ORPHAN NO MORE

Ethnobotany and genetic analysis of leaf yield and secondary metabolites content in *Gynandropsis gynandra* (Cleomaceae)

E. O. Dêêdi Sogbohossou

Propositions

1. The evolutionary history, reproduction strategy and medicinal properties of *Gynandropsis* (Akaya) are perfectly summarized by the Fon proverb "Akaya man non kou do da wiwi min" which translates as "May you live as long as Akaya".
(This thesis)

2. Beyond financial resources, a key element to promote orphan crops is collaboration between enthusiastic stakeholders.
(This thesis)

3. Assessing end-users feedback on research results should be a mandatory 'development' component of 'research for development' projects.

4. Advocacy for informed and accessible music education is as important as research-based music therapy.

5. The biggest contradiction of the century is the claim that climate change can be dealt with within the constraints of a capitalist system.

6. The only person worth competing with is the one we were yesterday.

Propositions belonging to the thesis entitled:
"Orphan no more: ethnobotany and genetic analysis of leaf yield and secondary metabolites content in *Gynandropsis gynandra* (Cleomaceae)"
EOD Sogbohossou
Wageningen, 16 December 2019

Orphan no more

Ethnobotany and genetic analysis of leaf yield and secondary metabolites content in *Gynandropsis gynandra* (Cleomaceae)

E. O. Dêêdi Sogbohossou

Thesis committee

Promotor
Prof. Dr M.E. Schranz
Professor of Biosystematics
Wageningen University & Research

Co-promotor
Dr E.G. Achigan-Dako
Associate professor, Faculty of Agronomic Sciences
University of Abomey-Calavi, Benin

Other members
Prof. Dr J.J.B. Keurentjes, Wageningen University & Research
Prof. Dr S. Neugart, Georg-August-Universität Göttingen, Germany
Prof. Dr P. Van Damme, Ghent University, Belgium
Dr A.B. Bonnema, Wageningen University & Research

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences.

Orphan no more

Ethnobotany and genetic analysis of leaf yield and secondary metabolites content in *Gynandropsis gynandra* (Cleomaceae)

E. O. Dêêdi Sogbohossou

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 16 December 2019
at 1.30 p.m. in the Aula.

E. O. Dêêdi Sogbohossou
Orphan no more: Ethnobotany and genetic analysis of leaf yield and secondary metabolites content in *Gynandropsis gynandra* (Cleomaceae)
214 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2019)
With references, with summaries in English and French

ISBN 978-94-6395-132-6
DOI https://doi.org/10.18174/501019

CONTENTS

Abbreviations

AFLP	Amplified Fragment Length Polymorphism
AGRA	Alliance for Green Revolution in Africa
AOCC	African Orphan Crops Consortium
CGIAR	Consultative Group for International Agricultural Research
cM	Centimorgan
CSI	Cultural Significance Index
DNA	Deoxyribonucleic acid
FAO	Food and Agriculture Organization of the United Nations
GBS	Genotyping-By-Sequencing
GC-MS	Gas Chromatography-Mass Spectrometry
GGE	Genotype main effect plus Genotype by Environment interaction
GO	Gene Ontology
HPLC	High-Performance Liquid Chromatography
ITS	Internal Transcribed Spacer
KASP	Kompetitive allele specific PCR
KENRIK	Kenya Resource Centre for Indigenous Knowledge
LC-MS	Liquid Chromatography-Mass Spectrometry
LD	Linkage Disequilibrium
LOD	Logarithm of odds
MAGIC	Multi-parent Advanced Generation Inter-Cross
MI	Management Index
NAD(P)	Nicotinamide Adenine Dinucleotide (Phosphate)
PCR	Polymerase Chain Reaction
PDA	Photodiode Array
PROSEA	Plant Resources of South-East Asia
PROTA	Plant Resources of Tropical Africa
PVE	Phenotypic Variation Explained
QTL	Quantitative Trait Locus
RAPD	Random Amplification of Polymorphic DNA
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeat
TILLING	Targeting Induced Local Lesions IN Genomes
UAC	University of Abomey-Calavi
USD	United States Dollar
USDA	United States Department of Agriculture
WHO	World Health Organization

CHAPTER 1

General Introduction

1

Introduction

In just 30 years, nine billion humans will have to compete for the limited available resources on our planet. Hopefully we can find and implement sustainable solutions to this crisis that incorporate curbing population growth, reducing income inequality, preserving biodiversity and promoting human well-being and health. The 2018 report of the FAO on the state of food security and nutrition in the world highlights that the prevalence of malnourished people in the world has been on the rise since 2014 increasing from 10.7 to 10.9% of the world population. At the same time, obesity is worsening with more than one in eight adults in the world being currently obese. Food insecurity contributes to malnutrition, including both obesity and undernutrition, with high rates coexisting in many countries. The higher cost of nutritious foods, the stress of living with food insecurity, and physiological adaptations to food restriction partially explain why food insecurity leads to increasing risks of overweight and obesity (FAO 2018). Moreover, overweight and obese individuals can also be affected by micronutrient (vitamin and mineral) deficiencies, often referred to as "hidden hunger" because there may be no visible signs of it. It is estimated that over two billion people in the world are affected by one or more forms of micronutrient deficiency (von Grebmer et al. 2014). Feeding the world will therefore not only be about quantity, but critically about ensuring a diversified diet with micronutrient rich crops. Such crops are too often overlooked in the evaluation of everyone's diet.

We currently rely on just thirty staple crops to provide 90% of food calories, while the global flora includes over 12,000 edible plants, including many fruits and leafy greens (Ozturk et al. 2018). In Africa alone, the Plant Resources of Tropical Africa Foundation listed approximately 8,600 edible species (Prota 2010). Khoury et al. (2014) analysed the evolution of global change in spread and abundance of fifty-two commodity crops over fifty years (1961-2009) and found that there was an increasing homogeneity in national food supplies. Oil commodities such as soybean, sunflower, palm oil, and rape/canola oil were among the crops showing the greatest average increase in relative abundance in national food supplies whereas "minor" crops including millets, rye, sorghum, yams, cassava, and sweet potato showed the largest declines.

Several locally important crops are often referred to as "minor", "underutilized", or "orphan" crops. They provide multiple benefits to local communities including nutrition and food, income and medicinal uses. Furthermore, they are often suitable for low-input agricultural systems and adapted to local and marginal environments. The paradox of poverty in tropical areas of the world, where the bulk of plant diversity is present, calls for smarter and more sustainable ways to use the local flora to attain food and nutritional security. During the last decades, orphan crops have increasingly become a focus of interest and research around the world. Minor staple crops, including cereals (e.g. sorghum, tef, fonio; Chanyalew et al. 2019; McCormick et al. 2018), legumes (e.g. cowpea, Bambara groundnut, pigeon pea; Bonthala et al. 2016; Spriggs et al. 2018; Varshney et al. 2017a) and roots/tubers (e.g. cassava, yam species; Umber et al. 2014; Wang et al. 2014b) are increasingly studied for breeding purposes (e.g. nutritional quality, yield, resistance to pests and diseases). However, several minor vegetables and fruits that remain the main sources of micronutrients in tropical regions of the world are still largely overlooked by agricultural research agendas.

Promoting orphan leafy vegetables to help achieve food security and climate –resilient agricultural systems

Most traditional dishes in Africa and other tropical areas of the world are composed of staple crops accompanied with vegetable preparations, either fresh or in soups, sauces and condiments **(Figure 1)**. These vegetables are a main source of micronutrients and are important in curbing hidden hunger (Cernansky 2015).

Figure 1. Mix of leafy vegetables cooked with egusi (*Citrullus mucosospermus*) to be accompanied with starchy preparations, here yam dough ("telibo") and rice cakes ("ablo"). ©Photo credit: E. G. Achigan-Dako

Most of these species are also used in local medicines as cures against various ailments from skin diseases to viral infections (Grubben and Denton 2004), suggesting that they are untapped reservoirs of metabolites with health-promoting benefits (Neugart et al. 2017). While most crops require several months from sowing to ripening of the edible parts (i.e. grains, pods, roots or floral parts), leafy vegetables can first be harvested within weeks after sowing. These crops can also be harvested multiple times during their cycle or planted multiple times per year, for example in West African vegetable production systems (Pasquini et al. 2009). Moreover, leafy vegetables are usually grown on small plots and are sometimes combined with other crops with a longer cycle, allowing an optimal utilization of available land resources. Given their short cycle and high planting densities, orphan leafy vegetable production is resilient to pockets of drought whereas plants can be easily watered in such situations. Marketing of orphan leafy vegetables is usually carried out by women, thus providing an additional source of income for their families. The low external inputs and labor costs required for leafy vegetable production also makes it both profitable for farmers and affordable for consumers (Weinberger and Pichop 2009). The volume of African vegetables sold in Kenya between 2004 and 2006 was estimated at 9,000 tonnes, with cash income of Ksh 80 million (UK£ 0.63 million) in informal and Ksh 150 million (UK£ 1.19 million) in formal markets (Muhanji et al. 2011).

Generating scientific information on leafy vegetables for their effective promotion and upscaling should build on existing traditional bodies of knowledge. However, the modernization process in many societies worldwide has led to the erosion of traditional local communities' knowledge on these species (Albuquerque et al. 2016; Cruz et al. 2014). Such a loss of knowledge and failure to pass information onto younger generations has been attributed to several factors including urbanization, migration and cultural changes (Cruz et al. 2014; Kidane et al. 2015; Leal et al. 2018). Therefore, there is an urgent need to preserve traditional knowledge through systematic documentation. Making relevant information on plant species knowledge and use

publicly available will help to preserve such knowledge. Moreover, patents cannot be applied on published data. Thus, dissemination of traditional knowledge through publications contributes to reduce cases of biopiracy and prevent the use of such knowledge for commercial purposes without access and benefit sharing with local communities.

In Africa, several recent initiatives to promote indigenous vegetables include, among others: (i) various small grants projects initiated by the International Plant Genetic Resources Institute (now Bioversity International) together with the Netherlands Ministry of Foreign Affairs, Development Cooperation and the Natural Resources Institute of the United Kingdom (NRI) to organize series of workshops and publish working documents to promote African Indigenous vegetables (Chweya and Eyzaguirre 1999); (ii) the EU programme IndigenoVeg, which drew together a network of leading EU and Sub-Saharan African researchers in order to promote the production of local vegetables in urban and peri-urban areas by coordinating and integrating current research efforts on the role played by local vegetables varieties in food security and livelihoods, and research conducted in the field of urban and peri-urban agriculture (Pasquini et al. 2007); (iii) the World Vegetable Center, Regional Center for Africa (AVRDC-RCA) which works in close collaboration with National Agricultural Research and Extension Systems (NARES), Regional and International organizations, Non-governmental Organizations (NGOs), the private sector, and farmers in Africa for research, development, training and information on local vegetables (Dinssa et al. 2016); (iv) the Darwin Initiative programme led by a consortium of stakeholders from Benin, Mali and the United Kingdom that conducted the inventory of the diversity of traditional vegetables in Benin and Mali (Achigan-Dako et al. 2010); (v) The Recipes for Success project funded by GlobalHort to promote selected leafy vegetables in Benin, Kenya and Tanzania (https://www.agropolis.fr/pdf/actu/2011-globalhort-recipes-for-success-Montpellier.pdf) ; (vi) the MicroVeg project implemented by the International Development Research Center (IDRC), the International Plant Nutrition Institute (IPNI) in Benin, Nigeria and Kenya (www.microveg.org); (vii) the African Orphan Crops Consortium which is committed to (re-)sequencing for free the genome 100 accessions for each of 101 crops used on the continent, including 20 vegetable crops and provide training in plant breeding/bioinformatics to African researchers involved in orphan crops breeding (Hendre et al. 2019); and (viii) the Cleome project with Benin, Kenya and the Netherlands that will be developed further in another paragraph.

Traditional knowledge and use of orphan crops can be leveraged to inform innovation of derived food products. However, such efforts need to go hand-in-hand with omics-assisted breeding projects targeting improved crop yield and quality. That begins with baseline information about the natural variation of key-traits, followed by the downstream discovery of functional molecular markers or candidate genes for the traits of interest. This approach can be greatly accelerated by exploiting and utilizing recent advances in -omics technologies and particularly by leveraging comparative approaches.

This thesis aims at developing a breeding program for *Gynandropsis gynandra*, an orphan leafy vegetable widespread in tropical and subtropical Africa and Asia. The species belongs the Cleomaceae, the sister family of the Brassicaceae which includes the well-studied model and crop species such as *Arabidopsis thaliana* and diverse Brassica crops. While *G. gynandra* has become a model for studying C_4 photosynthesis (Külahoglu et al. 2014; Marshall et al. 2007),

there has been a lack of research into crop-related traits such as the production of carotenoids (pro-vitamin A) and tocopherols (vitamin E), or their improvement by molecular breeding.

The key-aspects and aims for establishing a breeding program for orphan crops are introduced and discussed in detail in **Chapter 2**. Thus, in the remaining following sections of this General Introduction, we explain why we chose the orphan crop *G. gynandra* for our research project and present the overall methodological framework of the thesis.

Gynandropsis gynandra, an orphan leafy vegetable and more

Taxonomy, origin, and botanical description

Gynandropsis gynandra belongs to the Cleomaceae family which encompasses eighteen genera and 150-200 species (Patchell et al. 2014). The genus *Cleome* and other genera of the family have been traditionally considered as the subfamily Cleomoïdeae of the family Capparaceae (Inda et al. 2008). In recent phylogenetic studies, Cleomoïdeae were found to be more closely related to the Brassicaceae than Capparoideae (Hall 2008; Hall et al. 2002; Iltis et al. 2011). Thus, three monophyletic families were delineated: Capparaceae, Cleomaceae, and Brassicaceae. The morphological characters specific to the Cleomaceae include their mostly herbaceous habit, compound leaves and dry and dehiscent capsules with a replum. Seeds are uniformly strongly curved, reniform or horseshoe-shaped to conduplicate with a deep invagination of the testa, an incurved embryo with always strictly incumbent cotyledons (Iltis et al. 2011).

Gynandropsis gynandra (L.) Briq (1914) was also known as *Cleome gynandra* L. (1753), *Cleome pentaphylla* L. (1763) and *Gynandropsis pentaphylla* (L.) DC. (1824). *Cleome gynandra* was reclassified as the sole member of the monotypic genus *Gynandropsis* based on the presence of a long androgynophore, the characteristic from which its name was derived (Iltis 1960). However, such a structure has also been observed in *Cleome speciosa* and *Podandrogyne*. Many authors considered, therefore, this character as questionable and the species was renamed *Cleome gynandra* (Iltis 1960; Sánchez-Acebo 2005). Recent molecular studies, however, supported the reestablishment and conservation of *Gynandropsis* as a separate monotypic genus (Feodorova et al. 2010; Hall 2008; Patchell et al. 2014). Partially because of its wide-distribution range and diverse phenotypes, the species has an astounding forty-five synonyms used in literature (GBIF 2019). This plethora of names desperately calls for a taxonomic revision of the genus *Gynandropsis*.

The vernacular names of the species are noteworthy as they tell a story about how enslaved Africans migrating to the New World kept generic plant names from their communities of origin (van Andel et al. 2014). For example, *G. gynandra* is called "Sambo" or "Somboé" in Ewe and Mina communities of the coastal areas of West Africa while the same name "Sambo" has been reported by Iltis (1960) in Barbados. Likewise, the vernacular name "Akaya" in Fon and Mahi communities of Benin is declined by people in Martinique, Suriname and French Guiana as "Akaja", "Akaïa" or "Caïa". The name "Mozambue" or "Mozembue," used in Angola, have their equivalent in "Massambee" and "Mozambe" in Guadeloupe, Barbados, and several other Pacific Islands (Iltis 1960). While the plant was reported as a popular vegetable in French Guiana and Suriname in the 18th century, it is nowadays considered as a weed in these countries (van Andel et al. 2016).

The origin and dispersal of *G. gynandra* are still uncertain. Feodorova et al. (2010) suggested that the origin of the entire Cleomaceae family could be Central Asia with subsequent dispersal to southern Africa, Australia and North, Central and South America. The speciation event leading to *G. gynandra* likely occurred in southern Africa (Feodorova et al. 2010). However, the phylogeographic analysis of the latter authors was based on a single ITS region and with only four accessions of *G. gynandra* from Australia and South Africa, raising the need for further studies integrating more diverse germplasm collections and the use of whole-genome sequences to reconstruct the demographic history of the species.

G. gynandra occurs throughout the tropics and subtropics. In Africa, it is mainly found near human settlements, possibly feral escapes from earlier cultivation. It is often found in cultivated or fallowed fields, along roadsides, in fence rows, and along irrigation canals and ditches. *G. gynandra* occurs from sea level up to 2400 m altitude and requires warm conditions; growth is hampered below 15°C, which is typical of C_4 species. It thrives on a wide range of soils, mostly on sandy to clayey loam, provided they are deep and well-drained with a pH 5.5–7.0. It prefers soils with high organic matter and adequate mineral reserves (Mnzava and Chigumira Ngwerume 2004). The species tolerates some drought, but drought stress hastens maturity and senescence (Onyango et al. 2013b).

G. gynandra is an erect annual herb up to 150 cm tall that is strongly branched (**Figure 2A**), with a long taproot and few secondary roots. The stem is glandular-pilose or rarely glabrous. Leaves are alternate, palmately compound with 3, 5 or 7 leaflets with many observers describing the leaves as cannabis-like (**Figure 2B**). Stipules are absent and the petiole is 2–10 cm long and with glands. Leaflets are almost always sessile, obovate to elliptical or lanceolate, 2–10 cm × 1–4 cm, cuneate at base, rounded to obtuse, acute or acuminate at apex, margins finely toothed, sparsely to distinctly hairy. The inflorescence is a terminal raceme up to 30 cm long and bracteate (**Figure 2C**). Flowers are bisexual, with petals being white or tinged with purple. The pedicel is 1.5–2.5 cm long, with four sepals that are free, ovate to lanceolate and up to 8 mm long. It also has four petals that are elliptical to obovate, up to 1.5 cm long and clawed. There are six staminal filaments, that are purple and fused basally with the gynophore and thus produce an androgynophore 5-22 mm long (**Figure 2D**), the free filaments 8-22 mm long. The ovary is superior and is stalked and 2-celled. The fruit is a long, narrow, cylindrical capsule up to 12 cm × 1 cm, stalked and beaked, usually green or yellow, dehiscing from below when dry (**Figure 2E**). Seeds are subglobose, 1–1.5 mm in diameter, grey to black and irregularly ribbed (**Figure 2F**) (Grubben et al. 2004b; Iltis 1960).

A source of livelihoods for vegetable growers

G. gynandra is harvested in the wild in some communities, while in others the species is cultivated and sold at local markets. In Kenya, uprooted whole plants put together in bunches, are mainly sold in open-air markets but can also be found in supermarkets and greengrocer stores as a result of recent efforts to revalorize the crop in urban settings (Onyango et al. 2013a). The IndigenoVeg project reported that the selling price in 2006 was 0.30 USD/kg on average in Uganda with an average profit margin of 44.3%; while in Kenya, the selling price was 0.41 USD/kg on average with an average profit margin of 39.6% (Weinberger and

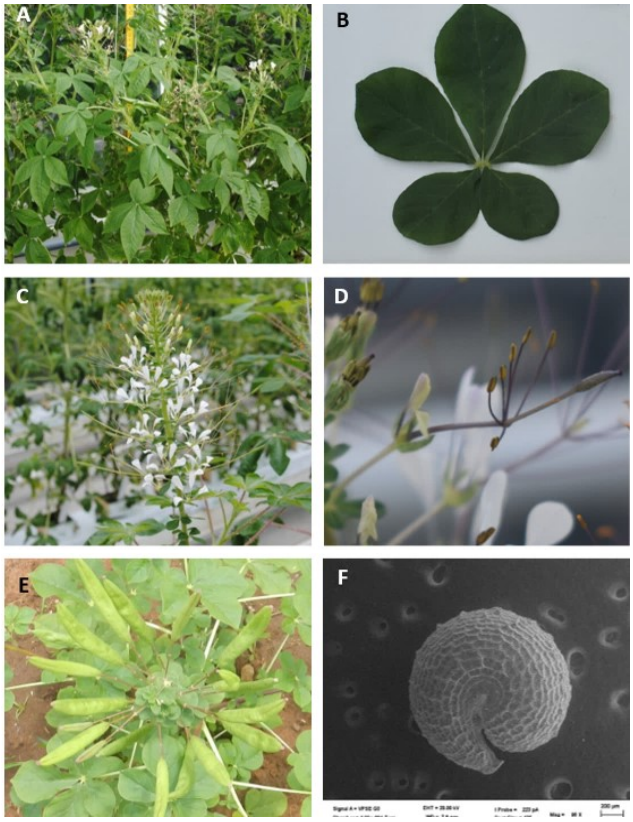


Figure 2. Morphology of *Gynandropsis gynandra*: (A) Plant at flowering stage; (B) Leaf with five leaflets; (C) Inflorescence with opened flowers. (D) Blooming flower with petals and bracteates at the base, inserted stamens on the androgynophore and the gynoecium at the tip. (E) Immature fruits; (F) Subglobose seed with ribbed coat. ©Photo credit: Deedi Sogbohossou & Jelila Blalogoe

Pichop 2009). The species contributed 15-40% of the total income of the farmers producing and selling it.

A survey conducted in Southern Benin in 2015 with 150 growers over two rainy seasons revealed that the selling price varied between 0.2 USD/kg to 0.6 USD/kg whereas the average profit margin varied between 170% in the beginning of the rainy season and around 24% at the end of the rainy season when lots of other vegetables were available (Matro et al. in prep.). The low production supply of the vegetable and its limited shelf life which results in important post-harvest losses were the main constraints for vegetable sellers. The longest period of storage was two days in the outlets. Post-harvest losses due to wilting, rotting and lesions during harvesting and transportation contributed to profit losses (Onyango et al. 2013c). A possible solution was found by Gao et al. (2011) who isolated and characterised a polyphenol oxidase from *G. gynandra* that could retard the browning reaction of the leaves and thus extend shelf-life.

A nutritious leafy vegetable

G. gynandra is a rich source of nutrients including vitamins, minerals (calcium and iron) and protein (Yang and Keding 2009). *G. gynandra* leaves, young shoots and occasionally flowers are eaten boiled as potherb, relish, stew or side dish (Grubben et al. 2004b; Mnzava and Chigumira Ngwerume 2004). The leaves can also be blanched, made into small balls and sun-dried for preservation (Flyman and Afolayan 2006). In some countries, the leafy vegetable is particularly sought for during extreme drought or famine periods, and, therefore plays a significant role in maintaining household food security (Ekpong 2009). The consumption of *G. gynandra* leaves can significantly contribute to the Recommended Nutrient Intake (RNI; WHO 2004) of vitamin A, ascorbic acid, folate and iron (Abukutsa-Onyango et al. 2010; Schönfeldt and Pretorius 2011; Steyn et al. 2001b; Uusiku et al. 2010; van Jaarsveld et al. 2014). However, variability was high in reported levels of vitamin A, ascorbic acid, riboflavin, iron, calcium and magnesium across studies. This is potentially due to differences in harvesting and storage conditions, sample preparation, analytical methods, stage of maturity, as well as the cultivars/genotypes used. A comparative study between *G. gynandra*, *Brassica oleracea* and *Beta vulgaris* revealed higher levels of phosphorus, potassium, calcium, iron, zinc, ascorbic acid, total phenolics, and flavonoids in *G. gynandra* than in the two other vegetables (Moyo et al. 2018). Moreover, the species had ratios of phytate:iron and phytate:calcium below the critical thresholds and could therefore be considered as a good source of bioavailable iron and calcium. In contrast, the phytate:zinc ratio was much higher than the critical threshold and thus zinc bioavailability was compromised (Gowele et al. 2019). The assessment of the levels in provitamin A carotenoids, tocopherols, ascorbic acid, minerals, and phytate in 13 different African leafy vegetables revealed that *G. gynandra* had average levels of carotenoids and tocopherols, and high levels of ascorbic acid (Gowele et al. 2019). Thus, improving the levels of carotenoids and tocopherols should be a key target for the improvement of *Gynandropsis* cultivars.

Uses in folk medicine and pharmacological properties

Leaves and seeds of *G. gynandra* are used in indigenous medicine in many countries (Bala et al. 2010). In India, a decoction of the root is used as febrifuge. Additionally, leaves are used as a wound disinfectant. Inhaling the aroma of the grinded fresh leaves is believed to relieve headache, whereas leaf juice and oil are used to treat earaches, stomachaches, epileptic fits and as an eye wash (Bala et al. 2010). Seeds have been reported to have anthelmintic properties and the oil is used as a fish poison. In rural Kenya, *G. gynandra* is believed to improve eyesight, provide energy and cure marasmus (Chweya and Mnzava 1997). Nandi people in Kenya also use a decoction of the whole plant to cure malaria, facilitate childbirth and relieve stomach congestion (Jeruto et al. 2008). In Uganda, leaves are used to treat snake bites (Namukobe et al. 2011). In Benin, uses for the treatment of malaria (Yetein et al. 2013) and anaemia (Allabi et al. 2011) were reported.

Secondary metabolites discovery in the leaves of *G. gynandra* corroborated the diverse uses of the species in health care. The leaves exhibit high concentrations of alkaloids, steroids (Ajaiyeoba 2000), glucosinolates, flavonoids, gallotannins, iridoids and other phenolic compounds (Moyo et al. 2013). Methanol extracts of *G. gynandra* had moderate antifungal and antibacterial activities (Muchuweti et al. 2007). Bala et al. (2010) reported on the anticancer properties of *G. gynandra*.

The main flavonoids found in the leaves of the species include quercetin, kaempherol and traces of isorhamnetin (Omondi et al. 2017b; Yang et al. 2008). Quercetin and kaempherol have been shown to have antithrombogenic activities (Nijveldt et al. 2001). *G. gynandra* also contains glucosinolates, sulfur-compounds that confer to Brassicales their pungency and are also involved in plant defense mechanisms against herbivores (Neugart et al. 2017; Omondi et al. 2017b).

A model to study whole-genome duplication and C_4 photosynthesis

For a long time, the interest of the scientific community for *G. gynandra* was not due to its importance as a crop but rather as a model to study eudicot C_4 photosynthesis (Külahoglu et al. 2014; Marshall et al. 2007). The knowledge generated in fundamental studies related to polyploidy and C_4 photosynthesis can be mainstreamed into molecular breeding efforts. A good example would be that carotenoids (pro-vitamin A) and tocopherols (vitamin E) are all photosynthetic pigments that contribute to photosynthesis efficiency in the plant.

Multiple and independent polyploidy events have occurred throughout plant evolution, often correlating with lineage diversification and the appearance of novel traits (Schranz et al. 2012; Soltis et al. 2009). Recent polyploidy events are detectable at the karyotype level for many crop species including for example wheat (*Triticum aestivum* L.), cotton (*Gossypium hirsutum* L.), oilseed rape (*Brassica napus* L.), peanut (*Arachis hypogaea* L.), potato (*Solanum tuberosum* L.), sweetpotato (*Ipomoea batatas* L.) among others. However, when polyploidy occurred in ancestral species, traces of these ancient whole-genome duplication events were difficult to identify in their descendant lineages as most plants would reduce their genome back to a more stable diploid state through gene loss (or fractionation), chromosomal rearrangements and overall a reduced chromosome number (Barker et al. 2009). Therefore, detecting and dating ancient polyploidy events is a challenging task. In the absence of fossil records in the taxa of interest, the most common approach would be to infer timing of polyploidy events using synonymous: non-synonymous substitution ratios in whole-genomes or Expression Sequence Tags (ESTs). The rationale behind the molecular dating approach is that nucleotide substitutions that do not result in a change in a protein sequence and hence do not induce a change in phenotype (synonymous) are less likely to be subject to natural selection than substitutions that result in a change in protein sequence or alter the protein function (non-synonymous) (Hu and Banzhaf 2008). Synonymous substitutions are therefore expected to accumulate at a constant rate over time and hence genes that "appeared" at the same evolutionary time-point are likely to have similar Ks values. A peak in the number of duplicate genes with Ks values in a narrow interval would therefore be a signal of a polyploidy event. Most flowering plant lineages underwent one or more rounds of ancient polyploidy (Jiao et al. 2011).

For example, the Brassicaceae and the Cleomaceae shared at least four ancient polyploidy events: ζ which occurred near the origin of seed plants; ε which was shared by the angiosperms; the ancient hexaploidy At-γ which was shared by almost all eudicots and At-β restricted to part of the Brassicales as it was not detected in the papaya genome (Jiao et al. 2011). An ancient whole-genome duplication event At-α is restricted to Brassicaceae while in Cleomaceae, evidence of a triplication Th-α (Cheng et al. 2013) shared by *Tarenaya hassleriana* and *G. gynandra* was found based on transcriptome analysis (van den Bergh et al. 2014). The occurrence of independent ancient polyploidy events in these closely related lineages provide an opportunity to analyse the

relative timing of polyploidization, patterns of gene retention and loss, and genome evolution (Schranz and Mitchell-Olds 2006).

The whole-genome triplication in the Cleomaceae could be linked to the occurrence of C_4 plants in this family (van den Bergh et al. 2014). C_4 plants are among the most successful species in terms of biomass production in warm climates (Sage and Stata 2015). They exhibit higher water, radiation and nitrogen use efficiency than C_3 species resulting in higher yields (Ding et al. 2015). C_4 photosynthesis is therefore considered as an important trait which could potentially help to cope with the adverse effects of climate change on food production. Despite several studies undertaken to dissect the C_4 pathway using systems biology and comparative omics (Covshoff et al. 2016; Kumar et al. 2015; Wang et al. 2014a), many questions remain unresolved. For example, regulatory and interaction networks underlying specific C_4 anatomical features such as venation patterns, differentiation between bundle sheath and mesophyll cells, plasmodesmatal abundance, are still poorly understood (Ding et al. 2015; Nelson 2011). C_4 studies have been interested in a few genera including but not limited to *Sorghum*, *Zea*, *Flaveria* and *Setaria* (Heimann et al. 2013; Mallmann et al. 2014; Schulze et al. 2013; Wang and Gao 2015).

G. gynandra has been used as a eudicot C_4 model species because it is the closest C_4 relative of the C_3 model species *A. thaliana,* allowing for easy comparisons to the tremendous molecular and genetic knowledge base of Arabidopsis (Brown et al. 2005; Marshall et al. 2007). Three biochemical subtypes of C_4 plant species have been identified based on the predominant decarboxylase used during the C_4 cycle: the NADP-malic enzyme (NADP-ME), the NAD-malic enzyme (NAD-ME) or the phosphoenolpyruvate carboxykinase (PEPCK). Unlike most C_4 model species which are NADP-ME, *G. gynandra* is a photosynthetic NAD-ME subtype (Koteyeva et al. 2015 and 2011; Marshall et al. 2007; Muhaidat et al. 2007). Two sister species of *G. gynandra*, *Cleome angustifolia* and *C. oxalidea* also show $\delta_{13}C$ values characteristic of C_4 photosynthesis. C_4 photosynthesis evolved many times in the Cleomaceae, twice with atriplicoid-type anatomy in compound leaves with flat, broad leaflets in the pantropical species *G. gynandra* and the Australian species *C. oxalidea*, and once by forming a single Kranz unit in compound leaves with narrow leaflets in the African species *C. angustifolia* (Koteyeva et al. 2011). Comparing *G. gynandra* and its sister C_3 species, *Tarenaya hassleriana* (previously known as *Cleome hassleriana* and sometimes confused with *Cleome spinosa*), Kocacinar (2015) observed, as one would expect, about two times higher net assimilation rates in the C_4 than the C_3 species, more efficient use of high irradiation use and an increase of 14% of the net photosynthetic rate in the C_4 and a decrease of 15% in the C_3 species in response to an increased leaf temperature from 30 to 40°C. Also, water use efficiency (WUE) in *G. gynandra* was 3.5 times higher than in *T. hassleriana*. *T. hassleriana* has been considered as a C_3 model species for comparative transcriptomics analysis with *G. gynandra* (Bräutigam et al. 2011; Külahoglu et al. 2014; van den Bergh et al. 2014). Both species were compared to investigate the contribution of the whole genome triplication event in the Cleomaceae to evolution of photosynthesis within the family. In absence of a higher retention of C_4 related orthologs in *G. gynandra* compared with *T. hassleriana*, van den Bergh et al. (2014) concluded that the shift to C_4 photosynthesis in *G. gynandra* was triggered by recruitment of existing genes through regulatory modifications. Some *Cleome* species including *C. foliosa*, *C. africana* and *C. paradoxa* have been characterised as intermediate C_3-C_4 species (Marshall et al. 2007) and could be included in further comparative studies in order to elucidate the evolutionary history of C_4 photosynthesis in the genus. The genome of *G. gynandra*

could significantly propel research on C_4 photosynthesis and allow clarifying whether polyploidy events favoured the occurrence of C_4 photosynthesis in the Cleomaceae family. An analysis of the *Gynandropsis* genome would also be invaluable for identifying candidate genes related to key crop-related traits and accelerate the development of a molecular breeding program. However, it is first critical to establish the traits of interest for the most important users: farmers and consumers.

The Cleome project: farmers' preferences and definition of breeding objectives

G. gynandra is mainly cultivated for home consumption, although in some regions the plant has a commercial value. Understanding consumers' perceptions on the species and farmers' constraints will contribute to defining meaningful breeding objectives. The present thesis was initiated as part of a project entitled "Utilizing the genome of the vegetable species *Cleome gynandra* (spider plant) for the development of improved cultivars for the West and East African markets" (in short: The Cleome project). The consortium involved in the project included Wageningen University and Research (WUR) in the Netherlands, the Faculty of Agronomic Sciences of the University of Abomey-Calavi (FSA/UAC) in Benin, the World Vegetable Center based in Taiwan, the African Orphan Crops Consortium (AOCC) based in Kenya, the Kenyan Resource Center for Indigenous Knowledge (KENRIK), and the NGO Hortitechs Development in Benin. The project ran from January 2015 to December 2017.

At the inception meeting of this project in Benin, we discussed with women organizations from Dogbo, Southern Benin, an area where the species is cultivated in home gardens and used since centuries both as a vegetable and a medicinal plant. Their main concerns were: (1) low leaf yields associated with early flowering of the plant and reduced leaf size; and (2) incidence of pests and diseases (Cleome Consortium 2016). Similar concerns were raised by farmers in Kenya where the lack of seeds and the poor germination were also mentioned (Onyango et al. 2013c). Other constraints for women in Benin were related to the lack of financial and institutional support to expand their activities. When asked about the reasons why the species was preferred over other vegetables, farmers in both Benin and Kenya mentioned: (1) the nutritional value of the crop; (2) the medicinal properties of the plant, especially the antibiotic ones and the fact the regular consumption of the crop by elderly people kept them healthy; and (3) the rapid growth of the vegetable which could be harvested already four weeks after sowing. In Benin, it is common to harvest the vegetables two to three times through successive cuttings every two weeks, a practice that provides quicker returns than other commonly cultivated crops such as cereals and tubers (Cleome Consortium 2016). Building on farmers' concerns and priorities, the project team considered the development of late-flowering varieties with improved nutritional and nutraceutical values as well as good seed germination and resistance to pests and diseases as the main breeding objectives. It was also important to further investigate farmers' traditional knowledge on the species and to characterise existing production/management systems.

From January 2018 until the end of our PhD study, our research was financially supported by a personal grant from the Faculty for the Future Fellowship of the Schlumberger Foundation. This funding allowed us to expand our research to the molecular level and create a connection between farmers' preferences, phenotypic data and more fundamental knowledge about the genome of the species and the potential genes involved in our traits of interest.

Research questions and objectives of this thesis

This thesis started within the Cleome project and aimed overall at generating baseline information about *G. gynandra* in order to facilitate subsequent breeding efforts for improved yield and nutritional quality. To contribute to achieving this goal, we particularly focused on answering the following research questions:

i) What would be the pathways to tackle the challenges inherent to orphan leafy vegetable breeding?

ii) How much do drivers of the management of the species affect the selection of use and conservation options by local communities?

iii) What is the extent of natural variation in morphological traits and vitamin content of the species and how are nutrient and morphological traits related to geographical provenances?

iv) To which extent is the morphological variation in the species correlated with variation of the metabolome and genome?

v) What are the molecular mechanisms underlying flowering time and vitamin biosynthesis in the species?

To address these questions, we followed a multidisciplinary research approach which is detailed below.

Research approach

Germplasm collection and documentation

The regions of diversity of *G. gynandra* include Africa and Asia (Zeven and Zhukovsky 1975). At the start of the Cleome Project in January 2015, germplasm collections of *G. gynandra* were available in a number of local, regional and international gene-banks. For example, a total of 295 accessions were maintained at the World Vegetable Center including 112 from East and Southern Africa and 183 from Asia. Thirty-one accessions from Southern Africa were recorded at United States Department of Agriculture (USDA) genebank. Thirty-one accessions from Southern Africa were recorded at United States Department of Agriculture (USDA) genebank. Germplasm collections are also maintained in Botswana, Kenya, Namibia, Tanzania, Zambia and Zimbabwe (Mnzava and Chigumira Ngwerume 2004). However, there was an obvious gap in accessions from West Africa. As the region was part of the target market for *G. gynandra* breeding programs, we conducted germplasm collection missions in Benin, Togo, Ghana and Niger in collaboration with local researchers **(Figure 3).** A total of 164 new accessions from these countries were collected and are now maintained at the Laboratory of Genetics, Horticulture and Seed Science of the Faculty of Agronomic Sciences, University of Abomey-Calavi (Benin). In addition to collecting and geo-referencing plant material, we also discussed with local communities about the uses and management of the species after requiring their prior informed consent. Our observations on the variation in management practices and relative importance of the species for local communities from harvest in the wild to full stand cultivation in irrigated intensive vegetable production systems led us to further investigations reported in

Chapter 3 of this thesis.

Assessment of the genetic variation in traits of interest using a multidisciplinary approach

Breeding a crop for specific traits requires comprehensive knowledge about these traits. In this research, we combined different sets of data derived from morphological characterisation, targeted analysis of vitamins, untargeted metabolomics and genome-wide diversity analyses to reconstruct the origin and diversification of *G. gynandra* with an emphasis on important traits of interest for breeders.

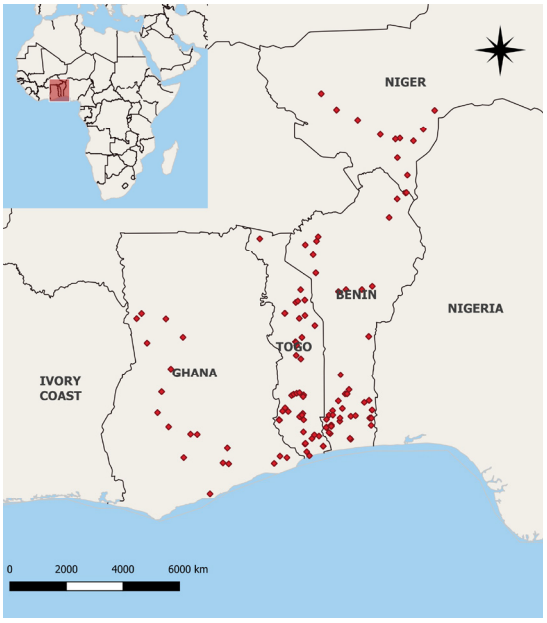


Figure 3. Map showing locations of germplasm collections in Benin, Togo, Ghana and Niger. Each red diamond represents a location where accessions were collected between 2015-2016.

The germplasm collection used in this thesis was a set of 76 accessions including accessions from East/Southern Africa, West Africa and Asia. The morphological characterisation of the collection allowed us to assess the variation in a range of quantitative and qualitative traits including leaf, flower, stem related traits as well as flowering time. The study was conducted from December 2016 to April 2017 under greenhouse conditions at Wageningen University. Our results were consistent with those of Wu et al. (2017), which included accessions from East/Southern Africa and Asia. Phenotyping of national or regional collections for morphological traits were also extensively performed including East Africa (Omondi et al. 2017b) and the Upper East region of Ghana (Kwarteng et al. 2018).

We also explored the variation in vitamin content in the collection through a targeted analysis of carotenoids, tocopherols and ascorbic acid which are respectively provitamin A, vitamin E and vitamin C related compounds. The choice of these compounds was dictated not only by their importance for human health but also by their contribution to the protection of the plants against oxidative damage. They are also important targets for molecular breeding.

Apart from the targeted analysis of vitamins, we were also interested in getting a broader view of the secondary metabolites synthetized in the plant. Secondary metabolites differ across plant lineages and biosynthetic pathways can often be lineage-specific (such as glucosinolates in Brassicales). This amazing chemical diversity emerged during the radiation of plant lineages in terrestrial habitats, most probably from a simple ancestral metabolism (Weng et al. 2012). Various types of genome rearrangements occurring during plant evolution, including local and whole-genome duplications, contributed to the extant diversity in plant secondary metabolites (Hofberger et al. 2013; Scossa et al. 2016). Secondary metabolites play an ecological function in plants and are involved in defense against herbivores, microbes, viruses and all kinds of predators (Saito and Matsuda 2010). Several secondary metabolites are also involved in plant-plant interactions (allelopathy) (Wink 2008) and responses to abiotic stresses such as drought stress cold or high salinity (Akula and Ravishankar 2011). From the human perspective, secondary metabolites in plants affect their taste and aroma and are also important for human health. Some of these metabolites are toxic for humans including oxalates, phytates, nitrates, tannins, saponins and alkaloids. In vegetables, the presence of oxalic acid is particularly detrimental to humans as this compound binds to minerals and decreases their bioavailibity (Lo et al. 2018). Other secondary metabolites play a key-role in preventing and treating chronic diseases such as cardiovascular diseases, diabetes mellitus, atherosclerosis, Crohn's disease, cancer, ulcerative colitis (Shayganni et al. 2016) and neurodegenerative disorders (Apetz et al. 2014). Although it might seem impossible to capture all the secondary metabolites produced in a plant, metabolomics is an analytical approach focused on generating the least biased and most comprehensive qualitative and quantitative overview of the metabolites present in a tissue, organ or whole plant (Hall 2018). Metabolomics appeared as the missing link in -omics technologies including genomics, transcriptomics and proteomics. We conducted the metabolomic fingerprinting of a subset of 48 accessions of *G. gynandra* from our three target regions for a more comprehensive assessment of the occurrence and natural variation in semi-polar and volatile compounds levels in the species. We discuss the potential metabolites involved in the taste and aroma of the leaves as well as the presence of metabolites that could play a role in human nutrition and defense against herbivores.

A hybrid assembly of the genome of *G. gynandra* was generated by combining Illumina sequencing and 10X Genomics technology. Genome annotation was performed combining *ab initio* gene prediction with comparative annotation using well-studied Brassicaceae. The genome also served as a critical starting point for analyzing natural variation and for molecular breeding. we analysed the whole-genome re-sequencing with Ion Proton and Illumina of 53 accessions (including the 48 ones used for the metabolic analyses). Subsequently, we performed SNP discovery and analysis. This included identifying and utilizing SNP markers for genetic mapping and subsequent quantitative trait loci (QTL) detection in a bi-parental population.

Association studies of traits of interest can be performed based on the analysis of the associations between markers, such as SNPs, across the genome and phenotypes in distinct plant accessions and/or varieties sampled to represent the natural variation in the trait of interest. However, this approach makes it difficult to detect rare variants unless the sampling size is adequate at the local level is large enough for that purpose. Moreover, looking for associations with single markers can mask the effects of multiple major loci in linkage disequilibrium (LD) with each other (Korte and Fallow. 2013). The confounding effect of population structure could make true causative

SNPs difficult to identify because they are in linkage disequilibrium with many loci in the genome (Brachi et al. 2011). Because of the low sample size and the high level of population structure (e.g. Asia vs. Africa) in our collection, we adopted a linkage mapping approach.

Linkage mapping consists in genotyping bi-(or multi-)parental populations using a pre-defined set of markers or sequencing the individuals directly and determining the marker set subsequently. Based on the co-segregation of markers in the population, the markers can be clustered together into linkage groups and spaced according to the genetic distances between them. Thus, associations are detected between traits and regions or intervals of linked markers which are often located near candidate genes (Kulwal et al. 2018). In our study, an F_2 population for which the parental lines had been re-sequenced was used for the genetic mapping and QTL detection. A targeted Genotyping-by-Sequencing approach was used for the genotyping of the population in an attempt to circumvent the low reliability of genotyping-by-sequencing data in F_2 populations (van den Bergh 2017) and the potentially high costs associated with other genotyping platforms (e.g. KASP, Hybrid capture next-generation sequencing).

Thesis outline

Our thesis aims at providing baseline knowledge on *G. gynandra* to inform breeding strategies for improved nutritional quality in the species using a multidisciplinary approach. The thesis is organized in eight chapters (**Figure 4**). **Chapter 1** (this chapter) provides background information about orphan crops and *G. gynandra* as well as a short presentation of the research framework of this thesis. **Chapter 2** presents a conceptual framework for orphan leafy vegetables breeding from the documentation of end-users preferences to the value chain development and mobilization of funding for sustainable breeding programs. Emphasis is put on leveraging modern -omics technologies to accelerate the breeding process. *G. gynandra* is used to illustrate how the devised strategies could be successfully applied considering the specificities of each crop. **Chapter 3,** the first step of the development of the breeding program, dives into the current state of knowledge and use of *G. gynandra* in seven socio-linguistic groups in Benin and Togo. The drivers of management of the species and the interventions needed in these communities to better promote the species are discussed. In **Chapter 4**, we examine the variation in morphology and nutrient content in a collection of 76 accessions of *G. gynandra*. Our analyses revealed differences in flowering time, plant height, leaf area but also levels of carotenoids and tocopherols between accessions from East/Southern Africa, West Africa and Asia. We also found a six-fold variation in ascorbic acid which is independent from the geographic origin of the accessions. In **Chapter 5**, we investigate the metabolic variation in the leaves of 48 accessions of *G. gynandra* from East/Southern Africa, West Africa and Asia. Semi-polar metabolites were detected using a liquid-chromatography mass-spectrometry platform while volatile metabolites were detected using a gas-chromatography mass-spectrometry approach platform. Semi-polar metabolites confirmed differences between accessions from East/Southern Africa, West Africa and Asia in terms of metabolic profiles. Volatile compounds were likely linked to the taste and aroma of the leaves. Our results are discussed in the light of their implications for human health, taste preferences and plant defense against herbivores. **Chapter 6** describes the draft reference genome of *G. gynandra* and provides an overview of the genetic diversity of 53 accessions based on resequencing data from the three abovementioned regions with demonstrated differences in morphological

features and metabolic profiles. We analyse the genetic diversity in the collection, revealing strong population genetic structure in the germplasm and association between metabolic profiles and genome wide nucleotide diversity. The genome of the species also provides insight in ancient polyploidy events that occurred in the Cleomaceae family. We discuss the implications of the results for breeding and from an evolutionary perspective. **Chapter 7** presents the first linkage map in *G. gynandra* using a segregating F_2 population. We identified QTLs associated with flowering time, plant height, leaf area and the biosynthesis of carotenoids, tocopherols and vitamin C and possible candidate genes underlying those traits. We also make suggestions for improving the quality of the genetic map and for validating the candidate genes. Finally, in **Chapter 8**, we discuss our findings and their contribution to current scientific knowledge on *G. gynandra*, highlight the limitations and challenges encountered during this research project and provide recommendations for further research aiming at improving specific traits in *G. gynandra* and orphan vegetable species in general.

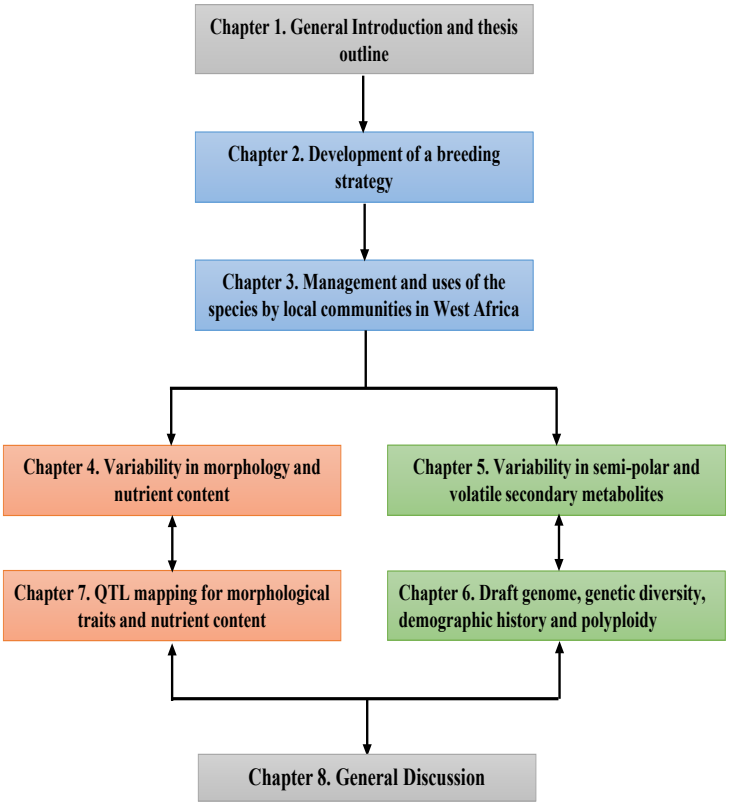


Figure 4. Outline of this thesis.

CHAPTER 2

A roadmap for breeding orphan leafy vegetable species: a case study of *Gynandropsis gynandra* (Cleomaceae)

E. O. Dêêdi Sogbohossou[1,2], Enoch G. Achigan-Dako[2], Patrick Maundu[3], Svein Solberg[4], Edgar M. S. Deguenon[5], Rita H. Mumm[6], Iago Hale[7], Allen Van Deynze[8], M. Eric Schranz[1]

[1]Biosystematics Group, Wageningen University, Postbus 647 6700AP, Wageningen, The Netherlands

[2] Laboratory of Genetics, Horticulture and Seed Sciences, Faculty of Agronomic Sciences, University of Abomey-Calavi, BP 2549 Abomey-Calavi, Republic of Benin

[3] Kenya Resource Center for Indigenous Knowledge (KENRIK), Centre for Biodiversity, National Museums of Kenya, Museum Hill, P.O. Box 40658, Nairobi, 00100, Kenya

[4] World Vegetable Center (AVRDC), P.O. Box 42, Shanhua, Tainan, 74199 Taiwan

[5] NGO Hortitechs Developpement, 02 BP 1111, Cotonou, Benin

[6] Department of Crop Sciences, University of Illinois, Urbana, IL 61801

[7] Department of Agriculture, Nutrition, and Food Systems, University of New Hampshire, Durham, NH, 03824

[8] Department of Plant Sciences, University of California, Davis, USA, 95616

Published in *Horticulture Research*, 2018, 5(1): 2. DOI: 10.1038/s41438-017-0001-2

Abstract

Despite an increasing awareness of the potential of "orphan" or unimproved crops to contribute to food security and enhanced livelihoods for farmers, coordinated research agendas to facilitate production and use of orphan crops by local communities are generally lacking. We provide an overview of the current knowledge on leafy vegetables with a focus on *Gynandropsis gynandra*, a highly nutritious species used in Africa and Asia, and highlight general and species-specific guidelines for participatory, genomics-assisted breeding of orphan crops. Key steps in genome-enabled orphan leafy vegetables improvement are identified and discussed in the context of *G. gynandra* breeding, including: (1) germplasm collection and management; (2) product target definition and refinement; (3) characterisation of the genetic control of key traits; (4) design of the 'process' for cultivar development; (5) integration of genomic data to optimize that 'process'; (6) multi-environmental participatory testing and end-user evaluation; and (7) crop value chain development. The review discusses each step in detail, with emphasis on improving leaf yield, phytonutrient content, organoleptic quality, resistance to biotic and abiotic stresses and post-harvest management.

Keywords: Orphan crop, leafy vegetables, *Gynandropsis gynandra*, food security, nutrients, genomics-assisted breeding

1. Introduction

One of the main challenges for agriculture in the coming decades is to meet the nutritional requirements of the nine billion people expected by 2050 (Kahane et al. 2013). World population growth, coupled with the effects of climate variability and increasing competition for water and land resources, makes achieving nutritional security an even more daunting task. While over 5,000 plant species are recorded as food plants (RBG Kew 2016), less than 20 species provide most of the world's food; and three cereals – rice, wheat, and maize – account for ~60% of calories and ~56% of proteins that humans consume directly from plants (Jacobsen et al. 2015; Lenné and Wood 2011). The bulk of edible species in the world are therefore non-commodity crops that are mostly overlooked by research and development initiatives. Thus, they are often referred to as orphan, minor, neglected, underutilized and/or unimproved crops. Orphan crops are also often indigenous, native species or those introduced centuries ago that are still used locally or even regionally, with much untapped potential to increase nutritional security (Gahukar 2014; Mayes et al. 2012). Such species contribute to regional diets, are often adapted to local environmental stresses, and may already be integrated into existing production systems. Yet, there is little investment to improve their productivity or quality.

Adding value to orphan crops can lead to better livelihoods and improved income generation, especially for smallholder farmers. Such species may also contribute to enhanced climate change mitigation via increased hardiness, reduced external inputs, and subsequent reduction of the carbon footprint of agriculture (Barbieri et al. 2014; Jacobsen et al. 2015; Padulosi et al. 2014). Despite this potential, orphan crops improvement has largely been absent from the global agricultural research agenda, presumably because the relevance of any given orphan crop species is highly geographically and culturally specific. Public agricultural funds are rarely allocated to enable orphan crop research and development, leaving farmers often unsupported in their quest for better use of local agrobiodiversity. Several challenges impede the utilization and conservation strategies of orphan crops, including low productivity, limited variety development, lack of consumer awareness, absence of a value chain, and loss of knowledge. Ongoing efforts in Africa and Asia to overcome such bottlenecks include the documentation of knowledge by the Plant Resources of Tropical Africa (PROTA) and the Plant Resources of South-East Asia (PROSEA) Programmes (prota4u.org; proseanet.org); the germplasm conservation and improvement efforts at the World Vegetable Center; the assembly and definition of genetic diversity of 101 orphan crop genomes and training of plant breeders by the African Orphan Crops Consortium Initiative (AOCC, africanorphancrops.org) and the Alliance for the Green Revolution for Africa (AGRA) to accelerate improvement of neglected and unimproved species of importance for local communities in Africa. How to translate these efforts into tangible breeding outputs for African markets remains an important issue that requires thorough attention. The urgent need to reduce malnutrition and hunger triggers the consideration of orphan leafy vegetables as a viable strategy recommended by FAO and WHO (Smith and Eyzaguirre 2007) to nourish the overgrowing world population. Strategies adopted to develop orphan leafy vegetables value chains should be aligned with the needs of local populations for access to nutritious and affordable food crops, well adapted to local conditions and available year-round.

This chapter serves both as a review of current knowledge and as a roadmap for the genome-enabled development of orphan leafy vegetables. These nutritious, short-cycle crops represent the bulk of African orphan crops (Maundu et al. 2009) and substantially contribute to local communities' safety nets during food shortage. The demand for these crops is increasing in urban areas of Africa as affordable and available sources of nutrients. Thus, they constitute a significant share of local vegetable markets. The diversity of these species across the continent, including the wide variation in production and consumption patterns, calls for the development of appropriate breeding strategies to meet both farmer and consumer preferences. However, for most of these species, basic knowledge is still lacking related to their reproductive biology, physiology, resistance/tolerance levels to biotic and abiotic stresses, the degree of natural variation, and genetic basis underlying traits of interest. Genomic resources are also lacking for leafy species, which have received less attention than other groups of orphan crops such as legumes (Varshney et al. 2009; Varshney et al. 2012), grain crops, millets (Cannarozzi et al. 2014; Varshney et al. 2017b), and root and tuber crops (Doungous et al. 2015; Utsumi et al. 2012). Additional considerations are therefore required for breeding of these species to highlight knowledge gaps and direct future efforts.

Throughout this review, the following questions are addressed: Why do we need to improve orphan leafy vegetables? What are the research gaps hindering production and promotion of these species? What would be the key components of a successful breeding program for orphan leafy vegetables, taking advantage of modern advances in genomics? To showcase ways and processes to develop cultivars of useful orphan leafy vegetables for Africa, we used spider plant (*Gynandropsis gynandra* (L.) Briq. syn. *Cleome gynandra* L.) as an example.

2. Why breed orphan leafy vegetables?

Orphan leafy vegetables play a significant role in livelihoods, nutrition and health in marginal areas of Africa. These crops are mostly grown and commercialized by women and contribute to income generation (Diouf and Ba 2014; Olabode et al. 2017; Weinberger and Pichop 2009). Urban and peri-urban orphan vegetable production employs vulnerable groups, often migrants who came to cities in search of jobs. In Senegal, the contribution of these species to the income of households can be as high as 100% (Diouf and Ba 2014; Oluoch et al. 2009). In East Africa, these species are most commonly cultivated and sold in local markets, supermarkets or green grocery stores, providing income to various stakeholders along the value chain (Onyango et al. 2013a). The average profit margin is estimated to be 30-45% of the selling price (Olabode et al. 2017; Weinberger and Pichop 2009). For example, *G. gynandra* contributes as much as 15-40% of the total income of some small-scale farmers in Kenya. The price for fresh leaves ranges from 0.40-0.50 USD/kg during the rainy season but can double in value during the dry season when vegetables are less readily available (Onyango et al. 2013a; Weinberger and Pichop 2009). A survey conducted on 861 indigenous vegetable retailers sampled in seven African countries revealed an annual turnover of 5.5 million USD. Weinberger and Pichop (2009) estimated that African indigenous vegetables market is worth billions of USD across Sub-Saharan Africa.

Orphan leafy vegetables can provide affordable and locally available sources of nutrients including vitamins, minerals and proteins (Abukutsa-Onyango et al. 2010; Chweya and Mnzava

1997; FAO 1990; Jiménez-Aguilar and Grusak 2015; Schönfeldt and Pretorius 2011; Steyn et al. 2001a; Uusiku et al. 2010; van Jaarsveld et al. 2014). Beyond their nutritional value, orphan leafy vegetables are also used as medicinal plants in various communities (Achigan-Dako et al. 2010; Grubben et al. 2004a; Kimiywe et al. 2007; Mensah et al. 2008). For instance, various parts of *G. gynandra* are used to strengthen the immune systems of women and children, as well as to cure wounds, diverse inflammations, digestive disorders, epileptic fits, and malaria (Allabi et al. 2011; Namukobe et al. 2011; Shanmugam et al. 2012). Its high vitamin (e.g. provitamin A, vitamin C) and micronutrient content (e.g. iron) makes the species particularly important in expecting and lactating mothers as well as in child development. Pharmacological studies revealed high concentrations of glucosinolates, flavonoids, tannins, iridoids and other phytochemicals in the leaves (Moyo et al. 2013; Yang et al. 2008), conferring to the plant proven antifungal, antibacterial, antiviral, anticarcinogenic, analgesic, febrifuge and anti-inflammatory properties (Bala et al. 2014; Ghogare et al. 2009).

Besides their great potential as both food and medicine, local landraces of orphan leafy vegetables are an asset to cope with climate variability. They are resistant to adverse environmental factors and can be easily grown in drought-prone areas with low rainfall (Capuno et al. 2015). However, attempts to breed more productive cultivars have been limited so far despite many features which make these species conducive to genetic improvement. Most of these species have a short cycle from 3 to 5 months and are predominantly self- or out-crossing with a certain rate of out- or self-pollination, which makes them amenable to different breeding strategies (Grubben and Denton 2004; Jansen van Rensburg et al. 2015). For example, *G. gynandra* has a short life cycle of 3 to 4 months, with plants tending to flower very early, within 4-6 weeks from planting. The species also shows substantial variation in reproductive characteristics relevant to domestication and crop improvement. Flowering is gradual, starting with the terminal shoot and followed by the axillary shoots, and may last for more than two months (K'opondo et al. 2005). The species is both self- and cross-pollinated (K'opondo et al. 2005), where cross-pollination is facilitated by wind or insects (e.g. honeybees, thrips and butterflies) (K'opondo et al. 2005; Raju and Rani 2016). Two types of flowers are commonly observed in *G. gynandra*: 1) a staminate type consisting of a residual ovary devoid of ovules, which can contribute to cross-pollination; and 2) a hermaphroditic type consisting of a functional ovary and stamens, which permit self-pollination (Chigumira Ngwerume et al. 1998; Raju and Rani 2016). Under stress, individuals bearing flowers with infertile reduced stamens were also observed. The species is self-compatible and facultative autogamous, with allogamy occurring occasionally (Raju and Rani 2016). Such characteristics are advantageous for the species, as they allow fruit set whether or not pollinators are available and additionally offer flexibility in breeding methods that can be applied to improve *G. gynandra*.

During the last thirty years, the World Vegetable Centre released 13 cultivars of orphan leafy vegetables obtained by single seed descent or mass selection in Tanzania, Uganda, Kenya and Mali. These include five African nightshades (*Solanum scabrum*), five amaranths (*Amaranthus* spp.), two Ethiopian mustards (*Brassica carinata*) and one jute mallow (*Corchorus olitorius*) (Afari-Sefa et al. 2012; Dinssa et al. 2016). Leveraging genomics-assisted breeding strategies to sustain current breeding efforts in orphan leafy vegetables in a concerted manner would therefore be beneficial for the development of commercial value chains for these species.

3. Developing breeding programs for orphan leafy vegetables

Developing breeding programs for orphan leafy vegetables begins with cultivar development based on consumer preferences and adequate adaptation to various ecological conditions, with precise product targets being dictated by individual market regions. Typically, smallholder farmers seek full-season varieties with high leaf yield, resistance to diseases and pests, and abiotic stress resistance (e.g. drought, heat and salinity tolerance). Retailers and consumers seek good appearance, long shelf-life, superior taste and aroma, high nutritional value, and affordability (Afari-Sefa et al. 2012). As both growers' and consumers' preferences are important in defining breeding objectives and product targets, they should be investigated at an early stage of the breeding program to guide germplasm collection and characterisation strategies, and then prioritized in later stages once information on genetic diversity and breeding constraints becomes available. Cultivar development with diverse stakeholder participation (including farmers, retailers, and consumers) can enable breeders to create varieties with desired traits, reduced adoption bottlenecks, and broad acceptability (Adeniji and Aloyce 2013). Furthermore, genomic tools have the potential to accelerate the entire cultivar development process, provided the product target is stakeholder-driven and well-defined.

In the next section, we propose a breeding framework based on a multi-disciplinary approach which takes advantage of modern advances in genomics and breeding to guide concerted, inclusive efforts by researchers to ensure improved nutritional outcomes for consumers. The following steps are identified as milestones for achieving a successful improvement program for orphan leafy vegetables: (1) germplasm assembly, characterisation, and management; (2) definition of product targets, (3) characterisation of the genetic control of key traits; (4) design of the process of cultivar development; (5) integration of genomics data to optimize that process; (6) multi-environment participatory trials and end-user evaluation; and (7) crop value-chain development (**Figure 1**).

3.1. Germplasm assembly, characterisation, and management

Knowledge of the taxonomy, distribution, and ecology of a crop species is a prerequisite for proper germplasm assembly. Such information is available for most orphan leafy vegetables (Grubben and Denton 2004) and should be sought for in species with such knowledge gaps.

With information of the taxonomy and distribution of the species, regions can be prioritized and germplasm collection strategies developed. All collections should be accompanied by detailed passport data, including accurate geo-referencing, habitat characterisation, and sampling methods, in order to facilitate the downstream interpretation of genetic data of genebank accessions (Volk and Richards 2011). Sampling strategies must be defined that ensure an optimal coverage of habitats relevant to key traits of interest (Brown 1989; Odong et al. 2013). Ethnobotanical data related to the importance, cultivation, and utilization of the species should also be recorded at each collection site, as such data help bring function-related structure to the collection of genetic diversity. Orphan leafy vegetables are usually conserved *ex situ*. *Ex situ* germplasm collections of some orphan leafy vegetables already exist in local, regional, and international genebanks but need to be expanded. For example, 295 accessions

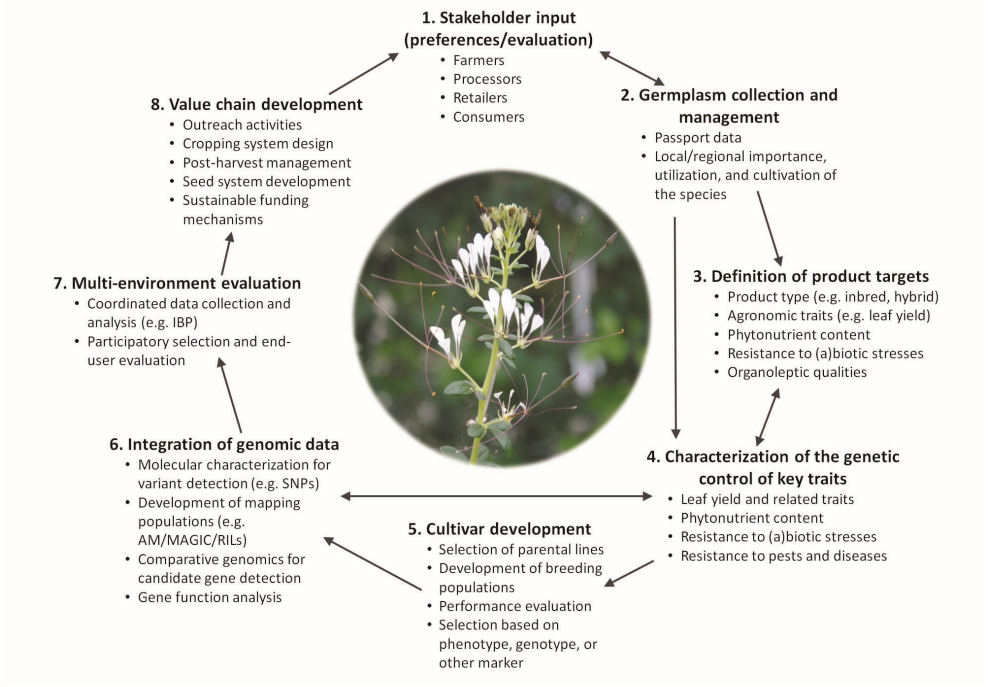


Figure 1. Schematic of an integrated breeding program for orphan leafy vegetables.

of *G. gynandra* are currently maintained at the World Vegetable Center, including 112 accessions from Eastern and Southern Africa and 183 from Asia (http://seed.worldveg.org/). Thirty-one accessions from Southern Africa are held within the National Plant Germplasm System of the USDA (http://www.ars-grin.gov/). Collections are also maintained in Botswana, Kenya, Namibia, Tanzania, Zambia, and Zimbabwe (Mnzava and Chigumira Ngwerume 2004). More recently, we assembled accessions from Benin, Togo, Ghana, Niger, Burkina Faso, and Kenya, resulting in a collection of 164 accessions from West Africa and 52 from Kenya. This new collection is currently maintained at both UAC and KENRIK and will soon be integrated into larger genebanks, such as that at the World Vegetable Center. In building this important new germplasm collection from West Africa and Kenya to support *G. gynandra* breeding programs, a standardized collection form was developed for the species (**Supplemental File 1**). Leveraging this initial work, future collection missions in South and Central America as well as Australia could help enhance the global collection of *G. gynandra* diversity. Available germplasm of orphan leafy vegetables must be continually enriched with purified lines, cultivars developed by research institutions, collections from farmers, as well as plant material collected from the wild. Conservation strategies to maintain genetic diversity depend on the mode of reproduction and fecundity of materials. Self-pollinated accessions are maintained as single plants when they are pure breeding lines but also as populations in the case of landraces and diverse materials. Cross-pollinated species are mainly maintained as populations paying attention to inbreeding depression and genetic drift. Collections should be performed and distributed in accordance with the national and international germplasm exchange policies such as the International

Treaty on Plant Genetic Resources for Food and Agriculture.

Germination and dormancy can be an important constraint for the successful conservation and utilization in orphan leafy vegetables germplasm as is the case for *G. gynandra* and *S. scabrum* (Oluoch et al. 2009). In *G. gynandra,* germination percentages between 25 - 65% were reported for seeds collected from research organisations in Kenya, compared with 15% from farmers' fields (Oluoch et al. 2009); and light exposure has been shown to inhibit seed germination (Muasya et al. 2009). Furthermore, a variable after-ripening period, ranging from three months (Ekpong 2009) to two years (Ochuodho and Modi 2005), has been shown to increase the germination rate up to 90%. Various pre-germination treatments, including imbibition with potassium sulfate (K_2SO_4) or gibberellin (GA_3), and germination at 30°C in the darkness, also improved germination rates (Motsa et al. 2015; Muasya et al. 2009; Ochuodho and Modi 2005; Zharare 2012). Protocols for proper seed conservation and efficient germination, without need for an extended after-ripening period, are needed to shorten the breeding cycle of *G. gynandra* and avoid the unintended erosion of *ex situ* genetic diversity due to selection against poor germination.

Linking genotypic and phenotypic variation to socio-ecological contexts is one means of gaining insight into the adaptation processes under different climatic conditions as well as the impacts of domestication on the species (Glaszmann et al. 2010). The long-term perspective should combine phenotypic and genotypic characterisation data for the development of core collections to be shared among genebanks for regionally-specific breeding goals (Glaszmann et al. 2010). However, the development of molecular markers for characterisation of the genetic diversity in orphan leafy vegetables should be pursued. To date, mainly second-generation molecular markers including random amplification of polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLPs), and microsatellite or simple sequence repeats (SSR) have been used to investigate genetic diversity in orphan leafy vegetable species (Omondi et al. 2016) such as *Corchorus olitorius* (Benor et al. 2012; Mir et al. 2008; Roy et al. 2006), *Brassica carinata* (Teklewold and Becker 2006; Warwick et al. 2006), *Solanum scabrum* (Manoko et al. 2008) and *G. gynandra* (Omondi et al. 2017a). Although these preliminary studies gave base information, the markers used are not economical for breeding. Discovery of single nucleotide polymorphisms (SNPs), insertions/deletions (indels) and copy number variation is yet to be exploited for most orphan leafy vegetables, although some exceptions have been reported for genetic characterisation of *Vigna unguiculata* (Huynh et al. 2013), *Brassica carinata* (Zhang et al. 2017) and *Amaranthus* spp. (Maughan et al. 2011; Maughan et al. 2009) which are also valued as pulse, oilseed and pseudo-cereal respectively.

3.2. Definition of breeding goals and objectives

The type of cultivar to be developed is a critical decision, based on the reproductive system of the crop, the presence of hybrid vigour and access to male sterility systems, as well as the seed distribution systems either already in place or to be developed in the market region. The aim is to develop a uniform, reproducible product that stably expresses all the target traits of interest and can be produced at low cost. Such a product can take the form of a pure line variety, an open-pollinated variety, a hybrid (e.g. single cross, 3-way cross, or double cross), or, in some cases, a clonally propagated cultivar. In Africa, where farmers may choose to save seed rather

than purchase seeds each year, a pure line variety may be preferred, if reproductive mechanisms allow it, although, maintaining performance and purity of cultivars through well-managed seed production is essential to gain the value of improved cultivars.

As is true for most crops, important target traits for orphan leafy vegetables include increased yield, higher nutritional content, resistance to pests, and tolerance to relevant stresses such as heat, drought and salinity. In addition, the maturity of any improved cultivar must be aligned with individual market regions; and consumer preferences must be honoured. For example, our recent investigations of farmers' preferences in Benin and Kenya revealed distinct regional flavour preferences for *G. gynandra*: no bitterness in East Africa and slight bitterness and spiciness with strong aroma in West Africa. Baseline data, together with ethnographical studies, are therefore essential in identifying not only promising breeding populations but also the environments and methods required for the meaningful evaluation of target traits.

3.3. Characterisation of the genetic control of key traits

Although some information on farmer and consumer preferences is available, little is known about the genetic control of key traits of interest; yet such knowledge is critical for designing a breeding program. Factors such as the number of genes controlling trait expression, the type of gene action involved (e.g. additive, dominant, epistastic), the magnitude of genetic and phenotypic variances, the possible interactions with the environment, and the heritability of key traits can influence the design of the 'process' by which improved cultivars will be developed. That such information is scarcely available for orphan crops implies that knowledge of genetic variances (additive, dominant, and epistatic) and heritability of traits of interest should be generated. Genetic variances are commonly estimated using procedures such as diallel, nested, and factorial designs; and the phenotypic evaluation of traits must take into account environmental and market specificity. Detailed descriptions of common estimation procedures are provided by Dudley and Moll (1969) and Fehr (1987).

Furthermore, other characteristics correlated with key traits may be utilized to devise more efficient and cost-effective breeding strategies. For example, simply-measured traits of an individual plant, such as height or total leaf number, may serve as reliable proxy metrics for a key system-level trait like leaf yield per area. Typically, the product target includes a balance of traits related to yield potential (e.g. resource partitioning and traits to protect that potential including defensive traits such as pest resistance and stress tolerance). It is important to understand as much as possible the genetic architecture of target traits from the start of a breeding program. For example, in the very early stages of a breeding program, association mapping using natural populations can be performed to explore population structure and the genetic control of the target traits, once phenotypic and genotypic data become available (Xu et al. 2017). Such preliminary results can guide subsequent efforts of molecular dissection of complex key traits; and more will be learned as strategic mapping populations are developed and genomic technologies are employed to map genes and estimate the magnitudes of their effects (Moose and Mumm 2008).

The phenotypic characterisation of selected traits in orphan leafy vegetables should be achieved using a set of standardized protocols, developed and shared among research institutes

in an effort to facilitate meaningful data comparison across environments. Morphological characterisation data is available in some species including *Amaranthus* spp. (Erum et al. 2012; Gerrano et al. 2015b; Sogbohossou and Achigan-Dako 2014), *G. gynandra* (Wu et al. 2017), *S. scabrum* (Stoilova et al. 2015), *V. unguiculata* (Gerrano et al. 2015a) and *C. olitorius* (Denton and Nwangburuka 2012; Ghosh et al. 2013). A standard list of morphological descriptors has been developed for *G. gynandra* by researchers at the World Vegetable Center and revised by the Cleome Consortium (**Supplemental File 2**). Gene banks and other institutions working on *G. gynandra* are encouraged to use and participate in the ongoing refinement of such a list of phenotypic descriptors. Descriptors used by the World Vegetable Center for some genera of importance including *Basella*, *Celosia*, *Cleome* [*Gynandropsis*], *Corchorus*, *Ocimum*, *Solanum*, *Talinum* and *Vigna* are available (http://seed.worldveg.org/download) and could be used for large-scale characterisation of target species.

Affordable techniques for high-throughput phenotyping (Araus and Cairns 2014; Pereyra-Irujo et al. 2012) should also be considered, especially as such methods are developed for other species with similar plant architecture. For example, the phenomics software Tomato Analyser (Rodríguez et al. 2010) developed for tomato was used to characterise other *Solanum* species including *S. macrocarpon* (Plazas et al. 2014). These methods are useful if objective data can be collected efficiently with high correlation to targeted phenotype.

3.3.1. Leaf yield-related traits

Leaf yield is the first and foremost target trait in leafy vegetables breeding. Breeding leafy vegetables requires proper characterisation and strong correlations between phenotypic traits and leaf yield. Such studies were undertaken for some species including *G. gynandra* (Omondi 1990), *A. tricolor* (Sarker et al. 2014; Sarker et al. 2015) and *B. carinata* though not extensively. In *G. gynandra*, Omondi (1990) estimated low heritabilities for leaf yield and yield-related traits, including plant height, number of leaves, leaf length, and leaf width, but a high heritability for days to flowering, a measure of maturity. Chweya and Mnzava (1997) reported a negative correlation between days to flowering and leaf dry weight in the species, indicating that leaf yield in *G. gynandra* may be maximized with full-season varieties. Further investigations including several genotypes and locations are required to validate these observations and assess other important traits. High heritabilities and genetic advances were estimated for leaf yield and positively correlated traits such as plant height, number of leaves, and stem diameter in *A. tricolor* (Sarker et al. 2014; Shukla et al. 2006). Therefore, leaf yield in *A. tricolor* could be significantly improved through direct selection for these traits. Further studies are required to document and better understand farmers' preferences in this respect. For instance, farmers who adopt a multiple harvesting strategy could be interested in plants developing many branches in a short amount of time, while late-flowering plants may not have this feature. Indeed, Chweya and Mnzava (1997) reported that good moisture supply at the early stages of plant growth promotes fast vegetative growth with reduced branching, while plant stress promotes early branching. Harvest index should also be considered, as late-maturing types may exhibit a high total biomass while the proportion of edible biomass may be low. The ability to regenerate after cuttings and the cost-effective number of cuttings should also be considered when selecting for leaf yield in regions where farmers adopt multiple cuttings (**Figure 2**). For example, in Kenya,

where whole plants of *G. gynandra* are uprooted 4-5 weeks after sowing, emphasis should be more put on fast vegetative growth rather than cutting ability. Finally, to date, there is little information available on the potential for hybrid vigour in leafy vegetables. A diallel crossing scheme could be used to develop materials to evaluate general and specific combining abilities, which could be exploited in cultivar improvement.

3.3.2. Phytonutrient content and consumer preferences

The significant impact of micronutrient deficiencies on human health, especially in developing countries, is gaining recognition (Afari-Sefa et al. 2012). Breeding for high-yielding varieties of orphan leafy vegetables that are both nutrient-rich and low in anti-nutritional factors may be an effective means of achieving biofortification, while simultaneously contributing to diet diversification and rural livelihoods.

Figure 2. Farmers harvesting *Gynandropsis gynandra* in a peri-urban garden (Benin). ©Photo credit: Edgar Deguenon

In a collection of one hundred accessions of *G. gynandra* from West Africa, East Africa, and Asia, nutritional content is being assessed by the Cleome Consortium using a multi-platform metabolomics approach able to provide a comprehensive characterisation of qualitative and quantitative variation in a wide range of metabolites. Analytical platforms include, but are not limited to: High Performance Liquid Chromatography-Photodiode Array-Fluorescence for apolar compounds (e.g. carotenoids, tocopherols, and ascorbic acid); Liquid Chromatography-Mass Spectrometry for semi-polar compounds (e.g. alkaloids, glucosinolates, flavonoids, and gallotannins); and Gas Chromatography-Mass Spectrometry for volatile compounds (e.g. amino acid derivatives, fatty acid derivatives, and terpenes). The results of such untargeted metabolomics

methods may be combined with proximate and mineral analyses to finely characterise nutrient content variation in germplasm collections and guide selection. Results from such analyses will therefore provide rational nutritional targets in *G. gynandra* as well as a basis for meeting them. With proper passport data, intraspecific variation in metabolic profiles may also be linked with cultural and/or geographical information. Moreover, the discovery and identification of metabolites with health-promoting properties could motivate further pharmacological studies on the species and provide incentive for utilisation of *G. gynandra* as a nutraceutical food. While there is available information on the nutritional value of orphan leafy vegetables (Schönfeldt and Pretorius 2011; Uusiku et al. 2010), such comprehensive analyses on natural variation in nutrient contents on wide germplasm collections are scarce and should be considered.

To help ensure adoption of improved varieties, the development of nutritional traits and nutrient-rich lines must be pursued with full knowledge of consumers' preferences. It is therefore essential that country- and/or region-wide organoleptic tests be performed, taking advantage wherever possible of existing partnerships and prioritizing areas of extant demand for target species. For example, the bitterness of *G. gynandra* is not desired in eastern African countries, as evidenced by the various cooking methods used to attenuate it. In Botswana, the leaves are initially blanched in water and the water discarded and replaced with a fresh supply (Flyman and Afolayan 2006). In Kenya, milk is added to the leaves in a pot and left overnight to improve the taste. Leaves are also mixed with those of other species, including *Amaranthus* spp., *Solanum* spp., *Basella alba* and *Brassica carinata* (Onyango et al. 2013b). In contrast, in West Africa and especially in Benin, bitter taste is more tolerated and even appreciated. In this region, bitter leaf (*Gymnanthemum amygdalinum*) is a popular vegetable and hence the bitterness in *G. gynandra* is not perceived as a negative trait. Such regional differences should be taken into account in breeding programs, especially given the objective of actively promoting and increasing the use of orphan crops. For *G. gynandra*, organoleptic tests may be conducted in West Africa and East Africa based on standard criteria selected with trained tasting panels (e.g. bitterness, spiciness, odour, texture). For instance, established correlations between bitterness or odour and specific metabolites could allow the early selection of *G. gynandra* lines with preferred taste profiles.

Cooking methods have a significant impact on the realized phytochemical content and antioxidant capacity of vegetables (Kunyanga et al. 2012; Palermo et al. 2014). Across sub-Saharan Africa, leafy vegetables are usually boiled (Mnzava and Chigumira Ngwerume 2004), blanched, or made into small balls and sun-dried for preservation (Flyman and Afolayan 2006). The differential effects of these and other cooking practices on nutrient content is not well studied in orphan leafy vegetables. Given the diversity of cooking methods used, it is important to assess the impact of common preparation methods on the bioavailability of specific phytonutrients and total antioxidant capacity in order to identify and recommend best cooking practices. Wide dissemination of results and open dialogue with consumers will be needed to understand and influence changes in traditional cooking practices while preserving culinary diversity. Progress in traits associated with phytonutrient content and flavour will also benefit from improved knowledge of their underlying genetic control, their relevant metabolic pathways, and the physiological processes involved. There is an opportunity to leverage comparative genomics between orphan leafy vegetables and well-studied relatives as well as metabolomics strategies toward this end.

3.3.3. Resistance to biotic stresses

Orphan crops are generally well-adapted to their environment and some species have developed chemical defences against specific pests. For instance, methanol extracts and volatile emissions of aerial parts of *G. gynandra* have been shown to have a strong acaricidal effect, especially on the two-spotted spider mite *Tetranychus urticae* (Kapsoot et al. 2015; Nyalala et al. 2013) as well as on both *Rhipicephalus appendiculatus* and *Amblyomma variegatum*, two livestock ticks occurring in Africa (Lwande et al. 1999; Malonza et al. 1992). The use of *G. gynandra* as a companion crop in plots of snap bean (*Phaseolus vulgaris*) significantly reduced the incidence of thrip species *Megalurothrips* and *Frankliniella occidentalis* (Waiganjo et al. 2007). Volatile compounds with significant repellent activity include aldehydes, terpenes, and isothiocyanates (Lwande et al. 1999; Nyalala et al. 2013), the latter being breakdown products of glucosinolates which occur after foliar disruption (Fahey et al. 2001).

Perhaps unsurprising, given the close relationship between *G. gynandra* and the Brassicaceae, most of the pests reported for the species also cause damage to cruciferous crops. For example, *Bagrada hilaris* (Palumbo et al. 2015) and *Phyllotreta* spp. (Soroka and Grenkow 2013) are serious economic pests of *Brassica* species, and such invasive species can be expected to have a stronger incidence in vegetable production systems where both *G. gynandra* and cruciferous crops are grown. Other pests commonly affecting orphan leafy vegetables production include caterpillars (e.g. *Helicoverpa armigera*, *Plutella xylostella, Spodoptera* spp.,), nematodes (*Meloidogyne* spp.), thrips, aphids (*Aphis* spp.) and whitefly (*Bemisia tabaci*). Judicious crop associations by leafy vegetable growers can therefore reduce the incidence of some pests and diseases.

Several insect pests of *Brassica* spp. are reported to be preferentially attracted to genotypes with specific metabolic profiles, related to levels of glucosinolates, amino-acids, and sugars (Kim et al. 2013; Nikooei et al. 2015), a result indicating both opportunities and difficult trade-offs. Indeed, innate chemical defences may be enhanced by orphan leafy vegetable breeders via direct selection under insect pressure or indirectly via metabolomics analyses. In either case, the factors affecting expression of insect resistance must be taken into account (Fritsche-Neto et al. 2012). Development of pest resistance must be prioritized on a regional basis.

With sufficient information on economically important diseases in orphan leafy vegetables and prioritization in terms of impact, effective field and greenhouse screening, and laboratory techniques can be developed for identifying tolerant and resistant genotypes. To the extent that such pests and pathogens are shared with more studied crops (e.g. economic *Brassica* spp.), such methods may already be well established. As with insect pests, resistance to pathogens is often associated with specific metabolites in plants. Thus metabolomics approaches may also be useful in detecting resistance-related compounds, particularly for use as biomarkers for selection. In general, morphological and molecular differentiation of pathogenic races and biotypes is a prerequisite for the efficient breeding of durable forms of resistance; and breeders must remain vigilant about the potential impacts of resistance to specific diseases on desired agronomic traits. Strategies for biotic stresses should integrate both management and breeding technologies.

3.3.4. Resistance to abiotic stresses

In the tropics, high-temperature conditions are often prevalent during the growing season and,

with a changing climate, crops in such areas will be subjected to increased temperature stress during developmental and productive phenostages (Kunchge et al. 2012). Phenotyping for drought and heat tolerance under controlled conditions can be used to pre-screen lines for further verification in the field. Even though field experiments are subject to variation, it is possible to closely monitor environmental parameters or design semi-controlled conditions using specific methods (e.g. rain shelters, irrigation, and enclosures) (Cattivelli et al. 2008; Langridge and Reynolds 2015; Lopes et al. 2014). The developmental stages at which these stresses occur, their duration, and their severity are also key factors for tolerance/resistance evaluation.

Besides yield evaluation under abiotic stresses, an understanding of the physiological processes underlying tolerance of abiotic stresses is needed to determine the physiological and morphological traits to include in selection criteria. For example, variation across genotypes in water uptake, photosynthetic efficiency, and water use efficiency traits like leaf conductance, photosynthetic assimilation rate, chlorophyll content, leaf thickness, leaf nitrogen content, and stable carbon isotope ratio could be investigated, particularly as they relate to leaf yield (Hall et al. 2005; Kumar et al. 2012). Leaf relative water content, wilt, and differential plant growth following drought-stress have been suggested as indicators of water stress in lettuce screening for drought tolerance (Knepper and Mou 2015); and such parameters could be considered to develop screening methods tailored to each vegetable species. C_4 species such as *G. gynandra*, *A. cruentus* and *A. tricolor* would develop different drought escape mechanisms compared with C_3 species. The available information on the genetic control of drought and heat tolerance in well-studied sister species, including *A. thaliana* (Bechtold et al. 2013; Bouchabke et al. 2008; Liang et al. 2011), *Brassica* spp. (Wu et al. 2012; Zhang et al. 2014) (e.g. *Brassica rapa*, *B. oleracea*, *B. napus*, *B. juncea* and *B. nigra*) for *G. gynandra* or *Chenopodium quinoa* (Jacobsen et al. 2003; Ruiz-Carrasco et al. 2011) for *Amaranthus* spp., should be used to facilitate the genetic characterisation of these traits.

3.4. Design of the process of cultivar development

Thorough knowledge about the reproductive biology of the targeted species will be a prerequisite for cultivar development. For predominantly self-pollinating species with some amount of outcrossing, one sensible product target would be pure line varieties (**Figure 3**). Controlled crosses can be performed via manual emasculation and outcrossing during breeding can be controlled by covering flowers. Later, varieties can be scaled up for seed distribution in isolated plantings to prevent outcrossing. If evidence is found for hybrid vigour in terms of leaf yield, hybrid cultivars could be pursued as a means to significantly increase leaf yield in the short-term. This can be specifically pursued for species like *G. gynandra* and *Solanum* spp. which are both self- and out-crossing. However, as mentioned before, for such a strategy to be commercially viable, seed production and distribution systems must be able to support hybrid seed production. Very likely, some mode of male sterility would need to be developed to facilitate hybrid seed production.

Once the form of the product has been decided (e.g. pure lines vs. hybrids), the critical elements of a breeding program focus on the choice of parents and the evaluation of progeny vis-a-vis well-defined trait targets. Pairs of lines that perform well for all of the key traits (high phenotypic value) yet offer diverse favorable alleles (high genotypic variance) represent ideal

parental combinations for crosses. Often referred to as a 'good-by-good' cross, such an ideal situation is rarely possible when breeding efforts are first initiated in a crop species.

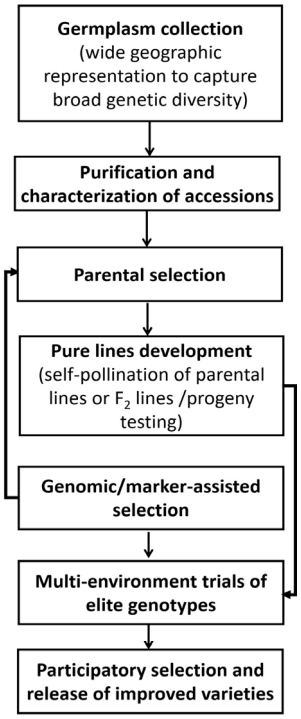


Figure 3. A breeding pipeline for the development of pure line varieties of *Gynandropsis gynandra.*

Instead, potential sources of favorable alleles for all the key traits of interest in available germplasm (e.g. leaf yield, time to maturity, resistance to pests and diseases, tolerance to heat and drought) should be identified and a plan for efficiently combining them set out. To begin, a testing system to evaluate the progeny of breeding crosses should be devised to accurately and precisely estimate the heritiabilities of and correlations among the traits in the refined product target, as well as to identify any related traits that could be a basis of selection toward the product target. The choice of parents will be based on the results of germplasm characterisation, as discussed in section 3.3, and the progeny selected based on performance for key traits will be tested at multiple locations representative of the range of agro-climatic conditions under which the final product will be cultivated (see section 3.6. for further details). Some orphan leafy vegetables (e.g. *G. gynandra*, *S. macrocarpum*, *Amaranthus* spp. and *C. olitorius*) are fast-growing plants and the whole production cycle (seed to seed) takes from four to six months, thus allowing two to three selection cycles per year.

Because product targets are never about a single trait, an approach to multi-trait selection must be determined. Such a strategy must take into account the relative importance of the traits for end-users, the nature of those traits (qualitative vs. quantitative), the complexity of their genetic control (e.g. additive, dominant, epistatic, additive x dominant, etc.), their degree of

heritability (low vs. high), their correlations to one another, and the selection intensity imposed by the breeder. One efficient means of breeding species of interest for multiple traits (e.g. maturity, leaf yield, and vitamin content) could involve the use of selection indices, particularly if significant negative correlations are found among target traits. To maximize gain via multiple selection cycles per year, it is likely that some selection will need to be done in off-seasons, when phenotypic data may be less reliable. In such cases, marker-assisted recurrent selection using markers with significant effects (Beyene et al. 2016) or genomic selection (Guo et al. 2012; Massman et al. 2013) with dense, genome-wide sets of markers could be pursued. In the latter case, training populations could be developed based on available association panels; and indeed, genomic selection methods have been used in other crops for accurate phenotypic prediction of a wide range of traits (Lipka et al. 2014; Schmidt et al. 2016; Schulthess et al. 2016). Such genotype-based approaches will increase in their efficacy as deeper knowledge about target trait-related gene function and variation is attained. When dealing with specifically known, single genes, marker-assisted backcrossing could be explored for cultivar development (Feng et al. 2015; Iftekharuddaula et al. 2016; Vishwakarma et al. 2016). Such an approach seems particularly promising for introgressing disease resistance genes in species of interest and pyramiding them in high-yielding, nutrient-rich cultivars.

3.5. Integration of genomic data

Molecular markers associated with traits or as a tool for whole-genome selection are a valuable asset for efficient breeding. Key genomic resources available in *G. gynandra* include a draft reference genome (Schranz *et al.*, unpublished) and a transcriptome atlas of the species (Külahoglu et al. 2014). Building upon these research efforts will provide opportunities for increasingly detailed analysis of genetic diversity within the species as well as accelerated trait development in breeding programs.

To facilitate the development of a dense, genome-wide set of molecular markers, 100 diverse *G. gynandra* accessions are being sequenced by the AOCC; and their sequence data will be aligned to the reference genome for global variant calling. Other orphan leafy vegetables on the agenda of AOCC include *Amaranthus blitum, A. cruentus, A. tricolor, Basella alba, Brassica carinata, Celosia argentea, Corchorus olitorius, Crassocephalum rubens, Moringa oleifera, Solanum scabrum* and *Talinum fruticosum*. On the basis of both genotypic and phenotypic data, divergent parents can be identified and Recombinant Inbred Lines (RIL) populations developed to generate genetic linkage maps, enable QTL analyses, and highlight genomic regions of particular interest to breeding programs. They also enable marker-assisted backcrossing programs for foreground selection of traits and background selection for quick recovery of the recurrent parent genotype. In addition to bi-parental populations and association mapping panels, Multi-parent Advanced Generation Inter-Cross (MAGIC) lines could also be used to dissect traits of interest (Platt et al. 2010; Sallam and Martsch 2015). The multiple cycles of inter-crossing among multiple founder lines in MAGIC populations give greater opportunities for recombination, permitting both greater precision in QTL location (Huang et al. 2015) and an increased probability of favourable combinations of alleles from the multiple parents. MAGIC populations have been successfully used for QTL mapping in durum wheat (Milner et al. 2015), barley (Sannemann et al. 2015) and rice (Bandillo et al. 2013), to name a few. As developing MAGIC lines is time- and

resource-intensive, this approach could be used to complement classical linkage and genome-wide association mapping.

Genotyping services such as DNA extraction and Kompetitive Allele Specific PCR (KASP) genotyping for single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) are provided for AOCC-designated orphan crops by the LGC Group (www.lgcgroup.com), a member of the AOCC. For orphan crops with no reference genome, recent whole-genome analyses and *de novo* SNP calling pipelines like GBS-SNP-CROP (Melo et al. 2016) present highly cost-effective alternatives for immediate implementation. Services for such whole-genome analyses are available through the Biosciences East Central Africa (BecA) hub in Nairobi. In the case of *G. gynandra*, the draft reference genome enables variant calling and genotypic characterisation via a number of open-source pipelines and complementary bioinformatics tools.

In terms of genomic analysis, including functional characterisation, advantage should be taken of the significant synteny among orphan species and well-studied crops. Genomics-assisted breeding in *G. gynandra* and *B. carinata* could tap into available information on *Brassica* spp., and *Arabidopsis thaliana* genomes (Schranz and Mitchell-Olds 2006; van den Bergh et al. 2014), *S. macrocarpon*, *S. aethiopium* and *S. scabrum* on *S. lycopersicum* and *S. tuberosum* genomes; vegetable amaranths (*A. blitum*, *A. cruentus*, *A. tricolor* and *A. dubius*) on the recently published *A. hypochondriacus* (Clouse et al. 2016) and *Chenopodium quinoa* genomes (Jarvis et al. 2017). Exploiting the available comparative data on physiology, genetics, and "omics" in well-studied crops is an attractive avenue for candidate gene identification in orphan leafy vegetables. In addition, reverse genetics approaches such as Targeting Induced Local Lesions IN Genomes (TILLING) can be used as a high-throughput approach for functional genomics based on well-characterised genes in closely-related species. Gene editing techniques could also be utilized if needed.

3.5. Multi-environment testing and end-user evaluation

3.6.1. Locations and seasons

Multi-location evaluations must be carried out throughout the breeding process, with emphasis on the end-user acceptability of resulting advanced lines of the vegetables of interest. During the cultivar development process, decisions on the number and locations of testing sites for evaluation of populations and developed lines should take into account the range of agro-climatic conditions under which the species is cultivated, the existing breeding stations or experimental farms, and the available resources to allocate to such experiments. For example, in West Africa, *G. gynandra* is cultivated in both semi-arid regions (e.g. southern regions of Niger and Burkina Faso, northern Benin, Togo and Ghana) and sub-humid areas (Central and Southern Benin, Togo and Ghana), under both rain-fed and irrigated cultivation systems. The suite of testing sites should therefore be representative of the target market region, and the selection process should account for region-specific breeding targets; and running multi-year trials would help ensure accurate evaluation of variety performance. Importantly the sites selected should reliably and effectively differentiate lines for their target environments. Many different tools are available to assist breeders throughout the selection process (e.g. GGE Biplot, R package "selectiongain") (Mi et al. 2014; Yan et al. 2007). These can be used to optimize the selection and testing environments. The Breeding Management System software developed by the Integrated Breeding Platform (www.integratedbreeding.

net) is particularly noteworthy, as it facilitates the statistical analysis of genotype x environment interactions, employs mixed model analysis to compute estimates of heritability, and offers Best Linear Unbiased Predictors (BLUPs) and Estimates (BLUEs) to facilitate selection. The platform also serves as an efficient repository for the vast amount of phenotypic, genotypic, and genomic data generated by a breeding program.

3.6.2. Agronomic practices pertaining to optimal performance of cultivars

A breeding program seeks to develop new high-yielding cultivars adapted to the range of environments in the target region; therefore, investigating the impact of agronomic practices on desired traits is a critical aspect of multi-environment testing of both breeding lines and products destined for release. Whenever possible, the agronomic practices utilized in testing should reflect existing cropping systems in the market region. More generally, the various impacts of different fertilization schemes (organic or chemical fertilizers, doses and frequencies of applications), soil tillage regimes (tillage vs. no tillage), irrigation practices (doses and frequencies of water supply, waterlogging), planting densities, and harvesting modes (e.g. rooting, cutting, leaf picking) on growth, regrowth (when applicable), and overall yield are important factors to take into account throughout the breeding process. Once genetically improved varieties are created, a final step before release involves identifying agronomic practices to maximize cultivar performance and productivity.

3.6.3. End-user evaluation

The foremost goal of a breeding program is the wide adoption of released cultivars that satisfy the needs of both producers and consumers. End-users should therefore be involved at different stages of the breeding process, including the up-front definition of breeding objectives, the selection of populations or lines with superior characteristics, and the evaluation of final products which also serve as starting points for further breeding efforts.

The outputs of the breeding program should be first presented to target producers for evaluation and selection in coordinated structured trials. The "mother and baby" trial design (Snapp 2002), which consists of within-site replicates of a set of cultivars/accessions on research stations (mother) and single replicate satellite trials of subsets of cultivars (babies) in larger plots on farmers' fields, could be adopted as an integrated way to involve farmers, local seed companies, as well as research institutes. Such a design allows elicitation of the evaluation criteria of growers, assessment of the impact of farmer practices on cultivar performance, and development of guidelines for farmer field schools. Consumer preferences may be determined based on tasting panels and/or test cultivars brought to points-of-sale to assess their marketability and storability. End-user preferences may differ according to socio-cultural, ecological, and economic contexts; therefore, these dimensions should be taken into account to define the geographical areas for such exercises.

3.7. Orphan leafy vegetables value chain development

Translating the research efforts for orphan leafy vegetables breeding into concrete outputs for end-users requires the creation of sustainable collaboration frameworks for stakeholders along the value chain. Analysis of constraints and opportunities for the development of these

species as commercial crops should involve researchers, farmer organizations, seed companies, traders, policy-makers, and consumers. Outreach activities could include promotional campaigns highlighting the nutritional benefits and commercial opportunities provided by the species, in partnership with restaurants, schools, and the media. Other key components of value chain development for orphan crops include the development of cropping systems with agronomic practices that facilitate maximum productivity; the creation of post-harvest management best practices, addressing such issues as appropriate harvesting time, drying, and packaging methods to ensure optimal shelf-life; and the provision of adequate resources for farmer training. Concerted and coordinated effort for the improvement of vegetable seed systems by researchers, seed companies, and farmers is needed to ensure the development of standardized seed quality control regulations and the release and distribution of readily-available, high-quality seeds. A well-established seed production system using established techniques including a controlled seed multiplication system with isolation, genetic and purity checks as well as a seed distribution system is essential. Outreach about the value of improved cultivars is critical through extension and early adopter farmer trials. The importance of good seed stewardship cannot be overstated. To this end, AGRA has created over 114 seed companies through its programs to address these needs in Africa. This includes training of growers to produce high quality seed.

Ultimately, the ongoing funding of breeding programs should rely on collaboration between local seed companies and other stakeholders involved in orphan leafy vegetables value chain development. Financial and technical support provided by international organizations involved in orphan crop breeding, such as the Consultative Group for International Agricultural Research (CGIAR) and the African Orphan Crops Consortium, among others, is also critical for capacity building and strengthening of local plant breeding capacity.

4. Conclusion and the road forward

Large-scale cultivation and increased commercialization of orphan leafy vegetables is desirable given their excellent nutritional properties and economic potential to stabilize nutritional food security in developing countries. Large-scale production cannot, however, be achieved without concerted efforts from stakeholders across the value chain, from research to production to marketing to end-use. This review highlighted the recent efforts for catalysing the systematic breeding and promotion of large-scale cultivation of these species in West and East Africa. The synthesis emphasized the need for (i) raising awareness of the potential of orphan leafy vegetables to contribute to food and nutritional security in Africa; (ii) increased and coordinated germplasm collection and characterisation of the species; (iii) investigation of the genetic, physiological, biochemical, and molecular processes underlying key traits of interest; (iv) traditional genetics and genome-enabled research targeting trait development; (v) breeding efforts taking advantage of the advances in "omics" disciplines and the available comparative resources in related species; and (vi) expanded collaboration among local researchers, value chain stakeholders and international organizations interested in orphan crops to sustain technical and financial support for orphan crops breeding programs.

Acknowledgements

This study was supported by the Applied Research Fund of the Netherlands Organization for Science under the Project "Utilizing the genome of the vegetable species *Cleome gynandra* for the development of improved cultivars for the West and East African markets" (Project Number: W.08.270.350) and the African Orphan Crops Consortium.

Supplemental Files

Supplemental Information can be found on the Horticulture Research website https://doi.org/10.1038/s41438-017-0001-2.

Supplemental File 1. *Gynandropsis gynandra* specimen collection form

Supplemental File 2. Morphological descriptors of *Gynandropsis gynandra*

CHAPTER 3

Drivers of management of spider plant (*Gynandropsis gynandra*) across different socio-linguistic groups of Benin and Togo

E. O. Dêêdi Sogbohossou[1,2], Enoch G. Achigan-Dako[2], Tinde van Andel[1,3], M. Eric Schranz[1]

[1]Biosystematics Group, Wageningen University, Wageningen, the Netherlands

[2]Laboratory of Genetics, Horticulture and Seed Sciences, Faculty of Agronomic Sciences, University of Abomey-Calavi, Abomey-Calavi, Republic of Benin

[3] Naturalis Biodiversity Center, Leiden, the Netherlands

Published in *Economic Botany,* 2018, 72(4): 411-435. DOI: 10.1007/s12231-018-9423-5

Abstract

We investigated the relationships between the cultural importance of spider plant (*Gynandropsis gynandra*), a neglected leafy vegetable in West Africa, and the different management regimes of the species among six socio-linguistic groups in Benin and one in Togo. Semi-structured interviews were conducted with 428 respondents. Cultural significance and management indices were used to quantify the importance of the species for each respondent. In addition to food uses, *G. gynandra* was used to cure 42 different diseases. Regression tree analysis revealed that the cultural importance and level of management of the species were strongly associated with ethnicity and gender, and to a lesser extent to age, education, income and land tenure. Socio-linguistic groups with similar cultural background had convergent perceptions of the cultural importance of the species and described similar management practices. An analysis of farmers' willingness to change their current management practices revealed that migration, market opportunities and external intervention might significantly affect future management decision-making processes. We discussed community-oriented approaches to upscale the species cultivation in the region. Our study highlights how cultural importance influences current and future management intensity and illustrates how ethnobotanical research can guide research for development strategies to enact positive changes in communities' management of traditional leafy vegetables.

Key-words: neglected and underutilized species, traditional knowledge, management, ethnicity, West Africa.

Introduction

Wild edible plants constitute an essential food source used around the world and contribute to diet diversification and livelihoods in several communities (Ong and Kim 2017; Pawera et al. 2017). New trends in nutrition and the rising popularity of functional foods led to a regain of interest for these species and documentation of their utilization (Łuczaj et al. 2012; Romojaro et al. 2013). Some studies reported a decrease in the knowledge and consumption of wild edible plants associated with modernization of communities' lifestyles (Menendez-Baceta et al. 2017) and urbanization (Leal et al. 2018; Reyes-García et al. 2005). Among the Ati Negrito community in Philippines, knowledge and use of wild edible plants was positively correlated with age, past experience of hunger, household size, but negatively affected by education and access to media and social services (Ong and Kim 2017). However, in some contexts, modernization and acculturation enriched plant knowledge and use through interaction with different communities and diversification of the sources of knowledge. Plant knowledge and use was rather greatly influenced by personal experiences, degree of interest for plants, age and occupation (Mathez-Stiefel and Vandebroek 2012; Quinlan and Quinlan 2007; van den Boog et al. 2017). Participation in wild food collecting also differs regionally and may vary according to the income level, age, gender, opportunities to collect wild food, and cultural factors (Schulp et al. 2014). Cruz et al. (2013) found that while knowledge about wild edible plants was related to age, their current use was not associated with age, gender or occupation. Migration patterns also play an important role in dynamics of plant knowledge and use, and induced changes vary depending on differences in social, cultural, economic, institutional and ecological contexts between migrants' home country/region and their host country/region (Medeiros et al. 2012). Knowledge exchange and significant cultural interactions may occur and migrant communities may adapt to the flora, knowledge and traditions of the host country (Ceuterick et al. 2008; Volpato et al. 2009). This process is illustrated by the case of Tyrolean migrants and their descendants who migrated from Austria to Australia, Brazil and Peru, 50 years, 80 years and 150 years ago, respectively. Investigation of their knowledge of medicinal plants revealed that in all three countries they abandoned specific medicinal plants and related practices from the original pharmacopoeia when the plants were neither available nor cultivated in the country of arrival (Pirker et al. 2012). Migrants might also actively preserve the flora and knowledge from their home country or region as an attempt to conserve their cultural identity, as for example the case of the Dominicans living in New York City (Vandebroek and Balick 2012) or the Surinamese in the Netherlands (van Andel and Westers 2010).

Associations between the body of knowledge developed by local communities and the harvesting strategies adopted have been increasingly investigated for informed conservation and management strategies of those useful plants. Typologies of plant management types take into account a gradient of complexity of practices and artificial selection (Vodouhè and Dansi 2012). González-Insuasti and Caballero (2007) described different management strategies including gathering from the wild, incipient non-selective management, incipient selective management and occasional cultivation as a gradient of manipulation of plant resources. The intensity of management of a plant is determined not only by its biological characteristics but also its cultural and/or economic importance and its availability (Blancas et al. 2013; González-Insuasti

et al. 2008; N'Danikou et al. 2015). Many plant resources are managed in a variety of incipient forms and these forms may coexist for populations of the same species (Blancas et al. 2010; González-Insuasti and Caballero 2007). For example, local resources management decisions in the Raramuri community in Mexico were based on a culturally embedded understanding of ecological processes. Different management regimes including selective harvesting, pruning and favourable environmental modifications were applied to different groups of wild edible plants to ensure their sustainable use (LaRochelle and Berkes 2003).

Human cultural values and traditional ecological knowledge of plant resources are therefore crucial for making management decisions in order to ensure or increase the availability and/or quality of desired plant resources (Blancas et al. 2013). However, despite good ethnoecological knowledge and awareness of sustainable harvesting practices, market forces might bring communities to abusive harvesting of wild plant resources to meet both market demand and their immediate economic needs (Ghimire et al. 2004; Sundriyal and Sundriyal 2004). González-Insuasti et al. (2011) suggested a quantitative approach to investigate the determinants of management intensity tested for 20 edible plants under different management forms within a rural community of the Tehuacan Valley, Mexico. Management forms were influenced by socio-cultural variables such as age, education and occupation as well as food preferences. The type of land tenure was also important, as commercially important species tended to be managed in communal areas whereas non-commercial food species were managed more intensely in private areas (González-Insuasti et al. 2011).

Gender is also a decisive factor influencing management practices in communities where there is a gender differentiation of farm activities (Ekué et al. 2010; Vodouhè and Dansi 2012). Cruz et al. (2013) found that the degree of management of native wild edible plants in a local community in Brazil could be mainly explained by age and occupation. Younger respondents had lower motivation to tolerate wild edible plants than older ones. Farmers were also more incline to manage the species than non-farmers. Blancas et al. (2013) developed an index to quantify management intensity of plant species based on eleven indicators related to energy invested, types of tools used, complexity of regulations and institutions, artificial selection intensity and plant species biology. Investigation of management motives revealed that risk indicators on plant resource availability including life cycle, reproductive system, distribution, number of uses and regulation of uses significantly influenced management intensity.

The existence of various human cultures, their distinct culinary costumes, the variable range of socio-economic situations, the differential availability and forms of access to resources, help to understand that management of plant resources at each location depends on multiple factors that need to be analysed particularly for species and communitarian contexts (Arellanes et al. 2013). Cross-cultural studies of plant use and management described different scenarios with emphasis on the interconnections between socio-cultural and ecological systems. Traditional communities with a strong connection with their environment and who depend on natural resources for livelihood had deeper plant knowledge and frequently used these species compared with urban communities (Monteiro et al. 2006; Soares et al. 2017). Communities sharing the same cultural background are also likely to have convergent patterns of plant knowledge and use. Exceptions arise when communities of the same socio-cultural background live in contrasting

phytogeographical areas. Migrant communities are obliged to adapt their management practices to the flora of their new environment and domestication may become a strategy to ensure availability of the culturally important species (Kujawska and Pardo-de-Santayana 2015; Segnon and Achigan-Dako 2014). Likewise, communities with different cultural backgrounds may share the same uses of wild food plants when living in the same environment (Ghorbani et al. 2012).

The way complex interactions between local context, cultural values, and traditional knowledge might affect wild edible plant resources management in the future has been given less attention. Moreover, drivers of such change need to be carefully investigated to determine leverage actions and enhance sustainable utilization. A comparison of records of wild edible plants uses in Belarus between the 19th and the 21th centuries revealed a conservation of edible fruits use but a decrease in knowledge and use of wild edible vegetables (Łuczaj et al. 2013). Taste appreciation is likely to have played a crucial role in the preservation of some wild vegetables over time (Łuczaj et al. 2013; Ong and Kim 2017). Urbanization coupled with increased knowledge exchange between different communities (Georgian and Emshwiller 2013), communities' migration and adaptation to new environments, market opportunities (Arellanes et al. 2013; Reyes-Garcia et al. 2005), education and patterns of knowledge transmission within communities (van den Boog et al. 2017) are all factors which can affect decision-making processes related to wild edible plants management.

The current management spectrum of spider plant (*Gynandropsis gynandra* (L.) Briq.), a neglected leafy vegetable in Benin and Togo (West Africa), offers a good example to study current management intensity of the species and understand how its management and use could change in different communities to inform strategies for sustainable utilization and conservation. Spider plant is cultivated in some communities in East and West Africa (Abasse et al. 2007; Oluoch et al. 2009) as well as Asia (Arora 2014). The species is rather found wild or feral in other parts of Africa (Dovie et al. 2007; Kidane et al. 2015; Segnon and Achigan-Dako 2014). Initiatives for promoting the species in commercial farming systems are rapidly emerging, especially in East Africa (Onyango et al. 2013c) as a result of increased recognition of its multiple nutritional and health benefits (Omondi et al. 2017b).

The present study aims at documenting the current management practices on *Gynandropsis gynandra* in local communities of Benin and Togo to explore drivers for future positive change. Using quantitative ethnobotanical data, we: 1) examined the variation in the traditional knowledge and management practices of *G. gynandra*; 2) identified factors that influenced such traditional knowledge and management practices; 3) analysed farmers' willingness to improve their current management practices, i.e. collectively moving the current practices to a higher management level in the process of domesticating *G. gynandra*. We hypothesized that: (1) current management practices are shaped by factors including ethnicity, economic and cultural values of the species, and existence of local market-oriented vegetable production systems; and (2) willingness to improve the species' management in the future is influenced by ethnicity and cultural importance of the species.

Materials and methods

Study area and sampling strategy

Based on information available on *G. gynandra* distribution and use in literature (Achigan-Dako et al. 2010; Akoègninou et al. 2006) and preliminary observations made during germplasm collection missions conducted from April to June 2015 in Benin and Togo (Cleome Consortium 2016), six socio-linguistic groups were selected in Benin: the Fon (17.6% of the country's population), Adja (8.7%), Holli (1.4%) in southern Benin and Waama (1%) Gourmantche (0.5%) and Zerma (0.1%) in the North (**Figure 1**). These six socio-linguistic groups use *G. gynandra* as a food plant and were chosen to represent the different management regimes. The Ewe, Adja and Fon are the three most important communities of the Gbe language group (Hounkpati 1991) while the Holli belong to the Ede language group (CENALA 2003). Both Gbe and Ede linguistic groups form the Kwa language family (CENALA 2003). The Waama and the Gourmantche belong to the Gur language family (CENALA 2003). The Zerma living in Northern Benin are migrants from Southern Niger (Dosso, Tillaberi), engaged in the trade of cereals and manufactured products (Walther 2014). The Zerma belong to the Songhay language group which is considered to be unrelated to any other known language or language group (Muhammad 2016). Ewe farmers in Togo cultivate *G. gynandra* for commercial purposes. Thus, the Ewe community was selected as a reference group for the comparison of management practices in Benin.

From July to August 2016, semi-structured interviews were conducted with 57 to 68 respondents per socio-linguistic group, using purposive sampling (Albuquerque et al. 2014; Tongco 2007). This sampling method allowed us to select respondents who know the species and consume it or used to consume it in each socio-linguistic group and to obtain a reasonable compromise between sample size and available resources to conduct our study. Although our sampling is biased because not representative of the whole community, this method was proved to be the most adequate for ethnobotanical studies which require respondents with specific qualifications for hypothesis testing (Albuquerque et al. 2014; Tongco 2007). The surveyed localities were chosen with the help of extension services agents in each municipality, while the local authorities at each site helped us establish the first contact with consumers of the species. We shared full disclosure on the nature of the research, received authorization from local authorities in each locality in Benin and Togo, explained the objectives of the study and requested prior informed consent from all participants, and ensured participants' confidentiality by anonymizing their identities in databases and publications.

Interviews

Our questionnaire (**Appendix 1**) contained general socio-demographic characteristics (age, gender, ethnic group, education level, occupation, number of years of local residency, area of origin (if migrant), monthly income, land ownership). Other questions were related to the economic importance of *G. gynandra*. In areas where the vegetable was sold on local markets,

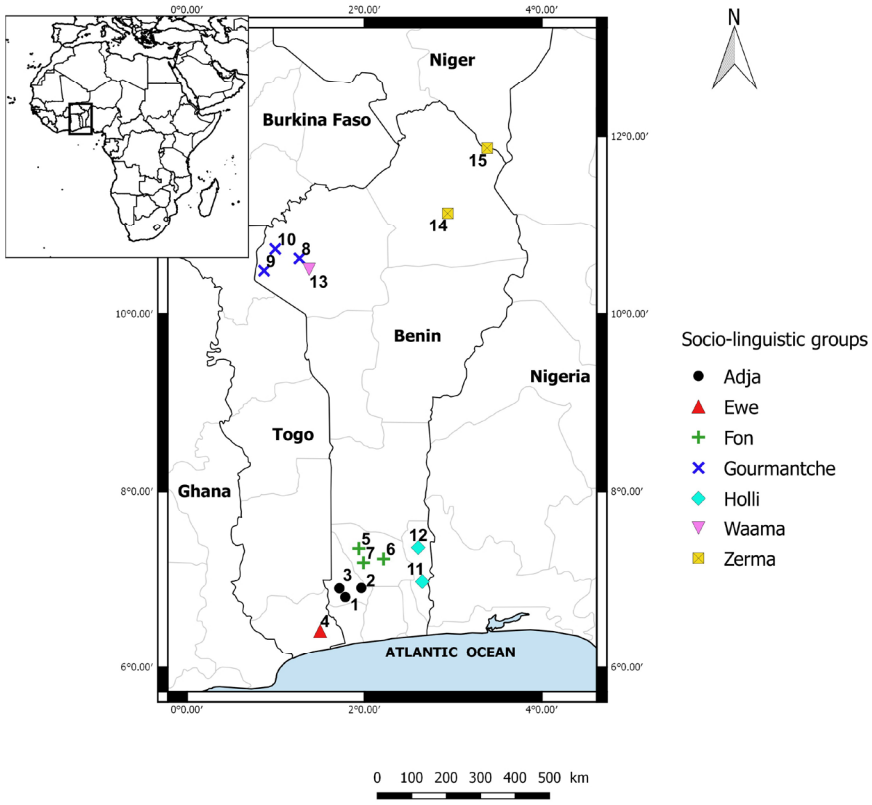


Figure 1. Map of the surveyed districts and socio-linguistic groups interviewed in Benin and Togo.

informants were asked about its market price. A contingent valuation was used when the species was not commercialized. Respondents were asked how much a given amount of *G. gynandra* would be worth in the market or for how much they would be willing to sell it. Local measurement units such as baskets and bunches were weighted to determine the average price of *G. gynandra* leaves per kilogram.

Cultural importance of the species

We modified the cultural significance indices (CSI) (Blancas et al. 2013; Pieroni 2001) to quantify the cultural importance of *G. gynandra*. The CSI was calculated based on eleven variables (**Table 1**) according the following formula:

$$CSI=RU \times S \times AI \times FU \times LC \times PPU \times FP \times TA \times H \times C \times MV \times 10^{-4}$$

The CSI was considered as a quantitative measure of the knowledge and use of *G. gynandra*. Higher CSI values indicated a greater cultural importance of the species.

Table 1. Variables and states considered to compute the cultural significance index.

Variables	Modalities (scores)	Explanation
Number of reported uses (RU)	Absolute number	The number of uses cited for the species by each informant. We included this variable as suggested by Blancas et al. (2013). Reyes-García et al. (2006) previously quantified the cultural value of a species based on the number of uses reported, the proportion of informants who mentioned the species as useful and the proportion of informants who mentioned each use of the species. In our case, the proportions of informants are irrelevant as we computed the values per respondent.
Source of knowledge (S)	Parents (4); Other community members (3); People from different community (2); Mass media (1)	The source of knowledge (S) indicates patterns of transmission of knowledge about the species. We considered the transmission from parents to children and from other members of the same community as the best indicators of the cultural dimension of such knowledge as previously reported in literature (Lozada et al. 2006; Reyes-García et al. 2009). When several sources of knowledge were cited, the average score was considered.
Perceived availability of the species (AI)	Very abundant (5); Abundant (4); Moderately abundant (3); Rare (2); Very rare (1)	The perceived availability is an important factor explaining the cultural importance of edible plants (Cruz et al. 2014; Pieroni 2001) and depends on the demand in the species . Available useful plant species tend to be more widely used in some communities than the rarer ones (de Albuquerque 2006; Hart et al. 2017) and their decreasing availability might trigger the domestication process if the demand cannot be satisfied (Vodouhè and Dansi 2012). In our case, the perceived availability of the species might contribute to explain patterns of knowledge and use.
Frequency of uses (FU)	More than once a week (5); More than once a month but less than once a week (4); Once a month (3); More than once a year but less than once a month (2); Once a year (1); Not used during the past 3 years (0.5)	The frequency of uses (FU) and the last day of consumption (LC) provide an indication on how often the respondent uses the species and how important it is in their daily diet. It is used to illustrate the actual use of the species and as such, complements the reported number of uses which only reflects the knowledge of the respondents (de Lucena et al. 2007).
Last day of consumption (LC)	More than 3 years ago (0.5); More than one year ago (1); Less than 1 year ago (2); less than 6 months ago (3); within the last month (4); within the last week (5)	
Plant parts used (PPU)	Leaves (1); Flower buds (1); Leaves and stems (1); Young pods (1); Seeds (1)	The plant parts used (PPU) (Pieroni 2001) were summed up for the respondents who used multiple organs of the species.
Food preparation (FP)	Processed food (1.5); Single in sauce or stew (1); Condiment (1); Raw (1); Mixed with other vegetables (1)	Cooking and processing practices are part of the cultural identity of local communities (Sansanelli et al. 2017). We attributed a higher score to processing of the species as they indicate a value addition to the plant (e.g. drying and grinding leaves for long term storage).
Taste appreciation (TA)	Very poor (1); Poor (2); Fair (3); Good (4); Very good (5)	The taste appreciation (Pieroni 2001) is in our case is specific to each respondent and was considered as an indicator of how palatable the species is for the respondent.

Harvest type (H)	Opportunistic (1); Dedicated (2)	The harvesting mode of the species was also considered as an indicator of its importance: harvesting exclusively the species would suggest a higher cultural importance while harvesting the species only when it is found while conducting other activities would indicate that the species is not important enough to plan or mobilize a specific time for its harvest (Blancas et al. 2013).
Commercialization possibilities (C)	None (0.5); existing but not exploited (1); leaves sold/bought on local markets (2); leaves and seeds sold/bought on local markets (3)	We included commercialization possibilities (González-Insuasti et al. 2008) to investigate whether or not some respondents considered selling the species or if it is already bought and sold in the surveyed areas in response to a high market demand. The modalities were modified to take into account the various situations described by the respondents.
Medicinal Value (MV)	Not medicinal (1); Healthy food without therapeutic specification (3); Medicinal (5)	The medicinal value was used to take into account the gradients of perceived "health-promoting" properties of the species.

Table 2. Variables used to estimate the management intensity index.

Variables	Categories (scores)	Explanation
Maintenance Labour (ML)	Cleaning or weeding (1); Grooves water penetration (1); Soil tillage (1); Removing dead or diseased plants (1); Fertilization (1); Irrigation (1); Fumigation with pesticides (1); Fences (1); Other dissuasive measures (1)	Maintenance practices on managed plots as suggested by Blancas et al. (2013). Only the modalities cited by our respondents are listed here.
Artificial selection (AS)	Odour (1), Leaf form (1), Leaf colour (1), Stem colour (1), Leaf size (1), Phenological differences (1), Texture (1)	Criteria mentioned for selection of particular variants on managed plots for leaf/seed harvesting and propagation.
Reaction to harvest without permission (RH)	No reaction (0); Yes, admonition applies (1); Yes, applies monetary penalty (2); Yes, complaints to authorities (3)	Indicates how important the managed plots are for the respondents. Adapted from the variable "Collective regulation" used by Blancas et al. (2013) to assess the rules and agreements for accessing and protecting plant resources. The modalities used in our case are the ones cited by the respondents.
Proximity to collection/cultivation sites (PC)	More than 5 km (0.5); Up to 5 km (1); Up to 1 km (1.5); Less than 100 m (2)	The distance from the residence area of the respondent to collection/cultivation site provides an indication of the importance of the resource and was suggested by Blancas et al. (2013) as an indication of management intensity. In our case, we gave a higher value to closer sites as most respondents explained that spider plant is a seasonal vegetable which requires intensive management over a short period of time and can also become invasive if grown in the fields.
Management Forms (MF)	Simple collection from the wild with no particular selection (1); Tolerance, protection or promotion of wild plants without selection (2); Transplantation of selected wild individuals to gardens (3); Rain-fed cultivation (4); Intensive irrigated cultivation all year-round (5)	Adapted from González-Insuasti et al. (2008) and Blancas et al. (2013) to reflect the different situations observed in the case of spider plant in the surveyed areas.

Current management practices

The management practices were described and categorized based on a management intensity index modified from Blancas et al. (2013) and González-Insuasti et al. (2011). Indicators used to compute the management index are presented in **Table 2**. When several modalities of a variable were adopted by the same informant, the scores were summed. Individual management indices were summed to calculate the Management Index:

$$MI = ML+AS+RH+PC+MF$$

Farmers' intention to improve current management forms

In order to assess farmers' intention to improve current management practices through higher investment of energy, intensive artificial selection and land allocation (Blancas et al. 2013; González-Insuasti et al. 2008), the interview included the following questions: are you satisfied with your current level of exploitation of *G. gynandra*? If not, what would you like to improve and how? What are the perceived constraints related to the implementation of your decisions? These questions allowed better understanding of the motivation underlying farmers' intentions to modify their management schemes in the future. In order to avoid the willingness to please biasing farmers' responses, it was well explained to them that we would not provide a reward in kind or in cash in response to their motivation to cultivate the species. Social desirability bias may lead to a tendency to provide positive answers to environmental questions, especially in studies related to behavioural intention (Floress et al. 2018). Information provided by farmers was used to compute an expected management scheme based on management forms' typology adapted from González-Insuasti et al. (2008) based on our observations during the exploratory survey. Five steps were identified including: simple collection with no selection (1); tolerance, protection or promotion without selection (2); transplantation of selected spontaneous individuals (3); rainfed cultivation (4); intensive irrigated cultivation all-year round (5).

Data analysis

The seven socio-linguistic groups were compared for their number of reported uses and the frequencies of uses of the species based on Kruskal-Wallis tests. A regression tree analysis was performed using the R packages "tree" (Ripley 2016) and "maptree" (White and Gramacy 2012) to predict the variation in Cultural Significance Index based on both quantitative and qualitative socio-demographic variables. The most significant variables among the dependent variables gender, age, education level, ethnicity, land tenure, land area, and level of income were selected. A generalized linear model with Poisson distribution allowed us to investigate the socio-demographic factors influencing the management intensity in each community. The deviance explained by the model (D^2) was determined using the R package "modEvA" (Barbosa et al. 2016). As ethnicity was the main factor explaining the variation in the CSI and the management intensity, a principal component analysis was performed using the components of the CSI, the management intensity and the economic value of the species to better describe the characteristics of the socio-linguistic groups. The Kendall's rank correlation (τ) was used to test the association between current and expected management regimes. To model the projected changes in management practices according to time (current and expected management forms)

and socio-linguistic groups, a generalized estimating equation was fitted to the management forms [simple collection with no selection (1); tolerance, protection or promotion without selection (2); transplantation of selected spontaneous individuals (3); rainfed cultivation (4); intensive irrigated cultivation all-year round (5)]. Generalized estimating equations are commonly used in epidemiology and social sciences for multivariate logistic regression of ordinal longitudinal outcome variables (Nooraee et al. 2014; Parsons et al. 2009).

Results

Characteristics of respondents

Women represented 72.0% of the respondents and were dominant across all socio-linguistic groups except for the Ewe, of which 73.0% of respondents were male (**Table 3**). *G. gynandra* users were between 14 and 90 years old, with an average age of 44.9 ± 15.7 years. A high proportion of the Beninese respondents were illiterate (73.0%), while for the Ewe ethnic group in Togo, this was only 33.0%. Literate male producers dominate the *G. gynandra* production system in the Ewe community, while female illiterate producers were more involved in wild harvesting and home gardening. The main occupation of all respondents was farming and/or trading agricultural products. Landowners represented 66.0% of the respondents, except in the Ewe group, where most of the market gardeners leased their land (62.0%).

Knowledge and diversity of uses of *G. gynandra*

Gynandropsis gynandra was used by all socio-linguistic groups as food (100.0% of respondents) and medicine (50.7% of respondents). Spiritual uses of the species were mentioned by less than 5.0% of respondents in all groups, except for the Gourmantche and Waama communities who did not mention these uses at all. The number of reported uses varied significantly among the groups (Kruskal-Wallis $\chi^2 = 124.6$, $df = 6$, $p < 0.001$). The highest number of uses was reported by Zerma respondents (3.2 ± 1.5). The Ewe (1.7±0.8), Gourmantche (1.5±0.8), Holli (1.4±0.6) and Waama (1.3±0.7) stated the lowest number of uses.

In addition to food, *G. gynandra* was used to cure 42 different diseases, classified in 12 disease categories according to the International Classification of Diseases and related health problems (WHO 2016) (**Appendix 2**). General symptoms and signs not elsewhere classified were grouped in "general health" category. Most of the Adja (81.0%) and Zerma (81.0%) respondents were aware of the medicinal uses of the species followed by the Fon (56.0%). Only 9.0% of the Waama reported medicinal uses for the species. The most cited diseases treated with *G. gynandra* included malaria, headache, earache, anaemia, stomach-ache and malnutrition. The citation frequency of medicinal uses per disease category and socio-linguistic group are presented in **Figure 2**.

The categories of diseases cured, and their relative importance varied across the communities. Most respondents (88.0%) obtained their knowledge on *G. gynandra* uses from their parents, 7.0% from community members other than their parents and 5.0% from people from other communities. Among the Fon respondents, 61.0% received their knowledge from their parents, 28.0% from community members other than their parents and 10.0% from people from other communities.

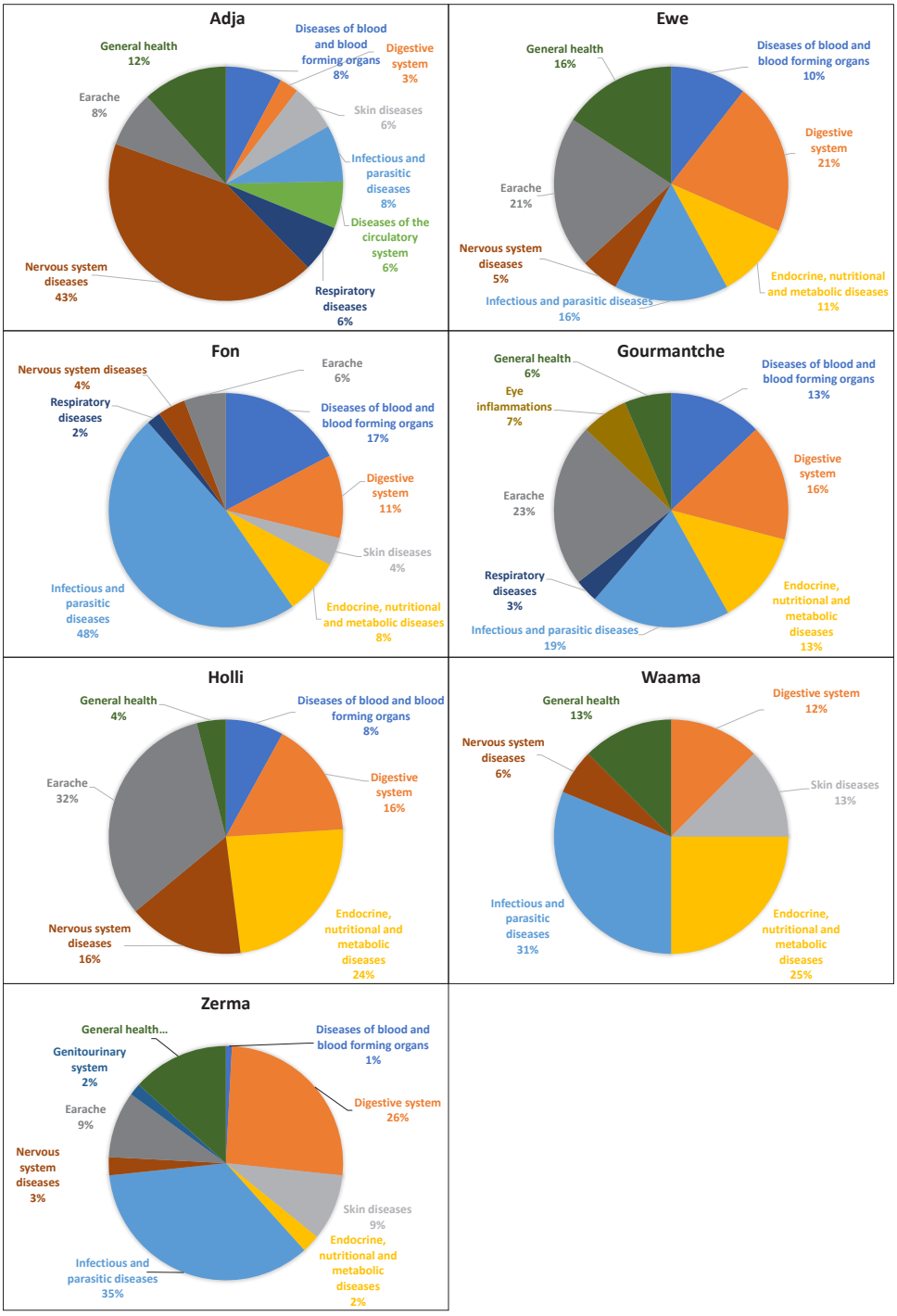


Figure 2. Frequency of citations of categories of diseases treated with *Gynandropsis gynandra* in each socio-linguistic group.

Table 3. Socio-demographical characteristics of respondents in seven socio-linguistic communities of Benin and Togo.

Variables	Modalities	Adja (N = 62) n (%)	Ewe (N = 60) n (%)	Fon (N = 57) n (%)	Gourmantche (N = 61) n (%)	Holli (N = 60) n (%)	Waama (N = 60) n (%)	Zerma (N = 68) n (%)	Total (N = 428) n (%)
Age (years)	Mean ± sd	44.5 ± 12.4	41.4 ± 15.0	48.0 ± 16.2	45.4 ± 18.7	43.8 ± 14.8	43.9 ± 13.8	47.2 ± 17.5	44.9 ± 15.7
Gender	Male	10 (16)	44 (73)	11 (19)	11 (18)	19 (32)	14 (23)	11 (16)	119 (28)
	Female	52 (84)	16 (27)	46 (81)	50 (82)	41 (68)	46 (77)	57 (84)	309 (72)
Instruction level	Illiterate	52 (83.8)	20 (33.3)	47 (82.4)	41 (67.2)	51 (85)	44 (73.4)	56 (82.3)	311 (73.4)
	Primary	-	23 (38.4)	6 (10.5)	5 (8.2)	6 (10)	8 (13.3)	9 (13.2)	57 (13.6)
	Secondary	7 (11.3)	15 (25)	3 (5.3)	15 (25.6)	3 (5)	8 (13.3)	3 (4.5)	54 (13.3)
	University	-	2 (3.3)	1 (1.8)	-	-	-	-	3 (1)
	Literate (local language)	3 (4.9)	-	-	-	-	-	-	3 (1)
Land ownership	No land	3 (4.8)	1 (1.7)	6 (10.5)	13 (21.4)	6 (10.2)	1 (1.7)	13 (19.1)	43 (10.1)
	Purchase	9 (14.5)	-	8 (14)	1 (1.6)	2 (3.4)	-	3 (4.4)	23 (5.4)
	Lease	11 (17.7)	37 (61.6)	1 (1.8)	-	14 (23.7)	-	6 (8.8)	69 (16.2)
	Loan	2 (3.2)	1 (1.7)	4 (7)	1 (1.6)	1 (1.7)	-	7 (10.3)	16 (3.7)
	Gift	2 (3.2)	1 (1.7)	5 (8.8)	-	0 (0)	2 (3.3)	5 (7.4)	15 (3.5)
	Inheritance	35 (56.5)	20 (33.3)	33 (57.9)	46 (75.4)	37 (61)	57 (95)	34 (50)	262 (61.1)

Factors influencing cultural importance of *G. gynandra*

The regression tree of the cultural significance index based on the respondents' socio-economic characteristics explained 37.1% of deviance in the cultural significance index and revealed that ethnicity was the main factor explaining the cultural importance attributed to the species (19.3% of deviance) (**Figure 3**). The splits based on other variables resulted in much lower deviance (<4%) and therefore poorly contribute to explaining the variation in the index. The socio-linguistic groups were split into two categories: the Adja and Ewe who gave a higher cultural importance to the species were in the first category and other socio-linguistic groups in the second.

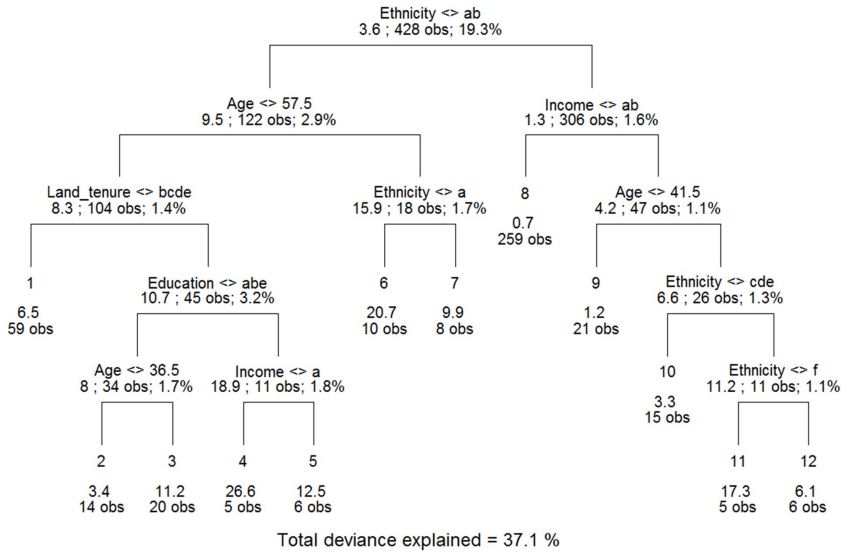


Figure 3. Regression tree of cultural significance index as a function of demographic variables for use of *Gynandropsis gynandra*. Terminal nodes display from top to bottom: terminal node number, average CSI for the subsample and the number of observations in the subsample. Categorical predictors include: Ethnicity (a=Adja; b= Ewe; c=Fon; d=Gourmantche; e= Holli; e= Waama; f= Zerma); Income (a= < 50 USD; b= 50 -170 USD; c=170-330 USD; d= > 330 USD; 1 USD=590 F CFA on average, July 2016); land tenure (a=no access; b= purchase; c= lease; d = loan; e=gift; f=heritage); education (a=illiterate; b=primary school; c=secondary school; d= university; e=literate (local language).

Among the Adja and Ewe, the species had a higher cultural importance for older respondents (> 57 years) than for younger respondents (< 57 years). Moreover, the older Ewe respondents had a CSI twice lower than their Adja counterparts. Among younger Adja and Ewe respondents, land tenure, education and income were also important factors of differentiation (**Figure 3**). Respondents who inherited family land or had no access to land gave a higher cultural importance to *G. gynandra* than those who bought or leased the land (mostly Ewe market gardeners).

Non-schooled respondents and those who only attended primary school had also a higher CSI than respondents with higher education levels. In this latter category, respondents with low monthly income (less than 50 USD) had a higher CSI than those with higher monthly income.

Among the other socio-linguistic groups, most of the respondents (85%) had low monthly income (less than 140 USD) and fell into one category with a low CSI. Among respondents with higher income (only 15%), older respondents (> 41 years) had a higher CSI than younger ones. The older Fon, Gourmantche and Holli respondents had a lower CSI than their Waama and Zerma counterparts.

Variation in the management intensity, cultural and economic importance of *G. gynandra*

Four management regimes of *G. gynandra* were observed during the survey: (1) wild collection with no selection, (2) tolerance, protection or promotion of plants around houses; (3) rainfed cultivation in home gardens and (4) intensive irrigated cultivation all-year round. The significant factors explaining the variation in management intensity were ethnicity (D^2=65.9%, $p<0.001$) and land tenure (D^2=17.6%, $p<0.05$).

The species was intensively cultivated by Ewe market gardeners in Southern Togo (96.0% of Ewe respondents; management intensity index (MI): 14.9 ± 1.4) (**Figure 4**), where the price of the leaves varied between 0.25 USD per kilogram in the rainy season and 1.1 USD /kg in the dry season (1 USD=590 F CFA on average, July 2016). Seeds were also sold on local markets; prices varied from 5 to 16.7 USD/kg, depending on the season.

Figure 4. Ewe woman harvesting spider plant leaves (*Gynandropsis gynandra*) in a market garden in Vogan County (Togo). ©Photo credit: Deedi Sogbohossou

Rainfed cultivation in home gardens was common among the Adja (93.0% of respondents; MI: 11.0 ± 2.7). Leaves were sold in bunches with a price that varied between 0.2 and 0.5 USD/kg. Seeds were occasionally sold but most of the farmers spared some of their own plants to harvest the seeds. The Gourmantche (90.2%; MI: 4.2±1.3) and the Fon (80.7%; MI: 4.1±3.1) mainly harvested the plant in the wild at the beginning of the rainy season along roadsides or in fallow lands.

Among Waama people, simple collection (46.7%) and tolerance or protection of spontaneous populations (48.3%) were the most common management practices (MI: 5.7±2.5). Simple collection in the wild (63.3%) and rainfed cultivation (23.3%) were the dominant management practices among the Holli (MI: 4.8 ±2.9). Our Zerma respondents pointed out that as they recently migrated from Niger to Northern Benin, they had limited access to agricultural land. They said that *G. gynandra* was not popular among the other socio-linguistic groups living near them in Northern Benin. Therefore, 54.0% of our Zerma informants harvested the plant in the wild and only 20.0% cultivated it in home gardens (MI: 5.0 ±2.7).

Land tenure had a significant influence on the MI (D^2=17.6%, $p<0.05$) with the highest average value observed for lease (11.3 ± 4.8) mainly adopted by Ewe (61.6%) respondents. Other land tenure regimes including purchase (7.0 ±4.9), inheritance (6.6 ± 4.1), loan (6.1 ± 3.5), gift (4.6 ± 3.2) were associated with lower MI scores. People with no land had the lowest MI (4.6 ± 3.0).

A significant variation was noticed in the perceived availability of the species across socio-linguistic groups (Kruskal-Wallis $\chi^2 = 194.3$, df = 6, $p < 0.001$). While in Fon, Adja, Ewe, Gourmantche and Waama communities, *G. gynandra* was regarded as a common species, the Holli and Zerma noticed that it was becoming rare in their region. The reasons cited to explain this rarity included: (1) increasing urbanisation, intensive crop production and use of herbicides, which reduced the natural habitat of the species (34.0% of respondents); (2) lack of knowledge about the reproductive biology (33.0%); (3) erosion of knowledge (10.0%); (4) low economic value (8.5%); (5) low germination ability of the seeds (8.0%) and (6) grazing by goats, pigs and other animals (7.0%). No selection of phenotypes was reported for the species.

The principal component analysis revealed that the variables related to the use of the species, the economic importance, and the management intensity were correlated with the first principal component (33.2% of variation explained) while the number of reported uses and the medicinal value of the species were correlated with the second principal component (12.9% of variation explained) (**Table 4**). The frequency of uses and the last day of consumption were also significantly correlated with the third principal component, which explained 10.0% of the variation.

The projection of the respondents on dimensions 1, 2 and 3 (**Figure 5)** supported the differences in knowledge, uses and management of *G. gynandra* among socio-linguistic groups. The species had both high cultural and economic importance for the Adja and Ewe, which presented positive coordinates on the first dimension. The difference between both groups lies in the fact that the Adja also had a strong knowledge on the uses of the species (positive coordinates on dimension 2), while for the Ewe the species was mostly a vegetable. For all other groups, the species was both culturally and economically less important, apart from the Zerma, who showed a significant knowledge on the vegetable that they brought with them during their migration from Niger to Benin.

Table 4. Scores of 14 variables related to cultural importance, management intensity and economic importance on 3 first principal components for use of *Gynandropsis gynandra.*

	Dim1	Dim2	Dim3
Cultural importance indicators			
Commercialization possibilities	0.85	-0.02	-0.17
Harvest type	0.68	0.04	-0.11
Frequency of uses	0.61	0.01	0.64
Last day of consumption	0.61	0.03	0.58
Perceived availability	0.60	-0.18	0.20
Plant parts used	0.37	0.03	-0.36
Taste appreciation	0.37	0.23	0.28
Source of knowledge	0.20	0.09	0.09
Medicinal value	-0.01	0.90	-0.07
Number of reported uses	-0.01	0.93	-0.06
Food Preparation	-0.18	0.18	0.44
Management index	0.85	0.05	-0.24
Economic value of leaves	0.83	-0.04	-0.27
Economic value of seeds	0.78	0.01	-0.21
% of variance explained	33.16	12.96	10.04
Cumulative % of variance	33.16	46.12	56.16

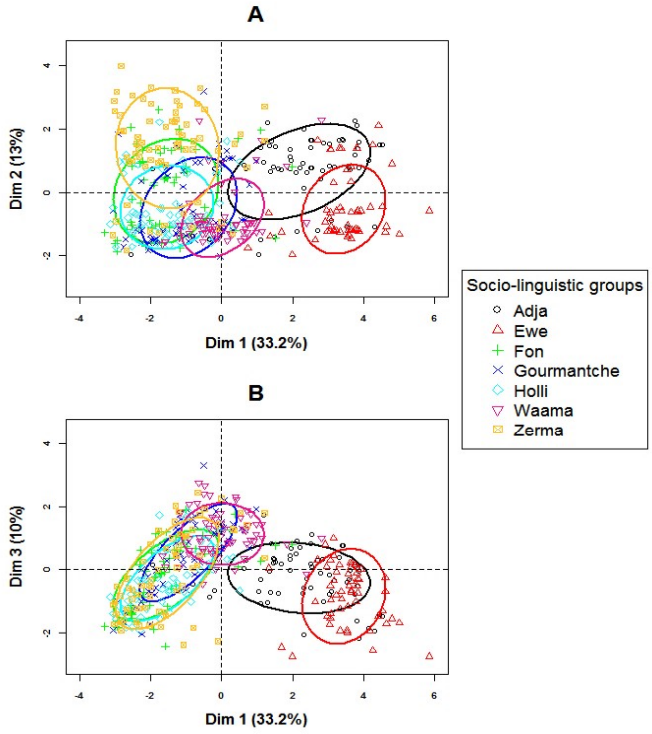


Figure 5. PCA scores plot of 428 respondents from the different socio-linguistic groups based on knowledge and use and economic importance of *Gynandropsis gynandra.* (A) Projection on principal components 1 and 2. (B) Projection on principal components 1 and 3.

Willingness to improve current management practices

A significant and positive correlation was found between current and projected management regimes of *G. gynandra* (τ=0.52, $p<0.001$). Among the Adja, who cultivated the species in home gardens, 73.0% were willing to shift from rainfed cultivation to intensive year-round cultivation. The Ewe would not significantly change their practices over time as they already intensively cultivated the species. In all the other socio-linguistic groups, rainfed cultivation was considered as an interesting option for the future. The variation in management forms across socio-linguistic groups is depicted in **Figure 6.**

The generalized estimating equation computed on the management forms used as repeated measures revealed a significant variation across socio-linguistic groups and across time as well as

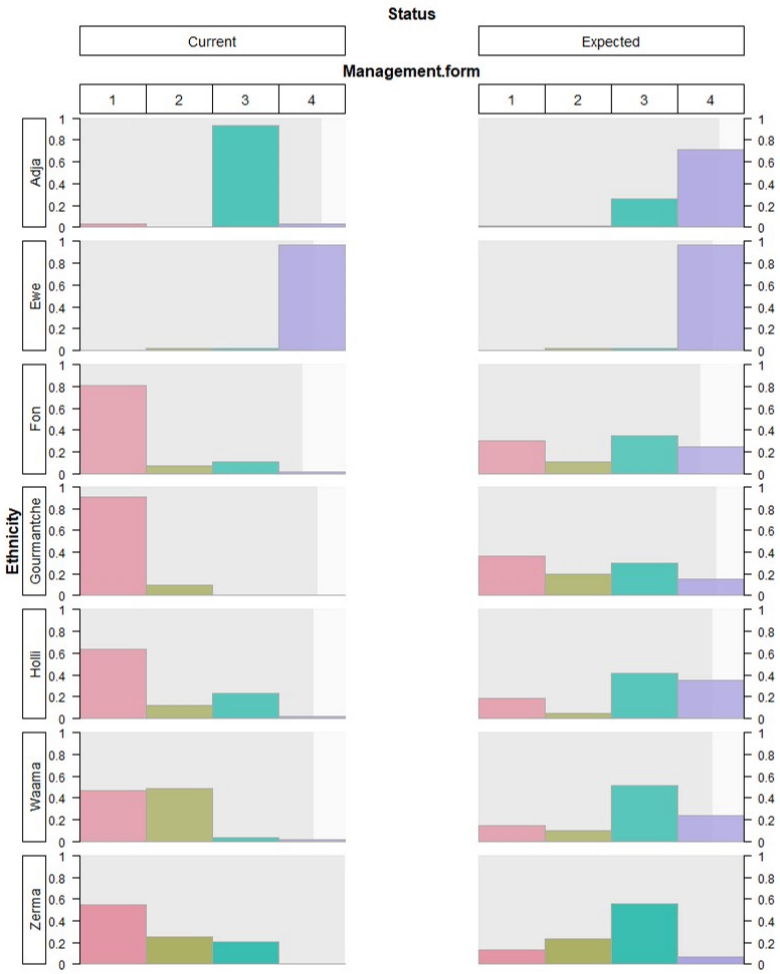


Figure 6. Frequencies for current and expected management forms in each socio-linguistic group for use of *Gynandropsis gynandra*. Management forms: (1) wild collection with no particular selection, (2) tolerance, protection or promotion of plants around houses; (3) rainfed cultivation in home gardens and (4) intensive irrigated cultivation all-year round.

for the [time x socio-linguistic groups] interaction (**Table 5** and **Figure 7**).

Table 5. Association between current and future management schemes across socio-linguistic groups (CI: confidence interval).

	Estimates	Standard error	Odd ratios	95% CI	Z-value	P >|Z|
β1	-2.17	0.16	0.11	0.08 - 0.15	-13.85	< 2e-16***
β2	-1.20	0.12	0.30	0.24 - 0.38	-9.98	< 2e-16***
β3	1.19	0.11	3.29	2.66 - 4.06	11.08	< 2e-16***
Current (reference)						
Future	-1.87	0.23	0.15	0.10 - 0.24	-8.11	< 2e-16***
Adja (reference)						
Ewe	-4.60	0.74	0.01	0.01 - 0.04	-6.23	< 2e-16***
Fon	3.73	0.41	41.53	18.46 - 93.44	9.01	< 2e-16***
Gourmantche	4.65	0.52	104.99	37.52 - 293.78	8.86	< 2e-16***
Holli	2.65	0.30	14.18	7.85 - 25.60	8.79	< 2e-16***
Waama	2.44	0.23	11.49	7.33 - 18.03	10.63	< 2e-16***
Zerma	2.40	0.26	10.97	6.63 -18.15	9.33	< 2e-16***
Future: Adja (reference)						
Future: Ewe	1.87	0.23	6.51	4.14 - 10.23	8.11	< 2e-16***
Future: Fon	-1.18	0.46	0.31	0.13 - 0.76	-2.55	0.013*
Future: Gourmantche	-1.44	0.62	0.24	0.07 - 0.79	-2.33	0.019*
Future: Gourmantche	-1.44	0.62	0.24	0.07 - 0.79	-2.33	0.019*
Future: Holli	-1.05	0.39	0.35	0.16 - 0.75	-2.69	0.007**
Future: Waama	-0.54	0.36	0.58	0.29 - 1.18	-1.49	0.136ns

***: p<0.001 ; **: p<0.01; * :p<0.05; ns: non-significant

The significant negative regression coefficient (Z-value=-8.1, p<0.001) for the projected management practices suggested that the level of these practices would overall increase compared with the current management practices. The Adja were used as a reference for comparison with other socio-linguistic groups. The interaction between time and socio-linguistic groups suggests that the Ewe (Z-value=8.1, p<0.001) would show significantly less change in management practices over time than the Adja. The negative interaction regression coefficients for the Fon, Gourmantche and Holli suggested significantly higher rate of change in these groups compared to the Adja respondents. The non-significant interaction for the Zerma (Z-value=0.6, p=0.537) and Waama (Z-value=-1.5, p=0.136) indicated that the management practices in these groups would not significantly change in the future compared to the Adja. However, the model suggested a higher change rate for the Waama than the Adja and a lower change rate for the Zerma than the Adja.

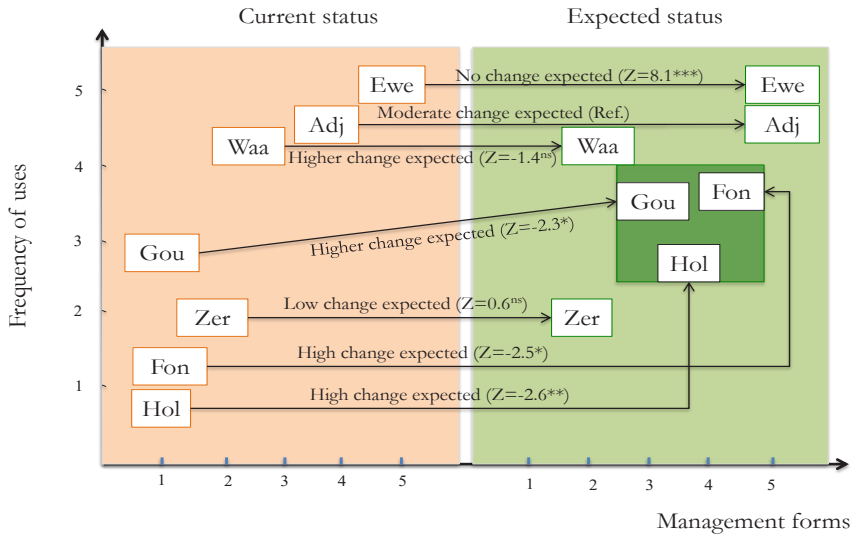


Figure 7. Current and expected future management intensity (τ = 0.52, $p < 0.001$) of *G. gynandra* **by socio-linguistic groups as main drivers (D^2 = 65.9%) in Benin and Togo.** Ewe = Ewe community, Adj: Adja community, Waa: Waama community, Gou: Gourmantche community, Zer: Zerma community, Fon: Fon community, Hol: Holly community. The greener area is where the leverage activities should be intensified.

Discussion

Factors influencing the use and management of *G. gynandra*

We hypothesized that the variation in use and management of the species would be influenced by ethnicity, cultural and economic importance as well as the presence of market-oriented vegetable production. Our results indeed demonstrated strong ethnicity-driven patterns of knowledge, use, and valuation of spider plant. Ethnicity was also raised as an important factor in previous studies investigating variation in indigenous plant species knowledge (Ekué et al. 2010; Kidane et al. 2015). We observed that the cultural and economic importance varied across socio-linguistic groups and could be also explained by ethnicity.

The economic importance of the species was high among Adja and Ewe socio-linguistic groups. Both socio-linguistic groups are closely related and sometimes considered as the same ethnic community (Landoh et al. 2016) which may explain these similarities. While in the Ewe community, the species was only found among market gardeners who cultivate a wide diversity of vegetables, in the case of the Adja, the species was cultivated in rainfed home gardens in pure stands and sold on local markets. We can therefore speculate that even in the absence of market-oriented vegetable production, local communities would domesticate the vegetable as soon as it becomes a popular and valued species. The differences noticed between both socio-linguistic groups in terms of cultural importance of the species suggested that urbanization could have resulted in less attention for the medicinal and cultural value of the species. Urbanization may jeopardize, increase or have no effect on traditional knowledge transmission and conservation

(Gaoue et al. 2017; Júnior et al. 2016). In our case, further comparison of traditional plant knowledge among urban and rural Ewe farmers would be required to test how urbanization affected their knowledge of the species.

Even though spider plant was perceived as abundant among the Fon, Gourmantche and Waama groups – with strong awareness of usefulness – these communities did not consider the species as economically important and would only harvest it when needed for consumption or medicinal uses. The differences observed between knowledge and use among the different communities could be explained by discrepancies between their theoretical ethnobotanical knowledge and their practical uses of the species (Godoy et al. 2009; Reyes-García et al. 2006). Albuquerque (2006) distinguished the notions of "mass knowledge" and "stock knowledge" which refer respectively to the total number of useful plants known by a community and the species actually used by them. In other words, some communities have a good knowledge of the species but would only use it as a last resort and prefer using other species with similar functions (e.g. food, medicine). This situation might change if there is an increasing market demand for the species.

Other communities have a high preference for the species which, in the case of the Ewe and Adja led to its cultivation and commercialization probably due to an increasing demand in the species. As previously reported by N'Danikou et al. (2011), the market importance of wild edible plants is also a main incentive for species' use and conservation. For example, our Zerma respondents, who originate from Niger, where the species is used and cultivated, cited the highest number of uses compared with any other socio-linguistic group. However, because of the low popularity and the lack of marketing opportunities for the species in Northern Benin, they used the species less frequently than the others and were less interested in its cultivation. Similar trends were observed for *Artemisia absinthium,* a plant species cultivated in Haiti for food, medicinal and spiritual uses and with a high economic value. The species lost popularity over time among migrant Haitians in Cuba and its cultivation in home gardens declined substantially. Moreover the species which was the main ingredient of the *Tifey*, a local beverage used for multiple purposes, was progressively replaced by other species more abundant in the host country (Volpato et al. 2009).

In the case of the Holli, spider plant was overall considered as rare and this community subsequently also had a low number of reported uses as well as frequency of uses and overall cultural importance. The link between species availability, knowledge and actual use is therefore not straightforward but rather context specific. In our case, the availability of the species was evaluated by the respondents based on how they perceived the balance between their demand and the abundance of the species in their community. To better account for the complexity of the concept of species availability, Gaoue et al. (2017) suggested a multi-dimensional index incorporating ecological, socio-cultural, economic, and political aspects and drivers of availability.

Patterns of plant species uses and management often arise from complementarity between socio-cultural attributes and biogeographical factors. Communities sharing the same environment but different cultural background may have convergent plant uses (Segnon and Achigan-Dako 2014; van Andel et al. 2015). We observed patterns of convergence in use and management between communities sharing a same cultural background and belonging to the same geographical area. For the Gourmantche and the Waama who belong to the Gur language family (CENALA 2003)

and live in the Sudanian semi-arid region, the species has a lower cultural importance and is less intensively managed compared with the Adja and Ewe who belong to the Kwa language family and live in the Guineo-Congolian humid region. Although our data supports the hypothesis that management and uses are shaped by ethnicity and geographical proximity, an exception is that among the Fon who also belong to the Kwa language family and are geographically close to the Adja and Ewe (CENALA 2003), the species is much less popular as a vegetable, but more as a medicinal plant, and thus managed less intensively. The name of the species is also the same among Ewe and Adja suggesting strong cultural affinities, and different for the Fon who might be more culturally distant from those socio-linguistic groups. Wartena (2006) explained that the Fon and the Adja shared their climate, geological conditions, trade opportunities, and cultural origins of their populations. However, divergences between the two socio-linguistic groups was traced back to pre-colonial times, first in trade- and culture contacts with neighbours and then in the socioeconomic and cultural practices of the slave raiding and trading Fon on the one hand and of the agrarian Adja hiding from slave raiders on the other. The author also described in details current differences in the traditional beliefs, agricultural practices and food preferences between these communities.

A commonality between all the surveyed socio-linguistic groups is the knowledge about the medicinal uses of the species. The use of the species to cure malaria and anaemia respectively by the Fon and the Mahi were previously reported in southern Benin (Allabi et al. 2011; Yetein et al. 2013). Medicinal uses of the species were also reported in other parts of the world including India (Bala et al. 2010; Shanmugam et al. 2012), Uganda (Oryema et al. 2010) and Kenya (Jeruto et al. 2008). The use of the species as a medicinal plant across several socio-linguistic groups in Benin suggests its proven efficacy. Its occurrence in anthropogenic areas as an herb which grows at the onset of the rainy season makes it easily available when needed. Such characteristics are in accordance with the "availability" and the "ecological apparency" hypotheses commonly used to explain the predominance of weeds in medicinal flora. The availability hypothesis states that accessible or locally abundant plants are more likely to be used as medicinal plants. The ecological apparency hypothesis speculates that species with short lifespan (non-apparent species) develop "inexpensive" qualitative defence compounds against herbivores (secondary metabolites) (Albuquerque 2006; Gaoue et al. 2017). Those compounds also have beneficial health-promoting properties. Spider plant leaves are indeed rich in glucosinolates and flavonoids which are involved in response against environmental stresses (Omondi et al. 2017b) but also have pharmacological properties (Bala et al. 2014; Moyo et al. 2013).

During our sampling, we noticed that the management of the species was gender-specific. Most of our respondents were female. However, in the Ewe community, with a high economic orientation to the cultivation of spider plant, producers were predominantly males. In the other socio-linguistic groups, women managed the species for household consumption in rainfed cultivation in home gardens and sold it occasionally on local markets. In Benin, it was reported that women are typically involved in the domestication of leafy vegetables and medicinal plants while fruit domestication is controlled by men (Vodouhè and Dansi 2012). The predominance of males in commercial production is explained by the fact that commercial vegetable production requires important financial resources for land lease, irrigation and purchase of various inputs. For male Ewe respondents, vegetable production was a full-time job. Only few women were

encountered during our survey in Togo and they were much more involved as hired labourers and vegetable traders. Several studies conducted in urban areas of West Africa reported the male dominance in vegetable production, while women lead vegetable marketing (Drechsel et al. 2013; Gockowski et al. 2003). However, peri-urban and rural vegetable production is often dominated by women who mainly grow African leafy vegetables as these require lower amounts of external inputs and financial resources than non-indigenous vegetables (Dinssa et al. 2016). Adja women who lived in peri-urban and rural areas were considered by their male counterparts as more knowledgeable about the uses and growth of the species. The same behaviour was noticed for all the other socio-linguistic groups in Benin, where male farmers considered that "only women manage the species for culinary purposes". These trends suggest an increasing interest from men when the species becomes economically important. Research for development initiatives aiming at promoting African leafy vegetables should therefore put an emphasis on ensuring women's access to key agricultural production factors including land, labour and financial asset.

In the Adja community, we observed a strong willingness to shift towards intensive year-round cultivation of the species, which would require higher investment of resources from women. The commercial cultivation of the species by the Ewe in Togo is a unique opportunity to identify and address the main constraints related to agronomy of the species. Rainfed cultivation of the species in home gardens was considered as an attractive option among the Fon, Holli, Gourmantche and Waama. These groups have access to land but external factors including farmers' training on adequate cultivation practices and leverage activities should also be taken into account to support their efforts. Strategies for integration of wild edible plants in the market should take into account the issues of production, storage and processing, organization of wild edible plants supply chains and negative perceptions of their consumption, often associated with poverty and low social status (Leal et al. 2018). The promotion of *Moringa oleifera* in several African countries is an example of successful diffusion of innovation consistent with existing social and cultural practices (Thurber and Fahey 2009). Intensive cultivation of the species year-round in the surveyed communities would require interventions to create public awareness about the nutritional and health benefits of the species and other wild leafy vegetables. Potential consumers including women, youths, and elite class citizens with high income in urban and peri-urban areas should be targeted in order to increase the demand in the vegetable. Participatory breeding efforts to provide farmers with high-yielding and nutritious varieties as well as investigation on optimal seed storage and conservation are also necessary to sustain the species cultivation.

Limitations of our study

The purposive sampling method in ethnobotany has been often criticized as the results cannot be generalized to the whole community (Tongco 2007). In our study, this method was however the most convenient as we needed to identify people who knew the species and were willing to participate in the study. Another unavoidable bias in our study was the gender skewness of our sample. In most households visited, men directed us to women who are in charge of collecting or cultivating vegetables for cooking and commercialization. Because of the subjective character of some of the variables collected, repeated interviews in time would allow better triangulation

of our data. However, this approach would be time-consuming and costly.

The Cultural Significance Index and the Management Index were both adapted to the context of our study. The choice of the Cultural Significance Index was guided by the need to analyse the perceptions of our respondents. Other cultural importance indices developed in literature were used to compare species and were limited to comparisons between the number of respondents who mentioned some uses of a species in defined use-categories (Tardío and Pardo-de-Santayana 2008). However, we do not pretend to have exhaustively taken into account all the aspects to consider when evaluating the cultural significance or the management intensity of a plant species. The selection of different variables and codifications might lead to different results. For example, determining the relative cultural importance of spider plant compared with other leafy vegetables in the surveyed communities would have allowed to better assess the species availability, use and management. Such information coupled with a discrete choice experiment (Espinosa-Goded et al. 2010; Schulz et al. 2014) and an analysis of market opportunities could also improve our evaluation of respondents' willingness to integrate the species in their home gardens.

Conclusion

Our study provides insight in the knowledge, use and management of *G. gynandra* in seven socio-linguistic groups in Benin and Togo. Communities had a good knowledge of the nutritional and medicinal value of the species. However, cultural importance of the species and management practices across these communities were diverse and strongly influenced by ethnicity and gender, and to a lesser extent by age, education and access to land. We observed overall a convergence of knowledge and management practices between communities with similar socio-cultural and ecological contexts. Migration and available market opportunities also significantly influenced the management practices adopted by our respondents. In areas where *G. gynandra* is still harvested from the wild, promotion of the species among consumers and farmers' training, especially women, are expected to trigger the species' cultivation. Efficient selection of neglected plant species to promote in a given region for successful commercial cultivation should take into account not only their nutritional value and pharmacological properties but also communities' willingness to intensify management of these species with a particular emphasis on potential socio-cultural and economic constraints which could arise during such an intensification process.

Acknowledgements

This study was financially supported by the WOTRO Applied Research Fund (Project Number W.08.270.350). We would like to thank Chaldia Agossou, Jelila Blalogoe and Xavier Matro for their assistance during data collection and Alcade Segnon for his insightful comments. We are grateful to the respondents for kindly sharing their knowledge with us. We also appreciated the support of the local authorities in all the surveyed areas.

Supplemental information

Supplemental files are available at: https://doi.org/10.1007/s12231-018-9423-5

Appendix 1. Questionnaire on uses and management of *Gynandropsis gynandra* in Benin and Togo

Appendix 2. Proportions of respondents per medicinal use of *Gynandropsis gynandra*

CHAPTER 4

Association between vitamin content, plant morphology and geographical origin in a worldwide collection of the orphan crop *Gynandropsis gynandra* (Cleomaceae)

E. O. Dêêdi Sogbohossou[1,2], Dieke Kortekaas[1], Enoch G. Achigan-Dako[2], Patrick Maundu[3], Tsvetelina Stoilova[4], Allen Van Deynze[5], Ric C. H. de Vos [6], M. Eric Schranz[1]

[1] Biosystematics Group, Wageningen University, Postbus 647 6700AP, Wageningen, Netherlands

[2] Laboratory of Genetics, Horticulture and Seed Science, Faculty of Agronomic Sciences, University of Abomey-Calavi, BP 2549 Abomey-Calavi, Republic of Benin

[3] Kenya Resource Center for Indigenous Knowledge (KENRIK), Centre for Biodiversity, National Museums of Kenya, Museum Hill, P.O. Box 40658, Nairobi, 00100, Kenya

[4] World Vegetable Center (AVRDC), P.O. Box 42, Shanhua, Tainan, 74199 Taiwan

[5] Department of Plant Sciences, University of California, Davis, USA, 95616

[6] Bioscience, Wageningen Plant Research, Wageningen UR, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.

Published in *Planta*, 2019, 250: 933. DOI: 0.1007/s00425-019-03142-1

Abstract

We examined the variation in carotenoids, tocopherols, chlorophylls and ascorbic acid as well as morphological traits in a worldwide germplasm of 76 accessions of the orphan leafy vegetable *Gynandropsis gynandra* (Cleomaceae) simultaneously grown under controlled greenhouse conditions. The levels of carotenoids and tocopherols in the leaves varied significantly across the accessions analysed and were linked with their geographical origin and morphological variation. The main carotenoids included lutein, β-carotene, α-carotene and violaxanthin. A two to three-fold variation was observed for these pro-vitamin A compounds. The main tocopherols detected were α-tocopherol and γ-tocopherol, with a twenty-fold variation. A nine-fold variation in vitamin C level, independent of geographical origin, was observed. Overall, accessions were grouped into three clusters based on variation in nutrient content and morphology. West African accessions were short plants with small leaves, with relatively high tocopherol and low carotenoid contents; Asian accessions were also short plants but with broad leaves and with relatively low carotenoid and high tocopherol contents; while East/Southern African plants were tall with high contents of both carotenoids and chlorophylls and low tocopherol contents. Carotenoid levels were positively correlated with plant height as well as foliar and floral traits but negatively correlated with tocopherols. The absence of a significant correlation between vitamin C and other traits indicated that breeding for high carotenoid or tocopherol content may be coupled to both improved leaf yield and vitamin C content. Our study provides baseline information on the natural variation available for traits of interest for breeding for enhanced crop yield and nutrient content in *G. gynandra*.

Key-words: *Gynandropsis gynandra;* leafy vegetable; vitamins; morphology; natural variation; geography.

Introduction

Tropical areas of the world are endowed with a rich diversity of edible plant species with tremendous potential to contribute to food security, human health and poverty alleviation (Cullis and Kunert 2017; Pinela et al. 2017). However, most of those species are niche or orphan crops with a regional or local importance and often overlooked in breeding programs (Sogbohossou et al. 2018a). With the recent impetus for improved food quality, there is an increasing interest in diet diversification and breeding for highly nutritious crop cultivars. A critical step of this process is to understand the natural variation in phytochemical profiles in the crop of interest. Plant species with broad geographic distributions experience a wide range of biotic and abiotic conditions which thus can drive the evolution of variation in morphology, phytochemical profiles and interactions with other organisms (Millett et al. 2018). A broad sampling of accessions representative of the distribution and environment of the crop is important for maximizing the potential morphological and biochemical (nutritional content) variation. Broad diversity is also needed to understand the potential genetic and environmental effects on the observed phenotypes. Such wide-scale studies are particularly critical for emerging orphan crops which have an untapped potential for improved human nutrition and health. Orphan vegetables are particularly important for local communities as they can easily cultivated year-round and are affordable sources of nutrients (Yang and Keding 2009). Moreover, they substantially contribute to local economies. For example, in Sub-Saharan Africa, the market of indigenous vegetables is worth billion of USD (Weinberger and Pichop 2009). Unfortunately, most studies that analyse the nutritional value of these species often focus only on local germplasm collections (Kengni et al. 2003; Kumssa et al. 2017).

Gynandropsis gynandra (syn. *Cleome gynandra)* is an orphan leafy vegetable belonging to the Cleomaceae family, the sister family of Brassicaceae. The species is a C_4 plant that is widespread in tropical areas of the world (van den Bergh et al. 2014). In several communities in Africa and Asia, the plant is harvested in the wild or cultivated in home gardens. Leaves, young shoots and occasionally flowers are eaten boiled in sauces or stews, but leaves can also be blanched and dried for preservation (Flyman and Afolayan 2006; van Den Heever and Venter 2007). There is an increasing awareness of the health benefits of the regular consumption of indigenous leafy vegetables which is enhanced by their availability and affordability. Thus, *G. gynandra* and other orphan leafy vegetables are increasingly cultivated and sold in urban areas of several African countries. To keep pace with the growing demand in these vegetables, value chain development for these species should go together with breeding efforts. Providing farmers with high-yielding and nutritious leafy vegetable cultivars which meet consumers' requirements is therefore an overriding challenge for breeders (Sogbohossou et al. 2018a).

Important nutrients in orphan leafy vegetables that can be improved and bred for include carotenoids (pro-vitamin A), tocopherols (vitamin E) and ascorbic acid (vitamin C). Carotenoids and some of their metabolites play a protective role in many Reactive Oxygen Species (ROS) mediated disorders, such as cardiovascular diseases, several types of cancer and neurological, photosensitive or eye-related disorders (Fiedor and Burda 2014). Tocopherols have vitamin E activity in mammals, with α-tocopherol and γ-tocopherol exhibiting the greatest effects. These activities involve protecting membrane lipids from ROS damage, acting as anti-inflammatory

agents, and protecting against degenerative diseases (Miyazawa et al. 2011; Usoro and Mousa 2010). The intake of vitamin E is generally low and very similar in both developing and industrialized countries (Jiang 2014). Ascorbic acid plays a critical role in cancer treatment, in the prevention and/or therapy of asthma, allergic rhinitis, atopic dermatitis, cardiovascular diseases, obesity, neurodegenerative diseases, hypertension and autoimmune diseases (Combs Jr and McClung 2017). For improving crops such as *G. gynandra,* it is important to establish their nutritional qualities and variation in these and other morphological traits in a broad collection of germplasm.

Previous studies on individual accessions have demonstrated that *G.gynandra* leaves can be rich in nutrients including carotenoids, vitamin C and minerals (Neugart et al. 2017; Odhav et al. 2007b; Omondi et al. 2017b; Schönfeldt and Pretorius 2011; Uusiku et al. 2010; van Jaarsveld et al. 2014). A set of thirty accessions of *G. gynandra* originating from East and Southern Africa showed variation in levels of minerals and health-related compounds including flavonoids and glucosinolates, as well as in 14 morphological traits (Omondi et al. 2017b). A two-fold variation in glucosinolate content and a six-fold variation in total flavonoid content in the leaves was observed while the most notable variation in minerals levels were observed for iron (five-fold) and manganese (six-fold). A more extensive set of 24 morphological characters assessed on a germplasm of 242 accessions from Asia and East-Southern Africa at the World Vegetable Center revealed differences in morphology between African and Asian accessions (Wu et al. 2017). A set of nine accessions of *G. gynandra* from East Africa, West Africa and Asia was used to analyse variation in C_4 photosynthesis related traits coupled with a phylogenetic analysis based on sequences of the Internal Transcribed Spacer region and demonstrated trait variation that clustered according to geographical origin. The accessions from Asia and West Africa were more closely related to each other than with East African accessions (Reeves et al. 2018).

The aim of the present study was to investigate the natural variation in both morphological features and nutrient content, with focus on vitamins, within *G. gynandra*, in order to provide baseline information for multiple traits breeding strategies in the future. The plant material used included 76 accessions that cover a wider geographic distribution than the aforementioned studies, including accessions from West Africa, East and Southern Africa as well as Asia. By dedicated nutrient analysis, we investigated the variation in carotenoids, tocopherols (vitamin E) and ascorbic acid (vitamin C) and their association with morphological variation. We specifically addressed the following questions: What is the extent of variation in morphological traits and levels of nutrients in the collection? Is there a relationship between the morphological characteristics of the species and the production of nutrients? Can both morphological traits and nutrient content discriminate the accessions according to their geographic origin? We also discuss the significance of our results for *G. gynandra* towards breeding for improved nutritional value and leaf yield.

Materials and methods

Plant material

A total of 76 diverse accessions of *G. gynandra* were selected from germplasm collections of the World Vegetable Center; the Laboratory of Genetics, Horticulture and Seed Science in Benin; the University of Ouagadougou in Burkina Faso and the Kenya Resource Center for Indigenous Knowledge. The selection of these accessions considered the variation in plant architecture and geographical origin in each collection. The detailed list of accessions and their provenance are presented in **Table 1**. The plants were grown under irrigated conditions with a defined light period (16h day/8h night) and temperature (24°C day/20°C night) in a greenhouse at Wageningen University from December 2016 to April 2017. The seeds were germinated on coco peat and three seedlings per accession were selected one week after sowing and transplanted on rockwool blocks. Three to five leaves per individual were harvested 8 weeks after sowing, pooled per accession, immediately frozen in liquid nitrogen, ground into a fine powder and stored at -80°C until nutrient analysis. A pooled sample of powder of all accessions was prepared and used as technical replicates for determining analytical variation during the biochemical analyses.

Table 1. List of accessions of *Gynandropsis gynandra* and their countries and region of origin.

Code	Accessions	Country of origin	Region	Source
AS1	Gyn	Malaysia	Asia	Wageningen University
AS2	TOT1048	Thailand	Asia	World Vegetable Center
AS3	TOT1480	Thailand	Asia	World Vegetable Center
AS4	TOT3514	Laos	Asia	World Vegetable Center
AS5	TOT3527	Laos	Asia	World Vegetable Center
AS6	TOT3534SC	Laos	Asia	World Vegetable Center
AS7	TOT3536	Laos	Asia	World Vegetable Center
AS8	TOT4935	Thailand	Asia	World Vegetable Center
AS9	TOT4937	Thailand	Asia	World Vegetable Center
AS10	TOT4976	Thailand	Asia	World Vegetable Center
AS11	TOT5799	Thailand	Asia	World Vegetable Center
AS12	TOT7196	Malaysia	Asia	World Vegetable Center
AS13	TOT7197	Malaysia	Asia	World Vegetable Center
AS14	TOT7198	Malaysia	Asia	World Vegetable Center
AS15	TOT7199	Malaysia	Asia	World Vegetable Center
AS16	TOT7200SC	Malaysia	Asia	World Vegetable Center
AS17	TOT7441	Laos	Asia	World Vegetable Center
AS18	TOT7449	Laos	Asia	World Vegetable Center

AS19	TOT7462	Laos	Asia	World Vegetable Center
AS20	TOT7486	Laos	Asia	World Vegetable Center
AS21	TOT7505	Laos	Asia	World Vegetable Center
AS22	TOT8996	Taiwan	Asia	World Vegetable Center
ESA1	BAR 1807B	Kenya	East and Southern Africa	National Museums of Kenya
ESA2	BGM 2107	Kenya	East and Southern Africa	National Museums of Kenya
ESA3	ELG 19/07A	Kenya	East and Southern Africa	National Museums of Kenya
ESA4	GA-01	Kenya	East and Southern Africa	National Museums of Kenya
ESA5	HBY/2307b	Kenya	East and Southern Africa	National Museums of Kenya
ESA6	KF-14	Kenya	East and Southern Africa	National Museums of Kenya
ESA7	KRC/2507a	Kenya	East and Southern Africa	National Museums of Kenya
ESA8	KSI 2407A	Kenya	East and Southern Africa	National Museums of Kenya
ESA9	KTI/0602	Kenya	East and Southern Africa	National Museums of Kenya
ESA10	KW-01	Kenya	East and Southern Africa	National Museums of Kenya
ESA11	NRK-2007A	Kenya	East and Southern Africa	National Museums of Kenya
ESA12	RW-SF-10	Rwanda	East and Southern Africa	World Vegetable Center
ESA13	TNZ 2107B	Kenya	East and Southern Africa	National Museums of Kenya
ESA14	TOT6420	Tanzania	East and Southern Africa	World Vegetable Center
ESA15	TOT6421	Tanzania	East and Southern Africa	World Vegetable Center
ESA16	TOT6439	Zambia	East and Southern Africa	World Vegetable Center
ESA17	TOT6440	South Africa	East and Southern Africa	World Vegetable Center
ESA18	TOT6441	South Africa	East and Southern Africa	World Vegetable Center
ESA19	TOT6442	South Africa	East and Southern Africa	World Vegetable Center
ESA20	TOT8887	Uganda	East and Southern Africa	World Vegetable Center
ESA21	TOT8888	Uganda	East and Southern Africa	World Vegetable Center
ESA22	TOT8889	Uganda	East and Southern Africa	World Vegetable Center
ESA23	TOT8890	Uganda	East and Southern Africa	World Vegetable Center
ESA24	TOT8891	Uganda	East and Southern Africa	World Vegetable Center
ESA25	TOT8892	Uganda	East and Southern Africa	World Vegetable Center
ESA26	TOT8915	Malawi	East and Southern Africa	World Vegetable Center
ESA27	TOT8916	Malawi	East and Southern Africa	World Vegetable Center
ESA28	TOT8917	Malawi	East and Southern Africa	World Vegetable Center
ESA29	TOT8918	Malawi	East and Southern Africa	World Vegetable Center
ESA30	TOT8925G	Kenya	East and Southern Africa	World Vegetable Center
ESA31	TOT8931	South Africa	East and Southern Africa	World Vegetable Center

ESA32	TOT8933	Zambia	East and Southern Africa	World Vegetable Center
ESA33	TOT8998	Uganda	East and Southern Africa	World Vegetable Center
ESA34	TT-00	Kenya	East and Southern Africa	National Museums of Kenya
WA1	INC-04-015	Burkina-Faso	West Africa	University of Ouagadougou
WA2	INC-06-015	Burkina-Faso	West Africa	University of Ouagadougou
WA3	INC-08-015	Burkina-Faso	West Africa	University of Ouagadougou
WA4	ODS-14-008	Benin	West Africa	University of Abomey-Calavi
WA5	ODS-15-002	Benin	West Africa	University of Abomey-Calavi
WA6	ODS-15-013	Benin	West Africa	University of Abomey-Calavi
WA7	ODS-15-020	Benin	West Africa	University of Abomey-Calavi
WA8	ODS-15-033	Benin	West Africa	University of Abomey-Calavi
WA9	ODS-15-038	Benin	West Africa	University of Abomey-Calavi
WA10	ODS-15-044	Benin	West Africa	University of Abomey-Calavi
WA11	ODS-15-045	Togo	West Africa	University of Abomey-Calavi
WA12	ODS-15-065	Togo	West Africa	University of Abomey-Calavi
WA13	ODS-15-068	Togo	West Africa	University of Abomey-Calavi
WA14	ODS-15-084	Togo	West Africa	University of Abomey-Calavi
WA15	ODS-15-094	Togo	West Africa	University of Abomey-Calavi
WA16	ODS-15-100	Togo	West Africa	University of Abomey-Calavi
WA17	ODS-15-104	Ghana	West Africa	University of Abomey-Calavi
WA18	ODS-15-111	Ghana	West Africa	University of Abomey-Calavi
WA19	ODS-15-114	Ghana	West Africa	University of Abomey-Calavi
WA20	ODS-15-121	Ghana	West Africa	University of Abomey-Calavi

Morphological characterisation

The morphological characterisation of the accessions was conducted using a set of 27 descriptors including 15 quantitative and 12 qualitative scores (**Supplemental Material 1**). Quantitative traits assessed included: days to germination, flowering time, plant height, stem diameter, leaflet length, leaflet width, leaf area, petiole length, filament length, gynophore length, androphore length, pedicel length, pod length, pod width and 1000-seeds weight. Qualitative traits observed included: plant habit, branching habit, stem colour, stem hairiness, leaf blade lobbing, leaf colour, leaf pubescence, leaf oiliness, leaflet shape, petiole colour, flower colour and leaf margin shape. The flowering time was recorded as the number of days from germination to flowering.

Extraction and analysis of lipid-soluble isoprenoids

Five hundred milligrams fresh weight of ground leaves were freeze-dried overnight and extracted with 4.5 ml of methanol/chloroform (1:1, v/v) containing 0.1% butylated hydroxytoluene (BHT) as antioxidant and 10 μM Sudan 1 as internal standard. After vortexing, samples were sonicated for 15 min and centrifuged at 2500 rpm for 10 min. Then, 2 ml of the supernatant was transferred to a new tube and dried under a gentle nitrogen flow. The dried samples were dissolved in 1 ml of ethanol containing 0.1% BHT, sonicated for 5 min and centrifuged at 2000 g for 5 min. The clear supernatant was transferred to amber vials for high-performance liquid chromatography (HPLC) analysis.

HPLC analysis was performed according to Wahyuni et al. (2011) using a YMC Pack reverse-phase C30 column (250 x 4.6 mm; 5 μm) coupled to a 20 x 4.6 mm C30 guard (YMC Inc. Wilmington, NC, USA), maintained at 35 °C. Chromatography was carried out on a Waters system consisting of a No. 2890 quaternary pump, No. 2996 photodiode array detector (PDA) and No. 2475 fluorescence detector. The mobile phase used was methanol, tert-methyl butyl ether and water:methanol (1:4, v/v) containing 0.2% ammonium acetate. Flow rate was 1 ml/min. Data were collected and analysed using the Waters Empower software. Carotenoids and chlorophylls were detected by setting the PDA to scan from 220 to 700 nm. Quantification of compounds was based on their chromatographic peak areas at specific wavelengths: both β-carotene, α-carotene and lutein at 478 nm, violaxanthin at 440 nm, chlorophyll b at 470 nm, and chlorophyll a at 665 nm. Tocopherols were detected by excitation at 296 nm and emission at 340 nm (**Supplemental Material 2**). Quantitative determination of compounds was conducted by comparison with dose-response curves constructed from authentic standards.

Extraction and analysis of ascorbic acid

The extraction and analysis of ascorbic acid (vitamin C) was as described by Wahyuni et al. (2011). An extraction solution of 5% metaphosphoric acid in purified water containing 1 mM diethylenetetriaminepentaacetic acid (a metal chelator) was prepared. Three hundred milligrams of frozen and ground material were weighed in 2 ml Eppendorf tubes and 1.2 ml of ice-cold extraction solution added. The extracts were vortexed, sonicated for 15 minutes and centrifuged at 14,000 g for 20 minutes. The supernatants were filtered through 0.2 μm polytetrafluoroethylene filters and pipetted into amber vials for HPLC-PDA analysis using the same Waters HPLC system as described above. Separation was made at 30°C using a YMC-Pro C_{18} column (YMC Europe GmbH; 150 x 3.9 mm) with 50mM KH_2PO_4 buffer (pH 4.4) as eluent at a flow of 0.5 ml/min. Quantification was made based on absorbance at 260 nm, using a calibration curve of an authentic L-ascorbic acid standard from Merck. Retention time of chromatographic peak of ascorbic acid in plant extracts was verified by co-elution with the authentic standard (**Supplemental Material 2**).

Statistical analysis

Multivariate analysis of variance (MANOVA) followed by univariate analysis of variance (ANOVA) were performed to test variations among regions of origin for all the quantitative variables. The p-value ≤ 0.05 was used to denote significant differences between mean values of regions and

determined by one-way analysis of variance (ANOVA). The Bonferroni post hoc test was used to denote statistically significant values at $p \leq 0.05$. Factorial analysis of mixed data (FAMD) combined with a hierarchical cluster analysis (HCPC) from the package "FactoMineR" (Husson et al. 2013) were performed on both quantitative and qualitative data to identify discriminating traits and group the accessions. The correlations between analysed variables were assessed using Pearson correlation coefficient and their significance was evaluated by Student t-test.

Results and discussion

Morphological variation in the collection

We found significant levels of variation for all our morphological traits in our collection of *G. gynandra* that includes 22 accessions from Asia, 34 from East and Southern Africa and 20 from West Africa (**Table 1**). The variability in quantitative morphological descriptors in the collection is summarized in **Table 2**.

Table 2. Descriptive statistics for the morphological quantitative descriptors in the collection of *Gynandropsis gynandra.*

Morphological characters	Mean	Standard deviation	Coefficient of variation (%)	Range
Days to germination	6.30	1.62	25.65	4-14
Flowering time (days)	53.95	7.48	13.86	42-73
Leaf area (cm^2)	32.74	17.24	52.64	3.72-125.16
Plant height (cm)	80.61	31.43	38.98	19.00-168.00
Stem diameter (mm)	9.44	2.04	21.62	2.81-15.43
Leaflet length (cm)	5.72	1.50	26.19	2.15-12.09
Leaflet width (cm)	2.55	0.69	27.20	0.70-4.63
Petiole length (cm)	8.42	2.99	35.49	1.43-26.00
Filament length (mm)	15.45	3.32	21.51	7.18-29.43
Gynophore length (mm)	12.82	3.96	30.92	4.95-24.31
Androphore length (mm)	14.46	3.50	24.21	5.96-31.47
Pedicel length (mm)	15.39	3.69	23.98	7.97-26.83
Pod length (mm)	74.02	20.49	27.69	39.54-142.59
Pod width (mm)	7.63	2.60	34.06	1.85-17.15
1000-seed weight (mg)	1.17	0.28	23.54	0.54-1.99

All the quantitative traits showed significant variation among accessions. The coefficient of variation among the traits was greater than or equal to 25% except for flowering time (14%), filament length (22%), androphore length (24%) and 1000-seeds weight (24%). The highest coefficient of variation was obtained for leaf area (52%). Despite the low coefficient of variation observed for flowering time, this trait ranged from 42 days in ODS-15-144 (WA19) to 73 days after germination in TOT8915 (ESA26), providing significant variation for potential changes within a breeding program. Overall, all accessions in our study flowered later than reported in previous studies: for different sets of accessions from Eastern and Southern Africa, flowering time varied between 31 and 40 days in the field in Kenya (Omondi et al. 2017b) and between 20 and 48 days in the field in Taiwan (Wu et al. 2017). The shorter flowering time in these studies might be explained by the variation in growing temperature, day-length and fertilization schemes used but could also be due to the fact that different sets of accessions

were used in these studies.

All the quantitative morphological traits showed significant differences with geographical origin except stem diameter and androphore length (**Figure 1**). For example, East-Southern African accessions germinated earlier than West African ($p<0.01$) and Asian ($p<0.001$) accessions (**Figure 1a**). Early flowering occurred in West African accessions while late-flowering accessions were from East and Southern Africa (**Figure 1b**). West African accessions had lower leaf dimensions compared with accessions from the other regions ($p<0.001$) (**Figure 1d-g**). East and Southern African accessions had flowers with longer gynophores ($p<0.001$) and filaments ($p<0.05$) compared with accessions from the other regions (**Figure 1h**) and similar trends were observed for pod length ($p<0.001$) and seed weight ($p<0.001$) (**Figure 1 k,m**). Asian accessions had shorter ($p<0.001$) but wider pods ($p<0.001$) than African accessions (**Figure 1 k,l**). Wu et al. (2017) used the same set of quantitative descriptors excluding leaf area and days to germination to characterise a collection of 242 accessions of *G. gynandra* from East-Southern Africa (n=174) and Asia (n=44). Overall, the authors found that accessions from East and Southern Africa had significantly higher average values than Asian accessions for all quantitative descriptors except leaflet width. In our case, even though East-Southern African accessions had higher average values than Asian accessions for leaf area, flowering time, petiole length and filament length, there was no significant difference between both regions for those traits. The 30 accessions from East and Southern Africa characterised in Kenya (Omondi et al. 2017b) had shorter plants (66.6±7.2 cm) compared with our results (109.1±25.0cm) and those of Wu et al. (2017) (115±27 cm). Similar to the case of flowering time, the differences in plant height could also be explained by a difference in environmental conditions, cultivation practices and the germplasm collections used. The variation in 1000-seed weight was comparable between all the studies and ranged from 1.44 to 1.71 g in Kenya (Omondi et al. 2017b), 1.0 and 2.2 g in Taiwan (Wu et al. 2017) and 0.7 to 1.8 g in our study.

The collection exhibited a wide variation in qualitative morphological descriptors. The geographical origin of the accessions was significantly associated with plant habit, leaflet shape, leaf colour and stem hairiness. Accessions with a spreading growth habit originated from West Africa (35%) and Asia (4.5%) while all accessions from East-Southern Africa had an upright (17.6%) or intermediate (82.4%) habit.

Most Asian accessions (77.3%) had deltoid leaflets while African accessions had either elliptic or lanceolate leaflets. Only East African accessions had dark green leaves (44.1%) while accessions from the other regions had green or light green leaves. None of the accessions from Asia had a glabrous stem and all exhibited various levels of stem hairiness from moderately hairy to densely hairy. In contrast, West African accessions had a high percentage of glabrous stems (65%) and none of them had a densely hairy stem.

Variation in carotenoids, tocopherols and ascorbic acid

The carotenoids detected in the leaves included lutein, violaxanthin, α- and β-carotene. Chlorophylls a and b were also quantified and provided a reliable measure of the "greenness" of the leaves. Total carotenoid concentrations varied from 36.9 μg/g FW in ODS-15-094 (WA15) to 95.8 μg/g FW in BAR1807B (ESA1). The predominant carotenoids were lutein

(21-56 µg/g FW with a coefficient of variation of 19.7%) and β-carotene (11.8-34.9 µg/g FW with a coefficient of variation of 19.5%), accounting for an average of 56% and 35% of total carotenoids, respectively. Overall, East and Southern African accessions had higher levels of carotenoids and chlorophylls compared with West African and Asian accessions except for violaxanthin for which no significant differences were observed among regions (**Figure 2a-f**).

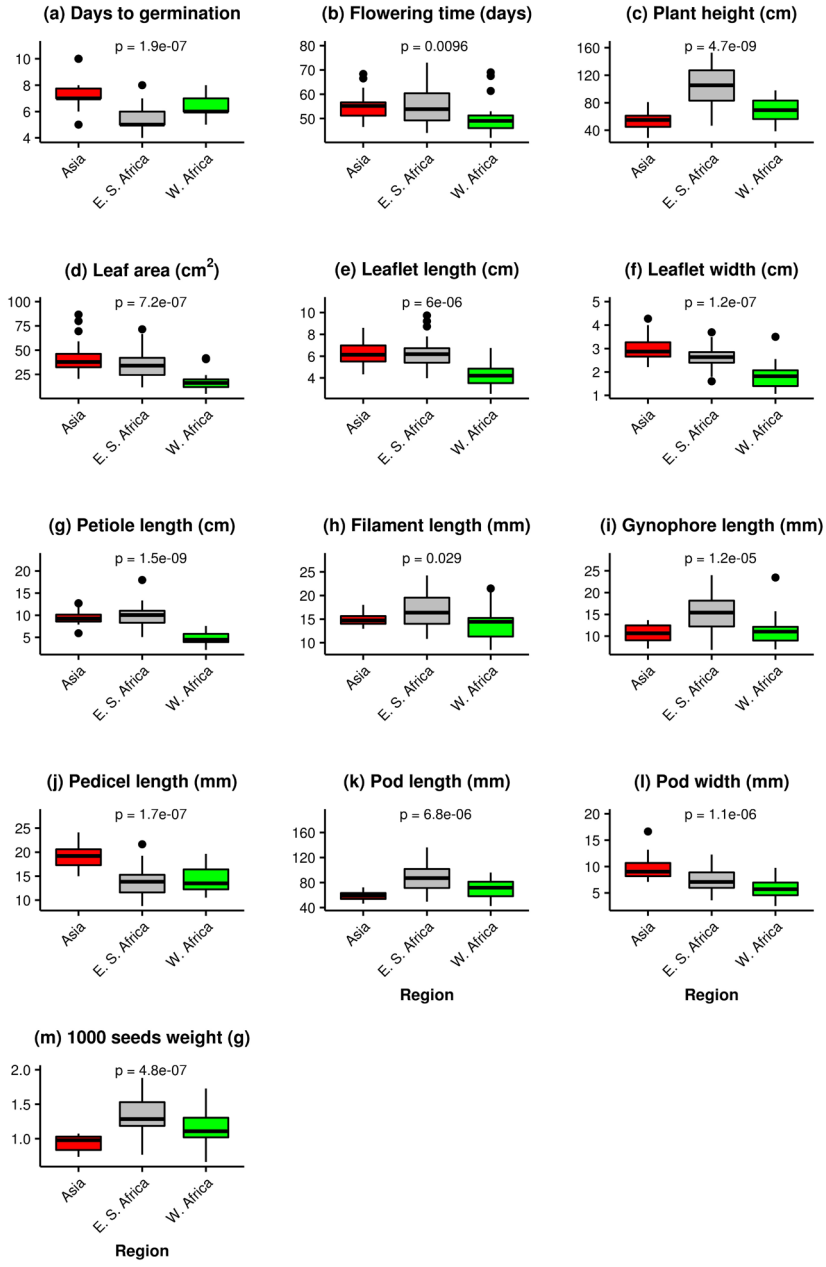


Figure 1. Box plots showing natural variation in morphological descriptors among Asian (n=22), East-Southern African (n=34) and West African accessions (n=20) of *Gynandropsis gynandra*.

Previous studies on the species reported higher levels of β-carotene: 41.2 μg/g FW (Schönfeldt and Pretorius 2011), 59.4 μg/g FW (van Jaarsveld et al. 2014), 57.1 and 64.7 μg/g DW (Neugart et al. 2017), though analyses were performed on single genotypes.

Such differences in absolute levels of metabolites between studies can be explained by the differences in growing conditions and developmental stages of leaves, the exact genotype analysed, as well as in the specific procedure and method used for analysis.

The main forms of tocopherols determined in all accessions were α- and γ-tocopherols (vitamin E). Of the many different forms of vitamin E, α-tocopherol is the most abundant form in nature and is the form preferably retained in human body (Jiang 2014). The concentration of α-tocopherol, the predominant form in *G. gynandra* leaves, ranged from 1.8 μg/g FW in TOT8889 (ESA22) to 37.2 μg/g FW in ODS-15-013 (WA6) with a coefficient of variation of 56.6% while the concentration in γ-tocopherol ranged from 0.02 μg/g FW in TOT6421 (ESA15) and TOT8892 (ESA25) to 0.4 μg/g FW in TOT7198 (AS14) with a coefficient of variation of 73.5%. West African accessions had higher average levels of α-tocopherol compared with accessions from the other regions (**Figure 2g**). In contrast, Asian accessions had the highest levels of γ-tocopherol (**Figure 2h**). To our knowledge, this is the first report of vitamin E quantification in *G. gynandra*. Levels of α-tocopherol comparable with our results were reported in other vegetables including lettuce (1.5-3.6 μg/g FW), watercress (7.2 μg/g FW) (Cruz and Casal 2013) and broccoli (27.2 μg/g FW) (Granado-Lorencio et al. 2007).

A nine-fold variation in ascorbic acid (vitamin C) was observed in the collection, ranging from 173.7 μg/g FW in TOT7197 (AS13) to 1556.8 μg/g FW in TOT4935 (AS8) with an average of 849.1 μg/g FW and a coefficient of variation of 31.9%. No geographic pattern of variation was associated with ascorbic acid levels in this collection (**Figure 2i**). Jiménez-Aguilar and Grusak (2015) also indicated high levels of ascorbic acid in *G. gynandra*, ranging from 1060 to 1400 μg/g FW in eight accessions collected in Zambia. The large variability in the levels of the selected nutrients in the collection provides an important opportunity for breeding towards cultivars with higher levels of these compounds. However, correlations between nutrients and morphological traits are critical for the design of appropriate breeding strategies.

Relationships between accessions and trait correlations

Accessions were classified in three main clusters based on the factorial analysis of mixed data including the 27 morphological descriptors and the levels of nutrients (**Figure 3a**). The first cluster included only accessions from West Africa (100%) (**Figure 3a**). Cluster 2 predominantly consisted of accessions from Asia (92%) with few accessions from East and Southern Africa (8%). Accessions in cluster 3 were mainly from East and Southern Africa (97%) with some accessions from West Africa (3%) (**Figure 3c**). Overall, clusters 1 and 2 made of accessions from West Africa and Asia respectively were closer than cluster 3 which encompasses mainly accessions from East Africa. Similar patterns of relationships between *G. gynandra* accessions from the 3 regions were revealed by Reeves et al. (2018) based on Internal Transcribed Spacer sequences. Wu et al. (2017) previously compared morphological

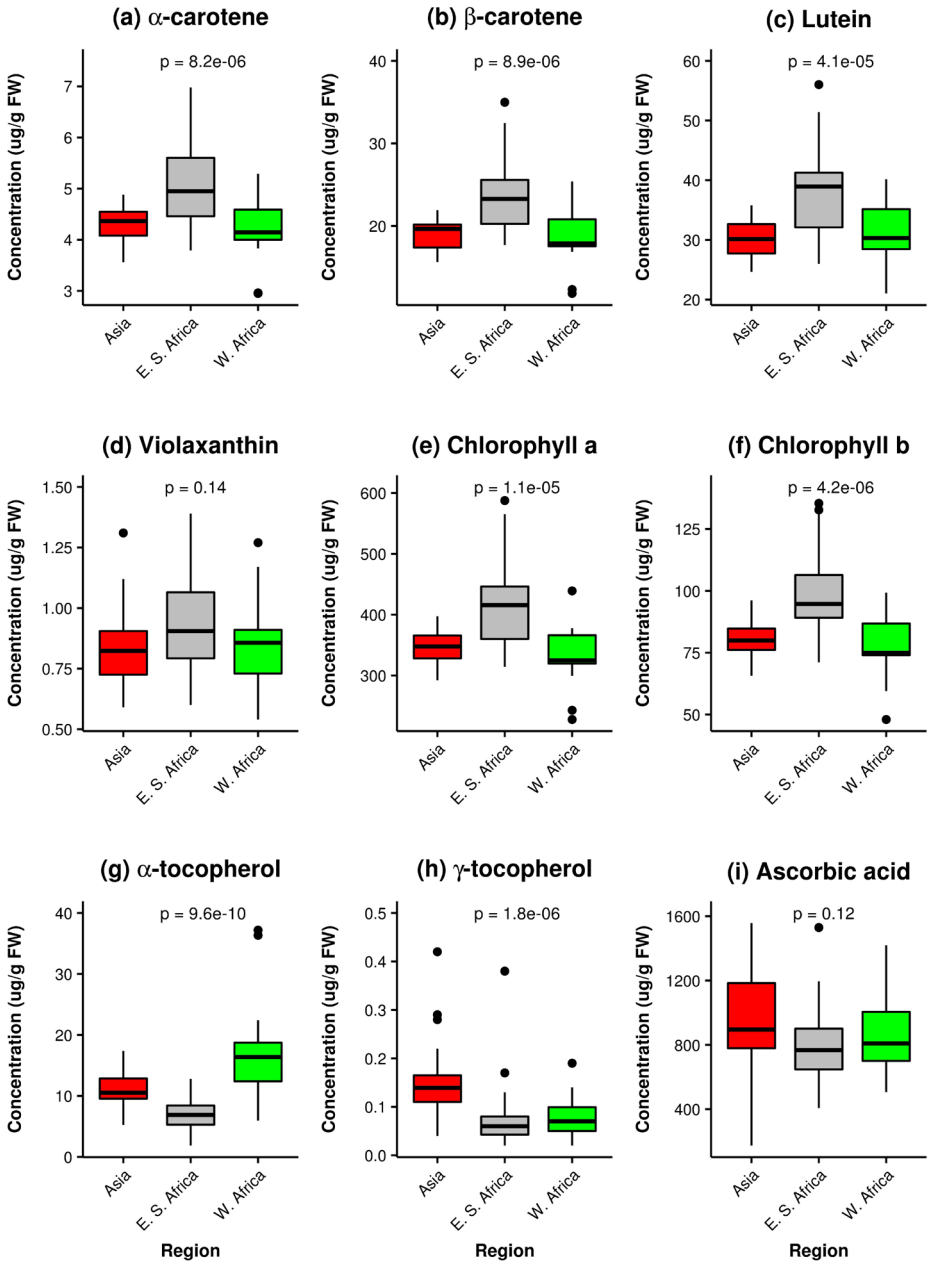


Figure 2. Box plots showing natural variation in levels of carotenoids, chlorophylls, tocopherols and ascorbic acid among Asian (n=22), East and Southern African (n=34) and West African (n=20) accessions of *Gynandropsis gynandra.*

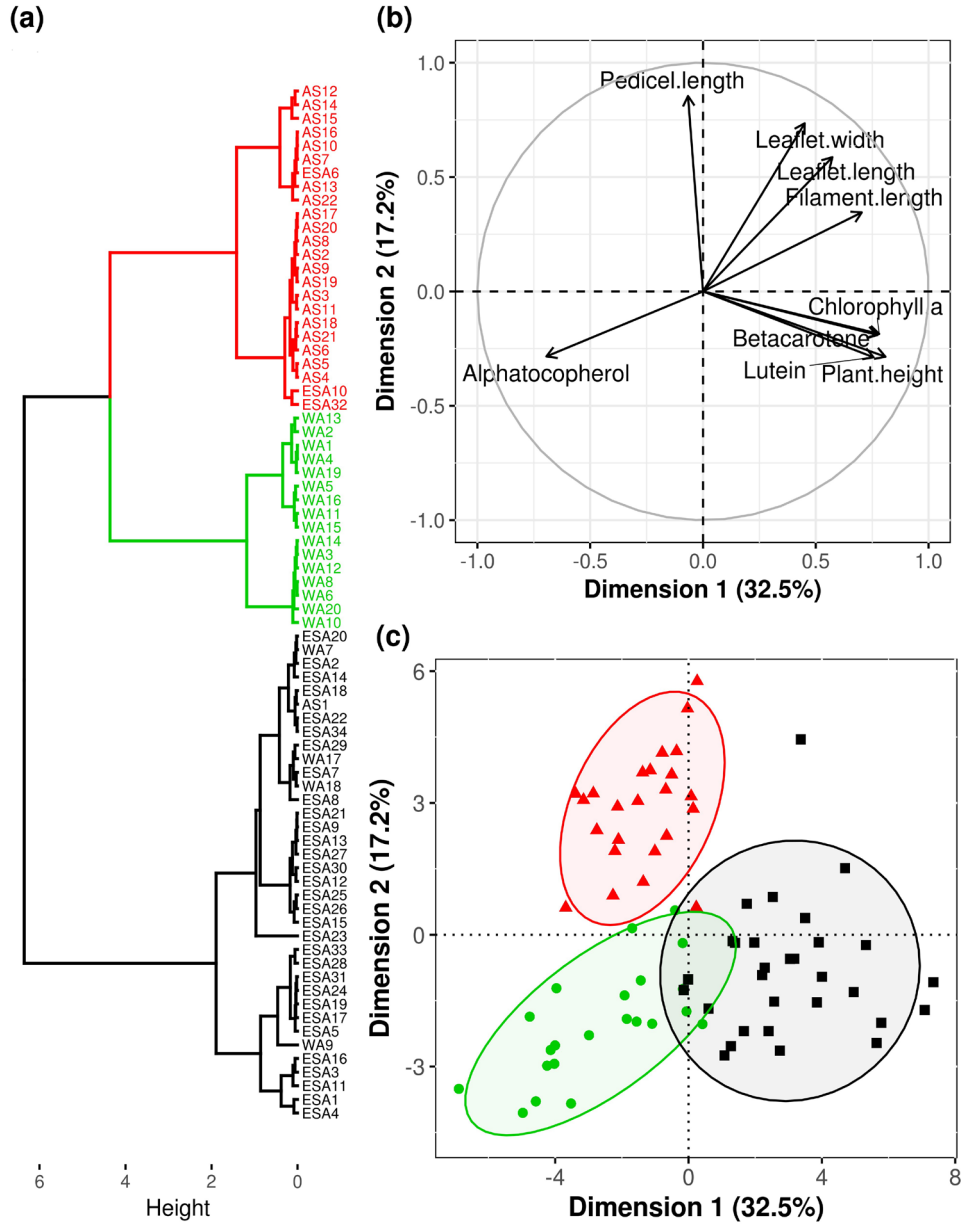


Figure 3. Clustering of the 76 accessions of *Gynandropsis gynandra* based on morphological descriptors and levels of nutrients. (a) Dendrogram based on hierarchical clustering of both quantitative and qualitative data; (b) Correlation circle based on Principal Component Analysis; only variables significantly represented on the principal components (cos2 > 0.5) are displayed. (c) Score plot of accessions based on hierarchical clustering on principal components.

Table 3. Description of clusters of accessions of *Gynandropsis gynandra* according to regions and qualitative descriptors.

Descriptors	Modalities	Cluster 1	Cluster 2	Cluster 3
Region	East Africa	0%	8%	97%
	West Africa	100%	0%	3%
	South-East Asia	0%	92%	0%
Growth habit	Upright	50%	42%	80%
	Intermediate	23%	54%	17%
	Spreading	27%	4%	3%
Branching habit	Sparse	73%	12%	33%
	Intermediate	23%	88%	60%
	Abundant	4%%	0%	7%
Stem colour (0= entirely green to 5= entirely purple)	0	45%	12%	20%
	1	9%	4%	7%
	2	23%	21%	13%
	3	13%	42%	7%
	4	5%	17%	20%
	5	5%	4%	33%
Stem hairiness	Glabrous	64%	0%	4%
	Scantly hairy	27%	29%	53%
	Moderately hairy	9%	33%	33%
	Very hairy or woolly	0%	38%	10%
Leaf colour	Light green	10%	4%	0%
	Green	90%	96%	50%
	Dark green	0%	0%	50%
Terminal leaflet shape	Deltoid	4%	76%	0%
	Elliptic	41%	20%	83%
	Lanceolate	55%	4%	17%
Leaf margin	Entire	68%	42%	80%
	Ondulate	23%	54%	20%
	Serrate	9%	4%	0%
Flower colour	White	95%	75%	80%
	Pink	5%	4%	4%
	Purple	0%	4%	16%
	Yellow	0%	17%	0%

features between African and Asian accessions of *G. gynandra* and showed differentiation based on geographic origin. However, their study included only accessions from East and Southern Africa and none from West Africa. All the qualitative traits except leaf blade lobbing, leaf pubescence and leaf oiliness contributed to classify the accessions (**Table 3**).

Morphological quantitative traits including plant height, leaflet length, leaflet width, filament length and pedicel length as well as levels of β-carotene, lutein, chlorophyll a and α-tocopherol were the most important variables which contributed to the differentiation of the accessions (**Figure 3b**). Cluster 1 was characterised by short plants with small leaves, high levels in α-tocopherol, and relatively low levels in carotenoids. Cluster 2 included accessions with short plants, broad leaves and long pedicels and relatively low levels of carotenoids and high levels of tocopherols while Cluster 3 was made of accessions with tall plants and high levels of carotenoids and chlorophylls (**Figure 3c; Table 4**).

Table 4. Average value (mean ±sd) of quantitative morphological descriptors and nutrient content in the clusters of *Gynandropsis gynandra* and the entire collection.

Descriptors	Cluster 1	Cluster 2	Cluster 3	Entire collection	P-value
Morphological descriptors					
Days to germination	6.5±0.9	7.0±1.5	5.6±1.9	6.3±1.6	0.002
Days to flowering	50.4±7.1	55.4±6.1	55.4±8.1	53.9±7.5	0.029
Leaf area (cm^2)	16.5±7.7	42.3±17.1	37.0±14.1	32.7±17.2	<0.0001
Plant height (cm)	71.3±20.5	53.6±11.9	109.1±25.0	80.6±31.4	<0.0001
Stem diameter (mm)	8.7±2.8	9.5±1.3	9.9±1.8	9.4±2.0	0.121
Leaflet length (cm)	4.2±1.0	6.2±1.1	6.4±1.2	5.7±1.5	<0.0001
Leaflet width (cm)	1.8±0.5	3.0±0.5	2.7±0.5	2.6±0.7	<0.0001
Petiole length (cm)	4.9±1.7	9.4±1.6	10.2±2.3	8.4±3.0	<0.0001
Filament length (mm)	13.5±2.9	14.8±1.4	17.4±3.7	15.4±3.3	<0.0001
Gynophore length (mm)	10.7±2.2	10.6±2.1	16.2±3.8	12.8±4.0	<0.0001
Androphore length (mm)	14.0±3.2	13.9±2.2	15.2±4.4	14.5±3.5	0.336
Pedicel length (mm)	13.8±2.4	18. 9±2.6	13.7±3.2	15.4±3.7	<0.0001
Pod length (mm)	72.5±16.7	59.1±8.1	87.3±21.5	74.0±20.5	<0.0001
Pod width (mm)	6.0±2.1	9.3±2.6	7.5±2.1	7.6±2.6	<0.0001
1000 seeds weight (mg)	1.1±0.2	0.9±0.1	1.4±0.2	1.2±0.3	<0.0001
Nutrients					
Lutein	31.1±4.8	30.1±3.2	39.0±6.9	33.9±6.7	<0.0001
α-carotene	4.2±0.5	4.3±0.3	5.2±0.8	4.6±0.8	<0.0001
β-carotene	18.8±3.1	19.1±1.8	24.0±4.2	20.9±4.1	<0.0001
Violaxanthin	0.8±0.2	0.8±0.1	0.9±0.2	0.9±0.2	0.032
Chlorophyll a	338.0±47.5	347.0±27.3	421.8±66.8	373.9±64.1	<0.0001
Chlorophyll b	78.2±11.0	79.9±7.2	100.5±17.8	87.6±16.8	<0.0001
α-tocopherol	16.3±8.1	10.9±2.9	6.9±2.5	10.9±6.2	<0.0001
γ-tocopherol	0.08±0.04	0.2±0.1	0.07±0.03	0.1±0.07	<0.0001
Ascorbic acid	827.4±179.6	938.8±341.5	793.2±257.5	849.1±272.8	0.136

The Pearson correlation matrix indicated significant positive correlations between lutein, α-carotene, β-carotene, chlorophyll a and chlorophyll b (**Figure 4**). Significantly positive correlations were observed between levels of carotenoids and chlorophylls and plant height,

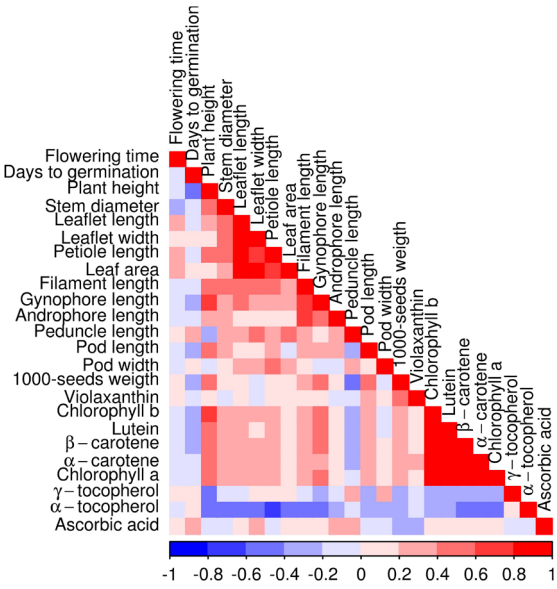


Figure 4. Correlation plot based on Pearson correlation between quantitative morphological descriptors and nutrient levels in 76 accessions of *Gynandropsis gynandra.*

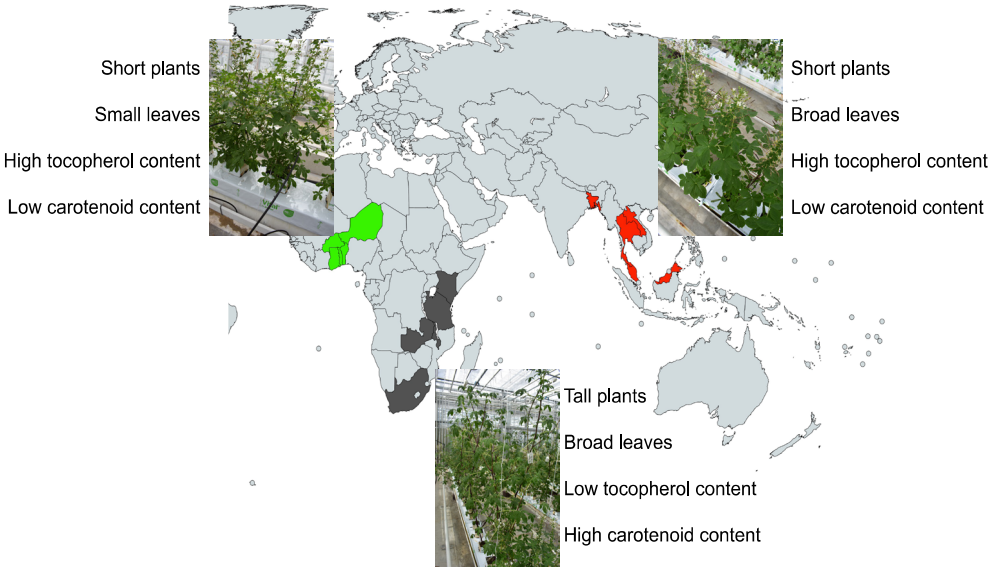


Figure 5. Summary of main differences in morphology and nutrient content between accessions from each region. West African countries are coloured in green, East African countries in grey and Southeast Asian countries in red.

leaflet length, petiole length, filament length, gynophore length, pod length and 1000 seeds weight, suggesting that taller and bigger accessions with dark green leaves are likely to have higher carotenoids content. In contrast, strong negative correlations were observed between α-tocopherol and plant height, stem diameter, leaflet length, leaflet width, petiole length, leaf area, filament length, gynophore length and androphore length, carotenoids and chlorophylls. No significant correlation was found between ascorbic acid (vitamin C) and other nutrients. The summary of the differences between the three clusters of accessions is presented in **Figure 5.** Both carotenoids and tocopherols are derived from the plastidic isoprenoid biosynthetic pathway (DellaPenna and Pogson 2006). Since carotenoids, tocopherols and chlorophylls are photosynthetic pigments, we expected positive correlations among those compounds. However, Reeves et al. (2018) observed variation in photosynthesis related traits in *G. gynandra* which corroborate the differences that we observe in terms of carotenoids and tocopherols content. The authors found that, compared with East African accessions, both Asian and West African accessions had higher water use efficiency, lower density of stomata and veins, and larger bundle sheath areas and cell sizes. Moreover, the Asian and West African accessions had increased expression of genes encoding C_4 enzymes. Such genetic variation could be induced by geographic or environmental isolation of different populations adapted to local conditions and with limited gene flow among them (Kleessen et al. 2012; Prunier et al. 2017). Several studies documented the geographical patterns of intraspecific differentiation in plant morphology (Pucher et al. 2015; Upadhyaya et al. 2017) and phytochemical profiles (Bellomo and Fallico 2007; da Silva et al. 2017; Masetti et al. 2017; Shepherd et al. 2017). We hypothesize that the observed trade-off between carotenoids and tocopherols associated with geographical origin could be also explained by differential expression of the specific genes involved in either carotenoid or tocopherol synthesis. Another explanation can be that isoprenoid levels are dependent on plant developmental stage, with higher levels in the late-flowering accessions from East Africa as compared to the early flowering West African and Asian accessions. The effect of developmental stage on the levels of these compounds in the leaves should therefore be investigated. For example, Ma et al. (2016) reported a decrease in the levels of carotenoids in *Chrysanthemum morifolium* leaves during flower development. Tracing back the origin and colonization routes of the species could also be helpful to elucidate the differences observed in the populations from the three regions. Feodorova et al. (2010) hypothesized that the speciation event leading to *G. gynandra* occurred in South Africa but their sampling was limited to accessions from South Africa and Australia. The absence of correlation between vitamin C and the other nutrients will facilitate breeding for high levels of both vitamin C and carotenoids or tocopherols in the species.

Conclusions

Our results show that morphological traits and levels of nutrients analysed varied significantly among accessions of *G gynandra*, depending on geographic regions of origin. We also established correlations between morphological traits and some of the targeted nutrients. Such information is critical to design breeding programs aiming at improving the nutritional value of the species with emphasis on pro-vitamin A, vitamins C and E. The negative correlation observed between carotenoid and tocopherol levels requires further investigation of the genetic regulation of

the biosynthetic pathways underlying these metabolites in *G. gynandra*. Moreover, differences in these nutrients across plant developmental stages should be assessed. The large variation in vitamin C between accessions and the absence of correlation between this compound and the other nutrients specifically provides a great opportunity for breeding towards new varieties with improved nutrient compositions. Based on the present results, lines with contrasting morphology and metabolite contents from East Africa and Asia have been selected and crossed for the development of mapping populations to accelerate breeding efforts in the species. Moreover, we are currently analyzing the variation in the global metabolome of this species using a subset of 48 accessions and such information would be valuable to determine anti-nutritional factors and their potential correlation with health-promoting compounds in the species. Analyzing the variation in the collection at the metabolome and the genome levels could also allow determining whether the geographic patterns of variation observed in the collection at the morphological and biochemical levels are simply due to environmental plasticity or indeed driven by genetic differentiation due to isolation by distance. Last but not least, since *G. gynandra* is usually eaten as a cooked vegetable, the impact of cooking on the levels of the nutrients in this species and the potential contribution of the different compounds to recommended daily intakes should also be investigated.

Acknowledgements

We thank the Genetics and Plant Breeding Unit of the University of Ouaga 1 Prof. Joseph Ki-Zerbo for providing us with accessions from Burkina Faso; Harry Jonker and Bert Schipper for their assistance with the HPLC analyses and Maarten Peters for the greenhouse management. This study was supported by the Applied Research Fund of the Netherlands Organization for Science under the Project "Utilizing the genome of the vegetable species *Cleome gynandra* for the development of improved cultivars for the West and East African markets" (Project Number: W.08.270.350), the Schlumberger Foundation Faculty for the Future Fellowship awarded to E.O.D.S. and the African Orphan Crops Consortium.

Supplemental information

Supplemental files are available at https://doi.org/10.1007/s00425-019-03142-1

Supplemental Material 1. Morphological data and nutrient content of 76 accessions of *Gynandropsis gynandra.*

Supplemental Material 2. HPLC chromatograms of nutritional compounds detected in *Gynandropsis gynandra* leaves.

CHAPTER 5

Natural variation in secondary metabolite production in the leafy vegetable spider plant (*Gynandropsis gynandra* L. (Briq.))

E. O. Dêêdi Sogbohossou[1,2], Enoch G. Achigan-Dako[2], Roland Mumm[3], Ric C. H. de Vos[3], M. Eric Schranz[1]

[1] Biosystematics Group, Wageningen University, Postbus 647 6700AP, Wageningen, Netherlands

[2] Laboratory of Genetics, Horticulture and Seed Sciences, Faculty of Agronomic Sciences, University of Abomey-Calavi, BP 2549 Abomey-Calavi, Republic of Benin

[3] Bioscience, Wageningen Plant Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.

Abstract

Leafy vegetables are sources of health-promoting secondary metabolites and are often used as both food and medicine. However, limited data are available on the occurence and abundance of these secondary metabolites to promote the utilization of those resources for food and nutrition. In the present study, we investigated the global metabolite variation in the leaves of 48 accessions of *Gynandropsis gynandra* using two complementary analytical platforms: liquid-chromatography mass spectrometry (LC-MS) for an untargeted comparison of semi-polar metabolites and gas-chromatography mass spectrometry (GC-MS) for an untargeted comparison of volatile organic compounds. Our results revealed large variation in 936 semi-polar compounds including flavonoids, terpenoids, glucosinolates and hydroxycinnamic acid derivatives. Unsupervised multivariate analysis indicated that the variation in relative levels of semi-polar metabolite profiles was mainly driven by geography, suggesting incipient speciation in *G. gynandra*, with the accessions from both West Africa and Asia forming a group clearly separated from the East/Southern African accessions. Volatile organic compounds detected included different sesquiterpenes, aldehydes, ketones, and sulfur-containing isothiocyanates. The variation in these volatiles was however not location-specific, but most likely linked to the taste and odour of the accessions. The relative abundance in glucosinolates and associated volatile sulfur compounds in the leaves allowed to cluster the accessions in three main groups that could be used for further plant-herbivore interaction studies. This study showed large diversity in phytochemical composition of *G. gynandra* accessions and provides a basis for towards improvement of human health or plant defense-related compounds in this leafy tropical vegetable.

Key-words: Spider plant; phytochemicals; health; plant defense; geography.

Introduction

Spider plant (*Gynandropsis gynandra*) is a leafy vegetable belonging to the Cleomaceae family, the sister family of Brassicaceae. Although it is considered a major leafy vegetable in several communities in Africa and Asia, *G. gynandra* has been for a long time understudied. The recent surge in demand for diversified and nutrient-rich diets in both advanced economies and developing countries has led to an increasing interest for "orphan" or "neglected" crops. These species are not only nutritious but may also be an important source of income for local communities. Moreover, such species are adapted to local climates: they are more tolerant to harsh conditions including heat, drought, flood, and pests and diseases. Re-introducing these so-called orphan crops in agri-food systems in developing countries could reduce rising food insecurity in these regions and contribute to more resilient agricultural ecosystems (Baldermann et al. 2016; Chiurugwi et al. 2019). Hence, *G. gynandra* may be referred to as a "super" leafy vegetable with a tremendous potential to contribute to food and nutrition, especially in regions where it is cultivated (Cernansky 2015; Sogbohossou et al. 2018a).

G. gynandra leaves are rich in nutrients including carotenoids, vitamin C and minerals (Odhav et al. 2007a; Omondi et al. 2017b; Sogbohossou et al. 2019; Uusiku et al. 2010). The leaves, young shoots and flowers are boiled and eaten in sauces/stews or can be blanched and dried for preservation (Flyman and Afolayan 2006; van Den Heever and Venter 2007). *G. gynandra* is also used as a medicinal plant in several African and Asian countries. The leaves are used as disinfectants on wounds and taken as medicine for body aches and pains, eye infections, malaria, typhoid fever, anaemia and skin disorders (Bala et al. 2010; Chweya and Mnzava 1997; Shanmugam et al. 2012; Sogbohossou et al. 2018b; Yetein et al. 2013). Pharmacological investigations of the plant species identified several compounds with confirmed medicinal properties: the leaves contain high concentrations of alkaloids, steroids, glucosinolates, flavonoids, tannins, iridoids and other phenolic compounds (Moyo et al. 2018; Moyo et al. 2013). Quercetin and kaempferol are the main flavonoids while traces of isorhamnetin are also present in the leaves of *G. gynandra* (Omondi et al. 2017b; Yang et al. 2008). Glucosinolates: glucocapparin, glucobrassicin and 3-hydroxypropylglucosinolate, were detected in the plant, and have been reported to play a role in plant defence against herbivores. Upon disruption of the leaf tissues, the hydrolysis of glucosinolates by myrosinases releases volatile sulfuric compounds, mainly isothiocyanates that have repellent properties against pests (Neugart et al. 2017; Omondi et al. 2017b; Songsak and Lockwood 2002). These studies however only target specific compounds or classes of compounds, and the studies were mostly limited to accessions collected in East and Southern Africa.

Consumers' preferences in terms of taste and odour are variable depending on the countries or regions. In the West African countries Benin and Togo, bitterness of the leaves is considered as an indicator of the healing properties of the species. Bitter and odorous leaves are therefore preferred for both food consumption and medicinal use (Sogbohossou et al. 2018a). East African communities prefer sweeter varieties and use various methods to reduce the bitterness of the leaves (Flyman and Afolayan 2006; Sogbohossou et al. 2018a). Taking consumers' preferences into account in breeding programs, requires an extensive knowledge on the secondary metabolites influencing taste and odour of the leaves as well as their health-promoting properties.

In this study, we investigated the natural variation in the global metabolite composition of 48 *G. gynandra* accessions originating from both Asia and Africa (East/Southern and West Africa), using an untargeted metabolomics approach based on mass spectrometry. These kind of untargeted approaches are useful for detecting new and confirming previously identified secondary metabolites in plants. We specifically focused on: (1) assessing the occurrence and variation in semi-polar and volatile organic compounds in our collection; (2) identifying potential compounds that may be associated with taste, aroma and plant defence in the leaves. We hypothesize that: (1) the natural variation in secondary metabolites in *G. gynandra* is linked to the geographical origin of the accessions; (2) the use of untargeted metabolomics approaches allows detection of novel compounds related to human health, taste, aroma and plant defense in the leaves of *G. gynandra*.

Materials and methods

Plant material

A total of 48 accessions of *G. gynandra* that represented the geographic distribution and morphological diversity of seed collections from the World Vegetable Center and the Laboratory of Genetics, Horticulture and Seed Science in Benin were selected. The detailed list of accessions and their provenance is presented in **Table 1**. Three replicates per accession were grown on rockwool under irrigated conditions in a greenhouse at Wageningen University for 5 months (August to December 2015). Leaf samples were taken at 8 weeks: 3-5 leaves per plant were harvested and leaf samples pooled per accession, resulting in a total of 10-15 leaves per accession. Harvested leaves were frozen in liquid nitrogen, ground and stored at -80°C for further analysis.

Metabolite data acquisition and processing

Extraction and analysis of semi-polar secondary metabolites

Semi polar metabolites were extracted and analysed using the protocol described in van Treuren et al. (2018). A total of 500 mg fresh weight of ground leaf material was extracted with 1.5 ml of 99.87% methanol containing 0.13% formic acid (final concentrations of 75% methanol and 0.1% formic acid). Frozen samples were vigourously vortexed immediately after adding the extraction solution to the frozen material, sonicated for 15 min and centrifuged at maximum speed for 15 min. The resulting extracts were passed through a 0.2 µm polytetrafluoroethylene (PTFE) filter. To check the total technical variation including extraction, sample analysis and data-processing, five quality control samples were prepared from the pooled leaf material of several randomly selected accessions.

A liquid chromatography system (Water Acquity) coupled to a photodiode array detector (Waters, 240-600nm) and a high-resolution mass spectrometry (Orbitrap FTMS, Thermo Fisher Scientific), with a C18-reversed phase chromatography system (Luna C18 column, 2.0x150 mm; Phenomenex) and negative electrospray ionization system (m/z 95-1350), were used for the analysis. Five µl of the plant extract was injected and separated using a binary gradient of ultrapure water (A) and acetonitrile (B). Solvents A and B were both acidified with 0.1% formic acid, with a flow rate of 0.19 ml/min. The initial solvent was

Table 1. List of accessions used in this study and their countries/regions of origin.

#	Accession name	Region	Country
1	Gyn	Asia	Malaysia
2	TOT1048	Asia	Thailand
3	TOT1480	Asia	Thailand
4	TOT3514	Asia	Lao People's Democratic Republic
5	TOT3527	Asia	Lao People's Democratic Republic
6	TOT3534	Asia	Lao People's Democratic Republic
7	TOT3536	Asia	Lao People's Democratic Republic
8	TOT4935	Asia	Thailand
9	TOT4937	Asia	Thailand
10	TOT4976	Asia	Thailand
11	TOT5799	Asia	Thailand
12	TOT7196	Asia	Malaysia
13	TOT7197	Asia	Malaysia
14	TOT7198	Asia	Malaysia
15	TOT7199	Asia	Malaysia
16	TOT7200	Asia	Malaysia
17	TOT7441	Asia	Lao People's Democratic Republic
18	TOT7462	Asia	Lao People's Democratic Republic
19	TOT7486	Asia	Lao People's Democratic Republic
20	TOT8996	Asia	Taiwan
21	RW-SF-10	East/Southern Africa	Rwanda
22	TOT6420	East/Southern Africa	Tanzania
23	TOT6421	East/Southern Africa	Tanzania
24	TOT6422	East/Southern Africa	Tanzania
25	TOT6435	East/Southern Africa	Kenya
26	TOT6439	East/Southern Africa	Zambia
27	TOT6440	East/Southern Africa	South Africa
28	TOT6441	East/Southern Africa	South Africa
29	TOT6442	East/Southern Africa	South Africa
30	TOT8888	East/Southern Africa	Uganda
31	TOT8889	East/Southern Africa	Uganda
32	TOT8890	East/Southern Africa	Uganda
33	TOT8891	East/Southern Africa	Uganda
34	TOT8915	East/Southern Africa	Malawi

35	TOT8916	East/Southern Africa	Malawi
36	TOT8917	East/Southern Africa	Malawi
37	TOT8918	East/Southern Africa	Malawi
38	TOT8925	East/Southern Africa	Kenya
39	TOT8926	East/Southern Africa	Kenya
40	TOT8931	East/Southern Africa	South Africa
41	TOT8933	East/Southern Africa	Zambia
42	TOT8997	East/Southern Africa	Uganda
43	TOT8998	East/Southern Africa	Uganda
44	ODS-15-020	West Africa	Benin
45	ODS-15-044	West Africa	Benin
46	ODS-15-045	West Africa	Benin
47	ODS-15-061	West Africa	Togo
48	ODS-15-117	West Africa	Ghana

composed of 95% A and 5% B, increased linearly to 35% A and 65% B at 45 min and was maintained for 2 min. The column was washed with a solution of 25% A and 75% B for 5 min and equilibrated to 95% A and 5% B for 2 min, before the next injection. Xcalibur software (Thermo) was used to regulate the instruments and for data acquisition.

Extraction and analysis of volatile secondary metabolites

Volatile metabolites were extracted based on the protocol described by Wahyuni et al. (2013). An extraction solution (5 M CaCl2 + 25 mM Tris + 25 mM EDTA) was prepared, and 1.7 ml of the solution was added to 100 mg of frozen ground leaf material in a 10-ml standard headspace (DHS) vial with screw-cap. The tubes were vortexed and sonicated at 30 °C for 10 min and 1 ml of the extract was transferred into a 10 ml crimp vial (Waters), capped and directly used for headspace SPME-GC-MS (davinci) analysis. As for LC-MS, quality control samples consisting of a pool of leaf samples of randomly selected accessions were prepared and injected after eight accession samples to assess the analytical variation.

Mass spectral alignment, filtering and clustering

MetAlign software package (www.metalign.nl) was used for baseline correction, noise estimation, and ion-wise mass spectral alignment. The software output for the GC-MS and LC-MS data sets were subsequently processed with MSClust software (Tikunov et al. 2012), that combines mass signals derived from the same molecule (natural isotopes, adducts and fragment ions) resulting in reduction of data redundancy and reconstruction of mass spectra.

Putative identification of semi-polar and volatile metabolites

The identification of selected semi-polar metabolites was based on UV/Vis light absorbance spectra and molecular weight. Putative identification of these metabolites was obtained using different databases such as the Dictionary of Natural Products (http://dnp.chemnetbase.com),

KNApSAcK (http://kanaya.naist.jp/KNApSAcK) and in-house metabolite databases. Putative identification of volatile metabolites was performed by automatic matching of mass spectra compiled by the MSClust program, with the National Institute of Standards and Technology (NIST) mass spectral library entries, using the NIST MS Search v2.0 software (http://chemdata.nist.gov/mass-spc/ms-search/). The compound hit that showed the highest matching factor (MF) value (≥800) and the lowest deviation from the retention index (RI) value was used for the putative metabolite identity.

Multivariate analysis

Metabolites were coded with their centrotype numbers obtained from the MSClust software, that were preceded by either LC or GC depending on whether they resulted from LC-MS or GC-MS analysis. Multivariate statistical analyses were performed on the semi-polar and volatile metabolite data sets containing the intensity levels of all centrotypes for all accessions. Pre-treatment of the data was performed by log2 transformation and mean centering using the R package "RFmarkerdetector" (Palla and Armano 2016). The pre-treated data were subjected to analysis of variance followed by a post-hoc Fisher LSD test and principal component analysis (PCA) with the R package "mixOmics" (Le Cao et al. 2019). Heatmaps and hierarchical cluster analysis using the UPGMA method were performed using the "heatmap.2" function of the R package "gplots" (Warnes et al. 2016).

Results

Natural variation in semi-polar metabolites

The relative intensity of 936 metabolites were extracted from leaf samples using the untargeted LC-MS approach (**Supplemental Table 1**). PCA on the 48 accessions revealed a separation of the accessions according to their geographic origin (**Figure 1**). The first component (PC1) explained 28.8% of the total variation and separated accessions from East/Southern Africa from those from West Africa and Asia. The second principal component (PC2) explained 10.9% of the variation and separated accessions from West Africa from those from Asia. The third principal component (PC3) which explained 6.8% of the variation, separated accessions from West Africa from those from both East/ Southern Africa and Asia. Of the 936 metabolites, 626 (66.9%) had significantly different abundance levels between East/ Southern Africa and Asian/ West African accessions ($p<0.05$). From these 626 differential metabolites, 270 (43.1%) were more abundant in East/Southern African accessions than in West African/Asian accessions, with a fold variation of the average values between 2 (compound LC1645) and 78 (compound LC2379). On the other hand a total of 228 (36.4%) metabolites were more abundant in West African/Asian accessions compared with East/Southern African accessions, with a fold variation average between 2 (LC2896) and 425 (LC1917). Comparing metabolite levels in West African and Asian accessions revealed 444 (47.4%) significantly different metabolites out of the initial 936. Of these metabolites, 243 (54.7%) were more abundant in Asian accessions with a fold-variation from 2 (LC6772) to 793 (LC586) while 195 (43.9%) were more abundant in West African accessions with a fold-variation from 2 (LC2622) to 556 (LC6401).

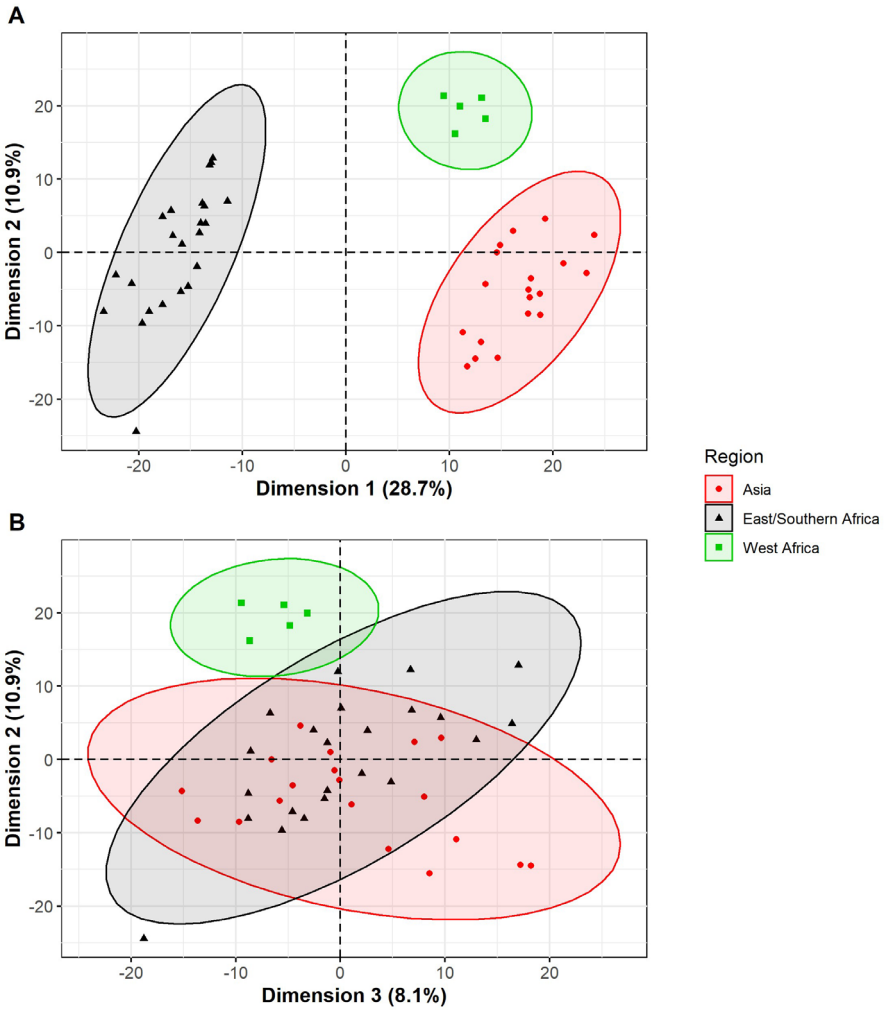


Figure 1. Principal component analysis of relative levels of 936 semi-polar compounds detected in the leaves of 48 accessions of *Gynandropsis gynandra*. (A) Projection of individuals on principal components 1 and 2. (B) Projection of individuals on principal components 2 and 3. Confidence ellipses at 0,95 level were drawn around points from the same geographical region of origin.

A heatmap (**Figure 2**) based on extremely differential metabolites among accessions with PCA loading scores > 0.7 on both PC1 and PC2, revealed the relationships amongst the accessions and metabolites in more details. Two main clusters of accessions (A1 and A2) and three major clusters of metabolites (B1, B2 and B3) were identified (**Figure 2**). Cluster A1 included the 20 accessions from Asia and 5 accessions from West Africa while the 23 accessions from East Africa were all in cluster A2. The metabolite cluster B1 comprised of compounds mainly present in East/Southern African accessions. Cluster B2 included compounds abundant in West African accessions and in most Asian accessions, that were also present at moderate levels in some East/Southern African accessions. Compounds in cluster B3 were highly abundant in Asian and West African accessions but were present at very low levels in East/Southern African accessions.

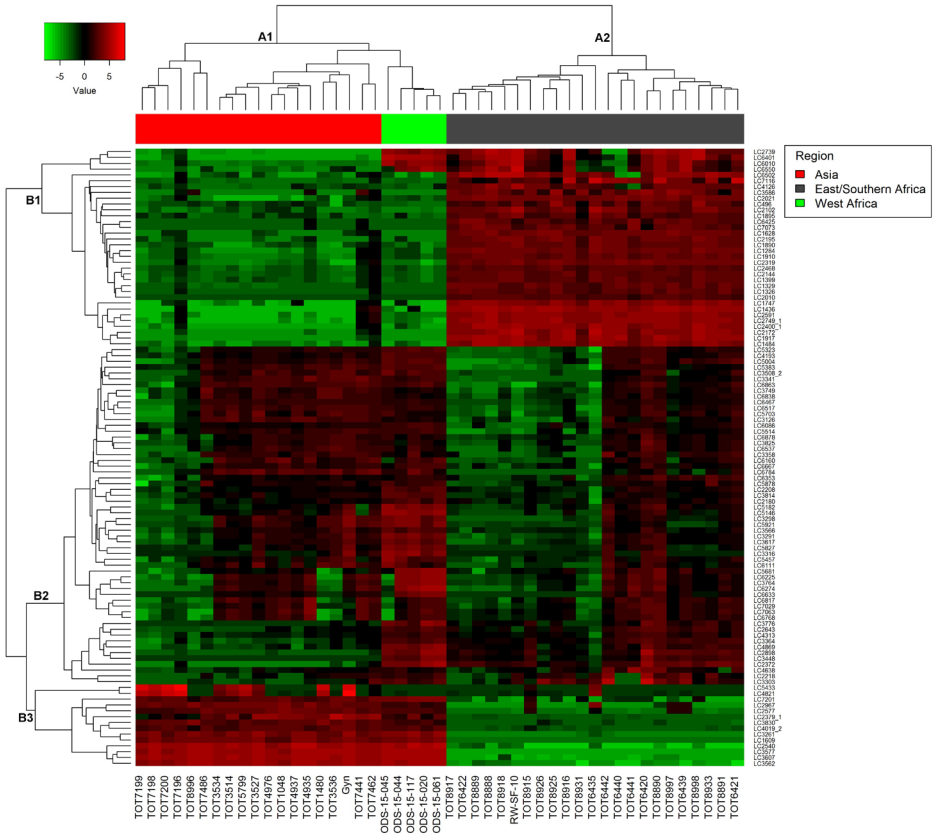


Figure 2. Heatmap of the log2-transformed and mean-centered abundance values of the top 107 significant semi-polar metabolites obtained from PCA loadings in 48 accessions of *Gynandropsis gynandra.*

Classes of metabolites that were abundant in Asian and West African accessions but in low levels in East/Southern African accessions (**Figure 3, a-e**) were: hydroxycinnamic acid derivatives (e.g. LC2540: 1,2-Dihydroxy-1,2,3-propanetricarboxylic acid 1-O-Caffeoyl(E) and LC3607: 1,2-Dihydroxy-1,2,3-propanetricarboxylic acid 1-O-Feruloyl(E)), terpenoids (e.g. LC3341: 1,15-Dihydroxy-11(13)-eudesmen-12,6-olide and LC2765: 2,3-dihydroxy-4-megastigmen-9-one) and flavonoids (e.g. LC3830: 5,7-Dihydroxy-4'-methoxyflavanone). East/Southern African and West African accessions (**Figure 3f**) exhibited high levels of quercetin-3-rutinoside (LC3890), a flavonoid, while glucocapparin (LC880; methylglucosinolate) was abundant in Asian accessions (**Figure 3g**).

Classes of metabolites with high levels in East/Southern African accessions (**Figure 3, h-k**) included flavonoids (LC2021: 5,6,7,8-tetrahydroxy-2H-1-benzopyran-2-one and LC2468: 5,7,8-trihydroxy-2H-1-benzopyran-2-one), and phenolic compounds (LC2400: ellagic acid and LC2749: 2-trans-O-feruloylglucaric acid). In West African accessions, specific compounds with high levels included 1,3-dihydroxy-12,6-eudesmanolide (LC5323), a sesquiterpenoid, and methylsyringin (LC3291), a phenyl glycoside. Glucobrassicin (LC3829; 3-indolylmethylglucosinolate) was rather present at low levels in West African accessions, when compared to East/Southern African and Asian accessions (**Figure 3, l-n**).

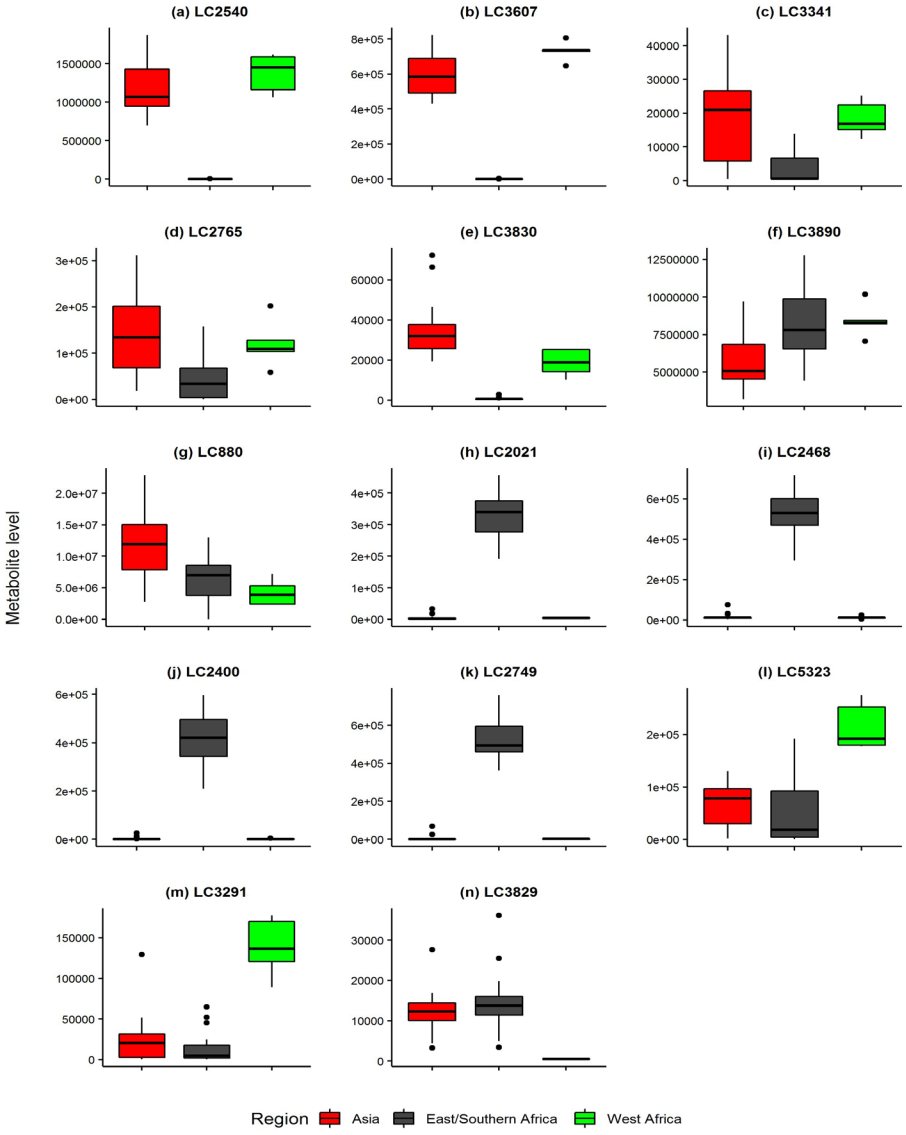


Figure 3. Box plots showing the variation in relative levels of 14 selected semi-polar metabolites in the leaves of 48 accessions of *Gynandropsis gynandra*. Putative identities: LC2540: 1,2-Dihydroxy-1,2,3-propanetricarboxylic acid,1-O-Caffeoyl(E); LC3607: 1,2-Dihydroxy-1,2,3-propanetricarboxylic acid, 1-O-Feruloyl(E)]; LC3341: 1,15-Dihydroxy-11(13)-eudesmen-12,6-olide; LC2765: 2,3-dihydroxy-4-megastigmen-9-one; LC3830: 5,7-Dihydroxy-4'-methoxyflavanone; LC3890 quercetin-3-rutinoside; LC880: glucocapparin; LC2021: 5,6,7,8-tetrahydroxy-2H-1-benzopyran-2-one; LC2468: 5,7,8-trihydroxy-2H-1-benzopyran-2-one; LC2400: ellagic acid; LC2749: 2-trans-O-Feruloylglucaric acid; LC5323: 1,3-dihydroxy-12,6-eudesmanolide; LC3291: methylsyringin; LC3829: glucobrassicin.

The geographic patterns of clustering of semi-polar metabolites suggests that the variation in these metabolites results from genetic adaptation processes of this species. Since our analyses were performed on accessions collected in their natural habitat and self-pollinated in controlled conditions for 4 to 6 generations, we hypothesize that the metabolic profiles are mostly under genetic control.

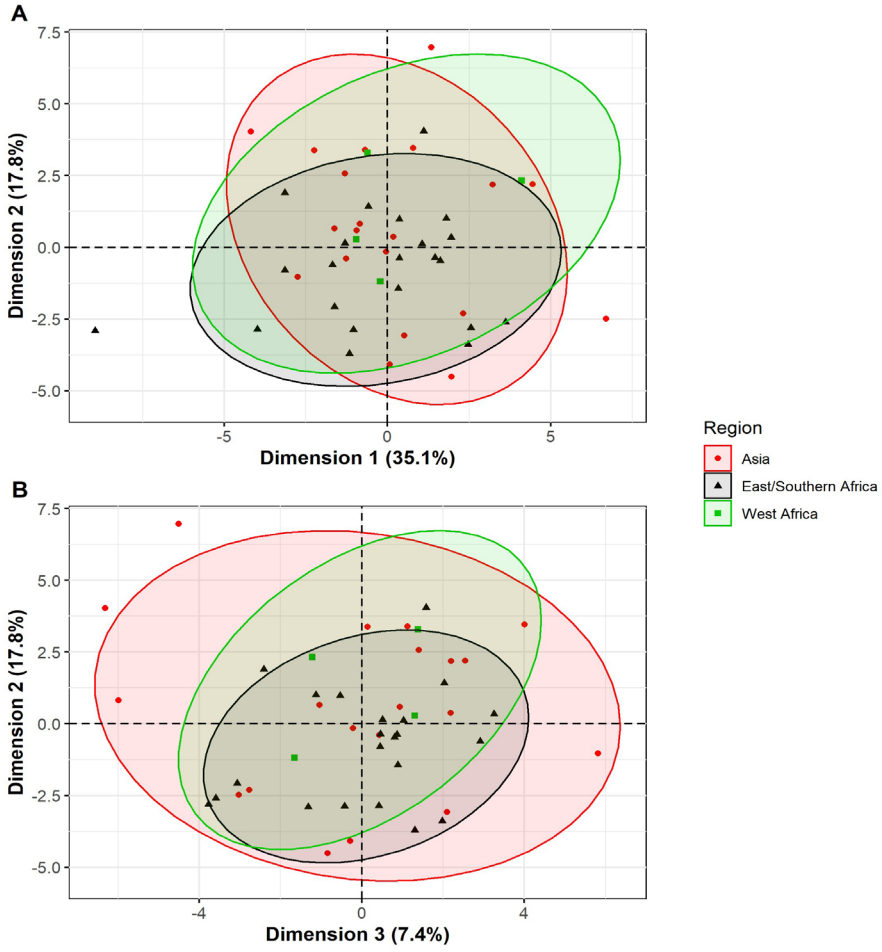


Figure 4. Principal component analysis of relative levels of 130 volatile compounds detected in the leaves of 46 accessions of *Gynandropsis gynandra***.** (A) Projection of individuals on principal components 1 and 2. (B) Projection of individuals on principal components 2 and 3. Confidence ellipses at 0,95 level were drawn around points from the same geographical region of origin.

Natural variation in volatile metabolites in *Gynandropsis gynandra*

A total number of 130 volatile metabolites was detected in the leaves of our accessions (**Supplemental Table 2**). Accessions TOT6440 and ODS-15-117 were detected as outliers and removed from the analyses. PCA on the remaining 46 accessions revealed no clear separation of the accessions based on their geographic origin (**Figure 4**).

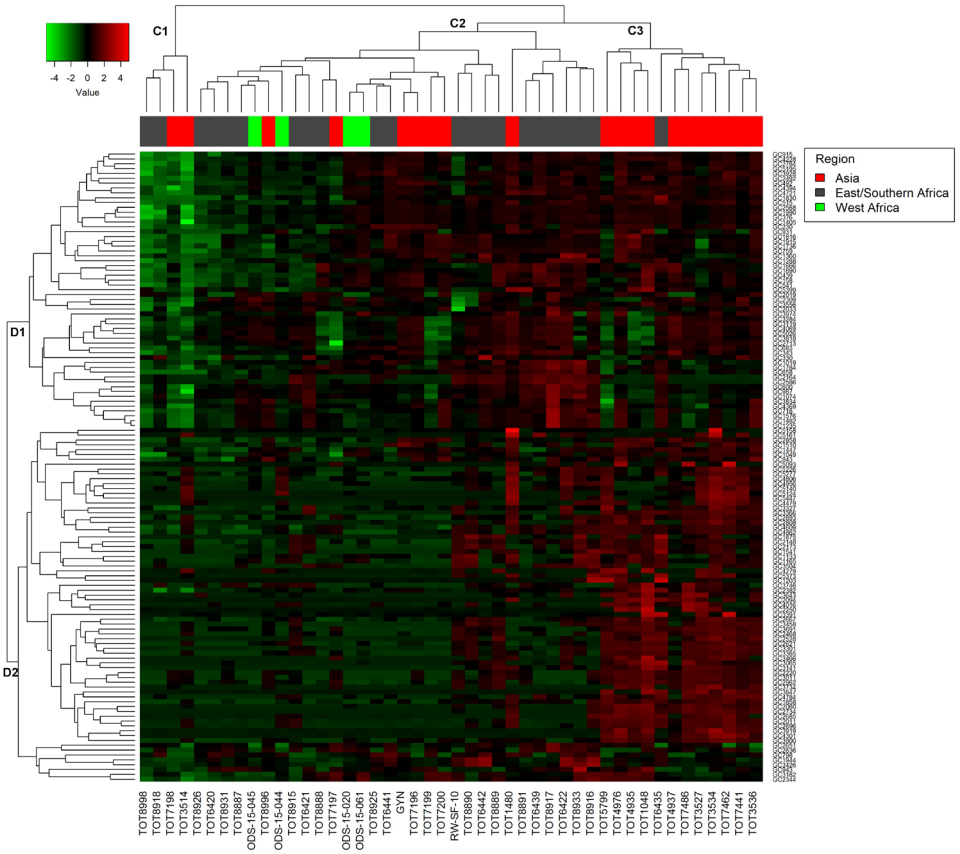


Figure 5. Heatmap of the log2 transformed and mean-centered abundance values of the 130 volatile metabolites detected in the leaves of 46 accessions of *Gynandropsis gynandra*.

A heatmap generated based on the log2 transformed and mean-centered metabolite levels indicated three clusters of accessions (C1, C2 and C3) and two main clusters of metabolites (D1 and D2) **(Figure 5)**. Cluster C1 consisted of 4 accessions (2 from East Africa and 2 from Asia) with overall low levels of volatiles. Cluster C2 consisted of 30 accessions (19 from East Africa, 7 from Asia and 4 from West Africa) with high levels of D1 compounds but low levels of D2 compounds. Cluster C3 on the other hand, was made up of 12 accessions (11 from Asia and one from East Africa) with moderate to high levels of volatiles.

Of the 132 volatiles detected, 60 were putatively identified. Volatile metabolites present in cluster

D1 mainly included aldehydes (e.g. GC987: (E)-2-Pentenal; GC1235: (Z)-3-hexenal, GC1482: (E)-2-hexenal, GC2713: (E,E)-2,4,-heptadienal and GC4069: 1-, 2,6,6-trimethyl-cyclohexene-1-carboxaldehyde), ketones (e.g. GC3119: 3,5-Octadien-2-one and GC2492: 6-methyl-5-hepten-2-one), monoterpenes (e.g. GC2858: 1,8-cineole, syn. eucalyptol) and alcohols (e.g. GC1074: (Z)-2-penten-1-ol and GC2784: 2-ethyl-1-hexanol).

The metabolite cluster D2 mainly included esters (e.g. GC2148: propyl 2-methylbutanoate; GC2528: 2-Methylpropanol 3-methylbutanoate; GC2667: Isobutyl isovalerate and GC3365: 2-Methylbutyl 2-methylbutanoate), sesquiterpenes (e.g. GC4862: alpha-humulene; GC4956: bicyclosesquiphellandrene; GC5093: (E,E)- alpha-Farnesene and GC5277: cis-Calamenene) and sulfuric compounds (e.g. GC798: methyl thiocyanate; GC843: methyl isothiocyanate; GC931: dimethyl disulfide, GC1447: isopropyl isothiocyanate and GC1784: 2-ethylthiophene).

Natural variation in glucosinolates and other plant defense related compounds in *Gynandropsis gynandra*

In our study, the glucosinolates putatively identified were: glucobrassicin or indolylmethyl glucosinolate (LC3829) and glucocapparin or methylglucosinolate (LC880). Volatile sulfuric compounds identified included dimethyldisulfide (GC931), methylisothiocyanate (GC843), isopropylisothiocyanate (GC1447), methylthiocyanate (GC798) and 2-ethylthiophene (GC1784). A closer look at the metabolic profiles **(Figure 6)** indicated 5 clusters of accessions (E1 to E5), and 2 clusters of metabolites: the volatile sulfuric compound cluster (F1) and the glucosinolate cluster (F2).

Accessions in clusters E1 and E2 have low levels of volatile sulfuric compounds overall. Cluster E1 included accessions with high levels of the two glucosinolates while cluster E2 consisted of accessions with high levels of glucobrassicin and low levels of glucocapparin. In cluster E3, accessions were all from West Africa and, had moderate levels of volatile sulfuric compounds and low levels of the glucosinolates. Accessions in cluster E4 had moderate levels of volatiles, high levels of glucobrassicin and low levels of glucocapparin. Cluster E5 was composed of accessions with moderate to high levels of all volatile sulfuric compounds.

Accessions from clusters E2 and E5 may be interesting for genetic studies on both volatile sulfuric compounds and glucocapparin content while accessions from cluster E3 may possibly be combined with accessions from clusters E2, E4 or E5 to study the genetic basis of glucobrassicin biosynthesis in the species.

Discussion

G. gynandra leaves contains different classes of semi-polar metabolites. The overall variation in semi-polar metabolites including phenylpropanoids, flavonoids, glucosinolates and terpenes, positively correlated with the different geographic regions of the accessions in our collection. Accessions from West Africa and Asia were more similar and clearly distinct from East/Southern African accessions which had unique metabolite signatures. A similar geographical pattern of differentiation in *G. gynandra* was previously demonstrated using a combination of morphological traits and isoprenoid levels (Sogbohossou et al. 2019) and a set of

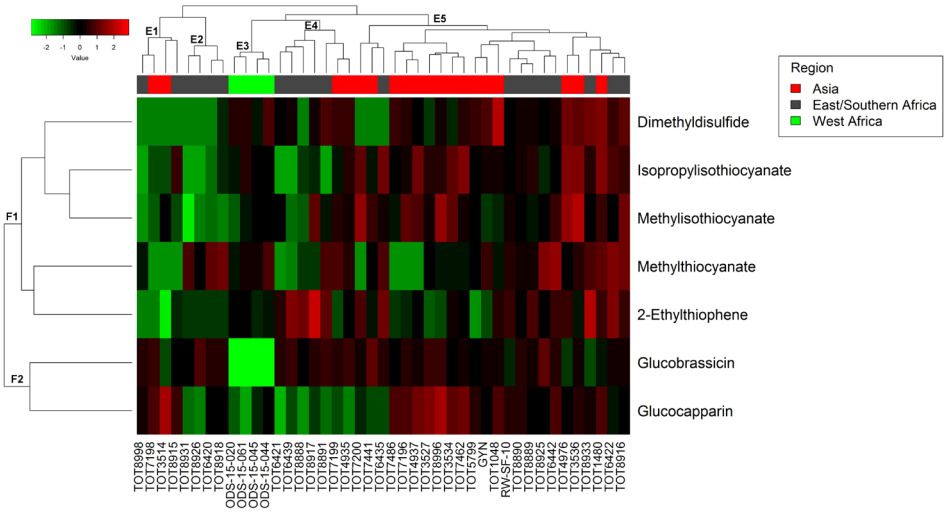


Figure 6. Heatmap of the log2 transformed and mean-centered abundance values of glucosinolates and volatile sulfur compounds in 46 accessions of *Gynandropsis gynandra.*

photosynthesis related traits (Reeves et al. 2018). The similarity between the various studies suggest that geographical-specific patterns may also be reflected at genome level. Wahyuni et al. (2013) found comparable patterns of species-specific profiles of semi-polar metabolites in the fruits of *Capsicum* spp. The authors suggested that differences in the regulation of metabolic pathways, the activity of rate-limiting enzymes or the substrate specificity of specific modifying enzymes could explain the observed differences in metabolite levels. In our study, the differences between accessions possibly indicate an on-going speciation event in *G. gynandra.* Another hypothesis is that these patterns in metabolite profiles may be due to accessions' selection based on consumers' preferences in different geographical locations. Further genomic analyses on the observed genetic patterns, genomic signatures of selection between populations, structural variation and demographic history reconstruction, using a broader collection of accessions could help unravel this question.

The potential benefits of *G. gynandra* leaf consumption in human health, especially in relation to its potential antibiotic and anti-inflammatory activities (Sogbohossou et al. 2018b; Yetein et al. 2013), were confirmed by the presence of different flavonoids, terpenoids, hydroxycinnamic acid derivatives as well as various other phenolic compounds known for their antioxidant, biocide or anti-inflammatory properties (Baron 2018; Maurya and Devasagayam 2010). Flavonoids were previously identified in two accessions of *G. gynandra* (Neugart et al. 2017) and the present study confirmed the presence of quercetin-3-rutinoside, kaempferol and hydroxycinnamic acids. Hydroxycinnamic acid reportedly possess potent antioxidant and anti-inflammatory properties, and may potentially be used for diabetes and hyperlipidaemia prevention (Alam et al. 2016). Coumarin derivatives including 5,6,7,8-tetrahydroxy-2H-1-benzopyran-2-one and 5,7,8-trihydroxy-2H-1-benzopyran-2-one are known for their pharmacological properties (Dandriyal et al. 2016; Witaicenis et al. 2014). Coumarin derivatives also have interesting anticancer properties: they

have minimum side effects and multi-drug reversal activity (Dandriyal et al. 2016). Ellagic acid and derived metabolites have antioxidant, estrogenic and/or anti-estrogenic, anti-inflammatory and anti-carcinogenic properties (Bell and Hawthorne 2008; Landete 2011; Vattem and Shetty 2005). It should be noted that most metabolites detected by the LC-MS analysis are not currently annotated, but may include other flavonoids or health-related phenolics.

The profiles of volatile organic compounds did not positively correlate with the geographical origin of the accessions. However, compounds abundant in cluster C3 including isobutyl isovalerate (GC2667), butanoic acid, 2-methyl-, 3-methylbutyl ester (GC3301), butanoic acid, 2-methyl-, 2-methylbutyl ester (GC3365), isobutyl isobutyrate (GC1858), propanoic acid, 2-methyl-, 3-methylbutyl ester (GC2685) and propanoic acid, 2-methyl-, 2-methylpropyl ester (GC2011) were characterised by a sweet fruity flavour (The Good Scents Company 2019). Cubenol (GC5447: spicy), linalool (GC3327: floral, citrus scent), trans-beta-Ocimene (GC2893: green, tropical, woody), caryophyllene (GC4609: spicy, woody) (The Good Scents Company 2019) were other abundant compounds with specific flavour and taste in the C3 cluster. Volatile compounds abundant in cluster C2 were characterised by pungent and spicy smells. Some of these compounds include 1-Penten-3-ol (GC600: pungent, horseradish-like), 2-Pentenal, (E)- (GC987: pungent, green, fruity apple-like), (E,E)-2,4-hexadienal (GC2033: pungent fatty green), decanal (GC3974:sweet, waxy, orange peel), 3,5-octadien-2-one (GC3284: fruity green grassy), 2-Penten-1-ol, (Z)- (GC1074; mustard horseradish) and 2-Hexenal, (E)- (GC1576: sharp, penetrating fresh leafy green, spicy) (The Good Scents Company 2019).

Of the 31 volatiles in *G. gynandra* previously reported (Nyalala et al. 2013), 16 were identified in our study and they were mainly isothiocyanates, terpenes and aldehydes. Nyalala et al. (2013) highlighted the inactivity of spider mites exposed to 2,4-heptadienal or β-cyclocitral, (Z)-2-pentenol, or methyl isothiocyanate, all compounds detected in our study. Natural variation in glucosinolates and their catabolic sulfur-containing volatile compounds also provides a basis for further investigation of herbivore interactions with *G. gynandra*. All glucosinolates share a common chemical backbone: an S-glycosylated thiohydroximate sulfate ester, with a variable side chain (Fahey et al. 2001). In our study we identified an aliphatic glucosinolate: glucocapparin and an indole glucosinolate: glucobrassicin. Omondi et al. (2017) identified 3-hydroxypropyl glucosinolate as the main glucosinolate present in the different plant organs of 30 accessions of *G. gynandra* collected from various East/Southern African countries while Neugart et al. (2017) observed mainly glucocapparin and only traces of indole glucosinolates (glucobrassicin and 4-methoxyglucobrassicin) in *G. gynandra* leaves. Glucosinolate levels are also strongly influenced by plant developmental stages as was shown in *Aethionema arabicum* for which glucosinate levels in the leaves drastically decreased after flowering (Mohammadin et al. 2017). The discrepancies between the studies mentioned above may be explained by differences in accessions analysed, and possibly also due to differences in growth conditions and plant developmental stages. The relative low levels of glucosinolates observed in West African lines at the time of sampling may be as a result of relatively early flowering in these lines (Sogbohossou et al. 2019).

Overall, our results provide clear evidence of genetic diversity in leaf phytochemicals, specifically in non-volatile compounds. This information can be applied in crop breeding for the selection of parental accessions that may produce lines with beneficial characteristics such as: medicinal

value, taste, odour, and for plant defence against herbivores or other biotic stresses. Further annotation of leaf metabolites and investigating natural variation of key metabolites in different developmental stages, food processes and in response to different environmental stresses, may also contribute to breeding for improved quality in *G. gynandra*.

Acknowledgements

We thank the World Vegetable Center and especially Andreas Ebert for providing seeds of the accessions from East Africa and Asia. We also thank Maarten Peters for the greenhouse management and Henriette van Eekelen for helping with the mass spectrometry analyses. This study was supported by the Applied Research Fund of the Netherlands Organization for Science under the Project "Utilizing the genome of the vegetable species *Cleome gynandra* for the development of improved cultivars for the West and East African markets" (Project Number: W.08.270.350) and the Schlumberger Foundation Faculty for the Future Fellowship awarded to E.O.D.S.

Supplemental files

Supplemental files are available upon request.

Supplemental Table 1. Semi-polar metabolite composition in leaves of 48 accessions of *Gynandropsis gynandra*

Supplemental Table 2. Volatile metabolite composition in leaves of 46 accessions of *Gynandropsis gynandra*

CHAPTER 6

The draft genome of *Gynandropsis gynandra* (Cleomaceae) provides insights in whole-genome duplication, population genetic structure and genome-metabolome associations

E. O. Dêêdi Sogbohossou[1,2], Andrea Brautigam[3], Erik van den Bergh[4], Enoch G. Achigan-Dako[2], Prasad S. Hendre[5], Ramni Jamnadass[5], Samuel Muthemba[5], Robert Kariba[5], Allen Van Deynze[5], Julian Hibberd[6], Andreas P. M. Weber[7], M. Eric Schranz[1]

[1] Biosystematics Group, Wageningen University, Droevendaalsesteeg 1, 6708PB, Wageningen, Netherlands

[2] Laboratory of Genetics, Horticulture and Seed Science, Faculty of Agronomic Sciences, University of Abomey-Calavi, BP 2549 Abomey-Calavi, Republic of Benin

[3] Faculty of Biology, Bielefeld University, Bielefeld, Germany

[4] Wageningen Data Competence Center, Wageningen University, Droevendaalsesteeg 1, 6708PB, Wageningen, Netherlands

[5] African Orphan Crops Consortium (AOCC), World Agroforestry (ICRAF), Nairobi, Kenya

[6] Department of Plant Sciences, Cambridge University, Cambridge CB2 3EA, United Kingdom

[7] Institute of Plant Biochemistry, Cluster of Excellence on Plant Science (CEPLAS), Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

Abstract

Orphan crops are plant species used around the world for food, medicine and/or fuel which contribute to users' livelihoods, but have not been included in mainstream agricultural research and development agenda. Rapid advances in -omics technologies provide an unprecedented opportunity to accelerate breeding of new and orphan crops which hold the potential in addressing food and nutritional insecurity in the world. Here, we present the draft genome and the whole-genome resequencing of 53 accessions of *Gynandropsis gynandra* (Cleomaceae), a leafy vegetable and medicinal plant used in Africa and tropical Asia. *G. gynandra*, one of the target crops of the African Orphan Crops Consortium, is an important source of vitamins and health-promoting secondary metabolites and is readily used as a model crop to study C_4 photosynthesis. The reference genome of the species was assembled into 1693 scaffolds with an N50 of 1.38 Mb. A survey of the genome for ancient polyploidy events revealed that *G. gynandra* underwent a whole-genome duplication instead of a whole-genome triplication as previously hypothesized. SNP calling based on the re-sequenced diversity panel yielded 10.8 million SNPs. Genetic diversity analyses showed that variation at the genome level was associated with the geographic origin of the accessions, and a strong differentiation between populations was identified, consistent with morphological characterisation and metabolome profiling. Accessions from West Africa and Asia exhibited a higher level of linkage disequilibrium suggesting that these populations diverged more recently than those of East/Southern Africa. Thus, our results suggest an African origin for the species, a hypothesis that should be further tested with a more comprehensive set of accessions. The high genetic diversity in East/Southern African and Asian populations constitute a tremendous opportunity for breeding programs. The reference genome coupled with SNP markers developed will facilitate the investigation of the genetic basis of agronomic and quality traits in *G. gynandra* and accelerate the development of high-yielding and nutrient-rich cultivars.

Key-words: genome, resequencing, origin, whole-genome duplication, breeding, *Gynandropsis gynandra*.

Introduction

With the evident challenges of sustainably feeding the growing world population, there is an urgent need for more diverse and nutrient-rich diets to achieve food security (Khoury et al. 2014). Crops used by diverse communities around the world at local or regional scales, often referred to as "orphan" or "neglected" are largely overlooked by the mainstream scientific community. Nonetheless, such species constitute important sources of nutrients, have numerous therapeutical properties, provide income to local communities, are adapted to harsh environments, and are increasingly sought for (Chiurugwi et al. 2019; Hendre et al. 2019). The global flora includes over 12,000 edible plants including many fruits and leafy greens (Ozturk et al. 2018). In Africa alone, the Plant Resources of Tropical Africa Foundation listed approximately 8,600 edible species (Prota 2010). Hence, several research teams are developing breeding programs targeting orphan crops, especially cereals, legumes, and fruit trees (Chang et al. 2018; Chanyalew et al. 2019; Debieu et al. 2017; Varshney et al. 2009; Varshney et al. 2017b). Recent advances in omics technologies made possible an accelerated breeding of emerging crops for improved yield, quality and resistance to pests and diseases (Chiurugwi et al. 2019; Ribaut and Ragot 2019). For example, the African Orphan Crops Consortium (AOCC) offered to sequence one hundred accessions of each of 101 priority crops for Africa, including 20 vegetable species, selected based on their importance and potential contribution to food security and livelihoods on the continent (Hendre et al. 2019). Vegetables are an important group of orphan crops requiring more attention. They are affordable and easily available sources of nutrients given their short cycle. They can be cultivated on small plots, requiring low inputs and thus present low capital risks, and provide income to small-scale farmers, especially women (Conti et al. 2019). This article is focused on *Gynandropsis gynandra* (spider plant), one of the most common "orphan" vegetable species in the tropical world and a target species of the AOCC initiative (Hendre et al. 2019).

G. gynandra (2n=34) is a plant species that belongs to the Cleomaceae family (Chweya and Mnzava 1997; Iltis et al. 2011). The plant is a leafy vegetable cultivated in Africa and Asia and also used as a medicinal plant in these regions. Spider plant is generally found to be distributed all over Africa near agricultural land and human settlements but grown in drier zones of South Africa, Zimbabwe, Niger and Burkina Faso (Mnzava and Chigumira Ngwerume 2004; Sogbohossou et al. 2018a).The species is an economically important leafy vegetable in several communities around the world and a source of provitamin A, vitamins C and E, calcium and iron (Sogbohossou et al. 2019; van Den Heever and Venter 2007) and diverse health-promoting compounds including glucosinolates, flavonoids and phenylpropanoids (Neugart et al. 2017; Nyalala and Grout 2015; Omondi et al. 2017b). Despite the wide geographical distribution of the species, *G. gynandra* has for a long time been considered as an "orphan" or "neglected" species because of a lack of research efforts to develop genetic and genomic resources for the crop. Developing genomic resources for *G. gynandra* will open diverse research avenues.

First, the genome of the species is an important resource for breeding programs targeting various traits from higher leaf yield to increased secondary metabolites production and disease resistance.

Second, *G. gynandra* is a C_4 plant and the Cleomaceae family spans both C_3 and C_4 plants (Bayat et al. 2018; Feodorova et al. 2010; Koteyeva et al. 2011; Marshall et al. 2007). *G. gynandra* has been used

in several studies as a C_4 model plant and compared with its sister species *Tarenaya hassleriana*, a C_3 plant for which the draft genome was recently released (Bräutigam et al. 2011; Cheng et al. 2013). Further comparative studies based on the genomes of *G. gynandra* and sister species will shed light on evolution of C_4 photosynthesis in plants.

Third, the Cleomaceae and the Brassicaceae are sister clades and underwent separate whole-genome duplication or triplication events which are thought to have been involved in the birth of novel traits in both families (Schranz and Mitchell-Olds 2006; van den Bergh et al. 2014). So far, evidence of the whole-genome triplication (Th-α) in Cleomaceae was found in *T. hassleriana*, the only species from the Cleomaceae family with a full genome sequence available (Cheng et al. 2013). With the genome of more Cleomaceae species becoming available, the impact of polyploidy on species and trait evolution can be investigated at a broader scale.

The origin of *G. gynandra* is still unknown. Recent phylogeographical analysis of the Cleomaceae suggests that the family originated from Central Asia while the speciation event leading to *G. gynandra* likely occurred in South Africa (Feodorova et al. 2010). However, the study only relied on two *G. gynandra* accessions from Australia and two from South Africa. The natural variation in morphology, vitamin content and leaf secondary metabolites in a germplasm collection of accessions revealed clustering of accessions according to their regions of origin: East/Southern Africa, West Africa and Asia (Sogbohossou et al. 2019; **Chapter 5**).

Here, we present the draft genome of *G. gynandra* and an analysis of whole-genome re-sequencing data of 53 accessions from East/Southern Africa, West Africa and Asia. This study provides the first comprehensive analysis of molecular variation in this important orphan crop. We hypothesize that: (1) the genome of *G. gynandra* bears evidence of the whole-genome triplication Th-α detected originally in the sister species *T. hassleriana*; (2) the genome-wide diversity in the species reveals patterns of geographic differentiation of accessions; (3) there is a congruence between classifications of accessions based on genomic data and metabolome data.

Materials and methods

Draft genome sequencing and assembly

The reference line 'GYN' from Malaysia was inbred by hand-pollination for four generations and then used for draft genome sequencing. A first version of the genome (v1.0) was generated using Illumina sequencing. We constructed 8 different insert-size paired-end (PE) libraries of 250bp, 350bp, 500bp, 800Kb, 2Kb, 5Kb, 10Kb, and 20Kb. After library construction, we used HiSeq2000 to sequence paired ends reads for each library and generated in total 209.57Gb raw data. Low quality reads, reads with adapter sequences and duplicated reads were removed and corrected reads were used to complete assembly of the genome by SOAPdenovo software (Li et al. 2010). In addition, SSPACE software (Boetzer et al. 2010) was used to build scaffolds.

Draft genome annotation

Repeats and transposable elements in the genome were masked with Repeat modeler/repeat masker and RepeatProteinMask (Tarailo-Graovac and Chen 2009). Firstly, *ab initio* prediction program RepeatModeler (version 1.1.0.4) was employed to build the *de novo* repeat library based

on genome, then contamination and multi-copy genes in the library were removed. Using this library as database, RepeatMasker was ran to find homolog repeats in the genome and classify them. Three approaches were used for gene prediction: homology search with closely related species including *Arabidopsis thaliana, A. lyrata, Brassica rapa, Thellungiella parvula* and *Tarenaya hassleriana*; *de novo* prediction using AUGUSTUS(Stanke and Morgenstern 2005), SNAP (Korf 2004) and GlimmerHMM (Majoros et al. 2004); and evidence-based annotation using transcriptomes from 15 different tissues of *G. gynandra*. Consensus gene sets were combined using the program GLEAN (Mackey et al. 2005).

Hybrid genome sequencing and assembly

An improved version of the genome (v2.0) was generated by combining 10X and Illumina sequencing. High molecular weight genomic DNA extraction, sample indexing, and barcoded libraries preparation were performed by 10x Genomics (Pleasanton, CA, USA) according to the Chromium Genome User Guide and as published elsewhere (Weisenfeld et al. 2017). The libraries were sequenced with Illumina HiSeq 2500 with 125 bp PE reads and the raw reads were assembled using the 10X Genomics Supernova software (version 1.0) (Weisenfeld et al. 2017). For scaffolding of the already assembled v1.0 genome, ARCS (Yeo et al. 2017) was used to add barcodes to read identifiers, map reads against the reference genome, use the barcode information to find the reads linking contigs and assemble them in scaffolds.

Hybrid genome annotation

Annotation from the initial genome described above was carried over to the new hybrid 10X/Illumina assembly using flo (same species annotation lift over pipeline - https://github.com/wurmlab/flo). The hybrid genome (v2.0) (fasta and annotation) is available at https://genomevolution.org/coge/GenomeInfo.pl?gid=52885. This last version of the genome was used in all subsequent analyses.

Synteny analysis for detection of ancient polyploidy signals in Cleomaceae

We used the genomes of *T. hassleriana*, *G. gynandra* and *Cleome violacea* available on CoGe (respectively ID 34654; 52885 and 33141) for our analyses. Using the SynMap tool (Lyons et al. 2008) on CoGe, we plotted the self-self synteny dotplots for each species as well as the associated Ks plots coloured according to the value ranges of the detected syntenic gene pairs. To illustrate the effect of ancient polyploidy events on the number of gene copies in each species, we identified regions syntenic to the gene AT5G15840 (CONSTANS) of *Arabidopsis thaliana* in the Cleomaceae species using the SynFind tool in CoGe and syntenic regions were further visualized with the GEvo tool in CoGe (Tang et al. 2015). The CONSTANS gene is involved in regulation of flowering under long days (Yoo et al. 2005).

Whole-genome re-sequencing

A set of 48 accessions of *G. gynandra* from Asia and East/Southern Africa were provided by the World Vegetable Center, Arusha, Tanzania. DNA for resequencing was extracted using Isolate II Plant DNA extraction kit from Bioline Meridian Biosciences, Memphis, TN, USA (https://

www.bioline.com/us/downloads/dl/file/id/879/isolate_ii_plant_dna_kit_product_manual.pdf) . Fresh leaves (~100 mg) were ground in liquid nitrogen and 400 µL lysis buffer was added followed by thorough vortexing. Ten µL of RNase A (1 mg/mL) was added to this mix and incubated at room temperature for 30-45 minutes. The mixture was filtered by centrifuging for 2 min at 11,000 g. To the filtrate, 450 µL of binding buffer was added and mixed by pipetting for 5-8 times. This solution was loaded into a binding column and centrifuged for 1 min at 11,000 g, flow-through was then discarded and the column was washed using 700 µL of wash buffer followed by centrifuging for 1 min at 11,000 g. The DNA was finally eluted in 100 mL TE (Tris-Cl, pH 7.8 and EDTA, pH 8.0).

The library of Ion Proton runs was prepared using Ion Express Plus gDNA Fragment Library Preparation kit (ThermoFisher Scientific, Life Technologies Corporation, Frederick, MD, USA, Catalogue No. 4471269) as follows. Initially 1 µg DNA was fragmented in a reaction volume of 50 µL using 10 µL Ion Shear Plus 10X Reaction Buffer, 10 µL Ion Shear Plus Enzyme Mix II and the remaining amount of deionized water (MilliQ), mixed thoroughly by rapid pipetting and incubated for 15 minutes at 37°C. To stop the reaction, 5 µL of Ion Shear Stop Buffer was added and mixed by continuous vortexing or pipetting for 5-10 seconds and then the reaction mix was stored on ice. The fragments were then purified by adding 99 µL of Agencourt AMPure XP Reagent (Beckman Coulter, Brea, CA, USA, Catalogue No. A63880) incubated at room temperature for 5 minutes followed by separation of beads on a magnetic separator. The clear supernatant was then discarded, and the bead pellet was washed by rinsing with 500 uL of fresh 70% ethanol. The ethanol wash was repeated once again. Finally, the separated bead pellet was air dried at room temperature for 5-10 minutes, finally the fragmented DNA bound to the beads was eluted in 25 µL of TE buffer.

The purified DNA fragments were further ligated to bar coded adapters using Ion Xpress Plus Fragment Library Adapters 1-16 kit (Thermo Fisher Scientific, Life Technologies Corporation, Frederick, MD, USA, Catalogue No. 4471250). In brief it was done as follows. In a reaction volume of 100 µL, 25 µL of the fragment elute was mixed with 10 µL Ligase buffer, 2 µL Ion P1 Adapter, 2 µL of Ion Xpress Barcode X, 2 µL of dNTP mix, 2 µL of DNA ligase, 8 µL of Nick Repair Polymerase and remaining volume adjusted using deionized water (MilliQ) and incubated at 25 ^{0}C for 15 minutes. The reaction was stopped by denaturation at 72 ^{0}C for 5 minutes followed by immediate transfer to ice bath. The ligated fragments were then purified using 120 uL of Agencourt AMPure XP Reagent followed by two washes of 500 uL of 70% ethanol and final elution in 20 µL of TE. The barcoded adapter ligated DNA fragments were then passed through Pippin Prep System (Sage Science, Catalogue No. PIP0001) using Pippin Prep Kit CDF 2010 (Sage Science, Catalogue No. CDF2010) to elute library size of 200 base-read, which had a target base pair setting of 270 bp. After the elution, the library was once again purified using 90 uL of Agencourt AMPure XP Reagent followed by two washes with 500 µL of 70% ethanol and finally eluted in 50 uL of TE. The library was quality checked on Agilent 2100 BioAnalyser system (Agilent Technologies, Santa Clara, CA, USA, Part No. G2939BA) to confirm the library size and Qubit fluorometer (ThermoFisher Scientific, Life Technologies Corporation, Frederick, MD, USA, Catalogue No. Q33226) for quantity. The libraries that passed quality check were loaded into Ion Chef (ThermoFisher Scientific, Life Technologies Corporation, Frederick, MD, USA, Catalogue No. 4484177) using the Ion PI Hi-Q Chef kit

(ThermoFisher Scientific, Life Technologies Corporation, Frederick, MD, USA, Catalogue No. A27198) for library amplification, Ion Sphere Particles (ISP) recovery, enrichment, and chip loading. The ready to use chips were loaded into the Ion Proton System for Next-Generation Sequencing (ThermoFisher Scientific, Life Technologies Corporation, Frederick, MD, USA, Catalogue No. 4476610) for a run length of 200 bases on PI chip using Ion PI Hi-Q Sequencing 200 Kit (ThermoFisher Scientific, Life Technologies Corporation, Frederick, MD, USA, Catalogue No. A26772).

Young leaves were sampled for DNA extraction from five accessions from West Africa grown in Unifarm, Wageningen in May 2019. DNA extraction was performed using a modified CTAB method (Allen et al. 2006). An amount of 1.0 µg DNA per sample was used as input material for the DNA sample preparations. Sequencing libraries were generated using NEBNext® DNA Library Prep Kit following manufacturer's instructions and indices were added to each sample. Genomic DNA was randomly fragmented to a size of 350bp by shearing, then DNA fragments were end polished, A-tailed, and ligated with the NEBNext adapter for Illumina sequencing, and further PCR enriched by P5 and indexed P7 oligos. The PCR products were purified (AMPure XP system) and resulted libraries were analysed for size distribution by Agilent 2100 Bioanalyser and quantified using real-time PCR. Libraries were sequenced using Illumina Novaseq.

SNP calling and annotation

The raw reads generated for each accession were checked for quality and trimmed using Trimmomatic (Bolger et al. 2014). The reads were aligned to the reference genome using the BWA MEM algorithm (Li 2013) and Samtools (Li et al. 2009). Duplicate reads were filtered and read groups added using GATK (version 4.1.2.0). Variant calling and filtering was performed with the HaplotypeCaller, VariantFiltration and SelectVariants tools of GATK (Van der Auwera et al. 2013). Variant quality scores were extracted from the vcf file using the VariantToTable tool and quality score distribution plotted in R to define filtering cutoffs. SNPs distributions and filtering thresholds per quality score are presented in **Supplemental Figure 1**. Moreover, clusters of 3 or more SNPs in 10 bp windows were also filtered. SNP annotation was performed based on genomic locations and predicted coding effects based on the genome annotation using snpEff 4.4 (Cingolani et al. 2012).

Linkage disequilibrium estimation

The pair-wise linkage disequilibrium between genome-wide SNPs based on allele frequencies correlations was estimated and visualized with the software PopLDdecay (Zhang et al. 2018). Only SNPs with pairwise distance less than 1000 kb of each scaffold were considered for the average LD decay plot.

Population structure and genetic differentiation

After pruning for linkage disequilibrium (MAF > 0.05; LD > 0.5) using PLINK2 with a window size of 50 SNPs and a step size of 5 SNPs, population structure was estimated based on the remaining SNPs using ADMIXTURE (v1.23) software (Alexander and Lange 2011). Like the popular software STRUCTURE, ADMIXTURE uses a model-based algorithm to estimate

the ancestry of unrelated individuals. The number of underlying population groups (K) was estimated from 2 to 10 using the maximum likelihood estimation approach with a fast numerical optimization algorithm. The cross-validation method of Alexander and Lange (2011) was used to determine the most likely number of population group (K). The admixture plots were plotted using the R package PopHelper (Francis 2017). A principal coordinate analysis was performed using PLINK (Slifer 2018) and plotted in R. A maximum likelihood tree was computed with SNPhylo (Yang et al. 2018a) to better visualise the genetic distances between our accessions. The genetic differentiation between different populations was assessed by calculating FSt values between populations with VCFtools (Danecek et al. 2011).

Comparison between clustering of accessions based on genomic data and metabolomics data

A subset of 42 accessions were included in a previous study aiming at investigating the natural variation in secondary metabolites in the leaves of *G. gynandra* (**Chapter 5**). Two categories of secondary metabolites were detected: semi-polar metabolites using a liquid-chromatography mass spectrometry platform (LC-MS) and volatile metabolites using a gas chromatography-mass spectrometry (GC-MS) as previously described by Wahyuni et al. (2013).

A distance matrix based on combined, log2 transformed and scaled, LC-MS and GC-MS data was computed in R. The genetic distance between accessions was computed using the package "VCF2Dis" (https://github.com/BGI-shenzhen/VCF2Dis). The correlation between distance matrices was assessed with the mantel test function from the R package "ade4" (Dray and Dufour 2007). Similarities between clustering of accessions with both datasets were compared using the Fowlkes-Mallows index and the baker correlation computed with the R package "dendextend "(Galili 2015). The same package was used to plot dendrograms based on both datasets.

Results

Genome assembly and annotation

An inbred line (GYN) of *G. gynandra* was used for genome sequencing combining Illumina and 10X platforms. Compared with sister species from the Cleomaceae family (*T. hassleriana*: ~290 Mb; *C. violacea*: ~280 Mb), the genome of *G. gynandra* is relatively large (~750 Mb, 2n=34). While the first version of the genome (v1.0) had 3219 scaffolds with an N50 of 447.2 kbp, the improved genome (v2.0) had 1693 scaffolds with an N50 of 1.4 Mbp (**Table 1**).

Integration of the various gene prediction approaches resulted in 27,154 well supported gene models classified in 15,955 gene families with an average transcript length of 2,137 bp and on average 5 exons per gene. Of the identified gene families, 79% were shared with *T. hassleriana* and the Brassicaceae *A. thaliana* and *B. rapa*, 14% were shared only with *T. hassleriana* and 7% were unique to *G. gynandra* **(Figure 1)**. A total of 87.7% of gene models matched with conserved motifs or homologs in at least one of the public protein databases including 67.3% in Swissprot, 67.9% in InterPro, 86.8% in TrEMBL, 52% in KEGG and 50.9% in GO.

Table 1. Summary statistics of the genome assembly and annotation

Genome version	v1.0	v2.0
Assembly		
Number of scaffolds	3219	1693
Genome size	739,804,406	739,819,666
Longest scaffold	3,513,483	5,871,673
Shortest scaffold	180	180
Average scaffold size	229,824.29	436,987.39
Median scaffold size	135,052	100,592
N50 of scaffolds	447,194	1,376,588
GC content	33.94	33.94
Annotation		
Genes	27154	27154
CDS	196,140	195,914
Exon	185,218	184,983
Five-prime UTR	25812	25791
mRNA	34701	34383
Three-prime UTR	22703	22680

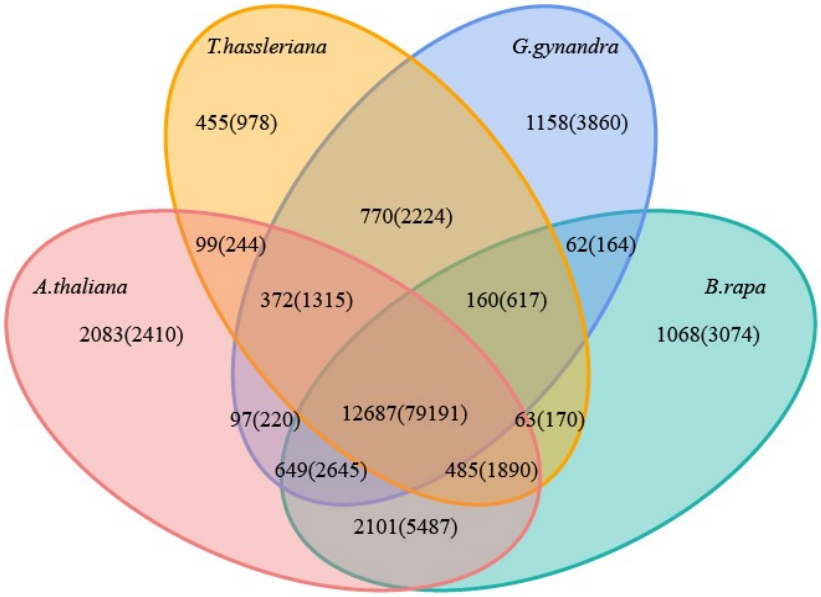


Figure 1. Venn diagram illustrating the shared and unique gene families from *Gynandropsis gynandra, Tarenaya hassleriana* **(Cleomaceae),** *Arabidopsis thaliana* **and** *Brassica rapa* **(Brassicaceae). The number of genes in shared gene families are in parentheses.**

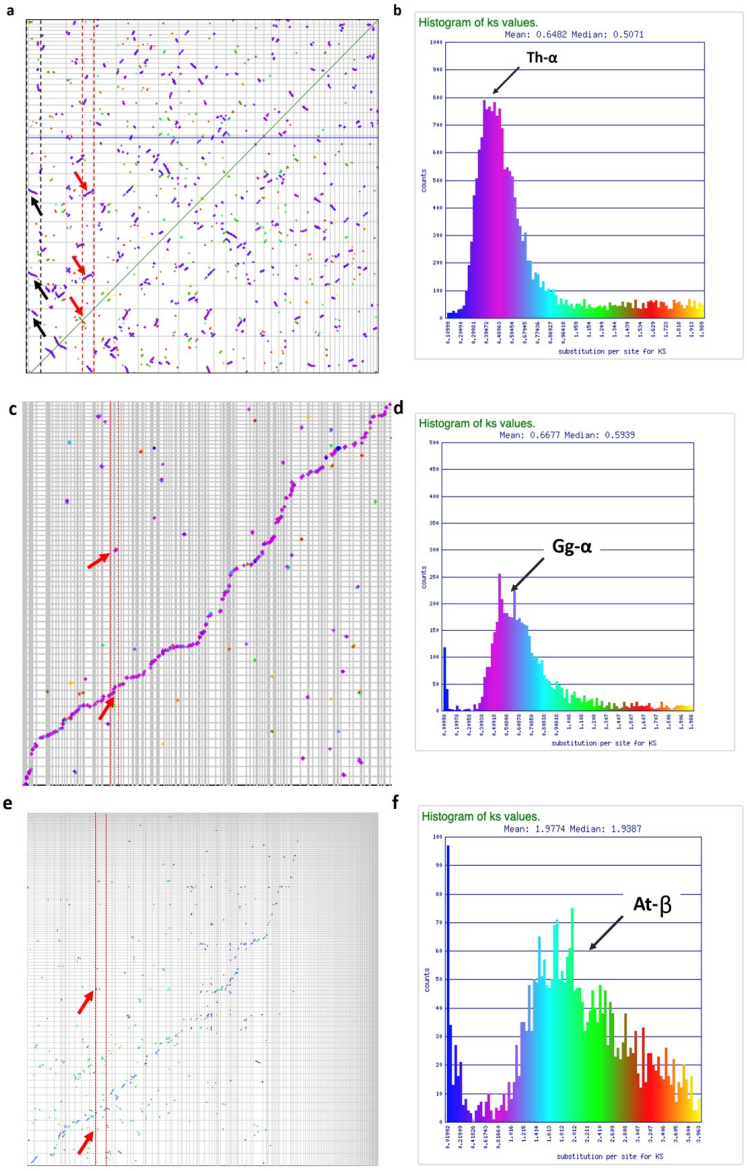


Figure 2. Polyploidy events in *Tarenaya hassleriana*, *Gynandropsis gynandra* **and** *Cleome violacea*. (a) Syntenic dotplot of self–self *Tarenaya hassleriana* genome comparison. Horizontal and vertical gray lines separate scaffolds. Coloured dots are syntenic gene pairs. Red and black dashed lines band and red and black arrows point to 3:3 syntenic signals. (b) Histogram of Ks values of syntenic gene pairs in *Tarenaya hassleriana*. Ks values are coloured by range and the same colours apply to syntenic gene pairs on the dotplots (c) Syntenic dotplot of self–self *Gynandropsis gynandra* genome comparison. Horizontal and vertical gray lines separate scaffolds. Coloured dots are syntenic gene pairs identified through collinearity. Red dashed lines band and red arrows point to an example of 2:2 syntenic signals. (d) Histogram of Ks values of syntenic gene pairs in *Gynandropsis gynandra*. Ks values are coloured by range and the same colours apply to syntenic gene pairs on the dotplots (e) Syntenic dotplot of self–self *Cleome violacea* genome comparison. Coloured dots are syntenic gene pairs identified through collinearity. Red dashed lines band and red arrows point to an example of 2:2 syntenic signals. (f) Histogram of Ks values of syntenic gene pairs in *Cleome violacea*. Ks values are coloured by range and the same colours apply to syntenic gene pairs on the dotplots.

Evidence of a whole-genome duplication event in *Gynandropsis gynandra*

Based on the version v2.0 of the genome, we generated synteny dotplots to determine whether the *Th-a* whole-genome triplication event that occurred in *T. hassleriana* was shared with *G. gynandra*. We included *Cleome violacea*, another species from the Cleomaceae family for which whole-genome sequence was available, in the analysis. The *Th-a* triplication was detected with a synteny dotplot based on the self-comparison of the *T. hassleriana* genome with three pairs of aligned genes in the same regions **(Figure 2a)**.

Our results showed however, that there is no evidence of a triplication in *G. gynandra* but rather of a duplication that we called *Gg-a* **(Figure 2c)** although the event occurred at the same time as *T. hassleriana* (similar Ks values) **(Figure 2b,d)**. Comparison of Ks plots in *T. hassleriana* and *G. gynandra* reveals a larger peak in *T. hassleriana* which suggests that an additional genome was added to the species after the *Gg-a* shared by both species. In *C. violacea*, gene pairs resulting from the more ancient *At-β* duplication were detected **(Figure 2e,f)**. Elucidating ancient polyploidy events in sister species of *G. gynandra* allows a better understanding of evolutionary relationships between these species. Such information can facilitate translational genomics between *G. gynandra* and well-studied plants such as *Brassica* crops and *A. thaliana*. The phylogenetic relationships between Brassicaceae and Cleomaceae and the polyploidy events which occured in both lineages are presented in **Figure 3.** The synteny analysis between *A. thaliana* and our three Cleomaceae genomes for the CONSTANS gene AT5G15840 revealed two syntenic regions in *C. violacea,* two in *G. gynandra* and three in *T. hassleriana* (**Figure 4)**.

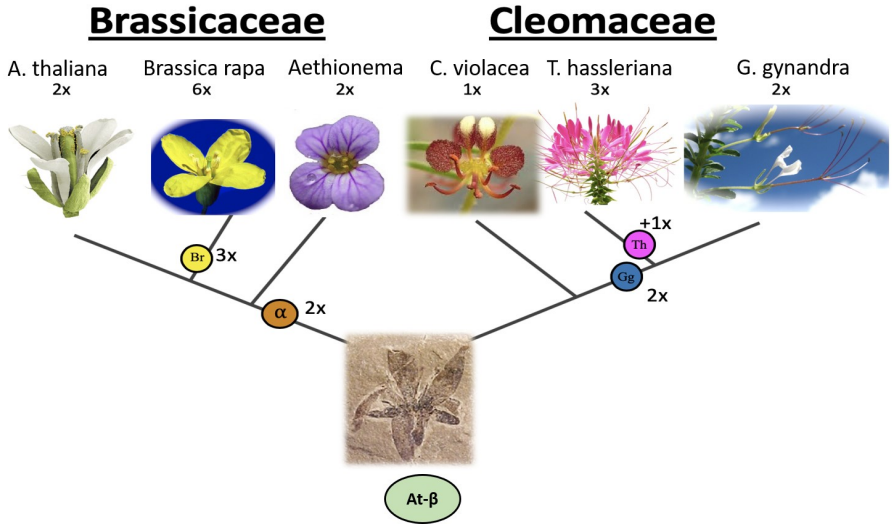


Figure 3. Phylogenetic relationships between Brassicaceae and Cleomaceae species/genera and ancient polyploidy events detected in both lineages: the At-β (in green) shared by Brassicaceae and Cleomaceae; the At-α shared by all Brassicaceae in brown, the Br-α event (in yellow) in *Brassica* spp.; in Cleomaceae, the Gg-α shared by *G. gynandra* and *T. hassleriana* (in blue) and a potential genome addition (in pink) in *T. hassleriana* explaining the Th-α triplication observed in the species.

A

Organism	Genome	Type	Name	Chr	Synteny Score	SynMap
Arabidopsis thaliana	(vTAIR10 unmasked)	query	AT5G15840.1.TAIR10	5	0	Dotplot
Cleome violacea	(v1 unmasked)	syntelog	Clevi.0012s0752.1.v1.1	scaffold_12	27	Dotplot
Cleome violacea	(v1 unmasked)	proxy for region	pos 1136081	scaffold_42	4	Dotplot
Gynandropsis gynandra	(vv4_10X_included unmasked)	syntelog	Cg01801.1	scaffold32	20	Dotplot
Gynandropsis gynandra	(vv4_10X_included unmasked)	syntelog	Cg03703.1	scaffold86	14	Dotplot
Tarenaya hassleriana	(v5.1, WeberBrautigam corrected unmasked)	proxy for region	pos 972007	scaffold00030	18	Dotplot
Tarenaya hassleriana	(v5.1, WeberBrautigam corrected unmasked)	proxy for region	pos 7489111	scaffold00006	9	Dotplot
Tarenaya hassleriana	(v5.1, WeberBrautigam corrected unmasked)	syntelog	Th2v18633	scaffold00015	7	Dotplot

B

Figure 4. SynFind analysis of CONSTANS gene (AT5G15840) across *A. thaliana, C. violacea, G. gynandra* **and** *T. hassleriana*. **(A) SynFind table output illustrating eight matching regions in the selected species. Result can be regenerated: https://genomevolution.org/r/190df. (B) GEvo visualization of the compiled syntenic regions. Each panel represents a syntenic region in** *A. thaliana, C. violacea, G. gynandra* **and** *T. hassleriana*, **from top to bottom.** Green arrows in each panel represent gene models, and boxes on top of the gene models are sequence matches. For the top A. thaliana panel, there are seven tracks of sequence matches, two for *C.violacea*, two for *G. gynandra* and three for *T. hassleriana,* respectively. Result can be regenerated: https://genomevolution.org/r/190dp.

Variant calling and annotation in the *Gynandropsis gynandra* collection

We re-sequenced the genomes of 53 accessions of *G. gynandra* including 24 accessions from Asia, 24 accessions from East/Southern Africa and five from West Africa, the most important regions of diversity of the species. The average sequencing coverage depth per sample (**Table 2**) was 6x when aligned to the reference genome (v2.0) (https://genomevolution.org/coge/GenomeInfo.pl?gid=52885). The variant calling procedure yielded a total of 29,537,715 SNPs. After filtering SNPs for quality and removing SNP clusters of 3 or more SNPs in 10 bp windows, 10,874,872 SNPs were left and distributed over 1473 of the 1693 scaffolds of the genome, corresponding on average to 15 SNPs per kb. About 54.8% of the SNPs were in intergenic regions, 31.5% in exons, 11.2% in introns, 2.5% in UTRs and 0.1% in splice sites. Coding SNPS included 57.3% synonymous, 42.0% missense and 0.7% nonsense.

Table 2. Accessions, country and region of origin and genome coverage.

Lab ID	Accession	Country	Region	Genome Size	Output per accession	Coverage X
1	TOT3527	Laos	Asia	0.7	2.3	3.3
2	TOT3514	Laos	Asia	0.7	2.3	3.3
3	TOT6441	South Africa	East/Southern Africa	0.7	2.3	3.3
4	TOT6439	Zambia	East/Southern Africa	0.7	1.4	2.0
5	TOT6435	Kenya	East/Southern Africa	0.7	18.6	26.6
6	TOT6421	Tanzania	East/Southern Africa	0.7	3.7	5.3
7	TOT5799	Thailand	Asia	0.7	1.1	1.6
8	TOT3534	Laos	Asia	0.7	3.3	4.7
9	TOT1480	Thailand	Asia	0.7	2.8	4.0
10	TOT1048	Thailand	Asia	0.7	3.6	5.1
11	TOT6440	South Africa	East/Southern Africa	0.7	3.1	4.4
12	TOT6422	South Africa	East/Southern Africa	0.7	3.3	4.7
13	TOT6420	Tanzania	East/Southern Africa	0.7	2.9	4.1
14	TOT4976	Thailand	Asia	0.7	3.2	4.6
15	TOT4937	Thailand	Asia	0.7	1.3	1.9
16	TOT4935	Thailand	Asia	0.7	1.3	1.9
17	TOT3536	Laos	Asia	0.7	1.3	1.9
18	TOT4447	Bangladesh	Asia	0.7	2.4	3.4
19	TOT4728	Bangladesh	Asia	0.7	1.9	2.7
20	TOT4489	Bangladesh	Asia	0.7	1.18	1.7
21	TOT8917	Malawi	East/Southern Africa	0.7	3	4.3
22	TOT8916	Malawi	East/Southern Africa	0.7	3.9	5.6
23	TOT8915	Malawi	East/Southern Africa	0.7	3.4	4.9
24	TOT8892	Uganda	East/Southern Africa	0.7	2.7	3.9
25	TOT8891	Uganda	East/Southern Africa	0.7	4.2	6.0
26	TOT8890	Uganda	East/Southern Africa	0.7	4	5.7
27	TOT8889	Uganda	East/Southern Africa	0.7	2.4	3.4
28	TOT8888	Uganda	East/Southern Africa	0.7	2.9	4.1
29	TOT8887	Uganda	East/Southern Africa	0.7	2.9	4.1
30	TOT7505	Laos	Asia	0.7	3.4	4.9
31	TOT8918	Malawi	East/Southern Africa	0.7	4.4	6.3
32	TOT8925	Kenya	East/Southern Africa	0.7	4.1	5.9
33	TOT8926	Kenya	East/Southern Africa	0.7	3.9	5.6
34	TOT8931	South Africa	East/Southern Africa	0.7	2.6	3.7
35	TOT8933	Zambia	East/Southern Africa	0.7	3.3	4.7
36	TOT8996	Taiwan	Asia	0.7	2.6	3.7
37	RW-SF-10	Rwanda	East/Southern Africa	0.7	2.3	3.3
38	RW-SF-04	Rwanda	East/Southern Africa	0.7	2.3	3.3
39	TOT7197	Malaysia	Asia	0.7	1.7	2.4
40	TOT7198	Malaysia	Asia	0.7	3.6	5.1
41	TOT7199	Malaysia	Asia	0.7	2.7	9.0
42	TOT7200	Malaysia	Asia	0.7	2.9	9.7
43	TOT7441	Laos	Asia	0.7	3.1	4.4
44	TOT7449	Laos	Asia	0.7	3.5	5.0
45	TOT7462	Laos	Asia	0.7	4.5	6.4
46	TOT7486	Laos	Asia	0.7	10.1	14.4
47	TOT6442	South Africa	East/Southern Africa	0.7	3.8	5.4

48	TOT7196	Malaysia	Asia	0.7	1.3	1.9
49	ODS-15-020	Benin	West Africa	0.7	13.3	19.0
50	ODS-15-044	Benin	West Africa	0.7	18.7	26.7
51	ODS-15-045	Benin	West Africa	0.7	16.3	23.3
52	ODS-15-061	Togo	West Africa	0.7	16.9	24.1
53	ODS-15-117	Togo	West Africa	0.7	16.1	23.0

Linkage disequilibrium decay

Linkage disequilibrium was estimated between all SNP markers for the 53 accessions and was plotted per region of origin of the accessions. The East/Southern African population group showed lower average maximum LD (0.62) than the West African (0.78) and Asian (0.9), and the LD decayed more rapidly in the Asian and East/Southern African populations. The squared correlations of allele frequencies r^2 of the West African accessions decreased to half of the maximum value at approximately 200 kb physical distance, in contrast in the East/Southern African and Asian accessions this had occurred by 50 kb **(Figure 5).**

Population structure

After pruning SNPs for linkage disequilibrium based on correlation threshold of 0.5 in SNP windows of 50 SNPs and a step of 5 SNPs, 752,305 SNPs were used to examine the relationships between the 53 accessions. We combined three different complementary approaches: principal component analysis, neighbour-joining tree building, and admixture analysis. Principal component analysis separated the three populations. The first dimension explained 24% of variation and partially separated accessions from East/Southern Africa, and West Africa and Asian accessions.

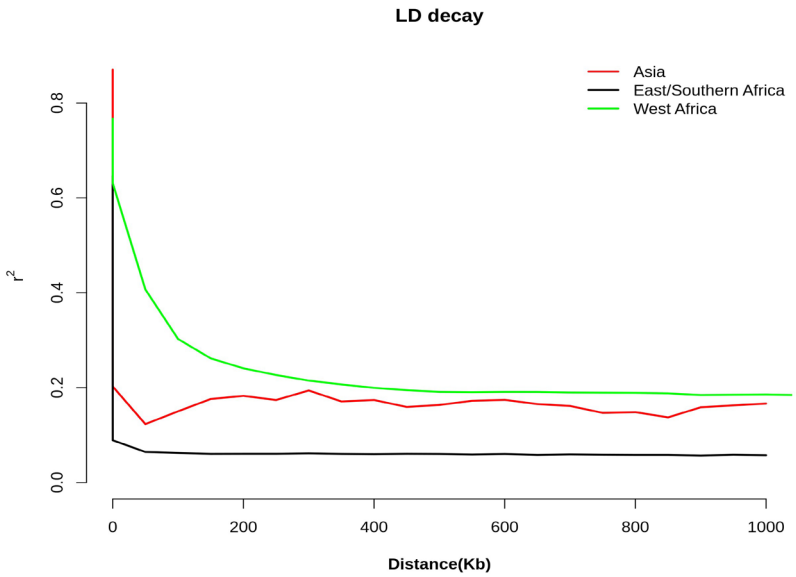


Figure 5. Average linkage disequilibrium decay (r^2) over pair-wise distance in West African, East/Southern African and Asian accessions of *Gynandropsis gynandra*.

The second dimension (7.6% variation explained) separated West African accessions from East/Southern African and Asian ones (**Figure 6a**) while the third dimension (4.1% variation explained) clustered together accessions from West Africa and Asia (**Figure 6b**). The neighbour-joining tree (**Figure 6c**) also revealed clear patterns of clustering with accessions from West Africa and Asia more closely related with each other than accessions from East/Southern Africa. Within Asian and West African clusters, branches were shorter compared with the East/Southern African cluster, suggesting more genetic diversity in this latter cluster. Results from ADMIXTURE corroborated results obtained from the previous two approaches. The optimal number of clusters was three **(Figure 6d; Supplemental Figure 2).**

Five accessions from Asia, TOT3534 (Laos), TOT3514 (Laos), TOT3527 (Laos), TOT4447 (Bangladesh) and TOT7486 (Laos) had the same genetic background as East/Southern African

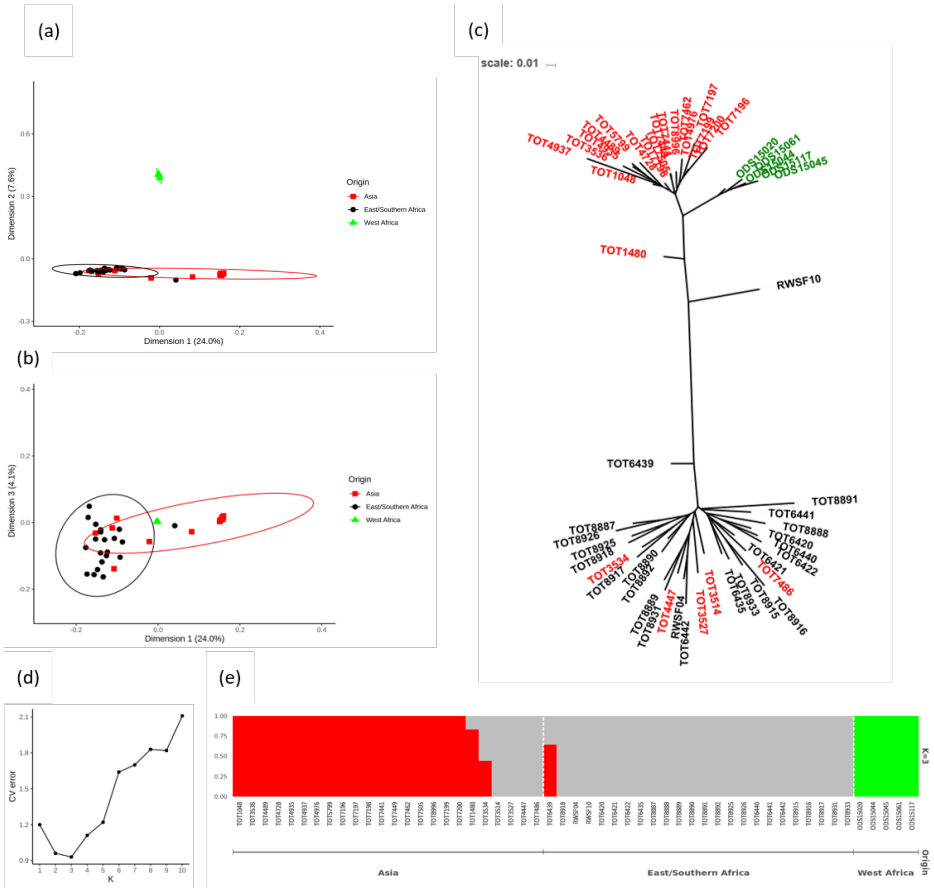


Figure 6. Population structure in 53 accessions of *Gynandropsis gynandra***.** (a) Projection of accessions on dimensions 1 (24.0%) and 2 (7.6%) resulting from principal component analysis. (b) Projection of accessions on dimensions 1 (24.0%) and 3 (4.1%) resulting from principal component analysis. (c) Unrooted maximum likelihood tree showing the distances between accessions. Asian accessions are in red, East/Southern African ones in black and West African ones in green. (d) Cross-validation error plot for admixture plots for K values from 1 to 10. (e) Admixture plot (K=3) showing the proportions of admixture in the 53 accessions. Asian, East/Southern African and West African accessions are separated by a dotted white line.

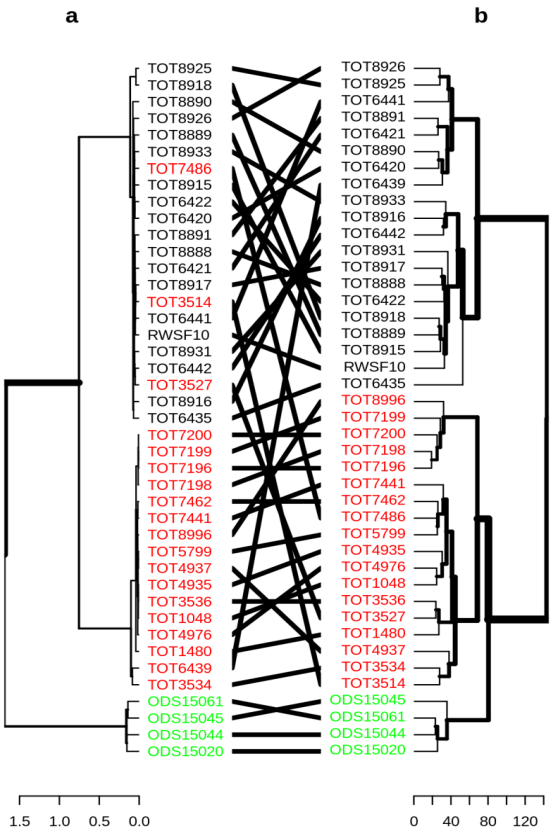


Figure 7. Dendrograms of 42 accessions of Gynandropsis gynandra based on (a) genetic distances computed with 752305 SNPs; (b) distance matrix of metabolite levels in the leaves of the accessions from both LC-MS and GC-MS analyses. Accessions from East Africa are in black, accessions from Asia in red and accessions from West Africa in green. The thickness of the branches denotes the level of confidence in the clustering.

accessions, as also shown on the PCA plots and the neighbor-joining tree (**Figures 6 a-e**). One accession from East/Southern Africa, TOT6439, had the same admixture group with accessions from Asia. West African accessions were in a separate cluster **(Figure 6e)**.

The genetic differentiation between populations revealed that there was a higher genetic differentiation between accessions from West Africa and Asian populations (0.77) than between East/Southern African and West African on one hand (0.58) and East/Southern African and Asian (0.51) on the other. The high Fst values between all the populations suggests an ancient split between these populations and a high differentiation between them. East/Southern African accessions had a higher genetic diversity (π=0.07) than Asian ones (π=0.02) while surprisingly, the highest genetic diversity was observed in West Africa accessions (π=0.15) for this region despite the low sample size (5 accessions com+pared with 24 in East/Southern Africa and 24 in Asia).

Comparison of classification of accessions based on genome and metabolome data

We compared dendrograms based on genetic distance between accessions and dissimilarity between accessions based on secondary metabolites data generated previously for 42 of the 53 accessions (**Chapter 5**). Different tests were applied to assess the level of relatedness between the two the datasets. The Mantel test applied on both distance matrices revealed a positive and significant correlation between them (r^2=0.28; p=0.001). The tests of Fowlkes-Mallow index (FM index=0.81; p<0.01) and the Baker correlation between dendrograms (r^2=0.52) supported the existence of similarities between clustering of accessions based on both datasets. However, there were differences in the topologies of the dendrograms resulting from both datasets **(Figure 7).** The genomic data clustered West African accessions separately from East/Southern African and Asian accessions (**Figure 7a**) while the metabolite data clustered together accessions from West Africa and Asia (**Figure 7b**). Moreover, three Asian accessions were clustered with East/Southern African accessions as previously shown by the population structure analyses even though the metabolite data clustered them with the other Asian accessions.

Discussion

Whole-genome sequencing, especially of orphan crops, provides new perspectives of genome evolution, population evolution, trait genetics, and genic information, which can be applied to develop modern and efficient breeding programs. The present study presents the genome of the orphan leafy vegetable *G. gynandra* and the assessment of the genetic diversity and population structure in a set of 53 re-sequenced accessions of this species.

Signature of a whole-genome duplication event in *Gynandropsis gynandra*

In the Cleomaceae family, evidence of an ancient genome triplication (*Th-α*) was previously found in *T. hassleriana*, a sister species of *G. gynandra* (Cheng et al. 2013). The triplication was found to be independent of the Brassicaceae-specific duplication (At-α) and nested *Brassica* triplication (Br-α) (Cheng et al. 2013; Schranz and Mitchell-Olds 2006). In the absence of multiple genome sequences for Cleomaceae species, it was impossible to adequately place this putative Cleomaceae-specific polyploidy event. However, previous transcriptome analyses in *T. hassleriana* and *G. gynandra* suggested that both species shared the *Th-α* polyploidy event although surprisingly, the C_4 *G. gynandra* had 16% less C_4 genes than the C_3 *T. hassleriana* (van den Bergh et al. 2014). Our synteny analysis on the genome of *G. gynandra* here presented and the draft genome of *C. violacea* (Edger et al. in prep) revealed that the Th- α triplication is not present in *C. violacea* and appears as a duplication event in *G. gynandra* (Gg-α). Based on our observations, we hypothesize that both *G. gynandra* and *T. hassleriana* first underwent the same the *Gg- α* and that *T. hassleriana* subsequently acquired an additional genome copy. As new genome sequences become available for the Cleomaceae, it will be possible to clarify the evolutionary history of the family.

Genetic diversity, population structure and origin of *Gynandropsis gynandra*

The natural variation in morphology, vitamin content and leaf secondary metabolites production in *G. gynandra* accessions from East/Southern Africa, West Africa and Asia revealed

geographic patterns of differentiation (Sogbohossou et al. 2019; **Chapter 5**). Re-sequencing of 53 accessions from those three regions of diversity of the species confirmed a high level of genetic differentiation between accessions driven primarily but not exclusively by geography. The low level of genetic diversity in Asian accessions and the presence of Asian samples with an East/Southern African background suggested a relatively recent flow of accessions from East/Southern Africa to Asia. The slower linkage disequilibrium decay in Asian and West African accessions could be an indicator of a strong founder effect in these populations, whereby a small number of accessions from East/Southern Africa contributed to the establishment of West African and Asian populations. This hypothesis is further supported by the low Fst values between East/Southern African accessions and both West African and Asian accessions and the high Fst value between Asian and West African accessions. The lower extent of LD in the East/Southern African accessions could be an indicator of a higher rate of recombination due to the older age, larger population size and greater diversity in that population. Our findings corroborate the biogeographical study of Cleomaceae which considered South Africa as the origin of the speciation event of *G. gynandra* (Feodorova et al. 2010). A similar pattern of out-of-Africa migration was found in *A. thaliana* (Durvasula et al. 2017).

Genome and metabolome congruence for accessions clustering

The existence of patterns of geographic differences between accessions was established with both genomic and metabolome data. Apart of the presence of three Asian accessions in the East/Southern African clade, the high level of similarity between the clustering results provides an indication that genome-wide association studies for secondary metabolites, morphological traits and vitamin content could be highly prone to population structure biases. Correcting for population structure or using Random Forest for selection of markers associated with traits of interest has proven efficient in genome-wide association studies in populations of *Brassica rapa* with high levels of population structure (Pino Del Carpio et al. 2011). The use of linkage mapping would be an efficient approach which will allow us to improve the current genome assembly and to map genes involved in traits of interest including leaf-yield related traits and secondary metabolites (e.g. carotenoids, tocopherols, glucosinolates, flavonoids) production.

The fact that *G. gynandra* always occurs near human settlements makes it hard to clearly distinguish cultivated populations from wild ones, which sometimes, escaped from earlier cultivation (Chweya and Mnzava 1997; Sogbohossou et al. 2018b). The maximum likelihood phylogenetic tree was consistent with the morphological and metabolic data which suggested that accessions from West Africa and Asia were more similar to each other than they were to the East/Southern African ones. Moreover, we also noticed during several crosses in full diallel designs between accessions that reproductive barriers are present between accessions from the different regions. For example, reciprocal crosses between Asian and West African accessions resulted in fertile F_1 populations. In contrast, crosses with East/Southern African accessions as maternal parent and West African or Asian accessions resulted in sterile F_1 while the opposite resulted in fertile F_1 populations and F_2 populations with high levels of male sterility (reduced or aborted stamens). A similar case of unilateral intraspecific incompatibility was reported between *Brassica rapa* accessions from Turkey and Japan and was due to recognition between a

pistil receptor SUI1 and a pollen ligand PUI1, two cryptic incompatibility genes (Takada et al. 2017). Overall the tremendous genetic diversity in *G. gynandra* in Africa provides an opportunity for the development of improved cultivars for different markets/environments.

Conclusions

The genome presented here provides a better understanding of the polyploidy history in the Cleomaceae and sheds light into the possible scenarios of step-wise ancient polyploidy events of *T. hassleriana* and *G. gynandra*. A set of 10.8 million SNPs developed based on the 53 resequenced accessions provides a valuable resource for genotyping. We also achieved a better understanding regarding population structure of germplasm and suggest that the East/Southern African region is the center of origin of *G. gynandra* with subsequent dispersion in West Africa and Asia. Moreover, the consistency between the genome and the metabolome data confirms a previous hypothesis that the variation in the metabolome of the species is highly driven by genetic differences. Further studies with more extensive sets of accessions are required to elucidate the origin and migration routes of the species. The high genetic differentiation between regions combined with the observations of reproductive incompatibilities between accessions from different regions suggests an incipient speciation event in the species. Investigating the existence of structural variation in the genome and potential cross-incompatibility genes in accessions from the East/Southern African accessions vs. West African/Asian accessions could help clarifying the molecular mechanisms driving our observations.

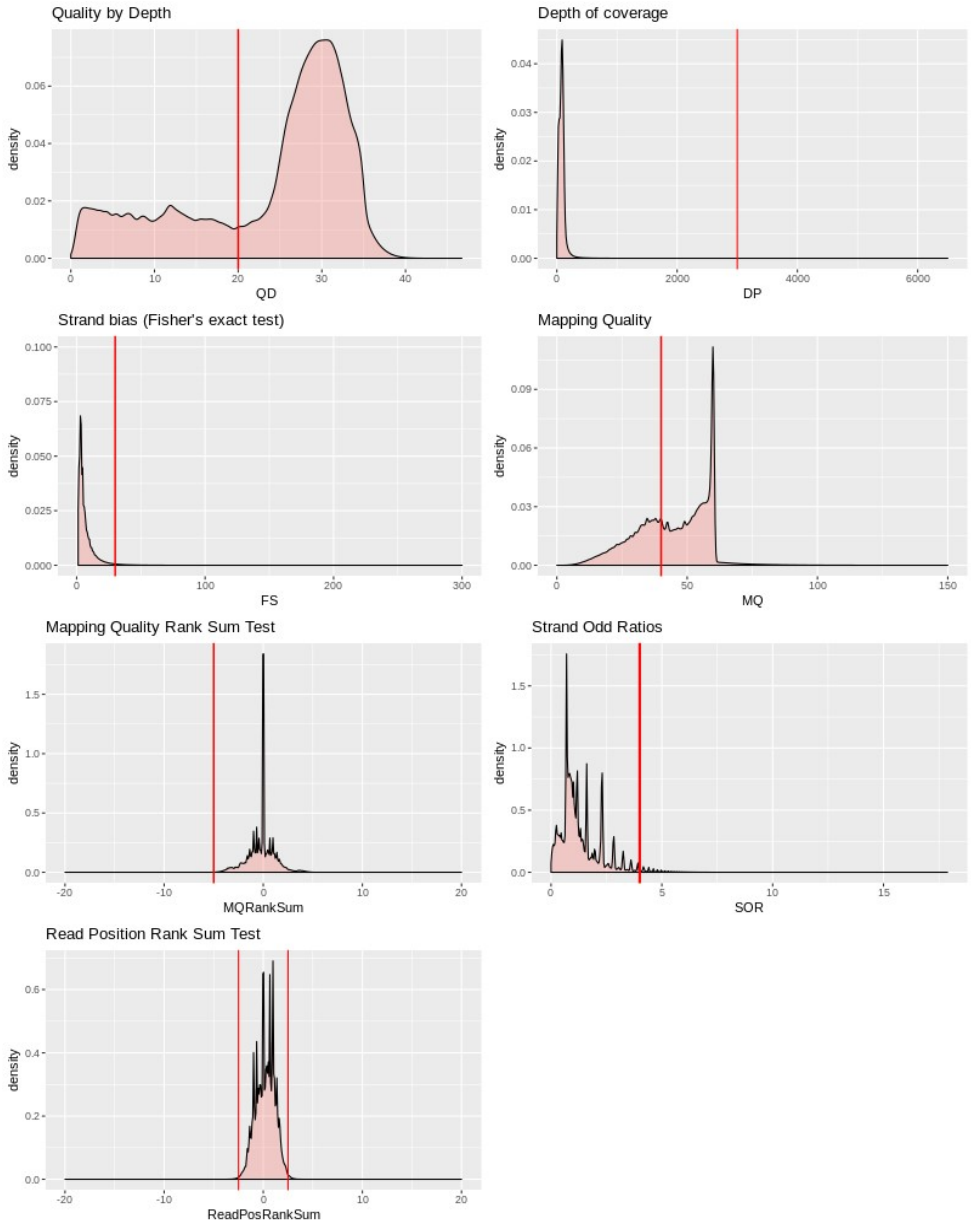
Acknowledgements

We thank Andreas Ebert from the World Vegetable Center for providing the seeds of the accessions from East/Southern Africa and Asia. This work was supported by the African Orphan Crops Consortium (AOCC), the Applied Research Fund of the Netherlands Organization for Science under the Project "Utilizing the genome of the vegetable species *Cleome gynandra* for the development of improved cultivars for the West and East African markets" (Project Number: W.08.270.350) and the Schlumberger Foundation Faculty for the Future Fellowship awarded to E.O.D.S. The authors acknowledge Thermo Fisher Scientific, Life Technologies Corporation, Frederick, MD, USA, for donation of Ion Protons and Ion Chefs to the AOCC genomics lab, ICRAF, Nairobi. A.P.M.W. was funded through the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC-2048/1 – Project ID: 390686111, and ERA-CAPS project C4BREED (WE 2231/20-1).

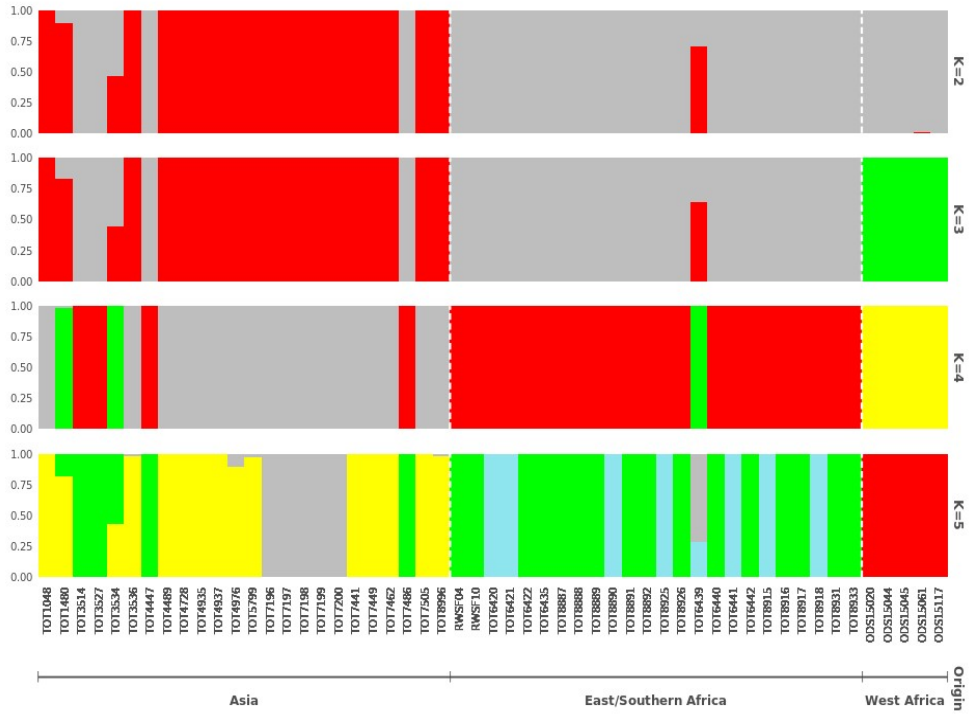
Supplemental material

Supplemental Figure 1. SNP distribution per quality parameter.

Supplemental Figure 2. Population structure for K values from 2 to 5 of 53 accessions of *Gynandropsis gynandra* from Asia, East/Southern Africa and West Africa.



Supplemental Figure 1. SNP distribution per quality parameter. Filter thresholds are denoted by red lines.



Supplemental Figure 2. Population structure for K values from 2 to 5 of 53 accessions of *Gynandropsis gynandra* from Asia, East/Southern Africa and West Africa.

CHAPTER 7

Mapping Quantitative Trait Loci controlling vitamin content, flowering time and morphological traits in the orphan leafy vegetable *Gynandropsis gynandra* (Cleomaceae)

E. O. Dêêdi Sogbohossou[1,2], Frank Becker[1], Prasad Hendre[3], Enoch G. Achigan-Dako[2], Allen Van Deynze[3], Julian Hibberd[4], Ric C. H. de Vos [5], M. Eric Schranz[1]

[1] Biosystematics Group, Wageningen University, Droevendaalsesteeg 1, 6708PB, Wageningen, Netherlands

[2] Laboratory of Genetics, Horticulture and Seed Science, Faculty of Agronomic Sciences, University of Abomey-Calavi, BP 2549 Abomey-Calavi, Republic of Benin

[3] Department of Plant Sciences, University of California, Davis, USA, 95616

[4] Department of Plant Sciences, Cambridge University, Cambridge CB2 3EA, United Kingdom

[5] Bioscience, Wageningen Plant Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.

Abstract

Spider plant (*Gynandropsis gynandra)* is a leafy vegetable valued in Africa and Asia for its nutritional and medicinal properties. While the species has recently become a model to study C_4 photosynthesis, less is known about the genetic basis controlling traits of interest for farmers and consumers such as yield and nutrition. There is extensive variation in morphology and metabolic profiles in the broad germplasm collection that can be utilized for breeding. The objective of this study was to construct the first genetic map in *G. gynandra* and identify SNPs associated with leaf-yield related traits and nutrient content in the species. An F_2 population derived from a cross between TOT7200 from Malaysia and TOT8917 from Malawi, two accessions with contrasting morphological features and metabolite profiles, was phenotyped for plant height, leaf area, flowering time, and for levels of carotenoids, tocopherols and ascorbic acid. A set of 1309 SNPs selected based on the re-sequenced genomes of the parental lines was used for targeted genotyping-by-sequencing in the F_2 population. The genetic map was constructed with a final set of 269 segregating SNPs. A total of twelve quantitative trait loci were identified for leaf area, plant height, flowering time, carotenoid content, tocopherol content and ascorbic acid content. QTLs with pleiotropic effects were identified on linkage groups 3, 7, 9 and 16. Candidate genes identified based on homology with Arabidopsis included genes for flowering time (*FTIP1, CDKG2, MRF1, AGL24*), carotenoid biosynthesis (*CRTISO, CYP97B3*), tocopherol biosynthesis (*ABC1K1, CHLP, TAT1, VTE2*), leaf area (*DOT3*) and plant height (*AGL24*). No obvious candidate gene for ascorbic acid content was identified. Further validation of the QTLs will facilitate marker-assisted breeding for higher leaf-yield and nutrient content in the species.

Key-words: Breeding; genetic map; nutrient content; leaf yield; spider plant

Introduction

There are thousands of crops which have been used for centuries at local or regional scales as sources of food, fibre, fodder, gums & resins, medicine, housing and which continue to provide livelihood options to farmers and communities (Hendre et al. 2019). Many of these crops have slowly faded from traditional and commercial farming landscapes due to the rise of monoculture-based high-input agriculture and are not included in the mainstream research agenda (Tadele 2019). Such crops are thus are called orphan or neglected crops. However, they continue to contribute to maintaining agrobiodiversity and provide livelihood options to smallholder and subsistence farmers. Recent advances in -omics combined with an increasing need to promote crop genetic diversity to mitigate climate change effects and attain food security led to a revived interest in these neglected species. In Africa, leafy vegetables are important components of local diets and provide nutrients and alternative sources of income for local farmers, particularly women (Gowele et al. 2019; Olabode et al. 2017). Scaling up cultivation and developing value chains for these species requires research in the development of new cultivars meeting consumers' expectations.

Among the traditional leafy vegetables with rejuvenated interest in Africa, spider plant (*Gynandropsis gynandra*) (2n = 2x =34) stands out as source of health-promoting compounds (Omondi et al. 2017; Sogbohossou et al. 2018; Sogbohossou et al. 2019) but also as a model crop to study C_4 photosynthesis (Brown et al. 2005; Marshall et al. 2007). The species belongs to the Cleomaceae, the sister family of the Brassicaceae, and genetic studies of the crop could therefore make use of comparative genomics approaches with well-studied *Arabidopsis* and *Brassica* crops. The germplasm of *G. gynandra* exhibits significant variation in the amounts of minerals, vitamins and secondary metabolites such as glucosinolates and flavonoids (Neugart et al. 2017; Omondi et al. 2017b; Sogbohossou et al. 2019). Current -omics resources available for the species include a transcriptome atlas (Külahoglu et al. 2014), the draft reference genome of the species (**Chapter 6**), metabolomics data and resequencing data of 53 diverse accessions (**Chapters 5 and 6**) as well as a set of Simple Sequence Repeat (SSR) markers developed and validated by Omondi et al. (2017a). Recent studies on a worldwide germplasm of the species revealed geographic patterns of variation at the morphological level but also in terms of vitamins, secondary metabolites and photosynthetic efficiency (Reeves et al. 2018; Sogbohossou et al. 2019). The collection exhibited a two-fold variation in carotenoid content, a 20-fold variation in tocopherol content and a nine-fold variation in ascorbic acid content.

Carotenoids carrying at least one unsubstituted β-ionone ring, such as α- and β-carotene, are vitamin A precursors in humans. Lutein, the most abundant carotenoid in plant leaf tissues along with zeaxanthin are concentrated in the human retina where they enhance visual performance and protect eyes from diseases as age-related macular degeneration (Cazzonelli 2011; Esteban et al. 2015). Both vitamins C and E are increasingly used in cosmetics as photoprotectors to mitigate skin damage due to exposition to UV light (Aguilera et al. 2012).

In plants, carotenoids are pigments present in plastids where they are involved in photosynthesis, photoprotection, and attraction of seed dispersers, pollinators or mutualist consumers (Esteban et al. 2015). They also serve as precursors for two important phytohormones, abscisic acid (ABA) and strigolactones, which are key regulators for plant development and stress responses

(Yuan et al. 2015). Tocopherols present in leaves have vitamin E activity and protect plants against oxidative damage through free radicals scavenging. Ascorbic acid (vitamin C) also plays a vital role in plants as it can improve tolerance against abiotic stresses by enhancing plant growth, rate of photosynthesis, transpiration, oxidative defense potential and photosynthetic pigments (Akram et al. 2017; Wang et al. 2015) Low heritability for vitamin E and moderate heritability for vitamin C were reported in tomato, making those traits hard to breed for in the species (Schauer et al. 2008). Since carotenoids, tocopherols and ascorbic acid are involved in stress response mechanisms in plants, these compounds are sensitive to variation in environmental conditions and circadian rhythm. Genetic analysis of these traits requires appropriate phenotyping strategies under controlled conditions.

Besides the significant variation in nutrients in *G. gynandra*, Sogbohossou et al. (2019) highlighted plant height, leaf area and flowering time as the most variable leaf yield-related traits in the species. Flowering time has been extensively studied in plants as a major adaptive trait and is the result of an interplay between developmental and environmental signals (Jung and Müller 2009; Li et al. 2016). Flowering time is an important trait in leafy vegetables as early flowering negatively affects leaf yield and quality (Takada et al. 2019). Farmers pointed out early flowering in *G. gynandra* as a major constraint for cultivation (Onyango et al. 2013b). Breeding for late-flowering varieties with high nutrient content and leaf yield in *G. gynandra* requires a thorough understanding of the genetic mechanisms underlying these traits with emphasis on their heritability, the gene actions involved and the extent of genotype by environment interaction.

Translational genomics between *G. gynandra* and *Arabidopsis thaliana* can help elucidate the genetic mechanisms underlying flowering time, leaf size and vitamin content in the species. Such a strategy can be implemented by taking advantage of the Arabidopsis Information Resource (https://www.arabidopsis.org/) and the recent database of flowering time genes (FLOR-ID) (Bouché et al. 2016). Given the genetic diversity within *G. gynandra*, different approaches could be used for mapping of traits of interest. However, as abovementioned traits are strongly correlated with the geographic origin of the accessions (Sogbohossou et al. 2019), genome-wide association mapping will likely give rise to spurious associations and make it harder to pinpoint genomic regions of interest due to population structure (Brachi et al. 2011). Linkage mapping could therefore be considered as a preferable option to identify quantitative trait loci and candidate genes. Other confirmatory approaches, including bulked segregant analysis, gene expression analysis, and mutation breeding, might be subsequently used either separately or combined for functional validation of candidate genes.

The present paper aimed to: (1) develop the first linkage map in *G. gynandra* based on an F_2 population resulting from a cross between two contrasting accessions from Asia and Africa; (2) identify quantitative trait loci for morphological traits as well as vitamin content including carotenoids, tocopherols and ascorbic acid and (3) identify potential candidate genes underlying the traits of interest based on homology searches with *A. thaliana*.

Materials and methods

Plant material and DNA isolation

We developed an F_2 population TOT7200 x TOT8917 consisting of 219 individuals, that was grown along with the parents in the greenhouse of Wageningen University, The Netherlands from March to June 2018. Seven individuals of each parent were randomized over the greenhouse. Those two parents were chosen for their contrasting phenotypes in term of height, flowering time, leaf area and vitamin content (Sogbohossou et al. 2019) and self-pollinated for five generations. The female parent was derived from selfing TOT7200, an accession from Malaysia with short and early flowering plants with relatively small leaves that have high tocopherol content and low carotenoid content (**Figure 1A**). The male parent was derived from selfing TOT8917, an accession from Malawi constiting of tall late-flowering plants with relatively broad leaves, low tocopherol content amd high carotenoid content (**Figure 1B**). The plants were grown under irrigated conditions with 24 °C day/20 °C night temperatures and a defined photoperiod (16 h day/8 h night) using artificial light. Young leaves were sampled for each individual and silica-dried for DNA extraction. Ten to fifteen leaves were sampled on 10-week-old plants between 9.00 and 11.00 am and directly frozen in liquid nitrogen for subsequent metabolic analyses.

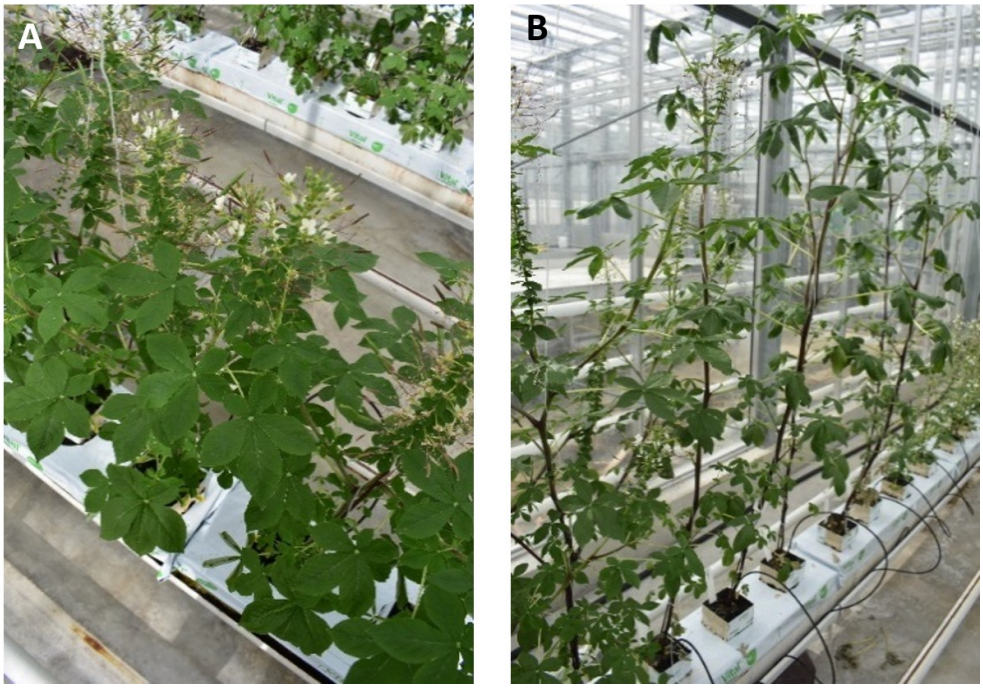


Figure 1. Parental lines crossed to generate the F_2 mapping population. (A) Individuals of the maternal line TOT7200; (B) Individuals of the paternal line TOT8917.

Phenotyping

Morphological characterisation

The morphological traits collected included plant height at 10 weeks and leaf area on three leaves per plant. Flowering time was measured in number of days from sowing to flowering. Flowers are hermaphroditic but because of the high number of male sterile plants in the population (reduced or aborted stamens), male sterility was scored as a binary trait (0: fertile plant and 1: sterile plant).

Extraction and analysis of carotenoids and tocopherols

Carotenoids and tocopherols detection and quantification was performed as described by (Sogbohossou et al. 2019). Briefly, 500 mg fresh weight of ground leaves was freeze-dried overnight and extracted with 4.5 ml of methanol/chloroform (1:1, v/v) containing 0.1% butylated hydroxytoluene (BHT) as antioxidant and 10 µM Sudan 1 as internal standard. After vortexing, samples were sonicated for 15 min and centrifuged at 2500 rpm for 10 min. Then, 2 ml of the supernatant was transferred to a new tube and dried under a gentle nitrogen flow. The dried samples were dissolved in 1 ml of ethanol containing 0.1% BHT, sonicated for 5 min and centrifuged at 2000 g for 5 min. The clear supernatant was transferred to amber vials for high-performance liquid chromatography (HPLC) analysis. HPLC analysis was performed according to Wahyuni et al. (2011) using a YMC Pack reverse-phase C30 column (250 × 4.6 mm; 5 µm) coupled to a 20 × 4.6 mm C30 guard (YMC Inc. Wilmington, NC, USA), maintained at 35°C. Chromatography was carried out on a Waters system consisting of a no. 2890 quaternary pump, no. 2996 photodiode array detector (PDA) and no. 2475 fluorescence detector. The mobile phase used was methanol, tert-methyl butyl ether and water/methanol (1:4, v/v) containing 0.2% ammonium acetate. Flow rate was 1 ml/min. Data were collected and analysed using the Waters Empower software.

Carotenoids were detected by setting the PDA to scan from 220 to 700 nm. Measurements for β-carotene and lutein were taken at 478 nm and violaxanthin at 440 nm. Tocopherols were detected by excitation at 296 nm and emission at 340 nm. Quantitative determination of compounds was conducted by comparison with dose–response curves constructed from authentic standards. To check the technical variation, including extraction, sample analysis and data processing, quality control samples were prepared from pooled leaf material of several randomly chosen accessions, extracted using the same procedure and injected after every ten sample extracts.

Extraction and analysis of ascorbic acid

The extraction and analysis of ascorbic acid (vitamin C) was made as previously described by (Wahyuni et al. 2011). An extraction solution of 5% metaphosphoric acid in purified water containing 1 mM diethylenetriaminepentaacetic acid was prepared. Three hundred milligrams of frozen and grinded material was weighed in cold 2-ml Eppendorf tubes and 1.2 ml of ice-cold extraction solution added. The extracts were vortexed, sonicated for 15 min and centrifuged at 2500 g for 20 min. The supernatants were filtered through 0.2-µm polytetrafluoroethylene filters and pipetted into amber vials for HPLC–PDA analysis using the same Waters HPLC system as

described above. Separation was made at 30 °C using a YMC-Pro C18 column (YMC Europe GmbH; 150 × 3.9 mm) with 50 mM KH_2PO_4 buffer (pH 4.4) as eluent at a flow of 0.5 ml/min. Quantification was made based on absorbance at 260 nm, using a calibration curve of an authentic l-ascorbic acid standard from Merck. Retention time of chromatographic peak of ascorbic acid in plant extracts was verified by co-elution with the authentic standard. As performed for the previous analysis, quality control samples from pooled leaf material of all accessions were extracted using the same procedure and injected after every ten sample extracts.

Whole-genome assembly and annotation

The genome of the accession Gyn from Malaysia was sequenced on Illumina and Chromium 10X genomics platforms and a hybrid assembly was performed (**Chapter 6**). The repeat masked genome assembly was used for de novo predictions with AUGUSTUS (version 2.03) (Stanke and Morgenstern 2005), GlimmerHMM (version 3.02) (Majoros et al. 2004), and SNAP (version 2.0) (Korf 2004). The protein sequences from *Arabidopsis thaliana, Brassica rapa, Carica papaya, Glycine max, Theobroma cacao,* and *Vitis vinifera* were mapped to the *G. gynandra* genome using tBLASTn, with an expected value cutoff of 10^{-5}, and Genewise (version 2.2.0) (Birney et al. 2004) was used for gene annotation. RNA reads generated from a previous transcriptome analysis of the species (Külahoglu et al. 2014) were mapped back to the reference genome using Tophat (version 1.0.14) (Trapnell et al., 2009), implemented with bowtie version 0.12.5, and the transcripts were assembled according to the genome using Cufflinks (version 0.8.2) (Trapnell et al., 2012). All the predictions were combined using GLEAN (Mackey et al. 2005) to produce the consensus gene sets.

Whole-genome re-sequencing of parental lines

Genomic DNA was extracted from young leaves of the two parents using a modified CTAB method. Briefly, approximately 50 mg of frozen young leaves was weighed per plant and placed into 2 ml tubes containing five glass beads (Ø 3 mm). The frozen sample was homogenized using beads (Retch, type MM2) for 2 minutes at 80 rpm. Liquid nitrogen freezing and homogenization were repeated until the leaf material had turned into a fine powder. One ml of CTAB buffer (2% CTAB, 2% PVP-40, 100 mM Tris-HCL, 1.4 M NaCl, 20 mM EDTA) and 5 µl β-mercaptoethanol were added with subsequent incubation for 30 min at 65°C. 500 µl of 24:1 chloroform:isoamylalcohol was then added, vortexed and centrifuged at 14.000 rpm for 4 min. The supernatant was removed, and the chloroform extraction was repeated. DNA was precipitated using 70% isopropanol at −20°C overnight, after which the DNA was pelleted at 14.000 rpm for 5 min. DNA was then re-suspended in TE buffer and treated with RNAse (Qiagen). DNA purification was performed using the Wizard DNA clean-up system (Promega Corp.) in combination with a vacuum manifold (Promega Corp.). The DNA was dissolved in 75 µl of pre-heated elution buffer (Qiagen). DNA extractions were visualized on 1% agarose gels, and the quantity was measured using a Qubit and Qubit DNA HS kit (Thermo Scientific).

DNA for resequencing was extracted using Isolate II Plant DNA extraction kit from Bioline Meridian Biosciences, xMemphis, TN, USA (https://www.bioline.com/us/downloads/dl/file/id/879/isolate_ii_plant_dna_kit_product_manual.pdf) . In brief, the lyophilised (~20 mg) or

fresh leaves (~100 mg) were ground in liquid nitrogen, 400 µL lysis buffer was added followed by thorough vortexing, 10 mL of RNase A (concentration of 1 mg/mL) was added to this mix. The mixture was incubated at room temperature for 30-45 minutes and filtered by centrifuging for 2 min at 11,000 g. To the clear filtered solution, 450 µL of binding buffer was added and mixed by pipetting for 5-8 times. This solution was loaded into binding column and centrifuged for 1 min at 11,000 g, flow through discarded and the column washed using 700 mL of wash buffer followed by centrifuging for 1 min at 11,000 g. The DNA was finally eluted in 100 mL TE (Tris-Cl, pH 7.8 and EDTA, pH 8.0).

The library of Ion Proton runs was prepared using Ion Express Plus gDNA Fragment Library Preparation kit (ThermoFisher Scientific, Life Technologies Corporation, Frederick, MD, USA, Cat. No. 4471269) as follows. Initially 1 µg DNA was fragmented in a reaction volume of 50 µL using 10 µL Ion Shear Plus 10X Reaction Buffer, 10 µL Ion Shear Plus Enzyme Mix II and the remaining amount of deionized water (MilliQ), mixed thoroughly by rapid pipetting and incubated for 15 minutes ate 37^0 C. To stop the reaction, 5 µL of Ion Shear Stop Buffer was added to it and mixed by continuous vortexing or pipetting for 5-10 seconds and then the reaction mix was stored on ice. The fragments were then purified by adding 99 µL of Agencourt AMPure XP Reagent (Beckman Coulter, Brea, CA, USA, Cat. No. A63880) incubated at room temperature for 5 minutes, the beads separated on a magnetic stand, the clear supernatant discarded by pipetting out, the bead pellet washed by rinsing in 500 µL of fresh 70% ethanol, the solution separated again on magnetic stand, and washed once again with 70% ethanol, the separated bead pellet air dried at room temperature for 5-10 minutes, and finally the fragmented DNA bound to the beads eluted in 25 µL of TE buffer.

The purified DNA fragments were ligated to bar coded adapters using Ion Xpress Plus Fragment Library Adapters 1-16 kit (Thermo Fisher Scientific, Life Technologies Corporation, Frederick, MD, USA, Cat. No. 4471250). In a reaction volume of 100 µL, 25 µL of the fragment elute was mixed with 10 µL Ligase buffer, 2 µL Ion P1 Adapter, 2 µL of Ion Xpress Barcode X, 2 µL of dNTP mix, 2 µL of DNA ligase, 8 µL of Nick Repair Polymerase and remaining volume made using deionized water (MilliQ) and incubated at 25 ^{0}C for 15 minutes. The reaction was stopped by denaturation at 72 ^{0}C for 5 minutes followed by immediate transfer to ice bath. The ligated fragments were purified using same procedure as described earlier using 120 µL of Agencourt AMPure XP Reagent followed by two washes of 500 µL of 70% ethanol and final elution in 20 µL of TE. The bar code adapter ligated DNA fragments were then passed through Pippin Prep System (Sage Science, Cat. No. PIP0001) using Pippin Prep Kit CDF 2010 (Sage Science, Cat. No. CDF2010) to elute library size of 200 base-read, which had a target base pair setting of 270 bp. After the elution, the library was once again purified using 90 µL of Agencourt AMPure XP Reagent followed by two washes of 500 µL of 70% ethanol and final elution in 50 µL of TE. The library was quality checked on BioAnalyser system for library size and Qubit for quantity. The quality passed libraries were loaded into Ion Chef (ThermoFisher Scientific, Life Technologies Corporation, Frederick, MD, USA, Cat. No. 4484177) using Ion PI Hi-Q Chef kit (ThermoFisher Scientific, Life Technologies Corporation, Frederick, MD, USA, Cat. No. A27198) for library amplification, Ion Sphere Particles (ISP) recovery, enrichment, and chip loading. The ready to load chips were loaded into the Ion Proton System for Next-Generation Sequencing (ThermoFisher

Scientific, Life Technologies Corporation, Frederick, MD, USA, Cat. No. 4476610) for a run length of 200 bases on PI chip using Ion PI Hi-Q Sequencing 200 Kit (ThermoFisher Scientific, Life Technologies Corporation, Frederick, MD, USA, Cat. No. A26772).

SNP calling and selection

The raw reads generated for each individual were checked for quality and trimmed using Trimmomatic (Bolger et al. 2014). The reads were aligned to the reference genome using the BWA MEM algorithm(Li 2013) and Samtools (Li et al. 2009). Duplicate reads were filtered and read groups added using GATK (version 4.0) Variant calling and filtering was performed with the HaplotypeCaller and the VariantFiltration tool of GATK (Van der Auwera et al. 2013). In order to select SNPS to be used for the characterisation of the F_2 population, heterozygous SNPs and SNPs with no call were discarded. Non-polymorphic SNPs between both parents were removed using vcffilterjdk (Lindenbaum and Redon 2017) and JEXL expressions. SNP clusters in 150 bp windows were also removed. We subsequently used a set of re-sequenced genomes of 48 accessions: 24 from East and Southern Africa and 24 from Asia to select SNPs that were present in at least 12 accessions from the same region of origin as the parents. Furthermore, we identified SNPs that were present in genomic regions of interest for our targeted traits (carotenoids, tocopherols, ascorbic acid biosynthesis pathways and potential flowering time genes) irrespective of whether they were unique to the parental lines or shared by other accessions from the same region. The final set consisted of 1309 SNPs that were used for the genotyping-by-sequencing.

Targeted genotyping-by-sequencing

The genotyping of the population was performed using a targeted GBS approach developed by LGC Group (https://www.biosearchtech.com/services/sequencing/targeted-genotyping-by-sequencing-seqsnp). Based on the SNP positions and the assembled reference genome, primers were designed for each SNP and 75 bp regions were sequenced with 200X coverage using NextSeq 500 v2. All library groups were demultiplexed using the Illumina bcl2fastq 2.17.1.14 software (Illumina 2013). Adapter remnants were clipped from all reads and reads with less than 65 bases were discarded. Quality trimming was performed by removing reads containing Ns and trimming reads at 3' end to get a minimum average Phred quality score of 30 over a window of ten bases. Reads with length less than 65 base pairs were again discarded. Read count per allele per SNP were then generated for all the samples. Alignment of quality trimmed reads against the reference genome was performed using Bowtie2 (Langmead and Salzberg 2012). Variant discovery and genotyping of samples was then performed with Freebayes v1.0.2-16 (Garrison and Marth 2012). The ploidy level was set to 2 and genotypes were filtered for a minimum coverage of 8 reads.

Genetic linkage map construction

After removing SNPs with MAF<0.05 and SNPs that were ambiguous in parents and those with more than 10% missing genotype data, 585 SNPs remained from the initial 1309 but only 269 segregating 1:2:1 or 3:1 were used to build the genetic map. The genetic map was constructed with JoinMap 5 (van Ooijen 2006) using regression mapping and the Kosambi

function and visualized with MapChart (Voorrips 2002).

QTL analysis and visualization

A single marker regression followed by a composite interval mapping with 3 covariates were performed using R/qtl. The genome-wide significant LOD threshold was set based on the 95 percentile of 1000 permutations with alpha values of 0.01, 0.05 and 0.1. An approximate 95% Bayesian credible interval expanded to the closest markers was computed to assess the location of the QTLs. The composite interval mapping results were confirmed with a stepwise analysis using the stepwiseqtl function of R/qtl. The fitqtl function was used to determine the effect and percentage variation explained for each qtl as well as interactions between QTLs. A single marker regression on binary traits was fitted using R/qtl to detect QTLs for presence/absence of male-fertile flowers.

Results

Performance of the population, correlations and heritability in phenotypic traits

The F_2 population exhibited a wide range of variation and transgressive segregation for all the measured traits (**Figure 2; Table 1**). Of the 219 plants, 157 (71.69%) were male sterile. There was a two-fold variation in flowering time. A four-fold variation was observed for total carotenoids, as

Table 1. Summary of phenotypic traits in parental lines and F_2 population.

Trait	TOT7200	TOT8917	F_2		Broad-sense heritability (H^2)
	Mean	Mean	Mean	Range	
Flowering time (days)	33.20 ± 2.39^a	43.17 ± 2.04^b	36.62 ± 5.31^a	26.00 - 52.00	0.82
Plant height (cm)	106.82 ± 6.13^a	179.67 ± 5.01^b	129.24 ± 31.53^a	57.00 - 204.00	0.97
Leaf area (cm^2)	40.96 ±7.46^a	53.02 ± 8.67^a	42.97 ± 22.10^a	8.96 - 123.74	0.87
Lutein (µg/g FW)	21.00 ± 1.36^a	30.30 ± 1.83^b	24.14 ± 4.60^c	8.86 - 37.95	0.88
Beta-carotene (µg/g FW)	50.03 ± 3.81^a	63.56 ± 5.89^b	56.65 ± 11.54^b	21.53 - 88.25	0.82
Violaxanthin (µg/g FW)	4.94 ± 1.01^a	12.11 ± 2.00^b	7.08 ± 2.31^c	2.59 - 13.81	0.53
Total carotenoids (µg/g FW)	75.97 ± 4.81^a	105.97 ± 5.97^b	87.86 ± 16.98^c	33.82 - 135.11	0.90
Alpha-tocopherol (µg/g FW)	61.54 ± 9.24^a	35.30 ± 9.87^b	44.40 ± 22.91^b	10.83 - 122.05	0.83
Carotenoids/tocopherols ratio	1.25 ± 0.14^a	3.24 ± 1.02^b	2.52 ±1.35^b	0.52 - 6.95	0.71
Ascorbic acid (µg/g FW)	417.17 ± 137.71^a	473.69 ± 61.62^a	428.09 ± 171.98^a	73.83 - 956.88	0.62

*Mean values with different letters in exponent are significantly different at 5% threshold

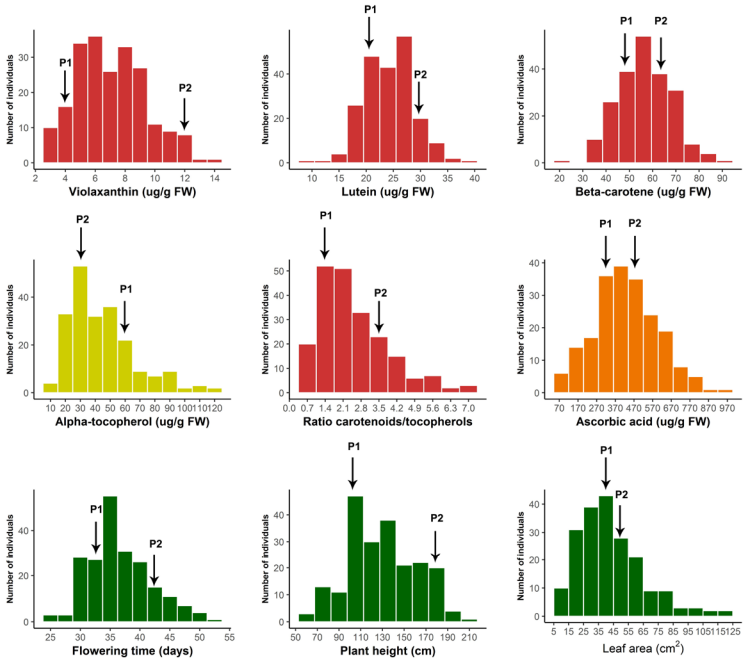


Figure 2. Distribution of metabolite levels and leaf-yield related traits in F_2 population of *Gynandropsis gynandra* (P1=TOT7200; P2=TOT8917).

well as lutein and beta-carotene variation when singled-out. Leaf area and ascorbic acid were the most variable traits with a thirteen-fold variation. As expected, the male parent TOT8917 (Malawi) had significantly higher values for plant height and carotenoid levels than the female TOT7200 (Malaysia) and a significantly lower tocopherol content. There were no significant differences between both parents for leaf area and ascorbic acid content although TOT8917 had higher values on average than TOT7200 for those two traits. The average values of the F_2 population for plant height and flowering time were not significantly different from the earlier-flowering and shorter female parent TOT7200. The population had on average similar values with TOT8917 for β-carotene, α-tocopherol and carotenoids-tocopherols ratio. All traits except violaxanthin (H^2=0.53) and ascorbic acid (H^2=0.62) levels had a broad-sense heritability above 0.7 (**Table 1**). Leaf-yield related traits i.e. plant height, leaf area and flowering time had a high and significant positive correlation between them as was also the case for all carotenoids (**Table 2**). There was also a significant positive correlation between plant height and ascorbic acid (R^2=0.16), lutein (R^2=0.23), violaxanthin (R^2=0.27) and total carotenoid content (R^2=0.18) There was no significant correlation between alpha-tocopherol and other traits.

The results of the first two principal component analysis on all the traits revealed that the first principal component explained 34.6% of variation while the second principal component explained 20.3% of variation. The projections of all the traits on the principal components are summarized in **Supplemental Table 1**. The first principal component was the dimension representing carotenoids while the second principal component represented leaf-yield related traits.

Table 2. Pearson correlation matrix of phenotypic traits measured in F_2 population of *Gynandropsis gynandra*.

Trait	Flowering time	Plant height	Leaf area	Lutein	Beta-carotene	Violaxanthin	Total carotenoids	Alpha-tocopherol	Carotenoids/ tocopherols ratio
Plant height	0.48***								
Leaf area	0.44***	0.57***							
Lutein	0.09	0.23***	0.07						
Beta-carotene	-0.02	0.11	0.04	0.82***					
Violaxanthin	0.03	0.27***	0.09	0.53***	0.57***				
Total carotenoids	0.01	0.18**	0.06	0.90***	0.98***	0.67***			
Alpha-tocopherol	-0.02	-0.01	0.09	-0.04	0	-0.05	-0.02		
Carotenoids/ tocopherols ratio	0.1	0.05	-0.08	0.35***	0.31***	0.19**	0.33***	-0.79***	
Ascorbic acid	-0.01	0.16*	0.12	0.13	0.1	0.07	0.11	0.07	-0.05

*$p < .001$ '***', $p < .01$ '**', $p < .05$ '*'*

Linkage map construction and segregation distortion

After removing SNPs with MAF<0.05, SNPs that were ambiguous in parents and those with more than 10% missing genotype data, 585 SNPs remained from the initial 1309 SNPs. The genetic map was constructed with 269 SNPs segregating 1:2:1 or 3:1 in the whole population and distributed over 17 linkage groups likely corresponding to the 17 chromosomes of *G. gynandra* (**Supplemental Figure 1**). Unfortunately, due to the low number of SNPs, linkage groups 8, 11, 15, 16 and 17 included only 3 or 2 SNPs. The linkage map had a total length of 859.1 cM with an average interval size of 3.20 cM. Based on the genome size of 740 Mbp, 1 cM on the genetic map spanned on average 861.3 kbp. From the 316 SNPs with high segregation distortion, 136 were skewed towards the female parent, 55 towards the male parent while 125 had a high heterozygosity.

QTL analysis

Twelve QTLs distributed over 8 of the 17 linkage groups were detected (**Figure 3; Table 3**). We found QTLs for all traits except for violaxanthin, beta-carotene and total carotenoid contents and male-sterility. The phenotypic variation explained ranged from 4.9 to 27.5% for all the QTLs. One QTL had a major effect (PVE > 25%), six had an intermediary effect (10% < PVE < 25%) and five a minor effect (PVE < 10%) (**Table 3**). QTLs additive models resulted in major effects for flowering time, tocopherols and carotenoids/tocopherols ratio, intermediary effect for plant height, leaf area and ascorbic acid and a minor effect for lutein content (**Table 3**).

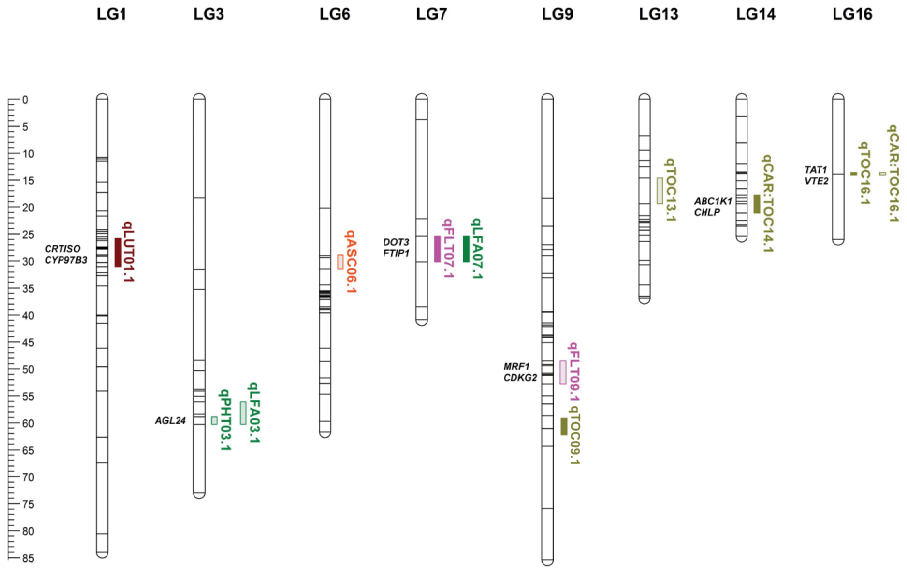


Figure 3. Quantitative trait loci detected for plant height and leaf area (green), flowering time (pink), carotenoids (red), alpha-tocopherol (olive green) and ascorbic acid (orange). QTLs are represented by coloured boxes on the right of the linkage group (Bayesian 95% credible interval around QTL peak expanded to the closest markers) and potential candidate genes detected are listed on the left of the QTLs or QTL clusters. QTLs for which the TOT7200 allele increases the phenotypic value are depicted with solid filled boxes while hatched boxes depict QTLs for which the TOT8917 allele increases the phenotypic value.

Table 3. QTLs detected by composite interval mapping in the F_2 population of *Gynandropsis gynandra*.

Trait	QTL name	Linkage group / QTL position	Marker interval	LOD score[a]	AE[b]	PVE[c] (%)	Total PVE (%)
Flowering time	qFLT07.1	7/30.2	sc194_840 – sc328_1043	6.2***	-2.15	10.6	27.9
	qFLT09.1	9/49.4	sc354_1078 – sc191_1449	8.8***	2.7	15.6	
Plant height	qPHT03.1	3/60.3	sc84_549 – sc623_1824	9.8***	22.7	22.2	22.2
Leaf area	qLFA03.1	3/60.3	sc113_1388 – sc623_1824	4.9***	0.9	13.0	21.8
	qLFA07.1	7/30.2	sc194_840 – sc328_1043	4.6**	-0.5	9.9	
Lutein	qLUT01.1	1/27.7	sc30_302 - sc282_1612	5.7***	-2.07	9.8	9.8
Alpha-tocopherol	qTOC09.1	9/61.1	Sc21_232	5.6***	-2.76	7.7	37.5
	qTOC13.1	13/19.4	sc47_410 – sc455-1707	4.2***	0.46	6.1	
	qTOC16.1	16/13.9	sc43_385	16.6***	-1.17	27.5	
Ratio carotenoids/ tocopherols	qCAR:TOC14.1	14/18.9	sc154_1427 – sc320_1488	3.6*	-0.5	6.4	31.8
	qCAR:TOC16.1	16/13.9	sc43_385	14.5***	0.8	24.7	
Ascorbic acid	qASC06.1	6/11.3	sc240_906 – sc31_312	7.4***	92.5	14.7	14.7

[a] Logarithm of odds: $p < .001$ '***', $p < .01$ '**', $p < .05$ '*'

[b] Additive effect of the QTL. Positive values indicate effect from TOT8917 and negative values indicate effect from TOT7200

[c] Phenotypic variation explained

In the following section, we considered as "favorable" alleles, alleles that contributed to higher values in all our traits of interest. Tall, late-flowering accessions, with broad leaves, high levels of carotenoids, tocopherols and ascorbic acid would therefore be considered as desirable even though, the desirable alleles for plant height can vary depending on growers' preferences.

Two QTLs with moderate effect were found for flowering time (qFLT07.1; qFLT09.1) on linkage groups 7 and 9 and together explained 27.9% of phenotypic variation. qFLT07.1 received the favorable allele from the female TOT7200 while qFLT09.1 had the opposite pattern. Those QTLs had an additive effect (**Supplemental Figure 2A**) but no significant interaction (p=0.22). A single QTL for plant height was located on linkage group 3 (qPHT03.1; 22.2% PVE) with the favorable allele from the male parent TOT8917. Two QTLs were detected for leaf area: a moderate one on linkage group 3 (qLFA03.1) and a minor one on linkage group 7 (qLFA07.1) which explained 21.8% of phenotypic variation when combined. Positive alleles for these QTLs came from different parents: LFA03.1 had the favorable allele from TOT8917 while the favorable allele of qLFA07.1 came from the female TOT7200 (**Figure 3**). Although there was no significant interaction between these QTLs (p=0.45), the highest leaf area was obtained from a combination of the male homozygote genotype for qLFA07.1 and the hybrid genotype for qLFA03.1 **(Supplemental Figure 2B).**

A single minor QTL related to lutein content was found on linkage group 1 (qLUT01.1; 9.8% PVE). Three QTLs were detected for alpha-tocopherol on linkage groups 9 (qTOC09.1; 7.7%PVE), 13 (qTOC13.1; 6.1% PVE) and 16 (qTOC16.1; 27.5% PVE) which combined explained 37.5% of phenotypic variation. The female allele was favorable for qTOC09.1 and qTOC16.1 while for qTOC13.1, the male one was favorable. Although interactions between pairs of these QTLs were not significant, the combination of the female homozygote of qTOC09.1 with either the hybrid of qTOC13.1 or the male homozygote of qTOC16.1 yielded the highest tocopherol levels **(Supplemental Figure 2C).** The carotenoids/tocopherols ratio revealed a first QTL on linkage group 14 (qCAR:TOC14.1; 6.7%) that did not appear for both carotenoids and tocopherols QTL analyses and a second QTL on linkage group 16 which overlapped with the QTL for alpha-tocopherol (qCAR:TOC16.1; 24.7% PVE). The two QTLs explained 31.8% of phenotypic variation when combined and had an additive effect but no significant interaction (p=0.61) **(Supplemental Figure 2D)**.

Some of the detected QTLs were co-located suggesting the presence of genes with pleiotropic effects at these locations or tight linkage. On linkage group 3, QTLs for plant height and leaf area overlapped and their favorable alleles were received from TOT8917. A second cluster was present on linkage group 7 where QTLs for leaf area and flowering time overlapped. QTLs for tocopherols and carotenoids-tocopherols ratio were co-located on linkage group 16.

Candidate gene identification

A list of known genes related to our traits of interest in *Arabidopsis thaliana* was retrieved based on GO terms from TAIR (https://www.arabidopsis.org/) and the Arabidopsis flowering time genes database (FLOR-ID) (Bouché et al. 2016). Protein sequences for each gene were blasted against the reference genome of *G. gynandra* using an expected value cutoff of 10^{-20}. Genes

present within a centimorgan (on average 861.3 kbp) window from a marker were retained as potential candidate genes.

We identified candidate genes for flowering time (*FTIP1, CDKG2, MRF1*), carotenoid biosynthesis (*CRTISO, CYP97B3*), tocopherol biosynthesis (*ABC1K1, CHLP, TAT1, VTE2*), leaf area (*DOT3*) and a candidate gene with potential pleiotropic effects (*AGL24*) co-localized with QTLs for flowering time and plant height. No candidate gene was detected for ascorbic acid (**Figure 3**). The complete list of candidate genes per QTL is presented as **Supplemental Table 2**.

Discussion

In this study, linkage mapping was proved to be an efficient approach for QTL and candidate gene identification in *G. gynandra.* In our F_2 population, transgressive segregation was observed in all the measured traits and we identified QTLs for most traits except beta-carotene and violaxanthin content. A high correlation was found between flowering time, plant height and leaf area and between the carotenoids as reported in a previous study (Sogbohossou et al. 2019). However, the negative correlation observed between carotenoids and tocopherols previously observed in a diversity panel of the species (Sogbohossou et al. 2019) was not found in this F_2 population. We therefore hypothesize that this negative correlation was not driven by antagonistic biosynthesis mechanisms within the plant itself but rather by population structure probably due geographic adaptation or human selection.

The low number of SNPs used to construct the genetic map (269 SNPs) is an important limitation to the reliability of the position of the QTLs and the regions they spanned. To improve the quality of the map, we are currently genotyping a second population with a set of 9000 good quality SNPs, this time, without considering their presence in other accessions from the region of origin of the parents. The presence of QTL clusters on linkage groups 3, 7 and 16 suggests the presence of genes with pleiotropic effects or tightly linked genes influencing our traits of interest at these locations. However, improving the coverage and resolution of our genetic map using a higher density of markers could further help pinpoint the genomic regions spanned by those QTLs.

We identified two candidate genes related to carotenoids biosynthesis and particularly lutein in *G. gynandra* co-localized with the QTLs: *CRTISO* and *CYP97B3*. The carotenoid isomerase gene (CRTISO) is involved in all-trans-lycopene formation which is later cyclized to carotenes (δ-, γ-, α-, β-carotene). Carotenes are then oxygenated to xanthophylls (e.g. lutein, violaxanthin) (Nisar et al. 2015). While *CYP97A* and *CYP97C1* were for a long time reported as the main genes involved in lutein biosynthesis, *CYP97B3*, was also identified as an important gene in the pathway (Kim et al. 2010). Mutants with overexpression of the gene *CYP97B3* in *A. thaliana* had up to 150% higher levels of zeinoxanthin and lutein than the wild type. Heme-containing cytochrome P450 (CYP) monooxygenases are known to be involved in secondary metabolite production in plants, usually in the form of hydroxylases even though many of them have an unknown function (Kim et al. 2010). We expect the candidate genes here identified in *G. gynandra* to have the same functions as in *A. thaliana.*

Tocopherol candidate genes included *ABC1K1*, *CHLP*, *TAT1* and *VTE2*. *ABC1K1* (activity of bc1 complex kinase 1) and *ABC1K3* are involved in the regulation of light-mediated

development in plants and α-tocopherol accumulation under high-light (Eugeni Piller et al. 2014; Yang et al. 2016). *CHLP* encodes for geranygeranyl reductase, an enzyme that catalyzes the reduction of geranylgeranyl pyrophosphate to phytyl pyrophosphate, a critical step on the biosynthesis pathway of carotenoids, tocopherols and chlorophylls. Plants with low *CHLP* expression partially accumulated geranylated chlorophyll a, had lower chlorophyll and tocopherol levels and hence had an increased sensitivity to prolonged exposition to high light and a delayed growth (Graßes et al. 2001; Tanaka et al. 1999). Tyrosine aminotransferases (TATs) catalyze the reversible reaction between tyrosine and 4-hydroxyphenylpyruvate (HPP) and in *A. thaliana*, *TAT1* (or *TAT7*) was responsible of 33-47% of tocopherols synthetized in the leaves (Riewe et al. 2012; Wang et al. 2019). *VTE2,* also known as *HPT,* encodes homogentisate phytyltransferase, an enzyme responsible for the early condensation of homogentisic acid (HGA) and phytyl diphosphate (PDP) to form 2-methyl-6-phytyl-1,4-benzoquinol (MPBQ) considered as the committed step of tocopherol biosynthesis (DellaPenna and Pogson 2006; Wunnakup et al. 2018). Arabidopsis *vte2* mutants exhibited severe seedling growth defects and a 4- and 100-fold increase in levels of lipid hydroperoxides and hydroxy fatty acids, respectively, compared with the wild type (Sattler et al. 2004).

The defectively organized tributaries gene *DOT3* was co-localized with leaf area and flowering time QTLs on linkage groups 7. *DOT3* is responsible of leaf vein spacing patterns in *A. thaliana*, a trait also reported to be variable in *G. gynandra* between East African and Asian accessions along with stomatal density and water use efficiency (Reeves et al. 2018). The authors found similar vein spacing patterns in Asian and West African accessions that had small leaves compared with East African accessions. We therefore hypothesize that vein spacing and leaf area are linked in *G. gynandra*. Arabidopsis *dot3* mutants have defects in shoot and primary root growth and produce an aberrant parallel venation pattern in juvenile leaves. Plants with the more severe dot3–2 allele were smaller in size throughout development, and less fertile than the wild type (Petricka et al. 2008). The co-localisation of leaf area and flowering time is most likely due to the positive correlation between both traits.

Flowering time potential candidate genes included *FTIP1* co-localized with qFLT07.1 as well as *MRF1* and *CDKG2* co-localized with qFLT09.1. Flowering time interacting protein 1 (*FTIP1)* regulates the transport of the Flowering Time locus T (FT) protein, which plays a key-role in flowering induction, from the phloem to the shoot apical meristem (Abe et al. 2015; Liu et al. 2012). Morn-motif repeat protein regulating flowering 1 (*MRF1*) is a photoperiod responsive gene that promotes flowering under long day conditions (You et al. 2019). As our experiments were conducted under long days, it is expected that *MRF1* influenced flowering in the F_2 population. Cyclin-dependent kinase G2 (*CDKG2)* acts as a negative regulator of the salinity stress response and plant flowering transition in *A. thaliana* (Ma et al. 2015). Agamous-like 24 (*AGL24*) was present in the region on linkage group 3 spanning 2 QTLs involved in plant height and flowering time. *AGL24* is a MADS-box transcription factor involved in both floral induction and floral development (Lee et al. 2008; Sun et al. 2016). The fact that *AGL24* is also involved in apical meristem differentiation into an inflorescence and that flowering time and plant height are highly correlated in *G. gynandra* might explain the presence of the gene in the same region as QTLs for both traits. It was not clear from our different experiments with the species whether *G. gynandra* was photo-period sensitive and there was no evidence in the

literature about day-length sensitivity in the species. However, the presence of two photoperiod related flowering genes *FTIP1* and *MRF1* among our candidate genes suggests that photoperiod plays an important role in flowering induction in *G. gynandra*.

The male sterility issue in the F_2 population hindered the development of a RIL population and could indicate incompatibilities between the genomes of both accessions and on-going reproductive isolation. Similar cases of intraspecific incompatibility were reported in other plant species. In *Arabidopsis lyrata*, observations of phenotypic male sterility in F_2 generations were due to a cytoplasmic male sterility system in the species (Aalto et al. 2013). Negative epistatic interactions between alleles that have evolved independently in each parental line, a phenomenon known as Bateson–Dobzhansky–Muller incompatibility (BDMI), are the most common cause of intrinsic postzygotic isolation (Lafon-Placette and Köhler 2015). However, other mechanisms such as underdominance and gene duplication, transposition and gene loss, or epigenetic mechanisms may also induce intrinsic postzygotic isolation. BDMIs were long thought to arise from deactivation of gene duplicates or as a by-product of ecological selection (Lafon-Placette and Köhler 2015; Seehausen et al. 2014). From our observations, crosses where the maternal line was East African and the paternal line from West Africa or Asia resulted systematically in low seed set and sterile F_1 while the reciprocal crosses provided fertile F_1s. However, in the present F_2 population, about 75% of the individuals were male-sterile. In addition, seed set in fertile individuals was low compared with the parental lines. A deeper analysis of nuclear–cytoplasmic incompatibility in crosses between several African and Asian accessions could elucidate the phenotypic male-sterility issues in those lineages. Moreover, besides the possibility to cross different accessions from the same region to develop RIL populations, elaborating a protocol for Doubled Haploid induction in the species would be beneficial to fully exploit the natural intraspecific variation in several traits in future genetic studies and understand their genetic control.

The present study provides QTLs and candidate genes for flowering time, leaf area, plant height, carotenoids, tocopherols and ascorbic acid content in *G. gynandra*. Our results need to be validated with future studies including different mapping populations, bulked segregant analysis, mutant lines, growth under other environments (such as short-days), fine mapping of traits, as well as transcript abundance analysis of candidate genes. The presence of QTL clusters for our traits of interest and positive correlations between some traits indicates the possibility to breed simultaneously for all of them. Markers identified in this study could be used for molecular screening of breeding populations and selection of individuals of interest. Future studies will be focused on refining the genetic map and improving the reference genome of the species to narrow-down the genomic regions of interest. Functional genomics approaches will be helpful to validate the functions of our candidate genes and the allelic effects described in this study. Regarding flowering time genes, photoperiod sensitivity in *G. gynandra* is a topic that also requires further investigation. The effect of agricultural practices including irrigation, fertilization schemes, harvest and storage methods and duration, processing on our targeted nutrients should also be documented.

While this study represents a first step towards *G. gynandra* breeding, the path to the development of nutrient-rich and high yielding varieties is still long and will require transparent

communication and close collaboration between researchers and other stakeholders of the leafy vegetable value chain in targeted countries.

Acknowledgements

We thank the World Vegetable Center and especially Andreas Ebert for providing seeds of the parental accessions. We also thank Maarten Peters for the greenhouse management, Niccolo Bassetti for helping with leaf sampling and Bert Schipper for his assistance with HPLC analyses. This study was supported by the Schlumberger Foundation Faculty for the Future Fellowship awarded to E.O.D.S. and the African Orphan Crops Consortium.

Supplemental material

Supplemental Table 1. Coordinates on principal components of phenotypic traits measured in the *Gynandropsis gynandra* F_2 population TOT7200 x TOT8917

Supplemental Table 2. List of candidate genes identified from QTL analysis on the F_2 TOT7200 x TOT8917 in *Gynandropsis gynandra*

Supplemental Figure 1. Genetic map from *Gynandropsis gynandra* F_2 population TOT7200 x TOT8917

Supplemental Figure 2. QTLs effects and interactions

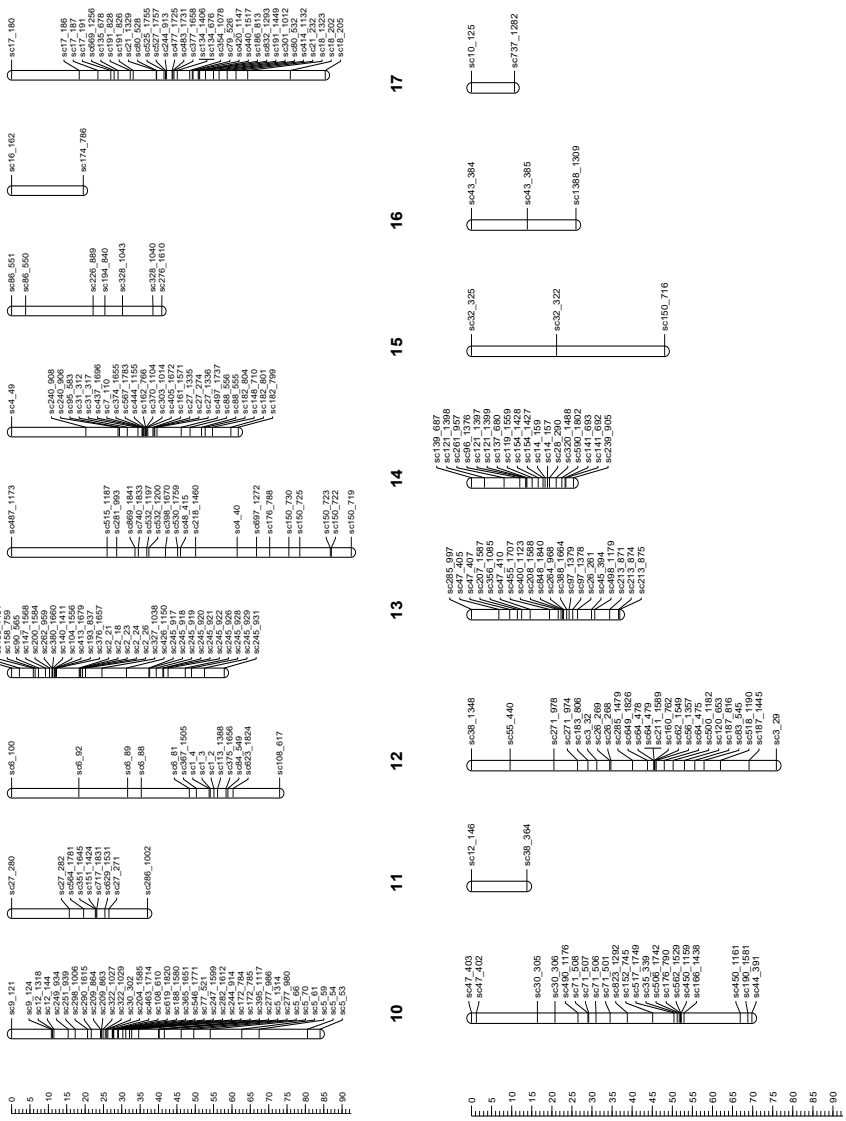
Supplemental Table 1. Coordinates on principal components of phenotypic traits measured in the *Gynandropsis gynandra* F_2 population TOT7200 x TOT8917

Trait	PC1	PC2	PC3	PC4
Flowering time	0.06	0.66	-0.42	-0.21
Plant height	0.27	0.76	-0.27	0.02
Leaf area	0.13	0.81	-0.19	-0.01
Lutein	0.90	-0.02	0.10	-0.03
Beta-carotene	0.93	-0.10	0.19	-0.05
Violaxanthin	0.72	0.01	0.15	-0.10
Total carotenoids	0.97	-0.07	0.18	-0.06
Alpha-tocopherol	-0.14	0.40	0.84	-0.16
Carotenoids/tocopherols ratio	0.45	-0.38	-0.75	0.10
Ascorbic acid	0.15	0.24	0.18	0.93
Proportion of variance explained (%)	34.7	20.3	17.0	9.6

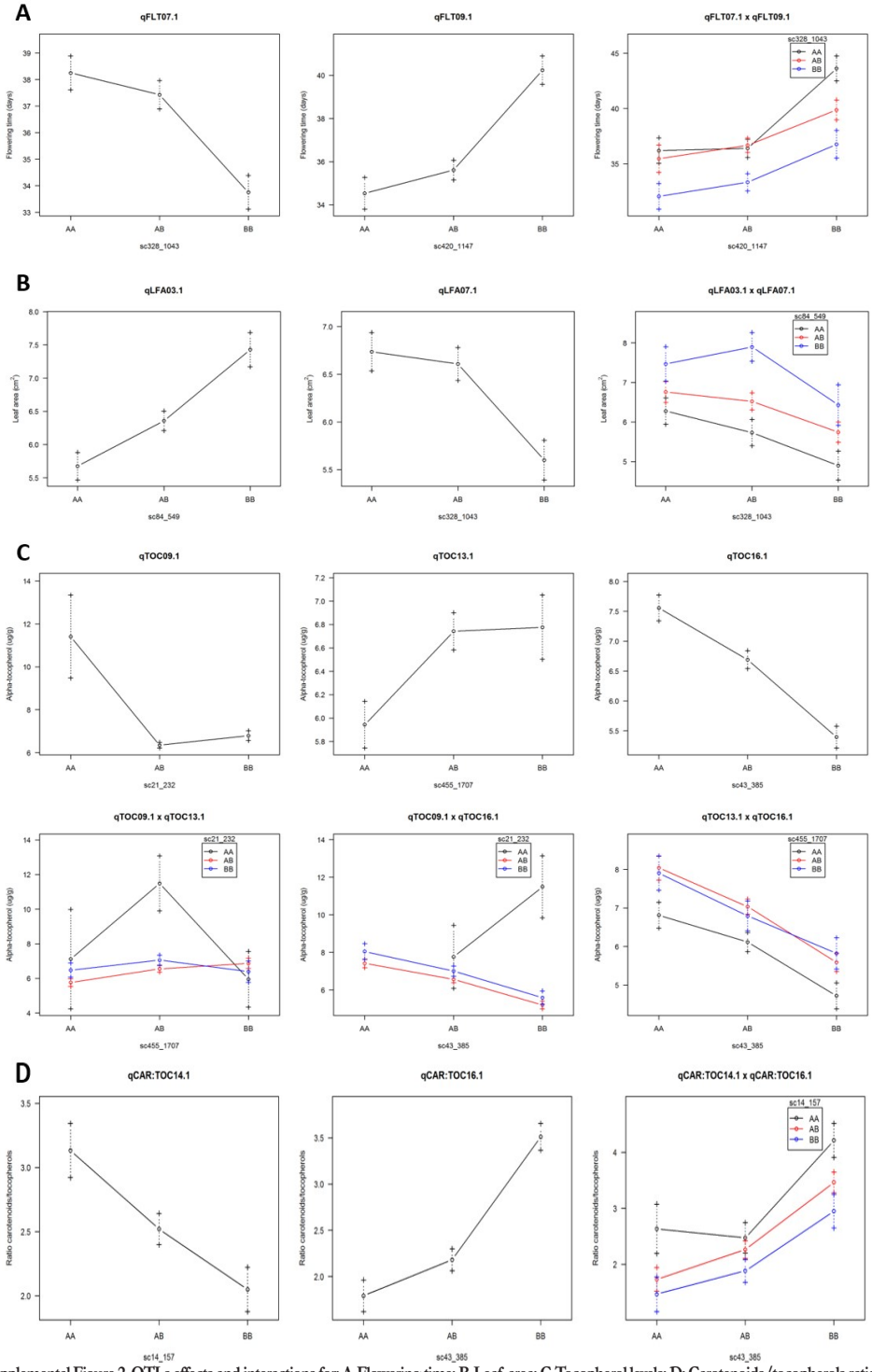
Supplemental Table 2. List of candidate genes identified from QTL analysis on the F_2 TOT7200xTOT8917 in *Gynandropsis gynandra*

Marker	QTL(s)	Scaffold	Position	*G. gynandra*	*A. thaliana*	Gene name	Position	Distance to marker (kbp)	E-value	Gene description TAIR
sc30_302	qLUT01.1	scaffold30	991361	Cg03906.1	AT4G15110.1	CYP97B3	1242938	252	2E-77	Heme-containing cytochrome P450 probably involved in the biosynthesis of xanthophylls. Hydroxylates beta-rings of alpha-carotene. May also have activity on beta-rings of beta-carotene.
sc30_302	qLUT01.1	scaffold30	991361	Cg03873.1	AT1G06820.1	CRTISO	968093	23	1E-44	Encodes carotenoid isomerase. Catalyzes the isomerization of poly-cis-carotenoids to all-trans-carotenoids. Together with PDS and ZDS, CRTiso is required to complete the synthesis of lycopene from phytoene
sc113_1388	qLFA03.1	scaffold113	538061	Cg26427.1	AT4G24540.1	AGL24	44392	494	2E-24	Encodes a MADS-box protein involved in flowering. Regulates the expression of SOC1 and is also upregulated by SOC1. Binds with IMK3 kinase domain. Phosphorylated by IMK3; likely to be a target for IMK3 kinase domain.
sc194_840	qFLT07.1 qLFA07.1	scaffold194	1111300	Cg12444.1	AT5G10250.1	DOT3	1611818	501	2E-32	Encodes a protein with an N-terminal BTB/POZ domain and a C-terminal NPH3 family domain. dot3 mutants have defects in shoot and primary root growth and produce an aberrant parallel venation pattern in juvenile leaves.
sc328_1043	qFLT07.1 qLFA07.1	scaffold328	1384944	Cg19287.1	AT5G06850.1	FTIP1	1416290	31	0	Encodes an endoplasmic reticulum protein that is involved in the transport of the florigen FT from companion cells to sieve elements, thus affecting FT transport through the phloem to the SAM.

sc79_526	qFLT09.1	scaffold79	2151823	Cg19861.1	AT1G21920.1	MRF1	1784541	367	2E-69	MRF1 is related to SET7/9 proteins but contains an atypical SET domain. It is expressed in phloem and mutants have a weak late-flowering phenotype.
sc440_1517	qFLT09.1	scaffold440	295582	Cg24883.1	AT1G67580.1	CDKG2	426873	131	3E-31	Protein kinase superfamily protein
sc154_1427	qCAR:TOC14.1	scaffold154	990041	Cg12821.1	AT4G31390.1	ABC1K1	1687837	698	7E-87	Kinase that can phosphorylate the tocopherol cyclase VTE1, a key enzyme of tocopherol (vitamin E) metabolism and involved in the recycling of oxidated alpha-tocopherol quinone, possibly stabilizing it at plastoglobules.
sc320_1488	qCAR:TOC14.1	scaffold320	165727	Cg18089.1	AT1G74470.1	CHLP	303815	138	2E-103	Catalyzes the reduction of geranylgeranyl diphosphate to phytyl diphosphate, providing phytol for both tocopherol and chlorophyll synthesis.
sc43_385	qTOC16.1 qCAR:TOC16.1	scaffold43	598908	Cg03335.1	AT5G53970.1	TAT1	93573	505	1E-69	Encodes a cytosolic tyrosine aminotransferase which is strongly induced upon aging and coronatine treatment. AtTAT1 prefers Tyr as an amino donor but can also use Phe, Trp, His, Met, and Leu. The mRNA is cell-to-cell mobile.
sc43_385	qTOC16.1 qCAR:TOC16.1	scaffold43	598908	Cg03445.1	AT2G18950.1	VTE2	767735	169	8E-53	Encodes homogentisate phytyltransferase involved in tocopherol biosynthesis. Has impact on seed longevity and plays a role in the adaptation to low temperature stress, notably phloem loading.



Supplemental Figure 1. Genetic map from *Cynandropsis gynandra* F_2 population TOT7200xTOT8917.



Supplemental Figure 2. QTLs effects and interactions for: A-Flowering time; B-Leaf area; C-Tocopherol levels; D: Carotenoids/tocopherols ratio.

CHAPTER 8

General discussion

The present thesis aimed to change the status of spider plant (*Gynandropsis gynandra*) from "orphan" to "emerging" crop; with an emphasis on generating baseline information on the species to enable the development of breeding programs for improved leaf yield and nutritional quality. At the inception of the thesis, efforts to analyse germplasm collections of the species were scattered and only one more recent study (Wu et al. 2017) characterised the morphology in a worldwide collection of accessions from East/Southern Africa and Asia. Here, we discuss the main findings of our research journey which started with documenting local communities' knowledge on spider plant in Benin and Togo (West Africa) and ultimately allowed us detecting candidate genes underlying flowering time and vitamin biosynthesis in the species.

Lessons learned for *G. gynandra* breeding

Promoting orphan crops: cultural context matters

Non-staple crops that serve an important purpose in the diets and livelihoods of local communities around the world are sometimes referred to as "orphan crops." Promoting these species as "emerging crops" requires consideration of their cultural importance for specific communities. Thorough documentation of traditional knowledge and uses of these crops can be used to suggest recipes and uses to new consumers and guide pharmacological investigations on those crops. A potential challenge associated with the promotion of orphan crops is the existence of food taboos associated with specific crops in some communities. Food taboos go beyond avoidance of food due to simple preferences. They are often associated with spiritual, ecological or health related motives, and are also part of the cultural identity of specific groups (Briones Alonso et al. 2018; Meyer-Rochow 2009). These aspects should therefore be accounted for in promotion strategies for emerging crops and used to better define target groups, markets and uses. Consumers' perceived values and beliefs associated with such crops are important to understand food choices and preferences. For example, in Micronesia, promoting green leafy vegetables consumption to alleviate vitamin A deficiency failed because local communities were not used to consuming green leafy vegetables and considered them as food for pigs (Englberger 2012). In the case of *G. gynandra*, we did not observe any rejection of the species in the surveyed areas. However, it was primarily perceived as a medicinal plant in some communities and more as a vegetable with occasional medicinal uses in others. Such perceptions were also reflected in the management intensity of the resource (**Chapter 3**).

Besides investigating knowledge, utilization and cultural acceptance of orphan crops, traditional production techniques and their impact on crop diversity and livelihoods are also relevant for mainstreaming of good existing practices into modern production systems designed for the crops of interest. Key elements to take into account for leafy vegetables include for instance nursery practices, planting density, harvesting time, number of harvests (Houdegbe et al. 2018), crop protection (Francisco 2018), seed viability and seed distribution systems (Blalogoe 2019; Sohindji 2018). An example is the uprooting of *G. gynandra* in Kenya and the harvest through successive multiple cuttings in Benin and Togo. These differences call for the adoption of tailored production practices and the development of specific cultivars with good performance in each type of production system (Houdegbe et al. 2018). Our thesis highlights the importance of putting any orphan crop promotion project in the appropriate socio-economic and cultural context (**Chapter**

3) to make sure that farmers and consumers' preferences and needs are prioritised.

Untargeted metabolomic profiling of orphan crops to inform breeding strategies for health-benefits, taste and plant defence mechanisms

Plants produce a diversity of phytochemicals that are involved in defence against biotic agents and responses to environmental stresses; which also constitute a wide range of natural products with benefits for human health (Wink 2015; Yang et al. 2018b). The latest developments in metabolomics has allowed for comprehensive surveys of diverse metabolites produced in various plants (Carreno-Quintero et al. 2013). Detecting specific secondary metabolites can contribute to the development of biomarkers and health-promoting compounds. It can also help identifying plant defence mechanisms against pests and diseases.

The detection and analysis of diverse secondary metabolites in *G. gynandra* (**Chapter 5**) provided scientific support to back up claims by users in various parts of Africa and Asia about the medicinal antibiotic, anti-inflammatory and nutritious properties of the species (Allabi et al. 2011; Imanirampa and Alele 2016; Jeruto et al. 2008; Neugart et al. 2017; Sogbohossou et al. 2018b). We putatively identified different classes of semi-polar secondary metabolites in *G. gynandra* including terpenoids, sesquiterpenoids, glucosinolates, flavonoids and diverse phenolic compounds. Many of these compounds have been investigated in drug development programs for human health. Specifically, hydroxycinnamic acid (e.g. ferulic acid, caffeic acid) derivatives were detected in high levels in the leaves of West African and Asian accessions. These compounds reportedly possess potent antioxidant, anti-inflammatory properties and have potential therapeutic benefits for diabetes and hyperlipidaemia prevention (Alam et al. 2016). Coumarin derivatives including 5,6,7,8-tetrahydroxy-2H-1-benzopyran-2-one and 5,7,8-trihydroxy-2H-1-benzopyran-2-one were abundant in East African accessions. Coumarin derivatives are known for their pharmacological properties including antioxidant and anti-inflammatory activities (Dandriyal et al. 2016; Witaicenis et al. 2014). Coumarin derivatives have also been explored as anticancer agents and they possess minimum side effects (Dandriyal et al. 2016). Ellagic acid and related derived metabolites, were abundant in East African accessions, and have been reported to have diverse activities including antioxidants, estrogenic, anti-inflammatory and anti-carcinogenic (Bell and Hawthorne 2008; Landete 2011; Vattem and Shetty 2005). Glucosinolates, sulfur-containing compounds found in the order Brassicales (Fahey et al. 2001) were as expected identified in *G. gynandra* (Neugart et al. 2017; **Chapter 5**). Degradation products of glucosinolates (isothiocyanates) have demonstrated anti-cancer properties and also can prevent cardio-vascular diseases (Cartea and Velasco 2008; Traka and Mithen 2009).

Volatile compounds identified in *G. gynandra* included aldehydes, ketones, alcohols, esters and sesquiterpenes. Some of these compounds are potentially associated with a sweet and fruity flavour while others were rather associated with spiciness and pungency (**Chapter 5**). Further sensory evaluations of accessions are required to evaluate the impact of the presence of these metabolites on the perceived taste and odour of the leaves. Such results could inform breeding programs aimed at selecting cultivars with strong odours for consumers in Benin and Togo, and cultivars with a sweet fruity flavour for East African countries. The claims by Beninese farmers that bitter accessions with a strong odour are healthier could find an explanation based on our

results. Pungency in the species correlated with the presence of sulfur-containing compounds including glucosinolates and isothiocyanates. The bitter taste of the leaves could be triggered by the presence of flavonoids (Roland et al. 2013) and hydroxycinnamic acids (Jiang and Peterson 2010). Volatile metabolites are also involved in plant defense mechanisms and the natural variation in volatile sulfur-containing compounds in the species provided some indication about the accessions to select for future analysis of plant-insect interactions.

Based on the semi-polar and volatile metabolite composition and relative levels in the germplasm collection of 48 accessions, we also found evidence of geographical patterns of clustering of accessions (**Chapter 5**) suggesting a link between genetic diversity and secondary metabolites production. This is a hypothesis which we confirmed in **Chapter 6**. Thus, secondary metabolites can also be used to provide a reliable assessment of the genetic diversity in a germplasm collection, elucidate molecular networks in *G. gynandra* and potentially be used to classify accessions of unknown geographic origin.

Integrative omics for *Gynandropsis gynandra* breeding for nutritional quality and yield

Breeding programs for conventional crops benefit from extensive published knowledge on phenotypes, genome data, gene action and mode of inheritance of key-traits. The lack of such knowledge for orphan crop breeding requires rapid generation of information about the natural variation in traits of interest as well as the genetic mechanisms underlying those traits. As exemplified in this thesis, genome sequencing opens the avenue to diverse investigations including the evolutionary history of the target species/family, rapid marker discovery and genetic mapping of traits of interest (**Chapters 6 and 7**). Moreover, knowledge on closely related well-known species such as *A. thaliana* and *Brassica* crops (i.e. comparative genomics) facilitated the identification of candidate genes in *G. gynandra* which need to be further validated. We also demonstrated that integrating different datasets including phenotypic, metabolic and genomic data provides a helicopter view of the overall variation in traits of interest, the genetic diversity in the studied germplasm and downstream genetic mapping and candidate gene identification. In this thesis, the low sample size of the collection hindered the use of association mapping approaches. To overcome this, we adopted a linkage mapping approach. However, a weakness of this study was the low number of markers used in building the genetic map (**Chapter 7**). A second F_2 population genotyped with a higher marker density (nine thousand SNPs) has been generated but remains to be analysed.

QTL analysis for carotenoid and tocopherol production allowed the identification of candidate genes present at critical steps of these respective pathways (**Chapter 6**). Developing stable populations such as Recombinant Inbred Lines, Doubled Haploids or Back-cross populations could be used to validate the QTLs and to investigate the interactions between QTLs and environment. Gene silencing and gene expression analyses under different conditions should also be considered for QTL validation. Biofortification in crops have been implemented using classical breeding strategies or genetic engineering (Jiang et al. 2017). Whether this latter approach effectively met consumers' requirements and reduced malnutrition in the socio-economic and cultural context of Africa remains debatable (de Valença et al. 2017; Hummel et

al. 2018; Schnurr et al. 2018). Recent developments in genome editing using the CRISPR-Cas9 system provide exciting avenues for crop quality improvement as exemplified with the high-amylose rice (Sun et al. 2017) or the lycopene-rich tomato (Li et al. 2018).

Appropriate breeding strategies following QTL validation should be defined taking into account: (1) farmers' perspectives in terms of other important traits to breed for besides nutrient content; (2) the cost-efficiency of adopting hybrid varieties vs inbred lines and the existing seed systems in the target countries; (3) GMOs and genome editing regulations in target countries; (4) the impact of biofortified varieties on human health; (5) consumers' taste preferences and willingness to pay for the biofortified varieties.

Regarding leaf yield, appropriate phenotyping platforms to accurately monitor plant growth and development would tremendously improve genetic mapping of traits of interest. In our case, we took a snapshot of the populations at one time-point which only partially reflects the total variation between individuals in an F_2 population. The link between flowering time, leaf area and plant height established in the diversity panel and the F_2 population provides the possibility to breed for those traits simultaneously as late-flowering plants are desired in *G. gynandra*. Other traits needing further investigation include branching patterns, internode length as well as the ability to regrow after cutting.

Breeding for orphan crops should go hand in hand with appropriate data management and there is a need for data publication, storage and sharing. The breeding management system (BMS) is an example of platform supporting day-to-day breeding activities for most crops, making it a very attractive option for digitalising breeding programmes at an institutional level (Ribaut and Ragot 2019). Making generated data on orphan crops publicly available will facilitate collaboration between researchers and the generation of relevant information for promotion of those crops among potential end-users. In the case of this thesis, open-access publications and availability of the data fostered new collaborations with other research groups that are further discussed below.

Is there an on-going speciation event occurring in *Gynandropsis*?

Gynandropsis is a monotypic genus in the Cleomaceae family. Differences in morphological traits (Sogbohossou et al. 2019), secondary metabolites (**Chapter 5**), genome-wide nucleotide diversity (**Chapter 6**) and Internal Transcribed Spacer sequences (Reeves et al. 2018) illustrated differences between accessions from East Africa, West Africa and Asia. The strong geographic differentiation between accessions suggested ongoing isolation by distance in the species with post-zygotic reproductive barriers between East/Southern African accessions and West African/Asian accessions, especially when the East/Southern African accessions were used as maternal line (**Chapter 7**). Geographically isolated populations experience divergent selection owing to differences in local abiotic and biotic environments. Adaptations to these distinct environments contribute directly or indirectly to reproductive isolation and speciation (Maron et al. 2019). Whole-genome duplication (**Chapter 6**) could contribute to reproductive isolation and eventually speciation with fixation of one gene copy via genetic drift in one population and neofunctionalization/sub-functionalization of the other copy in the second population (Ouyang and Zhang 2018). Further investigation of the reproductive mechanisms leading to male sterility in hybrids and F_2 individuals coupled with high quality genome assembly of selected accessions

displaying reproductive isolation could help elucidate the incipient speciation.

The roadmap for *Gynandropsis gynandra* breeding: how far have we come?

In **Chapter 2**, we devised a roadmap for orphan crops breeding (**Figure 1**). In this section, we discuss the contribution of this thesis to the implementation of the roadmap, present an overview of on-going efforts at different institutions based on the research results described in this thesis, and provide recommendations for future research.

Contribution of this thesis

At the inception of this thesis, farmers' priorities and consumers' preferences were investigated in the target markets in Benin and Kenya. In Benin, as the species was mainly cultivated and sold in Adja communities and farmers and consumers were the same people. Emphasis was put on late-flowering, resistance against pests and diseases, taste and odour, as well as medicinal properties.

Germplasm collection and documentation in *G. gynandra* was facilitated by the elaboration of a standardized germplasm collection form and a list of morphological descriptors (**Chapter 2**; Sogbohossou et al. 2018a) publicly available for use. The germplasm collection missions conducted in West Africa (Benin, Togo, Ghana and Niger) filled the gap in the region and could be further extended to other countries in the region and other parts of the world such as South America and Australia.

The need to provide farmers with stable cultivars that can be easily propagated led us to consider

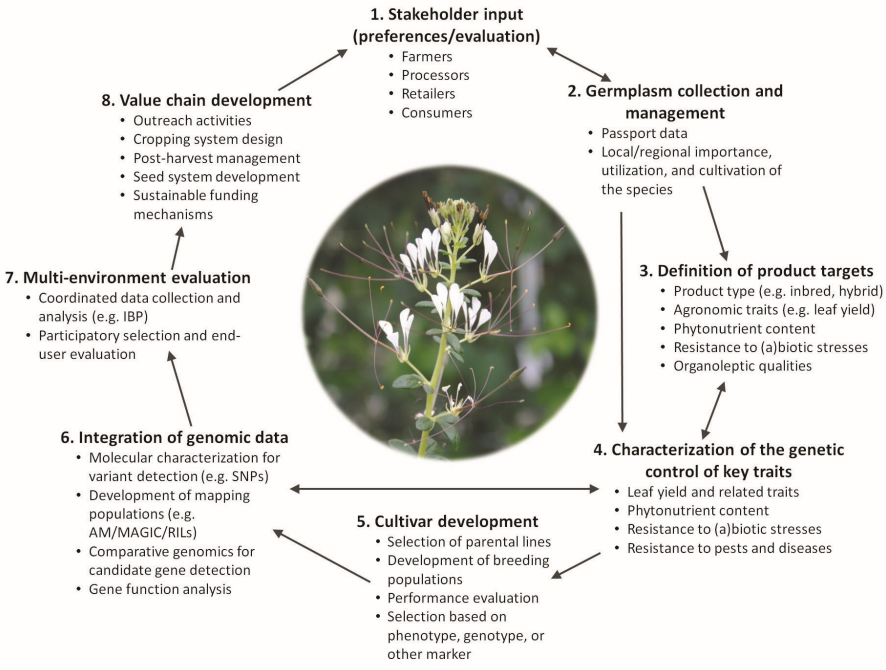


Figure 1. Proposed integrated breeding pipeline for orphan leafy vegetables (Sogbohossou et al. 2018a)

inbred lines as the best product target for promotion of *G. gynandra* (**Chapter 2**; Sogbohossou et al. 2018a). Hybrid cultivars would be an option once the species is well established as a commercial crop and once we better understand the potential genomic incompatibilities between lines from different geographical regions.

The measured variation in leaf-yield related traits and vitamins (carotenoids, tocopherols, ascorbic acid) in the germplasm collection (**Chapter 3**; Sogbohossou et al. 2019) informed the choice of parental lines we used for QTL detection and candidate gene identification (**Chapter 7**). Despite the low number of markers used, linkage mapping yielded preliminary results that can be used as a start point for further QTL validation. However, the sterility issues observed in the F_2 population hindered the development of inbred lines. Further studies could use a doubled haploid approach or prioritize crosses between accessions from the same region.

The draft genome of *G. gynandra* and the resequencing of 53 accessions allowed detection of 10.8 million good quality SNPs (**Chapter 6**) that will facilitate omics-assisted breeding in the species. Genotyping of mapping populations, candidate gene detection and functional genomics are therefore possible. The combination of morphological, metabolic and genomic data pointed to a strong geographic differentiation of accessions with reproductive barriers between accessions from East Africa and West Africa/Asia. The hypothesis of an on-going speciation event in the species could be further tested with a more comprehensive set of accessions to inform future breeding strategies.

Additional research outside of the scope of this thesis, included the testing of different agronomic practices for high yield production (Houdegbe et al. 2018), the description of the pollination biology (Zohoungbogbo et al. 2018) and characterizing natural variation in C_4 photosynthesis-related traits (Reeves et al. 2018) in *G. gynandra*. Many C_4 related traits were variable such as bundle sheath size, vein density, gas-exchange parameters and carbon isotope discrimination, which together suggest natural selection for water use efficiency of the considered accessions (Reeves et al. 2018) and thus potential targets for future crop improvement.

Short/medium term research perspectives: what is being done?

This thesis provides a picture of research achieved after 4 years after the start of our PhD. However, publications from this thesis opened the road to new collaborations with various research teams that are either using *G. gynandra* as a model species for fundamental research or contributing to the development of a breeding program for the species. The activities listed below are based on the germplasm collection presented in this thesis:

At Wageningen University:

- The re-sequencing and Hi-C assembly of the genome of four accessions including the reference line and one accession from each region of interest (East Africa, West Africa and Asia) to improve the quality of the reference genome and detect potential large structural variations;

- A second F_2 population is being genotyped/phenotyped with nine thousands SNPs for the exact same traits described in **Chapter 6**.

At the University of KwaZulu-Natal (South Africa) and the University of Abomey-Calavi (Benin)

- Multi-location trials are on-going in Benin and South Africa using a diallel mating system to assess the heritability of flowering time and other leaf-yield related traits as well as the general and specific combining ability of several *G. gynandra* lines. However, the same issues of reproductive barriers between accessions from different regions (West Africa x East Africa; East Africa x Asia; and East Africa x West Africa) also arose.

- Optimal temperatures for seed germination and long-term storage of *G. gynandra* are being assessed;

- The natural variation in seed structure and mineral content was assessed in the same collection described in **Chapter 4** of this thesis.

At Hibberd Lab, Cambridge University:

- The natural variation in photosynthesis-related traits is being investigated in order to identify candidate genes responsible for specific C_4 phenotypes;

- The spectrum of regulatory networks governing the evolution of C_4 photosynthesis is also being analysed.

Long-term research perspectives: what is needed?

Research gaps remaining to be addressed for *G. gynandra* breeding include but are not limited to: (1) the taxonomic revision of the species; (2) assessing metabolite accumulation during plant growth and development; (3) annotating more secondary metabolites detected in the leaves and resolving the molecular pathways involved in the biosynthesis of such metabolites using genome-wide association mapping or linkage mapping; (4) assessing the pharmacological properties of the novel compounds identified in the species; (5) assessing the impact of harvest, storage and cooking practices on the bioavailability of vitamins and secondary metabolites; (6) investigating relationships between resistance to pests and diseases and glucosinolates content; (7) developing speed breeding protocols for uniform flowering of the accessions and to safeguard late-flowering accessions; (8) developing efficient germination and seed conservation protocols; (9) identifying cost-efficient and accurate genotyping approaches.

Implications of this thesis for development and capacity building

The Cleome project that funded the first two years of this thesis contributed to the promotion of *G. gynandra* among farmers and consumers in urban settings in Benin and Kenya. Scientific evidence of the presence of a wide range of health-promoting compounds in the species coupled with the utilization of the species in local pharmacopeia can strengthen campaigns to raise consumers' awareness on the importance of consuming spider plant and other local leafy vegetables. Farmers' evaluation of identified late-flowering accessions grown in various environments was a first step in the cultivar development process and will contribute to refining breeding goals. Seed companies should also be associated to ensure appropriate seed multiplication, seed germination and distribution to farmers.

Knowledge about the nutrient content of the species is of important value for industries. Production of vitamin C and other health-promoting compounds from a short cycle plant is a phytochemical avenue to embrace. As more genetic information will be generated, it may be possible in the future to directly produce vitamins and other health-promoting compounds from spider plant. The species might also have applications in cosmetics for skin protection and regeneration as indicated by its traditional use to cure skin diseases.

The Cleome project also prompted other complementary research and capacity building actions including the training of six MSc students in Benin, one in Kenya and one in South Africa. Students explored various topics related to *G. gynandra* such as the reproductive biology of the species (Zohoungbogbo et al. 2018), appropriate agronomic practices for cost-effective production and high leaf-yield (Houdegbe et al. 2018), effects of genotypes and pre-treatments on seed germination (Sohindji 2018), the variability in seed morphology and seed mineral content in accessions from diverse origins (Blalogoe 2019), pests and diseases associated with *G. gynandra* and farmers' pest management practices in Benin (Francisco 2018). Moreover, two students in Namibia and in South Africa are currently doing their PhD, one of them pursuing our efforts towards genetic analysis of flowering time and other leaf yield-related traits in a multi-parental population and the second focusing on unravelling the genetics of anthocyanin production in the species. These capacity building actions opened the door to generate more knowledge towards accelerating spider plant breeding to enable adequate cultivar development.

Concluding remarks

This thesis presented a strategy for orphan leafy vegetables breeding with preliminary results related to the genetic variation in leaf-yield related traits and secondary metabolites production in *G. gynandra*. The different analyses at the morphological, metabolic and genomic levels pointed to an on-going speciation event in the species. The draft genome provides a basis for functional genomics of *G. gynandra* and accelerated breeding. The genetic map constructed and the QTLs and candidate genes identified for target traits in breeding programs constitute an important step towards varietal development in *G. gynandra*. The knowledge generated in this thesis raised several research questions related to *G. gynandra* breeding requiring the integration of different fields of expertise including plant and seed physiology, biochemistry, bioinformatics, pharmacology, taxonomy, and plant breeding.

References

Aalto EA, Koelewijn H-P, Savolainen O (2013) Cytoplasmic male sterility contributes to hybrid incompatibility between subspecies of *Arabidopsis lyrata*. G3: Genes|Genomes|Genetics 3:1727-1740. 10.1534/g3.113.007815

Abasse TA, Gouzayé A, Woltering L, Pasternak D (2007) The role of indigenous leafy vegetables on daily diet and rural and urban economy in Niger. Acta Horticulturae 752:35-40. 10.17660/ActaHortic.2007.752.2

Abe M, Kaya H, Watanabe-Taneda A, Shibuta M, Yamaguchi A, Sakamoto T, Kurata T, Ausín I, Araki T, Alonso-Blanco C (2015) FE, a phloem-specific Myb-related protein, promotes flowering through transcriptional activation of FLOWERING LOCUS T and FLOWERING LOCUS T INTERACTING PROTEIN 1. The Plant Journal 83:1059-1068. 10.1111/tpj.12951

Abukutsa-Onyango MO, Kavagi P, Amoke P, Habwe FO (2010) Iron and protein content of priority African indigenous vegetables in the Lake Victoria Basin. Journal of Agricultural Science and Technology 4:29

Achigan-Dako EG, Pasquini MW, Assogba Komlan F, N'danikou S, Yédomonhan H, Dansi A, Ambrose-Oji B (2010) Traditional vegetables in Benin. Institut National des Recherches Agricoles du Bénin, Imprimeries du CENAP, Cotonou

Adeniji OT, Aloyce A (2013) Farmers' participatory identification of horticultural traits: developing breeding objectives for vegetable amaranth in Tanzania. Journal of Crop Improvement 27:309-318. 10.1080/15427528.2013.768318

Afari-Sefa V, Tenkouano A, Ojiewo CO, Keatinge J, Hughes JdA (2012) Vegetable breeding in Africa: constraints, complexity and contributions toward achieving food and nutritional security. Food Security 4:115-127

Aguilera J, de Gálvez MV, Sánchez C, Herrera-Ceballos E (2012) Changes in photoinduced cutaneous erythema with topical application of a combination of vitamins C and E before and after UV exposure. Journal of dermatological science 66:216-220

Ajaiyeoba EO (2000) Phytochemical and antimicrobial studies of *Gynandropsis gynandra* and Buchholzia coriaceae extracts. African Journal of Biomedical Research 3:161-165

Akoègninou A, Van der Burg WJ, Van der Maesen LJG (2006) Flore analytique du Bénin. Backhuys Publishers, Wageningen, The Netherlands

Akram NA, Shafiq F, Ashraf M (2017) Ascorbic Acid-A potential oxidant scavenger and its role in plant development and abiotic stress tolerance. Frontiers in Plant Science 8. 10.3389/fpls.2017.00613

Akula R, Ravishankar GA (2011) Influence of abiotic stress signals on secondary metabolites in plants. Plant Signaling & Behavior 6:1720-1731. 10.4161/psb.6.11.17613

Alam MA, Subhan N, Hossain H, Hossain M, Reza HM, Rahman MM, Ullah MO (2016) Hydroxycinnamic acid derivatives: a potential class of natural compounds for the management of lipid metabolism and obesity. Nutrition & Metabolism 13:27. 10.1186/s12986-016-0080-3

Albuquerque UP (2006) Re-examining hypotheses concerning the use and knowledge of medicinal plants: a study in the Caatinga vegetation of NE Brazil. Journal of Ethnobiology and Ethnomedicine 2:30. https://doi.org/10.1186/1746-4269-2-30

Albuquerque UP, Cruz da Cunha LVF, de Lucena RFP, Alves RRN, Ramos MA, Alencar NL (2014) Methods and techniques used to collect ethnobiological data. In: Albuquerque UP, Cruz da Cunha LVF, de Lucena RFP, Alves RRN (eds) Methods and techniques in ethnobiology and ethnoecology. Springer, New York, pp 15-37. 10.1007/978-1-4614-8636-7_2

Albuquerque UP, Nóbrega Alves RR, Júnior WSF, Santoro FR, Vandebroek I, Albuquerque UP (2016) Urbanization, Modernization, and Nature Knowledge. In: Introduction to Ethnobiology. Springer International Publishing, Cham, pp 251-256. 10.1007/978-3-319-28155-1_37

Alexander DH, Lange K (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12:246. 10.1186/1471-2105-12-246

Allabi AC, Busia K, Ekanmian V, Bakiono F (2011) The use of medicinal plants in self-care in the Agonlin region of Benin. Journal of Ethnopharmacology 133:234-243. http://dx.doi.org/10.1016/j.jep.2010.09.028

Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF (2006) A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nature Protocols 1:2320-2325. 10.1038/nprot.2006.384

Apetz N, Munch G, Govindaraghavan S, Gyengesi E (2014) Natural compounds and plant extracts as therapeutics against chronic inflammation in Alzheimer disease: a translational perspective. CNS & Neurological Disorders - Drug Targets (Formerly Current Drug Targets) 13:1175-1191

Araus JL, Cairns JE (2014) Field high-throughput phenotyping: the new crop breeding frontier. Trends in Plant Science 19:52-61. http://dx.doi.org/10.1016/j.tplants.2013.09.008

Arellanes Y, Casas A, Arellanes A, Vega E, Blancas J, Vallejo M, Torres I, Rangel-Landa S, Moreno AI, Solís L, Pérez-Negrón E (2013) Influence of traditional markets on plant management in the Tehuacán Valley. Journal of Ethnobiology and Ethnomedicine 9:38. https://doi.org/10.1186/1746-4269-9-38

Arora RK (2014) Diversity in underutilized plant species – an Asia-Pacific perspective. Bioversity International, New Delhi, India

Bala A, Chetia P, Dolai N, Khandelwal B, Haldar PK (2014) Cat's whiskers flavonoid attenuated oxidative DNA damage and acute inflammation: its importance in lymphocytes of patients with rheumatoid arthritis. Inflammopharmacology 22:55-61. 10.1007/s10787-013-0193-5

Bala A, Kar B, Haldar PK, Mazumder UK, Bera S (2010) Evaluation of anticancer activity of *Cleome gynandra* on Ehrlich's Ascites carcinoma treated mice. Journal of Ethnopharmacology 129:131-134. http://dx.doi.org/10.1016/j.jep.2010.03.010

Baldermann S, Blagojević L, Frede K, Klopsch R, Neugart S, Neumann A, Ngwene B, Norkeweit J, Schröter D, Schröter A, Schweigert FJ, Wiesner M, Schreiner M (2016) Are neglected plants the food for the future? Critical Reviews in Plant Sciences 35:106-119. 10.1080/07352689.2016.1201399

Bandillo N, Raghavan C, Muyco PA, Sevilla MAL, Lobina IT, Dilla-Ermita CJ, Tung C-W, McCouch S, Thomson M, Mauleon R, Singh RK, Gregorio G, Redoña E, Leung H (2013) Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice 6:11. 10.1186/1939-8433-6-11

Barbieri R, Costa Gomes J, Alercia A, Padulosi S (2014) Agricultural biodiversity in Southern Brazil: integrating efforts for conservation and use of Neglected and Underutilized Species. Sustainability 6:741

Barbosa AM, Brown JA, Jimenez-Valverde A, Real R (2016) modEvA: Model evaluation and analysis. R package version 1.3.2.

Barker MS, Vogel H, Schranz ME (2009) Paleopolyploidy in the Brassicales: analyses of the Cleome transcriptome elucidate the history of genome duplications in Arabidopsis and Other Brassicales. Genome Biology and Evolution 1:391-399. 10.1093/gbe/evp040

Baron EP (2018) Medicinal properties of cannabinoids, terpenes, and flavonoids in Cannabis, and benefits in migraine, headache, and pain: an update on current evidence and Cannabis science. Headache: The Journal of Head and Face Pain 58:1139-1186. 10.1111/head.13345

Bayat S, Schranz ME, Roalson EH, Hall JC (2018) Lessons from Cleomaceae, the sister of Crucifers. Trends in Plant Science 23:808-821. https://doi.org/10.1016/j.tplants.2018.06.010

Bechtold U, Albihlal WS, Lawson T, Fryer MJ, Sparrow PAC, Richard F, Persad R, Bowden L, Hickman R, Martin C, Beynon JL, Buchanan-Wollaston V, Baker NR, Morison JIL, Schöffl F, Ott S, Mullineaux PM (2013) Arabidopsis HEAT SHOCK TRANSCRIPTION FACTORA1b overexpression enhances water productivity, resistance to drought, and infection. Journal of Experimental Botany 64:3467-3481. 10.1093/jxb/ert185

Bell C, Hawthorne S (2008) Ellagic acid, pomegranate and prostate cancer — a mini review. Journal of Pharmacy and Pharmacology 60:139-144. 10.1211/jpp.60.2.0001

Bellomo MG, Fallico B (2007) Anthocyanins, chlorophylls and xanthophylls in pistachio nuts (Pistacia vera) of different geographic origin. Journal of Food Composition and Analysis 20:352-359. https://doi.org/10.1016/j.jfca.2006.04.002

Benor S, Demissew S, Hammer K, Blattner FR (2012) Genetic diversity and relationships in *Corchorus olitorius* (Malvaceae s.l.) inferred from molecular and morphological data. Genetic Resources and Crop Evolution 59:1125-1146. 10.1007/s10722-011-9748-8

Beyene Y, Semagn K, Mugo S, Prasanna BM, Tarekegne A, Gakunga J, Sehabiague P, Meisel B, Oikeh SO, Olsen M, Crossa J (2016) Performance and grain yield stability of maize populations developed using marker-assisted recurrent selection and pedigree selection procedures. Euphytica 208:285-297. 10.1007/s10681-015-1590-1

Birney E, Clamp M, Durbin R (2004) GeneWise and Genomewise. Genome Research 14:988-995. 10.1101/gr.1865504

Blalogoe J (2019) Variability in seed morphology and mineral composition in *Gynandropsis gynandra* (L.) Briq accessions from diverse origins. University of KwaZulu-Natal

Blancas J, Casas A, Pérez-Salicrup D, Caballero J, Vega E (2013) Ecological and socio-cultural factors influencing plant management in Náhuatl communities of the Tehuacán Valley, Mexico. Journal of Ethnobiology and Ethnomedicine 9:39. https://doi.org/10.1186/1746-

4269-9-39

Blancas J, Casas A, Rangel-Landa S, Moreno-Calles A, Torres I, Pérez-Negrón E, Solís L, Delgado-Lemus A, Parra F, Arellanes Y, Caballero J, Cortés L, Lira R, Dávila P (2010) Plant management in the Tehuacán-Cuicatlán valley, Mexico. Economic Botany 64:287-302. 10.1007/s12231-010-9133-0

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2010) Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27:578-579. 10.1093/bioinformatics/btq683

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics 30. 10.1093/bioinformatics/btu170

Bonthala VS, Mayes K, Moreton J, Blythe M, Wright V, May ST, Massawe F, Mayes S, Twycross J (2016) Identification of gene modules associated with low temperatures response in bambara groundnut by network-based analysis. PLOS ONE 11:e0148771. 10.1371/journal.pone.0148771

Bouchabke O, Chang F, Simon M, Voisin R, Pelletier G, Durand-Tardif M (2008) Natural variation in *Arabidopsis thaliana* as a tool for highlighting differential drought responses. PLoS One 3:e1705

Bouché F, Lobet G, Tocquin P, Périlleux C (2016) FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*. Nucleic Acids Research 44:D1167-D1171. 10.1093/nar/gkv1054

Brachi B, Morris GP, Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. Genome Biology 12:232. 10.1186/gb-2011-12-10-232

Bräutigam A, Kajala K, Wullenweber J, Sommer M, Gagneul D, Weber KL, Carr KM, Gowik U, Maß J, Lercher MJ, Westhoff P, Hibberd JM, Weber APM (2011) An mRNA blueprint for C_4 photosynthesis derived from comparative transcriptomics of closely related C_3 and C_4 species. Plant Physiology 155:142-156. 10.1104/pp.110.159442

Briones Alonso E, Cockx L, Swinnen J (2018) Culture and food security. Global Food Security 17:113-127. https://doi.org/10.1016/j.gfs.2018.02.002

Brown AHD (1989) Core collections: a practical approach to genetic resources management. Genome 31:818-824. 10.1139/g89-144

Brown NJ, Parsley K, Hibberd JM (2005) The future of C_4 research – maize, Flaveria or Cleome? Trends in Plant Science 10:215-221. http://dx.doi.org/10.1016/j.tplants.2005.03.003

Cannarozzi G, Plaza-Wüthrich S, Esfeld K, Larti S, Wilson YS, Girma D, de Castro E, Chanyalew S, Blösch R, Farinelli L (2014) Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (*Eragrostis tef*). BMC genomics 15:581

Capuno OB, Gonzaga ZC, Dimabuyu HB, Rom JC Indigenous vegetables for coping with climate change and food security. In, 2015. International Society for Horticultural Science (ISHS), Leuven, Belgium, pp 171-178. 10.17660/ActaHortic.2015.1102.21

Carreno-Quintero N, Bouwmeester HJ, Keurentjes JJB (2013) Genetic analysis of metabolome–phenotype interactions: from model to crop species. Trends in Genetics 29:41-50. http://dx.doi.org/10.1016/j.tig.2012.09.006

Cartea ME, Velasco P (2008) Glucosinolates in Brassica foods: bioavailability in food and significance for human health. Phytochemistry Reviews 7:213-229. 10.1007/s11101-007-9072-2

Cattivelli L, Rizza F, Badeck F-W, Mazzucotelli E, Mastrangelo AM, Francia E, Marè C, Tondelli A, Stanca AM (2008) Drought tolerance improvement in crop plants: An integrated view from breeding to genomics. Field Crops Research 105:1-14. http://dx.doi.org/10.1016/j.fcr.2007.07.004

Cazzonelli CI (2011) Carotenoids in nature: insights from plants and beyond. Functional Plant Biology 38:833-847. https://doi.org/10.1071/FP11192

CENALA (2003) Atlas et études sociolinguistiques du Bénin. Agence Intergouvernementale de la Francophonie, Cotonou, Benin

Cernansky R (2015) The rise of Africa's super vegetables. Nature 522:146-148

Ceuterick M, Vandebroek I, Torry B, Pieroni A (2008) Cross-cultural adaptation in urban ethnobotany: the Colombian folk pharmacopoeia in London. Journal of Ethnopharmacology 120:342-359. https://doi.org/10.1016/j.jep.2008.09.004

Chang Y, Liu H, Liu M, Liao X, Sahu SK, Fu Y, Song B, Cheng S, Kariba R, Muthemba S, Hendre PS, Mayes S, Ho WK, Yssel AEJ, Kendabie P, Wang S, Li L, Muchugi A, Jamnadass R, Lu H, Peng S, Van Deynze A, Simons A, Yana-Shapiro H, Van de Peer Y, Xu X, Yang H, Wang J, Liu X (2018) The draft genomes of five agriculturally important African orphan crops. GigaScience 8. 10.1093/gigascience/giy152

Chanyalew S, Ferede S, Damte T, Fikre T, Genet Y, Kebede W, Tolossa K, Tadele Z, Assefa K (2019) Significance and prospects of an orphan crop tef. Planta. 10.1007/s00425-019-03209-z

Cheng S, van den Bergh E, Zeng P, Zhong X, Xu J, Liu X, Hofberger J, de Bruijn S, Bhide AS, Kuelahoglu C, Bian C, Chen J, Fan G, Kaufmann K, Hall JC, Becker A, Bräutigam A, Weber APM, Shi C, Zheng Z, Li W, Lv M, Tao Y, Wang J, Zou H, Quan Z, Hibberd JM, Zhang G, Zhu X-G, Xu X, Schranz ME (2013) The *Tarenaya hassleriana* genome provides insight into reproductive trait and genome evolution of Crucifers. The Plant Cell 25:2813. 10.1105/tpc.113.113480

Chigumira Ngwerume F, Mvere B, Mhazo M Traditional vegetable improvement project agronomic trials. In: National workshop on traditional vegetables and underutilized crops, Harare, Zimbabwe, 1998.

Chiurugwi T, Kemp S, Powell W, Hickey LT (2019) Speed breeding orphan crops. Theoretical and Applied Genetics 132:607-616. 10.1007/s00122-018-3202-7

Chweya JA, Eyzaguirre PB (1999) The biodiversity of traditional leafy vegetables. IPGRI, Rome, Italy.

Chweya JA, Mnzava NA (1997) Cat's whiskers, *Cleome gynandra* L. Bioversity International,

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6:80-92. 10.4161/fly.19695

Cleome Consortium (2016) Utilizing the genome of *Cleome gynandra* for the development of improved cultivars for West and East African markets - Annual Progress Report 2015.

Cleome Consortium, Cotonou, Benin

Clouse JW, Adhikary D, Page JT, Ramaraj T, Deyholos MK, Udall JA, Fairbanks DJ, Jellen EN, Maughan PJ (2016) The Amaranth Genome: genome, transcriptome, and physical map assembly. The Plant Genome 9. 10.3835/plantgenome2015.07.0062

Combs Jr GF, McClung JP (2017) The vitamins: fundamental aspects in nutrition and health. 5th Edition edn. Elsevier, London, UK

Conti MV, Campanaro A, Coccetti P, De Giuseppe R, Galimberti A, Labra M, Cena H (2019) Potential role of neglected and underutilized plant species in improving women's empowerment and nutrition in areas of sub-Saharan Africa. Nutrition Reviews. 10.1093/nutrit/nuz038

Covshoff S, Szecowka M, Hughes TE, Smith-Unna R, Kelly S, Bailey KJ, Sage TL, Pachebat JA, Leegood R, Hibberd JM (2016) C_4 photosynthesis in the rice paddy: Insights from the noxious weed *Echinochloa glabrescens*. Plant Physiology 170:57-73. 10.1104/pp.15.00889

Cruz MP, Medeiros PM, Sarmiento-Combariza I, Peroni N, Albuquerque UP (2014) "I eat the manofê so it is not forgotten": local perceptions and consumption of native wild edible plants from seasonal dry forests in Brazil. Journal of Ethnobiology and Ethnomedicine 10:45. 10.1186/1746-4269-10-45

Cruz MP, Peroni N, Albuquerque UP (2013) Knowledge, use and management of native wild edible plants from a seasonal dry forest (NE, Brazil). Journal of Ethnobiology and Ethnomedicine 9:1. https://doi.org/10.1186/1746-4269-9-79

Cruz R, Casal S (2013) Validation of a fast and accurate chromatographic method for detailed quantification of vitamin E in green leafy vegetables. Food Chemistry 141:1175-1180. https://doi.org/10.1016/j.foodchem.2013.03.099

Cullis C, Kunert KJ (2017) Unlocking the potential of orphan legumes. Journal of Experimental Botany 68:1895-1903. https://doi.org/10.1093/jxb/erw437

da Silva GS, Canuto KM, Ribeiro PRV, de Brito ES, Nascimento MM, Zocolo GJ, Coutinho JP, de Jesus RM (2017) Chemical profiling of guarana seeds (*Paullinia cupana*) from different geographical origins using UPLC-QTOF-MS combined with chemometrics. Food Research International 102:700-709. https://doi.org/10.1016/j.foodres.2017.09.055

Dandriyal J, Singla R, Kumar M, Jaitak V (2016) Recent developments of C-4 substituted coumarin derivatives as anticancer agents. European Journal of Medicinal Chemistry 119:141-168. https://doi.org/10.1016/j.ejmech.2016.03.087

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Group GPA (2011) The variant call format and VCFtools. Bioinformatics 27:2156-2158. 10.1093/bioinformatics/btr330

de Albuquerque UP (2006) Re-examining hypotheses concerning the use and knowledge of medicinal plants: a study in the Caatinga vegetation of NE Brazil. J Ethnobiol Ethnomed 2. 10.1186/1746-4269-2-30

de Lucena RFP, de Lima Araújo E, de Albuquerque UP (2007) Does the local availability of woody Caatinga plants (Northeastern Brazil) explain their use value? Economic Botany 61:347-361. 10.1663/0013-0001(2007)61[347:Dtlaow]2.0.Co;2

de Valença AW, Bake A, Brouwer ID, Giller KE (2017) Agronomic biofortification of crops to fight hidden hunger in sub-Saharan Africa. Global Food Security 12:8-14. https://doi.org/10.1016/j.gfs.2016.12.001

de Vos RCH, Moco S, Lommen A, Keurentjes JJB, Bino RJ, Hall RD (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. Nature Protocols 2:778

Debieu M, Kanfany G, Laplaze L (2017) Pearl millet genome: lessons from a tough crop. Trends in plant science 22:911-913

DellaPenna D, Pogson BJ (2006) Vitamin synthesis in plants: tocopherols and carotenoids. Annual Review of Plant Biology 57:711-738. https://doi.org/10.1146/annurev.arplant.56.032604.144301

Denton OA, Nwangburuka CC (2012) Morphological diversity among *Corchorus olitorius* accessions based on single linkage cluster analysis and principal component analysis. Jordan Journal of Biological Sciences 5:191-196

Ding Z, Weissmann S, Wang M, Du B, Huang L, Wang L, Tu X, Zhong S, Myers C, Brutnell TP, Sun Q, Li P (2015) Identification of photosynthesis-associated C_4 candidate genes through comparative leaf gradient transcriptome in multiple lineages of C_3 and C_4 species. PLoS ONE 10. 10.1371/journal.pone.0140629

Dinssa FF, Hanson P, Dubois T, Tenkouano A, Stoilova T, Hughes J, Keating JDH (2016) AVRDC—The World Vegetable Center's women-oriented improvement and development strategy for traditional African vegetables in sub-Saharan Africa. European Journal of Horticultural Science 81:91-105

Diouf M, Ba CO Contribution of African Leafy Vegetables to food safety and income generation in Senegal. In, 2014. International Society for Horticultural Science (ISHS), Leuven, Belgium, pp 53-58. 10.17660/ActaHortic.2014.1021.4

Doungous O, Kalendar R, Adiobo A, Schulman AH (2015) Retrotransposon molecular markers resolve cocoyam (*Xanthosoma sagittifolium*) and taro (*Colocasia esculenta*) by type and variety. Euphytica 206:541-554. 10.1007/s10681-015-1537-6

Dovie DBK, Shackleton CM, Witkowski ETF (2007) Conceptualizing the human use of wild edible herbs for conservation in South African communal areas. Journal of Environmental Management 84:146-156. http://dx.doi.org/10.1016/j.jenvman.2006.05.017

Dray S, Dufour A-B (2007) The ade4 package: implementing the duality diagram for ecologists. Journal of statistical software 22:1-20

Drechsel P, Hope L, Cofie OO (2013) Gender mainstreaming: who wins? Gender and irrigated urban vegetable production in West Africa. wH2O: Journal of Gender and Water 2:15-17

Dudley JW, Moll RH (1969) Interpretation and use of estimates of heritability and genetic variances in plant breeding. Crop Science 9:257-262. 10.2135/cropsci1969.0011183X000900030001x

Durvasula A, Fulgione A, Gutaker RM, Alacakaptan SI, Flood PJ, Neto C, Tsuchimatsu T, Burbano HA, Picó FX, Alonso-Blanco C (2017) African genomes illuminate the early history and transition to selfing in *Arabidopsis thaliana*. Proceedings of the National Academy of

Sciences 114:5213-5218

Ekpong B (2009) Effects of seed maturity, seed storage and pre-germination treatments on seed germination of cleome (*Cleome gynandra* L.). Scientia horticulturae 119:236-240

Ekué MRM, Sinsin B, Eyog-Matig O, Finkeldey R (2010) Uses, traditional management, perception of variation and preferences in ackee (*Blighia sapida* K.D. Koenig) fruit traits in Benin: implications for domestication and conservation. Journal of Ethnobiology and Ethnomedicine 6:12. https://doi.org/10.1186/1746-4269-6-12

Englberger L (2012) Revisiting the vitamin A fiasco: going local in Micronesia. Sustainable Diets and Biodiversity FAO, Rome

Erum S, Naeemullah M, Masood S, Qayyum A, Rabbani MA (2012) Genetic divergence in *Amaranthus* collected from Pakistan. Journal of Animal and Plant Sciences 22:653-658

Espinosa-Goded M, Barreiro-Hurlé J, Ruto E (2010) What do farmers want from agri-environmental scheme design? A choice experiment approach. Journal of Agricultural Economics 61:259-273

Esteban R, Moran JF, Becerril JM, García-Plazaola JI (2015) Versatility of carotenoids: An integrated view on diversity, evolution, functional roles and environmental interactions. Environmental and Experimental Botany 119:63-75. https://doi.org/10.1016/j.envexpbot.2015.04.009

Eugeni Piller L, Glauser G, Kessler F, Besagni C (2014) Role of plastoglobules in metabolite repair in the tocopherol redox cycle. Frontiers in Plant Science 5. 10.3389/fpls.2014.00298

Fahey JW, Zalcmann AT, Talalay P (2001) The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. Phytochemistry 56:5-51. http://dx.doi.org/10.1016/S0031-9422(00)00316-2

FAO (1990) Utilization of tropical foods: fruits and leaves. FAO Food and Nutrition Paper 47

FAO (2018) 2018 - The state of Food Security and Nutrition in the world - building climate resilience for food security and nutrition. FAO, Rome

Fehr WR (1987) Principles of cultivar development. Volume 1. Theory and technique. Macmillan Publishing Company, New York

Feng F, Wang Q, Liang C, Yang R, Li X (2015) Enhancement of tocopherols in sweet corn by marker-assisted backcrossing of ZmVTE4. Euphytica 206:513-521. 10.1007/s10681-015-1519-8

Feodorova TA, Voznesenskaya EV, Edwards GE, Roalson EH (2010) Biogeographic patterns of diversification and the origins of C_4 in Cleome (Cleomaceae). Systematic Botany 35:811-826. https://doi.org/10.1600/036364410X539880

Fiedor J, Burda K (2014) Potential role of carotenoids as antioxidants in human health and disease. Nutrients 6:466-488. https://doi.org/10.3390/nu6020466

Floress K, Reimer A, Thompson A, Burbach M, Knutson C, Prokopy L, Ribaudo M, Ulrich-Schad J (2018) Measuring farmer conservation behaviors: challenges and best practices. Land Use Policy 70:414-418. https://doi.org/10.1016/j.landusepol.2017.11.030

Flyman MV, Afolayan AJ (2006) A survey of plants used as wild vegetables in four districts of Botswana. Ecology of food and nutrition 45:405-415. https://doi.

org/10.1080/03670240600985431

Francis RM (2017) pophelper: an R package and web app to analyse and visualize population structure. Molecular Ecology Resources 17:27-32. 10.1111/1755-0998.12509

Francisco R (2018) Monitoring pests and diseases affecting *Gynandropsis gynandra* (L.) Briq. (Cleomaceae) and farmers' management practices in Benin and Togo (West Africa). University of Abomey-Calavi

Fritsche-Neto R, Borém A, de Morais AA, Pinheiro JB (2012) Breeding for resistance to insect pests. In: Plant Breeding for Biotic Stress Resistance. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 103-125. 10.1007/978-3-642-33087-2_6

Gahukar RT (2014) Potential of minor food crops and wild plants for nutritional security in the Developing World. Journal of Agricultural & Food Information 15:342-352. 10.1080/10496505.2014.952429

Galili T (2015) dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics 31:3718-3720

Gao Z-J, Liu J-B, Xiao X-G (2011) Purification and characterisation of polyphenol oxidase from leaves of *Cleome gynandra* L. Food Chemistry 129:1012-1018. http://dx.doi.org/10.1016/j.foodchem.2011.05.062

Gaoue OG, Coe MA, Bond M, Hart G, Seyler BC, McMillen H (2017) Theories and major hypotheses in Ethnobotany. Economic Botany 71:269-287. 10.1007/s12231-017-9389-8

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:12073907

GBIF (2019) *Gynandropsis gynandra* Briq. GBIF Backbone Taxonomy. https://www.gbif.org/species/3054227. Accessed 2019-07-11

Georgian E, Emshwiller E (2013) Shared and separate knowledge among eight cultural groups based on ethnobotanical uses of *Rhododendron* (Ericaceae) in Yunnan Province, China. Economic Botany 67:191-202. 10.1007/s12231-013-9235-6

Gerrano AS, Adebola PO, Jansen van Rensburg WS, Laurie SM (2015a) Genetic variability in cowpea (*Vigna unguiculata* (L.) Walp.) genotypes. South African Journal of Plant and Soil 32:165-174. 10.1080/02571862.2015.1014435

Gerrano AS, Van Rensburg WSJ, Adebola PO (2015b) Genetic diversity of *Amaranthus* species in South Africa. South African Journal of Plant and Soil 32:39-46. 10.1080/02571862.2014.973069

Ghimire SK, McKey D, Aumeeruddy-Thomas Y (2004) Heterogeneity in ethnoecological knowledge and management of medicinal plants in the Himalayas of Nepal: implications for conservation. Ecology and Society 9:6. http://www.ecologyandsociety.org/vol9/iss3/art6/

Ghogare UR, Nirmal SA, Patil RY, Kharya MD (2009) Antinociceptive activity of *Gynandropsis gynandra* leaves. Natural Product Research 23:327-333. 10.1080/14786410802047862

Ghorbani A, Langenberger G, Sauerborn J (2012) A comparison of the wild food plant use knowledge of ethnic minorities in Naban River Watershed National Nature Reserve,

Yunnan, SW China. Journal of Ethnobiology and Ethnomedicine 8:17. https://doi.org/10.1186/1746-4269-8-17

Ghosh RK, Sreewongchai T, Nakasathien S, Phumichai C (2013) Phenotypic variation and the relationships among jute (*Corchorus* species) genotypes using morpho-agronomic traits and multivariate analysis. Australian Journal of Crop Science 7:830

Glaszmann JC, Kilian B, Upadhyaya HD, Varshney RK (2010) Accessing genetic diversity for crop improvement. Current Opinion in Plant Biology 13:167-173. http://dx.doi.org/10.1016/j.pbi.2010.01.004

Gockowski J, Mbazo'o J, Mbah G, Fouda Moulende T (2003) African traditional leafy vegetables and the urban and peri-urban poor. Food Policy 28:221-235

Godoy R, Reyes-Garcia V, Broesch J, Fitzpatrick IC, Giovannini P, Rodriguez MRM, Huanca T, Leonard WR, McDade TW, Tanner S (2009) Long-term (secular) change of ethnobotanical knowledge of useful plants: separating cohort and age effects. Journal of Anthropological Research 65:51-67. 10.3998/jar.0521004.0065.105

González-Insuasti MS, Caballero J (2007) Managing plant resources: how intensive can it be? Human Ecology 35:303-314. 10.1007/s10745-006-9063-8

González-Insuasti MS, Casas A, Méndez-Ramírez I, Martorell C, Caballero J (2011) Intracultural differences in the importance of plant resources and their impact on management intensification in the Tehuacán Valley, Mexico. Human Ecology 39:191-202

González-Insuasti MS, Martorell C, Caballero J (2008) Factors that influence the intensity of non-agricultural management of plant resources. Agroforestry Systems 74:1-15. 10.1007/s10457-008-9148-z

Gowele VF, Kinabo J, Jumbe T, Kirschmann C, Frank J, Stuetz W (2019) Provitamin A carotenoids, tocopherols, ascorbic acid and minerals in Indigenous Leafy Vegetables from Tanzania. Foods 8:35

Granado-Lorencio F, Olmedilla-Alonso B, Herrero-Barbudo C, Blanco-Navarro I, Pérez-Sacristán B, Blázquez-García S (2007) In vitro bioaccessibility of carotenoids and tocopherols from fruits and vegetables. Food Chemistry 102:641-648. https://doi.org/10.1016/j.foodchem.2006.05.043

Graßes T, Grimm B, Koroleva O, Jahns P (2001) Loss of α-tocopherol in tobacco plants with decreased geranylgeranyl reductase activity does not modify photosynthesis in optimal growth conditions but increases sensitivity to high-light stress. Planta 213:620-628. 10.1007/s004250100532

Grubben GJH, Denton OA (2004) Plant Resources of Tropical Africa 2. Vegetables. PROTA Foundation, Wageningen, Netherlands/Backhuys Publishers, Netherlands/CTA, Wageningen, Netherlands.

Grubben GJH, Denton OA, Grubben GJH (2004a) *Amaranthus cruentus* L. In: Plant Resources of Tropical Africa 2. Vegetables. Backhuys Publishers, Netherlands/CTA, Wageningen, Netherlands, pp 67-72

Grubben GJH, Denton OA, Mnzava NA, Chigumira Ngwerume F (2004b) *Cleome gynandra* L. In: Plant Resources of Tropical Africa 2. Vegetables. PROTA Foundation, Wageningen,

Netherlands/Backhuys Publishers, Netherlands/CTA, Wageningen, Netherlands, pp 191-195

Guo Z, Tucker DM, Lu J, Kishore V, Gay G (2012) Evaluation of genome-wide selection efficiency in maize nested association mapping populations. Theoretical and Applied Genetics 124:261-275. 10.1007/s00122-011-1702-9

Hall JC (2008) Systematics of Capparaceae and Cleomaceae: an evaluation of the generic delimitations of *Capparis* and *Cleome* using plastid DNA sequence data. Botany 86:682-696. doi:10.1139/B08-026

Hall JC, Sytsma KJ, Iltis HH (2002) Phylogeny of Capparaceae and Brassicaceae based on chloroplast sequence data. American Journal of Botany 89:1826-1842. 10.3732/ajb.89.11.1826

Hall NM, Griffiths H, Corlett JA, Jones HG, Lynn J, King GJ (2005) Relationships between water-use traits and photosynthesis in *Brassica oleracea* resolved by quantitative genetic analysis. Plant Breeding 124:557-564. 10.1111/j.1439-0523.2005.01164.x

Hall RD (2018) Plant metabolomics in a nutshell: potential and future challenges. In: Annual Plant Reviews online. pp 1-24. 10.1002/9781119312994.apr0461

Hart G, Gaoue OG, de la Torre L, Navarrete H, Muriel P, Macía MJ, Balslev H, León-Yánez S, Jørgensen P, Duffy DC (2017) Availability, diversification and versatility explain human selection of introduced plants in Ecuadorian traditional medicine. PLoS ONE 12. https://doi.org/10.1371/journal.pone.0184369

Heimann L, Horst I, Perduns R, Dreesen B, Offermann S, Peterhansel C (2013) A common histone modification code on C_4 genes in maize and its conservation in *Sorghum* and *Setaria italica*. Plant Physiology 162:456-469. 10.1104/pp.113.216721

Hendre PS, Muthemba S, Kariba R, Muchugi A, Fu Y, Chang Y, Song B, Liu H, Liu M, Liao X, Sahu SK, Wang S, Li L, Lu H, Peng S, Cheng S, Xu X, Yang H, Wang J, Liu X, Simons A, Shapiro H-Y, Mumm RH, Van Deynze A, Jamnadass R (2019) African Orphan Crops Consortium (AOCC): status of developing genomic resources for African orphan crops. Planta. 10.1007/s00425-019-03156-9

Hofberger JA, Lyons E, Edger PP, Chris Pires J, Eric Schranz M (2013) Whole genome and tandem duplicate retention facilitated glucosinolate pathway diversification in the mustard family. Genome biology and evolution 5:2155-2173

Houdegbe CA, Sogbohossou EOD, Achigan-Dako EG (2018) Enhancing growth and leaf yield in *Gynandropsis gynandra* (L.) Briq. (Cleomaceae) using agronomic practices to accelerate crop domestication. Scientia Horticulturae 233:90-98. https://doi.org/10.1016/j.scienta.2018.01.035

Hounkpati BCC (1991) A comparative phonology of Gbe. Foris Publications, Berlin

Hu T, Banzhaf W Nonsynonymous to synonymous substitution ratio (ka/ks): measurement for rate of evolution in evolutionary computation. In: International Conference on Parallel Problem Solving from Nature, 2008. Springer, pp 448-457

Huang BE, Verbyla KL, Verbyla AP, Raghavan C, Singh VK, Gaur P, Leung H, Varshney RK, Cavanagh CR (2015) MAGIC populations in crops: current status and future prospects. Theoretical and Applied Genetics 128:999-1017. 10.1007/s00122-015-2506-0

Hummel M, Talsma EF, Van der Honing A, Gama AC, Van Vugt D, Brouwer ID, Spillane C (2018) Sensory and cultural acceptability tradeoffs with nutritional content of biofortified orange-fleshed sweetpotato varieties among households with children in Malawi. PLOS ONE 13:e0204754

Husson F, Josse J, Le S, Mazet J (2013) FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R.

Huynh B-L, Close TJ, Roberts PA, Hu Z, Wanamaker S, Lucas MR, Chiulele R, Cissé N, David A, Hearne S, Fatokun C, Diop NN, Ehlers JD (2013) Gene pools and the genetic architecture of domesticated cowpea. The Plant Genome 6. 10.3835/plantgenome2013.03.0005

Iftekharuddaula KM, Ahmed HU, Ghosal S, Amin A, Moni ZR, Ray BP, Barman HN, Siddique MA, Collard BCY, Septiningsih EM (2016) Development of early maturing submergence-tolerant rice varieties for Bangladesh. Field Crops Research 190:44-53. http://dx.doi.org/10.1016/j.fcr.2015.12.001

Illumina (2013) bcl2fastq Conversion User Guide. San Diego, California

Iltis HH (1960) Studies in the Capparidaceae—VII old world Cleomes adventive in the new world. Brittonia 12:279-294

Iltis HH, Hall JC, Cochrane TS, Sytsma KJ (2011) Studies in the Cleomaceae I. On the separate recognition of Capparaceae, Cleomaceae, and Brassicaceae. Annals of the Missouri Botanical Garden 98:28-36. 10.3417/2007017

Imanirampa L, Alele PE (2016) Antifungal activity of *Cleome gynandra* L. aerial parts for topical treatment of *Tinea capitis*: an in vitro evaluation. BMC complementary and alternative medicine 16:194-194. 10.1186/s12906-016-1187-9

Inda L, Torrecilla P, Catalán P, Ruiz-Zapata T (2008) Phylogeny of *Cleome* L. and its close relatives *Podandrogyne* Ducke and *Polanisia* Raf. (Cleomoideae, Cleomaceae) based on analysis of nuclear ITS sequences and morphology. Plant Systematics and Evolution 274:111-126. 10.1007/s00606-008-0026-y

Jacobsen S-E, Mujica A, Jensen C (2003) The resistance of quinoa (*Chenopodium quinoa* Willd.) to adverse abiotic factors. Food Reviews International 19:99-109

Jacobsen S-E, Sørensen M, Pedersen SM, Weiner J (2015) Using our agrobiodiversity: plant-based solutions to feed the world. Agronomy for Sustainable Development 35:1217-1235. 10.1007/s13593-015-0325-y

Jansen van Rensburg WS, Zulu NL, Gerano AS, Adebola PO Seed production of African leafy vegetables: some experiences. In, 2015. International Society for Horticultural Science (ISHS), Leuven, Belgium, pp 121-126. 10.17660/ActaHortic.2015.1102.14

Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJA, Ohyanagi H, Mineta K, Michell CT, Saber N, Kharbatia NM, Rupper RR, Sharp AR, Dally N, Boughton BA, Woo YH, Gao G, Schijlen EGWM, Guo X, Momin AA, Negrão S, Al-Babili S, Gehring C, Roessner U, Jung C, Murphy K, Arold ST, Gojobori T, Linden CGvd, van Loo EN, Jellen EN, Maughan PJ, Tester M (2017) The genome of *Chenopodium quinoa*. Nature 542:307-312. 10.1038/nature21370

Jeruto P, Lukhoba C, Ouma G, Otieno D, Mutai C (2008) An ethnobotanical study of medicinal plants used by the Nandi people in Kenya. Journal of Ethnopharmacology 116:370-376. http://dx.doi.org/10.1016/j.jep.2007.11.041

Jiang D, Peterson DG (2010) Role of hydroxycinnamic acids in food flavor: a brief overview. Phytochemistry Reviews 9:187-193. 10.1007/s11101-009-9159-z

Jiang L, Wang W, Lian T, Zhang C (2017) Manipulation of metabolic pathways to develop vitamin-enriched crops for human health. Frontiers in Plant Science 8. 10.3389/fpls.2017.00937

Jiang Q (2014) Natural forms of vitamin E: metabolism, antioxidant, and anti-inflammatory activities and their role in disease prevention and therapy. Free Radical Biology and Medicine 72:76-90. https://doi.org/10.1016/j.freeradbiomed.2014.03.035

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473:97. 10.1038/nature09916

Jiménez-Aguilar DM, Grusak MA (2015) Evaluation of minerals, phytochemical compounds and antioxidant activity of Mexican, Central American, and African green leafy vegetables. Plant Foods for Human Nutrition 70:357-364. https://doi.org/10.1007/s11130-015-0512-7

Jung C, Müller AE (2009) Flowering time control and applications in plant breeding. Trends in Plant Science 14:563-573. https://doi.org/10.1016/j.tplants.2009.07.005

Júnior WSF, Santoro FR, Vandebroek I, Albuquerque UP (2016) Urbanization, modernization, and nature knowledge. In: Albuquerque UP, Nóbrega Alves RR (eds) Introduction to ethnobiology. Springer International Publishing, Switzerland, pp 251-256. 10.1007/978-3-319-28155-1_37

K'opondo F, Muasya R, Kiplagat O A review on the seed production and handling of indigenous vegetables (Spiderplant, Jute mallow and African nightshade complex). In: Proceedings of the third Horticulture Workshop on Sustainable Horticultural production in the Tropics. Maseno University, Maseno, 2005. pp 42-48

Kahane R, Hodgkin T, Jaenicke H, Hoogendoorn C, Hermann M, Hughes JdA, Padulosi S, Looney N (2013) Agrobiodiversity for food security, health and income. Agronomy for sustainable development 33:671-693

Kapsoot E, Mwangi M, Kamau A Repellence and toxicity effect of crude plant extracts on the two-spotted spider mite *Tetranychus urticae* on roses. In, 2015. International Society for Horticultural Science (ISHS), Leuven, Belgium, pp 155-164. 10.17660/ActaHortic.2015.1077.16

Kengni E, Tchouanguep MF, Mbofung CMF, Tchoundjeu Z, Asaah KE (2003) Intraspecific variation in morphological, sensory and nutritional characteristics of *Dacryodes edulis* fruits in West and Central Africa. Plant Genetic Resources 1:135-141. https://doi.org/10.1079/PGR2003020

Khoury CK, Bjorkman AD, Dempewolf H, Ramirez-Villegas J, Guarino L, Jarvis A, Rieseberg LH, Struik PC (2014) Increasing homogeneity in global food supplies and the implications for food security. Proceedings of the National Academy of Sciences 111:4001-

4006. 10.1073/pnas.1313490111

Kidane B, van der Maesen LJG, Asfaw Z, Sosef MSM, van Andel T (2015) Wild and semi-wild leafy vegetables used by the Maale and Ari ethnic communities in southern Ethiopia. Genetic Resources and Crop Evolution 62:221-234. 10.1007/s10722-014-0147-9

Kim J-E, Cheng KM, Craft NE, Hamberger B, Douglas CJ (2010) Over-expression of *Arabidopsis thaliana* carotenoid hydroxylases individually and in combination with a β-carotene ketolase provides insight into in vivo functions. Phytochemistry 71:168-178. https://doi.org/10.1016/j.phytochem.2009.10.011

Kim JK, Choi SR, Lee J, Park S-Y, Song SY, Na J, Kim SW, Kim S-J, Nou I-S, Lee YH, Park SU, Kim H (2013) Metabolic differentiation of Diamondback Moth (*Plutella xylostella* (L.)) resistance in cabbage (*Brassica oleracea* L. ssp. *capitata*). Journal of Agricultural and Food Chemistry 61:11222-11230. 10.1021/jf403441t

Kimiywe J, Waudo J, Mbithe D, Maundu P (2007) Utilization and medicinal value of indigenous leafy vegetables consumed in urban and peri-urban Nairobi. African Journal of Food, Agriculture, Nutrition and Development 7:1-15

Kleessen S, Antonio C, Sulpice R, Laitinen R, Fernie AR, Stitt M, Nikoloski Z (2012) Structured patterns in geographic variability of metabolic phenotypes in *Arabidopsis thaliana*. Nature Communications 3. https://doi.org/10.1038/ncomms2333

Knepper C, Mou B (2015) Semi-high throughput screening for potential drought-tolerance in lettuce (*Lactuca sativa*) germplasm collections. Journal of Visualized Experiments 2015. 10.3791/52492

Kocacinar F (2015) Photosynthetic, hydraulic and biomass properties in closely related C_3 and C_4 species. Physiologia Plantarum 153:454-466. 10.1111/ppl.12240

Korf I (2004) Gene finding in novel genomes. BMC Bioinformatics 5:59. 10.1186/1471-2105-5-59

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9, 29. doi:10.1186/1746-4811-9-29

Koteyeva NK, Voznesenskaya EV, Edwards GE (2015) An assessment of the capacity for phosphoenolpyruvate carboxykinase to contribute to C_4 photosynthesis. Plant Science 235:70-80. http://dx.doi.org/10.1016/j.plantsci.2015.03.004

Koteyeva NK, Voznesenskaya EV, Roalson EH, Edwards GE (2011) Diversity in forms of C_4 in the genus Cleome (Cleomaceae). Annals of Botany 107:269-283. 10.1093/aob/mcq239

Kujawska M, Pardo-de-Santayana M (2015) Management of medicinally useful plants by European migrants in South America. Journal of Ethnopharmacology 172:347-355. https://doi.org/10.1016/j.jep.2015.06.037

Külahoglu C, Denton AK, Sommer M, Maß J, Schliesky S, Wrobel TJ, Berckmans B, Gongora-Castillo E, Robin Buell C, Simon R, De Veylder L, Bräutigam A, Weber APM (2014) Comparative transcriptome atlases reveal altered gene expression modules between two Cleomaceae C_3 and C_4 plant species. Plant Cell 26:3243-3260. 10.1105/tpc.114.123752

Kulwal, P. L. 2018. Trait mapping approaches through linkage mapping in plants. In: Plant Genetics and Molecular Biology, eds. Varshney, R. K., M. K. Pandey, and A. Chitikineni, 53-

82. Cham: Springer International Publishing.

Kumar RA, Oldenburg DJ, Bendich AJ (2015) Molecular integrity of chloroplast DNA and mitochondrial DNA in mesophyll and bundle sheath cells of maize. Planta 241:1221-1230. 10.1007/s00425-015-2253-0

Kumar S, Sehgal SK, Kumar U, Prasad PVV, Joshi AK, Gill BS (2012) Genomic characterization of drought tolerance-related traits in spring wheat. Euphytica 186:265-276. 10.1007/s10681-012-0675-3

Kumssa DB, Joy EJM, Young SD, Odee DW, Ander EL, Broadley MR (2017) Variation in the mineral element concentration of *Moringa oleifera* Lam. and M. stenopetala (Bak. f.) Cuf.: Role in human nutrition. PLOS ONE 12:e0175503. https://doi.org/10.1371/journal.pone.0175503

Kunchge N, Kumar K, Firke P (2012) Vegetable crops (chili pepper and onion): approaches to improve crop productivity and abiotic stress tolerance. Improving Crop Resistance to Abiotic Stress, Volume 1 & Volume 2:951-978

Kunyanga CN, Imungi JK, Okoth MW, Biesalski HK, Vadivel V (2012) Total phenolic content, antioxidant and antidiabetic properties of methanolic extract of raw and traditionally processed Kenyan indigenous food ingredients. LWT - Food Science and Technology 45:269-276. http://dx.doi.org/10.1016/j.lwt.2011.08.006

Kwarteng AO, Abogoom J, Adu Amoah R, Nyadanu D, Nyam CK, Ghunney T, Awuah E, Ziyaaba JZ, Ogunsanya JO, Orhin EE, Asiedu DD (2018) Phenomic characterization of twenty-four accessions of spider plant (*Cleome gynandra* L.) the Upper East region of Ghana. Scientia Horticulturae 235:124-131. https://doi.org/10.1016/j.scienta.2018.02.046

Lafon-Placette C, Köhler C (2015) Epigenetic mechanisms of postzygotic reproductive isolation in plants. Current Opinion in Plant Biology 23:39-44. https://doi.org/10.1016/j.pbi.2014.10.006

Landete JM (2011) Ellagitannins, ellagic acid and their derived metabolites: A review about source, metabolism, functions and health. Food Research International 44:1150-1160. https://doi.org/10.1016/j.foodres.2011.04.027

Landoh DE, Ouro-Kavalah F, Yaya I, Kahn A-L, Wasswa P, Lacle A, Nassoury DI, Gitta SN, Soura AB (2016) Predictors of incomplete immunization coverage among one to five years old children in Togo. BMC Public Health 16:968. https://doi.org/10.1186/s12889-016-3625-5

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357

Langridge P, Reynolds MP (2015) Genomic tools to assist breeding for drought tolerance. Current Opinion in Biotechnology 32:130-135. http://dx.doi.org/10.1016/j.copbio.2014.11.027

LaRochelle S, Berkes F (2003) Traditional ecological knowledge and practice for edible wild plants: biodiversity use by the Rarámuri, in the Sirerra Tarahumara, Mexico. International Journal of Sustainable Development & World Ecology 10:361-375. 10.1080/13504500309470112

Le Cao K-A, Rohart F, Gonzalez I, Dejean S (2019) Omics Data Integration Project. R package version 662

Leal ML, Alves RP, Hanazaki N (2018) Knowledge, use, and disuse of unconventional food plants. Journal of Ethnobiology and Ethnomedicine 14:6. https://doi.org/10.1186/s13002-018-0209-8

Lee J, Oh M, Park H, Lee I (2008) SOC1 translocated to the nucleus by interaction with AGL24 directly regulates LEAFY. The Plant Journal 55:832-843. 10.1111/j.1365-313X.2008.03552.x

Lenné J, Wood D (2011) Agrobiodiversity management for food security: a critical review. CAB International, Wallingford, UK. 10.1079/9781845937614.0000

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079

Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J, Wang J (2010) De novo assembly of human genomes with massively parallel short read sequencing. Genome Research 20:265-272. 10.1101/gr.097261.109

Li X, Wang Y, Chen S, Tian H, Fu D, Zhu B, Luo Y, Zhu H (2018) Lycopene is enriched in tomato fruit by CRISPR/Cas9-mediated multiplex genome editing. Frontiers in Plant Science 9. 10.3389/fpls.2018.00559

Li Y-x, Li C, Bradbury PJ, Liu X, Lu F, Romay CM, Glaubitz JC, Wu X, Peng B, Shi Y, Song Y, Zhang D, Buckler ES, Zhang Z, Li Y, Wang T (2016) Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. The Plant Journal 86:391-402. 10.1111/tpj.13174

Liang Y, Zhang F, Wang J, Joshi T, Wang Y, Xu D (2011) Prediction of drought-resistant genes in *Arabidopsis thaliana* using SVM-RFE. PLoS one 6:e21750

Lindenbaum P, Redon R (2017) bioalcidae, samjs and vcffilterjs: object-oriented formatters and filters for bioinformatics files. Bioinformatics 34:1224-1225. 10.1093/bioinformatics/btx734

Lipka AE, Lu F, Cherney JH, Buckler ES, Casler MD, Costich DE (2014) Accelerating the Switchgrass (Panicum virgatum L.) breeding cycle using genomic selection approaches. PLoS ONE 9:e112227. 10.1371/journal.pone.0112227

Liu L, Liu C, Hou X, Xi W, Shen L, Tao Z, Wang Y, Yu H (2012) FTIP1 is an essential regulator required for Florigen transport. PLOS Biology 10:e1001313

Lo D, Wang Hsin I, Wu W, Yang R (2018) Anti-nutrient components and their concentrations in edible parts in vegetable families. CAB Reviews 13:1-30. 10.1079/PAVSNNR201813015

Lopes MS, Rebetzke GJ, Reynolds M (2014) Integration of phenotyping and genetic platforms for a better understanding of wheat performance under drought. Journal of Experimental Botany 65:6167-6177. 10.1093/jxb/eru384

Łuczaj Ł, Köhler P, Pirożnikow E, Graniszewska M, Pieroni A, Gervasi T (2013) Wild

edible plants of Belarus: from Rostafiński's questionnaire of 1883 to the present. Journal of Ethnobiology and Ethnomedicine 9:21. https://doi.org/10.1186/1746-4269-9-21

Łuczaj Ł, Pieroni A, Tardío J, Pardo-de-Santayana M, Sõukand R, Svanberg I, Kalle R (2012) Wild food plant use in 21st century Europe: the disappearance of old traditions and the search for new cuisines involving wild edibles. Acta Societatis Botanicorum Poloniae 81:359-370

Lwande W, Ndakala AJ, Hassanali A, Moreka L, Nyandat E, Ndungu M, Amiani H, Gitu PM, Malonza MM, Punyua DK (1999) *Gynandropsis gynandra* essential oil and its constituents as tick (*Rhipicephalus appendiculatus*) repellents. Phytochemistry 50:401-405. http://dx.doi.org/10.1016/S0031-9422(98)00507-X

Lyons E, Pedersen B, Kane J, Freeling M (2008) The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the rosids. Tropical Plant Biology 1:181-190. 10.1007/s12042-008-9017-y

Ma CH, Chu JZ, Shi XF, Liu CQ, Yao XQ (2016) Effects of enhanced UV-B radiation on the nutritional and active ingredient contents during the floral development of medicinal chrysanthemum. Journal of Photochemistry and Photobiology B: Biology 158:228-234. https://doi.org/10.1016/j.jphotobiol.2016.02.019

Ma X, Qiao Z, Chen D, Yang W, Zhou R, Zhang W, Wang M (2015) CYCLIN-DEPENDENT KINASE G2 regulates salinity stress response and salt mediated flowering in Arabidopsis thaliana. Plant Molecular Biology 88:287-299. 10.1007/s11103-015-0324-z

Mackey A, Liu Q, Pereira F, Roos D (2005) GLEAN-Improved eukaryotic gene prediction by statistical consensus of gene evidence. Genome Informatics 8:R13

Majoros WH, Pertea M, Salzberg SL (2004) TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20:2878-2879. 10.1093/bioinformatics/bth315

Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U (2014) The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*. eLife 3:e02478. 10.7554/eLife.02478

Malonza MM, Dipeolu OO, Amoo AO, Hassan SM (1992) Laboratory and field observations on anti-tick properties of the plant *Gynandropsis gynandra* (L.) brig. Veterinary Parasitology 42:123-136. http://dx.doi.org/10.1016/0304-4017(92)90108-L

Manoko MLK, Van Den Berg RG, Feron RMC, Van Der Weerden GM, Mariani C (2008) Genetic diversity of the African hexaploid species *Solanum scabrum* Mill. and *Solanum nigrum* L. (Solanaceae). Genetic Resources and Crop Evolution 55:409-418. 10.1007/s10722-007-9248-z

Maron JL, Agrawal AA, Schemske DW (2019) Plant–herbivore coevolution and plant speciation. Ecology 100:e02704. 10.1002/ecy.2704

Marshall DM, Muhaidat R, Brown NJ, Liu Z, Stanley S, Griffiths H, Sage RF, Hibberd JM (2007) *Cleome*, a genus closely related to *Arabidopsis*, contains species spanning a developmental progression from C_3 to C_4 photosynthesis. The Plant Journal 51:886–896. doi: 10.1111/j.1365-313X.2007.03188.x

Masetti O, Ciampa A, Nisini L, Sequi P, Dell'Abate MT (2017) A multifactorial approach in characterizing geographical origin of Sicilian cherry tomatoes using 1H-NMR profiling. Food Research International 100:623-630. https://doi.org/10.1016/j.foodres.2017.07.059

Massman JM, Jung H-JG, Bernardo R (2013) Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in Maize. Crop Science 53:58-66. 10.2135/cropsci2012.02.0112

Mathez-Stiefel S-L, Vandebroek I (2012) Distribution and transmission of medicinal plant knowledge in the Andean highlands: a case study from Peru and Bolivia. Evidence-Based Complementary and Alternative Medicine 2012:18. http://dx.doi.org/10.1155/2012/959285

Matro X, Sogbohossou EOD, Achigan-Dako EG, Maundu P, Van Deynze A, Schranz ME, Solberg S, Deguenon ES (in prep.) Contribution of spider plant (*Gynandropsis gynandra*) home gardens to household food security and income in Southern Benin

Maughan P, Smith S, Fairbanks D, Jellen E (2011) Development, characterization, and linkage mapping of single nucleotide polymorphisms in the grain amaranths (*Amaranthus* sp.). The plant genome 4:92-101

Maughan PJ, Yourstone SM, Jellen EN, Udall JA (2009) SNP discovery via genomic reduction, barcoding, and 454-pyrosequencing in amaranth. The plant genome 2:260-270

Maundu P, Achigan-Dako EG, Morimoto Y (2009) Biodiversity of African vegetables. In: Shackleton CM, Pasquini MW, Drescher AW (eds) African Indigenous Vegetables in Urban Agriculture. Earthscan, London, UK, pp 65-104

Maurya DK, Devasagayam TPA (2010) Antioxidant and prooxidant nature of hydroxycinnamic acid derivatives ferulic and caffeic acids. Food and Chemical Toxicology 48:3369-3373. https://doi.org/10.1016/j.fct.2010.09.006

Mayes S, Massawe FJ, Alderson PG, Roberts JA, Azam-Ali SN, Hermann M (2012) The potential for underutilized crops to improve security of food production. Journal of Experimental Botany 63:1075-1079. 10.1093/jxb/err396

McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, Kennedy M, Amirebrahimi M, Weers BD, McKinley B, Mattison A, Morishige DT, Grimwood J, Schmutz J, Mullet JE (2018) The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. The Plant Journal 93:338-354. 10.1111/tpj.13781

Medeiros PMd, Soldati GT, Alencar NL, Vandebroek I, Pieroni A, Hanazaki N, de Albuquerque UP (2012) The use of medicinal plants by migrant people: adaptation, maintenance, and replacement. Evidence-Based Complementary and Alternative Medicine 2012:11. http://dx.doi.org/10.1155/2012/807452

Melo ATO, Bartaula R, Hale I (2016) GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. BMC Bioinformatics 17:29. 10.1186/s12859-016-0879-y

Menendez-Baceta G, Pardo-de-Santayana M, Aceituno-Mata L, Tardío J, Reyes-García V (2017) Trends in wild food plants uses in Gorbeialdea (Basque Country). Appetite 112:9-16. http://dx.doi.org/10.1016/j.appet.2017.01.010

Mensah J, Okoli R, Ohaju-Obodo J, Eifediyi K (2008) Phytochemical, nutritional and medical properties of some leafy vegetables consumed by Edo people of Nigeria. African Journal of Biotechnology 7

Meyer-Rochow VB (2009) Food taboos: their origins and purposes. Journal of Ethnobiology and Ethnomedicine 5:18. 10.1186/1746-4269-5-18

Mi X, Utz HF, Technow F, Melchinger AE (2014) Optimizing Resource Allocation for Multistage Selection in Plant Breeding with R Package Selectiongain. Crop Science 54:1413-1418. 10.2135/cropsci2013.10.0699

Millett J, Foot GW, Thompson JC, Svensson BM (2018) Geographic variation in Sundew (Drosera) leaf colour: plant–plant interactions counteract expected effects of abiotic factors. Journal of Biogeography 45:582-592. https://doi.org/10.1111/jbi.13141

Milner SG, Maccaferri M, Huang BE, Mantovani P, Massi A, Frascaroli E, Tuberosa R, Salvi S (2015) A multiparental cross population for mapping QTL for agronomic traits in durum wheat (*Triticum turgidum* ssp. *durum*). Plant biotechnology journal 14: 735-748. doi:10.1111/pbi.12424

Mir RR, Rustgi S, Sharma S, Singh R, Goyal A, Kumar J, Gaur A, Tyagi AK, Khan H, Sinha MK, Balyan HS, Gupta PK (2008) A preliminary genetic analysis of fibre traits and the use of new genomic SSRs for genetic diversity in jute. Euphytica 161:413-427. 10.1007/s10681-007-9597-x

Miyazawa T, Nakagawa K, Sookwong P (2011) Health benefits of vitamin E in grains, cereals and green vegetables. Trends in Food Science & Technology 22:651-654. https://doi.org/10.1016/j.tifs.2011.07.004

Mnzava NA, Chigumira Ngwerume F (2004) Cleome gynandra L. In: Grubben GJH, Denton OA (eds) Plant Resources of Tropical Africa 2. Vegetables. PROTA Foundation, Wageningen, Netherlands/Backhuys Publishers, Netherlands/CTA, Wageningen, Netherlands, pp 191-195

Mohammadin S, Nguyen T-P, van Weij MS, Reichelt M, Schranz ME (2017) Flowering Locus C (FLC) is a potential major regulator of glucosinolate content across developmental stages of *Aethionema arabicum* (Brassicaceae). Frontiers in Plant Science 8. 10.3389/fpls.2017.00876

Monteiro JM, Albuquerque UPd, Lins-Neto EMdF, Araújo ELd, Amorim ELCd (2006) Use patterns and knowledge of medicinal species among two rural communities in Brazil's semi-arid northeastern region. Journal of Ethnopharmacology 105:173-186. http://dx.doi.org/10.1016/j.jep.2005.10.016

Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st Century crop improvement. Plant Physiology 147:969-977. 10.1104/pp.108.118232

Motsa MM, Slabbert MM, van Averbeke W, Morey L (2015) Effect of light and temperature on seed germination of selected African leafy vegetables. South African Journal of Botany 99:29-35. http://dx.doi.org/10.1016/j.sajb.2015.03.185

Moyo M, Amoo SO, Aremu AO, Gruz J, Šubrtová M, Jarošová M, Tarkowski P, Doležal K (2018) Determination of mineral constituents, phytochemicals and antioxidant qualities of *Cleome gynandra*, compared to *Brassica oleracea* and *Beta vulgaris*. Frontiers in Chemistry 5. 10.3389/fchem.2017.00128

Moyo M, Amoo SO, Ncube B, Ndhlala AR, Finnie JF, Van Staden J (2013) Phytochemical and antioxidant properties of unconventional leafy vegetables consumed in southern Africa. South African Journal of Botany 84:65–71. http://dx.doi.org/10.1016/j.sajb.2012.09.010

Muasya RM, Simiyu J, Muui C, Rao N, Dulloo M, Gohole L (2009) Overcoming seed dormancy in *Cleome gynandra* L. to improve germination. Seed Technology:134-143

Muchuweti M, Mupure C, Ndhlala A, Murenje T, Benhura MAN (2007) Screening of antioxidant and radical scavenging activity of *Vigna ungiculata*, *Bidens pilosa* and *Cleome gynandra*. American Journal of Food Technology 2:161-168

Muhaidat R, Sage RF, Dengler NG (2007) Diversity of Kranz anatomy and biochemistry in C_4 eudicots. American Journal of Botany 94:362-381. 10.3732/ajb.94.3.362

Muhammad A (2016) The derivational processes in the Zarma numeral system In: Ndimele O, Chan ESL (eds) The numeral systems of Nigerian languages. M&J Orbit Communications, Port Harcourt, Nigeria, pp 295-302

Muhanji G, Roothaert RL, Webo C, Stanley M (2011) African indigenous vegetable enterprises and market access for small-scale farmers in East Africa. International Journal of Agricultural Sustainability 9:194-202. 10.3763/ijas.2010.0561

N'Danikou S, Achigan-Dako EG, Wong J (2011) Eliciting local values of wild edible plants in Southern Benin to identify priority species for conservation. Economic Botany 65:381-395. 10.1007/s12231-011-9178-8

N'Danikou S, Achigan-Dako EG, Tchokponhoue DA, Agossou COA, Houdegbe CA, Vodouhe RS, Ahanchede A (2015) Modelling socioeconomic determinants for cultivation and in-situ conservation of *Vitex doniana* Sweet (Black plum), a wild harvested economic plant in Benin. Journal of Ethnobiology and Ethnomedicine 11:28. https://doi.org/10.1186/s13002-015-0017-3

Namukobe J, Kasenene JM, Kiremire BT, Byamukama R, Kamatenesi-Mugisha M, Krief S, Dumontet V, Kabasa JD (2011) Traditional plants used for medicinal purposes by local communities around the Northern sector of Kibale National Park, Uganda. Journal of Ethnopharmacology 136:236–245. doi:10.1016/j.jep.2011.04.044

Nelson T (2011) The grass leaf developmental gradient as a platform for a systems understanding of the anatomical specialization of C_4 leaves. Journal of Experimental Botany 62:3039-3048. 10.1093/jxb/err072

Neugart S, Baldermann S, Ngwene B, Wesonga J, Schreiner M (2017) Indigenous leafy vegetables of Eastern Africa — A source of extraordinary secondary plant metabolites. Food Research International 100:411-422. http://dx.doi.org/10.1016/j.foodres.2017.02.014

Nijveldt RJ, van Nood E, van Hoorn DEC, Boelens PG, van Norren K, van Leeuwen PAM (2001) Flavonoids: a review of probable mechanisms of action and potential applications. The American Journal of Clinical Nutrition 74:418-425

Nikooei M, Fathipour Y, Javaran MJ, Soufbaf M (2015) How different genetically manipulated *Brassica* genotypes affect life table parameters of *Plutella xylostella* (Lepidoptera: Plutellidae). Journal of Economic Entomology 108:515-524. 10.1093/jee/tov018

Nisar N, Li L, Lu S, Khin NC, Pogson BJ (2015) Carotenoid metabolism in plants. Molecular

Plant 8:68-82. 10.1016/j.molp.2014.12.007

Nooraee N, Molenberghs G, van den Heuvel ER (2014) GEE for longitudinal ordinal data: comparing R-geepack, R-multgee, R-repolr, SAS-GENMOD, SPSS-GENLIN. Computational Statistics & Data Analysis 77:70-83. http://dx.doi.org/10.1016/j.csda.2014.03.009

Nyalala S, Grout BWW (2015) Volatile emissions from actively-growing *Gynandropsis gynandra* and *Cleome hasseleriana* to control spider mites in protected rose cultivation. In: VI International Symposium on Rose Research and Cultivation 1064, 2015. pp 299-302

Nyalala SO, Petersen MA, Grout BWW (2013) Volatile compounds from leaves of the African spider plant (*Gynandropsis gynandra*) with bioactivity against spider mite (*Tetranychus urticae*). Annals of Applied Biology 162:290-298. 10.1111/aab.12021

Ochuodho J, Modi A (2005) Temperature and light requirements for the germination of *Cleome gynandra* seeds. South African Journal of Plant and Soil 22:49-54

Odhav B, Beekrum S, Akula U, Baijnath H (2007a) Preliminary assessment of nutritional value of traditional leafy vegetables in KwaZulu-Natal, South Africa. Journal of Food Composition and Analysis 20:430-435

Odhav B, Beekrum S, Akula US, Baijnath H (2007b) Preliminary assessment of nutritional value of traditional leafy vegetables in KwaZulu-Natal, South Africa. Journal of Food Composition and Analysis 20:430-435. https://doi.org/10.1016/j.jfca.2006.04.015

Odong TL, Jansen J, van Eeuwijk FA, van Hintum TJL (2013) Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. Theoretical and Applied Genetics 126:289-305. 10.1007/s00122-012-1971-y

Olabode IA, Adetula OA, Akinwumi GS, Layade AA (2017) Marketing analysis of indigenous leafy vegetables in the Tropics. International Journal of Vegetable Science 23:226-232. 10.1080/19315260.2016.1236055

Oluoch MO, Pichop G.N., Silué D., Abukutsa-Onyango M.O., Diouf M., C.M. S (2009) Production and harvesting systems for African Indigenous Vegetable. In: Shackleton CM, Pasquini, M.W., Drescher A. (ed) African indigenous vegetables in urban agriculture. Earthscan, London, UK, pp 145-175

Omondi CO (1990) Variation and yield prediction analyses of some morphological traits in six Kenya landraces population of spider flower (*Gynandropsis gynandra* (L.) Briq.). University of Nairobi

Omondi EO, Debener T, Linde M, Abukutsa-Onyango M, Dinssa FF, Winkelmann T (2016) Molecular markers for genetic diversity studies in African leafy vegetables. Advances in Bioscience and Biotechnology 7 (2016), Nr 3 7:188-197

Omondi EO, Debener T, Linde M, Abukutsa-Onyango M, Dinssa FF, Winkelmann T (2017a) Mating biology, nuclear DNA content and genetic diversity in spider plant (*Cleome gynandra*) germplasm from various African countries. Plant Breeding 10.1111/pbr.12485

Omondi EO, Engels C, Nambafu G, Schreiner M, Neugart S, Abukutsa-Onyango M, Winkelmann T (2017b) Nutritional compound analysis and morphological characterization of spider plant (*Cleome gynandra*) - an African indigenous leafy vegetable. Food Research

International 100:284-295. http://dx.doi.org/10.1016/j.foodres.2017.06.050

Ong HG, Kim Y-D (2017) The role of wild edible plants in household food security among transitioning hunter-gatherers: evidence from the Philippines. Food Security 9:11-24

Onyango C, Kunyanga C, Ontita E, Narla R, Kimenju J Production, utilisation and indigenous knowledge of spider plant in Kenya. In: African Crop Science Conference Proceedings, 2013a. pp 925-930

Onyango CM, Kunyanga CN, Ontita EG, Narla RD, Kimenju JW (2013b) Current status on production and utilization of spider plant (*Cleome gynandra* L.) an underutilized leafy vegetable in Kenya. Genetic Resources and Crop Evolution 60:2183–2189. DOI 10.1007/s10722-013-0036-7

Onyango CM, Kunyanga CN, Ontita EG, Narla RD, Kimenju JW (2013c) Production, utilisation and indigenous knowledge of spider plant in Kenya. African Crop Science Conference Proceedings 11:925-930

Oryema C, Ziraba RB, Omagor N, Opio A (2010) Medicinal plants of Erute county, Lira district, Uganda with particular reference to their conservation. African Journal of Ecology 48:285–298. doi: 10.1111/j.1365-2028.2009.01147.x

Ouyang Y, Zhang Q (2018) The molecular and evolutionary basis of reproductive isolation in plants. Journal of Genetics and Genomics 45:613-620. https://doi.org/10.1016/j.jgg.2018.10.004

Ozturk M, Hakeem KR, Ashraf M, Ahmad MSA (2018) Global perspectives on underutilized crops. Springer.

Padulosi S, Amaya K, Jäger M, Gotor E, Rojas W, Valdivia R (2014) A Holistic Approach to Enhance the Use of Neglected and Underutilized Species: The Case of Andean Grains in Bolivia and Peru. Sustainability 6:1283

Palermo M, Pellegrini N, Fogliano V (2014) The effect of cooking on the phytochemical content of vegetables. Journal of the Science of Food and Agriculture 94:1057-1070. 10.1002/jsfa.6478

Palla P, Armano G (2016) Multivariate analysis of metabolomics data using Random Forests. R package version 101

Palumbo JC, Prabhaker N, Reed DA, Perring TM, Castle SJ, Huang T-I (2015) Susceptibility of *Bagrada hilaris* (Hemiptera: Pentatomidae) to insecticides in laboratory and greenhouse bioassays. Journal of Economic Entomology 108:672-682. 10.1093/jee/tov010

Parsons NR, Costa ML, Achten J, Stallard N (2009) Repeated measures proportional odds logistic regression analysis of ordinal score data in the statistical software package R. Computational Statistics & Data Analysis 53:632-641. http://dx.doi.org/10.1016/j.csda.2008.08.004

Pasquini MW, Assogba-Komlan F, Vorster HJ, Shackleton CM, Abukuysa-Onyango MO (2009) The production of African indigenous vegetables in urban and peri-urban agriculture: a comparative analysis of case studies from Benin, Kenya and South Africa. In: African indigenous vegetables in urban and peri-urban agriculture. Earthscan, London, UK, pp 177-224

Pasquini MW, Young EM, Chadha M, Kuo G, Gowda CLL (2007) Networking to promote the sustainable production and marketing of indigenous vegetables through urban and peri-urban agriculture in sub-Saharan Africa (IndigenoVeg).

Patchell MJ, Roalson EH, Hall JC (2014) Resolved phylogeny of Cleomaceae based on all three genomes. Taxon 63:315-328. 10.12705/632.17

Pawera L, Łuczaj Ł, Pieroni A, Polesny Z (2017) Traditional plant knowledge in the White Carpathians: ethnobotany of wild food plants and crop wild relatives in the Czech Republic. Human Ecology 45:655-671. 10.1007/s10745-017-9938-x

Pereyra-Irujo GA, Gasco ED, Peirone LS, Aguirrezábal LAN (2012) GlyPh: A low-cost platform for phenotyping plant growth and water use. Functional Plant Biology 39:905-913. 10.1071/fp12052

Petricka JJ, Clay NK, Nelson TM (2008) Vein patterning screens and the defectively organized tributaries mutants in *Arabidopsis thaliana*. The Plant Journal 56:251-263. 10.1111/j.1365-313X.2008.03595.x

Pieroni A (2001) Evaluation of the cultural significance of wild food botanicals traditionally consumed in Northwestern Tuscany, Italy. Journal of Ethnobiology 21:89-104

Pinela J, Carvalho AM, Ferreira ICFR (2017) Wild edible plants: Nutritional and toxicological characteristics, retrieval strategies and importance for today's society. Food and Chemical Toxicology 110:165-188. https://doi.org/10.1016/j.fct.2017.10.020

Pino Del Carpio D, Basnet RK, De Vos RCH, Maliepaard C, Paulo MJ, Bonnema G (2011) Comparative methods for association studies: a case study on metabolite variation in a *Brassica rapa* core collection. PLOS ONE 6:e19624. 10.1371/journal.pone.0019624

Pirker H, Haselmair R, Kuhn E, Schunko C, Vogl CR (2012) Transformation of traditional knowledge of medicinal plants: the case of Tyroleans (Austria) who migrated to Australia, Brazil and Peru. Journal of Ethnobiology and Ethnomedicine 8:44. https://doi.org/10.1186/1746-4269-8-44

Platt A, Vilhjálmsson BJ, Nordborg M (2010) Conditions under which genome-wide association studies will be positively misleading. Genetics 186:1045-1052. 10.1534/genetics.110.121665

Plazas M, Andújar I, Vilanova S, Gramazio P, Herraiz FJ, Prohens J (2014) Conventional and phenomics characterization provides insight into the diversity and relationships of hypervariable scarlet (*Solanum aethiopicum* L.) and gboma (*S. macrocarpon* L.) eggplant complexes. Frontiers in Plant Science 5:318. 10.3389/fpls.2014.00318

Prota (2010) Promising African Plants. CTA Editions, Wageningen, Netherlands

Prunier R, Akman M, Kremer CT, Aitken N, Chuah A, Borevitz J, Holsinger KE (2017) Isolation by distance and isolation by environment contribute to population differentiation in *Protea repens* (Proteaceae L.), a widespread South African species. American Journal of Botany 104:674-684. https://doi.org/10.3732/ajb.1600232

Pucher A, Sy O, Angarawai II, Gondah J, Zangre R, Ouedraogo M, Sanogo MD, Boureima S, Hash CT, Haussmann BIG (2015) Agro-morphological characterization of West and Central African pearl millet accessions. Crop Science 55:737-748. https://doi.org/10.2135/

cropsci2014.06.0450

Quinlan MB, Quinlan RJ (2007) Modernization and medicinal plant knowledge in a Caribbean horticultural village. Medical Anthropology Quarterly 21:169-192. 10.1525/maq.2007.21.2.169

Raju AJS, Rani DS (2016) Reproductive ecology of *Cleome gynandra* and *Cleome viscosa* (Capparaceae). Phytologia Balcanica 22:15-28

RBG Kew (2016) The State of the World's Plants Report - 2016. Royal Botanic Gardens, Kew.

Reeves G, Singh P, Rossberg TA, Sogbohossou D, Schranz E, Hibberd J (2018) Quantitative variation within a species for traits underpinning C_4 photosynthesis. Plant Physiology 177:504-512. www.plantphysiol.org/cgi/doi/10.1104/pp.18.00168

Reyes-García V, Huanca T, Vadez V, Leonard W, Wilkie D (2006) Cultural, practical, and economic value of wild plants: a quantitative study in the Bolivian Amazon. Economic Botany 60:62-74. 10.1663/0013-0001(2006)60

Reyes-Garcia V, Vadez V, Byron E, Apaza L, Leonard WR, Perez E, Wilkie D (2005) Market economy and the loss of folk knowledge of plant uses: estimates from the Tsimane' of the Bolivian Amazon. Current Anthropology 46:651-656. https://doi.org/10.1086/432777

Reyes-García V, Vadez V, Huanca T, Leonard W, Wilkie D (2005) Knowledge and consumption of wild plants: a comparative study in two Tsimane'villages in the Bolivian Amazon. Ethnobotany Research & Applications 3:201-208

Ribaut J-M, Ragot M (2019) Modernising breeding for orphan crops: tools, methodologies, and beyond. Planta 250:971-977. 10.1007/s00425-019-03200-8

Riewe D, Koohi M, Lisec J, Pfeiffer M, Lippmann R, Schmeichel J, Willmitzer L, Altmann T (2012) A tyrosine aminotransferase involved in tocopherol synthesis in *Arabidopsis*. The Plant Journal 71:850-859. 10.1111/j.1365-313X.2012.05035.x

Ripley B (2016) tree: classification and regression trees. R package version 1.0-37.

Rodríguez GR, Moyseenko JB, Robbins MD, Huarachi Morejón N, Francis DM, van der Knaap E (2010) Tomato Analyser: A useful software application to collect accurate and detailed morphological and colorimetric data from two-dimensional objects. JoVE:e1856. doi:10.3791/1856

Roland WSU, van Buren L, Gruppen H, Driesse M, Gouka RJ, Smit G, Vincken J-P (2013) Bitter taste receptor activation by flavonoids and isoflavonoids: modeled structural requirements for activation of hTAS2R14 and hTAS2R39. Journal of Agricultural and Food Chemistry 61:10454-10466. 10.1021/jf403387p

Romojaro A, Botella MÁ, Obón C, Pretel MT (2013) Nutritional and antioxidant properties of wild edible plants and their use as potential ingredients in the modern diet. International Journal of Food Sciences and Nutrition 64:944-952. 10.3109/09637486.2013.821695

Roy A, Bandyopadhyay A, Mahapatra AK, Ghosh SK, Singh NK, Bansal KC, Koundal KR, Mohapatra T (2006) Evaluation of genetic diversity in jute (*Corchorus* species) using STMS, ISSR and RAPD markers. Plant Breeding 125:292-297. 10.1111/j.1439-0523.2006.01208.x

Ruiz-Carrasco K, Antognoni F, Coulibaly AK, Lizardi S, Covarrubias A, Martínez EA,

Molina-Montenegro MA, Biondi S, Zurita-Silva A (2011) Variation in salinity tolerance of four lowland genotypes of quinoa (*Chenopodium quinoa* Willd.) as assessed by growth, physiological traits, and sodium transporter gene expression. Plant Physiology and Biochemistry 49:1333-1341. https://doi.org/10.1016/j.plaphy.2011.08.005

Sage RF, Stata M (2015) Photosynthetic diversity meets biodiversity: The C_4 plant example. Journal of Plant Physiology 172:104-119. http://dx.doi.org/10.1016/j.jplph.2014.07.024

Saito K, Matsuda F (2010) Metabolomics for functional genomics, systems biology, and biotechnology. Annual Review of Plant Biology 61:463-489. 10.1146/annurev.arplant.043008.092035

Sallam A, Martsch R (2015) Association mapping for frost tolerance using multi-parent advanced generation inter-cross (MAGIC) population in faba bean (*Vicia faba* L.). Genetica 143:501-514. 10.1007/s10709-015-9848-z

Sannemann W, Huang BE, Mathew B, Léon J (2015) Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. Molecular Breeding 35:1-16. 10.1007/s11032-015-0284-7

Sansanelli S, Ferri M, Salinitro M, Tassoni A (2017) Ethnobotanical survey of wild food plants traditionally collected and consumed in the Middle Agri Valley (Basilicata region, southern Italy). Journal of Ethnobiology and Ethnomedicine 13:50. https://doi.org/10.1186/s13002-017-0177-4

Sarker U, Islam MT, Rabbani MG, Oba S (2014) Genotypic variability for nutrient, antioxidant, yield and yield contributing traits in vegetable amaranth. J Food Agric Environ 12:168-174

Sarker U, Islam MT, Rabbani MG, Oba S (2015) Variability, heritability and genetic association in vegetable amaranth (*Amaranthus tricolor* L.). Spanish Journal of Agricultural Research 13:1-8. 10.5424/sjar/2015132-6843

Sattler SE, Gilliland LU, Magallanes-Lundback M, Pollard M, DellaPenna D (2004) Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. The Plant Cell 16:1419-1432. 10.1105/tpc.021360

Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D, Selbig J, Pleban T, Zamir D, Fernie AR (2008) Mode of inheritance of primary metabolic traits in Tomato. The Plant Cell 20:509-523. 10.1105/tpc.107.056523

Schmidt M, Kollers S, Maasberg-Prelle A, Großer J, Schinkel B, Tomerius A, Graner A, Korzun V (2016) Prediction of malting quality traits in barley based on genome-wide marker data to assess the potential of genomic selection. Theoretical and Applied Genetics 129:203-213. 10.1007/s00122-015-2639-1

Schnurr MA, Addison L, Mujabi-Mujuzi S (2018) Limits to biofortification: farmer perspectives on a vitamin A enriched Banana in Uganda. The Journal of Peasant Studies:1-20. 10.1080/03066150.2018.1534834

Schönfeldt HC, Pretorius B (2011) The nutrient content of five traditional South African dark green leafy vegetables—A preliminary study. Journal of Food Composition and Analysis 24:1141-1146. http://dx.doi.org/10.1016/j.jfca.2011.04.004

Schranz ME, Mitchell-Olds T (2006) Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. The Plant Cell Online 18:1152-1165

Schranz ME, Mohammadin S, Edger PP (2012) Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. Current Opinion in Plant Biology 15:147-153. https://doi.org/10.1016/j.pbi.2012.03.011

Schulp CJE, Thuiller W, Verburg PH (2014) Wild food in Europe: a synthesis of knowledge and data of terrestrial wild food as an ecosystem service. Ecological Economics 105:292-305. https://doi.org/10.1016/j.ecolecon.2014.06.018

Schulthess AW, Wang Y, Miedaner T, Wilde P, Reif JC, Zhao Y (2016) Multiple-trait- and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. Theoretical and Applied Genetics 129:273-287. 10.1007/s00122-015-2626-6

Schulz N, Breustedt G, Latacz-Lohmann U (2014) Assessing farmers' willingness to accept "greening": insights from a discrete choice experiment in Germany. Journal of Agricultural Economics 65:26-48. doi:10.1111/1477-9552.12044

Schulze S, Mallmann J, Burscheidt J, Koczor M, Streubel M, Bauwe H, Gowik U, Westhoff P (2013) Evolution of C_4 photosynthesis in the genus *Flaveria*: establishment of a photorespiratory CO_2 pump. The Plant Cell 25:2522-2535. 10.1105/tpc.113.114520

Scossa F, Brotman Y, de Abreu e Lima F, Willmitzer L, Nikoloski Z, Tohge T, Fernie AR (2016) Genomics-based strategies for the use of natural variation in the improvement of crop metabolism. Plant Science 242:47-64. https://doi.org/10.1016/j.plantsci.2015.05.021

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre G-P, Bank C, Brännström Å, Brelsford A, Clarkson CS, Eroukhmanoff F, Feder JL, Fischer MC, Foote AD, Franchini P, Jiggins CD, Jones FC, Lindholm AK, Lucek K, Maan ME, Marques DA, Martin SH, Matthews B, Meier JI, Möst M, Nachman MW, Nonaka E, Rennison DJ, Schwarzer J, Watson ET, Westram AM, Widmer A (2014) Genomics and the origin of species. Nature Reviews Genetics 15:176. 10.1038/nrg3644

Segnon AC, Achigan-Dako EG (2014) Comparative analysis of diversity and utilization of edible plants in arid and semi-arid areas in Benin. Journal of Ethnobiology and Ethnomedicine 10:80. https://doi.org/10.1186/1746-4269-10-80

Shanmugam S, Rajendran K, Suresh K (2012) Traditional uses of medicinal plants among the rural people in Sivagangai district of Tamil Nadu, Southern India. Asian Pacific Journal of Tropical Biomedicine 2:S429-S434. http://dx.doi.org/10.1016/S2221-1691(12)60201-9

Shayganni E, Bahmani M, Asgary S, Rafieian-Kopaei M (2016) Inflammaging and cardiovascular disease: management by medicinal plants. Phytomedicine 23:1119-1126. https://doi.org/10.1016/j.phymed.2015.11.004

Shepherd M, Savins D, Dowell A, Morrow S, Allen G, Southwell I (2017) Ecotype variation of methyl eugenol content in tea tree (*Melaleuca alternifolia* and *Melaleuca linariifolia*). Chem Biodiversity 14:e1700278. https://doi.org/10.1002/cbdv.201700278

Shukla S, Bhargava A, Chatterjee A, Srivastava A, Singh S (2006) Genotypic variability in vegetable amaranth (*Amaranthus tricolor* L) for foliage yield and its contributing traits over successive cuttings and years. Euphytica 151:103-110

Slifer SH (2018) PLINK: key functions for data analysis. Current Protocols in Human Genetics 97:e59. 10.1002/cphg.59

Smith FI, Eyzaguirre P African leafy vegetables: their role in the World Health Organization's global fruit and vegetables initiative. In: Oniang'o R, Grum M, Obel-Lawson E (eds) Developing African leafy vegetables for improved nutrition. Regional workshop, 6-9 December 2005, Nairobi, Kenya, 2007. Rural Outreach Program, pp 1-8

Snapp S (2002) Quantifying farmer evaluation of technologies: The mother and baby trial design. Quantitative analysis of data from participatory methods in plant breeding 9

Soares DTN, Sfair JC, Reyes-García V, Baldauf C (2017) Plant knowledge and current uses of woody flora in three cultural groups of the Brazilian semiarid region: does culture matter? Economic Botany 71:314-329. 10.1007/s12231-017-9393-z

Sogbohossou EOD, Achigan-Dako EG, Maundu P, Solberg S, Deguenon EMS, Mumm RH, Hale I, Van Deynze A, Schranz ME (2018a) A roadmap for breeding orphan leafy vegetable species: a case study of *Gynandropsis gynandra* (Cleomaceae). Horticulture Research 5:2. https://doi.org/10.1038/s41438-017-0001-2

Sogbohossou EOD, Achigan-Dako EG, van Andel T, Schranz ME (2018b) Drivers of management of spider plant (*Gynandropsis gynandra*) across different socio-linguistic groups in Benin and Togo. Economic Botany 72:411-435. 10.1007/s12231-018-9423-5

Sogbohossou EOD, Kortekaas D, Achigan-Dako EG, Maundu P, Stoilova T, Van Deynze A, de Vos RCH, Schranz ME (2019) Association between vitamin content, plant morphology and geographical origin in a worldwide collection of the orphan crop *Gynandropsis gynandra* (Cleomaceae). Planta. 10.1007/s00425-019-03142-1

Sogbohossou OED, Achigan-Dako EG (2014) Phenetic differentiation and use-type delimitation in *Amaranthus* spp. from worldwide origins. Scientia Horticulturae 178:31-42. http://dx.doi.org/10.1016/j.scienta.2014.08.003

Sohindji F (2018) Effect of genotypes and seed pre-treatments on germination of spider plant (*Gynandropsis gynandra* (L). Briq.). University of Abomey-Calavi

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, de Pamphilis CW, Wall PK, Soltis PS (2009) Polyploidy and angiosperm diversification. American Journal of Botany 96:336-348. 10.3732/ajb.0800079

Songsak T, Lockwood GB (2002) Glucosinolates of seven medicinal plants from Thailand. Fitoterapia 73:209-216. 10.1016/s0367-326x(02)00061-8

Soroka J, Grenkow L (2013) Susceptibility of Brassicaceous plants to feeding by Flea Beetles, *Phyllotreta* spp. (Coleoptera: Chrysomelidae). Journal of Economic Entomology 106:2557-2567. 10.1603/ec13102

Spriggs A, Henderson S, Hand M, Johnson S, Taylor J, Koltunow A (2018) Assembled genomic and tissue-specific transcriptomic data resources for two genetically distinct lines of Cowpea (*Vigna unguiculata* (L.) Walp) [version 2; peer review: 3 approved]. Gates Open Research 2. 10.12688/gatesopenres.12777.2

Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Research 33:W465-W467.

10.1093/nar/gki458

Steyn NP, Olivier J, Winter P, Burger S, Nesamvuni C (2001a) A survey of wild, green, leafy vegetables and their potential in combating micronutrient deficiencies in rural populations : research in action. 97:p.276-278

Steyn NP, Olivier J, Winter P, Burger S, Nesamvuni C (2001b) A survey of wild, green, leafy vegetables and their potential in combating micronutrient deficiencies in rural populations : research in action. 97:p.276-278

Stoilova T, Dinssa FF, Ebert AW, Tenkouano A (2015) The diversity of African leafy vegetables: Agromorphological characterization of subsets of AVRDC's germplasm collection vol 1102. 10.17660/ActaHortic.2015.1102.7

Sun L-M, Zhang J-Z, Hu C-G (2016) Characterization and expression analysis of PtAGL24, a SHORT VEGETATIVE PHASE/AGAMOUS-LIKE 24 (SVP/AGL24)-Type MADS-Box gene from trifoliate Orange (*Poncirus trifoliata* L. Raf.). Frontiers in Plant Science 7. 10.3389/fpls.2016.00823

Sun Y, Jiao G, Liu Z, Zhang X, Li J, Guo X, Du W, Du J, Francis F, Zhao Y, Xia L (2017) Generation of high-amylose rice through CRISPR/Cas9-Mediated targeted mutagenesis of starch branching enzymes. Frontiers in Plant Science 8. 10.3389/fpls.2017.00298

Sundriyal M, Sundriyal R (2004) Wild edible plants of the Sikkim Himalaya: marketing, value addition and implications for management. Economic Botany 58:300-315

Tadele Z (2019) Orphan crops: their importance and the urgency of improvement. Planta 250:677-694. 10.1007/s00425-019-03210-6

Takada Y, Murase K, Shimosato-Asano H, Sato T, Nakanishi H, Suwabe K, Shimizu KK, Lim YP, Takayama S, Suzuki G, Watanabe M (2017) Duplicated pollen–pistil recognition loci control intraspecific unilateral incompatibility in *Brassica rapa*. Nature Plants 3:17096. 10.1038/nplants.2017.96

Takada, S., A. Akter, E. Itabashi, N. Nishida, D. J. Shea, N. Miyaji, H. Mehraj, K. Osabe, M. Shimizu, T. Takasaki-Yasuda, T. Kakizaki, K. Okazaki, E. S. Dennis, and R. Fujimoto (2019) The role of FRIGIDA and FLOWERING LOCUS C genes in flowering time of *Brassica rapa* leafy vegetables. Scientific Reports 9(1): 13843. 10.1038/s41598-019-50122-2

Tanaka R, Oster U, Kruse E, Rüdiger W, Grimm B (1999) Reduced activity of Geranylgeranyl Reductase leads to loss of chlorophyll and tocopherol and to partially Geranylgeranylated chlorophyll in transgenic tobacco plants expressing antisense RNA for Geranylgeranyl Reductase. Plant Physiology 120:695-704. 10.1104/pp.120.3.695

Tang H, Bomhoff MD, Briones E, Zhang L, Schnable JC, Lyons E (2015) SynFind: Compiling syntenic regions across any set of genomes on demand. Genome Biology and Evolution 7:3286-3298. 10.1093/gbe/evv219

Tarailo-Graovac M, Chen N (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. Current Protocols in Bioinformatics 25:4.10.11-14.10.14. 10.1002/0471250953.bi0410s25

Teklewold A, Becker HC (2006) Geographic pattern of genetic diversity among 43 Ethiopian Mustard (*Brassica carinata* A. Braun) accessions as Revealed by RAPD Analysis. Genetic

Resources and Crop Evolution 53:1173-1185. 10.1007/s10722-005-2011-4

The Good Scent Company (2019) http://www.thegoodscentscompany.com/.

Thurber MD, Fahey JW (2009) Adoption of *Moringa oleifera* to combat under-nutrition viewed through the lens of the "diffusion of innovations" theory. Ecology of Food and Nutrition 48:212-225. 10.1080/03670240902794598

Tikunov YM, Laptenok S, Hall RD, Bovy A, de Vos RCH (2012) MSClust: a tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data. Metabolomics 8:714-718. 10.1007/s11306-011-0368-2

Tongco MDC (2007) Purposive sampling as a tool for informant selection. Ethnobotany Research & Applications 5:147-158

Traka M, Mithen R (2009) Glucosinolates, isothiocyanates and human health. Phytochemistry Reviews 8:269-282. 10.1007/s11101-008-9103-7

Umber M, Filloux D, Muller E, Laboureau N, Galzi S, Roumagnac P, Iskra-Caruana M-L, Pavis C, Teycheney P-Y, Seal SE (2014) The genome of African yam (*Dioscorea cayenensis-rotundata* complex) hosts endogenous sequences from four distinct badnavirus species. Molecular Plant Pathology 15:790-801. 10.1111/mpp.12137

Upadhyaya HD, Reddy KN, Ahmed MI, Kumar V, Gumma MK, Ramachandran S (2017) Geographical distribution of traits and diversity in the world collection of pearl millet [*Pennisetum glaucum* (L.) R. Br., synonym: *Cenchrus americanus* (L.) Morrone] landraces conserved at the ICRISAT genebank. Genetic Resources and Crop Evolution 64:1365-1381. 10.1007/s10722-016-0442-8

Usoro OB, Mousa SA (2010) Vitamin E forms in Alzheimer's disease: a review of controversial and clinical experiences. Critical Reviews in Food Science and Nutrition 50:414-419. https://doi.org/10.1080/10408390802304222

Utsumi Y, Sakurai T, Umemura Y, Ayling S, Ishitani M, Narangajavana J, Sojikul P, Triwitayakorn K, Matsui M, Manabe R-i, Shinozaki K, Seki M (2012) RIKEN Cassava Initiative: establishment of a cassava functional genomics platform. Tropical Plant Biology 5:110-116. 10.1007/s12042-011-9089-y

Uusiku NP, Oelofse A, Duodu KG, Bester MJ, Faber M (2010) Nutritional value of leafy vegetables of sub-Saharan Africa and their potential contribution to human health: A review. Journal of Food Composition and Analysis 23:499-509. https://doi.org/10.1016/j.jfca.2010.05.002

van Andel T, Ruysschaert S, Boven K, Daly L (2015) The use of Amerindian charm plants in the Guianas. Journal of Ethnobiology and Ethnomedicine 11:66. https://doi.org/10.1186/s13002-015-0048-9

van Andel T, Westers P (2010) Why Surinamese migrants in the Netherlands continue to use medicinal herbs from their home country. Journal of Ethnopharmacology 127:694-701. https://doi.org/10.1016/j.jep.2009.11.033

van Andel TR, van 't Klooster CIEA, Quiroz D, Towns AM, Ruysschaert S, van den Berg M (2014) Local plant names reveal that enslaved Africans recognized substantial parts of the New World flora. Proceedings of the National Academy of Sciences 111:E5346-E5353.

10.1073/pnas.1418836111

van Andel TR, van der Velden A, Reijers M (2016) The 'Botanical Gardens of the Dispossessed' revisited: richness and significance of Old World crops grown by Suriname Maroons. Genetic Resources and Crop Evolution 63:695-710. 10.1007/s10722-015-0277-8

van den Bergh E (2017) Comparative genomics and trait evolution in Cleomaceae, a model family for ancient polyploidy.

van den Bergh E, Külahoglu C, Bräutigam A, Hibberd JM, Weber APM, Zhu XG, Schranz ME (2014) Gene and genome duplications and the origin of C_4 photosynthesis: Birth of a trait in the Cleomaceae. Current Plant Biology 1:2-9. https://doi.org/10.1016/j.cpb.2014.08.001

van den Boog T, van Andel T, Bulkan J (2017) Indigenous children's knowledge about non-timber forest products in Suriname. Economic Botany 71:361-373. 10.1007/s12231-017-9400-4

van Den Heever E, Venter SL (2007) Nutritional and medicinal properties of *Cleome gynandra*. Acta Horticulturae 752:127–130

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA (2013) From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Current Protocols in Bioinformatics 43:11.10.11-11.10.33. 10.1002/0471250953.bi1110s43

van Jaarsveld P, Faber M, van Heerden I, Wenhold F, Jansen van Rensburg W, van Averbeke W (2014) Nutrient content of eight African leafy vegetables and their potential contribution to dietary reference intakes. Journal of Food Composition and Analysis 33:77-84. http://dx.doi.org/10.1016/j.jfca.2013.11.003

van Ooijen JW (2006) JoinMap 4, software for the calculationof genetic linkage maps in experimental populations. Plant Research International and Kyazma BV, Wageningen

van Treuren R, van Eekelen HDLM, Wehrens R, de Vos RCH (2018) Metabolite variation in the lettuce gene pool: towards healthier crop varieties and food. Metabolomics 14:146. 10.1007/s11306-018-1443-8

Vandebroek I, Balick MJ (2012) Globalization and loss of plant knowledge: challenging the paradigm. PLoS ONE 7. https://doi.org/10.1371/journal.pone.0037643

Varshney RK, Close TJ, Singh NK, Hoisington DA, Cook DR (2009) Orphan legume crops enter the genomics era! Current Opinion in Plant Biology 12:202-210. http://dx.doi.org/10.1016/j.pbi.2008.12.004

Varshney RK, Ribaut JM, Buckler ES, Tuberosa R, Rafalski JA, Langridge P (2012) Can genomics boost productivity of orphan crops? Nat Biotechnol 30. 10.1038/nbt.2440

Varshney RK, Saxena RK, Upadhyaya HD, Khan AW, Yu Y, Kim C, Rathore A, Kim D, Kim J, An S, Kumar V, Anuradha G, Yamini KN, Zhang W, Muniswamy S, Kim J-S, Penmetsa RV, von Wettberg E, Datta SK (2017a) Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. Nature Genetics 49:1082. 10.1038/ng.3872

Varshney RK, Shi C, Thudi M, Mariac C, Wallace J, Qi P, Zhang H, Zhao Y, Wang X, Rathore A, Srivastava RK, Chitikineni A, Fan G, Bajaj P, Punnuri S, Gupta SK, Wang H, Jiang Y, Couderc M, Katta MAVSK, Paudel DR, Mungra KD, Chen W, Harris-Shultz KR, Garg V, Desai N, Doddamani D, Kane NA, Conner JA, Ghatak A, Chaturvedi P, Subramaniam S, Yadav OP, Berthouly-Salazar C, Hamidou F, Wang J, Liang X, Clotault J, Upadhyaya HD, Cubry P, Rhone B, Gueye MC, Sunkar R, Dupuy C, Sparvoli F, Cheng S, Mahala RS, Singh B, Yadav RS, Lyons E, Datta SK, Hash CT, Devos KM, Buckler E, Bennetzen JL, Paterson AH, Ozias-Akins P, Grando S, Wang J, Mohapatra T, Weckwerth W, Reif JC, Liu X, Vigouroux Y, Xu X (2017b) Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. Nat Biotech advance online publication. 10.1038/nbt.3943

Vattem DA, Shetty K (2005) Biological functionality of ellagic acid: A review. Journal of Food Biochemistry 29:234-266. 10.1111/j.1745-4514.2005.00031.x

Vishwakarma MK, Arun B, Mishra VK, Yadav PS, Kumar H, Joshi AK (2016) Marker-assisted improvement of grain protein content and grain weight in Indian bread wheat. Euphytica 208:313-321. 10.1007/s10681-015-1598-6

Vodouhè R, Dansi A (2012) The "Bringing into Cultivation" phase of the plant domestication process and its contributions to in situ conservation of genetic resources in Benin. The Scientific World Journal 2012. http://dx.doi.org/10.1100/2012/176939

Volk GM, Richards CM (2011) Integration of georeferencing, habitat, sampling, and genetic data for documentation of wild plant genetic resources. HortScience 46:1446-1449

Volpato G, Godínez D, Beyra A (2009) Migration and ethnobotanical practices: the case of Tifey among Haitian immigrants in Cuba. Human Ecology 37:43-53. 10.1007/s10745-008-9211-4

von Grebmer K, Saltzman A, Birol E, Wiesmann D, Prasai N, Yin S, Yohannes Y, Menon P, Thompson J, Sonntag A (2014) 2014 Global Hunger Index: The challenge of hidden hunger. Welthungerhilfe, International Food Policy Research Institute, and Concern Worldwide, Bonn, Washington, D.C., and Dublin

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. Journal of Heredity 93:77-78. 10.1093/jhered/93.1.77

Wahyuni Y, Ballester A-R, Sudarmonowati E, Bino RJ, Bovy AG (2011) Metabolite biodiversity in pepper (*Capsicum*) fruits of thirty-two diverse accessions: Variation in health-related compounds and implications for breeding. Phytochemistry 72:1358-1370. http://dx.doi.org/10.1016/j.phytochem.2011.03.016

Wahyuni Y, Ballester A-R, Tikunov Y, de Vos RH, Pelgrom KB, Maharijaya A, Sudarmonowati E, Bino R, Bovy A (2013) Metabolomics and molecular marker analysis to explore pepper (*Capsicum* sp.) biodiversity. Metabolomics 9:130-144. 10.1007/s11306-012-0432-6

Waiganjo M, Muriuki J, Mbugua G Potential of indigenous leafy vegetables as companion crops for pest management of high-value legumes: a case study of *Gynandropsis gynandra* in Kenya. In: First International Conference on Indigenous Vegetables and Legumes. Prospectus for Fighting Poverty, Hunger and Malnutrition 752, 2007. ISHS, pp 319-321

Walther OJ (2014) Trade networks in West Africa: a social network approach. The Journal of

Modern African Studies 52:179-203. 10.1017/s0022278x14000032

Wang G-L, Xu Z-S, Wang F, Li M-Y, Tan G-F, Xiong A-S (2015) Regulation of ascorbic acid biosynthesis and recycling during root development in carrot (*Daucus carota* L.). Plant Physiology and Biochemistry 94:10-18. https://doi.org/10.1016/j.plaphy.2015.04.014

Wang L, Czedik-Eysenberg A, Mertz RA, Si Y, Tohge T, Nunes-Nesi A, Arrivault S, Dedow LK, Bryant DW, Zhou W, Xu J, Weissmann S, Studer A, Li P, Zhang C, LaRue T, Shao Y, Ding Z, Sun Q, Patel RV, Turgeon R, Zhu X, Provart NJ, Mockler TC, Fernie AR, Stitt M, Liu P, Brutnell TP (2014a) Comparative analyses of C_4 and C_3 photosynthesis in developing leaves of maize and rice. Nature Biotechnology 32:1158-1164. 10.1038/nbt.3019

Wang M, Toda K, Block A, Maeda HA (2019) TAT1 and TAT2 tyrosine aminotransferases have both distinct and shared functions in tyrosine metabolism and degradation in *Arabidopsis thaliana*. Journal of Biological Chemistry 294:3563-3576. 10.1074/jbc.RA118.006539

Wang S, Gao LZ (2015) Complete chloroplast genome sequence of green foxtail (*Setaria viridis*), a promising model system for C_4 photosynthesis. Mitochondrial DNA. 10.3109/19401736.2015.1079867

Wang W, Feng B, Xiao J, Xia Z, Zhou X, Li P, Zhang W, Wang Y, Møller BL, Zhang P, Luo M-C, Xiao G, Liu J, Yang J, Chen S, Rabinowicz PD, Chen X, Zhang H-B, Ceballos H, Lou Q, Zou M, Carvalho LJCB, Zeng C, Xia J, Sun S, Fu Y, Wang H, Lu C, Ruan M, Zhou S, Wu Z, Liu H, Kannangara RM, Jørgensen K, Neale RL, Bonde M, Heinz N, Zhu W, Wang S, Zhang Y, Pan K, Wen M, Ma P-A, Li Z, Hu M, Liao W, Hu W, Zhang S, Pei J, Guo A, Guo J, Zhang J, Zhang Z, Ye J, Ou W, Ma Y, Liu X, Tallon LJ, Galens K, Ott S, Huang J, Xue J, An F, Yao Q, Lu X, Fregene M, López-Lavalle LAB, Wu J, You FM, Chen M, Hu S, Wu G, Zhong S, Ling P, Chen Y, Wang Q, Liu G, Liu B, Li K, Peng M (2014b) Cassava genome from a wild ancestor to cultivated varieties. Nature Communications 5:5110. 10.1038/ncomms6110

Warnes GR, Ben B, Bonebakker L, Gentleman R, Huber Andy Liaw W, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, Venables B (2016) Various R programming tools for plotting data. R package version 3011

Wartena D (2006) Styles of making a living and ecological change on the Fon and Adja plateaux in South Benin. Wageningen University

Warwick SI, Gugel RK, McDonald T, Falk KC (2006) Genetic variation of Ethiopian Mustard (*Brassica carinata* A. Braun) germplasm in Western Canada. Genetic Resources and Crop Evolution 53:297-312. 10.1007/s10722-004-6108-y

Weinberger K, Pichop GN (2009) Marketing of African indigenous vegetables along urban and peri-urban supply chains in sub-Saharan Africa. In: African indigenous vegetables in urban agriculture. Earthscan, UK, pp 225-244

Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB (2017) Direct determination of diploid genome sequences. Genome Research 27:757-767. 10.1101/gr.214874.116

Weng J-K, Philippe RN, Noel JP (2012) The rise of chemodiversity in Plants. Science 336:1667-1670. 10.1126/science.1217411

White D, Gramacy RB (2012) maptree: Mapping, pruning, and graphing tree models. R package version 1.4-7.

WHO (2004) Vitamin and mineral requirements in human nutrition. 2nd edition. Report of a report of a joint FAO/WHO expert consultation, Bangkok, Thailand, 21–30 September 1998. WHO, Geneva, Switzerland

International Statistical Classification of Diseases and Related Health Problems 10th Revision. ICD-10 Version for 2016 (2016) http://apps.who.int/classifications/icd10/browse/2016/en. Accessed 20.12.2016

Wink M (2008) Plant Secondary Metabolism: Diversity, Function and its Evolution. Natural Product Communications 3:1934578X0800300801. 10.1177/1934578x0800300801

Wink M (2015) Modes of Action of Herbal Medicines and Plant Secondary Metabolites. Medicines 2:251-286

Witaicenis A, Seito LN, da Silveira Chagas A, de Almeida LD, Luchini AC, Rodrigues-Orsi P, Cestari SH, Di Stasi LC (2014) Antioxidant and intestinal anti-inflammatory effects of plant-derived coumarin derivatives. Phytomedicine 21:240-246. https://doi.org/10.1016/j.phymed.2013.09.001

Wu H, Wu X, Li Z, Duan L, Zhang M (2012) Physiological evaluation of drought stress tolerance and recovery in cauliflower (*Brassica oleracea* L.) seedlings treated with methyl jasmonate and coronatine. Journal of Plant Growth Regulation 31:113-123. 10.1007/s00344-011-9224-x

Wu T-h, Solberg SO, Yndgaard F, Chou Y-Y (2017) Morphological patterns in a world collection of *Cleome gynandra*. Genetic Resources and Crop Evolution 65:271-283. https://doi.org/10.1007/s10722-017-0529-x

Wunnakup T, Vimolmangkang S, De-Eknamkul W (2018) Transient expression of the homogentisate phytyltransferase gene from *Clitoria ternatea* causes metabolic enhancement of α-tocopherol biosynthesis and chlorophyll degradation in tomato leaves. Journal of Plant Biochemistry and Biotechnology 27:55-67. 10.1007/s13562-017-0415-9

Xu Y, Li P, Yang Z, Xu C (2017) Genetic mapping of quantitative trait loci in crops. The Crop Journal 5:175-184. https://doi.org/10.1016/j.cj.2016.06.003

Yan W, Kang MS, Ma B, Woods S, Cornelius PL (2007) GGE Biplot vs. AMMI Analysis of Genotype-by-Environment Data. Crop Science 47:643-653. 10.2135/cropsci2006.06.0374

Yang H-l, Dong L, Wang H, Liu C-l, Liu F, Xie C-x (2018a) A simple way to visualize detailed phylogenetic tree of huge genome-wide SNP data constructed by SNPhylo. Journal of Integrative Agriculture 17:1972-1978. https://doi.org/10.1016/S2095-3119(18)62023-4

Yang L, Wen K-S, Ruan X, Zhao Y-X, Wei F, Wang Q (2018b) Response of plant secondary metabolites to environmental factors. Molecules 23:762

Yang M, Huang H, Zhang C, Wang Z, Su Ye, Zhu P, Guo Y, Deng XW (2016) Arabidopsis atypical kinase ABC1K1 is involved in red light-mediated development. Plant Cell Reports 35:1213-1220. 10.1007/s00299-016-1953-7

Yang R-Y, Lin S, Kuo G (2008) Content and distribution of flavonoids among 91 edible plant species. Asia Pac J Clin Nutr 17:275-279

Yang R, Keding GB (2009) Nutritional contributions of important African indigenous vegetables. In: African indigenous vegetables in urban agriculture. Earthscan, pp 105-143

Yeo S, Coombe L, Warren RL, Chu J, Birol I (2017) ARCS: scaffolding genome drafts with linked reads. Bioinformatics 34:725-731. 10.1093/bioinformatics/btx675

Yetein MH, Houessou LG, Lougbégnon TO, Teka O, Tente B (2013) Ethnobotanical study of medicinal plants used for the treatment of malaria in plateau of Allada, Benin (West Africa). Journal of Ethnopharmacology 146:154-163. http://dx.doi.org/10.1016/j.jep.2012.12.022

Yoo SK, Chung KS, Kim J, Lee JH, Hong SM, Yoo SJ, Yoo SY, Lee JS, Ahn JH (2005) CONSTANS activates SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 through FLOWERING LOCUS T to promote flowering in Arabidopsis. Plant Physiology 139:770-778. 10.1104/pp.105.066928

You Y, Sawikowska A, Lee JE, Benstein RM, Neumann M, Krajewski P, Schmid M (2019) Phloem companion cell-specific transcriptomic and epigenomic analyses identify MRF1, a regulator of flowering. The Plant Cell 31:325-345. 10.1105/tpc.17.00714

Yuan H, Zhang J, Nageswaran D, Li L (2015) Carotenoid metabolism and regulation in horticultural crops. Horticulture Research 2:15036

Zeven AC, Zhukovsky PM (1975) Dictionary of cultivated plants and their centres of diversity: excluding ornamentals, forest trees and lower plants. Pudoc.

Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L (2018) PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics 35:1786-1788. 10.1093/bioinformatics/bty875

Zhang W, Hu D, Raman R, Guo S, Wei Z, Shen X, Meng J, Raman H, Zou J (2017) Investigation of the genetic diversity and Quantitative Trait Loci accounting for important agronomic and seed quality traits in *Brassica carinata*. Frontiers in Plant Science 8:615. 10.3389/fpls.2017.00615

Zhang X, Lu G, Long W, Zou X, Li F, Nishio T (2014) Recent progress in drought and salt tolerance studies in *Brassica* crops. Breeding Science 64:60-73. 10.1270/jsbbs.64.60

Zharare G (2012) Differential requirements for breaking seed dormancy in biotypes of *Cleome gynandra* and two *Amaranthus* species. African Journal of Agricultural Research 7:5049-5059

Zohoungbogbo HPF, Houdegbe CA, Sogbohossou DEO, Tossou MG, Maundu P, Schranz EM, Van Deynze A, Zoundjihekpon J, Achigan-Dako EG (2018) Andromonoecy in *Gynandropsis gynandra* (L.) Briq. (Cleomaceae) and effects on fruit and seed production. Genetic Resources and Crop Evolution 65:2231-2239. 10.1007/s10722-018-0687-5

Summary

Orphan crops have been used for centuries around the world as food, medicines and as sources of income. They also have ensured diversified and nutritious diets to local communities. However, these species have been largely absent from agricultural research agendas compared to the only a handful of commercial crops which cover 90% of human food demand. Reducing malnutrition and food insecurity in the world requires promoting novel and orphan crops. This thesis focused on developing a breeding strategy in the orphan leafy vegetable spider plant (*Gynandropsis gynandra*) for improved leaf yield and nutrient content.

The thesis starts with a reflection on the appropriate strategies for orphan leafy vegetable breeding using spider plant as a model (**Chapter 2**). *G. gynandra* belongs to the Cleomaceae family, the sister family of the Brassicaceae and is used as a vegetable and medicinal plant in Africa and Asia. A breeding program for the species should take into account the following components: (1) investigation of end-users preferences for the definition of breeding goals; (2) germplasm assembly, characterisation, and management; (3) the definition of product targets; (4) characterisation of the genetic control of key traits; (5) design of the process of cultivar development; (6) integration of genomics data to optimize that process; (7) multi-environment participatory trials and end-user evaluation; and (8) crop value chain development.

During germplasm collection in West Africa, observations of striking differences in the management systems of spiderplant, wild in some areas and cultivated all-year round under irrigated systems/fields/plots in others, led to an investigation of the variation in knowledge and management in seven communities in Benin and Togo (**Chapter 3**). Semi-structured interviews conducted with 428 respondents using cultural significance and management indices revealed that in addition to food uses, *G. gynandra* was used to cure 42 different diseases. The cultural importance and level of management of the species were strongly associated with ethnicity and gender. Socio-linguistic groups with similar cultural background had comparable perceptions of the cultural importance of the species and described similar management practices. An analysis of farmers' willingness to change their current management practices revealed that migration, market opportunities, and external intervention might significantly affect future management decision-making processes. The study highlighted how understanding socio-economic and cultural context can contribute to efficient design of research for development strategies aiming at inducing changes in local communities' management practices. Furthermore, the documented traditional knowledge on the species in surveyed areas will substantially contribute to promoting the species among potential consumers in urban areas.

Breeding programs require a good knowledge in the natural variation in traits of interest. Therefore, in **Chapter 4** variation in carotenoids, tocopherols and ascorbic acid as well as morphological traits in a worldwide germplasm of 76 accessions were investigated under greenhouse conditions. The levels of carotenoids and tocopherols accumulating in the leaves varied significantly across accessions and were linked with geographical origin and morphological variation. The main carotenoids included lutein, β-carotene, α-carotene and violaxanthin. A two-fold to three-fold variation was observed for these compounds. The main tocopherols detected were α-tocopherol and γ-tocopherol with a 20-fold variation. A nine-fold

variation in vitamin C concentration and independent of geographical origin was observed. Overall, the accessions were grouped into three clusters based on variation in nutrient content and morphology. West African accessions were short plants with small leaves and with high tocopherol contents and relatively low carotenoid contents; Asian accessions were short plants with broad leaves and with relatively low carotenoid and high tocopherol contents; while East/Southern African plants were tall with high contents of both carotenoids and chlorophylls and low tocopherol contents. Carotenoids were positively correlated with plant height as well as foliar and floral traits but negatively correlated with tocopherols. The absence of a significant correlation between vitamin C and other traits indicated that breeding for high carotenoids or tocopherols content may be coupled with improved leaf yield and vitamin C content.

To further investigate the variation in secondary metabolites in the species, metabolite profiling was performed on 48 accessions using liquid-chromatography mass spectrometry to detect semi-polar metabolites and gas-chromatography mass spectrometry for volatile compounds (**Chapter 5**). Results revealed large variation in 936 semi-polar compounds including flavonoids, terpenoids, glucosinolates and hydroxycinnamic acid derivatives. The variation in relative levels of semi-polar metabolite profiles was mainly driven by geography, suggesting incipient speciation, with the accessions from both West Africa and Asia forming a group clearly separated from the East/Southern African accessions. Volatile organic compounds detected included different sesquiterpenes, aldehydes, ketones, and sulphur-containing isothiocyanates. The variation in these volatiles was however not geography-specific, but likely linked to the taste and odour of the accessions. The relative abundance in glucosinolates and associated volatile sulphur compounds in the leaves allowed to cluster the accessions in three main groups that could be used for future plant-herbivore interaction studies.

Chapter 6 presents the current draft genome of *G. gynandra* coupled with ancient polyploidy event detection and the whole-genome re-sequencing of 53 accessions from East/Southern Africa, West Africa and Asia. The genome was assembled into 1693 scaffolds and SNP calling based on the resequenced diversity panel yielded 10.8 million SNPs. Genetic diversity analyses showed that variation at the genome level was associated with geographic origin of the accessions, and identified a strong differentiation between populations, consistent with morphological characterisation and metabolome profiling. Accessions from West Africa and Asia exhibited a higher level of linkage disequilibrium suggesting that these populations diverged more recently than East/Southern African ones. Thus, the results suggest an African origin for the species, a hypothesis that needs to be further investigated with a more comprehensive set of accessions.

In **Chapter 7**, SNPs called between two accessions with contrasting morphological and metabolic profiles re-sequenced in Chapter 6 allowed genotyping of 219 F_2 individuals using a targeted genotyping-by-sequencing approach. The high level of segregation distortion reduced considerably the number of SNPs effectively used to build the genetic map. However, twelve quantitative trait loci were identified for leaf area, plant height, flowering time, lutein content, tocopherol content and ascorbic acid content. QTLs with pleiotropic effects were identified on linkage groups 3, 7, 9 and 16. Candidate genes identified based on homology with Arabidopsis included flowering time (*FTIP1, CDKG2, MRF1*), carotenoid biosynthesis (*CRTISO, CYP97B3*),

tocopherol biosynthesis (*ABC1K1, CHLP, TAT1, VTE2*), leaf area (*DOT3*) and plant height (*AGL24*). No obvious candidate gene for ascorbic acid content was identified. Further validation of the QTLs is required for marker-assisted breeding for higher leaf-yield and nutrient content in the species. Finally, **Chapter 8** discussed the main findings in this thesis putting them in the broader context of orphan crops breeding, presented an overview of on-going collaborative research on *G. gynandra* by different institutions and defined research avenues for *G. gynandra* breeding. This thesis is expected to be a starting point for omics-assisted breeding in *G. gynandra* and promotion of the species in Africa and beyond.

Résumé

Les cultures orphelines sont utilisées depuis des siècles dans le monde entier comme aliment, médicament et source de revenus. Elles assurent également aux communautés locales des régimes alimentaires diversifiés et nutritifs. Cependant, ces espèces ont été largement absentes des agendas de la recherche agricole et seule une poignée de cultures commerciales couvre 90% de la demande alimentaire humaine. La réduction de la malnutrition et de l'insécurité alimentaire dans le monde nécessite donc la promotion des cultures orphelines. La présente thèse a porté sur le développement d'une stratégie d'amélioration du rendement en feuilles et de la teneur en éléments nutritifs chez le Caya blanc (*Gynandropsis gynandra*).

La thèse débute par une réflexion sur les stratégies appropriées pour la sélection des légumes-feuilles orphelins avec comme exemple principal le Caya blanc (**Chapitre 2**). Le Caya blanc appartient à la famille des Cleomaceae, la famille soeur des Brassicaceae, et est utilisé comme légume et plante médicinale en Afrique et en Asie. Les éléments-clés dont nous suggérons la prise en compte pour le développement d'un programme d'amélioration pour les légumes-feuilles incluent: (1) une enquête sur les préférences des utilisateurs finaux pour la définition des objectifs d'amélioration; (2) l'assemblage, la caractérisation et la gestion du matériel génétique; (3) la définition des types de variétés recherchés; (4) la caractérisation du contrôle génétique des caractères d'intérêt; (5) la conception du processus de développement des cultivars; (6) l'intégration des données génomiques pour optimiser ce processus; (7) les essais participatifs dans divers environnements et l'évaluation par les utilisateurs finaux; et enfin (8) le développement de chaînes de valeur pour ces cultures.

Au cours de la collecte de matériel génétique en Afrique de l'Ouest pour suppléer aux collections existantes dans d'autres régions du monde, l'observation de différences marquantes dans les systèmes de gestion du Caya blanc, récolté à l'état sauvage dans certaines zones et cultivé toute l'année dans des systèmes irrigués dans d'autres, a conduit à une enquête sur la variation des connaissances sur les modes de gestion de l'espèce dans sept communautés au Bénin et au Togo (**Chapitre 3**). Des entretiens semi-structurés menés auprès de 428 personnes interrogées et l'utilisation d'indices d'importance culturelle et de gestion ont révélé qu'en plus des utilisations alimentaires, *G. gynandra* était utilisé pour soigner 42 différentes maladies. L'importance culturelle et le niveau de gestion de l'espèce étaient fortement associés à l'appartenance ethnique et au genre. Une analyse de la volonté des agriculteurs de modifier leurs pratiques de gestion actuelles a révélé que la migration, les opportunités de marché et les interventions externes pouvaient influer considérablement sur les processus de prise de décisions liées la gestion de la plante. L'étude a montré comment la compréhension du contexte socio-économique et culturel pouvait contribuer à la conception de stratégies de développement efficaces visant à induire des changements dans les pratiques de gestion des communautés locales. En outre, les connaissances traditionnelles documentées sur l'espèce dans les zones étudiées contribueront considérablement à la promotion de l'espèce parmi les consommateurs potentiels des zones urbaines.

Les programmes d'amélioration des plantes exigent une bonne connaissance de la variation naturelle des caractères d'intérêt. Par conséquent, au **Chapitre 4**, la variation de la teneur en

caroténoïdes, tocophérols et acide ascorbique ainsi que de 27 divers caractères morphologiques au sein d'une collection mondiale de 76 accessions a été analysée en serre. Les teneurs en caroténoïdes et en tocophérols accumulées dans les feuilles ont varié de manière significative selon les accessions et étaient liées à l'origine géographique et à la diversité au plan morphologique. Les principaux caroténoïdes détectés incluent la lutéine, le β-carotène, l'α-carotène et la violaxanthine. La concentration en vitamine C était quant à elle indépendante de l'origine géographique. Globalement, les accessions ont été classées en trois groupes en fonction de la teneur en éléments nutritifs et de la morphologie. Les accessions d'Afrique de l'Ouest étaient des petites plantes ayant de petites feuilles, une teneur élevée en tocophérols et une teneur en caroténoïdes relativement faible; les accessions asiatiques étaient de petites plantes à feuilles larges et à teneur relativement faible en caroténoïdes et en tocophérols; tandis que les plantes d'Afrique orientale et australe étaient hautes avec des teneurs élevées en caroténoïdes et en chlorophylles et de faibles teneurs en tocophérols. L'absence de corrélation significative entre la vitamine C et les autres caractères étudiés indique que l'amélioration de la teneur en caroténoïdes ou en tocophérols peut être associée à une amélioration du rendement en feuilles et de la teneur en vitamine C.

Afin d'étudier plus en détail la variation en métabolites secondaires chez l'espèce, un profilage des métabolites secondaires a été réalisé sur les feuilles de 48 accessions en utilisant la spectrométrie de masse par chromatographie en phase liquide pour détecter les métabolites semi-polaires et la spectrométrie de masse par chromatographie en phase gazeuse pour les composés volatils (**Chapitre 5**). Les résultats ont révélé une grande variation chez 936 composés semi-polaires, notamment les flavonoïdes, les terpénoïdes, les glucosinolates et les dérivés de l'acide hydroxycinnamique. La variation des niveaux relatifs des profils de métabolites semi-polaires était principalement due à la géographie, ce qui suggère une spéciation naissante, les accessions d'Afrique de l'Ouest et d'Asie formant un groupe clairement séparé des accessions d'Afrique de l'Est. Les composés organiques volatils détectés comprenaient différents sesquiterpènes, aldéhydes, cétones et isothiocyanates. La variation de ces composés volatils n'était toutefois pas spécifique à la géographie, mais probablement liée au goût et à l'odeur des accessions. L'abondance relative des glucosinolates et des composés soufrés volatils associés dans les feuilles a permis de regrouper les accessions en trois groupes principaux qui pourraient être utilisés pour de futures études sur les interactions plantes-herbivores.

Le **Chapitre 6** présente le génome de *G. gynandra*, la détection de paleopolyploïdie et l'analyse de la diversité génétique au sein de 53 accessions d'Afrique de l'Est, d'Afrique de l'Ouest et d'Asie. Le génome a été assemblé en 1693 scaffolds et la détection de SNP avec les données de reséquençage des 53 accessions a généré 10,8 millions de SNP. Les analyses de diversité génétique ont montré que la variation au niveau du génome était associée à l'origine géographique des accessions et ont mis en évidence une forte différenciation entre les populations, compatible avec la caractérisation morphologique et le profil métabolomique. Les accessions d'Afrique de l'Ouest et d'Asie ont présenté un déséquilibre de liaison élevé, ce qui laisse supposer que ces populations ont divergé plus récemment que celles d'Afrique de l'Est. Ainsi, les résultats suggèrent une origine africaine pour l'espèce, une hypothèse qui doit être approfondie avec une plus grande collection d'accessions.

Au **Chapitre 7**, les SNP détectés entre deux accessions présentant des profils morphologiques et métaboliques contrastés et reséquencés au **Chapitre 6** ont permis le génotypage de 219 individus d'une population F_2 en utilisant une approche ciblée de génotypage par séquençage. Le niveau élevé de distorsion de la ségrégation a considérablement réduit le nombre de SNP utilisés pour établir la carte génétique. Cependant, douze loci associés à des caractères quantitatifs ont été identifiés, notamment, pour la surface foliaire, la hauteur de la plante, la période de floraison, la teneur en lutéine, en tocophérols et en acide ascorbique. Des QTL ayant des effets pléiotropes ont été identifiés sur les groupes de liaison 3, 7, 9 et 16. Les gènes candidats identifiés sur la base de leur homologie avec Arabidopsis incluent *FTIP1, CDKG2, MRF1* pour la floraison, *CRTISO* et *CYP97B3* pour la biosynthèse des caroténoïdes, *ABC1K1, CHLP, TAT1* et *VTE2* pour la biosynthèse des tocoph**é**rols, *DOT3* pour la surface foliaire ainsi que la hauteur de la plante (*AGL24*). Aucun gène candidat évident pour la teneur en acide ascorbique n'a été identifié. La validation des QTL est nécessaire pour la sélection assistée par marqueurs afin d'obtenir un rendement en feuilles et une teneur en éléments nutritifs plus élevés chez l'espèce.

Enfin, le **Chapitre 8** a examiné les principales conclusions de cette thèse en les plaçant dans le contexte plus général de la sélection de cultures orphelines, a présenté un aperçu de la recherche collaborative en cours sur *G. gynandra* avec différentes institutions et a défini des pistes de recherche pour la sélection de *G. gynandra*. La présente thèse constitue un point de départ pour la sélection assistée par «omiques **»** chez *G. gynandra* et la promotion de l'espèce en Afrique et dans le monde.

Acknowledgments

Finally a place where I can freely express myself!

Undertaking this PhD has been a truly life-changing experience for me and it would not have been possible without the support and guidance that I received from many interesting people ;)

First, I would like to thank my promotor **Eric** and my co-promotor **Enoch**, the most optimistic, funny and strict scientists I have ever met! Thank you **Eric** for your trust and guidance during this journey. You are a model for me as a scientist but also as a very charismatic and positive person. Your enthusiastic response to each of my baby steps helped me gain more confidence and motivation over the years. Thank you for believing in me and being always supportive. Dear **Enoch**, thank you for being my mentor for the past seven years and for encouraging me to start this PhD. I was surprised when you asked me during my MSc training to write a paper in English for publication but that was just the first of many challenges that you threw at me over the years to help me grow. Thank you for always giving me constructive feedback on my work and for that permanent positive attitude which is one of the greatest lessons I learned from you.

Mr Edgar Deguenon, thank you for giving me the chance for my Bachelor thesis to work with you and your community-supported agriculture network. You gave me the opportunity to test my abilities on my first "real" job at your NGO and ignited my passion for horticulture. It was a happy coincidence that six years later, you were leading the project that launched my PhD journey.

I would also like thank the members of the Cleome Consortium and the African Orphan Crops Consortium. **Allen Van Deynze**, thank you very much for your support, your mentorship and your very precise feedback during this journey. **Tsvetelina Stoilova**, **Patrick Maundu**, **Prasad Hendre**, **Iago Hale** and **Rita Mumm**, thank you all for the great collaboration and your mentoring. I have learnt a lot working with each of you.

My deepest gratitude goes to the fantastic four of the Metabolomics Lab, WUR: **Ric de Vos**, **Harry Jonker**, **Bert Schipper** and **Henriette van Eekelen**. Thank you all for training me for the HPLC analyses and for your infinite patience with the clumsy labmate that I was :)

Tinde, thank you very much for accepting to guide me for the ethnobotanical study and keeping the motivation on. Despite "out of scope" emails we finally made it!

My heartfelt thanks also go to the Cambridge C_4 team, **Julian Hibberd**, **Conor Simpson** and **Pallavi Singh** for the great collaboration. It was always a pleasure to work with you and our discussions helped me quite often put things into perspective.

Dear co-authors, **Erik van den Bergh**, **Andrea Brautigam**, and **Andreas Weber,** thank you for your collaboration and availability.

I am very grateful towards the Women of Dogbo (Southern Benin) for initiating me to the world of spider plant production and utilization as well as all the people who kindly accepted to participate in the ethnobotanical study or assisted us during germplasm collection missions. The hospitality and generosity of people in all the West African countries that we surveyed was

just mind-blowing!

My deepest gratitude goes to the **Schlumberger Foundation Faculty for the Future** team, especially **Roseline, Elisabeth, Regina** and **Eve**. Thank you all for this amazing opportunity that you gave me to achieve my dream and for the daily assistance with all kinds of requests.

Dear colleagues from Biosystematics, I really enjoyed working with you! These 4 years seem now to have been too short to get to know each other.

Namaste **Wilma**. From the first day of my PhD, you made a great impression on me with the bike tour of Wageningen and the cutlery you gathered for me. It was reassuring to know that you were always there to help me sort all kinds of things out. I also enjoyed having you as a Yoga tutor :) You are the Sun of the Group. Keep shining and meditating.

Sara, somehow you were at the beginning and the end of this adventure. Rick and you welcomed me when I first arrived in the Netherlands in the middle of the night and I was happy to share the silence room and the ups and downs of the last stretch of the PhD with you :) Wish you all the best in your career!

Setareh, you were an inspiration for me along this journey! You are a strong confident woman and I admire your focus in science but also in life in general. You rock!

Floris, ce fut un plaisir de parler français avec toi de temps en temps. Thanks for all the nice discussions! Your enthusiasm and spontaneity were always refreshing and your funny stories about Indonesia and Belgium, your adoptive countries, helped to clear up many foggy days.

Niccolo, it was nice to discuss with you about pretty much everything and your good vibes were always welcome. Molto grazie for saving my F_2 sampling day. I wish you a lot of fun with Brassicas and eggs!

Thijmen, it was nice to literally have each other's back for a while :) I passively learnt a few Dutch words from you that will remind me our several comforting discussions and aha! moments ;)

Wei, it was soothing to share with you all the bioinformatics tips and frustrations! Keep up the good energy!

Frank, the king of the lab, your enthusiasm about this project was very helpful along the way. It was a pleasure to work with you!

Lotte, your sense of humor brought a nice atmosphere to the open space. I really enjoyed our chitchats. Thanks for your support and for being on stage with me during the defence!

Kitty, you are welcome in Benin anytime for new and much more fun memories. Thanks a lot for your support and advice in the confusing moments.

Nina, you always had a nice word or was available when I needed your feedback. Thanks a lot!

Freek, many of the first presentations I followed in the group sounded a bit alien to me. Thank you for guiding me through the systematics and phylogeny concepts. It was great to discuss with you about science but also about culture and music. That concert at the Musis Sacrum really filled my soul and gave me energy for the writing period! Bedankt!

Eddie, **Klaas**, **Lars**, **Roel**, **Nynke**, **Sabrina**, **Robin**, **Marco**, **Casper**, **Patrick**, **Nora, Chloe,** thanks for all the "gezellig" moments we shared.

My dear students **Dieke**, **Jan**, **Bart** and **Chen**: I was very happy to work with you guys and I learned a lot with you. This thesis would have probably taken a different turn without your commitment on your respective projects. **Ilse** and **Nina**, your short course project was actually a life saver and I was really impressed by your team spirit and accuracy!

Maarten, you've been a true friend and always took good care of the spiderplants in Unifarm. I cannot thank you enough for everything.

My stays in the Netherlands were brightened by some wonderful souls who made me feel like home:

Phuong, my sister from another continent, thank you so much for your love and support. Our discussions helped me a lot put myself together and learn to be patient. It was nice to spend some time with you, Robin, the great Casper and more recently the little Viam :)

Falylath, thank you Sis for always being there when I needed, for printing and submitting this thesis while I was in Benin and the list could go on and on. Words are not enough express my gratitude towards you.

Severin, thank you for being there for me in the past two years. Thank you for everything. I learned a lot from you and I wish you all the best!

Deji, you are probably the most extroverted introvert I have ever met. I am grateful for the happy combination of circumstances that led to our friendship! Thanks for being my paranymph.

Erik R., it was always nice to spend some time with you far away from the university world. Thank you so much for the nice outings.

Bernard and **Joshua**, thanks for those Friday evenings at the embassy! They were priceless!

My heartfelt thanks to the Beninese Community in Wageningen, especially **Waliouuuu, Charles, Folachodé, Hermann, Yvon, Christel, Marius, Deo-Gratias, Osseni, Donald H., Imelda, Romaric, Gildas, Folakè**. Your daily support and brotherhood spirit were very much appreciated.

Allemaal bedankt voor uw vriendelijkheid!

Although I was a PhD student at WUR, I had the luxury to spend 19 months in Benin over the 49 months that this adventure lasted.

Dear colleagues of **GBioS**, you guys are awesome and inspire me daily to keep going. My immense gratitude goes to you all for that touching surprise birthday party on August 26th this year. I will forever cherish those memories.

Jelila, **Xavier**, **Rachidi**, **Herbaud**, **Soulemane, Fernand, Olafemi**, I am glad to see each of you thrive and it has been an honour to guide you at some point in your academic training.

N'Danikou, **Dedeou**, **Chaldia**, **Lys**, **Félicien**, **Ulrich**, **Charlotte**, **Innocent**, **Hervé** and

Nicodème, special thanks to you for your support and your availability.

Dear **Carlos**, your stamina on every project you run is unbelievable and I am happy that your PhD project somehow builds on the findings of this thesis.

Alcade, dear Brother, sometimes you remind me an octopus, doing a thousand things at the same time :) thanks a lot for your presence and support along the way.

Kênou, you are a great inspiration for me. Keep shining brighter and brighter.

Olafemi, I could write a whole book but for now I will just say thank you :)

The **MoBreed and Genes fellows**, big up to you!

I also want to thank some people outside the academic world who simply enlightened this journey and made it more fun.

My gratitude also goes to the Gletafa crew: **Ariel, Frida, Veronica, Igor, Kennedy, Isaac,** and of course **Mohamed** :) I hope we catch up soon for new challenges. Stay blessed!

Gladys G., you are just an awesome person to have as a friend! Thank you for being you!

Rock, dear cousin, thank you for hosting me for my Christmas and New Year times in Europe. You are awesome!

Harmony, lovely sister, you really bear very nicely your name and I will never be thankful enough for your presence n my life.

Laetitia et **Grégoire**, je ne sais comment vous remercier pour les appels vespéraux, les fous rires et tout le reste. Le Père vous bénisse! Gabin, la Vademecumoise te dis tout simplement merci.

Papa et **Maman**, merci infiniment pour tous les sacrifices consentis pour nous donner a tous les cinq une éducation de qualité. Merci de m'avoir toujours soutenue et encouragée dans tous mes projets. Je vous suis infiniment reconnaissante pour tout. Je vous aime très fort. Soyez bénis et comblés.

My dear siblings, **Medesu, Defodji, Etotepe** and **Missi**, these past 4 years made me realize how lucky I am to be part of you. Medesu, thank you for always checking on me and making sure I stay mentally healthy :) Etotepe, you are a role model for me as a female scientist and I cannot thank you enough for your encouragement along the way. The best is yet to come for you! Missi, your permanent optimism, curiosity and dedication to hard work inspire me daily. I wish I had half of your time management skills and stamina. **Olabissi**, thanks for your presence. You are not my little man anymore but now a first year undergraduate. It is a delight to see you bloom and I wish you all the best in your studies. Whatever the challenges, your dreams are valid!

Finally, I would like to thank all the people who participated in a way or the other in completing this journey but were not specifically acknowledged. Thank you all!

List of publications

Peer-reviewed papers

Sogbohossou E.O.D, D. Kortekaas, E.G. Achigan-Dako, P. Maundu, T. Stoilova, A. Van Deynze, R.C.H. de Vos, M. E. Schranz. 2019. Association between vitamin content, plant morphology and geographical origin in a worldwide collection of the orphan crop *Gynandropsis gynandra* (Cleomaceae). Planta 250(3):933-947. https://doi.org/10.1007/s00425-019-03142-1

Houdegbe, C.A., **E.O.D. Sogbohossou** and E.G. Achigan-Dako. 2018. Enhancing growth and leaf yield in *Gynandropsis gynandra* (L.) Briq.(Cleomaceae) using agronomic practices to accelerate crop domestication. Scientia Horticulturae 233: 90-98. https://doi.org/10.1016/j.scienta.2018.01.035

Zohoungbogbo H. P. F., C.A. Houdegbe, **E.O.D. Sogbohossou**, M.G. Tossou, Patrick Maundu, Eric M. Schranz, Allen Van Deynze, J. Zoundjihekpon, E.G. Achigan-Dako, 2018. Andromonoecy in *Gynandropsis gynandra* (L.) Briq. (Cleomaceae) and effects on fruit and seed production. Genetic Resources and Crop Evolution. https://doi.org/10.1007/s10722-018-0687-5

Reeves G., P. Singh, T. A. Rossberg, **E.O.D. Sogbohossou**, M. Eric Schranz, J. M. Hibberd. Natural variation within a species for traits underpinning C_4 photosynthesis. Plant Physiology 178 (1): 504-512. https://doi.org/10.1104/pp.18.00168

Sogbohossou, E.O. D., E.G. Achigan-Dako, T. van Andel and M.E. Schranz. 2018. Drivers of management of spider plant (*Gynandropsis gynandra*) across different socio-linguistic groups in Benin and Togo. Economic Botany: 1-25. https://doi.org/10.1007/s12231-018-9423-5

Sogbohossou E.O.D, E. G. Achigan-Dako, P. Maundu, S. Solberg, E. M. S. Deguenon, R. H. Mumm, I. Hale, A. Van Deynze, and M. E. Schranz. 2018. A roadmap for breeding orphan leafy vegetable species: a case study of *Gynandropsis gynandra* (Cleomaceae). Horticulture Research 5(1): 2. https://doi.org/10.1038/s41438-017-0001-2

Houdegbe, C.A., **E.O.D. Sogbohossou** and E.G. Achigan-Dako. 2016. Utilization and breeding perspective in the egusi gourd *Melothria sphaerocarpa* (Cogn.) H. Schaef. et SS Renner (syn: *Cucumeropsis mannii* Naudin). Genetic resources and crop evolution 63: 545-559. https://doi.org/10.1007/s10722-015-0361-0

Sogbohossou E.O.D., Achigan-Dako E. G., Assogba-Komlan F., Ahanchede A. 2015. Diversity and differential utilization of *Amaranthus* spp. along the urban-rural continuum of Southern Benin. Economic Botany 69(1):9-25. https://doi.org/10.1007/s12231-014-9294-3

Sogbohossou, E.O.D., Achigan-Dako, E.G. 2014. Phenetic differentiation and use-type delimitation in *Amaranthus* spp. from worldwide origins. Scientia Horticulturae 178:31-42. https://doi.org/10.1016/j.scienta.2014.08.003

Achigan-Dako, E.G., **Sogbohossou E.O.D.**, Maundu, P. 2014. Current knowledge on *Amaranthus* spp.: research avenues for improved nutritional value and yield in leafy amaranths in Sub-Saharan Africa. Euphytica 197(3):303-317. https://doi.org/10.1007/s10681-014-1081-9

In Preparation

E.O.D. Sogbohossou, E.G. Achigan-Dako, R. Mumm, R.C.H. de Vos, M.E. Schranz. Natural variation in secondary metabolites production in the leafy vegetable spider plant (*Gynandropsis gynandra* L. (Briq.)).

E.O.D. Sogbohossou, A. Brautigam, E. van den Bergh, E.G. Achigan-Dako, P.S. Hendre, R. Jamnadass, S. Muthemba, R. Kariba, A. Van Deynze, J. Hibberd, A.P.M. Weber, M. E. Schranz. The draft genome of *Gynandropsis gynandra* (Cleomaceae) provides insights in whole genome-duplication, population genetic structure and genome-metabolome associations.

E.O.D. Sogbohossou, F. Becker, P. Hendre, E.G. Achigan-Dako, A. Van Deynze, J. Hibberd, R.C.H. de Vos , M.E. Schranz. Mapping Quantitative Trait Loci controlling vitamin content, flowering time and morphological traits in the orphan leafy vegetable *Gynandropsis gynandra* (Cleomaceae).

About the author

Dêêdi was born on August, 26th 1991 in Cotonou, Benin Republic. Growing up, she gradually developed a passion for nature with influences from her father, an agronomist and her mother, a passionate of flowers and medicinal plants. After receiving her secondary school diploma in 2007, she obtained a national scholarship to pursue her Bachelor at the Faculty of Agronomic Sciences of the University of Abomey-Calavi (Benin). She completed her Bachelor in Agronomy in 2011 and later did her Master studies at the same Faculty. In April 2014, she defended her MSc thesis entitled "Leafy amaranths: Farmers' knowledge along an urban-rural continuum in Southern Benin and phenotypic characterisation" under the supervision of Dr. Enoch Achigan-Dako and graduated (cum laude).

Since April 2014, she is a research assistant at the Laboratory of Genetics, Horticulture and Seed Science (GBioS) led by Dr. Achigan-Dako. As part of her duties, she co-wrote and submitted in 2014 the Cleome Project to the Applied Research Fund Scheme of NWO-WOTRO. She also contributed to the elaboration of two Intra-Africa Mobility programs, MoBreed (mobreed.com) in 2016 and Genes (genes-intra-africa.org) in 2017, which aim at training high-profile PhD and MSc plant breeders on the continent. With the support of the Schlumberger Foundation and the University of Abomey-Calavi, she initiated with Dr Achigan-Dako and other colleagues, the application of the GBioS lab to the Seeding Labs Instrumental Access program (seedinglabs.org). The program tremendously contributed to equipping the laboratory with microscopes, fume-hoods, PCR machines among others, all things that she and other students did not have access to during their MSc training at the Faculty.

With financial support from the Cleome Project, in July 2015, she started a sandwich PhD at Wageningen University under the supervision of Prof. Eric Schranz. The project ended in December 2017 and having successfully applied to the Schlumberger Foundation Faculty for the Future Fellowship in 2018 and 2019, she could stay in the Netherlands to complete and submit her PhD. Her research, described in this thesis, focused on developing a breeding program for the emerging leafy vegetable *Gynandropsis gynandra*. She is currently looking for the next challenging project that will meet both her passion for neglected crops breeding and her thirst to learn.

Education statement

Education Statement of the Graduate School
Experimental Plant Sciences

Issued to: Eurydice Olga Dêêdi Sogbohossou
Date: 16 December 2019
Group: Biosystematics Group
University: Wageningen University

1) Start-Up Phase	date	cp
► **First presentation of your project**		
Ethnobotany and functional genomics of *Cleome gynandra* for improved nutritional value: a proposal	4 Nov 2015	1.5
► **Writing or rewriting a project proposal**		
Ethnobotany, metabolomics and functional genomics in *Cleome gynandra* for increased consumption in West Africa	Dec 2015	3.0
► **Writing a review or book chapter**		
A roadmap for breeding orphan leafy vegetable species: a case study of *Gynandropsis gynandra* (Cleomaceae). Hort. Res. 5:2 (2018). doi: 10.1038/s41438-017-0001-2	Jan 2016	6.0
► **MSc courses**		
Subtotal Start-Up Phase		10.5

2) Scientific Exposure	date	cp
► **EPS PhD student days**		
EPS PhD Get2gether	9-10 Feb 2017	0.6
EPS PhD Get2gether	11-12 Feb 2019	0.6
► **EPS theme symposia**		
EPS theme 4 'Genome biology',	15 Dec 2015	0.3
EPS theme 4 'Genome biology'	16 Dec 2016	0.3
EPS theme 4 'Genome biology'	25 Sep 2018	0.3
► **Lunteren Days and other national platforms**		
Annual Meeting Experimental Plant Sciences Lunteren	9-10 Apr 2018	0.6
Annual Meeting Experimental Plant Sciences Lunteren	8-9 Apr 2019	0.6
► **Seminars (series), workshops and symposia**		
Seminar: WEES Seminar, Prof Tinde van Andel 'Can the belief in magic plants lead to rainforest conservation?', WUR	17 Dec 2015	0.1
Seminar: Dr. Teemu Teeri 'Pelargonidin in flowers - why not?', WUR	14 Mar 2018	0.1
Seminar: Dr Casper van der Kooi "Evolution and optics of flower coloration"	20 Mar 2019	0.1
Seminar: Dr Enrico Scarpella 'Control of vein patterning by auxin', WUR	19 Jun 2019	0.1
Symposium: Genotype to phenotype modelling of plant adaptation, WUR	16 Nov 2017	0.3
Symposium: WUR PhD Symposium 2018	17 May 2018	0.3
► **Seminar plus**		
Ethnobotany masterclass, Prof Tinde van Andel	17 Dec 2015	0.1
► **International symposia and congresses**		
Annual Cleonomics Project Meeting, Wageningen, The Netherlands	30 Sep - 1 Oct 2016	0.5
Tropentag 2017, Bonn, Germany	20-22 Sep 2017	0.9
Annual Cleonomics Project Meeting, Nairobi, Kenya	28-29 Nov 2017	0.5
Plant Genome Evolution, Sitges, Spain	1-3 Oct 2017	0.9
Plant and Animal Genome, 2018, San Diego, California, USA	17 Jan 2018	0.2
European Plant Science Retreat, Utrecht, The Netherlands	3-6 Jul 2018	0.9
Plant Genetics and Breeding Technologies, Vienna, Austria	12-13 Jul 2018	0.6
Schlumberger Foundation Faculty for the Future Fellows and Alumnae Forum, Abu Dhabi, UAE	18-21 Nov 2018	0.9
► **Presentations**		
Talk: Diversity and uses of *Cleome gynandra* , an underutilized vegetable in Africa, Annual Cleonomics Project Meeting, WUR	30 Sep - 1 Oct 2016	1.0
Talk: Metabolic diversity in *Gynandropsis gynandra* and implications for breeding, EPS theme 4 'Genome biology', WUR	16 Dec 2016	1.0
Poster: Effect of agronomic practices on growth and leaf yield in spider plant (*Gynandropsis gynandra*), Tropentag 2017	20-22 Sep 2017	1.0
Talk: Orphan no more: combining ethnobotanical and omics-assisted breeding approaches for improved nutritional quality in *Gynandropsis gynandra* , Annual Cleonomics Project Meeting, Nairobi, Kenya	28-29 Nov 2017	1.0
Talk: Gynandropsis gynandra , a model for orphan crops. Plant and Animal Genome, USA	17 Jan 2018	1.0
Talk: Developing resources for omics-assisted breeding in G*ynandropsis gynandra* (Cleomaceae), an orphan leafy vegetable, WUR PhD symposium 2018	17 May 2018	1.0
Poster: Genetic variation in morphology and nutrient content in the orphan crop *Gynandropsis gynandra* (Cleomaceae) associated with geographical origin, European Plant Science Retreat, Utrecht, NL	3-6 Jul 2018	1.0
Talk: Genetic variation in morphology and nutrient content in *Gynandropsis gynandra* (Cleomaceae), an orphan leafy vegetable, Plant Genetics and Breeding Technologies, Vienna, Austria	12-13 Jul 2018	1.0
Poster: Developing resources for omics assisted breeding in spider plant (*Gynandropsis gynandra*), an orphan leafy vegetable. Schlumberger Foundation Faculty for the Future Fellows and Alumnae Forum, Abu Dhabi, UAE	18-21 Nov 2018	1.0
Talk: Genetic mapping of leaf yield and nutritional quality traits in the orphan crop *Gynandropsis gynandra* , Annual Meeting Experimental Plant Sciences Lunteren	8-9 Apr 2019	1.0
► **IAB interview**		
► **Excursions**		
Company Visit TomatoWorld	14 Oct, 2016	0.2
Subtotal Scientific Exposure		20.0

CONTINUED ON NEXT PAGE

3) In-Depth Studies	date	cp
► **Advanced scientific courses & workshops**		
Postgraduate course 'Bayesian Statistics' , WUR	17-18 Oct 2016	0.6
EMBL-EBI course 'Analysis of High-Throughput Sequencing Data', Hinxton, UK	7-10 Nov 2017	1.2
The Power of RNA-Seq, WUR	11-13 Jun 2018	0.9
Genome-wide signatures of selection and association studies, Physalia Courses, Berlin, Germany	22-26 Oct 2018	1.5
► **Journal club**		
Biosystematics Group Journal Club	2015-2019	3.0
► **Individual research training**		
Individual training at the Wageningen Plant Research Metabolomics laboratory, NL	1 Sep - 15 Oct 2015	3.0
Subtotal In-Depth Studies		10.2

4) Personal Development	date	cp
► **General skill training courses**		
Project and time management	22 Sep - 3 Nov 2016	1.5
PhD Competence Assessment	10 Nov 2016	0.3
Reviewing a scientific paper	23 Nov 2017	0.1
Scientific artwork with Photoshop and Illustrator	19-20 Mar 2017	0.6
WGS PhD workshop carousel 2018	25 May 2018	0.3
Bridging across cultural differences	28 Nov, 12 Dec 2018	0.7
WGS PhD workshop carousel 2019	24 May 2019	0.3
► **Organisation of meetings, PhD courses or outreach activities**		
► **Membership of EPS PhD Council**		
Subtotal Personal Development		3.8

TOTAL NUMBER OF CREDIT POINTS*	**44.5**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS with a minimum total of 30 ECTS credits.

** A credit represents a normative study load of 28 hours of study.*

The research described in this PhD thesis was funded by the Applied Research Fund of the Netherlands Organization for Science under the Project "Utilizing the genome of the vegetable species *Cleome gynandra* for the development of improved cultivars for the West and East African markets" (Project Number: W.08.270.350); the Schlumberger Foundation Faculty for the Future Fellowship and the LEB Foundation.

Financial support from Wageningen University & Research for printing this thesis is gratefully acknowledged.

Cover design by Olafemi Adjinda
Printed by Digiforce