RESEARCH ARTICLE

# Characterization of a sex-determining locus and development of early molecular assays in *Telfairia occidentalis* Hook. F., a dioecious cucurbit

Adrian Metry · Oyenike Arike Adeyemo ·
Sandra Grünig · Christian Parisod

**Abstract** Although there exist over 7000 crop species, only a few are commercially valuable and grown on a large scale in monocultures worldwide. However, underutilised crops (also called orphan crops) have significant potential for food security and *Telfairia occidentalis* Hook. F. (Cucurbitaceae) is one such orphan crop grown in West Africa for its nutritious leaves, oil and protein-rich seeds. In this dioecious crop, farmers like to eliminate male plants and keep mostly females to increase their yield. However, they face the challenge of determining sex due to limited morphological differences between females and males before flowering. This study used double digested restriction site-associated DNA sequencing data (ddRADseq) to examine the genetic diversity within and among landraces of *T. occidentalis*, identify common sex-determining loci, and establish reliable assays to characterize the sex of immature plants in the vegetative state. To differentiate males from females of *T. occidentalis*, two molecular assays were thereupon developed based on polymerase chain reaction (PCR) to genotype sex-specific sequence variation either through restriction by Mfe1 or the direct use of sex-specific primers. Both assays require standard laboratory conditions to reach a certainty of 94.3% for females and 95.7% for males from the studied samples. With the inclusion of additional landraces, medium to largescale farms growing *T. occidentalis* as a crop can readily benefit from an early determination of the sex of plants.

A. Metry · S. Grünig · C. Parisod (✉)
Department of Biology, University of Fribourg, Fribourg, Switzerland
e-mail: christian.parisod@unifr.ch

A. Metry
e-mail: adrian.metry@gmail.com

S. Grünig
e-mail: sandra.gruenig@unifr.ch

O. A. Adeyemo
Department of Cell Biology and Genetics, University of Lagos, Akoka, Yaba, Lagos, Nigeria
e-mail: aoadeyemo@unilag.edu.ng

## Introduction

World food security depends on only a small fraction of the thousands of crops available (Jaenicke et al. 2009; Martin et al. 2019). Many of the underutilized crop species, also known as orphan crops, have never been extensively cultivated locally or regionally. One underlying reason is that large investments in research and breeding are driven mainly by the commercial interests of developed countries and with little profit

from orphan crops (World Bank 2007). Underutilised crops do not appear to be sufficiently competitive with staple food crops grown in the same agricultural area. Therefore, farmers prefer already well-established crops to provide for the world food market (Padulosi and Hoeschle-Zeldon 2004).

The term "underutilised" may refer to an intrinsic lack of competitiveness, or to a crop that has not yet realized its full agricultural potential for particular stakeholders or in specific regions. For example, chickpeas (*Cicer arietinum*) are largely cultivated in Asia, but they are considered to be underutilised in Italy (Padulosi et al. 2002; Padulosi and Hoeschle-Zeldon 2004). Many orphan crops have great potential, to serve as an innovative, sustainable and safe food source under climate change conditions, especially for developing countries (Dawson et al. 2009; Raheem 2011; Chivenge et al. 2015; Mabhaudhi et al. 2019; Tadele 2019). Their nutrients, medicinal effects and biodiversity can contribute a great value to achieving the Millennium Development Goals and act against unbalanced diets. Since the large diversity of underutilised crops that are known regionally and locally is often also of great cultural value and supports social heterogeneity (Jaenicke and Hoeschle-Zeledon 2009), orphan crops adapted to their area of origin may outperform dominant world crops in many circumstances (Fahey 1998). Despite their huge potential, orphan crops have not received enough attention in terms of initial improvement in their quantity and quality, which would not only increase awareness and investment, but also enhance food security in underdeveloped countries (Tadele 2019).

Fluted pumpkin (*Telfairia occidentalis* Hook. F.), commonly known as Ugu, is one of many orphan crops with great potential and for which increased knowledge and resources would be beneficial to improve food security. Due to their high productivity, cultivated landraces of *T. occidentalis* are maintained by small farmers as a major nutritional food and source of income for their livelihood. It is a dioecious flowering plant in the *Cucurbitaceae* family (Fayeun et al. 2016a; Okoli and Mgbeogu 1983), whose sex expression is likely genetically determined (Grumet and Taft 2011). Thought to originate from Southern Nigeria, the species is mostly cultivated in West Africa for its nutritious and edible leaves as well as healthy oils and protein-rich seeds (Akoroda 1990; Akoroda et al. 1990; Okoli and Mgbeogu

1983; Badifu 1993). Individuals of *T. occidentalis* can reach a vine length of up to 4.7 m when flowering (Nwonuala and Obiefuna 2015) and female plants can produce two to five large fruits weighing 2–20 kg containing up to 200 flat, round seeds around 5 cm in diameter (Adeyemo and Tijani 2018; Okoli and Mgbeogu 1983). Since only female plants produce useful fruits containing seeds and larger succulent leaves than the males of *T. occidentalis*, they are considered more beneficial and male plants tend to be regarded as a waste of energy (Akoroda 1990; Chukwurah and Uguru 2010; Fayeun et al. 2016b). Despite reports of morphological and floral differences between females and males (Chukwurah and Uguru 2010; Fayeun et al. 2016a; Okoli and Mgbeogu 1983), neither morphological nor molecular traits have yet been identified to successfully support reliable sex determination of immature males versus females of *T. occidentalis* (Ndukwu et al. 2005; Fayeun and Odiyi 2015).

A few studies have used site-associated DNA sequencing data (RADseq) to determine plant sex (Gamble et al. 2015; Jeffries et al. 2018; Scharmann et al. 2019; Morgan et al. 2020), including *Cucurbitaceae* (Matsumura et al. 2014). Undoubtedly, RADseq appears to be versatile and robust enough to characterize genome-wide polymorphisms in non-model species that do not have an existing reference genome and therefore enable the reliable identification of loci associated with sex determination (Feron et al. 2021; Palmer et al. 2019; Peterson et al. 2012). In this study, we used single nucleotide polymorphisms (SNPs) to examine the structure of genetic variation within and among populations of selected landraces of *T. occidentalis* grown in Nigeria. We also identified possible sex-determining loci in this orphan crop and used consistent sequence variation between males and females among the landraces to develop, design and validate molecular assays for determining the sex of immature *T. occidentalis* plants.

## Materials and methods

### Plant samples and DNA extraction

For each cultivated landraces of *T. occidentalis* from five states in Nigeria (Lagos, Anambra, Cross River, Oyo, Ekiti; Table 1), which are representatives of the

**Table 1** Samples of *Telfairia occidentalis* collected from cultivated landraces in five States of Nigeria used in this study

| States (towns) | Sampling size | Samples with sex phenotyped | Females | Males |
|---|---|---|---|---|
| Anambra (Akwa) | 26 | 4 | 2 | 2 |
| Lagos (Lagos) | 29 | 21 | 9 | 12 |
| Oyo (Ibadan) | 14 | 11 | 7 | 4 |
| Cross River (Calabar) | 23 | 22 | 13 | 9 |
| Ekiti (Ado Ekiti) | 47 | 42 | 22 | 20 |
| Total | 139 | 100 | 53 | 47 |

species distribution, a whole fruit was collected and between 32 and 124 seeds were obtained (Table 1). Seeds were germinated and grown in nursery bags and transplanted in a protected field in the Post and Telecommunication area of Baruwa/Ipaja, Alimosho Local Government, Lagos, and in the field at University of Lagos, Nigeria in between December 2019, and July 2020. The sampled plants were allowed to grow to maturity and young leaves were collected from a total of 139 individual plants for DNA extraction. Leaf samples were immediately preserved and stored in silica gel for later transfer from Nigeria to Switzerland. For molecular analyses, we genotyped 139 plants, out of which a total of 100 plants reached sexual maturity and were phenotyped as male vs female. The sexes were identified by the morphology of fully developed distinct male/female flowers in the field.

Genomic DNA was extracted from 10 to 20 mg of dried leaf tissue for each of the 139 plant samples using the QIAGEN DNeasy plant kit as recommended by the manufacturer (Qiagen, Hombrechtikon, Switzerland). The quality and quantity of each DNA sample were checked using NanoDrop ND-1000 Spectrophotometer and Qubit 3 Fluorometer (both by Thermo Fischer Scientific). Electrophoresis on a 1% agarose gel was used to perform the DNA integrity of a subset of samples.

Generating the ddRAD sequencing data set

Three ddRAD libraries (Table S1) were prepared from the 139 DNA samples of *T. occidentalis*, together with one randomly selected technical replicate from each landrace (total of 144 samples), following the protocol of Peterson et al. (2012) as modified by Grünig et al. (2021). In brief, 150 ng of DNA was digested with the restriction enzymes EcoR1

and Mse1 (New England Biolabs) for 1 h at 37 °C. Reads from each sample were individually tagged by ligating digested DNA to EcoR1 adapters including one of 48 barcodes of five base pairs (bp), and to biotin-tagged MseI adapters including three different indexes (Table S2). Mse1 indexes also contained four degenerated bp which allowed PCR duplication to be identified. For each library, fragments around 550 bp were selected with AMPure XP beads (Agencourt) and stored with Dynabeads M-270 Streptavidin (Invitrogen). Each library was PCR amplified in a thermal cycler (Eppendorf Mastercycler) with initial denaturation at 98 °C, followed by ten cycles of 10 s at 98 °C, 30 s at 65 °C and 30 s at 72 °C.

The concentration of PCR-products was estimated using Qubit high-sensitivity assays (Thermo Fischer) and the distribution of fragment sizes in the library was determined using an Agilent 2100 Bioanalyzer (Agilent Technologies) following the manufacturer's protocol. The three libraries with a total of 144 randomly distributed samples ended up being pooled at equimolarity and paired-end sequenced in 300 cycles ($2 \times 150$ bp) on an SP flow cell of the Illumina NovaSeq 6000 at the next-generation sequencing (NGS) platform of the University of Bern.

Sequenced raw reads were examined for quality, demultiplexed and checked for intact restriction associated DNA cut sites and barcodes using *process radtags* (Catchen et al. 2011, 2013). PCR clones were identified by *clone filter* and removed (Catchen et al. 2011, 2013). Illumina adapters were trimmed, and reads were quality-filtered using *trimmomatic* (Bolger et al. 2014). Only reads of 100 bp or more, which also met an average Phred quality score of 15 within a four base sliding window, were kept.

Genotyping was performed following the dDocent pipeline (Puritz et al. 2014). Accordingly, a de novo catalogue was created based on reads that appeared

at least twice and in four individuals (first and second cutoffs) and that were clustered into contigs using *Rainbow* based on at least 10 reads with a minimum similarity set to 0.5 (Chong et al. 2012). The *Rainbow* assembly was checked for overlaps between newly assembled forward and reverse reads using *PEAR* (Zhang et al. 2014). Contigs from *Rainbow* were then aligned with *cd-hit-est*, allowing for 80%, 85% and 90% sequence similarity (Li and Godzik 2006) and resulting in three different de novo catalogues upon which the one based on 85% sequence similarity was selected as most robust for downstream analyses (Table S4). Raw reads were aligned to the de novo catalogue with the programme *Burrows-Wheeler Alignment tool (BWA;* Li and Durbin 2009) whereas mapping quality was filtered using *sambamba* (Tarasov et al. 2015). *Samtools* (Li et al. 2009) indexed the reference sequence and *HaplotypeCaller* and *GatherVcfs* from the *gatk4* package (McKenna et al. 2010) were then used for SNP calling and the gathering of generated VCF files into one *VCFtools* (Danecek et al. 2011) was then applied for filtering of what were earlier called SNPs.

The ddRAD sequencing error rate was calculated as the number of allelic differences between the technical replicates and the corresponding original sample divided by the total number of comparisons (Bonin et al. 2004).

Genetic diversity of the five landraces

Genetic variation from SNPs (18,507 SNPs on 8657 contigs) among individual samples of *T. occidentalis* was checked for population structure within and among landraces with a principal component analysis (PCA) using the R libraries "vcfR", "adegenet", "adegraphics", "ggplot2" and "reshape" in R Studio (version 1.3.1093).

Genetic variation was pruned from SNPs in linkage disequilibrium using *plink2* (Purcell and Chang, www.cog-genomics.org/plink/2.0/; Chang et al. 2015) and the pruned dataset (9077 SNPs out of 18,507) was used in STRUCTURE (version 2.3.3; Porras-Hurtado et al. 2013; Pritchard et al. 2000) with the admixture model of ancestry to infer the partitioning of genetic variation within *T. occidentalis*. Ten iterations of STRUCTURE for K = 1, 2, 3, 4 and 5 were performed with a burn-in period of 100,000 and 1,000,000 iterations. Results were processed using

STRUCTURE HARVESTER (Earl and vonHoldt 2012) and the K-value that best describes data was evaluated as in Evanno et al. (2005). STRUCTURE outputs were firstly aggregated with CLUMPP (version 1.1.2; Jakobsson and Rosenberg 2007) and then used to create the bar plots in R studio.

Identifying sex-determining loci

Genetic variation associated with sex phenotypes (i.e., putative sex-determining loci) was estimated using the hierarchical AMOVA framework in ARLEQUIN (version 3.5; Excoffier and Lischer 2010). Samples from the Anambra landrace were taken out of this analysis because the low number of females and males would not have contributed to strong evidence (Table 1). The partitioning of pruned genetic variation among 96 samples of *T. occidentalis* with known sex (51 females + 45 males) was thus estimated among four groups (i.e., landraces), with individuals of different sexes as nested populations (Excoffier and Lischer 2015). As a result, $F_{SC}$-values relate to the differentiation between males and females within and among landraces, identifying loci with high $F_{SC}$-values indicative of SNPs that are strongly differentiated between sexes and that may be part of or tightly linked to sex-determining loci (Charlesworth 2016; Holsinger and Weir 2009). The top 0.1% of most differentiated loci between sexes among landraces (i.e., SNPs with $F_{SC}$-values > = 0.4) were identified. Such a threshold appears rather conservative when heterozygosity is considered (as could be expected in the case of an XY sex-determining system, where a maximum $F_{SC}$ of 0.5 can be reached). The SNPs that reached an $F_{SC}$-value > = 0.4 were checked for their location along contigs and consistent association with either male or female samples (checked on the unpruned genetic dataset with 18,507 SNPs on 8657 contigs).

PCR-based sex-determining assay

As is clear from the results, one contig was identified as a top candidate associated with sex determination in *T. occidentalis* and the locus was thus further characterized.

PCR primers specific to that locus of interest were designed using primer3.ut.ee (version 4.1.0) and synthetized by microsynth.ch. Primers (each 10 mM)

were used for PCRs in 23.625 µl, with $5 \times$ GoTaq buffer (Promega), 10 mM dNTPs (Sigma Aldrich), ddH$_2$O and $0.125 \times$ GoTaq polymerase (Promega) on DNA templates from the 100 phenotyped samples. PCR amplification was carried out in a thermal cycler (Eppendorf Mastercycler) with the following conditions: 3 min of denaturation at 95 °C followed by 30 cycles of 30 s at 95 °C, 30 s at 54 °C and 30 s at 72 °C. To complete the sequence, the top candidate contig left ambiguous by paired-end $2 \times 150$ bp sequencing reads, PCR products from five females and one male were Sanger sequenced (service provided by Mycrosynth, Balgach, Switzerland).

To develop an assay based on cleaved amplified polymorphism sequences, the PCR product was digested by adding $0.1 \times$ of the restriction enzyme Mfe1-HF (NEB), $10 \times$ CutSmart Buffer (NEB) and ddH$_2$O resulting in a final volume of 25 µl for 2 h at 37 °C. Mfe1 cuts the sequence motif 5′CAATG 3′ that is mostly shared by female samples on the candidate locus, but fails to do so on the variant that is common to most male plants (i.e. 5′CCATG 3′). Digested PCR products were separated by electrophoresis on a 3% LE agarose gel in $1 \times$ TAE buffer for 2 h 15 min at 50 Volts with a 100 bp ladder, stained by GelRed and visualized under ultraviolet light.

A faster and cheaper assay based entirely on PCR was subsequently developed, relying on sex-specific primer pairs designed based on the different sets of SNPs reported along the candidate locus in the female and male plants investigated here. PCR conditions were similar to the assay described above, with $5 \times$ GoTaq Buffer (Promega), 10 mM dNTPs (Sigma Aldrich), 10 mM of each sex-specific primer (forward/reverse female, forward/reverse male), ddH$_2$O and $0.125 \times$ GoTaq polymerase (Promega) in a final volume of 25.625 µl, amplified by 3 min of denaturation at 95 °C followed by 30 cycles of 30 s at 95 °C, 30 s at 54 °C and 30 s at 72 °C. PCR products were visualized under ultraviolet light after electrophoresis on a 2% LE agarose gel in $1 \times$ TAE buffer for 2 h 15 min at 50 Volts with a 100 bp ladder and staining by GelRed.

A detailed protocol for both tests can be accessed in Supplementary Information 2.

Following visualization of PCR products, the two assays were ascertained by counting the number of times the phenotype (i.e., sex) predicted by the observed genotype matched the observed phenotype.

This ratio of observed sexes versus expected sexes was investigated for all 53 tested females and once for all 47 tested male *T. occidentalis* plant samples.

## Results

### Double digest restriction-site associated DNA sequencing data and genetic diversity

All 139 samples of *T. occidentalis* were successfully sequenced and genotyped following the ddRAD pipeline that resulted in 18,507 SNPs called on 8657 contigs. Although one technical replicate (A16r) was filtered out, the comparison of the four remaining ones estimates an average error rate of 4.4% (Table S5).

By using PCA (Fig. 1), variation based on those SNPs reveals three main genetic clusters among the five landraces of *T. occidentalis* investigated here. Samples belonging to the Anambra landrace cluster together and are differentiated from other samples along the first principal component of maximum genetic variation. The second principal component mainly differentiates Ekiti from the Cross River, Oyo and Lagos samples that thus appear more genetically coherent. One Anambra plant sample (A100) clusters together with Cross River, Oyo and Lagos (Fig. 1, supplementary material Figures S1 and S2) due to a likely mix-up of labels during either sowing of seeds or ddRAD library preparation.

### Population genetics

Model-based analyses of the pruned dataset with 9077 SNPs in low linkage disequilibrium using STRUCTURE indicated that genetic variation within *T. occidentalis* can be divided into three main clusters, confirming insights from PCA. As shown in Fig. 2, K=3 shows the most closely the genetic variation as supported by the delta K-value. Genetic structure among individual samples under K=3 revealed that Anambra and Ekiti each form a distinct group that can be genetically distinguished from others. Cross River, Oyo and Lagos appear to have a coherent genetic background and thus form a third genetically distinguishable population. All these findings provide strong evidence that the genetic data of *T. occidentalis* plant samples used in this study were can be divided into three main populations. As already

**Fig. 1** Principal component analysis of SNPs from ddRAD sequencing among 143 samples of *Telfairia occidentalis*. The different landraces are coloured differentially according to the panel. The first principal component axis (PC1) explains 8.3% and the second principal component axis (PC2) explains 5.6% of the overall genetic variation
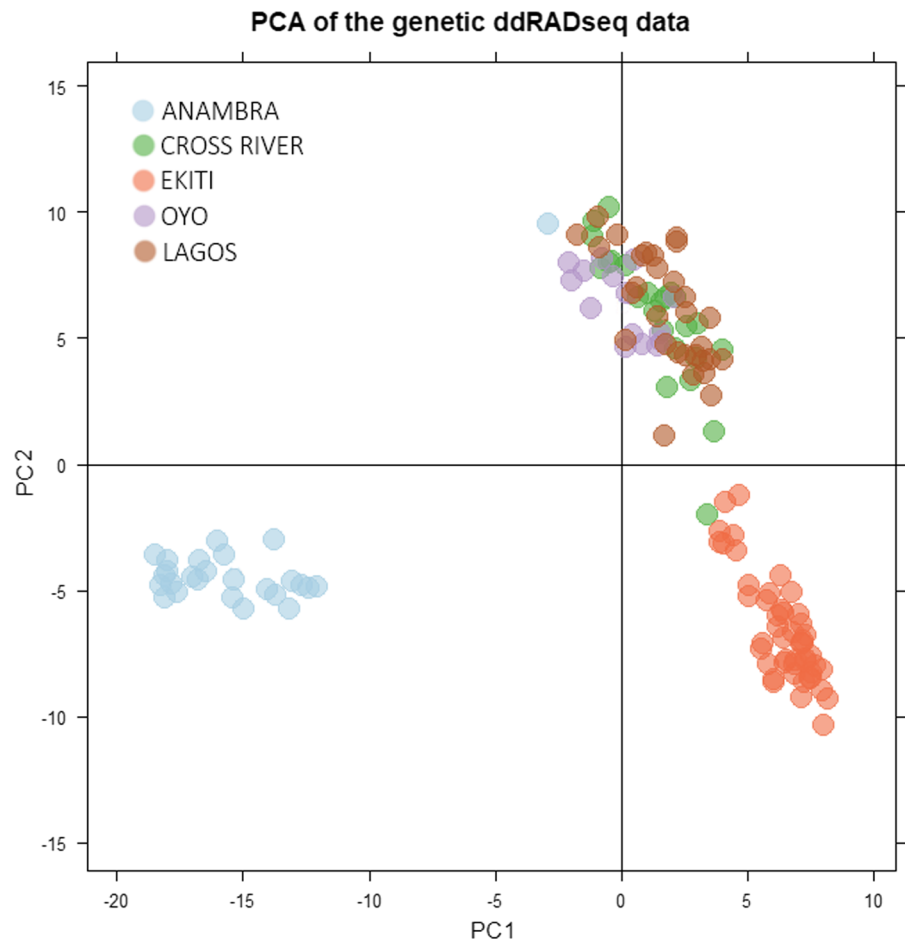


mentioned, plant sample A100 (Anambra) presents very similar genetic data to samples from Cross River, Oyo and Lagos and is thus unambiguously assigned to this cluster.

Sex-determining locus

The genetic data of 96 phenotyped plant samples (without Anambra landrace) was used to search for a sex determining locus. The programme ARLEQUIN produced $F_{SC}$-values for each of the 18,507 called SNPs (average $F_{SC}$-value ~ 0.05). Twenty SNPs with an $F_{SC}$-value > = 0.4 were identified of which ten were on contig number 4813 of the de novo reference catalogue. The other ten SNPs were distributed on nine different contigs. Contig 4813 is therefore the most promising sex-specific locus. All the SNPs of contig 4813 with their corresponding $F_{SC}$-values are listed in Table S6b. The $F_{SC}$-value

of ~ 0.46 (mean $F_{SC}$-value of the ten SNPs on contig 4813) is consistent with heterozygosity at the candidate locus in male plant samples. The mean expected heterozygosity is 0.08 for females and 0.53 for males (SNPs with an $F_{SC}$-value > = 0.4; Table S6a).

Figure 3 shows the most promising candidate contig 4813 as a sex-determining locus identified in this study. The alternative base for the SNP on contig 4813 at position 53 is a cytosine (Table 3). According to the genetic data, this SNP occurs in 94.3% of all plants of which the sex of the female phenotype is known in advance (50/53) and the alternative occurs in 93.6% of all plants of which the sex of the male phenotype is known in advance (44/47), which were used to generate the ddRAD libraries. The nucleotide sequence of the sex-associated candidate contig 4813 in Fig. 3 showed no significant matches on Web BLAST (last search 22.06.21).
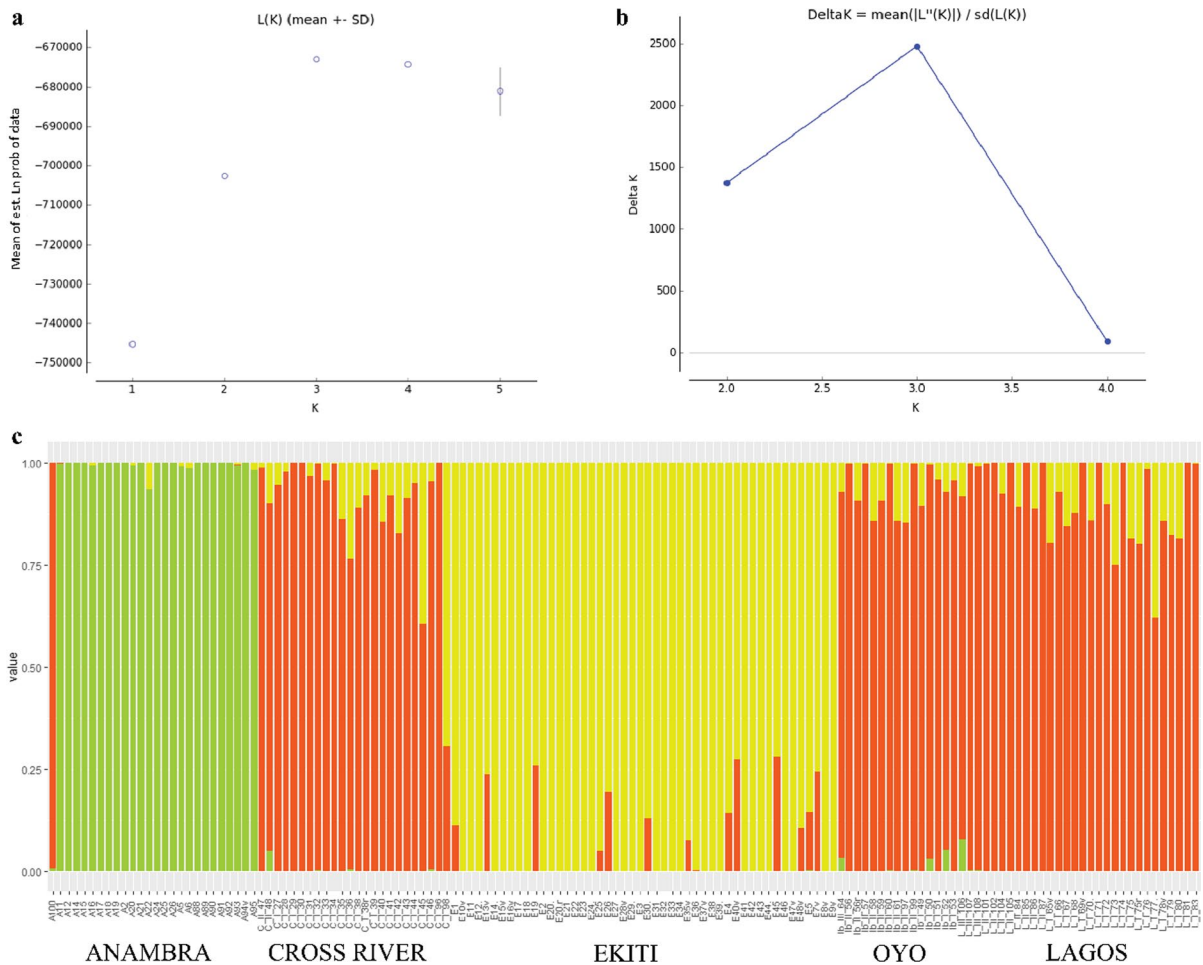
**Fig. 2** Genetic structure within *Telfairia occidentalis* based on STRUCTURE analyses of a pruned dataset of 9077 SNPs in low linkage disequilibrium. **a** Mean L(K) ± SD. Visualized with STRUCTURE HARVESTER. The value of the likelihood distribution L(K) levels off as soon as the real K-value is reached. **b** Visualized with STRUCTURE HARVESTER. Delta K showing a peak at an optimal K = 3. **c** Genetic data were obtained from the STRUCTURE HARVESTER processing and clustered with CLUMPP for visualisation in R Studio as a bar plot of the genetic data from every individual. All 143 individual names are listed along the X-axis. The Y-axis shows how much genetic variation is shared with the other individuals in 3 genetic clusters

Female-specific PCR primers designed for contig 4813 (Table 2) cover five SNPs that must be present in female plants (Fig. 3b). This is true in 49 out of 53 female plant samples (92.5%) which means that four plants have at least one variant that is most common in contig 4813 of male *T. occidentalis* plant samples. Genetic data shows that three of these four female plant samples carry all the male base variants. One female plant sample (C_I_29) carries just one SNP which most male plant samples share (cytosine instead of guanine) at position 41 (Table 3). In the case of the male specific PCR primers (Table 2), they cover four SNPs (Fig. 3c). The genetic data shows that 45 out of 47 male plants carry these four SNPs (95.7%). Two male plant individuals in this study have all four nucleotide variants that most female fluted pumpkin plant samples share.

### PCR-based sex-determination

The assay using cleaved amplified polymorphism sequences showed visually clear patterns of PCR products. Tested female plants all show a banding pattern consistent with homozygosity, with one

```
Contig 4813 5' - 3'

female      1  AGGGAAGAACCTTTTCTTGAGTAGATGAAGCATTTTGGGCGAAGATCTAACCAATTGTTT   60

male           AGGGAAGAACCTTTTCTTGAGTAGATGAAGCATTTTGGGCCAAGATCTAGCCCATTGTTT

female     61  TTCCAAATCTTTGTcGATTTATTTGTAATTGTTCCCACCATGCCAATGCTGCACTTTGgA  120

male           TTCCAAATCTTTGTaGATTTATTTGTAATTGGTCCCACCATGCCAATGCCGCACTTTGtA

female    121  GTTTGAAAGCTACCAGCTTTACCCTTTTGTGTTTGGGAGTGTTTGCGTAGTCGAAAAATT  180

male           GCTTGAAAGCTACCAGCTTTACCCTT-----------------------------------

female    181  TTCAACATTCTTGATCTAGTCAAGGAAGTTCTCTACATTCAATTTCCCATTGAAATGAGG  240

male           ------------------------------------------------------------

female    241  AAGATCGACTTTCATTTTGAAGTCAGATGTGTCTGCAAAGGTTGTtGTCTTCTTTCTAGT  300

male           -------------------------------------GTTGTcGTCTTCTTTCTAGT

female    301  GTTTCAAGAGGTTGTTGGTGCACATGATTTGGAAAGCAGCCTTCTCCACCTGAGGAATCA  360

male           GTTTCAAGAGGTAGTTGGTGTACATGATTTGGAAAGCAGCCTTCTCCACCTGAGGAATCA

female    361  GAATCTTGAATGTAGGCTCCTGGAAATTGGTACTCAAAAGGCCTTCTCATCATGTTTTCT  420

male           GAATCTTGAATGTAGGCTCCTGGAAATTGGTACTCAAAGGGCCGTCTCATCATGTTTTCT

female    421  TG  422

male           TG
```

**Fig. 3** Sex-associated candidate locus in *Telfairia occidentalis* (contig 4813) derived from ddRAD sequencing. Letters that are in bold and capital represent the ten SNPs with $F_{SC}$-values $>=0.4$ and small, bold letters are SNPs with $F_{SC}$-values $<0.4$. The female variant is 422 bp long, whereas the male variant is of unknown length and is represented as dashed lines. The boxed letters indicate the Mfe1 recognition site. The primers Tocc-CAPS-F and Tocc-CAPS-R (straight underlined sequences) do not include any sex-specific SNPs and result in a 328 bp PCR product to be digested using Mfe1, with males being expected to present three bands of 34 bp, 294 bp and 328 bp (heterozygosity), whereas females are expected to present one band of 34 bp and one of 294 bp. The primers Tocc-female-F and Tocc-female-R (wavy lines) include SNPs that are specific for most female samples of *T. occidentalis* and result in PCR products of 297 bp in length. The primers Tocc-male-F and Tocc-male-R (dotted lines) include male specific SNPs and result in PCR products of 120 bp in length

clear band at 294 bp and a lighter band at 34 bp (Fig. 4a). In contrast, tested male samples are all confirmed as heterozygotes at the targeted locus (contig 4813), with two strong bands (around 294 bp and 328 bp) and a lighter third band (around 34 bp). If the males were homozygous for contig 4813, only one band around 328 bp would be visible because Mfe1 does not detect the SNP variant at position 53 (Table 3) like in most male plants.

Sequence data predicts that the test could reveal the right sex 50 times out of 53 (94.3%) among *T. occidentalis* female plant samples and 44 times out of 47 (95.7%) among male plant samples. This indicates that one plant sample (L_I_80, Lagos) was considered

**Table 2** PCR primers used for sex-determining assay using Mfe1 as the restriction enzyme and two primers used for sex-determination via sex-specific PCRs in *Telfairia occidentalis*, with oligonucleotide sequences and corresponding names, ending with either -F (forward) or -R (reverse)

| Primer names | Primer sequences 5′–3′ |
| --- | --- |
| Tocc-CAPS-F[1] | GAGTAGATGAAGCATTTTGGGC |
| Tocc-CAPS-R[1] | GAGAAGGCTGCTTTCCAAATC |
| Tocc-female-F[2] | TTGGGCGAAGATCTAACCAATTG |
| Tocc-female-R[2] | CAAATCATGTGCACCAACAACC |
| Tocc-male-F[2] | GTGTTTCAAGAGGTAGTTGGTGTAC |
| Tocc-male-R[2] | AAAACATGATGAGACGGCCC |

**Table 3** Genetic variation (SNP) along with the top candidate contig 4813 as a sex-determining locus in *Telfairia occidentalis* and differentiation between sexes shown by $F_{SC}$ values $>= 0.4$ (marked with *)

| SNP number | SNP $F_{SC} >= 0.4$ | Position | Reference | Alternative |
| --- | --- | --- | --- | --- |
| 9434 | * | 41 | G | C |
| 9435 | * | 50 | A | G |
| 9436 | * | 53 | A | C |
| 9437 |   | 75 | C | A |
| 9438 | * | 92 | T | G |
| 9439 | * | 110 | T | C |
| 9440 |   | 119 | G | T |
| 9441 | * | 122 | T | C |
| 9442 |   | 162 | T | C |
| 9443 | * | 189 | T | A |
| 9444 | * | 197 | C | T |
| 9445 | * | 275 | A | G |
| 9446 | * | 280 | T | G |

Reference positions are mostly shared among females, whereas alternative ones are mostly shared by male samples



**Fig. 4** Banding patterns of molecular assays illustrating the sex determination of *Telfairia occidentalis* samples for 4 males and 4 females, along with a 100 bp ladder. **a** Cleaved amplified polymorphism sequences on 3% agarose gel electrophoresis **b** Direct PCRs on 2% agarose gel electrophoresis. This is an assemblage cropped from a larger gel picture used to genotype several samples and shown as supplementary information Fig. S4

male based on the flower phenotype, although its genotype predicts a female. Furthermore, the assay failed on two male samples (L_I_76 from Lagos and C_I_40 from Cross River) and the three female samples (Ib_I_52 and Ib_I_49 from Oyo, C_I_39 from Cross River). In the case of those two male plant individuals, their sequence data were the same as most female *T. occidentalis* plant individuals, but they were phenotyped as males. Vice versa, this applies to the three female plants where the assay failed.

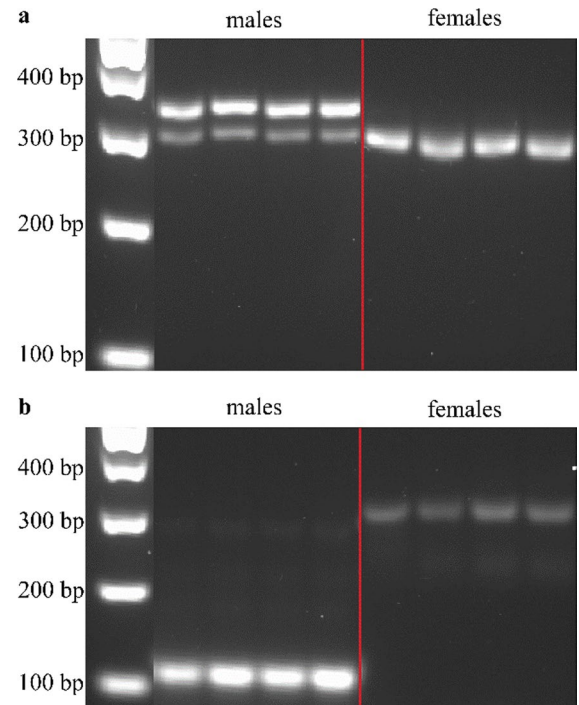The molecular assay using direct PCR via sex-specific primers successfully produced distinguishable banding patterns, with female samples showing one clear band at 297 bp, whereas males show one strong band at 120 bp and a lighter one at 297 bp, supporting a heterozygous genotype in males. Unlike the banding pattern generated via restriction enzymes, this assay via sex-specific primers generated a more complex pattern since primer dimers and other non-relevant bands are visible. Nevertheless, sex-specific patterns could be unambiguously identified.

Sex-determination via sex-specific primers successfully tested 50 out of 53 female samples. The female individual C_I_29 (Cross River) that had one SNP at position 41 as in most male samples (i.e., cytosine instead of guanine) was genotyped as a female here. The sex of 45 out of 47 male samples was correctly determined. These results suggest an assay certainty by sex-specific primers of 94.3% for females and 95.7% for male *T. occidentalis* plants.

Both sex-determining tests failed to adequately predict the phenotype based on the genotype on the same samples (i.e., L_I_76 (Lagos) and C_I_40 (Cross River) that should have shown a male genotype and Ib_I_52, Ib_I_49 (Oyo) and C_I_39 (Cross River) that should have presented a female genotype). The male individuals, on which the assays failed, have shown to carry all the SNPs that are usually expected in female *T. occidentalis* plant samples, while the converse is true for the female plant individuals Ib_I_52, Ib_I_49 (Oyo) and C_I_39 (Cross River). All the inconsistencies listed above can be seen in the supplementary information in Fig. S4.

Noticeably, both assays were performed accurately on the two female and two male samples from the Anambra landrace (Table 1) that were not used to initially identify contig 4813 as a candidate sex-determining locus.

## Discussion

### Genetic structure within *T. occidentalis* based on ddRAD sequencing

Reducing genome complexity for genotyping an underutilized crop using ddRAD sequencing based on rare (EcoR1) and frequent (Mse1) cutter restriction enzymes has here successfully captured the genetic variation of *T. occidentalis* that does not yet benefit from an already existing reference (Peterson et al. 2012; Palmer et al. 2019). Here, candidate sex-determining loci were detected through the partitioning of genetic variation between sexes while taking the variation due to genetic backgrounds of different landraces into account (i.e., $F_{SC}$-values in a hierarchical AMOVA framework), which advantageously relies on limited assumptions regarding the presence of sex chromosomes. Unlike approaches that assume the presence of sex-determining loci on X chromosomes but absence on Y chromosomes (or vice versa; as in Jeffries et al. 2018; Scharmann et al. 2019; Morgan et al. 2020), the analysis of $F_{SC}$ values also highlights sex-specific loci that suppress the development of the opposite sex, as is the case in dioecious plants (Charlesworth and Charlesworth 1978).

### Landraces of *Telfairia occidentalis*

Samples of *T. occidentalis* used to create ddRAD libraries were all collected from five different states of Nigeria and they were therefore expected to be genetically structured into five distinct groups. Multivariate PCA on SNPs (Fig. 1) as well as model-based inferences using STRUCTURE rather indicated three main genetic clusters (K = 3) within the species. The Anambra and Ekiti landrace samples were observed to be genetically homogenous and most distinct from the other genetic clusters, indicating little to no genetic exchanges for a longer period of time. In contrast, the remaining samples of Cross River, Oyo and Lagos landraces appear to share more genetic variation than other landraces and therefore group into a third genetic cluster. Although such genetic similarity may seem counter-intuitive at first glance, since the Cross River state is geographically distant from Lagos state and Oyo state, it should be kept in mind that *T. occidentalis* has probably been long indigenous in the regions of the Delta state, Imo state, Anambra state and up to Cross river state (7°–8° east and 5°–6° north). Plants cultivated by the Igbo people native to these regions and spread *Telfairia* to other parts of Nigeria (Akoroda 1990), so that the introduction of *T. occidentalis* in Lagos and Oyo could be too recent (Akoroda 1990) to yet be marked as patent genetic differentiation among samples from Cross River, Lagos and Oyo backgrounds. Given that Lagos and Oyo are geographically closer to each other, gene flow is more likely to have contributed to the genetic similarity of Oyo and Lagos (Wang et al. 2013). On the other hand, farmers in Lagos and Oyo may have collected seeds from Calabar (capital of Cross River state). Also, these varieties from Cross River, Lagos and Oyo states exhibited similar morphology in forms of growth habit and leaf (colour, shape, size), fruits and productivity and could be preferred by farmers (Oyenike Arike Adeyemo, personal observations). In contrast, diverse morphological traits were observed in the Anambra and Ekiti landraces. More comprehensive sampling of *T. occidentalis* landraces across its distribution range would provide valuable light on the domestication of fluted pumpkin, the significance of divergence among landraces and the potential for genetic resources in this underutilized crop.

## Sex phenotype versus genotype

The sex-determining assays developed here accurately predict the male vs female phenotype in at least 94% of all cases. In some cases, the genotype (and thus results of the sex-determining assay) predicted the opposite sex than revealed by the phenotype. Five samples (i.e., three females L_I_76, C_I_40, Ib_I_52 and two males Ib_I_49, C_I_39) revealed consistent incongruencies between genotype and phenotype. The underpinnings of such failure remain unclear and could either be due to mislabelling or be of biological origin.

Although there is only limited evidence that *T. occidentalis* presents sex chromosomes, a XY system was postulated based on cytogenetics, which seems here supported by the heterozygosity of the candidate contig 4813 associated with male determination that matches expectations of heterogametic XY males and homogametic XX females (Uguru and Onovo 2011). Therefore, it remains unclear to what extent heteromorphic sex chromosomes with inhibited recombination have yet evolved in a dioecious plant such as *T. occidentalis* (Ming et al. 2011; Renner 2014) and recombination within the candidate sex-determining locus should accordingly not be ruled out (Bergero and Charlesworth 2009; Charlesworth 2016). Similarly, it remains to be clarified to what extent sex determination in *T. occidentalis* is exclusively under genetic control. Genes that control the expression of ethylene, the sex-determining hormone in *Cucurbitaceae*, could likely be involved in *T. occidentalis*, as they are in melon (*Cucumis melo*), watermelon (*Citrullus lanatus*) or cucumber (*Cucumis sativus*; Boualem et al. 2015; Grumet and Taft 2011; Li et al. 2019; Zhang et al. 2020) that collectively support a conserved pathway of sex expression among *Cucurbitaceae*. A possible influence of environmental factors cannot be excluded (Golenberg and West 2013; Renner 2014) and may have had an impact on those five *T. occidentalis* samples showing inconsistency of genotype and phenotype. Finally, monoecious life forms of fluted pumpkin have already been reported (Akoroda et al. 1990) and, although not recently confirmed (Fayeun et al. 2016a, b), it should be considered that unlinked and undetected sex-determining loci could be involved that could have led to the unexpected phenotypes of the five inconsistent fluted pumpkin individuals from this study.

## Development of a sex determination assay

The two different assays described here to genetically establish the sex of *T. occidentalis* samples yielded similar results (94.3% for female and 95.7% for male plants) and appear equally time-consuming. Despite the fact that the restriction enzyme approach is slightly more laborious, it also leads to clearer banding patterns than the approach based on sex-specific primers (Fig. S4). Such molecular sex-determining assays are of great importance for farmers owning medium to large *Telfairia* plantations as to optimize the proportion and cultivation space devoted to female plants. Although sex-determining assays were designed based on Cross River, Ekiti, Oyo and Lagos landraces, assays were performed accurately in the landrace Anambra, suggesting transferability among landraces. Further phenotyping and genotyping analyses of other *T. occidentalis* landraces as well as the sister species, *Telfairia pedata* (Hook. F.) that is also a dioecious plant growing across eastern and south-eastern regions of Africa (Okoli 2007; Okoli and Mgbeogu 1983), would be valuable to foster the potential of underutilized *Telfairia* crops and the use of universal sex-determining tools for their plantlets.

**Declarations**

**Competing interest** The authors have not disclosed any competing interests.

# References

Adeyemo OA, Tijani HA (2018) Fluted pumpkin [*Telfairia occidentalis* (Hook F.)]: genetic diversity and landrace identification using phenotypic traits and RAPD markers. IFE J Sci 20:391–401. https://doi.org/10.4314/ijs.v20i2.19

Akoroda MO (1990) Ethnobotany of *Telfairia occidentalis* (Cucurbitaceae) among Igbos of Nigeria. Econ Bot 44:29–39

Akoroda MO, Ogbechie-Odiaka NI, Adebayo ML, Ugwo OE, Fuwa B (1990) Flowering, pollination and fruiting in fluted pumpkin (*Telfairia occidentalis*). Sci Hortic 43:197–206. https://doi.org/10.1016/0304-4238(90)90091-R

Badifu GIO (1993) Food potentials of some unconventional oilseeds grown in Nigeria: a brief review. Plant Foods Hum Nutr 43:211–224. https://doi.org/10.1007/BF01886222

Bergero R, Charlesworth D (2009) The evolution of restricted recombination in sex chromosomes. Trends Ecol Evol 24:94–102. https://doi.org/10.1016/j.tree.2008.09.010

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bonin A, Bellemain E, Bronken Eidesen P et al (2004) How to track and assess genotyping errors in population genetics studies. Mol Ecol 13(11):3261–3273. https://doi.org/10.1111/j.1365-294X.2004.02346.x

Boualem A, Troadec C, Camps C et al (2015) A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. Science 350(6261):688–691. https://doi.org/10.1126/science.aac8370

Catchen JM, Amores A, Hohenlohe P et al (2011) Stacks: building and genotyping loci de novo from short-read sequences. G3 Genes Genomes Genet 1:171–182. https://doi.org/10.1534/g3.111.000240

Catchen JM, Hohelohe PA, Bassham S et al (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124–3140. https://doi.org/10.1111/mec.12354

Chang CC, Chow CC, Tellier LC et al (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4:7. https://doi.org/10.1186/s13742-015-0047-8

Charlesworth D (2016) Plant sex chromosomes. Annu Rev Plant Biol 67:397–420. https://doi.org/10.1146/annurev-arplant-043015-111911

Charlesworth B, Charlesworth D (1978) A model for the evolution of dioecy and gynodioecy. Am Nat 112:975–997

Chivenge P, Mabhaudhi T, Modi A, Mafongoya P (2015) The potential role of neglected and underutilised crop species as future crops under water scarce conditions in sub-Saharan Africa. Int J Environ Res Public Health 12:5685–5711. https://doi.org/10.3390/ijerph120605685

Chong Z, Ruan J, Wu CI (2012) Rainbow: an integrated tool for efficient clustering and assembling RAD-seq reads. Bioinformatics 28:2732–2737. https://doi.org/10.1093/bioinformatics/bts482

Chukwurah NF, Uguru MI (2010) Juvenile morphological markers for maleness in Fluted Pumpkin (*Telfairia occidentalis* Hook F.). Agro-Sci J Trop Agric Food Environ Ext 9:90–96

Danecek P, Auton A, Abecasis G et al (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dawson IK, Hedley PE, Guarino L, Jaenicke H (2009) Does biotechnology have a role in the promotion of underutilised crops? Food Policy 34:319–328. https://doi.org/10.1016/j.foodpol.2009.02.003

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361. https://doi.org/10.1007/s12686-011-9548-7

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier L, Lischer HEL (2015) Manual Arlequin Ver 3.5.2. computational and molecular population genetics. University Bern. http://cmpg.unibe.ch/software/arlequin35/man/Arlequin35.pdf. Accessed 4 Dec 2020

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10(3):564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fahey JW (1998) Underexploited African grain crops: a nutritional resource. Nutr Rev 56:282–285. https://doi.org/10.1111/j.1753-4887.1998.tb01767.x

Fayeun LS, Odiyi AC (2015) Variation and heritability of marketable leaf yield components in fluted pumpkin. Sci Agric 11:8–14

Fayeun LS, Odiyi AC, Adebisi AM, Hammed LA, Ojo DK (2016a) Floral biology of fluted pumpkin (*Telfairia occidentalis* Hook. F.). Sci Biol 8:482–488. https://doi.org/10.15835/nsb849895

Fayeun LS, Ojo DK, Odiyi AC et al (2016b) Identification of facultative apomixis in fluted pumpkin (*Telfairia occidentalis* Hook F.) through emasculation method. Am J Exp Agric 10:1–10. https://doi.org/10.9734/AJEA/2016/15261

Feron R, Peron Q, Wen M et al (2021) RADSex: a computational workflow to study sex determination using restriction site-associated DNA sequencing data. Mol Ecol Resour 21:1715–1731. https://doi.org/10.1111/1755-0998.13360

Gamble T, Coryell J, Esaz T et al (2015) Restriction site-associated DNA sequencing (RAD-Seq) reveals an extraordinary number of transitions among Gecko sex-determining systems. Mol Biol Evol 32:1296–1309. https://doi.org/10.1093/molbev/msv023

Golenberg EM, West NW (2013) Hormonal interactions and gene regulation can link monoecy and environmental plasticity to the evolution of dioecy in plants. Am J Bot 100:1022–1037. https://doi.org/10.3732/ajb.1200544

Grumet R, Taft J (2011) Sex expression in cucurbits. In: Wang YH, Behera TK, Chittaranjan K, Yi-Hong W (eds) Genetics, genomics and breeding of cucurbits. Science Publishers, Boca Raton, pp 353–375

Grünig S, Fischer M, Parisod C (2021) Recent hybrid speciation at the origin of the narrow endemic *Pulmonaria helvetica*. Ann Bot 127:21–31. https://doi.org/10.1093/aob/mcaa145

Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat Rev Genet 10:639–650. https://doi.org/10.1038/nrg2611

Jaenicke H, Hoeschle-Zeledon I (2009) A strategic framework for underutilized plant species research and development, with special reference to Asia and the Pacific, and to Sub-Saharan Africa. Acta Hortic 818:333–342. https://doi.org/10.17660/ActaHortic.2009.818.49

Jaenicke H, Dawson IK, Guarino L, Hermann M (2009) Impacts of underutilized plant species promotion on biodiversity. Acta Hortic 806:621–628. https://doi.org/10.17660/ActaHortic.2009.806.77

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jeffries DL, Lavanchy G, Sermier R et al (2018) A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. Nat Commun 9:4088. https://doi.org/10.1038/s41467-018-06517-2

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li W, Godzik A (2006) Cd-Hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Li H, Hendsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li D, Sheng Y, Niu H, Li Z (2019) Gene interactions regulating sex determination in Cucurbits. Front Plant Sci 10:1–12. https://doi.org/10.3389/fpls.2019.01231

Mabhaudhi T, Chimonyo VGP, Hlahla S et al (2019) Prospects of orphan crops in climate change. Planta 250:695–708. https://doi.org/10.1007/s00425-019-03129-y

Martin AR, Cadotte MW, Isaac ME et al (2019) Regional and global shifts in crop diversity through the Anthropocene. PLoS ONE 14:1–18. https://doi.org/10.1371/journal.pone.0209788

Matsumura H, Miyagi N, Taniai N et al (2014) Mapping of the gynoecy in bitter gourd (*Momordica charantia*) using RAD-Seq analysis. PLoS ONE 9:1–10. https://doi.org/10.1371/journal.pone.0087138

McKenna A, Hanna M, Banks E et al (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110

Ming R, Bendahmane A, Renner SS (2011) Sex chromosomes in land plants. Ann Rev Plant Biol 62:485–514. https://doi.org/10.1146/annurev-arplant-042110-103914

Morgan EJ, Kaiser-Bunbury CN, Edwards PJ et al (2020) Identification of sex-linked markers in the sexually cryptic Coco de Mer: Are males and females produced in equal proportions? AoB PLANTS 12:1–9. https://doi.org/10.1093/aobpla/plz079

Ndukwu BC, Obute GC, Wary-Toby IL (2005) Tracking sexual dimorphism in *Telfairia occidentalis* Hooker f. (Cucurbitaceae) with morphological and molecular markers. Afr J Biotechnol 4:1245–1249

Nwonuala A, Obiefuna J (2015) Yield and yield components of fluted pumpkin (*Telfairia occidentalis* Hook) landrace. Int J Agric Innov Res 4:421–425

Okoli BE, Mgbeogu CM (1983) Fluted Pumpkin, *Telfairia occidentalis*: West African vegetable crop. Econ Bot 37:145–149

Okoli BE (2007) *Telfairia pedata* (Sm. Ex Sims) Hook. https://www.prota4u.org/database/protav8.asp?&g=pe&p=Telfairia+pedata+(Sm.+ex+Sims)+Hook (2 June 2021)

Padulosi S, Hodgkin T, Williams JT, Haq N (2002) Underutilized crops: trends, challenges and opportunities in the 21st century. In: Engels JMM, Ramanatha RV, Jackson AHD, Brown MT (ed) Managing plant genetic diversity. Proceedings of an international conference, Kuala Lumpur, Malaysia, 12–16 June 2000. CABI, Wallingford, pp 323–338. http://www.cabi.org/cabebooks/ebook/20023003597

Padulosi S, Hoeschle-Zeldon I (2004) Underutilized plant species: what are they? Leisa Mag 20:5–6

Palmer DH, Rogers TF, Dean R, Wright AW (2019) How to identify sex chromosomes and their turnover. Mol Ecol 28:4709–4724. https://doi.org/10.1111/mec.15245

Peterson BK, Weber JN, Kay EH et al (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7:1–11. https://doi.org/10.1371/journal.pone.0037135

Porras-Hurtado L, Ruiz Y, Santos C et al (2013) An overview of STRUCTURE: applications, parameter settings, and supporting software. Front Genet 4:1–13. https://doi.org/10.3389/fgene.2013.00098

Pritchard JK, Stephens M, Donelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Puritz JB, Hollenbeck CM, Gold JR (2014) *dDocent*: a RAD-seq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2:e431. https://doi.org/10.7717/peerj.431

Raheem D (2011) The need for agro-allied industries to promote food security by value addition to indigenous African food crops. Outlook Agric 40:343–349. https://doi.org/10.5367/oa.2011.0063

Renner SS (2014) The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. Am J Bot 101:1588–1596. https://doi.org/10.3732/ajb.1400196

Scharmann M, Grafe TU, Metali F, Widmer A (2019) Sex is determined by XY chromosomes across the radiation of dioecious *Nepenthes* pitcher plants. Evol Lett 3:586–597. https://doi.org/10.1002/evl3.142

Tadele Z (2019) Orphan crops: their importance and the urgency of improvement. Planta 250:677–694. https://doi.org/10.1007/s00425-019-03210-6

Tarasov A, Vilella AJ, Cuppen E et al (2015) Sambamba: fast processing of NGS alignment formats. Bioinformatics 31:2032–2034. https://doi.org/10.1093/bioinformatics/btv098

Uguru MI, Onovo JC (2011) Gender in fluted pumpkin (*Telfairia occidentalis* Hook. F.). Int J Plant Breed 5:64–66

Wang IJ, Richard EG, Jonathan BL (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. Ecol Lett 16:175–182. https://doi.org/10.1111/ele.12025

World Bank (2007) Agriculture for development 2008 World Development Report (c) The International Bank for Reconstruction and Development/The World Bank. https://openknowledge.worldbank.org/handle/10986/5990 (24 May 2021).

Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina paired-end read merger. Bioinformatics 30:614–620. https://doi.org/10.1093/bioinformatics/btt593

Zhang J, Guo S, Zhao H et al (2020) A unique chromosome translocation disrupting *ClWIP1* leads to gynoecy in watermelon. Plant J 101:265–277. https://doi.org/10.1111/tpj.14537