



Adoption and impact of improved amaranth cultivars in Tanzania using DNA fingerprinting

Rosina Wanyama¹ · Pepijn Schreinemachers² · Justus Ochieng⁵ · Omary Bwambo¹ · Roselyne Alphonse³ · Fekadu Fufa Dinssa¹ · Ya-ping Lin⁴ · Roland Schafleitner⁴

Received: 11 May 2022 / Accepted: 14 June 2023 / Published online: 11 July 2023
© The Author(s) 2023

Abstract

Traditional vegetables are very important for food and nutrition security in sub-Saharan Africa but have not received much attention in crop improvement research. One exception is amaranth (*Amaranthus* spp.) for which international breeding research began at the World Vegetable Center in Tanzania in 2004. This study is the first impact evaluation of this research program. It uses a representative sample of 1,355 amaranth-producing households from five agroecological zones of Tanzania to collect questionnaire-based data and plant material. The genetic identity of the amaranth cultivars used by the farmers was identified using DNA fingerprinting, while the impact on crop yield, production cost and sales was quantified using propensity score matching. The analysis shows that 66% of Tanzania's cultivated area under amaranth uses cultivars derived from the amaranth breeding program. These cultivars had a mean vegetable yield that was 6.1 ton/ha higher (+48%; $p=0.002$) than that of other cultivars, while there was no significant impact on the cost of production or quantity sold. This suggests tremendous impact of genetic improvement research in traditional vegetables.

Keywords DNA fingerprinting · Propensity score matching · Technology adoption · Variety · Vegetable

1 Introduction

Amaranth (*Amaranthus* spp., *mchicha* in Swahili) is a popular traditional vegetable with a long cultural tradition in East Africa. It is cultivated for its leaves and young stems, and sometimes for its grain. The leaves are rich in vitamin C and pro-vitamin A as well as in iron, zinc, and calcium – micronutrients that local diets are often deficient in (Yang & Keding, 2009). Increased consumption of green leafy

vegetables such as amaranth is important to achieving food and nutrition security. In Kenya and Tanzania, there has been an increasing demand for traditional African vegetables such as amaranth, especially among urban consumers (Cernansky, 2015; Chelang'a et al., 2013). This provides an opportunity for smallholder farmers, many of whom are women, to earn income and improve their livelihood through vegetable amaranth production (Mwadingeni et al., 2021).

Until about fifteen years ago there was no formal seed system for supplying improved amaranth cultivars. Farmers saved their own seed and exchanged it locally, which is still a common practice today. The World Vegetable Center (WorldVeg) established an amaranth breeding program in Tanzania in 2004 with the aim of creating new amaranth cultivars with preferred taste and enhanced performance under local biotic and abiotic constraints. To do this, WorldVeg researchers and partners collected amaranth germplasm from across Africa, Asia and the Americas and conserved this for public access in vegetable genebanks in Taiwan and Tanzania, which currently hold 867 amaranth accessions of 18 species (Dinssa et al., 2018; see also <https://genebank.worldveg.org/>). Accessions with farmer and consumer-preferred traits were chosen for purification and improved through selection and cross-breeding. Together with the Tanzania Agricultural Research

✉ Rosina Wanyama
rnwanyama@gmail.com

¹ World Vegetable Center, Eastern and Southern Africa, PO Box 10, Duluti, Arusha, Tanzania

² World Vegetable Center, East and Southeast Asia, PO Box 1010, Bangkok 10903, Thailand

³ Department of Agricultural Economics and Agribusiness, College of Economics and Business Studies, Sokoine University of Agriculture, PO Box 3007, Morogoro, Tanzania

⁴ World Vegetable Center, PO Box 42, Shanhua, Tainan 74199, Taiwan

⁵ Alliance of Bioversity International and CIAT, P.O. Box 2704, Arusha, Tanzania

Institute (TARI), promising lines were tested in multiple locations in Tanzania and the best performing lines were released for commercial seed production and distributed to farmers. Amaranth cultivars (that is, cultivated varieties) and selected lines have been promoted in Tanzania and other countries of sub-Saharan Africa including Malawi, Rwanda, Kenya, and Uganda by including them in large-scale seed distributions through several programs (Stoilova et al., 2019). In addition, by 2016, local seed companies in Kenya and Tanzania were commercializing improved cultivars and sold and estimated 2.9 tons of WorldVeg amaranth seed (Ochieng et al., 2019), which increased to 8 tons in Kenya alone in 2020 (Simlaw Seeds, personal communication).

A recent study using expert elicitation methods estimated that improved amaranth cultivars were planted on 51% of the amaranth area in Kenya and 70% in Tanzania with WorldVeg-related varieties planted on 20.0% of the amaranth area in Kenya and 47.1% in Tanzania (Ochieng et al., 2019). These data suggest that improved amaranth cultivars have made a substantial impact, but this has not been accurately quantified to date.

Previous studies have shown that estimates based on household surveys or expert opinion tend to overestimate the adoption of improved cultivars as compared to DNA fingerprinting, which is a more reliable method (Rabbi et al., 2015; Wossen et al., 2019). DNA fingerprinting has previously been applied to quantify the adoption of improved cultivars of staple food crops (e.g., Floro IV et al., 2018; Kosmowski et al., 2019; Maredia et al., 2016). For Tanzania, two previous studies applied DNA fingerprinting to quantify the adoption of improved maize and rice cultivars (Oruko et al., 2015; Wineman et al., 2020). Regarding the accuracy of alternative methods, Floro IV et al. (2018) in a study for Colombia, estimated that 9.2% of farmers cultivated improved cassava cultivars as based on DNA fingerprinting while this was 17.1% as based on household survey data. Also, Maredia et al. (2016) found that expert elicitation and household surveys overestimated the adoption of improved cassava and bean cultivars as compared to DNA fingerprinting. Their results showed 85% adoption of improved cassava varieties in Ghana as based on expert opinion, 57% as based on a household survey, but only 16% as based on DNA fingerprinting. For improved bean varieties in Zambia this was 80% adoption as based on household survey data, 68% as based on expert opinion, but only 4% as based on DNA fingerprinting.

Against this background, the objective of this study is to estimate the adoption and impact of improved amaranth cultivars in Tanzania that originated from the international breeding program of WorldVeg. We use DNA fingerprinting as the main method to quantify cultivar adoption, but also compare it against farmers' own assessment as measured from a farm household survey and expert elicitation

estimates as reported by Ochieng et al. (2019). Our study is the first application of DNA fingerprinting to cultivar adoption in of vegetables. The application to amaranth is challenging because it is a multispecies crop with at least six different species cultivated in Tanzania. Furthermore, while most previous studies stopped at cultivar identification and adoption, this study goes one step further and combines household survey data with cultivar identification to quantify the immediate impact of variety adoption on crop yield, production costs and sales. Our hypothesis is that improved cultivars give a higher mean yield while production costs are similar. A higher yield would allow farmers to sell a greater quantity.

2 Materials and method

2.1 Farm household survey

Survey data were collected in five of Tanzania's seven agroecological zones (AEZs) between October and November 2018. A multi-stage stratified sampling procedure was used to select households. In the first stage, five AEZs with the highest amaranth production as based on national statistics were selected, namely Eastern zone, Northern zone, Lake zone, Central Zone, and the Southern Highlands (Table 1).

In the second stage, regions with most amaranth production were selected from each AEZ: Dar-es-salaam and Pwani in Eastern zone, Arusha and Kilimanjaro in Northern zone, Mwanza in Lake zone, Dodoma in Central zone, and Iringa in the Southern Highlands (Fig. 1). In the third stage, three districts were selected from the selected regions in each AEZ. Within the selected districts, three wards known to be involved in vegetable or amaranth production were identified with the help of key informants (mostly District Agricultural, Irrigation and Cooperative Officers (DAICO) or horticultural officers). Two to four villages were then randomly selected from a complete list of villages. Amaranth farmers were listed with the help of a local lead farmer and a sample was drawn proportionate to the number of amaranth farmers. A total of 1,355 households were interviewed, about 270 from each zone.

Table 1 Distribution of the sample of amaranth-producing households in Tanzania, 2017–2018

Agroecological zone	Region(s)	Districts	Households
Eastern zone	Dar es Salaam, Pwani	3	271
Northern zone	Arusha, Kilimanjaro	3	270
Lake zone	Mwanza	3	270
Central zone	Dodoma	3	274
Southern Highlands	Iringa	3	270
Total		15	1,355

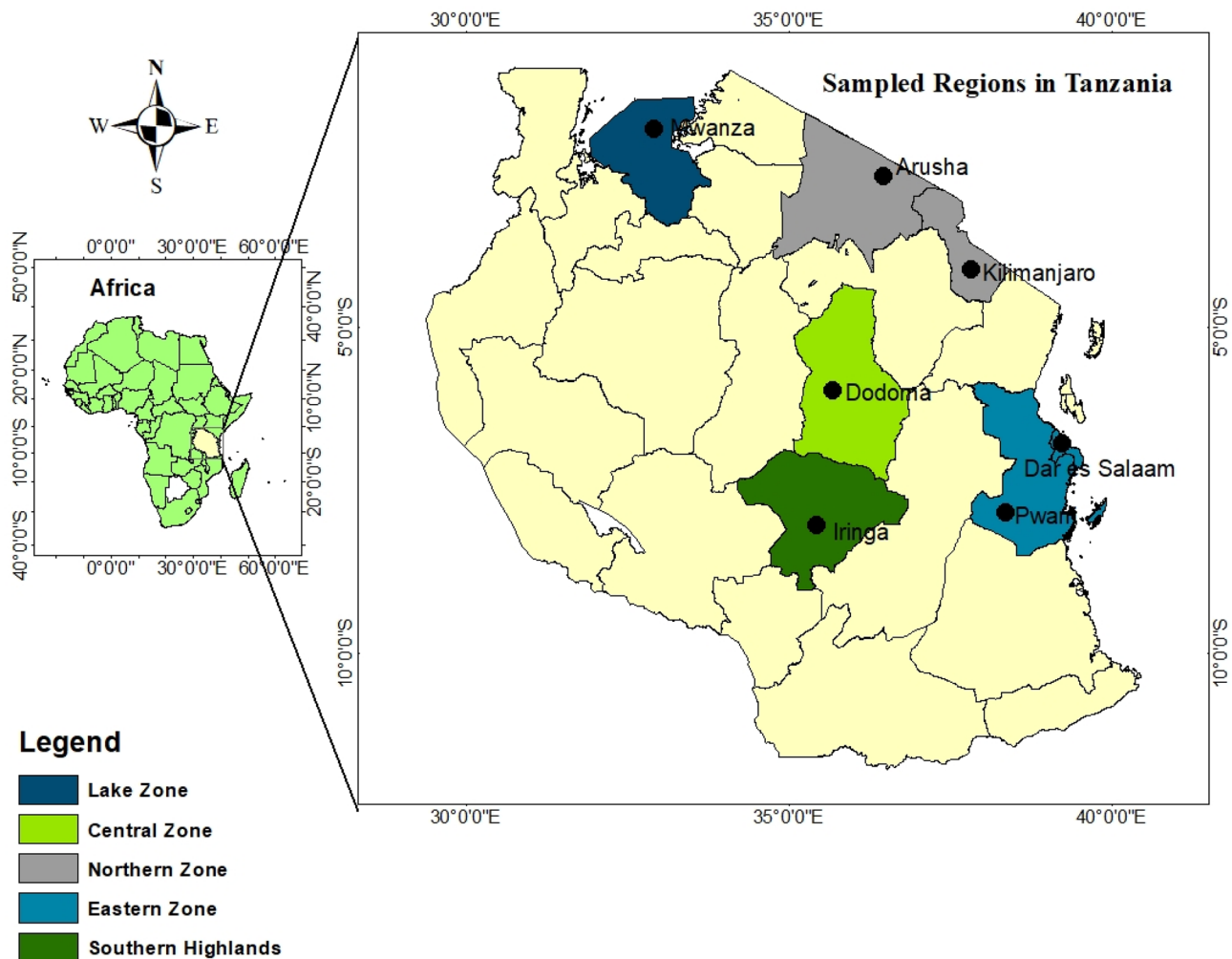


Fig. 1 Study areas in Tanzania

The respondent was the person in the household with most knowledge of amaranth production. The questionnaire contained sections on household demographics (household composition, age, gender, and education of all household members), crop production (cultivation techniques, management, harvesting, crop yield, and sales), and income, among others. The questionnaire was coded on tablet computers using SurveyCTO and interviews were conducted by trained enumerators. Respondents were informed about the purpose and scope of the interview, how the data will be used and by whom. Before the survey, each respondent was carefully informed that their participation was voluntary and was asked to sign a consent form before the interview or provide verbal consent where the respondent was unable to read and write.

2.2 DNA fingerprinting

Amaranth leaf samples were collected from farmers' fields at the same time as the survey data collection. Three to four pieces of young leaf approximately 3 cm long were sampled per plant and from each cultivar grown. One sample was taken from each field with the same plant species. Where farmers cultivated more than one variety of amaranth, each was given a different label. For farmers who had no amaranth in the field, a few seeds were collected and grown out at the WorldVeg research station in Arusha to collect the leaves. The leaf samples were dried before DNA extraction. DNA was extracted with Sbeadex Maxi Plant Kit (LGC, UK) according to the supplier manual. DNA of suitable quality for genotyping

by sequencing could be obtained for 1,007 out of 1,450 samples collected. The loss was high because samples had to be shipped from Tanzania to Taiwan for analysis.

DNA libraries and sequencing were completed by Diversity Arrays Technology Pty Ltd. (DarT, Australia) using DarTSeq™. DarTSeq routinely uses double-digested restriction enzyme (*PstI-MseI*) associated DNA sequencing to genotype the samples. The SNP calling followed the procedure described in Lin et al. (2021, 2022). The raw reads were filtered with a Phred quality score threshold of ≥ 20 in SolexaQA (Cox et al., 2010) and then de novo assembled using the “denovo-map.pl” pipeline in Stacks (version 2.54; Catchen et al., 2013), after which the SNPs were recalled. The SNP markers are available in Appendix Table A1 in the online data depository. Samples with a genotype missing rate $> 50\%$ were removed, resulting in a total of 955 samples kept in the DNA fingerprint analysis. The genotypes of 29 cultivars distributed to farmers in amaranth seed kits by WorldVeg and partners or released and commercialized as varieties were obtained from Lin et al. (2022) and included six amaranth species: *A. blitum*, *A. caudatus*, *A. cruentus*, *A. dubius*, *A. hypochondriacus*, and *A. retroflexus*.

A. dubius is a highly heterozygous tetraploid species (Lin et al., 2022) that is very different from the other amaranth species. We therefore used a different criterion to prepare the SNP matrix. In total, 22 k SNPs were available for the set of 29 reference genotypes. For *A. caudatus*, *A. cruentus*, *A. hypochondriacus*, *A. blitum* and *A. retroflexus* (a total of 24 cultivars), homozygous SNPs with minor allele frequency > 0.05 and missing data rate < 0.5 were kept, resulting in 2,063 SNPs, while for *A. dubius* (5 cultivars), also heterozygous SNPs were included, but SNPs with missing data were excluded, resulting in total in 9,020 SNPs. Samples with heterozygous SNP sites > 0.02 were grouped to *A. dubius*, and those with < 0.02 were not considered as *A. dubius* and instead grouped to the other amaranth species. After merging the SNP matrix for the known WorldVeg cultivars and the field samples, the SNPs were filtered again following the same principles described above. The SNP matrix was used to calculate genetic distance in TASSEL 5.0 (Bradbury et al., 2007). A field sample is claimed to be a WorldVeg cultivar if the genetic distance between itself and a known WorldVeg cultivar equals zero. The fingerprints obtained with these SNP matrices could not distinguish all field samples and released cultivars from each other. In order to resolve genetic distances between *A. hypochondriacus* accessions, 85 SNPs polymorphic in this group were selected and used to genotype field samples. Similarly, a set of 18 SNPs was chosen to further distinguish *A. dubius* accessions and cultivars from each other.

2.3 Propensity score matching

The use of cross-sectional data for impact evaluation may risk the presence of selection bias which occurs when characteristics that influence the outcomes also influence individuals’ decision to adopt a technology. We use propensity score matching (PSM) to correct for selection bias in observable household characteristics between adopters and non-adopters (Rosenbaum & Rubin, 1983). PSM has been widely used in the impact evaluation literature (Fischer & Qaim, 2012; Gautam et al., 2017; Gitonga et al., 2013; Ochieng et al., 2018).

PSM follows three steps. In the first step, adoption status is regressed upon a set of independent farm and household characteristics that simultaneously influence adoption and outcomes. The DNA fingerprinting results are used to separate households using amaranth cultivars that very likely originate from WorldVeg (labelled as “adopters”) from households that use other cultivars (“non-adopters”). WorldVeg cultivars are mostly improved while others are mostly local, though we cannot rule out the existence of other improved cultivars. Independent variables, selected based on our own knowledge of production systems in Tanzania, include: (a) gender of the household head which has an influence on decision making related to land allocation, type of crop to be cultivated, and is a proxy for access to information given that men have better access to information on new technologies than women; (b) age and education of the household head which have an influence on experience and knowledge in vegetable production; (c) household size as a driver of land-use intensity; (d) total area of land under cultivation which has an influence on the intensity of technology adoption, and also represents the wealth status of a household; (e) use of manure; (f) access to extension which is important for access to technical know-how; (g) membership in farmer groups as a proxy for knowledge sharing and innovativeness; and (h) location dummies. A logit model is estimated and predicted values are computed. These predicted values (i.e., the propensity scores) indicate the probability of a certain household adopting improved cultivars conditional on confounding factors.

In the second step, households are ranked according to their propensity scores and each adopter is compared with a set of most similar non-adopters using alternative matching algorithms (Caliendo & Kopeinig, 2008). We use the nearest-neighbor method, which matches each adopter to its most similar non-adopter. Observations without a good match are dropped from the analysis. We also use the radius and kernel matching methods as a robustness checks.

In the final step, the difference in outcome variables is calculated for each matched pair, after which these differences are averaged over the entire sample to obtain the average treatment effect. Three outcome indicators were used to

quantify impact. First, crop yield is measured as the quantity of amaranth leaves harvested expressed in tons per hectare. In Tanzania, amaranth leaves are the main output with only few households harvesting the grain and we therefore focus on leaf yield, calculated as the sum of harvested leaves in the short- and long-rain seasons. Second, the quantity sold is expressed in tons. Third, the cost of production is expressed in US dollars per hectare and calculated as the sum of costs incurred on land preparation, inputs (seed, fertilizer, pesticides etc.), and labor, specific to fields planted to amaranth.

Two key assumptions are tested while implementing the PSM technique. First, application of propensity scores requires that the distribution of covariates between adopters and non-adopters is similar or balanced (Caliendo & Kopeinig, 2008; Rosenbaum & Rubin, 1983). We check this using four tests: First, a two-sample t-test comparing covariate means before and after matching shows no significant difference after matching (Appendix Table A2). Second, a comparison of the pseudo- R^2 before and after matching shows it to decline from 0.057 before matching to 0.002 after matching indicating no systematic differences in the distribution of covariates. Third, the percentage bias after matching is less than 6% for all covariates after matching and 1.6% over all covariates, which is below the common thresholds of 20% and 10%, respectively. Fourth, after matching the variance ratio tests, Rubin's B should be less than 25 while Rubin's R should lie between 0.5 and 2 (Rubin, 2001). The results show that Rubin's B is 10.2 and while Rubin's R is 1.17 as required.

A second assumption is the presence of common support (overlap condition) in the propensity scores of adopters and non-adopters, which is confirmed through visual inspection (Appendix Figure A1).

3 Results

3.1 Sample characteristics

The means for adopters and non-adopters are not statistically different for any variable except household size, monthly per capita expenditure (a proxy for income), and cultivated area (Table 2). Non-adopters have a significantly higher monthly per capita expenditure (+ USD 3,060). Adopters have a significantly higher area under cultivation (+ 0.22 ha), but the average area under amaranth was not significantly different between adopters and non-adopters. The location variables show that adoption is higher in the Lake zone and Central zone of the country, and lower in the Eastern zone.

Seed use and cultivar identification.

Of the interviewed amaranth producers, 55% used own saved seed or seed obtained from other farmers—also known as the informal seed sector; while 42% bought seed

from stockists and local shops—that is, the formal sector (Table 3). Less than 1% of the amaranth seed used was obtained from public sector organizations (e.g., ministry of agriculture or research institutions). Seed from the informal sector may also be of improved cultivars as all cultivars are open-pollinated and farmers can save the seed.

Farmers were asked to tell the name of their amaranth cultivars during the 2017/2018 cropping period. Nine cultivars were listed, with adoption varying by AEZ (Table 4). The names can be confusing as some of the same cultivars are known under different names (e.g., *Madiira 1* and *Ex-Zim-Sel*) and many cultivars are known by their species name. *A. hypochondriacus* appears to be the most popular specie, cultivated by over 56% of farmers interviewed, followed by *A. dubius* (22%), and *A. cruentus* (15%). Popular cultivars identified by farmers include *Poli*, *Nguruma*, and *Madiira 2*, and a cultivar farmers call *A. hypochondriacus*. *Nguruma* is popular in the Northern zone of Arusha and Kilimanjaro, while *A. cruentus* is popular in the Eastern zone (Dar-es-salaam and Pwani) and Southern Highlands (Iringa).

Cultivar choice was driven by farmers' perceptions of product marketability (30%), high yield (17%) and early maturity for grain or seed growers (17%) (Appendix Table A3). These perceptions varied slightly across cultivars. For instance, farmers who cultivated *A. cruentus* and *A. dubius* species were more driven by the low cost of seed, and less by early maturity. *A. cruentus* was also preferred because it is relatively pest resistant. Those who cultivated *Madiira 1* were also driven by its taste and good pest resistance. Farmers who cultivated *Nguruma* (*A. dubius*) indicated that it is drought tolerant. This means that beyond marketability and yield, amaranth farmers prefer cultivars resistant to biotic and abiotic stresses. Most farmers (47%) for instance mentioned that pests and diseases are a challenge in amaranth production (Appendix Table A4). The strong preference placed on marketability shows that many households are interested in selling amaranth. The survey data show that 86% of farmers produce amaranth for both home consumption and selling (Table 5).

3.2 DNA fingerprinting

Table 6 shows amaranth cultivars or species grown by farmers as identified by DNA fingerprinting. Of the total samples assayed, 67.4% were found to contain genetic material originating from WorldVeg. Most of these cultivars were of *A. hypochondriacus* (46%), followed by *A. dubius* (14%), and *A. cruentus* (3.9%) and *A. blitum* (3.4%). Around 80% of the samples from the Lake and Central zones and 55% from the Southern Highlands were *A. hypochondriacus*, while the Northern zone had mainly *A. dubius*. About half of the samples from the Eastern zone did not originate from WorldVeg.

Table 2 Summary of household characteristics, sample means, 2017–2018

Household characteristic	Pooled	Adopters	Non-adopters	Difference ^a
Household head is male (1 = Yes)	0.85 (0.36)	0.84 (0.36)	0.82 (0.38)	-0.02 (0.03)
Age of the household head (years)	46.49 (12.91)	47.12 (13.10)	47.24 (12.79)	0.12 (0.93)
Education of the household head (years)	7.25 (3.63)	7.06 (3.52)	7.47 (3.28)	0.41 (0.25)
Farming is main occupation of head (1 = Yes, 0 otherwise)	0.81 (0.39)	0.82 (0.39)	0.80 (0.40)	-0.02 (0.03)
Person in charge of amaranth production is male (1 = Yes)	0.85 (0.36)	0.84 (0.36)	0.82 (0.38)	-0.02 (0.03)
Household size (persons)	5.19 (2.30)	5.32 (2.39)	4.80 (2.14)	-0.53*** (0.17)
Monthly per capita expenditure (1,000 USD)	15.64 (13.48)	14.90 (13.76)	17.97 (13.64)	3.06*** (0.99)
Accessed agricultural information (1 = Yes, 0 otherwise)	0.32 (0.47)	0.32 (0.47)	0.32 (0.47)	-0.00 (0.03)
Distance to the nearest extension service (km)	4.06 (9.26)	4.42 (9.73)	4.87 (9.11)	0.45 (1.69)
Sought agricultural information (1 = Yes, 0 otherwise)	0.16 (0.37)	0.16 (0.37)	0.16 (0.37)	-0.00 (0.03)
Has a savings/bank account (1 = Yes, 0 otherwise)	0.19 (0.40)	0.19 (0.39)	0.22 (0.41)	0.03 (0.03)
Received credit (1 = Yes, 0 otherwise)	0.16 (0.37)	0.17 (0.38)	0.18 (0.39)	0.01 (0.03)
Belongs to a farmer group (1 = Yes, 0 otherwise)	0.28 (0.45)	0.29 (0.46)	0.30 (0.46)	0.00 (0.03)
Area under cultivation (hectares)	1.03 (1.17)	1.00 (1.02)	0.78 (0.88)	-0.22*** (0.07)
Area under vegetables (hectares)	0.23 (0.31)	0.24 (0.34)	0.22 (0.31)	-0.02 (0.02)
Area under amaranth (hectares)	0.11 (0.18)	0.10 (0.16)	0.11 (0.17)	0.01 (0.01)
Eastern Zone (1 = Yes)	0.20 (0.40)	0.18 (0.38)	0.38 (0.49)	0.21*** (0.03)
Northern zone (1 = Yes)	0.20 (0.40)	0.17 (0.38)	0.21 (0.40)	0.03 (0.03)
Lake zone (1 = Yes)	0.20 (0.40)	0.24 (0.43)	0.10 (0.31)	-0.14*** (0.03)
Central zone (1 = Yes)	0.20 (0.40)	0.15 (0.36)	0.05 (0.22)	-0.10*** (0.02)
Southern Highlands (1 = Yes)	0.20 (0.40)	0.26 (0.44)	0.25 (0.44)	-0.00 (0.03)
Observations	1,355	592	287	879

Means with standard deviation in parentheses

Significant differences denoted by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

^aBased on t-test with standard errors in parentheses

Table 3 Sources of amaranth seed planted in Tanzania, 2017–2018

Source	Frequency	Proportion
Own saved and other farmers	881	0.55
Local shops, seed companies, seed stockists	684	0.42
NGOs/CBOs/Cooperatives	43	0.03
Public sector organizations	7	<0.01
Total	1,615	1.00

NGOs non-governmental organizations, CBOs community-based organizations

The total exceeds the sample size of 1,355 as farmers may obtain seed from multiple sources

The large group of *A. hypochondriacus* showed identical DNA fingerprints to three WorldVeg lines and one WorldVeg genebank accession. Similarly, the fingerprints of the *A. dubius* samples matched the fingerprints of one WorldVeg line and two WorldVeg genebank accessions. To assess whether the field samples were related to distributed WorldVeg cultivars or to genebank accessions, specific polymorphic marker sets for the *A. hypochondriacus* and *A. dubius* groups were used to determine the genetic distances between the WorldVeg cultivars, genebank accessions and field samples. It showed that 93% of the field samples attributed to the WorldVeg *A. hypochondriacus* group showed greatest genetic similarity to line TZSMN (VI062428), 3% were related to two other WorldVeg cultivars and 3% could not be distinguished from the genebank accession. From the *A. dubius* group, only 25% of the samples showed greatest genetic similarity to WorldVeg cultivar *Ex-Zan*. The other samples showed either greater or equal genetic similarity to two WorldVeg genebank accessions. All *A. blitum* field samples showed greatest similarity to a WorldVeg cultivar, and all *A. cruentus* field samples were related to genebank

Table 5 Farmers' purpose of producing amaranth in Tanzania, 2017–2018

Purpose	Frequency	Proportion
Home consumption only	93	0.07
Selling only	101	0.07
Home consumption and selling	1,161	0.86
Total	1,355	1.00

accessions. Overall, fingerprinting confirmed that 51.8% of the field samples show greatest genetic similarity to WorldVeg cultivars, and further 15.6% of the samples were related to WorldVeg genebank material. We compared the DNA fingerprinting results against the household survey data to determine the extent to which farmers were able to accurately identify their amaranth cultivars. The comparison is based on a subset of 708 households from whom only a single plant sample was collected. Figure 2 shows that farmers were correct 51% of the time.

3.3 Impact of improved cultivars

Table 7 summarizes the outcome variables used in the PSM model. Households that had adopted WorldVeg amaranth cultivars achieved a significantly higher mean yield (+4.65 tons), and lower cost of production (–USD 64.52) than non-adopters. There was no significant difference in the quantity of amaranth sold between adopters and non-adopters.

Controlling for other factors, the PSM model shows that the decision to use WorldVeg amaranth cultivars is largely influenced by locational variables. Households in the Northern zone, Southern Highlands, and Lake zone are more likely to use WorldVeg cultivars than those in the Eastern zone (Appendix Table A5). This may be because of differences in the suitability

Table 4 Amaranth cultivars based on farmers knowledge, in percentage of all respondents per agroecological zone, 2017–2018

Cultivar name according to farmers (species name)	Eastern zone	Northern zone	Lake zone	Central zone	Southern Highlands	Total
<i>Madiira 1 / Ex-Zim-Sel (A. cruentus)</i>	0.3	1.5	-	0.7	8.9	2.4
<i>Madiira 2 / AM 38-Sel (A. cruentus)</i>	22.1	2.2	6.1	5.4	20.7	11.7
<i>Poli / AH-TL-Sel (A. hypochondriacus)</i>	35.3	3.3	51.4	49.1	37.5	35.4
<i>Nguruma / UG-AM-9-ES13-2 (A. dubius)</i>	3.5	53.5	1.8	1.1	4.9	12.5
Known as <i>A. hybridus</i>	1.9	5.8	11.8	5.8	5.9	6.2
Known as <i>A. hypochondriacus</i>	26.6	1.1	23.6	35.0	17.4	20.9
Known as <i>A. cruentus</i>	-	2.2	1.1	1.1	2.0	1.2
Known as <i>A. dubius</i>	9.9	29.8	3.9	1.1	1.6	9.1
Known as <i>A. spinosus/ A. viridus</i>	0.3	0.7	-	-	1.0	0.4
Farmer does not know	-	-	0.4	0.7	-	0.2
Observations (<i>n</i>)	312	275	280	277	304	1,448

Number of observations exceeds the sample size as households can use multiple cultivars

Table 6 Amaranth cultivars as identified by DNA fingerprinting disaggregated by region, in percentage of all observations

	Eastern zone	Northern zone	Lake zone	Central zone	Southern Highlands	Total
Non-WorldVeg cultivars	51.4	36.4	17.3	14.3	32.4	32.7
WorldVeg cultivars	48.5	63.5	82.7	85.9	67.6	67.4
Of which by species:						
– <i>A. hypochondriacus</i>	25.7	1.2	79.8	81.9	54.7	46.0
– <i>A. dubius</i>	10.7	60.5	<0.1	1	0.9	14.1
– <i>A. cruentus</i>	12.1	1.8	<0.1	2	1.3	3.9
Of which by cultivars:						
<i>Madiira 1</i>	<0.1	1.2	<0.1	1	0.9	0.6
<i>Madiira 2</i>	<0.1	<0.1	<0.1	<0.1	0.4	0.1
Other	12.1	0.6	<0.1	1	<0.1	3.2
– <i>A. blitum</i>	<0.1	<0.1	2.9	1	10.7	3.4
Observations (<i>n</i>)	214	162	173	105	225	879

of cultivars or because of differences in efforts to introduce the cultivars to farmers. Evidently, only 6% of the observed variation in crop yield, cost of production and quantity sold is explained by the covariates as indicated by the R-squared.

Impact estimates show that households using WorldVeg cultivars achieved a mean amaranth leaf yield that was 6.1 tons/ha higher ($p=0.002$ using nearest neighbor matching) than those who used other cultivars (Table 8). This suggests a crop yield premium of 48%. The impact on the cost of production and quantity sold are not significant. Radius matching and kernel-based matching give the same results.

4 Discussion

4.1 Strengths and limitations of the study

This is the first application of DNA fingerprinting to quantify the adoption of vegetable cultivars and the first to combine DNA fingerprinting with propensity score matching

to quantify the impact of improved cultivars. DNA fingerprinting provides a unique way of quantifying adoption of improved crop cultivars and we used nationally representative data of amaranth-producing households in Tanzania.

The fingerprinting study applied highly stringent criteria to determine the relationship between field samples and WorldVeg material. An overestimation of the adoption of WorldVeg amaranth lines is therefore unlikely. However, the study may have missed WorldVeg materials that have not been documented as having been released in the region, which may result in underestimating the adoption of WorldVeg cultivars. While DNA fingerprinting is thought to give more accurate estimates of cultivar adoption compared household surveys or expert opinions, it needs to be kept in mind that the method assumes farmers' fields to be genetically uniform, while in reality, farmers' amaranth fields may have different genotypes. Amaranth is mainly self-pollinating, but seed saving and the use cultivar mixtures may introduce genetic variation. Consequently, a single sample per field may not represent the genotype of the whole field.

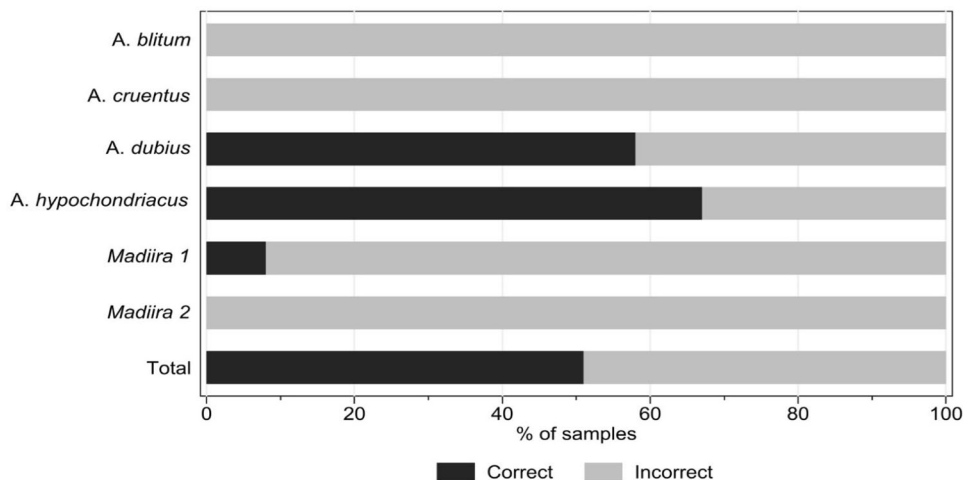
Fig. 2 Percentage of correctly identified amaranth cultivars by farmers

Table 7 Sample means of the outcome variables used

Variable	Pooled	Adopters	Non-adopters	Difference ^a
Crop yield (tons/ha)	16.37 (29.49)	17.38 (30.56)	12.73 (19.21)	4.65** (2.04)
Quantity sold (tons)	0.44 (0.84)	0.42 (0.78)	0.47 (1.02)	-0.05 (0.06)
Cost of production (USD)	163.56 (327.56)	151.27 (278.97)	215.79 (385.45)	64.52*** (22.85)
Observations (n)	1,355	592	287	879

Means with standard deviation in parentheses

Significant differences denoted by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. ^aBased on t-test with standard errors in parentheses

Vegetable amaranth has been collected by WorldVeg and other organizations in Tanzania and may have been used for developing improved cultivars. This opens up the possibility that germplasm collected from the region is still in use by local farmers and could have been erroneously classified as an improved WorldVeg cultivar by the fingerprinting study. The WorldVeg genebank holds 58 *Amaranthus* accessions collected from Tanzania and one accession, VI059412 collected from Tanzania in 2002, has an identical DNA fingerprint to three improved WorldVeg cultivars, namely *TZSMN 102*, *AH-NL* and *AH-TL*. Investigation of the genetic relationship among these materials with 85 markers specifically designed for this germplasm corroborated that 93% of the field samples fingerprinted as *A. hypochondriacus* showed greatest similarity to cultivar *TZSMN 102* and further 3% of the samples had greatest similarity with two other WorldVeg cultivars, while only 3% of the field samples was more or equally similar to genebank accession VI059412. This strongly indicates that most of these *A. hypochondriacus* field samples were indeed derived from improved WorldVeg lines rather than the original genebank accession. Similarly, the *A. dubius* accessions VI050451 and VI50448 originating from Tanzania could not be distinguished from WorldVeg cultivar *Ex-Zan*. Here, a smaller polymorphic marker set was available allowing the attribution of only 25% of the samples of the *A. dubius* group to the improved WorldVeg cultivar *Ex-Zan*, while the remaining samples were more or equally

similar to *A. dubius* genebank materials collected from the region. In conclusion, based on genetic similarity, 51.8% of the field samples are most likely derived from improved WorldVeg material.

As a possible weakness of the study, we note that the use of propensity score matching on cross-section data of a non-randomized intervention may not fully eliminate selection bias. Panel data would be preferred to better understand trends and drivers of cultivar adoption. The analysis also focused on a few relatively easy to measure outcome variables (mainly crop yield), while further analyses could investigate income and nutrition effects.

4.2 Comparison of methods

The DNA fingerprinting method showed that 67% of the amaranth samples originated from genetic material supplied by WorldVeg. The adoption rate for WorldVeg amaranth cultivars is estimated to be 66% of the total area under amaranth in Tanzania. This is 20% higher than what was reported earlier using expert elicitation techniques (Ochieng et al., 2019). While previous studies showed that expert elicitation tends to overestimate adoption rates (e.g., Maredia et al., 2016), our study shows that it can also underestimate adoption rates as the farm survey data showed the adoption of improved cultivars to be 51%, which is 16% lower than the DNA fingerprinting results.

Table 8 Impact of WorldVeg amaranth cultivars on crop yield, sales and cost of production in Tanzania using three alternative propensity score matching algorithms, 2017–2018

Outcome variable	Nearest neighbor matching			Radius matching			Kernel based matching		
	ATE	SE	P> z	ATE	SE	P> z	ATE	SE	P> z
Crop yield (tons/ha)	6.11 ***	1.94	0.002	6.11 ***	2.20	0.005	6.11 ***	2.30	0.008
Quantity sold (ton)	0.03	0.09	0.71	0.03	0.10	0.753	0.03	0.09	0.725
Cost (USD/ha)	17.09	31.93	0.592	17.09	24.30	0.482	17.09	28.78	0.553

ATE average treatment effect, SE bootstrapped standard error

Significant differences denoted by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5 Conclusion

In Tanzania, 67% of amaranth-producing households use amaranth cultivars that are likely derived from breeding lines and genebank accessions coming from the international amaranth breeding program of World Vegetable Center in Tanzania. The WorldVeg cultivars had a mean yield that is + 6.1 tons/ha higher (+ 48%) than other cultivars. This shows that international vegetable breeding research on traditional African vegetables can create tremendous impact as the lack of previous research in these crops can lead to relatively quick and sizeable improvements in crop performance.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12571-023-01378-0>.

Data availability Data for both DNA fingerprinting and household survey has been made available.

Declarations

Conflicts of interest The authors declared that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31–72. <https://doi.org/10.1111/j.1467-6419.2007.00527.x>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. <https://doi.org/10.1111/mec.12354>
- Cernansky, R. (2015). The rise of Africa's super vegetables. *Nature*, *522*(7555), 146–148. Retrieved February 7, 2022, from <https://www.nature.com/articles/522146a>
- Chelang'a, P., Obare, G., & Kimenju, S. (2013). Analysis of urban consumers' willingness to pay a premium for African Leafy Vegetables (ALVs) in Kenya: A case of Eldoret Town. *Food Security*, *5*(4), 591–595. <https://doi.org/10.1007/s12571-013-0273-9>
- Cox, M. P., Peterson, D. A., & Biggs, P. J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics*, *11*(1), 1–6. <https://doi.org/10.1186/1471-2105-11-485>
- Dinssa, F., Yang, R., Ledesma, D., Mbwambo, O., & Hanson, P. (2018). Effect of leaf harvest on grain yield and nutrient content of diverse amaranth entries. *Scientia Horticulturae*, *236*, 146–157. <https://doi.org/10.1016/j.scienta.2018.03.028>
- Fischer, E., & Qaim, M. (2012). Gender, agricultural commercialization, and collective action in Kenya. *Food Security*, *4*(3), 441–453. <https://doi.org/10.1007/s12571-012-0199-7>
- Floro IV, V. O., Labarta, R. A., Becerra López-Lavalle, L. A., Martínez, J. M., & Ovalle, T. M. (2018). Household determinants of the adoption of improved cassava varieties using DNA fingerprinting to identify varieties in farmer fields: A case study in Colombia. *Journal of Agricultural Economics*, *69*(2), 518–536. <https://doi.org/10.1111/1477-9552.12247>
- Gautam, S., Schreinemachers, P., Uddin, M. N., & Srinivasan, R. (2017). Impact of training vegetable farmers in Bangladesh in integrated pest management (IPM). *Crop Protection*, *102*, 161–169. <https://doi.org/10.1016/j.cropro.2017.08.022>
- Gitonga, Z. M., De Groote, H., Kassie, M., & Tefera, T. (2013). Impact of metal silos on households' maize storage, storage losses and food security: An application of a propensity score matching. *Food Policy*, *43*, 44–55. <https://doi.org/10.1016/j.foodpol.2013.08.005>
- Kosmowski, F., Aragaw, A., Kilian, A., Ambel, A., Ilukor, J., Yigezu, B., & Stevenson, J. (2019). Varietal identification in household surveys: Results from three household-based methods against the benchmark of DNA fingerprinting in southern Ethiopia. *Experimental Agriculture*, *55*(3), 371–385. <https://doi.org/10.1017/S0014479718000030>
- Lin, Y.-P., Wu, T.-H., Chan, Y.-K., van Zonneveld, M., & Schafleitner, R. (2021). De novo SNP calling reveals the candidate genes regulating days to flowering through interspecies GWAS of Amaranthus genus. *bioRxiv*. <https://doi.org/10.1101/2021.10.05.463269>
- Lin, Y.-P., Wu, T.-H., Chan, Y.-K., van Zonneveld, M., & Schafleitner, R. (2022). De novo SNP calling reveals the genetic differentiation and morphological divergence in genus Amaranthus. *The Plant Genome*, *15*(2), e20206. <https://doi.org/10.1002/tpg2.20206>
- Maredia, M. K., Reyes, B. A., Manu-Aduening, J., Dankyi, A., Hamazakaza, P., Muimui, K., Rabbi, I., Kulakow, P., Parkes, E., & Abdoulaye, T. (2016). *Testing alternative methods of varietal identification using DNA fingerprinting: Results of pilot studies in Ghana and Zambia* (International Development Working paper Issue 149). <https://doi.org/10.22004/ag.econ.246950>
- Mwadzingeni, L., Afari-Sefa, V., Shimelis, H., N'Danikou, S., Figlan, S., Depenbusch, L., Shayanowako, A. I. T., Chagomoka, T., Mushayi, M., Schreinemachers, P., & Derera, J. (2021). Unpacking the value of traditional African vegetables for food and nutrition security. *Food Security*. <https://doi.org/10.1007/s12571-021-01159-7>
- Ochieng, J., Afari-Sefa, V., Karanja, D., Kessy, R., Rajendran, S., & Samali, S. (2018). How promoting consumption of traditional African vegetables affects household nutrition security in Tanzania. *Renewable Agriculture and Food Systems*, *33*(2), 105–115. <https://doi.org/10.1017/S1742170516000508>
- Ochieng, J., Schreinemachers, P., Ogada, M., Dinssa, F. F., Barnos, W., & Mndiga, H. (2019). Adoption of improved amaranth varieties and good agricultural practices in East Africa. *Land Use Policy*, *83*, 187–194. <https://doi.org/10.1016/j.landusepol.2019.02.002>
- Oruko, L., Gracious, D., Tairo, F., & Amanquah, S. (2015). Diffusion and adoption of improved rice and maize varieties in Tanzania:

- application of genetic fingerprinting technique. Retrieved February 7, 2022, from <https://agra.org/wp-content/uploads/2020/07/Tanzania-Rice-and-Miawe-adoption-study-using-DNA-Fingerprinting.pdf>
- Rabbi, I. Y., Kulakow, P. A., Manu-Aduening, J. A., Dankyi, A. A., Asibuo, J. Y., Parkes, E. Y., Abdoulaye, T., Girma, G., Gedil, M. A., Ramu, P., Reyes, B., & Mareid, M. K. (2015). Tracking crop varieties using genotyping-by-sequencing markers: A case study using cassava (*Manihot esculenta* Crantz). *BMC Genetics*, *16*(1), 115. <https://doi.org/10.1186/s12863-015-0273-1>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>
- Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*, 169–188. <https://doi.org/10.1023/A:1020363010465>
- Stoilova, T., van Zonneveld, M., Roothaert, R., & Schreinemachers, P. (2019). Connecting genebanks to farmers in East Africa through the distribution of vegetable seed kits. *Plant Genetic Resources: Characterization and Utilization*, *17*(3), 306–309. <https://doi.org/10.1017/S1479262119000017>
- Wineman, A., Njagi, T., Anderson, C. L., Reynolds, T. W., Alia, D. Y., Wainaina, P., Njue, E., Biscaye, P., & Ayieko, M. W. (2020). A Case of Mistaken Identity? Measuring Rates of Improved Seed Adoption in Tanzania Using DNA Fingerprinting. *Journal of Agricultural Economics*, *71*(3), 719–741. <https://doi.org/10.1111/1477-9552.12368>
- Wossen, T., Abdoulaye, T., Alene, A., Nguimkeu, P., Feleke, S., Rabbi, I. Y., Haile, M. G., & Manyong, V. (2019). Estimating the productivity impacts of technology adoption in the presence of misclassification. *American Journal of Agricultural Economics*, *101*(1), 1–16. <https://doi.org/10.1093/ajae/aay01>
- Yang, R. Y., & Keding, G. B. (2009). Nutritional contributions of important African indigenous vegetables. In Shackleton, C. M., Pasquini, M. W., & Drescher, A. W. (Eds.), pp. 137–176. *Routledge*.



Rosina Wanyama is a Postdoctoral Scientist at the World Vegetable Center. She is an Agricultural Economist working with the impact evaluation team at the Center. Her previous work was focused on understanding the diets and consumption patterns of the urban poor households in Africa. Her current research work is based on projects promoting the use of regenerative technologies, inclusive vegetable value chains, vegeta-

ble business networks and home-grown school feeding programs. Rosina holds a PhD in Agricultural Economics at the University of Goettingen, Germany, and an MSc in Agricultural Economics from Egerton University, Kenya.



Pepijn Schreinemachers is an Agricultural Economist with the World Vegetable Center where he leads the flagship research programs on “Enabling Impact” and “Healthy Diets”. Before joining the Center in 2012, he was project coordinator of “The Uplands Program” at Hohenheim University. Pepijn holds a PhD from the Center for Development Research (ZEF), University of Bonn, Germany and an MSc in Development Studies from Wageningen University, the Netherlands.



Justus Ochieng is an Agricultural Economist and an Impact Evaluation professional with an extensive experience in the agriculture sector. His current research work involves agricultural value chain development and analysis, impact assessment (ex-ante and ex-post impact analysis), climate change and adaptation, technology adoption and scaling, and youth development. Justus holds a PhD in Agricultural Economics from the University of Kassel, Germany.



Omary Mbwambo is research associate at the World Vegetable Center. He has a background in Horticulture and currently working with the Vegetable Diversity and Improvement team at the Center. He is in-charge of farm operations where he is involved in capacity building of farmers and stakeholders through training, and technology demonstration and dissemination. Omary holds an MSc in Research Methods from Jomo Kenyatta University of Agriculture and Technology,

Kenya and a BSc in Horticulture from Sokoine University of Agriculture, Tanzania.



Roselyne Alphonse is a Senior Lecturer at the College of Economics and Business Studies (CoEBS) at the Sokoine University of Agriculture in Tanzania, and a Food System Research Network for Africa (FSNeT-Africa) fellow. Her research work has been focused on food

systems research, with specific interest on food quality interventions, impact evaluation, consumer preference, experimental economics, non-market valuation techniques and value chain analysis. Roselyne holds a PhD in Economics with specialization in Food, Consumer and Experimental Economics from the Norwegian University of Life Sciences (NMBU).



Fekadu Fufa Dinssa is currently a vegetable breeder at the World Vegetable Center (WorldVeg) Eastern and Southern Africa. Prior to joining WorldVeg, he worked as postdoc at International Center for Agricultural Research in the Dry Areas (ICARDA), Ethiopian Institute of Agricultural Research, Institute of Biodiversity of Ethiopia. He holds a PhD in plant breeding/genetics from Jordan University, MSc in plant breeding from Wageningen University and BSc in plant sciences

from Haramaya University. His research interest areas are participatory breeding and bring underutilized crops to the farming systems.



Ya-ping Lin is a Postdoctoral Scientist at the World Vegetable Center headquarters in Taiwan. She is an omics breeder who aims to apply omics data and omics technology into vegetable breeding programs. Currently her work focuses on pre-breeding of tomato and pepper. She is also a geneticist who studies the germplasm diversity and the genetic architecture of desired traits in *Solanum*, *Vigna*, and *Amaranthus* species. She obtained her PhD in Agronomy from the

National Taiwan University, Taipei, Taiwan.



Roland Schafleitner is Flagship Program Leader for Vegetable Diversity and Improvement and Head of Molecular Genetics at the World Vegetable Center based in Taiwan. Prior to that, he worked as researcher at the International Potato Center (CIP) in Lima, Peru, and at the Austrian Research Center Seibersdorf (now AIT) in Austria. He holds a PhD in Biochemical Technology and a MSc in Molecular Microbiology. His research interest is to mobilize vegetable bio-

diversity for diversification of agricultural production and crop improvement.