectively. Blue bar represents molecular process, green bar represents cellular component, and red bar represents biological process.

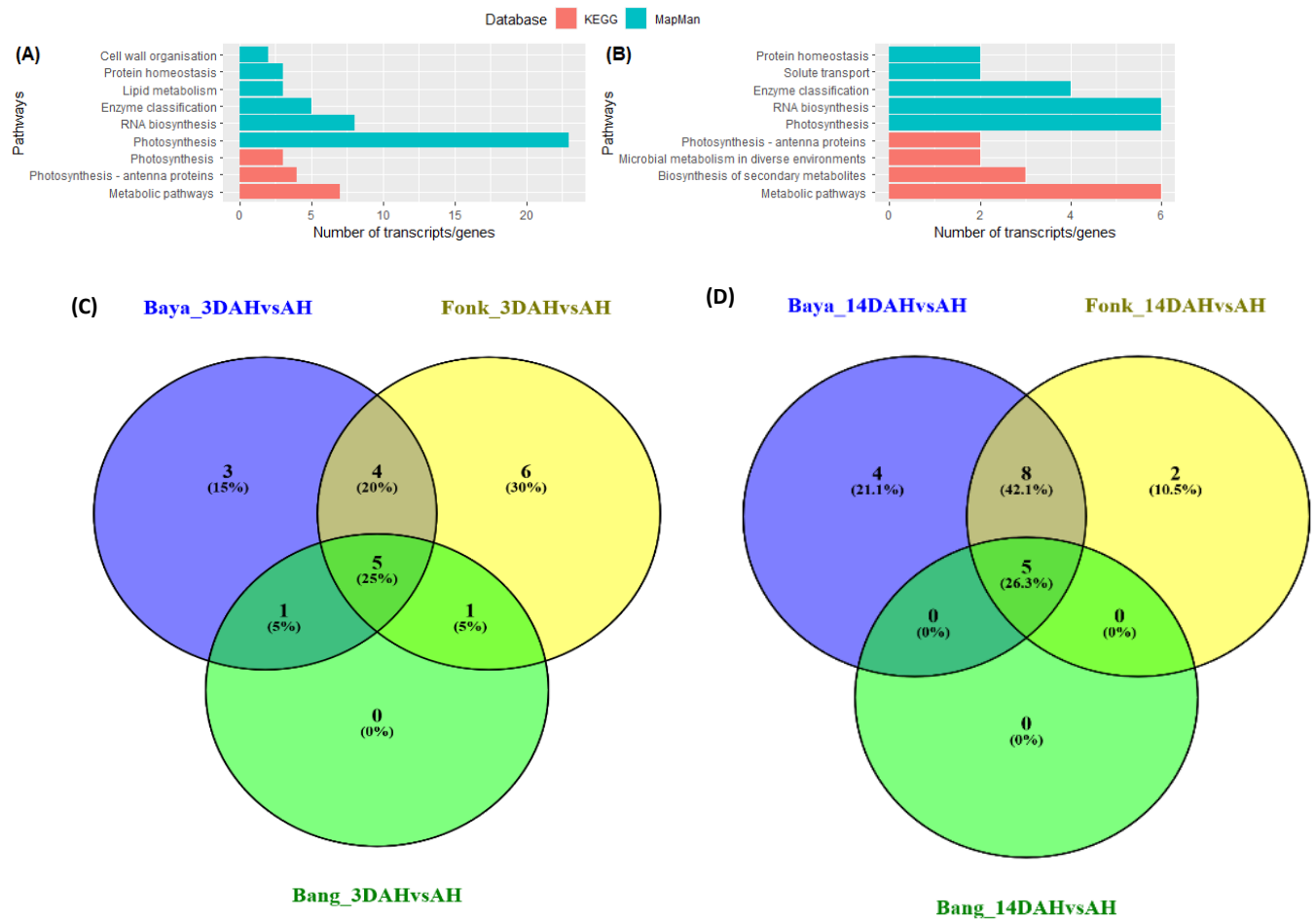


Figure 5. Functional classification of DEG after harvest. (A) and (B) the most enriched pathways of the combined analysis of three hardening accessions 3DAH and 14DAH, respectively. (C) and (D) Venn diagrams the most enriched pathways of each hardening accession 3DAH and 14DAH respectively. Green bars represent pathway annotation with MapMan database, and red bars represent pathway annotation with KEGG database.

2.4. Cluster Expression Analysis

Based on the differential expression analysis and functional annotation results, the PHH likely occurs 3DAH. Thus, differentially expressed up-regulated genes 3DAH in the hardening accessions and in the combined analysis of the three hardening accessions together were selected for cluster expression analysis. Their expressions during four tuber development stages were plotted (Figure 6). This analysis aims to identify groups of genes with similar expression patterns in all hardening accessions in relation to PHH. Two groups or clusters were identified amongst up-regulated DEGs 3DAH (Figure 6). The first pattern depicted a high peak 4MAE and then decreased AH and slightly increased 3DAH and 14DAH with an expression under

zero except for the accession Fonkouankem 1. This first pattern corresponds to cluster 1 in Bangou 1 and Fonkouankem 1 and cluster 2 in Bayangam 2 and in the mixture of the three hardening accessions (Figure 6 A, B, C, D). Conversely, for the second pattern, the expression was down 4MAE and AH, and sharply increased 3DAH and then decreased 14DAH. This second pattern corresponds to cluster 2 in Bangou 1 and Fonkouankem 1 and cluster 1 in Bayangam 2 and the mixture of the three hardening accessions. This second pattern, showing the highest peak 3DAH could be the group of genes that are co-regulated and involved in the PHH. Therefore, functional annotation of genes of clusters showing the second pattern were further investigated.

The top 3 accumulated pathways in the cluster 2 were photosynthesis (20 contigs) followed by solute transport (two contigs) and cell wall organization (one contig) in Bangou 1 (cluster 2) (Supplementary S6). For Bayangam 2 (cluster 1), the top 3 pathways were protein modification (eight contigs) followed by RNA biosynthesis (seven contigs) and phytohormone action (seven contigs). However, it is worth to outline that cell wall organization (four contigs) and secondary metabolism (three contigs) pathways were as well accumulated. On the contrary, in Fonkouankem 1 (cluster 2), cell wall organization (19 contigs) was the most enriched pathway followed by RNA biosynthesis (eight contigs) and photosynthesis, secondary metabolism, protein homeostasis, cytoskeleton organization and solute transport pathways with four contigs each of them. The mixture of the three hardening accessions (cluster 1) showed that photosynthesis was the most accumulated pathway (21 contigs) followed by protein homeostasis, lipid metabolism pathways with three contigs each of them and cell wall organization pathway with two contigs. In sum, genes encoding for photosynthesis, cell wall organization, protein modification and RNA biosynthesis, and secondary metabolism pathways are co-up-regulated after harvest and likely involved in the PHH on *D. dumetorum* tubers.

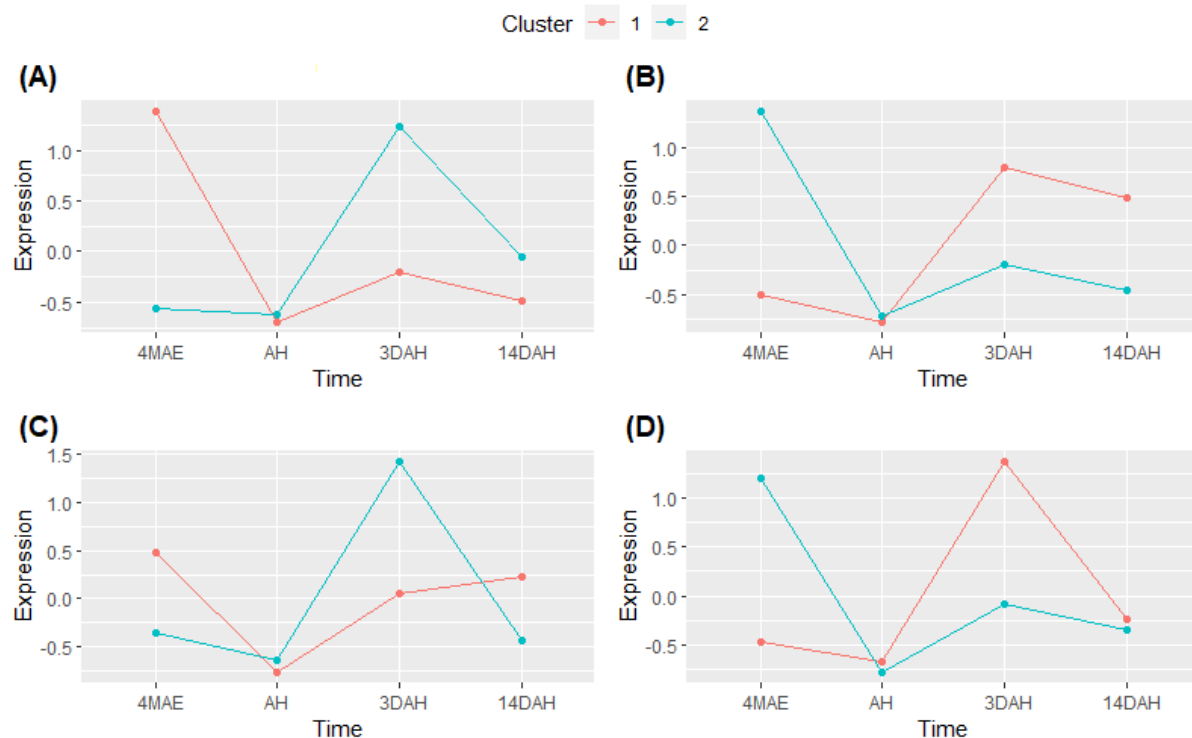


Figure 6. Cluster analysis of DEGs 3DAH among the different sampling time 4MAE and after harvest. (A) Bangou 1, (B) Bayangam 2, (C) Fonkouankem 1, (D) combined analysis of the three hardening accessions.

2.5. Comprehensive Analysis of Expression of Genes Potentially Involved in PHH

We opted for investigation of genes differentially expressed 3DAH in the accession Fonkouankem 1 (cluster 2) due to its high amount of up-regulated genes associated with cell organization and the combined analysis of the three hardening accessions together (cluster 1). In the mixture of the three hardening accessions, (cluster 1), a total of 20 transcripts homologous to the genes encoding for photosynthesis were observed as up-regulated differentially expressed three 3DAH, when all hardening accessions were analyzed together (Table 1), including chlorophyll a/b binding proteins LHCB1 (8 transcripts), LHCA4 (two transcripts) LHCB2 (two transcripts), photosystem II protein PSBX (two transcripts). Those genes response to light stimulus and may be the triggers of this phenomenon. Three transcripts associated with cell wall organization were found encoding for fasciclin-type arabinogalactan protein, COB cellulose and glucan endo-1,3-beta-glucosidase. They are likely involved in the reinforcement of the cell wall (hardening). One transcript homologous to the gene related to myoloblastosis (MYB) transcription factors (TFs) was included in this group. However, it is important to note that genes involved in lipid metabolism, namely lipase (three transcripts) were found in this group.

In Fonkouankem 1 (cluster 2) (Table 2), 18 up-regulated genes encoding for cell wall organization including xylan O-acetyltransferase (XOAT) (five transcripts), cellulose synthase (CESA) (three transcripts), corncob cellulose (COB) (two transcripts) were found in cluster 2. The MYB transcription factor was the most abundant (four transcripts) followed by TFs DREB and NAC with 2 transcripts each of them. Photosynthesis genes LHCB1, and LHCA4 were found with two transcripts each of them. However, genes encoding for phenolic metabolism were enriched with two genes cinnamate 4-hydroxylase (two transcripts) and phenylalanine ammonia lyase (two transcripts). Likewise, lipase (three transcripts) was recorded in this group.

In all hardening accessions and the combined analysis of the three hardening accessions together, annotation with several MYB database identified putative MYB genes (MYB54, MYB52, MYB73, MYB70, MYB44, MYB77, MYB46, MYB83, MYB9, MYB107, MYB93, MYB53, MYB92) associated with cell wall modifications (Supplementary S7).

Table 1. Candidate genes associated with PHH in *D. dumetorum* tuber in the combined analysis of the hardening accessions 3DAH vs AH.

Contig	LF2C	padj	Bin/KO	Gene\Name	Description
contig544.g2040	6.91	0.04740	21.4.1.1.3	FLA	fasciclin-type AGP
contig278.g50	8.89	0.02720	21.1.2.2	COB	regulatory protein
contig760.g29	18.35	0.03609	1.2.3/K05298	GAPA	glyceraldehyde-3-phosphate dehydrogenase
contig119.g125	8.07	0.00170	1.1.6.1.1	PGR5/PGRL1	complex.component PGR5-like
contig678.g379	7.72	0.00000	1.1.4.1.4/K08910	LHCA4	chlorophyll a/b binding protein 4
contig679.g24	7.98	0.00000	1.1.4.1.4/K08910	LHCA4	chlorophyll a/b binding protein 4
contig549.g218	6.53	0.00000	K02694	psaF	photosystem I subunit III
contig222.g1555	5.27	0.00626	1.1.4.2.8/K02695	psaH	photosystem I subunit VI
contig206.g10	5.55	0.00042	1.1.1.1.1/K08912	LHCB1	chlorophyll a/b binding protein 1
contig206.g11	7.98	0.00000	1.1.1.1.1/K08912	LHCB1	chlorophyll a/b binding protein 1
contig206.g6	7.12	0.00000	1.1.1.1.1/K08912	LHCB1	chlorophyll a/b binding protein 1
contig206.g8	7.58	0.00000	1.1.1.1.1/K08912	LHCB1	chlorophyll a/b binding protein 1
contig267.g402	5.81	0.00836	1.1.1.1.1/K08913	LHCB2	chlorophyll a/b binding protein 2
contig355.g38	5.82	0.01516	1.1.1.1.1/K08913	LHCB2	chlorophyll a/b binding protein 2
contig391.g20	6.24	0.00012	1.1.1.1.1/K08912	LHCB1	chlorophyll a/b binding protein 1
contig391.g26	6.94	0.00000	1.1.1.1.1/K08912	LHCB1	chlorophyll a/b binding protein 1
contig391.g28	5.72	0.00038	1.1.1.1.1/K08912	LHCB1	chlorophyll a/b binding protein 1
contig391.g29	7.65	0.00000	1.1.1.1.1/K08912	LHCB1	chlorophyll a/b binding protein 1
contig553.g402	4.31	0.04740	1.1.1.1.1/K08914	LHCB3	chlorophyll a/b binding protein 3
contig565.g52	7.56	0.02366	1.1.1.1.1/K08912	LHCB1	chlorophyll a/b binding protein 1
contig89.g1873	5.94	0.01452	1.1.1.6.2.1	ELIP	LHC-related protein group.protein
contig544.g1881	5.17	0.04740	1.1.1.2.13	1.1.1.2.13/PsbX	PS-II complex.component
contig544.g1970	5.05	0.00905	1.1.1.2.13	1.1.1.2.13/PsbX	PS-II complex.component
contig267.g494	20.89	0.00000	15.5.2.1/K09422	MYB	transcription factor

Table 2. Candidate genes associated with PHH in *D. dumetorum* tuber in Fonkouankem 1 3DAH vs AH.

Contig	LF2C	padj	Bin/Ko	GeneName	Description
contig557.g748	9.02	3.15E-09	21.1.1.1.1/K10999	CESA	cellulose synthase A
contig60.g53	8.86	3.44E-09	21.1.1.1.1/K10999	CESA	cellulose synthase A
contig73.g5	8.94	3.78E-09	21.1.1.1.1/K10999	CESA	cellulose synthase A
contig267.g188	23.39	5.99E-06	21.1.2.2	COB	regulatory protein
contig278.g50	14.51	2.99E-03	21.1.2.2	COB	regulatory protein
contig143.g88	17.83	2.27E-03	21.2.2.2.2	XOAT	xylan O-acetyltransferase
contig145.g17	17.90	2.01E-03	21.2.2.2.2	XOAT	xylan O-acetyltransferase
contig199.g1435	12.17	1.46E-04	21.2.2.2.2	XOAT	xylan O-acetyltransferase
contig920.g250	17.89	1.76E-04	21.2.2.2.2	XOAT	xylan O-acetyltransferase
contig922.g12	11.49	7.50E-03	21.2.2.2.2	XOAT	xylan O-acetyltransferase
contig750.g97	6.45	8.83E-04	21.6.1.7/K13066	COMT	caffeic acid 3-O-methyltransferase
contig646.g19	5.52	1.68E-02	K18368	CSE	caffeoylshikimate esterase
contig552.g180	5.18	1.60E-02	K00588	E2.1.1.104	caffeoyl-CoA O-methyltransferase
contig3.g487	5.66	4.55E-02	21.6.1.2/K09754	CYP98A	5-O-(4-coumaroyl)-D-quinic acid 3'-monooxygenase
contig199.g1672	10.14	3.21E-03	21.6.2.2/K05909	E1.10.3.2	Laccase
contig559.g139	26.23	4.72E-08	21.6.2.1	PMT	p-coumaroyl-CoA
contig119.g106	14.35	9.30E-03	K05350	bgIB	beta-glucosidase
contig390.g181	6.08	3.53E-02	21.3.2.2.2	BGAL	beta-galactosidase
contig678.g379	7.74	2.36E-04	1.1.4.1.4/K08910	LHCA4	chlorophyll a/b binding protein 4
contig679.g24	11.17	4.28E-06	1.1.4.1.4/K08910	LHCA4	chlorophyll a/b binding protein 4
contig206.g11	7.52	4.19E-07	1.1.1.1.1/K08912	LHCB1	chlorophyll a/b binding protein 1
contig391.g29	6.83	5.77E-04	1.1.1.1.1/K08912	LHCB1	chlorophyll a/b binding protein 1
contig546.g79	20.36	6.88E-04	15.5.7.2	DREB	transcription factor
contig771.g2	25.08	4.05E-05	15.5.7.2	DREB	transcription factor
contig267.g494	25.57	3.54E-02	15.5.2.1/K09422	MYB	transcription factor
contig678.g290	16.94	1.44E-02	K09422	MYB	transcription factor
contig693.g10	6.77	4.76E-02	K09422	MYB	transcription factor
contig750.g121	25.14	2.61E-07	K09422	MYB	transcription factor
contig158.g23	37.78	5.01E-06	15.5.17	NAC	transcription factor
contig556.g459	37.78	5.01E-06	15.5.17	NAC	transcription factor

2.6. Comprehensive Difference between Harden and Non-Harden Accessions

Pairwise comparisons of accessions that do harden to the accession that does not harden in different stages after harvest showed that up-regulated genes were enriched mostly in cellular process, cellular anatomical entity and intracellular terms 3DAH and 14DAH (Figure 7, Supplementary S8). Besides, KEGG enrichment revealed that metabolic pathways were the most enriched with ten, eight and five up-regulated genes 3DAH for Bayangam 2 vs Ibo, Fonkouankem vs Ibo and Bangou 1 vs Ibo, respectively (Figure 7 A, B, C). This pathway was followed by biosynthesis of secondary metabolites pathway with six, five, and five up-regulated

genes for Bayangam 2 vs Ibo, Fonkouankem 1 vs Ibo sweet 3 and Bangou 1 vs Ibo sweet 3 respectively. Those pathways were the most enriched as well 14DAH (Supplementary S9). MapMan annotation showed that cell wall organization pathway was predominantly enriched when comparing Bangou 1 to Ibo sweet 3 and Fonkouankem 1 vs Ibo sweet 3 3DAH. Whereas, protein modification pathway was particularly enriched for Bayangam 2 vs Ibo sweet 3. However, cell organization, protein modification and RNA biosynthesis pathways were in the top 7 of the most enriched pathways 3DAH. On the contrary, protein modification pathway was the most enriched irrespective of the comparison 14DAH (Supplementary S9). Venn diagram of the annotation revealed five common up-regulated genes potentially involved in the hardening process among the accessions that do harden comparing to the non-hardening accession Ibo sweet 3. Those genes encoding for chalcone synthase, diterpene synthase, a MYB transcription factor, XOAT, lignin laccase (Figure 7 D).

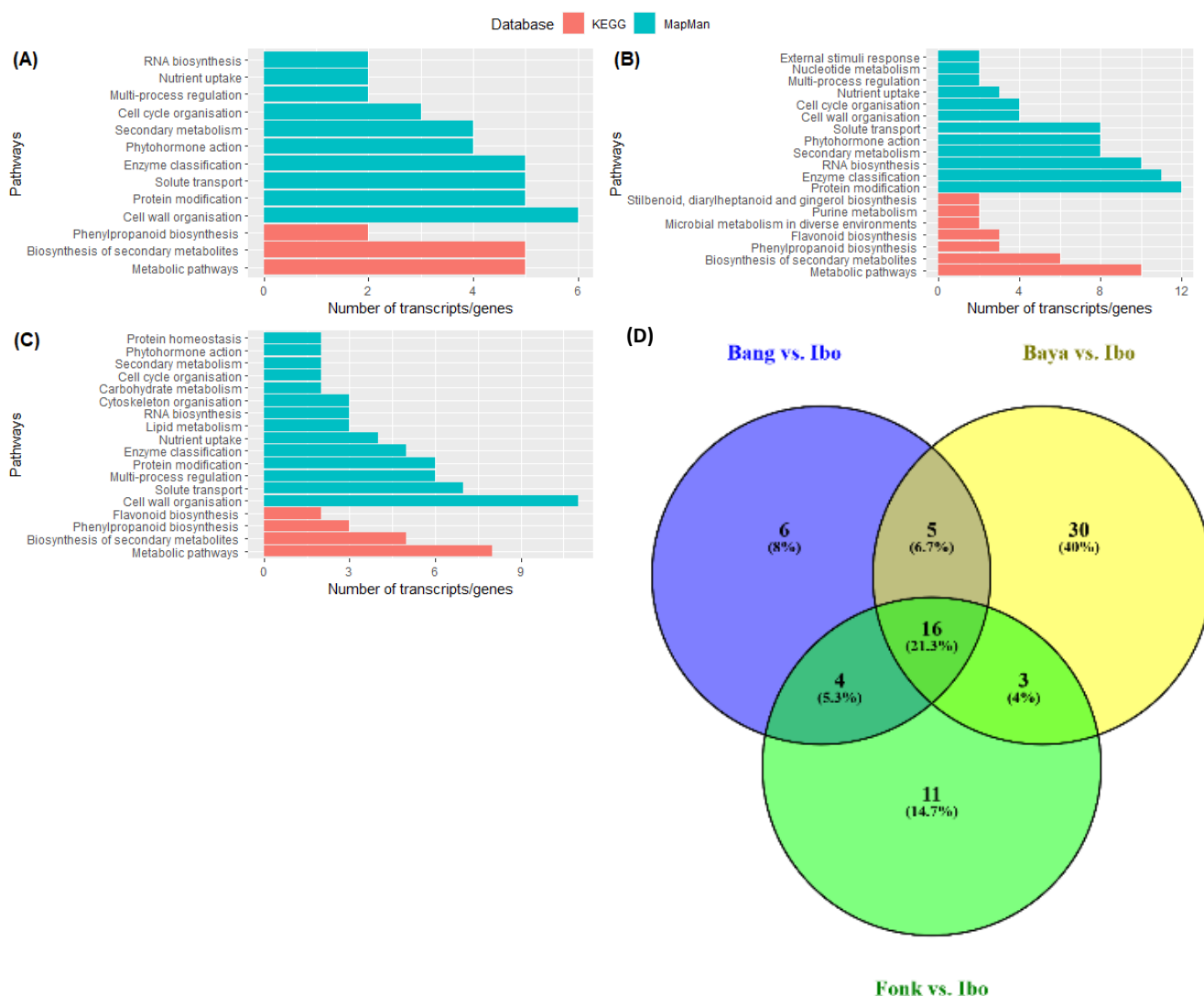


Figure 7. Functional classification of up-regulated DEG 3DAH based on the comparison of hardened accessions against the non-hardening accessions. (A), (B) and (C) the most enriched

pathways 3DAH on Bangou 1 vs Ibo sweet 3, Bayangam 2 vs Ibo sweet 3 and Fonkouankem 1 vs Ibo sweet 3, respectively. (D) Venn diagram of the most enriched pathways on Bangou 1 vs Ibo sweet 3, Bayangam 2 vs Ibo sweet 3 and Fonkouankem 1 vs Ibo sweet 3. Green represents pathway annotation with MapMan database, and red represents pathway annotation with KEGG database.

3. Discussion

The PHH of *D. dumetorum* tubers has been extensively studied in terms of the biochemical and physical aspects [1-17]. Based on our study we reported genes that were differentially expressed and up-regulated after harvest. This demonstrates that the PHH on *D. dumetorum* tuber is likely to be genetically regulated. Our results demonstrate that the number of up-regulated genes was abundant 3DAH and then decreased 14DAH. This suggest that PHH predominantly occurs in the first days after harvest. This is consistent with previous studies [1,8,18] showing a substantial increase of the hardness in the first three DAH.

Functional analysis via KEGG enrichment revealed that most genes differentially expressed were involved in pathways of secondary metabolites. These genes are involved in photosynthesis, RNA biosynthesis (transcription factors), and cell wall organization. In order to understand how this phenomenon occurs, GO enrichment revealed that many genes were involved in cellular processes, response to stimuli, and metabolic processes, as well as response to stress. These results firmly suggest that PHH in *D. dumetorum* is a cellular and metabolic process in response to stimuli leading to stress.

It has been reported that PHH in *D. dumetorum* is associated with an increase in sugar and structural polysaccharides (cellulose, hemicellulose, and lignin) [1]. Later, Medoua et al.[18] associated it with a decrease of phytate and total phenols. However, these authors failed to address causes of this phenomenon. Cellular processes are triggered by stimulus, an investigation of genes related to response to stimuli revealed that photosynthetic genes LHCB1,2,3 and LCH4 were up-regulated 3DAH. Those genes are light-harvesting chlorophyll a/b binding antenna responsible for photon capture. This suggests that *D. dumetorum* tubers are starting photosynthesis after harvest. In the field, *D. dumetorum* tubers turn green under the yam skin (on the surface) when they are exposed to sun light (Supplementary S10). Unlike potatoes, greening occurs only in the field but not in storage. This highlights the importance of water in this process. After harvest, tubers are exposed to the external environment with no possibility of water absorption. This likely leads to drought stress as supported by GO term analysis. In fact, a rapid decrease of water in tubers after harvest was reported [1-18], probably mainly due to photosynthetic activity of *D. dumetorum* tubers. Thus, the PHH of *D. dumetorum* tubers appears to be a mechanism to limit water loss.

Mechanisms to limit water loss in plants has been extensively associated with the reinforcement of the cell wall [19]. Medoua et al.[18] reported a decrease of water absorption by tubers after harvest suggesting that the cell wall permeability decreases during storage. Genes related to cell wall organisation XOAT, CESA, COB were predominantly up-regulated after harvest. This confirms biochemical changes associated with the PHH of *D. dumetorum* tubers [1-18]. An increase in various cell wall polysaccharide such as cellulose, hemicellulose and lignin during storage has been reported [1-18]. Cellulose synthase encodes for cellulose biosynthesis [20] and COB regulate the orientation of cellulose microfibrils, whereas, XOAT encodes for hemicellulose (xylan) biosynthesis [21]. These cell wall polysaccharides play an important role as a protective barrier in response to various environmental perturbations. Accumulation and deposition of these polysaccharides inside primary cell walls reinforces the strength and rigidity of the cell wall and are probably a key component of the plant response to environment factors [19]. It suggests that cellulose and lignin are key cell wall polymers responsible for cell wall rigidification during PHH in *D. dumetorum*.

Many biological processes are controlled by the regulation of gene expression at the level of transcription. Transcription factors are key players in controlling cellular processes. Among those TFs, the MYB family is large and involved in controlling diverse processes such as responses to abiotic and biotic stresses [22]. Our results demonstrated that TF from the MYB family was predominantly expressed and up-regulated after harvest. This result suggests that transcription factors from MYB family may be potentially involved in the mechanism of PHH. Guo et al.[23] demonstrated the role of an MYB TF family in response to water stress from stem of a birch tree through lignin deposition, secondary cell wall thickness and the expression of genes in secondary cell wall formation.

Pairwise comparison of the hardening accessions and the non-hardening accession confirmed that the PHH phenomenon is a cellular and metabolic process leading to the cell wall modification. However, it is interesting to note that protein modifications seem to occur predominantly after hardening from 3 to 14DAH. This could explain the poor sensory qualities of hardened tubers such as coarseness in the mouth [24]. Five common genes were found up-regulated in the hardening accessions and down-regulated in Ibo sweet 3 3DAH. Those genes are chalcone synthase, diterpene synthase, transcription factor MYB, xylan O-acetyltransferase and lignin laccase. Chalcone synthase is a key enzyme of the flavonoids/isoflavonoid biosynthesis pathway and is induced in plants under stress conditions [25]. Laccase catalyse the oxidation of phenolic substrates using oxygen as electron acceptor. Laccase has been recognized in the lignification process through the oxidation of lignin precursors. Indeed, Arcuri

et al[26] demonstrated an involvement of laccase genes in lignification as response to adaptation to abiotic stresses in *Eucalyptus*.

Based on our results, the PHH seems to be governed by differentially expressed genes in a metabolic network, which is attributed to the exposure to external environment or sun light. Therefore, a putative model of the hardening mechanism and the regulatory network associated was proposed (Figure 8). After harvest, yam tubers are exposed to the external environment particularly to sun light. This environmental factor acts as the first signal to stimulate the expression of photosynthetic genes involved in photon capture namely LHCB1, LHCB2, LHCB3 and LHCA4. The absorption of photons implies loss of electrons which is replaced by electrons from the splitting of water through photolysis [27]. This activity implies the necessity of a continued electron supply through the breakdown of water molecules. However, tubers are detached from roots with no possibility of water absorption. Therefore, a signal is given to reinforce the cell wall to avoid water loss from the tubers via the up-regulation of CESA, XOAT and COB genes. This reinforcement of the cell wall implies firstly, an accumulation of cell wall polysaccharide such as cellulose hemicellulose during the first days of storage. Secondly, probably from the third day after harvest, lignification process controlling laccase genes begins. This overall process is might to be controlled by a MYB TF.

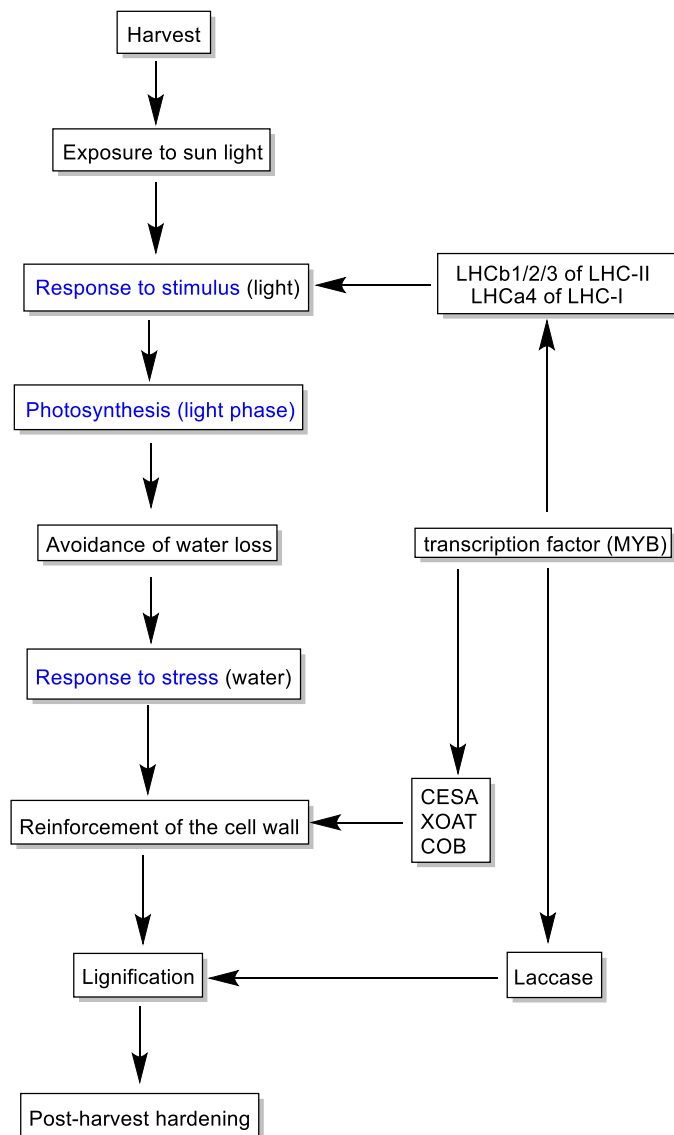


Figure 8. Putative mechanism of the PHH in *D. dumetorum*. Blue represents GO annotation.

4. Materials and Methods

4.1. Plant materials

Four accessions have been collected from various localities in the main growing regions of yam (West and South-West) in Cameroon and one from Nigeria based on the analysis of [9]. These accessions were characterized by many roots (Bangou 1, Bayangam 2, Fonkouankem 1) and few roots (Ibo sweet 3) on the tuber surface with yellow flesh color. Among these accessions, tubers of Ibo sweet 3 does not harden after harvest. Ten tubers of each accession were planted in pots in the greenhouse of the botanic garden of the University of Oldenburg under controlled conditions at 25 °C. They are available upon request.

4.2. Sample Preparation

Three tubers of each accession were randomly collected 4 months after emergence (4MAE), 9 months after emergence (Harvest time AH), 3 days after harvest (3DAH) and 14 DAH. After harvest, tubers were stored under shelter in greenhouse at 25 °C and light intensity of 1.76 $\mu\text{mol s}^{-1}\text{m}^{-2}$. Collected tubers were washed and their skin peeled off. Then, the samples were immediately frozen in liquid nitrogen and stored at – 80 °C prior to RNA isolation.

4.3. RNA-Seq Extraction

The stored tubers (– 80 °C) were lyophilized prior to further handling. Total RNA was extracted from 48 samples [(Four accessions (Bangou 1, Bayangam 2, Fonkouankem 1, Ibo sweet 3) x four stages (4MAE, AH, 3DAH, and 14DAH) x three biological replicates)] using innuPREP Plant RNA Kit (Analytik Jena AG, Germany). The RNA quality was analysed using a spectrophotometer (Nano-Drop Technologies, Wilmington, DE, USA). RNA Integrity Number (RIN) values were determined using a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA) to ensure all samples had a RNA integrity number (RIN) above 6.

4.4. Library Construction and Illumina Sequencing

We constructed cDNA libraries comprising 48 RNA samples using the Universal Plus mRNA-Seq offered by NuQuant (Tecan Genomics, Inc California, USA). Paired-end (2×150 bp) sequencing of the cDNA libraries was performed on an Illumina NovaSeq (Genewiz Germany GmbH, Leipzig).

4.5. Data Processing and Functional Analysis

Low quality reads were filtered using TrimGalore v 0.6.5 (<https://github.com/FelixKrueger/TrimGalore/releases>) with the following parameters --length 36 -q 5 --stringency 1 –e 0.1. The filtered reads were aligned to the reference genome of *D. Dumetorum* [10] with STAR v 2.7.3a [28] with default parameters. The aligned reads in BAM files were sorted and indexed using SAMtools v 1.9 [29]. The number of reads that can be assigned uniquely to genomic features were counted using the function SummarizeOverlaps of the R package GenomicAlignments v1.20.1 [30] with mode="Union", singleEnd=FALSE, ignore.strand=TRUE, fragments=TRUE as parameters.

Two programs DESeq2 [31] and edgeR [32] were deployed to analyze differentially expressed genes (DEGs) between conditions and the interaction conditions x accessions. Gene with p-adjusted value < 0.05 and log2 fold change > 2 were considered as significantly expressed genes. False discovery rate FDR threshold was < 0.05. We performed a basic time

course experiment to assess genes that change their expression after harvest using Deseq2 [31]. Metabolic pathway assignments of DEGs were based on the KEGG Orthology database using the KAAS system [33]. The final pathway analyses were mostly based on the tool Mercator4 and Mapman4 [34]. In addition, differentially expressed MYB genes were functionally annotated based on several datasets *Arabidopsis thaliana* MYBs [35], *Beta vulgaris* MYBs [36], *Musa acuminata* MYBs [37], *Croton tiglium* MYBs [38], *Dioscorea rotundata* MYBs and *Dioscorea dumetorum* MYBs via KIPes (<https://github.com/bpucker/KIPes>). GO term assignment and enrichment were performed using Blast2GO [39] via OmicsBox with cutoff 55, Go weight 5, e-value 1.e-6, HSP-hit coverage cutoff 80 and hit filter 500.

4.5. Cluster differential expressed genes analysis

Co-expression analysis was carried out to identify up-regulated genes with similar expression patterns using The k-means cluster method. Genes that are clustered together across different time points and conditions may be involved in the same regulatory processes [40]. The number of clusters was determined through the sum of squared error and the average silhouette width as described in this script (<https://2-bitbio.com/2017/10/clustering-rnaseq-data-using-k-means.html>).

5. Conclusions

This is the first study that investigates differentially expressed genes after harvest and during yam storage through RNA-Seq. The evidence from this study suggests that PHH in *D. dumetorum* is a cellular and metabolic process involving a combined action of several genes as a response to environmental stress from sun and water. Genes encoding for cell wall polysaccharide constituents were found significantly up-regulated suggesting that they are directly responsible for the hardening of *D. dumetorum* tubers. It is worth noticing that many genes encoding for light-harvesting chlorophyll a/b binding proteins were significantly up-regulated after harvest as well. This supports the idea that sunlight is the trigger element of the PHH manifested by the strengthening of cell walls in order to avoid water loss, which is useful for a putative photosynthetic activity. These findings add substantially to our understanding of hardening in *D. dumetorum* and provide the framework for molecular breeding against PHH in *D. dumetorum*.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Supplementary S1: Statistic of clean reads mapped to *D. dumetorum* reference genome, Supplementary S2: Results of DEG analysis per accession based on the pairwise comparisons (4MAE vs, AH, 3DAH vs AH, 14DAH vs AH, 14DAH vs 3DAH) resulting from the approach

STAR_DESeq2, Supplementary S3: Number of DEGs based on the combined analysis of the three hardening accessions 4MAE and after harvest, Supplementary S4: Results of DEG analysis based on the combined analysis of the 3 hardening accessions 4MAE and after harvest, Supplementary S5: Results of DEG analysis based on the comparison between Ibo sweet 3 and other accessions after harvest, Supplementary S6: Group resulting from Cluster analysis of DEGs 3DAH among the different sampling times for Bangou 1, Bayangam 2, Fonkouankem 1, and the combined analysis of the 3 hardening accessions, Supplementary S7: Phylogenetic tree of candidate MYB genes in Bangou 1, Bayangam 2, Fonkouankem 1, and the combined analysis of the three hardening accessions, Supplementary S8: GO enrichment of up-regulated DEG 3DAH and 14DAH based on the comparison of hardening accessions against the non-hardening accession, Supplementary S9: Functional classification of up-regulated DEG 14DAH based on the comparison of hardening accessions against the non-hardening accession. (A), (B) and (C) the most enriched pathways 14DAH in Bangou 1 vs Ibo sweet 3, Bayangam 2 vs Ibo sweet 3, and Fonkouankem 1 vs Ibo sweet 3, respectively. Green bars represent pathway annotation with the MapMan database, and red bars represent pathway annotation with the KEGG database, Supplementary S10: Greening of young *D. dumetorum* tuber exposed to sunlight as opposed to the non-greening one.

Author Contributions: Conceptualization, C.S. and D.C.A.; methodology, C.S., E.M., and S.L.; software, C.S.; validation, C.S.; formal analysis, C.S.; investigation, C.S., E.M.; resources, C.S.; data curation, C.S.; writing—original draft preparation, C.S.; writing—review and editing, D.C.A., S.L.; visualization, C.S.; supervision, D.C.A., S.L.; project administration, D.C.A., E.M.; funding acquisition, DCA., C.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Alexander von Humboldt-Stiftung, grant number 1128007-NGA-IP and by Deutscher Akademischer Austauschdienst DAAD, grant number 57299294.

Data Availability Statement: Data are available upon request.

Acknowledgments: We would like to thank Dr. Boas Pucker for helping in the annotation with KIPes.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sefa-Dedeh, S.; Afoakwa, E.O. Biochemical and textural changes in trifoliate yam *Dioscorea dumetorum* tubers after harvest. *Food Chem.* 2002, 79, 27–40, doi:10.1016/S0308-8146(02)00172-3.
2. Mbome Lape, I.; Treche, S. Nutritional quality of yam (*Dioscorea dumetorum* and *D. rotundata*) flours for growing rats. *J. Sci. Food Agric.* 1994, 66, 447–455, doi:10.1002/jsfa.2740660405.
3. Afoakwa, E.O.; Sefa-Dedeh, S. Chemical composition and quality changes occurring in *Dioscorea dumetorum* Pax tubers after harvest. *Food Chem.* 2001, 75, 85–91, doi:10.1016/S0308-8146(01)00191-1.
4. Iwu, M.M.; Okunji, C.; Akah, P.; Corley, D.; Tempesta, S. Hypoglycaemic Activity of Dioscoretine from Tubers of *Dioscorea dumetorum* in Normal and Alloxan Diabetic Rabbits. *Planta Med.* 1990, 56, 264–267.

5. Nimenibo-Uadia, R.; Oriakhi, A. Proximate, Mineral and Phytochemical Composition of *Dioscorea dumetorum* Pax. *J. Appl. Sci. Environ. Manag.* 2017, 21, 771, doi:10.4314/jasem.v21i4.18.
6. Medoua, G.N.; Mbofung, C.M.F. Hard-to-cook defect in trifoliate yam *Dioscorea dumetorum* tubers after harvest. *Food Res. Int.* 2006, 39, 513–518, doi:10.1016/j.foodres.2005.10.005.
7. Medoua, G.N. Potentiels nutritionnel et technologique des tubercules durcis de l'igname *Dioscorea dumetorum* (Kunth), Doctoral thesis, Ngaoundere University, Cameroon, 2005.
8. Siadjeu, C.; Panyoo, E.A.; Mahbou Somo Toukam, G.; Bell, M.; Nono, B.; Medoua, G.N. Influence of cultivar on the post-harvest hardening of trifoliate yam (*Dioscorea dumetorum*) tubers. *Adv. Agric.* 2016, 2016, 1–18.
9. Siadjeu, C.; Mayland-quellhorst, E.; Albach, D.C. Genetic diversity and population structure of trifoliate yam (*Dioscorea dumetorum* Kunth) in Cameroon revealed by genotyping-by-sequencing (GBS). *BMC Plant Biol.* 2018, 1–14.
10. Siadjeu, C.; Pucker, B.; Viehöver, P.; Albach, D.C.; Weisshaar, B. High contiguity de novo genome sequence assembly of trifoliate yam (*Dioscorea dumetorum*) using long read sequencing. *Genes (Basel)*. 2020, 11, doi:10.3390/genes11030274.
11. Wang, L.; Wang, Z.; Chen, J.; Liu, C.; Zhu, W.; Wang, L.; Meng, L. De Novo Transcriptome Assembly and Development of Novel Microsatellite Markers for the Traditional Chinese Medicinal Herb, *Veratrilla Baillonii* Franch (*Gentianaceae*). *Evolutionary Bioinformatics* 2015, 11, 39–45, doi:10.4137/EBO.S20942.Received.
12. Alves-Carvalho, S.; Aubert, G.; Carrère, S.; Cruaud, C.; Brochot, A.L.; Jacquin, F.; Klein, A.; Martin, C.; Boucherot, K.; Kreplak, J.; et al. Full-length de novo assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. *Plant J.* 2015, 84, 1–19, doi:10.1111/tpj.12967.
13. Becerra-Moreno, A.; Redondo-Gil, M.; Benavides, J.; Nair, V.; Cisneros-Zevallos, L.; Jacobo-Velázquez, D.A. Combined effect of water loss and wounding stress on gene activation of metabolic pathways associated with phenolic biosynthesis in carrot. *Front. Plant Sci.* 2015, 6, doi:10.3389/fpls.2015.00837.
14. Wu, Z.-G.; Jiang, W.; Mantri, N.; Bao, X.-Q.; Chen, S.-L.; Tao, Z.-M. Transcriptome analysis reveals flavonoid biosynthesis regulation and simple sequence repeats in yam (*Dioscorea alata* L.) tubers. *BMC Genomics* 2015, doi:10.1186/s12864-015-1547-8.
15. Bhattacharjee, R.; Gedil, M.; Sartie, A.; Otoo, E.; Dumet, D.; Kikuno, H.; Kumar, P.L.; Asiedu, R. *Dioscorea*. In Book Wild Crop Relatives: Genomic Breeding Resources Industrial Crops. ed. C. Kole, 2011, 71–96, doi:10.1007/978-3-642-21102-7_4.
16. Tamiru, M.; Natsume, S.; Takagi, H.; White, B.; Yaegashi, H.; Shimizu, M.; Yoshida, K.; Uemura, A.; Oikawa, K.; Abe, A.; et al. Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. *BMC Biol.* 2017, 15, 1–20, doi:10.1186/s12915-017-0419-x.
17. Medoua, G.N.; Mbome, I.L.; Agbor-Egbe, T.; Mbofung, C.M.F. Physicochemical changes occurring during post-harvest hardening of trifoliate yam (*Dioscorea dumetorum*) tubers. *Food Chem.* 2005, 90, 597–601, doi:10.1016/j.foodchem.2004.04.018.
18. Medoua, G.N.; Mbome, I.L.; Agbor-Egbe, T.; Mbofung, C.M.F. Study of the hard-to-cook property of stored yam tubers (*Dioscorea dumetorum*) and some determining biochemical factors. *Food Res. Int.* 2005, 38, 143–149.
19. Le Gall, H.; Philippe, F.; Domon, J.M.; Gillet, F.; Pelloux, J.; Rayon, C. Cell wall metabolism in response to abiotic stress. *Plants* 2015, 4, 112–166, doi:10.3390/plants4010112.
20. Speicher, T.L.; Li, P.Z.; Wallace, I.S. Phosphoregulation of the plant cellulose synthase complex and cellulose synthase-like proteins. *Plants* 2018, 7, 1–18, doi:10.3390/plants7030052.
21. Polko, J.K.; Kieber, J.J. The regulation of cellulose biosynthesis in plants. *Plant Cell* 2019, 31, 282–296, doi:10.1105/tpc.18.00760.

22. Ambawat, S.; Sharma, P.; Yadav, N.R.; Yadav, R.C. MYB transcription factor genes as regulators for plant responses: An overview. *Physiol. Mol. Biol. Plants* 2013, 19, 307–321, doi:10.1007/s12298-013-0179-1.
23. Guo, H.; Wang, Y.; Wang, L.; Hu, P.; Wang, Y.; Jia, Y.; Zhang, C.; Zhang, Y.; Zhang, Y.; Wang, C.; et al. Expression of the MYB transcription factor gene Bp1MYB46 affects abiotic stress tolerance and secondary cell wall deposition in *Betula platyphylla*. *Plant Biotechnol. J.* 2017, 15, 107–121, doi:10.1111/pbi.12595.
24. Medoua, G.N.; Lape Mbome, I.; Agbor-Egbe, T.; Mbofung, C.M.F. Antinutritional factors changes occurring in trifoliate yam (*Dioscorea dumetorum*) tubers after harvest. *Food Chem.* 2007, 102, 716–720, doi:10.1016/j.foodchem.2006.06.005.
25. Dao, T.T.H.; Linthorst, H.J.M.; Verpoorte, R. Chalcone synthase and its functions in plant resistance. *Phytochem. Rev.* 2011, 10, 397–412, doi:10.1007/s11101-011-9211-7.
26. Arcuri, M.L.C.; Fialho, L.C.; Vasconcellos Nunes-Laitz, A.; Fuchs-Ferraz, M.C.P.; Wolf, I.R.; Valente, G.T.; Marino, C.L.; Maia, I.G. Genome-wide identification of multifunctional laccase gene family in *Eucalyptus grandis*: potential targets for lignin engineering and stress tolerance. *Trees - Struct. Funct.* 2020, 34, 745–758, doi:10.1007/s00468-020-01954-3.
27. Wang, Y.; Suzuki, H.; Xie, J.; Tomita, O.; Martin, D.J.; Higashi, M.; Kong, D.; Abe, R.; Tang, J. Mimicking Natural Photosynthesis: Solar to Renewable H₂ Fuel Synthesis by Z-Scheme Water Splitting Systems. *Chem. Rev.* 2018, 118, 5201–5241, doi:10.1021/acs.chemrev.7b00286.
28. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 2013, 29, 15–21, doi:10.1093/bioinformatics/bts635.
29. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009, 25, 2078–2079, doi:10.1093/bioinformatics/btp352.
30. Lawrence, M.; Huber, W.; Pagès, H.; Aboyoun, P.; Carlson, M.; Gentleman, R.; Morgan, M.T.; Carey, V.J. Software for Computing and Annotating Genomic Ranges. *PLoS Comput. Biol.* 2013, 9, 1–10, doi:10.1371/journal.pcbi.1003118.
31. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014, 15, 1–21, doi:10.1186/s13059-014-0550-8.
32. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2009, 26, 139–140, doi:10.1093/bioinformatics/btp616.
33. Moriya, Y.; Itoh, M.; Okuda, S.; Yoshizawa, A.C.; Kanehisa, M. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 2007, 35, 182–185, doi:10.1093/nar/gkm321.
34. Schwacke, R.; Ponce-Soto, G.Y.; Krause, K.; Bolger, A.M.; Arsova, B.; Hallab, A.; Gruden, K.; Stitt, M.; Bolger, M.E.; Usadel, B. MapMan4: A Refined Protein Classification and Annotation Framework Applicable to Multi-Omics Data Analysis. *Mol. Plant* 2019, 12, 879–892, doi:10.1016/j.molp.2019.01.003.
35. Stracke, R.; Werber, M.; Weisshaar, B. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* 2001, 4, 447–456, doi:10.1016/S1369-5266(00)00199-0.
36. Stracke, R.; Holtgräwe, D.; Schneider, J.; Pucker, B.; Rosleff Sørensen, T.; Weisshaar, B. Genome-wide identification and characterisation of R2R3-MYB genes in sugar beet (*Beta vulgaris*). *BMC Plant Biol.* 2014, 14, 1–17, doi:10.1186/s12870-014-0249-8.
37. Pucker, B.; Pandey, A.; Weisshaar, B.; Stracke, R. The R2R3-MYB gene family in banana (*Musa acuminata*): Genome-wide identification, classification and expression patterns. *PLoS One* 2020, 15, 1–27, doi:10.1371/journal.pone.0239275.
38. Pucker, B.; Reiher, F.; Schilbert, H.M. Automatic identification of players in the flavonoid biosynthesis with application on the biomedical plant croton tiglium. *Plants* 2020, 9, 1–21, doi:10.3390/plants9091103.

-
39. Götz, S.; García-Gómez, J.M.; Terol, J.; Williams, T.D.; Nagaraj, S.H.; Nueda, M.J.; Robles, M.; Talón, M.; Dopazo, J.; Conesa, A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 2008, 36, 3420–3435, doi:10.1093/nar/gkn176.
 40. Yeung, K.Y.; Medvedovic, M.; Bumgarner, R.E. From co-expression to co-regulation: how many microarray experiments do we need? *Genome Biol.* 2004, 5, 1–11, doi:10.1186/gb-2004-5-7-r48.

Supplementary material. Siadjeu et al. (2021)

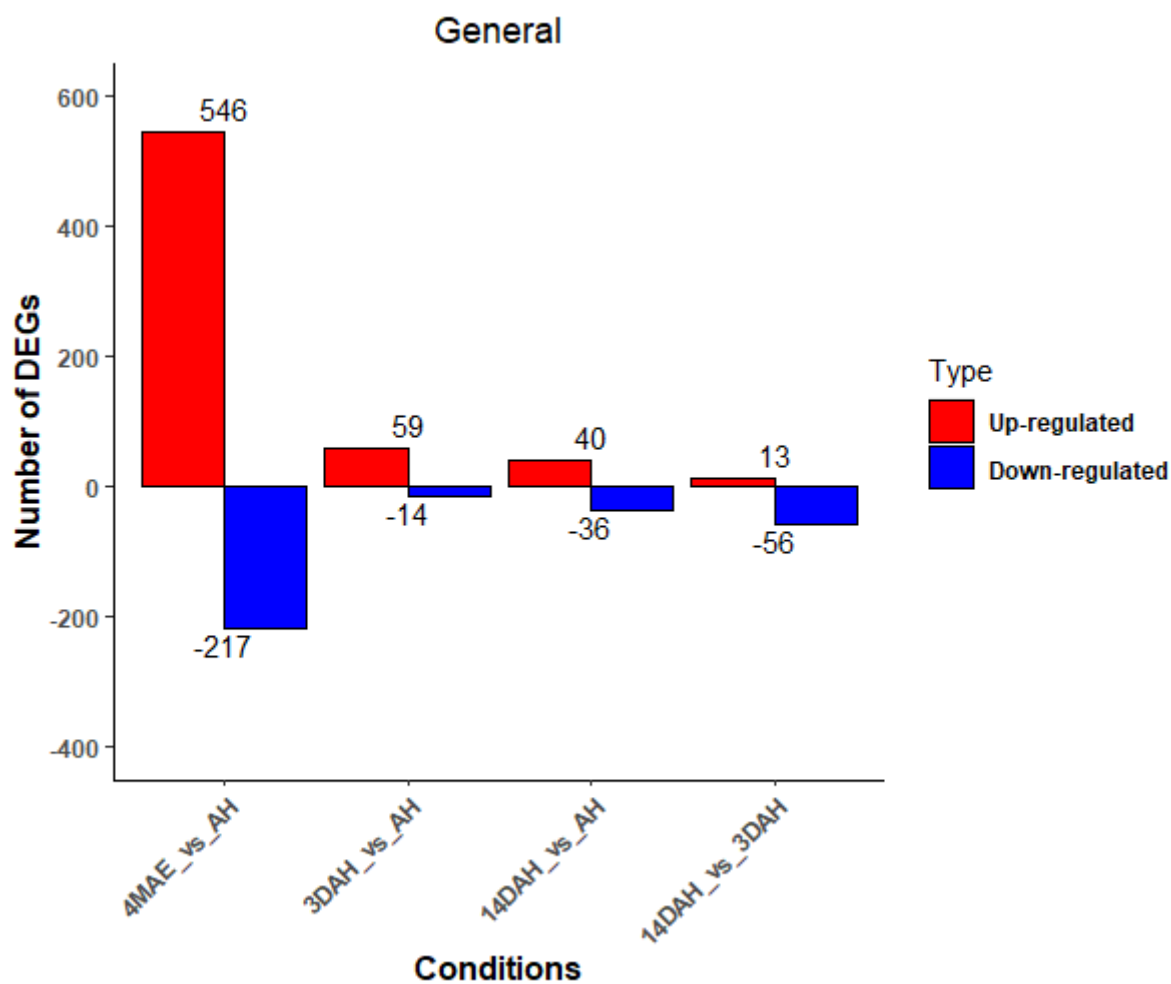
Supplementary S1: Statistic of clean reads mapped to *D. dumetorum* reference genome

Accessions	Conditions	Total cleaned reads	Total (%)	Uniquely mapped reads (%)	Total (%)
Bangou 1	4MAE	17893940	56.61	36.51	45.13
Bangou 1	4MAE	18949046	93.49	60.09	69.42
Bangou 1	4MAE	20612907	92.51	58.2	69.05
Bangou 1	AH	21284551	96.13	59	74.38
Bangou 1	AH	20481906	91.77	53.96	70.26
Bangou 1	AH	17929078	93.36	53.29	78.56
Bangou 1	3DAH	19902140	95.5	56.66	74
Bangou 1	3DAH	24032161	93.56	51.76	78.23
Bangou 1	3DAH	14752529	86.88	52.2	70.09
Bangou 1	14DAH	18279097	77.73	45.16	63.81
Bangou 1	14DAH	20344788	91.64	56.5	74.26
Bangou 1	14DAH	4536123	93.3	58.73	69.95
Bayangam 2	4MAE	53929557	96.63	63.53	71.97
Bayangam 2	4MAE	22569645	91.81	60.56	71.45
Bayangam 2	4MAE	24937580	88.9	57.83	65.31
Bayangam 2	AH	21708627	82.96	52.63	61.75
Bayangam 2	AH	23997962	97.29	53.7	80.6
Bayangam 2	AH	31335502	94.09	49.7	78.78
Bayangam 2	3DAH	24284434	94.56	54.66	72.63
Bayangam 2	3DAH	22565017	93.62	56.11	76.8
Bayangam 2	3DAH	20803207	96.21	58.49	74.35
Bayangam 2	14DAH	27593312	90.11	55.12	72.04
Bayangam 2	14DAH	18519729	95.9	62.63	71.94
Bayangam 2	14DAH	19713286	96	60.25	70.93
Fonkouankem 1	4MAE	22166474	94.25	62.53	71.25
Fonkouankem 1	4MAE	12111303	84.48	55.59	61.48
Fonkouankem 1	4MAE	20286805	85.76	56.75	61.04
Fonkouankem 1	AH	13126435	93.29	53.1	74.92
Fonkouankem 1	AH	19560305	92.51	55.3	73.54
Fonkouankem 1	AH	13307694	93.43	56	70.66
Fonkouankem 1	3DAH	22500916	92.21	56.99	71.71
Fonkouankem 1	3DAH	17371457	96.19	56.96	74.36
Fonkouankem 1	3DAH	11617175	95.44	57.81	67.16
Fonkouankem 1	14DAH	20060449	93.44	62.04	68.06
Fonkouankem 1	14DAH	15829240	86.93	56.82	71.64
Fonkouankem 1	14DAH	15279090	86.98	56.88	63.02
Ibosweet 3	4MAE	16672162	93.04	59.72	68.71
Ibosweet 3	4MAE	13924548	91.2	58.68	68.4
Ibosweet 3	4MAE	30873744	91.99	57.53	70.07
Ibosweet 3	AH	7881301	93.36	55.37	73.71
Ibosweet 3	AH	14683717	94.95	56.98	75.31
Ibosweet 3	AH	14787799	95.71	61.83	72.2
Ibosweet 3	3DAH	14413912	95.55	59.9	73.23
Ibosweet 3	3DAH	21994748	94.17	56.54	73.06
Ibosweet 3	3DAH	17962629	95.22	55.32	73.62
Ibosweet 3	14DAH	15590855	91.61	58.23	68.25
Ibosweet 3	14DAH	21780220	84.31	52.77	65.4
Ibosweet 3	14DAH	18583946	95.04	60.53	72.64

Supplementary material. Siadjeu et al. (2021)

Supplementary S2: Results of DEG analysis per accession based on the pairwise comparisons (4MAE vs, AH, 3DAH vs AH, 14DAH vs AH, 14DAH vs 3DAH) resulting from the approach STAR_DESeq2 (cf. external additional file)

Supplementary material. Siadjeu et al. (2021)



Supplementary S3: Number of DEGs based on the combined analysis of the three hardening accessions 4MAE and after harvest.

Supplementary material. Siadjeu et al. (2021)

Supplementary S4: Results of DEG analysis based on the combined analysis of the 3 hardening accessions 4MAE and after harvest, Supplementary S5: Results of DEG analysis based on the comparison between Ibo sweet 3 and other accessions after harvest (cf. external additional file)

Supplementary material. Siadjeu et al. (2021)

Supplementary S5: Results of DEG analysis based on the comparison between Ibo sweet 3 and other accessions after harvest (cf. external additional file)

Supplementary S6: Group resulting from Cluster analysis of DEGs 3DAH among the different sampling times. Bangou, Bayangam, Fonkouankem, combined analysis of the three hardening accessions

	Bangou	
Cluster	Functional annotation	Contig
1	Enzyme classification.EC_2 transferases.EC_2.4 glycosyltransferase	contig1.g403
1	not assigned.annotated	contig1.g721
1	not assigned.annotated	contig119.g1631
1	Phytohormone action.cytokinin.conjugation and degradation.UDP-dependent glycosyl transferase	contig119.g312
1	Phytohormone action.cytokinin.conjugation and degradation.UDP-dependent glycosyl transferase	contig119.g313
1	Enzyme classification.EC_1 oxidoreductases.EC_1.14 oxidoreductase acting on paired donor with incorporation or reduction of molecular oxygen	contig119.g563
1	not assigned.not annotated	contig119.g712
1	Protein modification.phosphorylation.TKL protein kinase superfamily.G-Lectin protein kinase families.protein kinase (SD-1)	contig126.g3
1	Cell wall organisation.hemicellulose.xylan.modification and degradation.xylan O-acetyltransferase (XOAT)	contig143.g88
1	Cell wall organisation.hemicellulose.xylan.modification and degradation.xylan O-acetyltransferase (XOAT)	contig145.g17
1	not assigned.annotated	contig199.g10
1	not assigned.annotated	contig199.g1511
1	Phytohormone action.gibberellin.modification and degradation.gibberellin 2-oxidase	contig222.g1163
1	Cell wall organisation.cellulose.cellulose-hemicellulose network assembly.regulatory protein (COB)	contig267.g188
1	not assigned.not annotated	contig267.g320
1	Solute transport.carrier-mediated transport.APC superfamily.potassium cation transporter (HAK/KUP/KT)	contig3.g258
1	Cell wall organisation.pectin.rhamnogalacturonan I.modification and degradation.alpha-L-arabinofuranosidase activities.bifunctional alpha-L-arabinofuranosidase and beta-D-xylosidase (BXL)	contig3.g604
1	not assigned.not annotated	contig338.g10
1	Enzyme classification.EC_1 oxidoreductases.EC_1.14 oxidoreductase acting on paired donor with incorporation or reduction of molecular oxygen	contig359.g7
1	Phytohormone action.cytokinin.biosynthesis.cytokinin phosphoribohydrolase	contig390.g141
1	not assigned.not annotated	contig391.g18
1	not assigned.not annotated	contig421.g166
1	Enzyme classification.EC_1 oxidoreductases.EC_1.10 oxidoreductase acting on diphenol or related substance as donor	contig421.g556
1	not assigned.annotated	contig463.g7
1	not assigned.annotated	contig488.g27
1	RNA biosynthesis.transcriptional regulation.AP2/ERF transcription factor superfamily.transcription factor (DREB)	contig546.g79
1	Enzyme classification.EC_2 transferases.EC_2.7 transferase transferring phosphorus-containing group	contig552.g93

	Bangou (continued)	
Cluster	Functional annotation	Contig
1	Enzyme classification.EC_2 transferases.EC_2.4 glycosyltransferase	contig554.g475
1	Protein modification.phosphorylation.CAMK protein kinase superfamily.SNF1-related protein kinase (SnRK3)	contig557.g723
1	Protein modification.phosphorylation.protein kinase (PEK)	contig557.g888
1	not assigned.not annotated	contig558.g1906
1	Phytohormone action.signalling peptides.CRP (cysteine-rich-peptide) category.RALF/RALFL-peptide activity.RALF/RALFL-precursor polypeptide	contig558.g2268
1	Cell cycle organisation.cell cycle control.cyclin-dependent regulation.cyclin activities.cyclin (CYCB)	contig558.g93
1	not assigned.not annotated	contig570.g234
1	Protein modification.phosphorylation.CAMK protein kinase superfamily.SNF1-related protein kinase (SnRK3)	contig60.g25
1	Enzyme classification.EC_2 transferases.EC_2.4 glycosyltransferase	contig606.g48
1	Phytohormone action.signalling peptides.CRP (cysteine-rich-peptide) category.RALF/RALFL-peptide activity.RALF/RALFL-precursor polypeptide	contig675.g16
1	not assigned.not annotated	contig678.g112
1	not assigned.annotated	contig678.g292
1	not assigned.annotated	contig678.g386
1	Phytohormone action.jasmonic acid.biosynthesis.13-lipoxygenase	contig685.g29
1	Enzyme classification.EC_2 transferases.EC_2.4 glycosyltransferase	contig689.g65
1	not assigned.annotated	contig711.g13
1	not assigned.not annotated	contig728.g796
1	RNA biosynthesis.transcriptional regulation.AP2/ERF transcription factor superfamily.transcription factor (DREB)	contig771.g2
1	External stimuli response.symbiont.symbiosis signalling pathway.symbiosis factor (VAPYRIN)	contig886.g8
1	Protein modification.phosphorylation.TKL protein kinase superfamily.G-Lectin protein kinase families.protein kinase (SD-1)	contig89.g2235
1	Cell wall organisation.hemicellulose.xylan.modification and degradation.xylan O-acetyltransferase (XOAT)	contig920.g250
1	not assigned.annotated	contig95.g584
2	not assigned.annotated	contig1.g808
2	not assigned.annotated	contig105.g26
2	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig206.g10
2	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig206.g11
2	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig206.g6
2	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig206.g8
2	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig267.g402

	Bangou (continued)	
Cluster	Functional annotation	Contig
2	not assigned.not annotated	contig330.g122
2	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig355.g38
2	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig391.g20
2	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig391.g26
2	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig391.g28
2	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig391.g29
2	Photosynthesis.photophosphorylation.photosystem II.PS-II complex.component PsbX	contig544.g1970
2	not assigned.annotated	contig544.g2227
2	not assigned.annotated	contig544.g2453
2	not assigned.annotated	contig546.g272
2	Photosynthesis.photophosphorylation.photosystem I.PS-I complex.component PsaF	contig549.g218
2	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig553.g402
2	Solute transport.carrier-mediated transport.MFS superfamily.anion transporter (NRT1/PTR)	contig558.g136
2	not assigned.annotated	contig558.g1884
2	not assigned.annotated	contig559.g41
2	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig565.g52
2	Photosynthesis.photophosphorylation.photosystem II.LHC-related protein groups.three-helix LHC-related protein group.protein (ELIP)	contig666.g65
2	Photosynthesis.photophosphorylation.photosystem I.LHC-I complex.component LHCa4	contig678.g379
2	Photosynthesis.photophosphorylation.photosystem I.LHC-I complex.component LHCa4	contig679.g24
2	not assigned.annotated	contig691.g220
2	Cell wall organisation.pectin.modification and degradation.pectate lyase	contig705.g980
2	not assigned.annotated	contig764.g854
2	not assigned.not annotated	contig773.g11
2	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig879.g4
2	Solute transport.carrier-mediated transport.MFS superfamily.anion transporter (NRT1/PTR)	contig88.g2
2	Photosynthesis.photophosphorylation.photosystem II.LHC-related protein groups.three-helix LHC-related protein group.protein (ELIP)	contig89.g1873
2	Photosynthesis.photophosphorylation.photosystem I.PS-I complex.component PsaD	contig89.g2104

	Bayangam	
Cluster	Functional annotation	Contig
1	Enzyme classification.EC_2 transferases.EC_2.4 glycosyltransferase	contig1.g403
1	not assigned.annotated	contig103.g12
1	Protein homeostasis.proteolysis.aspartic-type peptidase activities.A1-class protease (Pepsin)	contig106.g55
1	Cell wall organisation.lignin.monolignol glycosylation and deglycosylation.coniferin beta-glucosidase	contig119.g106
1	not assigned.annotated	contig119.g365
1	Phytohormone action.signalling peptides.NCRP (non-cysteine-rich-peptide) category.PIP/PIPL-peptide activity.PIP/PIPL-precursor polypeptide	contig192.g17
1	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig206.g10
1	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig206.g11
1	not assigned.not annotated	contig222.g686
1	not assigned.not annotated	contig235.g14
1	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig267.g402
1	Cell wall organisation.hemicellulose.xylan.modification and degradation.xylan O-acetyltransferase (XOAT)	contig279.g48
1	Enzyme classification.EC_2 transferases.EC_2.7 transferase transferring phosphorus-containing group	contig293.g19
1	not assigned.not annotated	contig349.g7
1	not assigned.annotated	contig351.g11
1	Phytohormone action.cytokinin.biosynthesis.cytokinin phosphoribohydrolase	contig390.g141
1	not assigned.not annotated	contig391.g18
1	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig391.g28
1	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig391.g29
1	not assigned.not annotated	contig421.g166
1	Protein modification.phosphorylation.TKL protein kinase superfamily.RLCK-IX receptor-like protein kinase families.receptor-like protein kinase (RLCK-IXb)	contig421.g532
1	Coenzyme metabolism.tetrapyrrol biosynthesis.chlorophyll metabolism.Mg-protoporphyrin IX monomethylester cyclase complex.catalytic component CRD1	contig423.g6
1	Protein homeostasis.proteolysis.metallopeptidase activities.carboxypeptidase activities.M28-class carboxypeptidase	contig457.g385
1	Protein homeostasis.proteolysis.metallopeptidase activities.carboxypeptidase activities.M28-class carboxypeptidase	contig458.g6

	Bayangam (continued)	
Cluster	Functional annotation	Contig
1	not assigned.not annotated	contig458.g8
1	not assigned.annotated	contig471.g1
1	Protein modification.phosphorylation.CAMK protein kinase superfamily.SNF1-related protein kinase (SnRK3)	contig494.g1
1	not assigned.annotated	contig521.g20
1	not assigned.annotated	contig536.g3
1	Solute transport.carrier-mediated transport.DMT superfamily.organic cation transporter (PUP)	contig542.g85
1	Nutrient uptake.nitrogen assimilation.nitrate assimilation.nitrate reductase	contig544.g1026
1	RNA biosynthesis.transcriptional regulation.MYB transcription factor superfamily.transcription factor (MYB)	contig544.g1666
1	not assigned.annotated	contig544.g2409
1	not assigned.annotated	contig544.g2453
1	Solute transport.carrier-mediated transport.APC superfamily.APC family.cationic amino acid transporter (CAT)	contig546.g471
1	Solute transport.carrier-mediated transport.MOP superfamily.MATE family.metabolite transporter (DTX)	contig549.g155
1	Protein modification.phosphorylation.TKL protein kinase superfamily.G-Lectin protein kinase families.protein kinase (SD-1)	contig553.g328
1	Enzyme classification.EC_2 transferases.EC_2.7 transferase transferring phosphorus-containing group	contig553.g374
1	Secondary metabolism.phenolics.flavonoid biosynthesis.chalcones.chalcone synthase activity.chalcone synthase (CHS)	contig554.g125
1	Phytohormone action.abscisic acid.transport.abscisic acid transporter (AIT)	contig557.g1019
1	Protein modification.phosphorylation.CAMK protein kinase superfamily.SNF1-related protein kinase (SnRK3)	contig557.g723
1	RNA biosynthesis.transcriptional regulation.transcription factor (WRKY)	contig559.g157
1	not assigned.annotated	contig559.g41
1	Cell wall organisation.pectin.homogalacturonan.modification and degradation.pectin methylesterase	contig565.g119
1	Phytohormone action.brassinosteroid.biosynthesis.6-deoxocastasterone 6-oxidase	contig592.g24
1	Enzyme classification.EC_2 transferases.EC_2.7 transferase transferring phosphorus-containing group	contig595.g104
1	not assigned.annotated	contig595.g86
1	Protein modification.phosphorylation.CAMK protein kinase superfamily.SNF1-related protein kinase (SnRK3)	contig60.g25
1	Enzyme classification.EC_2 transferases.EC_2.4 glycosyltransferase	contig606.g48
1	RNA biosynthesis.transcriptional regulation.MYB transcription factor superfamily.transcription factor (MYB)	contig631.g31
1	not assigned.not annotated	contig66.g23

	Bayangam (continued)	
Cluster	Functional annotation	Contig
1	Phytohormone action.signalling peptides.CRP (cysteine-rich-peptide) category.RALF/RALFL-peptide activity.RALF/RALFL-precursor polypeptide	contig675.g16
1	Photosynthesis.photophosphorylation.photosystem I.LHC-I complex.component LHCa4	contig678.g379
1	Photosynthesis.photophosphorylation.photosystem I.LHC-I complex.component LHCa4	contig679.g24
1	Enzyme classification.EC_2 transferases.EC_2.4 glycosyltransferase	contig689.g65
1	not assigned.annotated	contig691.g220
1	not assigned.not annotated	contig691.g303
1	Phytohormone action.signalling peptides.NCRP (non-cysteine-rich-peptide) category.PIP/PIPL-peptide activity.PIP/PIPL-precursor polypeptide	contig705.g1119
1	RNA biosynthesis.transcriptional regulation.transcription factor (NAC)	contig705.g1396
1	Cell wall organisation.pectin.modification and degradation.pectate lyase	contig705.g980
1	Secondary metabolism.phenolics.flavonoid biosynthesis.chalcones.chalcone synthase activity.chalcone synthase (CHS)	contig714.g10
1	Secondary metabolism.phenolics.flavonoid biosynthesis.chalcones.chalcone synthase activity.chalcone synthase (CHS)	contig714.g18
1	Phytohormone action.abscisic acid.transport.abscisic acid transporter (AIT)	contig72.g46
1	not assigned.annotated	contig758.g13
1	Protein modification.phosphorylation.TKL protein kinase superfamily.receptor-like protein kinase (RLCK-XI)	contig764.g288
1	RNA biosynthesis.transcriptional regulation.AP2/ERF transcription factor superfamily.transcription factor (ERF)	contig764.g945
1	not assigned.not annotated	contig773.g11
1	RNA biosynthesis.transcriptional regulation.MYB transcription factor superfamily.transcription factor (MYB)	contig782.g16
1	Protein modification.phosphorylation.TKL protein kinase superfamily.receptor-like protein kinase (RLCK-XI)	contig805.g33
1	Solute transport.carrier-mediated transport.MOP superfamily.MATE family.metabolite transporter (DTX)	contig837.g56
1	Solute transport.carrier-mediated transport.MFS superfamily.anion transporter (NRT1/PTR)	contig88.g2
1	Protein modification.phosphorylation.TKL protein kinase superfamily.G-Lectin protein kinase families.protein kinase (SD-1)	contig89.g2235
1	not assigned.annotated	contig89.g487
1	Enzyme classification.EC_2 transferases.EC_2.1 transferase transferring one-carbon group	contig891.g44
1	not assigned.annotated	contig918.g16
1	RNA biosynthesis.transcriptional regulation.AP2/ERF transcription factor superfamily.transcription factor (ERF)	contig921.g9

	Bayangam (continued)	
Cluster	Functional annotation	Contig
2	RNA biosynthesis.transcriptional regulation.transcription factor (C2H2-ZF)	contig1.g578
2	not assigned.annotated	contig1.g808
2	Cell wall organisation.hemicellulose.xylan.modification and degradation.xylan O-acetyltransferase (XOAT)	contig143.g89
2	Nucleotide metabolism.purines.catabolism.allantoinase	contig2.g458
2	Cell wall organisation.hemicellulose.xylan.biosynthesis.xylosyltransferase activities.xylosyltransferase (IRX10)	contig2.g807
2	Cell cycle organisation.cell cycle control.cyclin-dependent regulation.cyclin activities.cyclin (CYCB)	contig2.g845
2	RNA biosynthesis.transcriptional regulation.transcription factor (C3H-ZF)	contig207.g608
2	not assigned.annotated	contig267.g420
2	not assigned.not annotated	contig290.g11
2	Protein modification.phosphorylation.TKL protein kinase superfamily.receptor-like protein kinase (RLCK-VI)	contig294.g1079
2	Protein modification.phosphorylation.TKL protein kinase superfamily.receptor-like protein kinase (RLCK-VI)	contig302.g136
2	Vesicle trafficking.clathrin coated vesicle (CCV) machinery.CCV plasma membrane detachment.dynamin (DRP1)	contig343.g22
2	not assigned.annotated	contig355.g67
2	Protein modification.phosphorylation.TKL protein kinase superfamily.receptor-like protein kinase (RLCK-V)	contig383.g5
2	not assigned.annotated	contig390.g268
2	Enzyme classification.EC_2 transferases.EC_2.3 acyltransferase	contig395.g32
2	not assigned.annotated	contig421.g356
2	not assigned.annotated	contig488.g24
2	not assigned.annotated	contig540.g97
2	not assigned.annotated	contig544.g239
2	not assigned.not annotated	contig546.g206
2	Cell wall organisation.hemicellulose.xylan.biosynthesis.galacturonosyltransferase	contig549.g157
2	not assigned.annotated	contig552.g389
2	Enzyme classification.EC_2 transferases.EC_2.7 transferase transferring phosphorus-containing group	contig552.g93
2	not assigned.not annotated	contig553.g28

Bayangam (continued)		
Cluster	Functional annotation	Contig
2	Cell cycle organisation.cell cycle control.cyclin-dependent regulation.cyclin activities.cyclin (CYCB)	contig558.g93
2	RNA biosynthesis.transcriptional regulation.AP2/ERF transcription factor superfamily.transcription factor (AP2)	contig560.g108
2	not assigned.not annotated	contig662.g21
2	Phytohormone action.jasmonic acid.biosynthesis.13-lipoxygenase	contig677.g56
2	Cell cycle organisation.cytokinesis.phragmoplast disassembly.NACK-PQR signalling pathway.MAPKK-kinase (NPK/ANP)	contig678.g100
2	Nucleotide metabolism.deoxynucleotides.biosynthesis.ribonucleoside-diphosphate reductase heterodimer.small subunit	contig678.g662
2	not assigned.not annotated	contig678.g707
2	Enzyme classification.EC_2 transferases.EC_2.7 transferase transferring phosphorus-containing group	contig685.g24
2	Phytohormone action.jasmonic acid.biosynthesis.13-lipoxygenase	contig685.g29
2	External stimuli response.gravity.sensing and signalling.signalling protein factor (LAZY)	contig687.g17
2	External stimuli response.gravity.sensing and signalling.signalling protein factor (LAZY)	contig688.g16
2	Cell cycle organisation.cytokinesis.phragmoplast microtubule organization.microtubule plus-end-tracking protein (EB1)	contig691.g407
2	not assigned.not annotated	contig750.g72
2	not assigned.annotated	contig753.g10
2	Solute transport.carrier-mediated transport.OPT family.oligopeptide transporter (OPT)	contig764.g53
2	Cell wall organisation.cutin and suberin.cutin polyester biosynthesis.cutin synthase (DCR)	contig775.g1
2	Protein modification.phosphorylation.TKL protein kinase superfamily.receptor-like protein kinase (RLCK-X)	contig790.g11
2	Cell cycle organisation.cytokinesis.phragmoplast microtubule organization.microtubule-based motor protein (Kinesin-10)	contig896.g122
2	not assigned.not annotated	contig909.g49
2	Protein modification.phosphorylation.TKL protein kinase superfamily.protein kinase (LRR-III)	contig920.g254
2	not assigned.annotated	contig95.g584

	Fonkouankem	
Cluster	Functional annotation	Contig
1	not assigned.annotated	contig105.g26
1	not assigned.annotated	contig119.g1631
1	Solute transport.carrier-mediated transport.MFS superfamily.DHA-1 family.metal chelator transporter (TCR)	contig119.g1726
1	Protein modification.phosphorylation.TKL protein kinase superfamily.receptor-like protein kinase (RLCK-X)	contig294.g263
1	Protein modification.phosphorylation.TKL protein kinase superfamily.receptor-like protein kinase (RLCK-VI)	contig308.g7
1	RNA biosynthesis.transcriptional regulation.C2C2 transcription factor superfamily.transcription factor (GATA)	contig390.g530
1	Carbohydrate metabolism.trehalose metabolism.trehalase	contig418.g268
1	not assigned.not annotated	contig421.g238
1	not assigned.annotated	contig463.g7
1	not assigned.annotated	contig546.g491
1	Solute transport.carrier-mediated transport.MOP superfamily.MATE family.metabolite transporter (DTX)	contig549.g155
1	Protein modification.phosphorylation.CAMK protein kinase superfamily.SNF1-related protein kinase (SnRK3)	contig557.g723
1	not assigned.annotated	contig558.g1884
1	Protein modification.phosphorylation.TKL protein kinase superfamily.receptor-like protein kinase (RLCK-VI)	contig564.g1082
1	Protein modification.phosphorylation.CAMK protein kinase superfamily.SNF1-related protein kinase (SnRK3)	contig60.g25
1	not assigned.annotated	contig711.g13
1	not assigned.not annotated	contig728.g796
1	Solute transport.carrier-mediated transport.OPT family.oligopeptide transporter (OPT)	contig764.g53
1	Cell wall organisation.cutin and suberin.cutin polyester biosynthesis.cutin synthase (DCR)	contig775.g1
1	Protein modification.phosphorylation.TKL protein kinase superfamily.receptor-like protein kinase (RLCK-X)	contig790.g11
1	Phytohormone action.cytokinin.biosynthesis.cytokinin phosphoribohydrolase	contig820.g29
1	Carbohydrate metabolism.trehalose metabolism.trehalase	contig850.g21
1	Cytoskeleton organisation.microtubular network.alpha-beta-Tubulin heterodimer.component alpha-Tubulin	contig89.g113
1	Cytoskeleton organisation.microtubular network.alpha-beta-Tubulin heterodimer.component alpha-Tubulin	contig902.g19
1	Enzyme classification.EC_1 oxidoreductases.EC_1.10 oxidoreductase acting on diphenol or related substance as donor	contig915.g28
1	RNA biosynthesis.transcriptional regulation.MYB transcription factor superfamily.transcription factor (MYB)	contig94.g124
1	not assigned.annotated	contig95.g584

	Fonkouankem (continued)	
Cluster	Functional annotation	Contig
1	Protein modification.phosphorylation.TKL protein kinase superfamily.LRR-VI protein kinase families.protein kinase (LRR-VI-2)	contig95.g791
1	Enzyme classification.EC_2 transferases.EC_2.3 acyltransferase	contig95.g847
2	not assigned.annotated	contig1.g268
2	not assigned.annotated	contig1.g721
2	not assigned.annotated	contig105.g33
2	Cell wall organisation.lignin.monolignol glycosylation and deglycosylation.coniferin beta-glucosidase	contig119.g106
2	Solute transport.carrier-mediated transport.MFS superfamily.SP family.inositol transporter (INT)	contig119.g127
2	Solute transport.carrier-mediated transport.APC superfamily.AAAP family.amino acid transporter (AAP)	contig119.g1883
2	Secondary metabolism.phenolics.p-coumaroyl-CoA biosynthesis.phenylalanine ammonia lyase activity.phenylalanine ammonia lyase (PAL)	contig119.g44
2	not assigned.not annotated	contig119.g712
2	Cell wall organisation.hemicellulose.xylan.modification and degradation.xylan O-acetyltransferase (XOAT)	contig143.g88
2	Cell wall organisation.hemicellulose.xylan.modification and degradation.xylan O-acetyltransferase (XOAT)	contig145.g17
2	RNA biosynthesis.transcriptional regulation.transcription factor (NAC)	contig158.g23
2	not assigned.annotated	contig194.g56
2	not assigned.annotated	contig199.g10
2	Cytoskeleton organisation.microtubular network.Kinesin microtubule-based motor protein activities.motor protein (Kinesin-14)	contig199.g1288
2	Cell wall organisation.hemicellulose.xylan.modification and degradation.xylan O-acetyltransferase (XOAT)	contig199.g1435
2	Cell wall organisation.lignin.monolignol conjugation and polymerization.lignin laccase	contig199.g1672
2	not assigned.not annotated	contig2.g920
2	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig206.g11
2	Lipid metabolism.lipid degradation.triacylglycerol lipase activities.diacyl-/triacylglycerol lipase activities.lipase (LIP)	contig222.g16
2	Cell cycle organisation.cytokinesis.preprophase microtubule organization.microtubule-based motor protein (Kinesin-14)	contig222.g71
2	Cell wall organisation.cellulose.cellulose-hemicellulose network assembly.regulatory protein (COB)	contig267.g188
2	not assigned.annotated	contig267.g189
2	not assigned.annotated	contig267.g420

Fonkouankem (continued)		
Cluster	Functional annotation	Contig
2	RNA biosynthesis.transcriptional regulation.MYB transcription factor superfamily.transcription factor (MYB)	contig267.g494
2	Cell wall organisation.cellulose.cellulose-hemicellulose network assembly.regulatory protein (COB)	contig278.g50
2	Cell wall organisation.lignin.monolignol biosynthesis.p-coumaroyl shikimate/quinic 3-hydroxylase (C3H)	contig3.g487
2	not assigned.not annotated	contig330.g122
2	Cell wall organisation.pectin.rhamnogalacturonan I.modification and degradation.beta-galactosidase (BGAL)	contig390.g181
2	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig391.g29
2	Secondary metabolism.phenolics.p-coumaroyl-CoA biosynthesis.cinnamate 4-hydroxylase (C4H)	contig421.g274
2	Protein homeostasis.proteolysis.aspartic-type peptidase activities.A1-class protease (Pepsin)	contig43.g105
2	Secondary metabolism.phenolics.p-coumaroyl-CoA biosynthesis.cinnamate 4-hydroxylase (C4H)	contig446.g4
2	Protein modification.phosphorylation.CAMK protein kinase superfamily.SNF1-related protein kinase (SnRK3)	contig494.g1
2	Protein homeostasis.proteolysis.aspartic-type peptidase activities.A1-class protease (Pepsin)	contig544.g813
2	Enzyme classification.EC_1 oxidoreductases.EC_1.3 oxidoreductase acting on CH-CH group of donor	contig546.g147
2	not assigned.annotated	contig546.g359
2	not assigned.not annotated	contig546.g718
2	RNA biosynthesis.transcriptional regulation.AP2/ERF transcription factor superfamily.transcription factor (DREB)	contig546.g79
2	not assigned.annotated	contig552.g152
2	Cell wall organisation.lignin.monolignol biosynthesis.caffeoyl-CoA 3-O-methyltransferase (CCoA-OMT)	contig552.g180
2	Amino acid metabolism.biosynthesis.serine family.non-photorespiratory serine.phosphoglycerate dehydrogenase	contig552.g308
2	Protein translocation.peroxisome.importomer translocation system.receptor monoubiquitination system.component Pex22	contig553.g322
2	not assigned.not annotated	contig554.g128
2	not assigned.annotated	contig554.g687
2	RNA biosynthesis.transcriptional regulation.transcription factor (NAC)	contig556.g459
2	Cell wall organisation.cellulose.cellulose synthase complex (CSC).CSC components.catalytic component CesA	contig557.g748
2	Nutrient uptake.nitrogen assimilation.ammonium assimilation.glutamine synthetase activities.plastidial glutamine synthetase (GLN2)	contig558.g1807
2	not assigned.annotated	contig558.g1891

	Fonkouankem (continued)	
Cluster	Functional annotation	Contig
2	not assigned.not annotated	contig558.g1906
2	Cell wall organisation.lignin.monolignol conjugation and polymerization.p-coumaroyl-CoA:monolignol transferase (PMT)	contig559.g139
2	not assigned.annotated	contig559.g178
2	not assigned.annotated	contig559.g40
2	not assigned.annotated	contig559.g41
2	not assigned.not annotated	contig570.g234
2	Multi-process regulation.Rop-GTPase regulatory system.ROP GTPase effector activities.GTPase effector (RIC)	contig581.g131
2	not assigned.annotated	contig588.g101
2	Cell wall organisation.cellulose.cellulose synthase complex (CSC).CSC components.catalytic component CesA	contig60.g53
2	Cell wall organisation.lignin.monolignol biosynthesis.caffeoyl shikimate esterase (CSE)	contig646.g19
2	RNA biosynthesis.transcriptional regulation.MYB transcription factor superfamily.transcription factor (MYB)	contig678.g290
2	Photosynthesis.photophosphorylation.photosystem I.LHC-I complex.component LHCA4	contig678.g379
2	not assigned.annotated	contig678.g386
2	not assigned.annotated	contig678.g555
2	Photosynthesis.photophosphorylation.photosystem I.LHC-I complex.component LHCA4	contig679.g24
2	not assigned.annotated	contig679.g31
2	RNA biosynthesis.transcriptional regulation.MYB transcription factor superfamily.transcription factor (MYB)	contig693.g10
2	Cytoskeleton organisation.microtubular network.microtubule dynamics.cortical microtubule organisation.Exocyst complex recruiting factor (VETH)	contig705.g1283
2	Multi-process regulation.Rop-GTPase regulatory system.ROP GTPase effector activities.GTPase effector (RIC)	contig705.g607
2	not assigned.not annotated	contig714.g7
2	Lipid metabolism.lipid degradation.triacylglycerol lipase activities.diacyl-/triacylglycerol lipase activities.lipase (LIP)	contig729.g9
2	Cell wall organisation.cellulose.cellulose synthase complex (CSC).CSC components.catalytic component CesA	contig73.g5
2	Lipid metabolism.lipid degradation.triacylglycerol lipase activities.diacyl-/triacylglycerol lipase activities.lipase (LIP)	contig730.g9
2	RNA biosynthesis.transcriptional regulation.MYB transcription factor superfamily.transcription factor (MYB)	contig750.g121
2	Cell wall organisation.lignin.monolignol biosynthesis.caffeic acid O-methyltransferase (COMT)	contig750.g97

	Fonkouankem (continued)	
Cluster	Functional annotation	Contig
2	Secondary metabolism.phenolics.p-coumaroyl-CoA biosynthesis.phenylalanine ammonia lyase activity.phenylalanine ammonia lyase (PAL)	contig757.g12
2	not assigned.not annotated	contig765.g21
2	RNA biosynthesis.transcriptional regulation.AP2/ERF transcription factor superfamily.transcription factor (DREB)	contig771.g2
2	Solute transport.carrier-mediated transport.MOP superfamily.MATE family.metabolite transporter (DTX)	contig837.g56
2	Cytoskeleton organisation.microfilament network.actin polymerisation.Arps2/3 actin polymerization initiation complex.component ArpsC2	contig869.g12
2	Cytoskeleton organisation.microfilament network.actin polymerisation.Arps2/3 actin polymerization initiation complex.component ArpsC2	contig870.g14
2	Solute transport.carrier-mediated transport.MFS superfamily.anion transporter (NRT1/PTR)	contig88.g2
2	Cell wall organisation.hemicellulose.xylan.modification and degradation.xylan O-acetyltransferase (XOAT)	contig920.g250
2	Cell wall organisation.hemicellulose.xylan.modification and degradation.xylan O-acetyltransferase (XOAT)	contig922.g12
2	not assigned.not annotated	contig95.g795

	Combined analysis of the three hardening accessions	
Cluster	Functional annotation	Contig
1	not assigned.annotated	contig1.g808
1	Photosynthesis.photophosphorylation.cyclic electron flow.PGR5/PGRL1 complex.component PGR5-like	contig119.g125
1	not assigned.not annotated	contig119.g712
1	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig206.g10
1	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig206.g11
1	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig206.g6
1	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig206.g8
1	Photosynthesis.photophosphorylation.photosystem I.PS-I complex.component PsaH	contig222.g1555
1	Lipid metabolism.lipid degradation.triacylglycerol lipase activities.diacyl-/triacylglycerol lipase activities.lipase (LIP)	contig222.g16
1	Cell cycle organisation.cytokinesis.preprophase microtubule organization.microtubule-based motor protein (Kinesin-14)	contig222.g71
1	Protein homeostasis.proteolysis.serine-type peptidase activities.S8-class protease (subtilisin) families.protease (SBT1)	contig227.g44
1	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig267.g402
1	RNA biosynthesis.transcriptional regulation.MYB transcription factor superfamily.transcription factor (MYB)	contig267.g494
1	Cell wall organisation.cellulose.cellulose-hemicellulose network assembly.regulatory protein (COB)	contig278.g50
1	not assigned.not annotated	contig330.g122
1	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig355.g38
1	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig391.g20
1	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig391.g26
1	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig391.g28
1	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig391.g29
1	Protein homeostasis.proteolysis.metallopeptidase activities.carboxypeptidase activities.M28-class carboxypeptidase	contig457.g385
1	Protein homeostasis.proteolysis.metallopeptidase activities.carboxypeptidase activities.M28-class carboxypeptidase	contig458.g6
1	not assigned.not annotated	contig458.g8
1	Photosynthesis.photophosphorylation.photosystem II.PS-II complex.component PsbX	contig544.g1881
1	Photosynthesis.photophosphorylation.photosystem II.PS-II complex.component PsbX	contig544.g1970

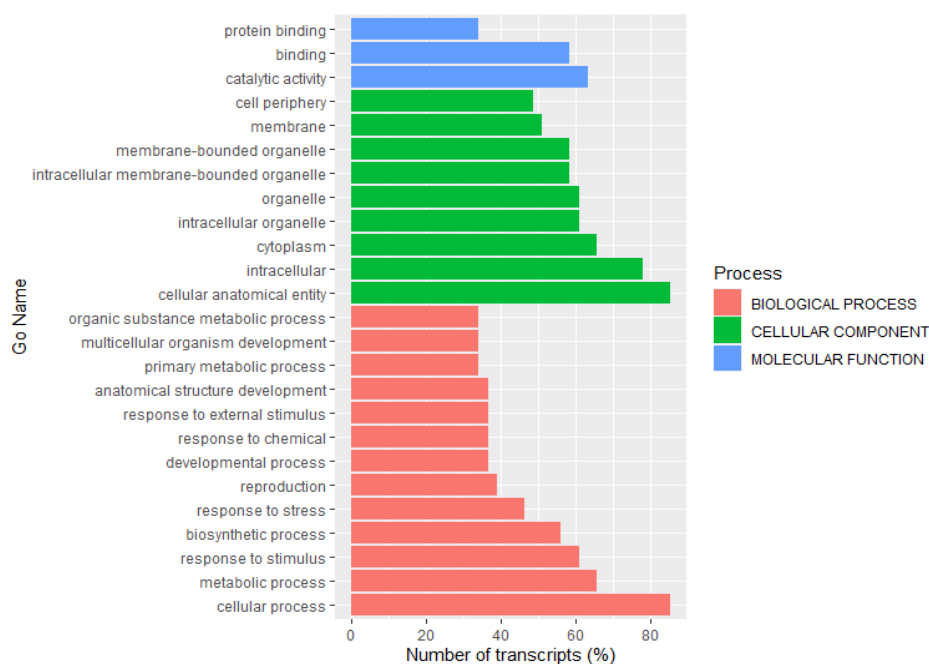
	Combined analysis of the three hardening accessions (continued)	
Cluster	Functional annotation	Contig
1	Cell wall organisation.cell wall proteins.hydroxyproline-rich glycoprotein activities.arabinogalactan-protein (AGP) activities.fasciclin-type AGP (FLA)	contig544.g2040
1	Photosynthesis.photophosphorylation.photosystem I.PS-I complex.component PsbF	contig549.g218
1	Enzyme classification.EC_2 transferases.EC_2.7 transferase transferring phosphorus-containing group	contig552.g93
1	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig553.g402
1	not assigned.not annotated	contig558.g1906
1	Enzyme classification.EC_3 hydrolases.EC_3.2 glycosylase	contig564.g28
1	Photosynthesis.photophosphorylation.photosystem II.LHC-II complex.component LHCb1/2/3	contig565.g52
1	not assigned.not annotated	contig570.g234
1	not assigned.not annotated	contig60.g17
1	Photosynthesis.photophosphorylation.photosystem I.LHC-I complex.component LHCA4	contig678.g379
1	Photosynthesis.photophosphorylation.photosystem I.LHC-I complex.component LHCA4	contig679.g24
1	not assigned.annotated	contig722.g1
1	Lipid metabolism.lipid degradation.triacylglycerol lipase activities.diacyl-/triacylglycerol lipase activities.lipase (LIP)	contig729.g9
1	Lipid metabolism.lipid degradation.triacylglycerol lipase activities.diacyl-/triacylglycerol lipase activities.lipase (LIP)	contig730.g9
1	Photosynthesis.calvin cycle.glyceraldehyde 3-phosphate dehydrogenase	contig760.g29
1	not assigned.not annotated	contig764.g413
1	Photosynthesis.photophosphorylation.photosystem II.LHC-related protein groups.three-helix LHC-related protein group.protein (ELIP)	contig89.g1873
1	not assigned.annotated	contig920.g331
2	Photosynthesis.photophosphorylation.photosystem II.photoprotection.non-photochemical quenching (NPQ).PsbS-dependent machinery.regulatory protein (PsbS)	contig25.g5
2	RNA biosynthesis.transcriptional regulation.transcription factor (TCP)	contig3.g405
2	not assigned.annotated	contig330.g32
2	Enzyme classification.EC_3 hydrolases.EC_3.4 hydrolase acting on peptide bond (peptidase)	contig456.g83
2	Solute transport.carrier-mediated transport.solute transporter (MTCC)	contig544.g105
2	Photosynthesis.photophosphorylation.photosystem II.photoprotection.non-photochemical quenching (NPQ).PsbS-dependent machinery.regulatory protein (PsbS)	contig546.g3084
2	Protein modification.phosphorylation.TKL protein kinase superfamily.protein kinase (LRR-XII)	contig557.g1157

	Combined analysis of the three hardening accessions (continued)	
Cluster	Functional annotation	Contig
2	RNA biosynthesis.transcriptional regulation.transcription factor (bHLH)	contig557.g843
2	not assigned.annotated	contig559.g40
2	Enzyme classification.EC_3 hydrolases.EC_3.4 hydrolase acting on peptide bond (peptidase)	contig588.g174
2	not assigned.annotated	contig588.g34
2	Enzyme classification.EC_1 oxidoreductases.EC_1.14 oxidoreductase acting on paired donor with incorporation or reduction of molecular oxygen	contig595.g148
2	RNA biosynthesis.transcriptional regulation.transcription factor (bHLH)	contig65.g48
2	RNA biosynthesis.transcriptional regulation.AP2/ERF transcription factor superfamily.transcription factor (ERF)	contig761.g12
2	RNA biosynthesis.transcriptional regulation.AP2/ERF transcription factor superfamily.transcription factor (ERF)	contig762.g1
2	not assigned.annotated	contig95.g785

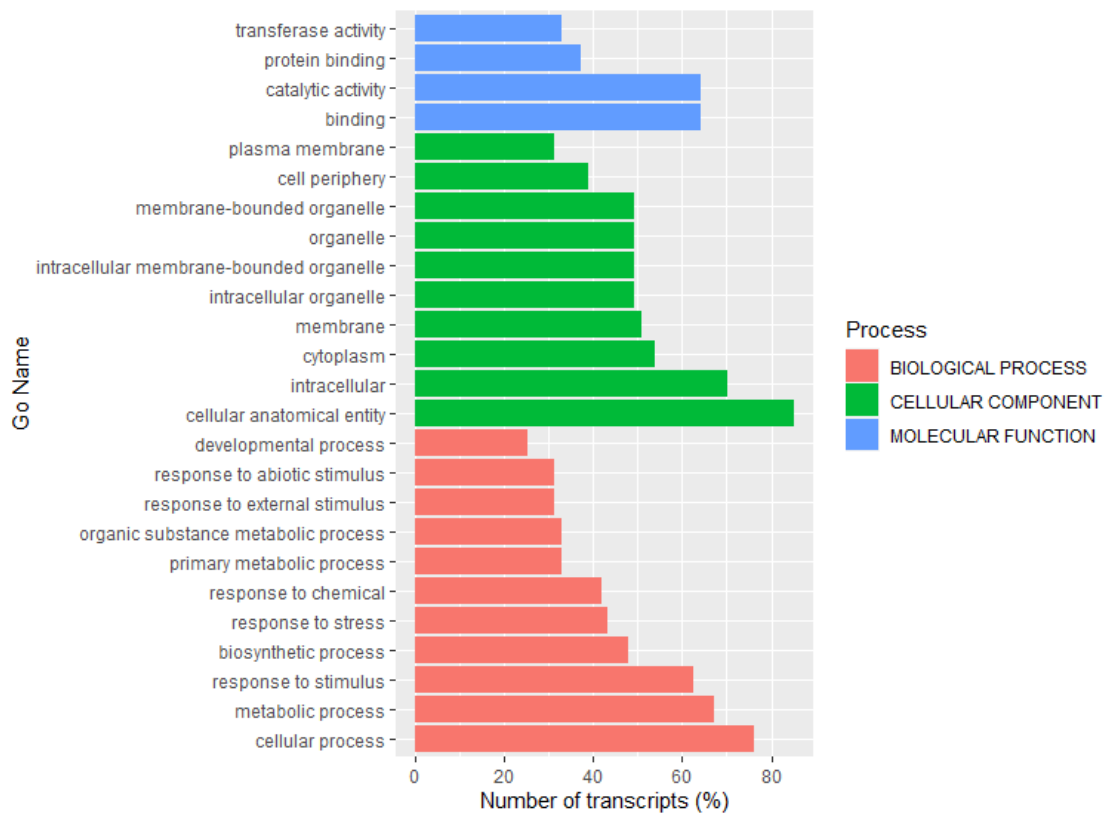
Supplementary material. Siadjeu et al. (2021)

Supplementary S7: Phylogenetic tree of candidate MYB genes in Bangou 1, Bayangam 2, Fonkouankem 1, and the combined analysis of the three hardening accessions (cf. external additional file)

Supplementary material. Siadjeu et al. (2021)

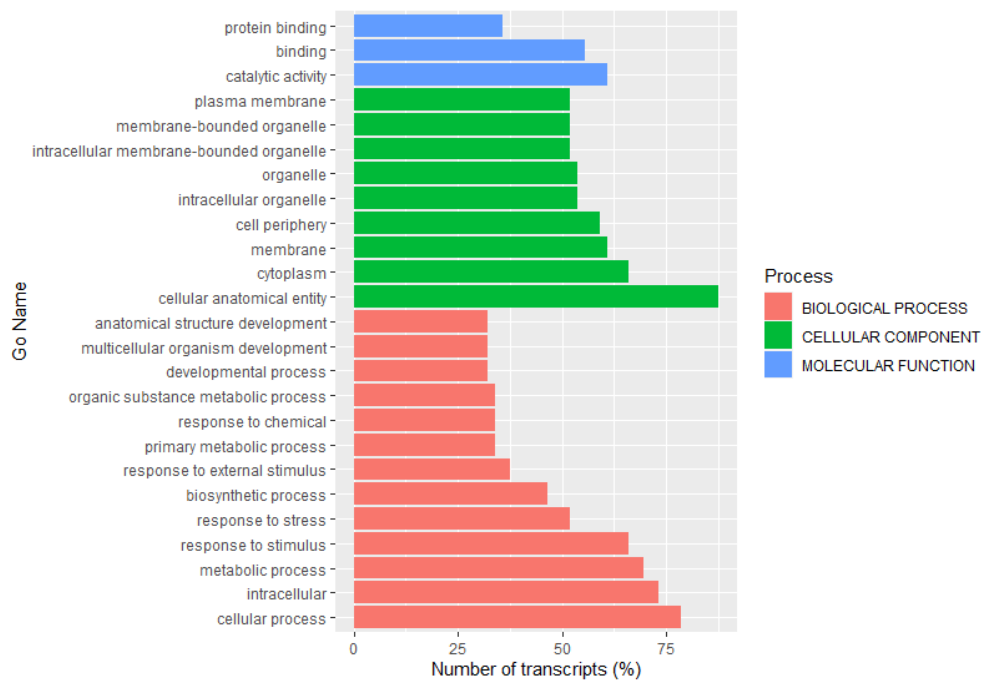


Bangou 1 vs. Ibo sweet 3 3DAHvsAH

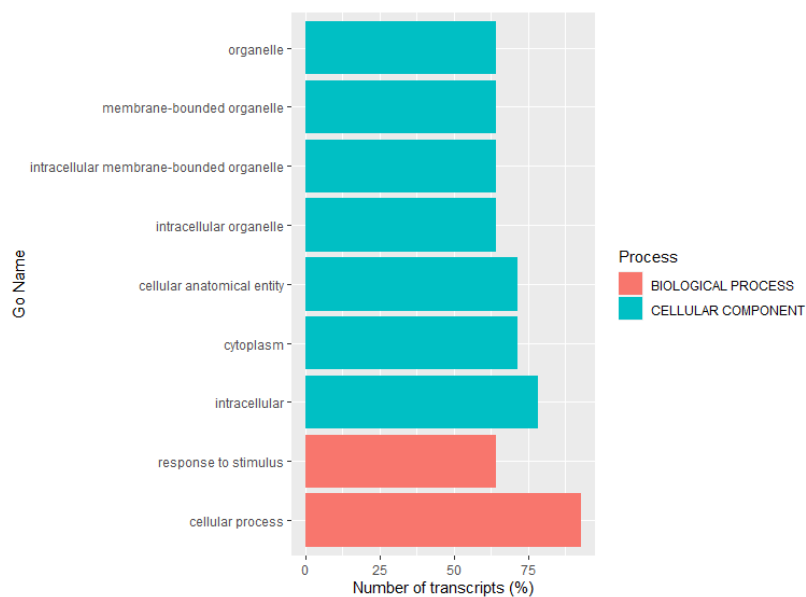


Bayangam 2 vs. Ibo sweet 3 3DAHvsAH

Supplementary S8: GO enrichment of up-regulated DEG 3DAH and 14DAH based on the comparison of hardening accessions against the non-hardening accession.

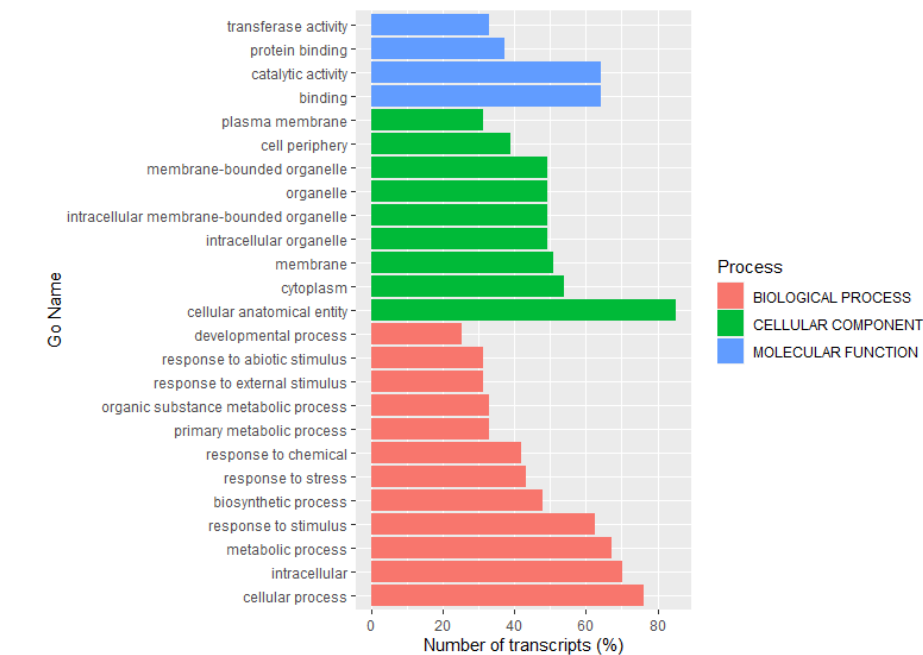


Fonkouankem 1 vs. Ibo sweet 3 3DAHvsAH

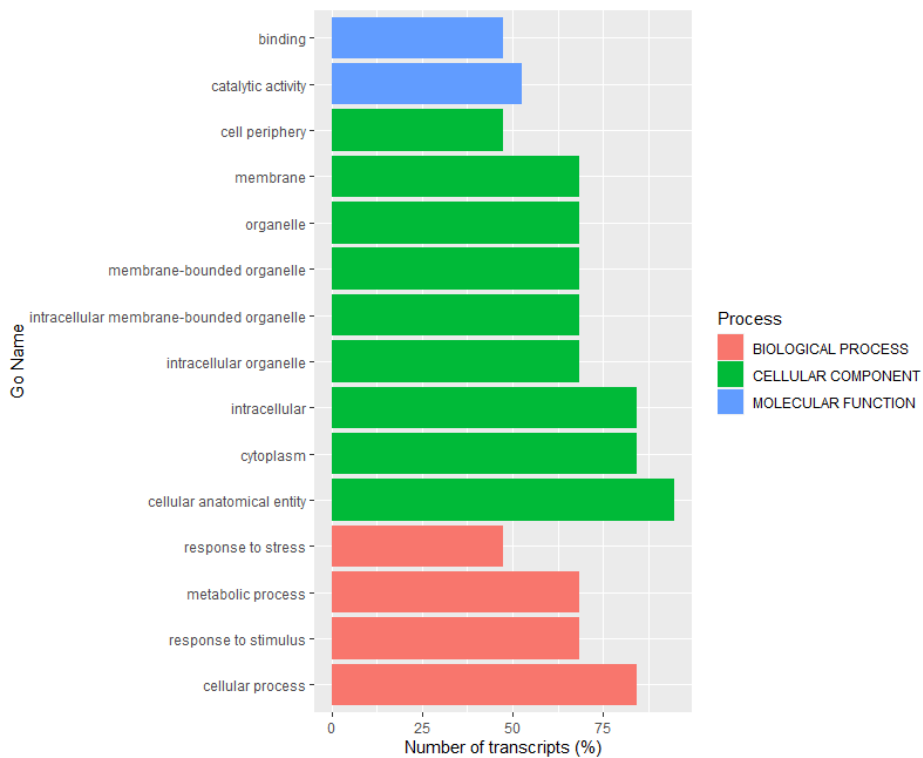


Bangou 1 vs. Ibo sweet 3 14DAHvsAH

Supplementary S8: GO enrichment of up-regulated DEG 3DAH and 14DAH based on the comparison of hardening accessions against the non-hardening accession.



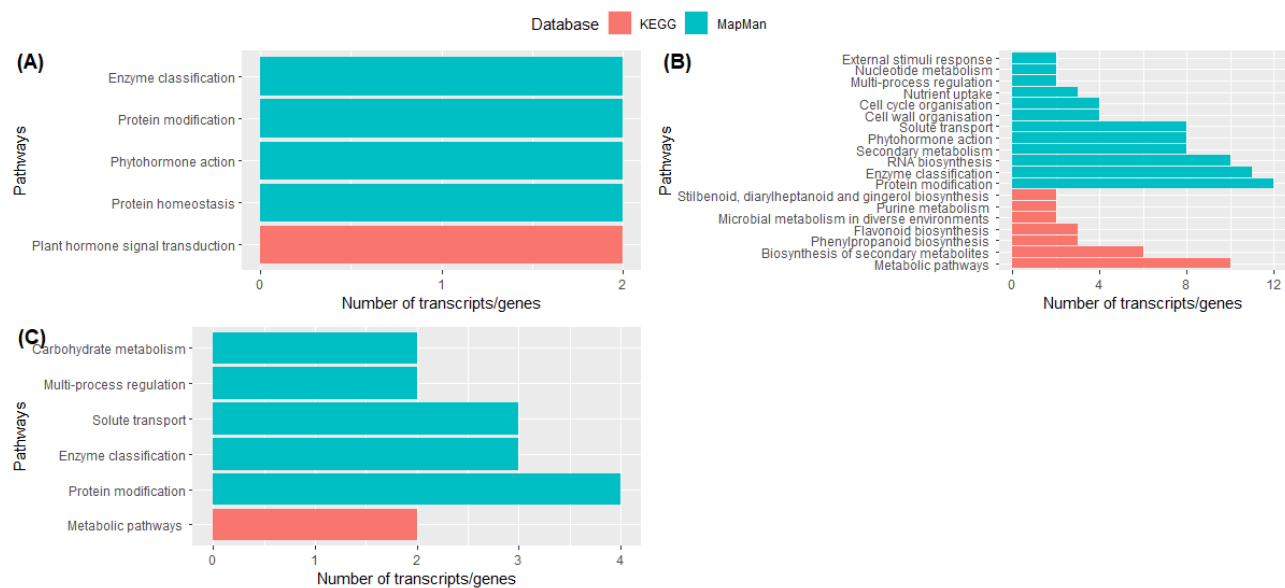
Bayangam 2 vs. Ibo sweet 3 14DAHvsAH



Fonkouankem 1 vs. Ibo sweet 3 14DAHvsAH

Supplementary S8: GO enrichment of up-regulated DEG 3DAH and 14DAH based on the comparison of hardening accessions against the non-hardening accession.

Supplementary material. Siadjeu et al. (2021)



Supplementary S9: Functional classification of up-regulated DEG 14DAH based on the comparison of hardening accessions against the non-hardening accession. (A), (B) and (C) the most enriched pathways 14 DAH in Bangou 1 vs. Ibo sweet 3, Bayangam 2 vs. Ibo sweet 3, and Fonkouankem 1 vs. Ibo sweet 3, respectively. Green bars represent pathway annotation with the MapMan database, and red bars represent pathway annotation with the KEGG database.

Supplementary material. Siadjeu et al. (2021)



Supplementary S10: Greening of young *D. dumetorum* tuber exposed to sunlight as opposed to the non-greening one

OVERALL CONCLUSION AND FUTURE PERSPECTIVES

My PhD thesis aimed to provide molecular genetic information for breeding *D. dumetorum* against the PHH. Unlike other yam species, *D. dumetorum* cultivation is limited by the PHH. My work was mainly concentrated on evaluating genetic diversity and population structure of *D. dumetorum* in Cameroon, releasing the first genome sequence of *D. dumetorum*, and finally identify genes involved in the PHH.

Phylogenetic and structure analysis revealed that *D. dumetorum* accessions in Cameroon form populations consisting of three clearly identified populations. High variability of *D. dumetorum* in Cameroon was reported. The Southwestern population was the most diverse. These results were obtained using genotyping by sequencing, which appears as a tool of choice to uncover intraspecific variability. All this information is of great importance for yams conservation. Intraspecific hybridization was detected among *D. dumetorum* and ploidy analysis showed that *D. dumetorum* in Cameroon was predominantly diploid. This constitutes useful resources for *D. dumetorum* improvement program. Although significant gene flows were observed between accessions, none was implying the non-hardening accession. Thus, there is a need to track the origin of the PHH in *D. dumetorum* and the ancestry of this accession. Hence, an extensive study on the phylogenetic of *D. dumetorum* across its distribution range is required.

The non-hardening accession represents a useful material for a breeding program for the PHH avoidance. Due to the outcrossing obligation and highly heterozygosity nature of yam, access to a complete well-annotated genome sequence is meaningful step towards among other genes identification. The first de novo nuclear genome sequence of *D. dumetorum* was reported using the non-hardening accession. Among protein encoding genes predicted, annotated genes, 68% were functional annotated. Almost ten thousand single orthologous gene copies were determined between *D. duemtorum* and *D. rotundata* genomes and could help to elucidate the evolutionary history of those species. Our genome sequence is high contiguity and phased to large extend and represents a reference genomic resource for identifying and developing molecular markers linked to interesting agronomical characters. Nevertheless, an improvement of the annotation and the assembly at the chromosome scale of *D. dumetorum* genome are still necessary.

The reference genome sequence was harnessed to identify putative genes involved in and controlling the PHH. Differential expressed and up regulated genes were detected after harvest, indicating that the PHH is likely controlled by genes. Gene ontology analysis of differential up regulated genes revealed that the PHH is a protecting mechanism of tubers against stress due to sun light and water losses. This mechanism seems to be triggered by sun light. Indeed, genes reacting to light excitation were found differential and up regulated. Hence, tubers activate this mechanism after exposure to sun light. Differential expressed and up regulated genes associated with photosynthesis, cell wall organization and transcription factor were mostly up regulated after harvest. In this regard, tuber reaction to sun light is related to genes (light-harvesting chlorophyll a/b binding proteins) involving in light capture during photosynthesis. Indeed, we found that *D. dumetorum* tubers become green when expose to sun light during growth. Photosynthesis, cell wall organization and transcription factor genes were co-expressed and up regulated 3 days after harvest, suggesting that the PHH likely appears three days after harvest. Whereas, genes encoding for protein modification were mostly up regulated from 3 to 14 days after harvest, impacting negatively the organoleptic qualities of hardening tubers. A putative mechanism of the PHH was suggested implying light-harvesting proteins, cell wall strengthening and lignification genes, and transcription factors from MYB family. Nonetheless, some genes were differentially expressed after harvest, but they were not annotated. Thus, it appears relevant to study in detail the physiology of the post-harvest of yam at the gene level.

In summary, a good contribution to molecular genetic of *D. dumetorum* which is the most neglected species among the main cultivated yam species was made. This paves the way to *D. dumetorum* improvement with respect to the PHH avoidance and other characters of interest. Thus, advanced breeding methods such as development of genomic markers, construction of genetic linkage maps to scan the genome the genome for quantitative trait loci (QTL) associated with genes conferring important characters could be now easily applicable on *D. dumetorum*.

REFERENCES

- Abraham, K., and Nair, P. G. (1990). Floral biology and artificial pollination in *Dioscorea alata* L. *Euphytica*, 48(1):45–51. doi: 10.1007/BF00028959
- Adaramola, T. F., Sonibare, M. A., Sartie, A., Lopez-Montes, A., Franco, J., and 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
- Aetion, D. A. R., and Ree, R. H. (2013). Inferring phylogeny and introgression using radseq data: an example from flowering plants (*Pedicularis*: *Orobanchaceae*). *PLoS One*, 62:689–706.
- Afoakwa, E. O., and Sefa-Dedeh, S. (2001). Chemical composition and quality changes occurring in *Dioscorea dumetorum* pax tubers after harvest. *Food Chemistry*, 75(1): 85–91. doi: 10.1016/S0308-8146(01)00191-1
- Agbor-Egbe, T., and Treche, S. (1995). Evaluation of the chemical composition of cameroonian yam germplasm. *Journal of Food Composition and Analysis*, 8(3):274–283. doi: 10.1006/jfca.1995.1020
- Alozie Y. E., Akpanabiatu M., Eyong E. U., Umoh I. B., Alozie G. (2009). Amino Acid Composition of *Dioscorea dumetorum* Varieties. *Pak. J. Nutr.*, 8:103–105.
- Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, 215:403–410. doi: 10.1016/S0022-2836(05)80360-2.
- Alves-Carvalho, S., Aubert, G., Carrère, S., Cruaud, C., Brochot, A. L., Jacquin, F., ... Burstin, J. (2015). Full-length de novo assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. *Plant Journal*, 84(1):1–19. doi: 10.1111/tpj.12967
- Ambawat, S., Sharma, P., Yadav, N. R., and Yadav, R. C. (2013). MYB transcription factor genes as regulators for plant responses: An overview. *Physiology and Molecular Biology of Plants*, 19(3):307–321. doi: 10.1007/s12298-013-0179-1
- Arcuri, M. L. C., Fialho, L. C., Vasconcellos Nunes-Laitz, A., Fuchs-Ferraz, M. C. P., Wolf, I. R., Valente, G. T., ... Maia, I. G. (2020). Genome-wide identification of multifunctional laccase gene family in *Eucalyptus grandis*: potential targets for lignin engineering and stress tolerance. *Trees - Structure and Function*, 34(3):745–758. doi: 10.1007/s00468-020-01954-3
- Arnau, G., Bhattacharjee, R., Mn S., Malapa R., Lebot V., Abraham K., ... Pavis C. (2017). Understanding the genetic diversity and population structure of yam (*Dioscorea alata* L.) using microsatellite markers. *PLoS One*, 1–17.
- Arnau, G., Nemorin, A., Maledon, E., and Abraham, K. (2009). Revision of ploidy status of *Dioscorea alata* L. (*Dioscoreaceae*) by cytogenetic and microsatellite segregation analysis. *Theoretical and Applied Genetics*. doi: 10.1007/s00122-009-0977-6
- Asiedu R., Ng S. Y. C., Bai K. V., Ekanayake I. J., and Wanyera N. M. W. (1998). Genetic improvement. In: Orkwor G. C., Asiedu R., and Ekanayake I. J., editors, *Food yams: Advances in research*. IITA and NRCRI, Nigeria, pp. 63–104
- Asiedu, R., Mignouna, H., Odu, B., and Hughes, J. A. (2003). Yam breeding. *Plant Virology in Sub-Saharan Africa, Processing of Conference Organized by IITA*, pp. 466–475.

- Asiedu, R., Wanyera, N., Ng, S.Y.C., and Ng, N.Q. (1997). Yams, In D. Fuccillo, L. Sears and P. Stapleton, Biodiversity in trust: conservation and use of plant genetic resources in CGIAR centres, Cambridge, UK: Cambridge University Press, pp. 57-66.
- Baasner J. S., Howard D., Pucker B. (2019). Influence of neighboring small sequence variants on functional impact prediction. *bioRxiv*. doi: 10.1101/596718.
- Bai, K.V., Ekanayake, I. J. (1998). Taxonomy, morphology and floral biology of yams. In R. A. and I. J. E. Orkwor, G. c. (Ed.), Food Yams: Advances in research, African Bo, pp. 13–37.
- Barton H. (2014). Yams: origins and development. *Encyclopaedia Glob. Archaeol.* 2014:7943–7947.
- Bautista O. K., 1990. Postharvest technology for southeast asian perishable crops. (Ed.) Technology and Livelihood Resource Center, Manila, Philippines, 302 p.
- Becerra-Moreno, A., Redondo-Gil, M., Benavides, J., Nair, V., Cisneros-Zevallos, L., and Jacobo-Velázquez, D. A. (2015). Combined effect of water loss and wounding stress on gene activation of metabolic pathways associated with phenolic biosynthesis in carrot. *Frontiers in Plant Science*, 6:1-15. doi: 10.3389/fpls.2015.00837
- Bhattacharjee, R., Gedil, M., Sartie, A., Otoo, E., Dumet, D., Kikuno, H., ... Asiedu, R. (2011). *Dioscorea*. Wild Crop Relatives: Genomic and Breeding Resources, Industrial Crops., pp. 71–96. doi: 10.1007/978-3-642-21102-7_4
- Burns, M., Hedin, M., Tsurusaki, N. (2018). Population genomics and geographical parthenogenesis in Japanese harvestmen (*Opiliones*, *Sclerosomatidae*, *Leiobunum*). *Ecol. Evol.*, 8:36–52.
- Chan, P. P., Lowe, T. M. (2019). tRNAscan-SE: searching for tRNA genes in genomic sequences. In: Kollmar M., editor. *Gene Prediction: Methods and Protocols*. Volume 1962. Springer; Berlin/Heidelberg, Germany, pp. 1–14.
- Cheng, C. Y., Krishnakumar, V., Chan A., Thibaud-Nissen, F., Schobel, S., Town, C. D. (2017). Araport11: A complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.*, 89:789–804. doi: 10.1111/tpj.13415.
- Cormier, F., Lawac, F., Maledon, E., Gravillon, M. C., Nudol, E., Mournet, P., Vignes, H., Chaïr, H., Arnau, G. (2019). A reference high-density genetic map of greater yam (*Dioscorea alata* L.) *Theor. Appl. Genet.*, 132:1733–1744. doi: 10.1007/s00122-019-03311-6.
- Coursey, D. G. (1967). Yams: an account of the nature, origins, cultivation and utilisation of the useful Members of the *Dioscoreaceae*. In London.
- Crean, D. E. C., and Haisman, D. R. (1963). The interaction between phytic acid and divalent cations during the cooking of dried peas. *Journal of the Science of Food and Agriculture*, 14(11):824–833. doi: 10.1002/jsfa.2740141109
- DaCosta, J. M., Sorenson, M. D. (2016). DdRAD-seq phylogenetics based on nucleotide, indel, and presence-absence polymorphisms: Analyses of two avian genera with contrasting histories. *Mol Phylogenet Evol.*, 94:122–35.
- Dansi, A., Mignouna, H. D., Pillay, M., and Zok, S. (2001). Ploidy variation in the cultivated yams (*Dioscorea cayenensis*-*Dioscorea rotundata* complex) from Cameroon as determined by flow cytometry. *Euphytica*, 119(3):301–307.
- Dao, T. T. H., Linthorst, H. J. M., and Verpoorte, R. (2011). Chalcone synthase and its functions in plant resistance. *Phytochemistry Reviews*, 10(3):397–412. doi: 10.1007/s11101-011-9211-7

- Darkwa, K., Olasanmi, B., Asiedu, R., and Asfaw, A. (2020). Review of empirical and emerging breeding methods and tools for yam (*Dioscorea* spp.) improvement: Status and prospects. *Plant Breeding*, 139(3): 474–497. doi: 10.1111/pbr.12783
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet.*, 12:499–510.
- De Donato, M., Peters, S. O., Mitchell, S. E., Hussain, T., Imumorin, I. G. (2013). Genotyping-by-Sequencing (GBS): A novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *PLoS One.*, 8:1-9.
- Degras, L. (1993). The yam: a tropical root crop. In *The yam: a tropical root crop*. London: Macmillan Press Ltd.
- Deschamps, S., Llaca, V., and May, G. D. (2012). Genotyping-by-Sequencing in plants. *Biology (Basel)*, 1:460–83.
- Dolezel, J., Greilhuber, J., Lucretti, S., Meister, A., Lysak, M. A., Nardi, L., Obermayer, R. (1998). Plant genome size estimation by flow cytometry: Inter-laboratory comparison. *Ann Bot.*, 82:17–26.
- Dumont, R., Hamon, P., and Seignobos, C. (1994). Les ignames au Cameroun. CIRAD-CA, (Repères: CIRAD, 3), 80 p.
- Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour.*, 4:359–61.
- Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5:1–19.
- Egesi, C. N., Asiedu, R., Egunjobi, J. K., Bokanga, M. (2003). Genetic diversity of organoleptic properties in water yam (*Dioscorea alata* L.). *J. Sci. Food Agric.*, 83:858–65.
- Egesi, C. N., Pillay, M., Asiedu, R., and Egunjobi, J. K. (2002). Ploidy analysis in water yam, *Dioscorea alata* L. germplasm. *Euphytica*, 128(2):225–230. doi: 10.1023/A:1020868218902
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., ... Mitchell S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 6:1–10.
- Emms, D. M., Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.*, 20:238. doi: 10.1186/s13059-019-1832-y.
- Evanno, G., Regnaut, S., Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol Ecol.*, 14:2611–20.
- FAOSTAT, (2018). Food and Agriculture Organization of the United Nations Statistics database, FAOSTAT. Retrieved from <http://www.fao.org/faostat/en/#data/QC>
- Ferede, R., Maziya-Dixon, B., Alamu, O. E., and Asiedu, R. (2010). Identification and quantification of major carotenoids of deep yellow-fleshed yam (tropical *Dioscorea dumetorum*). *Journal of Food, Agriculture and Environment*, 8(3-4):160–166.
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., and Smit, A.F. (2019). RepeatModeler2: Automated genomic discovery of transposable element families. *bioRxiv.*, doi: 10.1101/856591.
- Gamiette, F., Bakry, F., and Ano, G. (1999). Ploidy determination of some yam species (*Dioscorea* spp.) by flow cytometry and conventional chromosomes counting. *Genetic Resources and Crop Evolution*, 46(1):19–27. doi: 10.1023/A:1008649130567

- Garcia, E., Filisetti, T. M. C. C., Udaeta, J. E. M., and Lajolo, F. M. (1998). Hard-To-Cook Beans (*Phaseolus vulgaris*): Involvement of Phenolic Compounds and Pectates. *Journal of Agricultural and Food Chemistry*, 46(6):2110–2116. doi: 10.1021/jf970848f
- Girardin, O. (1996). Technologies après-récolte de l'igname : Etude de l'amélioration du stockage traditionnel en Côte d'Ivoire. Doctoral thesis, ETH Zürich, Zwitterland, 129 p.
- Girma, G., Hyma, K. E., Asiedu, R., Mitchell, S. E., Gedil, M., and Spillane, C. (2014). Next-generation sequencing based genotyping, cytometry and phenotyping for understanding diversity and evolution of guinea yams. *Theoretical and Applied Genetics*, 127(8):1783–1794. doi: 10.1007/s00122-014-2339-2
- Gompert, Z., and Mock, K. E. (2017). Detection of individual ploidy levels with genotyping-by-sequencing (GBS) analysis. *Mol Ecol Resour.*, 17:1156–67.
- Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., ... Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, 36(10):3420–3435. doi: 10.1093/nar/gkn176
- Govaerts R., Wilkin P., Saunders R. M. K. (2007). *World Checklist of Dioscoreales, Yams and their Allies*. Kew Publishing, Royal Botanic Gardens, Kew, UK.
- Guo, H., Wang, Y., Wang, L., Hu, P., Wang, Y., Jia, Y., ... Yang, C. (2017). Expression of the MYB transcription factor gene BplMYB46 affects abiotic stress tolerance and secondary cell wall deposition in *Betula platyphylla*. *Plant Biotechnology Journal*, 15(1), 107–121. doi: 10.1111/pbi.12595
- Hamon, P., Brizard, J. P., Zoundjihékpon, J., Duperray, C., and Borgel, A. (1992). Étude des index d'ADN de huit espèces d'ignames (*Dioscorea* sp.) par cytométrie en flux. *Canadian Journal of Botany*, 70(5):996–1000. doi: 10.1139/b92-123
- Hamon, P., Dumont, R., Zoundjihékpon, J., Ahoussou, N., Tio-Touré B. (1997). Les ignames. In: *L'amélioration des plantes tropicales*. (Eds.) Charrier A, Jacqot M, Hamon S. and Nicolas D. CIRAD-ORSTOM, Paris, pp. 385-400.
- Hamon, P., Dumont, R., Zoundjihekpon, J., Tio-Toure, B. and Hamon, S. (1995). Les ignames sauvages d'Afrique de L'Ouest. *Caractères Morphologiques*. ORSTOM-Editions, Paris, 81 p.
- Henderson, H. M., and Ankrah, S. A. (1985). The relationship of endogenous phytase, phytic acid and moisture uptake with cooking time in *Vicia faba* minor cv. Aladin. *Food Chemistry*, 17(1):1–11. doi: 10.1016/0308-8146(85)90087-1
- Hochu, I., Santoni, S., and Bousalem, M. (2006). Isolation, characterization and cross-species amplification of microsatellite DNA loci in the tropical American yam *Dioscorea trifida*. *Molecular Ecology Notes*, 6(1):137–140.
- Hui-Chen, C., Mei-Chen, C., Ping-Ping, L., Chih-Tsun, T., Fang-Ping, D. (1985). A cytotoxic study on Chinese *Dioscorea*, L.—The chromosome numbers and their relation to the origin and evolution of the genus. *J. Syst. Evol.*, 23:11–18.
- Huson, D. H., and Bryant D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.*, 23:254–67.
- IRAD, (2008). Second report on the state of plant genetic resources for food and agriculture in Cameroon. Institute of Agricultural Research and Development (IRAD), Yaoundé, Available from: <http://www.fao.org/docrep/013/i1500e/Cameroun.pdf>

- Iwu, M. M., Okunji, C., Akah, P., Corley, D., and Tempesta, S. (1990). Hypoglycaemic Activity of dioscoretine from tubers of *Dioscorea dumetorum* in normal and alloxan diabetic rabbits. *Planta Med.*, 56(7990):264–267.
- Janiak, A., Kwaśniewski, M., and Szarejko, I. (2016). Gene expression regulation in roots under drought. *J Exp Bot.*, 67:1003–14.
- Jones P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., ... Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30:1236–1240.
- Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., Eddy, S. R., Bateman, A., Finn, R. D., Petrov A. I. (2018). Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.*, 46:D335–D342. doi: 10.1093/nar/gkx1038.
- Keller, O., Kollmar, M., Stanke, M., Waack, S. (2011). A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics*, 27:757–763. doi: 10.1093/bioinformatics/btr010.
- Knuth, R. (1924). *Dioscoreaceae*. In: Engelm A., editor. *Das Pflanzenreich*. Engelmann, W.; Leipzig, Germany.
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.*, 27:722–736. doi: 10.1101/gr.215087.116.
- Kwoseh, C. K. (2000). Identification of resistance to major nematode pest of yams (*Dioscorea* spp.) in West Africa. University of Reading, UK.
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359. doi: 10.1038/nmeth.1923
- Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., ... Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Computational Biology*, 9(8):1–10. doi: 10.1371/journal.pcbi.1003118
- Le Gall, H., Philippe, F., Domon, J. M., Gillet, F., Pelloux, J., and Rayon, C. (2015). Cell wall metabolism in response to abiotic stress. *Plants*, 4(1):112–166. doi: 10.3390/plants4010112
- Lebot, V. (2009). Tropical root and tuber crops: cassava, sweet potato, yams and aroids. *Crop Production Science in Horticulture* no. 17, CABI Publishing, UK, 413 p.
- Leitch, I. J., Chase, M. W., Bennett, M. D. (1998). Phylogenetic Analysis of DNA C-values provides evidence for a small Ancestral genome size in flowering Plants. *Ann Bot.*, 82:85–94.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 20131303.3997
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34:3094–3100. doi: 10.1093/bioinformatics/bty191.
- Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H., Li, N., Li, Z., Chen, Y., Mu, D., Fan, W. (2013). Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *arXiv*. 20131308.2012
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):1–21. doi: 10.1186/s13059-014-0550-8

- Lowe, T. M., Eddy, S. R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, 25:955–964. doi: 10.1093/nar/25.5.955.
- Magwe-Tindo, J., Wieringa, J. J., Sonke, B., Zapfack, L., Vigouroux, Y., Couvreur, T. L. P., Scarcelli, N. (2019). Complete plastome sequences of 14 African yam species (*Dioscorea* spp.) Mitochondrial DNA Part B-Resour., 4:74–76. doi: 10.1080/23802359.2018.1536466.
- Malapa, R., Arnau, G., Noyer, J. L., and Lebot, V. (2005). Genetic diversity of the greater yam (*Dioscorea alata* L.) and relatedness to *D. nummularia* Lam. and *D. transversa* Br. as revealed with AFLP markers. *Genetic Resources and Crop Evolution*, 52(7):919–929. doi: 10.1007/s10722-003-6122-5
- Marcais, G., Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27:764–770.
- Mbome Lape, I., Treche, S. (1994). Nutritional quality of yam (*Dioscorea dumetorum* and *D. rotundata*) flours for growing rats. *Journal of the Science of Food and Agriculture*, 66(4):447–455. doi: 10.1002/jsfa.2740660405
- Mbome Lape, I. (1991). Utilisation de l'igname *Dioscorea dumetorum* comme base des bouillies de sevrage. In : Les bouillies de sevrage en Afrique centrale. Brazzaville : ORSTOM, 16 p.
- Medoua, G. N. (2005). Potentiels nutritionnel et technologique des tubercules durcis de l'igname *Dioscorea dumetorum* (Kunth). Doctoral thesis, Ngaoundere University, Cameroon.
- Medoua, G. N., and Mbofung, C. M. F. (2006). Hard-to-cook defect in trifoliate yam *Dioscorea dumetorum* tubers after harvest. *Food Research International*, 39(5):513–518. doi: 10.1016/j.foodres.2005.10.005
- Medoua, G. N., Mbome, I. L., Agbor-Egbe, T., and Mbofung, C. M. F. (2005). Study of the hard-to-cook property of stored yam tubers (*Dioscorea dumetorum*) and some determining biochemical factors. *Food Research International*, 38(2), 143–149. doi: 10.1016/j.foodres.2004.09.005
- Medoua, G. N., Mbome, I. L., Agbor-Egbe, T., and Mbofung, C. M. F. (2008). Influence of fermentation on some quality characteristics of trifoliate yam (*Dioscorea dumetorum*) hardened tubers. *Food Chemistry*, 107(3):1180–1186. doi: 10.1016/j.foodchem.2007.09.047
- Medoua, G. N., Mbome, I. L., Agbor-Egbe, T., Mbofung, C. M. F. (2005). Physicochemical changes occurring during postharvest hardening of trifoliate yam (*Dioscorea dumetorum*) tubers. *Food Chem.*, 90:597–601.
- Medoua, G. N., Mbome, I. L., Egbe, T. A., Mbofung, C. M. F. (2007). Salts soaking treatment for improving the textural and functional properties of trifoliate yam (*Dioscorea dumetorum*) hardened tubers. *J Food Sci.*, 72:464–9.
- Michael, T. P., Jupe, F., Bemm, F., Motley, S. T., Sandoval, J. P., Lanz, C., Loudet, O., Weigel, D., Ecker, J. R. (2018). High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nat. Commun.*, 9:541. doi: 10.1038/s41467-018-03016-2.
- Miège, J. (1954). Nombres chromosomiques et répartition géographique de quelques plantes tropicales et équatoriales. *ERev. Cytol. et Biol. Vég.*, 15(4):312–348.
- Mignouna, H. D., Abang, M. M., and Asiedu, R. (2003b). Harnessing modern biotechnology for tropical tuber crop improvement: Yam (*Dioscorea* spp.) molecular breeding. *African Journal of Biotechnology*, 2(12):475–485. doi: 10.5897/AJB2003.000-1097

- Mignouna, H. D., Abang, M. M., and Fagbemi, S. A. (2003a). A comparative assessment of molecular marker assays (AFLP, RAPD and SSR) for white yam (*Dioscorea rotundata*) germplasm characterisation. *Annals of Applied Biology*, 142(3):269–276. doi: 10.1111/j.1744-7348.2003.tb00250.x
- Mignouna, H. D., Abang, M. M., Asiedu, R. (2008). Genomics of yams, a common source of food and medicine in the tropics. In: Moore P.H., Ming R., editors. *Genomics of Tropical Crop Plants*. Springer; New York, NY, USA, pp. 549–570.
- Mignouna, H. D., Abang, M. M., Onasanya, A., Agindotan, B., Asiedu, R. (2002a). Identification and potential use of RAPD markers linked to Yam mosaic virus resistance in white yam (*Dioscorea rotundata* Poir.). *Ann. Appl. Biol.* 140:163–169
- Mignouna, H. D., Mank, R. A., Ellis, T. H. N., Van den Bosch, N., Asiedu, R., Ng, S. Y. C., and Peleman, J. (2002b). A genetic linkage map of Guinea yam (*Dioscorea rotundata* Poir.) based on AFLP markers. *Theoretical and Applied Genetics*, 105(5):716–725. doi: 10.1007/s00122-002-0911-7
- Mizuki, I., Tani, N., Ishida, K., and Tsumura, Y. (2005). Development and characterization of microsatellite markers in a clonal plant, *Dioscorea japonica* Thunb. *Molecular Ecology Notes*, 5(4):721–723. doi: 10.1111/j.1471-8286.2005.01020.x
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS : an automatic genome annotation and pathway reconstruction server. 35:182–185. doi: 10.1093/nar/gkm321
- Nawrocki, E. P., Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, 29:2933–2935. doi: 10.1093/bioinformatics/btt509.
- Nemorin, A. (2012). Acquisition de connaissances sur la génétique de l'espèce *Dioscorea alata* L. pour la production de variétés polyploïdes. Doctoral thesis, Antilles and Guyana University, Guadeloupe.
- Nemorin, A., David J., Maledon, E., Nudol, E., Dalon, J., Arnau, G. (2013). Microsatellite and flow cytometry analysis to help understand the origin of *Dioscorea alata* polyploids. *Ann Bot.*, 112:811–9.
- Ngo Ngwe, M. F., Omokolo, D. N., Joly, S. (2015). Evolution and Phylogenetic Diversity of Yam Species (*Dioscorea* spp.): Implication for Conservation and Agricultural Practices. *PLoS ONE*, 10:e0145364. doi: 10.1371/journal.pone.0145364.
- Nimenibo-Uadia, R., Oriakhi, A. (2017). Proximate, mineral and phytochemical composition of *Dioscorea dumetorium* Pax. *J. Appl. Sci. Environ. Manag.*, 21:771–774.
- Obidiegwu, J. E., Rodriguez, E., Loureiro, J., Muoneke, C. O., Asiedu, R. (2009). Estimation of the nuclear DNA content in some representative of genus *Dioscorea*. *Sci Res Easay*, 4:448–52.
- Onwueme, I.C. (1978). *The tropical tuber crops—yams, cassava, sweet potato and cocoyams*. John Wiley and Sons, Chichester, 3-97.
- Opara, L. (2003). Yams post-harvest operations. (ed.) FAO. *African Journal of Root and Tuber Crops*, 22.
- Paajanen, P., Kettleborough, G., Lopez-Girona, E., Giolai, M., Heavens, D., Baker, D., Lister, A., Cugliandolo, F., Wilde, G., Hein, I., ... Clark, M. D. (2019). A critical comparison of technologies for a plant genome sequencing project. *Gigascience*, 8:giy163. doi: 10.1093/gigascience/giy163.

- Petro, D., Onyeka, T. J., Etienne, S., and Rubens, S. (2011). An intraspecific genetic map of water yam (*Dioscorea alata* L.) based on AFLP markers and QTL analysis for anthracnose resistance. *Euphytica*, 179:405–416, doi: 10.1007/s10681-010-0338-1
- Pickrell, J. K., Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.*, 8(11).
- Polko, J. K., and Kieber, J. J. (2019). The regulation of cellulose biosynthesis in plants. *Plant Cell*, 31(2):282–296. doi: 10.1105/tpc.18.00760
- 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. *Sci. Rep.*, 6:29136. doi: 10.1038/srep29136.
- Pritchard, J. K., Stephens, M., Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155:945–59.
- Pucker, B. (2019). Mapping-based genome size estimation. *bioRxiv*. doi: 10.1101/607390.
- Pucker, B., Brockington, S. F. (2018). Genome-wide analyses supported by RNA-Seq reveal non-canonical splice sites in plant genomes. *BMC Genomics*, 19:980. doi: 10.1186/s12864-018-5360-z.
- Pucker, B., Feng, T., Brockington, S. (2019). Next generation sequencing to investigate genomic diversity in Caryophyllales. *bioRxiv*, doi: 10.1101/646133.
- Pucker, B., Holtgräwe, D., Rosleff Sørensen T., Stracke, R., Viehöver, P., Weisshaar, B. (2016). A *de novo* genome sequence assembly of the *Arabidopsis thaliana* accession Niederzenz-1 displays presence/absence variation and strong synteny. *PLoS ONE*, 11:e0164321. doi: 10.1371/journal.pone.0164321.
- Pucker, B., Holtgräwe, D., Stadermann, K. B., Frey, K., Huettel, B., Reinhardt, R., Weisshaar B. (2019). A chromosome-level sequence assembly reveals the structure of the *Arabidopsis thaliana* Nd-1 genome and its gene set. *PLoS ONE*, 14:e0216233. doi: 10.1371/journal.pone.0216233.
- Pucker, B., Holtgräwe, D., Weisshaar, B. (2017). Consideration of non-canonical splice sites improves gene prediction on the *Arabidopsis thaliana* Niederzenz-1 genome sequence. *BMC Res. Notes*, 10:667. doi: 10.1186/s13104-017-2985-y.
- Pucker, B., Ruckert, C., Stracke, R., Viehover, P., Kalinowski, J., Weisshaar, B. (2019). Twenty-five years of propagation in suspension cell culture results in substantial alterations of the *Arabidopsis thaliana* genome. *Genes*, 10:671. doi: 10.3390/genes10090671.
- Rabbi, I. Y., Kulakow, P. A., Manu-Aduening, J. A., Dankyi, A. A., Asibuo, J. Y., Parkes, E. Y., ... Maredia, M. K. (2015). Tracking crop varieties using genotyping-by-sequencing markers: A case study using cassava (*Manihot esculenta* Crantz). *BMC Genetics*, 16:1–11.
- Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol.*, 5:94–100.
- Ramachandran, K. (1968). Cytological studies in *Dioscoreaceae*. *Cytologia* 33: 401–410.
- Ramsey, J., Schemske, D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu Rev Ecol Syst.*, 29:467–501.
- Ranallo-Benavidez, T. R., Jaron, K. S., Schatz, M. C. (2019). GenomeScope 2.0 and Smudgeplots: Reference-free profiling of polyploid genomes. *bioRxiv*. 2019:747568. doi: 10.1038/s41467-020-14998-3.

- Ravi, V., Aked, J., and Balagopalan, C. (1996). Review on tropical root and tuber crops. I. Storage methods and quality changes. In *Critical reviews in food science and nutrition* (Vol. 36). doi: 10.1080/10408399609527744
- Reyes-Moreno, C., Rouzaud-Sandez, O., Milán-Carrillo, J., Garzón-Tiznado, J. A., and Camacho-Hernández, L. (2001). Hard-to-cook tendency of chickpea (*Cicer arietinum* L.) varieties. *Journal of the Science of Food and Agriculture*, 81(10):1008–1012. doi: 10.1002/jsfa.872
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140. doi: 10.1093/bioinformatics/btp616
- Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, 4:e2584.
- Rosso, M. G., Li, Y., Strizhov, N., Reiss, B., Dekker, K., Weisshaar, B. (2003). An *Arabidopsis thaliana* T-DNA mutagenised population (GABI-Kat) for flanking sequence tag based reverse genetics. *Plant Mol. Biol.*, 53:247–259.
- Ruggieri, V., Alexiou, K. G., Morata, J., Argyris, J., Pujol, M., Yano, R., Nonaka, S., Ezura, H., Latrasse, D., Boualem, A., ..., Garcia-Mas, J. (2018). An improved assembly and annotation of the melon (*Cucumis melo* L.) reference genome. *Sci. Rep.*, 8:8088. doi: 10.1038/s41598-018-26416-2.
- Salmela, L., Walve, R., Rivals, E., Ukkonen, E. (2017). Accurate self-correction of errors in long reads using de Bruijn graphs. *Bioinformatics*, 33:799–806.
- Saski, C. A., Bhattacharjee, R., Scheffler, B. E., Asiedu, R. (2015). Genomic resources for water yam (*Dioscorea alata* L.): Analyses of EST-sequences, de novo sequencing and GBS libraries. *PLoS One*, 10(7).
- Sattler, M. C., Carvalho, C. R., Clarindo, W. R. (2016). The polyploidy and its key role in plant breeding. *Planta*, 243:281–96.
- Sauphanor B., Bordat, D., Delvare, G., and Ratnadass, A. (1987). Les insectes des ignames stockées de Côte d'Ivoire. *Inventaire faunistique et éléments biologiques. L'Agronomie Tropicale* 42(4):305-312.
- Schwacke, R., Ponce-Soto, G. Y., Krause, K., Bolger, A. M., Arsova, B., Hallab, A., ... Usadel, B. (2019). MapMan4: A Refined Protein Classification and Annotation Framework Applicable to Multi-Omics Data Analysis. *Molecular Plant*, 12(6):879–892. doi: 10.1016/j.molp.2019.01.003
- Sealy, L., Renaudin, S., Gallant, D. J., Bouchet, B., and Brillouet, J. M. (1985). Ultrastructural study of yam tuber as related to postharvest hardness. *Food Microstructure*, 4(1).
- Sefa-Dedeh, S., Afoakwa, E.O. (2002). Biochemical and textural changes in trifoliate yam *Dioscorea dumetorum* tubers after harvest. *Food Chem.*, 79:27–40. doi: 10.1016/S0308-8146(02)00172-3.
- Shehata, A. M. E. T. (1992). Hard-to-cook phenomenon in legumes. In *Food Reviews International* (Vol. 8). doi: 10.1080/87559129209540938
- Siadjeu, C., Mahbou, S. T. G., Bell, J. M., and Nkwate, S. (2015). Genetic diversity of sweet yam *Dioscorea dumetorum* (Kunth) Pax revealed by morphological traits in two agro-ecological zones of Cameroon. *African Journal of Biotechnology*, 14(9):781–793. doi: 10.5897/AJB2014.14067
- Siadjeu, C., Mayland-Quellhorst, E., and Albach, D. C. (2018). Genetic diversity and population structure of trifoliate yam (*Dioscorea dumetorum* Kunth) in Cameroon

- revealed by genotyping-by-sequencing (GBS). *BMC Plant Biology*, 18(1). doi: 10.1186/s12870-018-1593-x
- Siadjeu, C., Panyoo, E.A., Toukam, G. M. S., Bell, J. M., Nono, B., Medoua, G. N. (2016). Influence of cultivar on the postharvest hardening of trifoliate yam (*Dioscorea dumetorum*) Tubers. *Adv. Agric.* 2016:2658983. doi: 10.1155/2016/2658983.
- Siadjeu, C., Pucker, B., Viehöver, P., Albach, D. C., and Weisshaar, B. (2020). High contiguity de novo genome sequence assembly of trifoliate yam (*Dioscorea dumetorum*) using long read sequencing. *Genes*, 11(3). doi: 10.3390/genes11030274
- Silva, L. R. G., Bajay, M. M., Monteiro, M., Mezette T. F., Nascimento, W. F., Zucchi, M. I., Pinheiro, J. B., Veasey, E. A. (2014). Isolation and characterization of microsatellites for the yam *Dioscorea cayenensis* (*Dioscoreaceae*) and cross-amplification in *D. rotundata*. *Genet Mol Res.*, 13(2):2766–71.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31:3210–3212. doi: 10.1093/bioinformatics/btv351.
- Siqueira, M. V. B. M., Marconi, T. G., Bonatelli, M. L., Zucchi, M. I., and Veasey, E. A. (2011). New microsatellite loci for water yam (*Dioscorea alata*, *Dioscoreaceae*) and cross-amplification for other *Dioscorea* species. *American Journal of Botany* 98(6):144–146. doi: 10.1007/s10455-012-9335-z
- Smit A.F.A., Hubley R., Green P. RepeatMasker Open-4.0. [(accessed on 2 December 2018)]; Available online: <http://www.repeatmasker.org>.
- Sonibare, M. A., Asiedu, R., and Albach, D. C. (2010). Genetic diversity of *Dioscorea dumetorum* (Kunth) Pax using Amplified Fragment Length Polymorphisms (AFLP) and cpDNA. *Biochemical Systematics and Ecology*, 38(3):320–334.
- Speicher, T. L., Li, P. Z., and Wallace, I. S. (2018). Phosphoregulation of the plant cellulose synthase complex and cellulose synthase-like proteins. *Plants*, 7(3):1–18.
- Sun, H., Ding, J., Piednoel, M., Schneeberger, K. (2018). findGSE: Estimating genome size variation within human and *Arabidopsis* using k-mer frequencies. *Bioinformatics*. 2018;34:550–557. doi: 10.1093/bioinformatics/btx637.
- Tamiru, M., Natsume, S., Takagi, H., White, B., Yaegashi, H., Shimizu, M., ... Terauchi, R. (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):1–20. doi: 10.1186/s12915-017-0419-x
- Tamiru, M., Yamanaka, S., Mitsuoka, C., Babil, P., Takagi, H., Lopez-Montes, A., ... Terauchi, R. (2015). Development of genomic simple sequence repeat markers for yam. *Crop Science*, 55(5):2191–2200. doi: 10.2135/cropsci2014.10.0725
- Tang, H., Zhang, X., Miao, C., Zhang, J., Ming, R., Schnable, J. C., Schnable, P. S., Lyons, E., Lu, J. (2015). ALLMAPS: Robust scaffold ordering based on multiple maps. *Genome Biol.*, 16:3. doi: 10.1186/s13059-014-0573-1.
- Terauchi, R., Chikaleke, V. A., Thottappilly, G., and Hahn, S. K. (1992). Origin and phylogeny of Guinea yams as revealed by RFLP analysis of chloroplast DNA and nuclear ribosomal DNA. *Theoretical and Applied Genetics*, 83(6–7):743–751. doi: 10.1007/BF00226693
- Terauchi, R., Konuma, A. (1994). Microsatellite polymorphism in *Dioscorea tokoro*, a wild yam species - *Genome*. *Genome*, 794–801.

- Torkamaneh, D., Laroche, J., Belzile, F. (2016). Genome-wide SNP calling from genotyping by sequencing (GBS) data: A comparison of seven pipelines and two sequencing technologies. *PLoS One*, 11:1–14.
- Tostain, S., Scarcelli, N., Brottier, P., Marchand, J. L., Pham, J. L., and Noyer, J. L. (2006). Development of DNA microsatellite markers in tropical yam (*Dioscorea* sp.). *Molecular Ecology Notes*, 6(1):173–175. doi: 10.1111/j.1471-8286.2005.01182.x
- Treche, S. (1989). Potentialités nutritionnelles des ignames (*Dioscorea* spp.) cultivées au Cameroun (C. E. et T. Editions de l'ORSTOM, Ed.). Paris.
- Treche, S., and Agbor-Egbe, T. (1996). Biochemical changes occurring during growth and storage of two yam species. *International Journal of Food Sciences and Nutrition*, 47(2):93–102. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8833173>
- Treche, S., and Delpuch, F. (1979). Evidence for the development of membrane thickening in the parenchyma of tubers of *Dioscorea dumetorum* during storage. *C. R. Acad. Sc. Paris, T.*, 288:67–70.
- Treche, S., and Delpuch, F. (1982). Le durcissement de *Dioscorea dumetorum* au Cameroun. In: *Yams-Ignames*. (Eds.) J. Miegé and S. N. Lyonga, Oxford: Clarendon Press, pp. 294–311.
- Treche, S., Mbome Lape, I., Agbor-Egbe, T. (1984). Variations de la valeur nutritionnelle au cours de la préparation des produits sèches à partir d'ignames cultivées. *Revue Science et Technique*, (Sci. Santé), (1) :7–22.
- Vaser, R., Sović, I., Nagarajan, N., Šikić, M. (2017) Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.*, 27:737–746. doi: 10.1101/gr.214270.116.
- Verity, R., Nichols, R. A. (2016). Estimating the number of subpopulations (K) in structured populations. *Genetics*, 203:1827–35.
- Viruel, J., Forest, F., Paun, O., Chase, M. W., Devey, D., Couto, R. S., Segarra-Moragues, J. G., Catalan, P., Wilkin, P. (2018). A nuclear Xdh phylogenetic analysis of yams (*Dioscorea* *Dioscoreaceae*) congruent with plastid trees reveals a new Neotropical lineage. *Bot. J. Linn. Soc.*, 187:232–246. doi: 10.1093/botlinnean/boy013.
- Viruel, J., Segarra-Moragues, J. G., Raz, L., Forest, F., Wilkin, P., Sanmartin, I., Catalan, P. (2016). 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.
- Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., Schatz, M. C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics*, 33:2202–2204. doi: 10.1093/bioinformatics/btx153.
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., ... Earl A. M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*, 9(11):e112963. doi: 10.1371/journal.pone.0112963.
- Wang, L., Wang, Z., Chen, J., Liu, C., Zhu, W., Wang, L., and Meng, L. (2015). *De novo* transcriptome assembly and development of novel microsatellite markers for the traditional chinese medicinal herb, 11:39–45. doi: 10.4137/EBO.S20942.Received
- Wang, Y., Suzuki, H., Xie, J., Tomita, O., Martin, D. J., Higashi, M., ... Tang, J. (2018). Mimicking natural photosynthesis: solar to renewable H₂ fuel synthesis by Z-scheme

- water splitting systems [Review-article]. Chemical Reviews, 118(10):5201–5241. doi: 10.1021/acs.chemrev.7b00286
- Wendel, J. F., Jackson, S. A., Meyers, B. C., Wing, R. A. (2016). Evolution of plant genome architecture. Genome Biol., 17:37. doi: 10.1186/s13059-016-0908-1.
- Westphal, E, Embrechts, J, Ferwerda, J. D., van Gils-Meeus, H. A. E., Mutsaers, H J W, Westphal-Stevels, J. M. C. (1985). Cultures vivrières tropicales avec référence spéciale au Cameroun (Pudoc). Wageningen.
- Wickland, D. P., Battu, G., Hudson, K. A., Diers, B. W., Hudson, M. E. (2017). A comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow, GB-eaSy. BMC Bioinformatics, 18:586.
- Wu, Z. G., Jiang, W., Mantri, N., Bao, X. Q., Chen, S. L., and Tao, Z. M. (2015). Transcriptome analysis reveals flavonoid biosynthesis regulation and simple sequence repeats in yam (*Dioscorea alata* L.) tubers. BMC Genomics, 16:346 doi: 10.1186/s12864-015-1547-8
- Zannou, A., Agbicodo, E., Zoundjihékpon, J., Struik, P. C., Ahanchédé, A., Kossou, D. K., and Sanni, A. (2009). Genetic variability in yam cultivars from the Guinea- Sudan zone of Benin assessed by random amplified polymorphic DNA. 8(1):26–36.
- Zhu, Y. L., Song, Q. J., Hyten, D. L., Van Tassell, C. P., Matukumalli, L. K., Grimm, D. R., ... Cregan, P. B. (2003). Single-nucleotide polymorphisms in soybean. Genetics, 163:1123–34.
- Zoundjihekpon, J., Essad, S., Toure, B. (1990). Dénombrement chromosomique dans dix groupes variétaux du complexe *Dioscorea cayenensis-rotundata*. Cytologia, 55:115–120.

CURRICULUM VITAE

Christian SIADJEU
Apenrader Str. 1
26127 Oldenburg
(+49) 1729498930
christian.siadjeu@uol.de

EDUCATION

10/2017 - Present	PhD student, University of Oldenburg, Oldenburg, Germany
05/2019	Transatlantic Summer School – Frontiers in Plant Sciences, Cluster of Excellence on Plant Sciences (CEPLAS), Düsseldorf, Germany
01/2017	Bioinformatics, African Centre of Excellence for genomics of Infectious Diseases ACEGID and African Center of Excellence (CETIC), Nigeria, Cameroon
10/2009-11/2012	M.Sc, Plant Biotechnology, University of Yaoundé 1, Yaounde, Cameroon Thesis: Neutral lipid content and fatty acid composition of oil palm during germination, Grade: 3,09 (being 4 the best grade)
10/2006-06/2009	B.Sc, Plant Biology, University of Yaoundé 1, Yaounde, Cameroon

WORK EXPERIENCE

10/2017 - Present	PhD student, University of Oldenburg, Oldenburg, Germany
01/2020 – 02/2020	Visiting researcher, International Institute of Tropical Agriculture IITA, Ibadan, Nigeria
01/2020 – 02/2020	Guest researcher, University of Ibadan, Ibadan, Nigeria
10/2017 - Present	Visiting researcher, University of Bielefeld, Bielefeld, Germany
05/2016 - 06/2019	Local expert, Press NO & SW, Bundesanstalt für Geowissenschaften und Rohstoffe (BGR, Germany), Yaounde, Cameroon
12/2014 - 05/2016	Research assistant, Centre for Food and Nutrition Research (IMPM), Yaoundé Yaoundé, Cameroon
02/2013-03/2014	Research assistant at Institute of Agricultural Research for Development (IRAD), Ekona Buea, Cameroon

01/2013 – 12/2016	Expert, Appropriate Development for Africa Foundation (ADAF), Yaounde, Cameroon
06/2011-12/2012	Research assistant in laboratory of biochemistry of Research Center specializing in the Oil Palm (CEREPAH), (CEREPAH)/IRAD, Douala, Dibamba, Cameroon
06/2011-07/2013	Teacher of physics, and chemistry at the Government bilingual high school of Mbongo, Dibamba, Cameroon

GRANT FUNDING & AWARDS

10/2019	Lab visiting grant, research group genetics and genomics of plants, University of Bielefeld, German Society of Plant Sciences (DBG), Germany
05/2019- 12/2021	Research Group Linkage Programme, Alexander von Humboldt foundation, Germany
08/2018	Travel grant, Universitätsgesellschaft Oldenburg (UGO), Oldenburg, Germany
06/2017	PKFokam Awards for Science & Technology 2017, PKFokam Institute of Excellence, Yaounde, Cameroon
03/2017- 03/2021	Research Grant-Doctoral Programmes in Germany, German Academy Exchange Service (DAAD), Germany
02/2017	Robert S. McNamara Fellowship Program, The World Bank, Washington, USA
12/2015	Third Jury Award of Dr. THE Charles Prize << Ma thèse en 180 sec >> of 22nd annual conference of Cameroon Bioscience Society (CBS), Yaounde, Cameroon
11/2015	Third Jury Award of the second edition of Science Slam competition (Goethe Institute, Yaounde, Cameroon)

TEACHING/MENTORING

01/2020 - Present	I am supervising one PhD student (Priscilla Aiyedun) from the University of Ibadan, Nigeria on transcriptome analysis of genes involved in anti-diabetic properties of yam using RNA-Seq
02/2020	Guest lecture on transcriptome analysis using RNA-Seq, Faculty of Pharmacy, University of Ibadan, Ibadan, Nigeria
02/2020	Transcriptomic data analysis from RNA-Seq, Workshop, International Institute of Tropical Agriculture, IITA, Ibadan,

Nigeria, workflow available on my GitHub account
https://github.com/Siadjeu/Yam_transcriptome

- 03/2019 Assessment of ploidy level using GBS from ipyrad, working group internal workshop, University of Oldenburg, Oldenburg, Germany
- 06/2019 Variant calling using stacks program, working group internal workshop, University of Oldenburg, Oldenburg, Germany, workflow available on my GitHub account
<https://github.com/Siadjeu/Calling-variant-Stacks>
- 04/2016 – 03/2017 I supervised the master thesis of Narcisse Tybeau Zoa Andela on the floral biology of yam (*Dioscorea dumetorum*), university of Yaounde 1, Yaounde, Cameroon
- 04/2014 – 03/2015 I supervised the master thesis of Marie Noel Ateko Tiokeng on artificial pollination of yam (*Dioscorea dumetorum*), University of Yaounde 1, Yaounde, Cameroon

LANGUAGE(S)

- English: fluent
- French: native speaker
- German: good
- Fe'e fe'e (a cameroonian language): native speaker

SKILLS & TOOLS

- R: Very good
- Python: Very good
- Bash: Very good
- Unix/Linux: Very good
- Perl: good
- High Performing Computer HPC: very good

BIOINFORMATIC

- Genome assembly and annotation
- Transcriptome assembly and annotation
- Variants calling for genetic diversity and population structure
- Differential expression analysis

PRESENTATIONS

- Transatlantic Summer School-Frontiers in Plant Sciences, Poster presentation (Molecular breeding bases of trifoliate yam (*Dioscorea dumetorum* Kunth Pax) against post-harvest hardening phenomenon of tubers, May 27-31, 2019, Maria in der Aue, Germany
- 18th Triennial Symposium of International Society for Tropical Roots Crops ISTRC, 2018. Poster presentation

(Genetic diversity and population structure of trifoliate yam (*Dioscorea dumetorum* Kunth) in Cameroon revealed by genotyping-by-sequencing (GBS), Cali Colombia.

- 22th Annual Conference of the Cameroon Bioscience Society, 2015. Poster presentation (Content of Neutral lipids and fatty acid composition of the oil palm seeds *Elaeis guineensis* Jacq. during germination). Yaounde, Cameroon

REVIEWING ACTIVITIES via publons

- Plant Breeding
- Physiologia Plantarum
- PeerJ

MEMBERSHIPS

09/2020 - Present	Alumni Netzwerk Subsahara Afrika e.V., Germany
11/2018 - Present	German Society for Plant Sciences (DBG), Germany

OFFICE IN ACADEMIC SELF-ADMINISTRATION

03/2020 - Present	PhD student representative committee, University of Oldenburg, Oldenburg, Germany
05/2019 - Present	PhD thesis committee “Promotionsausschuss”, University of Oldenburg, Oldenburg, Germany

INTERESTS

- Football
- Travelling
- Reading

PUBLICATIONS

- **Siadjeu, C.**; Pucker, B.; Viehöver, P.; Albach, D.C.; Weisshaar, B. (2020). High Contiguity *de novo* Genome Sequence Assembly of Trifoliate Yam (*Dioscorea dumetorum*) Using Long Read Sequencing. *Genes*, (2020) **11**: 274.
- **Siadjeu C**, Mayland-Quellhorst E, and Albach D.C. (2018). Genetic diversity and population structure of trifoliate yam (*Dioscorea dumetorum* Kunth) in Cameroon revealed by genotyping-by-sequencing (GBS). *BMC Plant Biology*, (2018) 18:359.

- **Siadjeu C.**, Panyoo A. E., Mahbou S. T. G., Bell J. M., Nono B., and Gabriel Medoua N. G. (2016): Influence of Cultivar on the Postharvest Hardening of Trifoliate Yam (*Dioscorea dumetorum*) Tubers. *Advances in Agriculture*, (2016) 1-7.
- **Siadjeu, C.**, MahbouSomoToukam, G., Bell, J. M. & Nkwate, S. (2015). Genetic diversity of sweet yam “*Dioscorea dumetorum*“(Kunth) Pax revealed by morphological traits in two agro-ecological zones of Cameroon. *African Journal of Biotechnology*, 14(9), 781-793.
- Toukam G. M. S., **Siadjeu C.**, Bell J. M., Nkwate S., Bomda J. (2015). Influence de quelques caractères agronomiques sur le rendement de l’igname sucrée (*Dioscorea dumetorum* Kunth Pax) au Cameroun. *Int. J. Biol. Chem. Sci.* 9(1): 141-154.

AUTHOR CONTRIBUTIONS

Article I and Article II are already published. Article II submitted to Plants (2021), manuscript ID, plants-1130765

Article I

C.S. contributed to design of the experiments, collection of yam accessions, data analysis and interpretation and manuscript writing. E.M.Q. DNA extraction and sequencing, D.C.A. data interpretation and manuscript writing.

Article II

C.S., B.P., D.C.A., and B.W. designed the study. C.S. collected the sample. B.P. performed DNA extraction, ONT sequencing, and genome assembly. P.V. performed Illumina sequencing. C.S. and B.P. processed the assembly. B.P. performed gene prediction and evaluation. C.S. and B.P. wrote the initial draft. B.W. and D.C.A. revised the manuscript. All authors read and approved the final version of the manuscript.

Article II

C.S., D.C.A., S.L. contributed to design of the experiments, C.S. collected the samples and performed data analysis and interpretation. E.M.Q. DNA performed cDNA library preparation, S.L. performed qPCR validation, C.S. wrote the first draft of the manuscript, D.C.A. and S.L. revised the manuscript.

Als Betreuer der Arbeit bestätige ich die Richtigkeit der Autorenbeiträge zu den aufgeführten Kapiteln bzw. Deren Manuscripten oder Veröffentlichungen.

Oldenburg, den.....

.....

Prof. Dr. Dirk C. Albach

Hiermit bestätige ich die Richtigkeit der Autorenbeiträge zu den aufgeführten Kapiteln bzw. Deren Manuscripten oder Veröffentlichungen.

Oldenburg, den.....

.....

Christian Siadjeu