

## Research Article

**Cite this article:** Steadman EM, Whitehouse KJ, Norton SL (2023). Molecular characterization of *Brassica* genebank germplasm confirms taxonomic identity and reveals low levels and source of taxonomic errors. *Plant Genetic Resources: Characterization and Utilization* **20**, 194–202. <https://doi.org/10.1017/S1479262123000035>

Received: 1 February 2022  
Revised: 20 October 2022  
Accepted: 2 February 2023  
First published online: 28 February 2023

### Key words:

*Brassica*; *ex-situ* seed banking; misclassification; molecular markers; plant genetic resources; taxonomy

### Author for correspondence:

Erica M. Steadman,  
E-mail: [erica.steadman@agriculture.vic.gov.au](mailto:erica.steadman@agriculture.vic.gov.au)

# Molecular characterization of *Brassica* genebank germplasm confirms taxonomic identity and reveals low levels and source of taxonomic errors

Erica M. Steadman , Katherine J. Whitehouse and Sally L. Norton

Australian Grains Genebank, Grains Innovation Precinct, Agriculture Victoria, Private Bag 260, Horsham, Victoria, 3401, Australia

## Abstract

Crop germplasm conserved in genebanks, are a fundamental resource of genetic diversity for crop improvement activities, underpinning future food security and sustainable agricultural practices. However, taxonomic errors in genebank germplasm (due to misclassification, contamination and poor data collation) restrict the effective use of this material for correct purpose. Earlier studies investigating species genetic diversity using genebank germplasm, have shown varying levels of taxonomic error within the *Brassica* species. In response to this reported taxonomic error of global collections, together with the availability of a multiplex PCR (MPCR) marker, targeting the specific chromosomes (A, B and C) of the six *Brassica* species in U's triangle, this study was undertaken to confirm the taxonomic identity of accessions within the Australian Grains Genebank's (AGG) long-term *Brassica* collection. A total of 5161 accessions were analysed with MPCR for taxonomic identification, of which, 4842 (93.8%) were confirmed to be consistent (correct) with their labelled taxonomy, while the remaining 319 (6.2%) were identified as taxonomically inconsistent (in-error). Through the evaluation of earlier regeneration and original seed of the error accessions with MPCR, we determined that 80.9% of the taxonomic errors were traced back to the original seed, while 19.1% of errors were the result of genebank seed regeneration handling practices. Results from this study directly enhance information of the AGG *Brassica* collection and shape directions for distribution, acquisition and regeneration practices within the AGG and potentially other global genebanks, which will facilitate in a more effective use of these valuable genetic resources by researchers and breeders.

## Introduction

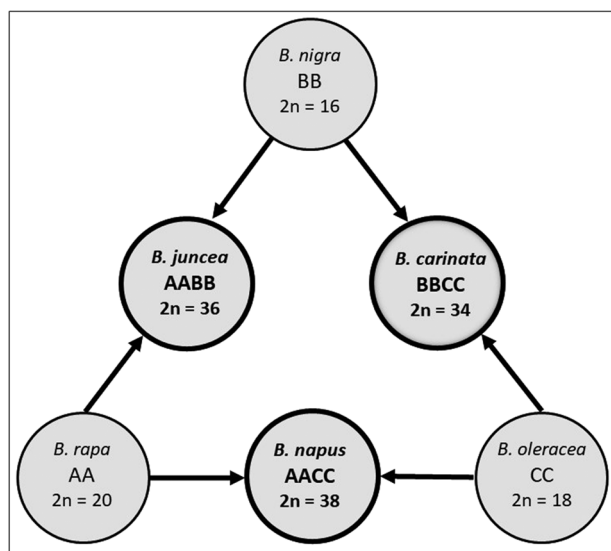
Domestication through plant breeding and selection primarily for yield over numerous years has resulted in a bottleneck or homogeneity of traits and a lack of diversity in many of today's agricultural crops (Gizlice *et al.*, 1994; Abbo *et al.*, 2003; Hyten *et al.*, 2006; Aguilar *et al.*, 2015; Olodo *et al.*, 2020). This, together with the rising impacts of climate change and the growing global population, has placed an increasing focus and importance on plant genetic resources (or crop germplasm) to improve current crop varieties (Dempewolf *et al.*, 2014; Pilling *et al.*, 2020; Crop Trust, 2021; Palanivel and Shah, 2021). Genebanks have been established by the worlds' scientific community with the mandate to acquire, conserve and share plant genetic resources between the world's researchers and plant breeders. The diversity in global genebank collections are comprised of traditional and modern varieties, advanced breeding lines, genetic stocks and crop wild relatives. These germplasm collections have been established over time through local and international donations or deposits from breeders, collectors, research institutes and other global collections, and form the genetic basis for crop improvement. Thus, *ex-situ* conservation of crop plant germplasm in genebanks is crucial to supporting our future; through alleviating global food insecurity and promoting sustainable agriculture.

To ensure an efficient continuum between the conservation and the utilization of plant genetic resources, it is critical that the users of crop germplasm receive material that is true to type. Crop breeders and researchers rely on the genetic integrity of material from genebanks, as research and breeding efforts using these genetic resources are expensive and time-consuming practices. Errors in genebank germplasm, due to misclassification, misidentification, mislabelling, contamination or poor data collation, restrict the effective use of this germplasm for correct purpose. Historically, genebank germplasm passport data has sometimes been incomplete or missing, with limited information recorded, and formal taxonomic identification of collected or donated material often not undertaken. Proper taxonomic classification and correct identification of germplasm is vitally important, but often remains a challenge for genebank managers due to the lack of specialist taxonomic expertise, and



reliance on morphological characters that are prone to subjective errors and are less accurate where there is trait ambiguity (Czajkowska *et al.*, 2019), and variability due to genotype-by-environment interactions (Ndjiondjop *et al.*, 2018). Genebank managers are heavily reliant on the accuracy of the passport information from donors, which can exacerbate inaccuracies unless appropriate validation is undertaken. Errors within genebanks can also occur, resulting in genetic contamination and/or mislabelling of accessions at various stages of genebank activities including registration, sampling and especially during seed regeneration or multiplication practices, or when handling large numbers of plant material (Girma *et al.*, 2012). Genetic contamination during seed regeneration is most likely to occur through inadequate control of cross pollination by interfertile species. Errors have been reported in several species including *Triticum* (Czajkowska *et al.*, 2019), *Oryza* (Ndjiondjop *et al.*, 2018), *Arabidopsis thaliana* (Anastasio *et al.*, 2011), *Theobroma cacao* L. (Motilal *et al.*, 2011), *Solanum* (Ellis *et al.*, 2018), *Dioscorea* (Girma *et al.*, 2012), *Lens* (Wong *et al.*, 2015) and *Brassica* (Mason *et al.*, 2015) and are severely detrimental to the effective conservation, dissemination and use of these crop germplasms.

The *Brassica* genus, belonging to the *Brassicaceae* family (formerly *Cruciferae*), contain the most genetically diverse collection of agriculturally important plant species, including oilseed, leaf vegetables, condiment and root vegetable crops such as rapeseed, broccoli, cabbage, bok choy, mustard and turnip. *Brassica* crops are major contributors to horticultural and agricultural economies worldwide, with 70.5 million metric tonnes of rapeseed alone produced globally in 2019, with a gross production value of US\$26.7 billion (FAO, 2021). The six main *Brassica* species have an interesting genetic relationship, which is described as U's triangle (Fig. 1), whereby three diploid species *Brassica rapa* (AA,  $2n = 20$ ), *Brassica nigra* (BB,  $2n = 16$ ) and *Brassica oleracea* (CC,  $2n = 18$ ) were determined to be the progenitors of three allotetraploid species, *Brassica juncea* (AABB,  $2n = 36$ ), *Brassica*



**Fig. 1.** The Triangle of U, showing the genetic relationships of the six species in the *Brassica* genus. The three diploid progenitor species *B. rapa*, *B. nigra* and *B. oleracea* with genomes AA, BB and CC respectively, through natural hybridization and chromosome doubling gave rise to three allotetraploid species *B. napus* (AACC), *B. juncea* (AABB) and *B. carinata* (BBCC). Diploid chromosome number ( $2n$ ) is shown. Adapted from U (1935).

*napus* (AACC,  $2n = 38$ ) and *Brassica carinata* (BBCC,  $2n = 34$ ), which arose through natural hybridization and chromosome doubling (U, 1935). Traditionally, classification of *Brassica* species has been based on morphological traits, such as floral architecture, the angle between the pedicel and rachis, leaf morphology and the presence of leaf hairs (Pradhan *et al.*, 2011). However, due to similarities in morphological characters and the wide ranges of morphotypes, it is difficult to delineate between some of the *Brassica* species (Annisa *et al.*, 2013; OECD, 2016; Shen *et al.*, 2018). For example, within the *B. rapa* species, several morphotypes exist including oilseed, root vegetable, leafy vegetable and fodder crop types (Guo *et al.*, 2014). Furthermore, many species in the broader *Brassicaceae*, including the wild radishes (*Raphanus*), woad (*Isatis*), dog mustard (*Erucastrum*), hoary mustard (*Hirschfeldia incana*) and white mustard (*Sinapis*), can also hybridize with important crop species within *Brassica* (Lefol *et al.*, 1996, 1997; Warwick *et al.*, 2003; FitzJohn *et al.*, 2007; Kumar *et al.*, 2013; Zhan *et al.*, 2017). Therefore, due to the economic importance of *Brassica*, together with the ability for hybrid introgression from wild relatives and the existing rich genetic diversity in the non-cultivated forms, *Brassica* is a key component of genebanks worldwide.

Recently, molecular studies investigating the centre of origin, genetic diversity and species identification of *Brassica* crop germplasm sourced from global collections, have identified taxonomic errors within the *Brassica* germplasm evaluated (Pradhan *et al.*, 2011; Annisa *et al.*, 2013; Chen *et al.*, 2013; Guo *et al.*, 2014; Mason *et al.*, 2015). For example, when characterizing *B. nigra* germplasm with simple sequence repeat (SSR) markers it was discovered that *B. nigra* accessions formed distinct groups associated with country/region of origin (Pradhan *et al.*, 2011). As a result of this study, 26.7% of the germplasm assessed were found to be mislabelled or misidentified as *B. nigra*. Similarly, Chen *et al.* (2013) who analysed the genetic diversity in a globally diverse set of *B. juncea* using SSRs found 3.3% of accessions to be mislabelled or misclassified. Furthermore, results from research investigating diversity (Annisa *et al.*, 2013) and centre of origin (Guo *et al.*, 2014) of *B. rapa* accessions using SSR markers identified 6.4 and 7.5% of the respective germplasm sets were misclassified. In addition, Mason *et al.* (2015) evaluated *Brassica* lines using an Illumina Infinium *Brassica* 60 K SNP (Single Nucleotide Polymorphism) array to assess for genetic diversity and species confirmation, and identified 18.0% of samples as misclassified. In response to these findings, together with the recent publication and public availability of *B. rapa*, *B. oleracea* and *B. napus* complete genome maps (Wang *et al.*, 2011; Chalhoub *et al.*, 2014; Liu *et al.*, 2014), a chromosome specific multiplex polymerase chain reaction (MPCR) marker assay for the rapid taxonomic identification of *Brassica* species in U's triangle was developed and validated by Koh *et al.* (2017), with the sensitivity to detect the *Brassica* A, B and C genomes of individual seeds within pooled seed samples due to its low detection limit (0.1 ng DNA).

Considering these past studies highlighting the prevalence of taxonomic error within global *Brassica* germplasm, and the recent development of a reliable, species specific MPCR marker assay, this study was undertaken to confirm taxonomic identity of *Brassica* germplasm at the Australian Grains Genebank (AGG) and detect any taxonomic error within the collection. The objectives of this study were to (a) evaluate all *Brassica* accessions in the AGG long-term collection for taxonomy using the MPCR marker assay; (b) evaluate lines with taxonomic error to determine the source of this error; and (c) use results from this study to develop

specific acquisition, curation, regeneration and distribution protocols for the AGG.

## Materials and methods

### Selection of germplasm

Seed of 5161 *Brassica* accessions (comprising of wild, landraces/traditional cultivars, breeding lines, Plant Breeders Rights (PBR) reference samples and advanced cultivars), representing six species (*B. rapa* [2303], *B. napus* [1188], *B. juncea* [1167], *B. oleracea* [262], *B. nigra* [133] and *B. carinata* [108]) were sampled from the AGG long-term collection (online Supplementary Table S1). Where possible, seed from the most recent regeneration cycle (Current Active) was used, however, in instances where low seed number occurred, the previous regeneration (if present) or the original seed lots were sampled (Original and Archive) (online Supplementary Table S1). From each accession, a 2 g seed sample was visually inspected, and characteristics recorded including seed size and colour (using the 'IBPGR Descriptors of *Brassica* and *Raphanus*' IBPGR, 1990), and any observed seed mixtures were also recorded (online Supplementary Table S1). As controls, seeds from six accessions (representing each of the species) whose taxonomic identity had been previously confirmed (Pradhan *et al.*, 2011; Mason *et al.*, 2015; Koh *et al.*, 2017) were sampled from the AGG, except for *B. oleracea* where the commercial variety 'Red Cabbage' was used to align with the previous study by Koh *et al.* (2017) (Table 1).

Following initial analysis, those accessions identified as being inconsistent (in-error), with their labelled taxonomic classification, were further evaluated using seed from an earlier regeneration cycle and from the original seed where possible (based on seed availability). These additional seed samples were sourced from the AGG long-term storage, sampled and visually inspected, as described above. In total, 5335 samples of seed (5161 for initial analyses and 174 for further evaluation), were analysed and are presented in this study (online Supplementary Table S1).

Further to the analyses, those accessions identified as producing a mixed species genotypic result (online Supplementary Table S1), were re-tested using single seed analysis, whereby a minimum of eight individual seeds per accession, visually inspected and representatively sampled to match the overall seed phenotype, were evaluated to determine the species constituents of the identified mixed seed sample.

### DNA extraction

Genomic DNA was extracted from 25 mg of seeds (eight seeds) per accession for a bulk DNA extraction or extracted individually

where single seed analysis was undertaken. Seeds were placed in 96 well Collection Microtube racks (Qiagen) – one accession per well for bulk extractions or one seed per well for single seed extractions – with seven wells per rack left blank for the controls (positive and negative). Seeds were rehydrated in 200 µl of sterile distilled water per well and left at room temperature overnight, after which the water was removed. DNA was extracted from *Brassica* seeds using the DNeasy 96 Plant Kit (Qiagen) on the QIAcube HT Robotic system (Qiagen), according to manufacturer's instructions. Seeds were ground in 400 µl of AP1 Lysis Buffer, RNase A and Reagent DX (Qiagen), with two tungsten beads (Qiagen) per well, on a Mixer Mill 300 (Retsch, Germany) at 30 Hz for 1.5 min. The rack was then rotated before grinding again. The DNA was eluted in 100 µl of AE Elution Buffer (Qiagen) and stored at –20 °C until required. DNA concentration was measured using a NanoDrop 2000 spectrophotometer (Thermo Scientific), according to manufacturer's instructions, and concentrations adjusted to 10–25 ng/µl for bulk DNA extractions and 5 ng/µl for single seed DNA extractions, with sterile distilled water.

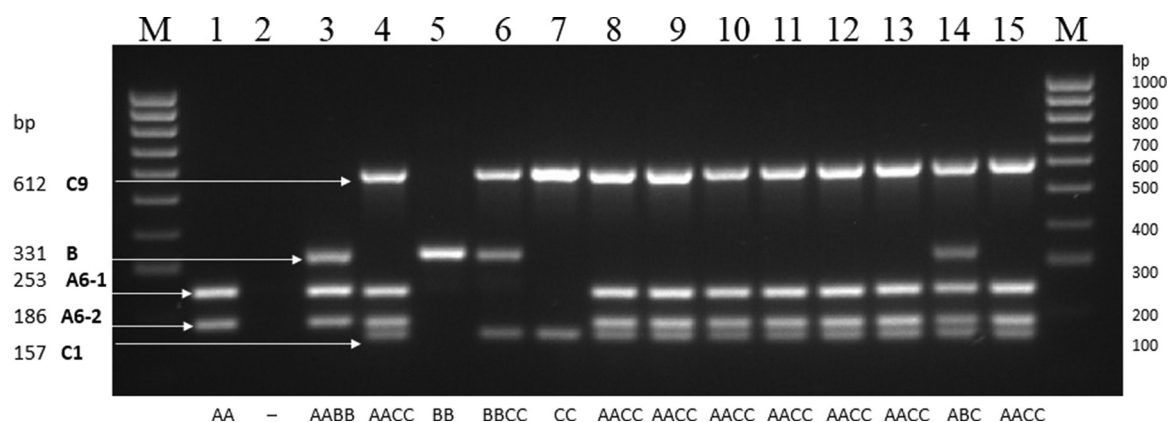
### Genotyping and analysis

Following extraction, the DNA was amplified using the MPCPCR assay developed by Koh *et al.* (2017). MPCPCR assays were performed on a Mastercycler thermocycler (Eppendorf) in 12.5 µl volumes, containing 2.5 µl of 5 × MyTaq Red Reaction Buffer (Bioline), 0.75 units of MyTaq HS DNA Polymerase (Bioline), 10–25 ng of template DNA, and primer mix (forward and reverse each) C1 0.36 µM, C9 0.32 µM, A6–1 0.16 µM, A6–2 0.24 µM and B 0.2 µM. Primer sequences and PCR cycling conditions are described in Koh *et al.* (2017). Negative (water; no DNA) and positive DNA controls (Table 1) were included with every MPCPCR assay run to ensure no DNA contamination and to confirm successful MPCPCR amplification (Fig. 2).

After amplification, the MPCPCR products were separated by gel electrophoresis on a 2.0% w/v agarose gel in 1 × TAE (40 mM Tris Acetate, 1 mM EDTA) buffer using a BioRad gel apparatus at 5 V/cm for 2 h, along with a molecular marker weight ladder (HyperLadder 100 bp, Bioline). The products were visualized under UV light, after staining with SYBR Safe DNA Gel Stain (Invitrogen, Thermo Fischer Scientific) and gel images were captured on a GelDoc XR+ imaging system (BioRad). Banding patterns of each accession were visually analysed, and results determined by comparing them to the bands of known fragment sizes of the *Brassica* controls and the known fragment sizes of the molecular weight ladder (HyperLadder 100 bp, Bioline) (Fig. 2).

**Table 1.** Six samples of known *Brassica* species identity used as controls in this study

Species	AGG Number	Primary accession name	Improvement status	Country of origin	Reference
<i>Brassica rapa</i>	AGG90210	Nagoaka C	Unknown	Unknown	Koh <i>et al.</i> (2017)
<i>Brassica napus</i>	AGG90553	Darmoor	Advanced Cultivar	Poland	Mason <i>et al.</i> (2015)
<i>Brassica juncea</i>	AGG93387	Domo	Advanced Cultivar	Canada	Mason <i>et al.</i> (2015)
<i>Brassica oleracea</i>		Red Cabbage	Commercial Variety	Australia	Koh <i>et al.</i> (2017)
<i>Brassica nigra</i>	AGG91057	BRA 185/78	Unknown	Greece	Pradhan <i>et al.</i> (2011)
<i>Brassica carinata</i>	AGG94024	BRA 926/78	Unknown	Ethiopia	Pradhan <i>et al.</i> (2011)



**Fig. 2.** Gel image of MPCR amplification of six positive controls of known species identity, a negative (water) control, and eight AGG long-term *Brassica* collection accessions, together with the molecular weight ladder (HyperLadder 100 bp, Bioline). M: Molecular weight ladder (HyperLadder100 bp, Bioline), (1) *B. rapa* control (AA) AGG90210, (2) negative control (H<sub>2</sub>O, no DNA), (3) *B. juncea* control (AABB) AGG93387, (4) *B. napus* control (AACC) AGG90553, (5) *B. nigra* control (BB) AGG91057, (6) *B. carinata* control (BBCC) AGG94024, (7) *B. oleracea* control (CC) Red Cabbage, (8) AGG90046, (9) AGG90059, (10) AGG90060, (11) AGG90061, (12) AGG90062, (13) AGG90063, (14) AGG90069, (15) AGG90070. Lanes 8–15 are *B. napus* labelled accessions which all gave a *B. napus* (AACC) result with MPCR, except Lane 14 (AGG90069) which gave an ABC result, this denotes a mixed bulk sample. MPCR results were resolved on a 2.0% agarose gel. Fragment sizes of the MPCR products (C9, B, A6-1, A6-2 and C1) are labelled on the left in base pairs (bp) and molecular weight ladder band sizes are labelled on right in bp.

## Results

### *Brassica* germplasm taxonomic ID screening and error source determination

The MPCR results, seed characteristics and a subset of the passport data for each of the accessions are presented in online Supplementary Table S1. Of the 5161 *Brassica* accessions screened, 4842 were identified as taxonomically correct when compared to their labelled taxonomy (93.8%), while the remaining 319 were identified as inconsistent (incorrect or mixed seed). These incorrect accessions accounted for 6.2% of the total germplasm analysed and included 129 *B. rapa* accessions, 56 *B. napus*, 77 *B. juncea*, 15 *B. oleracea*, 36 *B. nigra* and six *B. carinata* accessions (Table 2).

Of the 319 incorrect accessions, the error source for 283 was resolved. Of these 283 accessions, 229 (80.9%) were identified as the result of donor source errors (either misclassified or contaminated by the donor source), and 54 (19.1%) were due to inhouse genebank errors (mislabelling or poor sample handling or regeneration procedures) (Table 2). For the remaining 36 incorrect accessions, which included nine *B. rapa*, six *B. napus*, seven *B. juncea*, four *B. oleracea* and 10 *B. nigra*, the error source was

not able to be resolved due to the unavailability of seed from the original sample (Table 2).

### *Brassica rapa*

*Brassica rapa* labelled accessions contributed to 2303 (44.6%) of the total germplasm samples analysed by MPCR in this study. Of these, 2174 (94.4%) were identified as taxonomically correct, with the remaining 129 (5.6%) accessions identified as taxonomically incorrect (Table 2). Further MPCR evaluation of these incorrect accessions, using seed from a previous regeneration cycle and from the original seed, were carried out, with nine unable to be evaluated due to seed unavailability. For 120 accessions with an incorrect labelled taxonomy, 108 (90.0%) had an error source traced back to the original donor of the seed, with 12 (10.0%) linked back to genebank practices occurring in 2005, 2009, 2011 and 2014 (Tables 2 and 3). Of the 108 accessions showing original seed source error; 34 were taxonomically reidentified by our study as *B. napus*, 25 as *B. juncea*, six as *B. oleracea*, three as *B. carinata* and two accessions (AGG92013 and AGG97702) failed to produce banding patterns. The remaining 38 accessions were identified as contaminated through single

**Table 2.** Summary of the MPCR genotyping results for 5161 accessions from the Australian Grains Genebank long-term *Brassica* collection, showing the number and percentage of accessions tested and error accessions identified, and source of taxonomic error in this collection by Labelled Taxonomy group

Labelled taxonomy	No. of accessions tested	No. of error accessions (% in parentheses)	No. error accessions without original seed	No. error accessions with original seed	Extra samples tested – error source	No. accessions original seed source error (%)	No. accessions AGG practices source error (%)
<i>B. rapa</i>	2303	129 (5.6)	9	120	54	108 (90.0)	12 (10.0)
<i>B. napus</i>	1188	56 (4.7)	6	50	71	26 (52.0)	24 (48.0)
<i>B. juncea</i>	1167	77 (6.6)	7	70	19	65 (92.9)	5 (7.1)
<i>B. oleracea</i>	262	15 (5.7)	4	11	0	11 (100.0)	0 (0.0)
<i>B. nigra</i>	133	36 (27.1)	10	26	28	13 (50.0)	13 (50.0)
<i>B. carinata</i>	108	6 (5.6)	0	6	2	6 (100.0)	0 (0.0)
<b>Total</b>	<b>5161</b>	<b>319 (6.2)</b>	<b>36</b>	<b>283</b>	<b>174</b>	<b>229 (80.9)</b>	<b>54 (19.1)</b>

**Table 3.** Number of accessions with taxonomic error due to genebank regeneration handling practices, in the year the error occurred, as determined by MPCR evaluation, for each Labelled Taxonomy group

Labelled taxonomy	Number of error accessions	2004	2005	2006	2007	2008	2009	2011	2014	2015
<i>B. rapa</i>	12		6				2	2	2	
<i>B. napus</i>	24			16			5	1	1	1
<i>B. juncea</i>	5	1			2	2				
<i>B. oleracea</i>	0									
<i>B. nigra</i>	13						8	4	1	
<i>B. carinata</i>	0									
Total	54	1	6	16	2	2	15	7	4	1

seed MPCR analysis and were mixtures of *B. rapa* with either *B. napus* or *B. juncea* (35), a mixture of *B. rapa*, *B. napus* and *B. oleracea* (1) a mixture of *B. napus* and *B. oleracea* (1) and a mixture of *B. carinata* and *B. oleracea* (1) seeds (online Supplementary Table S1).

#### *Brassica napus*

*Brassica napus* labelled accessions contributed to 1188 (23.0%) of the total germplasm analysed in this study. Of these, 1132 (95.3%) were identified as taxonomically correct, with the remaining 56 (4.7%) accessions identified as taxonomically incorrect (Table 2). Following further MPCR evaluation, the source of error was successfully identified in 50 of these incorrect accessions, with 26 (52.0%) traced back to the original donor of the seed. The remaining 24 (48.0%) were linked back to genebank practices mostly occurring in 2006 (16) and 2009 (5), with only one occurrence each in 2011, 2014 and 2015 (Tables 2 and 3). Of the 26 incorrect *B. napus* accessions showing original seed source error, 16 were taxonomically reidentified by our study as *B. rapa* accessions, three *B. juncea*, two *B. oleracea* and two *B. nigra*. There was one accession (AGG92102) that failed to produce bands, and two accessions were shown to be mixtures of *B. rapa* and *B. napus* by single seed MPCR analysis (online Supplementary Table S1).

#### *Brassica juncea*

*Brassica juncea* labelled accessions contributed to 1167 (22.6%) of the total germplasm evaluated by MPCR in this study. Of these, 1090 (93.4%) were identified as taxonomically correct, with the remaining 77 (6.6%) accessions identified as being taxonomically incorrect (Table 2). Following further MPCR evaluation, the source error was successfully identified in 70 of these incorrect accessions, with 65 (92.9%) traced back to the original donor of the seed and five (7.1%) linked to genebank practices in 2004, 2007 and 2008 (Tables 2 and 3). Of the 65 *B. juncea* accessions showing original seed source error, 13 were taxonomically reidentified by our study as *B. carinata*, six as *B. napus*, 10 as *B. nigra* and 31 identified as *B. rapa* (online Supplementary Table S1). The remaining accessions were identified as being seed mixtures by single seed MPCR analysis, with three comprising a mixture of *B. juncea* and *B. napus*, one a mixture of *B. juncea* and *B. carinata* and the other a mixture of *B. carinata* and *B. rapa* seeds (online Supplementary Table S1).

#### *Brassica oleracea*

*Brassica oleracea* labelled accessions contributed to 262 (5.1%) of the total germplasm evaluated by MPCR in this study. Of these, 247 (94.3%) were identified as being taxonomically correct, with the remaining 15 (5.7%) accessions identified as being taxonomically incorrect (Table 2). Following further MPCR evaluation, the source error was successfully identified in 11 of these incorrect accessions, with all (100.0%) errors traced back to the original donor of the seed (Tables 2 and 3). Of the 11 incorrect *B. oleracea* labelled accessions, three were taxonomically reidentified by our study as *B. rapa*, two as *B. napus*, two as *B. carinata*, three as *B. juncea* and one a mixture between *B. oleracea* and *B. juncea* (online Supplementary Table S1).

#### *Brassica nigra*

*Brassica nigra* labelled accessions contributed to 133 (2.6%) of the total *Brassica* germplasm evaluated by MPCR in this study. Of these, 97 (72.9%) were identified as taxonomically correct, with the remaining 36 (27.1%) accessions identified as being taxonomically incorrect. Following further MPCR evaluation, the source error was successfully identified in 26 of these incorrect accessions, with 13 (50.0%) traced back to the original donor seed and 13 (50.0%) due to genebank handling practices (Table 2); with a high proportion of errors occurring in the 2009 regeneration year (8), and the remaining errors in 2011 (4) and 2014 (1), as shown in Table 3. Of the 13 accessions with original donor source error, 10 were taxonomically reidentified by our study to be *B. juncea*, one as *B. rapa*, one as *B. napus* and one accession, using single seed MPCR analysis, was found to be a mixture of *B. nigra* and *B. rapa* seed (online Supplementary Table S1).

#### *Brassica carinata*

*Brassica carinata* labelled accessions contributed to 108 (2.1%) of the total germplasm evaluated by MPCR in this study. Of these, 102 (94.4%) were identified as taxonomically correct, with the remaining six (5.6%) accessions identified as being taxonomically incorrect. Following further MPCR evaluation, the source of error for these incorrect accessions was successfully identified with all errors (100.0%) originating from the original donor seed source (Tables 2 and 3). The inconsistent accessions were taxonomically reidentified by our study as *B. juncea* (4), *B. rapa* (1) and *B. nigra* (1) (online Supplementary Table S1).

## Discussion

Genebanks conserve diverse genetic material that are valuable sources of traits for plant breeders and researchers, however it is imperative that genebanks conserve and distribute germplasm with the correct taxonomic identity to ensure effective and efficient development of new, more resilient plant varieties for current and future growing conditions. In our study, the multiplex PCR marker assay (MPCR) (Koh *et al.*, 2017) targeting specific chromosomes (A, B and C) of the *Brassica* complex was used to confirm species taxonomic identification of *Brassica* accessions sampled from the AGG, a widely used germplasm collection located in Horsham, Victoria, Australia. Of the 5161 *Brassica* accessions evaluated in this study, 4842 (93.8%) were confirmed to have the correct taxonomic identity, while the remaining 319 accessions were identified as taxonomically incorrect, with the overall level of taxonomic error of the AGG *Brassica* collection being 6.2% (Table 2). Of the six *Brassica* species (in U's triangle) included in our study, we identified taxonomic error rates for *B. napus* of 4.7%, *B. oleracea* of 5.7%, *B. carinata* of 5.6%, *B. rapa* of 5.6%, *B. juncea* of 6.6% and *B. nigra* of 27.1% (Table 2). Previous studies using SNP and SSR markers to investigate species confirmation, genetic diversity and centre of origin also reported, to varying degrees, the presence of taxonomic error within *Brassica* germplasm, selected from global genebanks (Pradhan *et al.*, 2011; Annisa *et al.*, 2013; Chen *et al.*, 2013; Guo *et al.*, 2014; Mason *et al.*, 2015). Although these studies found varying levels of taxonomic error, they sampled fewer accessions and less species than ours. Thus, our study, to date, is the most comprehensive assessment in resolving the taxonomic identity, and detecting taxonomic errors and their source, of accessions from a single genebank, across all six *Brassica* species in the U's triangle.

To identify the source, or origin, of these taxonomic errors, seed from previous regenerations as well as the original seed sample (where available) were assessed with MPCR and the results were compared against those from the initial screening. Of the 319 accessions that were identified in the initial screening as taxonomically incorrect, 54 were due to errors resulting from genebank regeneration handling practices (Table 2). For these accessions identified as an internal genebank error, historical regeneration records (N Sawyer, pers. comm.) were used to determine where errors may have occurred. For the 13 *B. nigra* accessions (Table 2), eight (61.0%) occurred in 2009 (Table 3), with all but one identified to be *B. napus* or mixed for *B. napus* and *B. nigra* (online Supplementary Table S1). Similarly, for the 24 *B. napus* accessions (Table 2), 16 (66.7%) occurred in 2006 (Table 3). Historical regeneration records indicated that these *B. napus* accessions were either substituted with *B. carinata* seed or mixed/contaminated with *B. carinata* seeds (online Supplementary Table S1). Errors, or mistakes, can occur during any sample handling stage of the regeneration process without good quality control measures in place, for example, at subsampling, sowing, harvest, threshing and seed cleaning. Mistakes that fail to be identified are at risk of being passed on in subsequent regeneration events. Evidence of this can be seen in accessions: AGG90506, AGG90568, AGG90655, AGG90661, AGG91073, AGG91105 and AGG92441 (online Supplementary Table S1). For example, for *B. napus* accession AGG90506, MPCR analysis identified that an error was made in the 2006 regeneration cycle, with resultant seed from this cycle consisting of *B. napus* mixed or contaminated with *B. carinata* seed (online Supplementary

Table S1). This contaminated 2006 regeneration seed lot was subsequently used as the parent seed for a regeneration cycle in 2010, perpetuating this error. Therefore, genebanks that fail to prevent or detect contamination errors will have continual issues with incorrect germplasm, with potentially significant downstream ramifications for research and breeding programmes following distribution.

To the best of our knowledge, there have been no other studies providing evidence of the source of misclassification in *Brassica* germplasm, however similar work has been performed with *Oryza* species in a study by Ndjiondjop *et al.* (2018). Here, DArTseq-based SNP technology was used for the development of species and sub-species (ecotype) specific diagnostic quality control markers for genotyping a large set of rice (*Oryza*) germplasm, representing three African rice species (*O. glaberrima*, *O. barthii* and *O. longistaminata*) and Asian rice (*O. sativa* spp. *indica* and *japonica*), from the AfricaRice genebank. They identified 3.1% of the accessions (97/3134) as being taxonomically incorrect. From these 97 misclassified accessions, 37 were from regenerated seed lots (which included *O. glaberrima* [32], *O. longistaminata* [1] and *O. sativa* [4]). When they compared the genotypic data of the original collections with the regenerated seed lots, they identified that most of these errors (35/37) were caused during seed regeneration, multiplication, seed processing and handling. The remaining two samples were due to the original seed lots also being misclassified, probably because of species misclassification during germplasm collection. The error attributable to genebank regeneration and seed handling practices accounted for 36.1% (35/97) of the overall errors found in the *Oryza* germplasm study by Ndjiondjop *et al.* (2018). While in a separate study (Ellis *et al.*, 2018), errors in genebank germplasm due to genebank handling, regeneration and manipulation of germplasm were detected when 250 potato (*Solanum*) landrace accessions representing paired samples consisting of original mother plants and their in vitro counterparts, from the International Potato Center (CIP) genebank were analysed. These accessions were SNP fingerprinted using the Infinium 12 K V2 Potato Array and pairwise comparisons of the SNP fingerprint profiles between the in vitro plantlets and the mother plants revealed errors and genetic mismatches in 11 of the 250 (4.4%) accessions.

Despite having standard operating procedures for the management of plant genetic resources in *ex-situ* genebanks, human error can occur at numerous stages during genebank activities and processes, e.g., seed handling, labelling, storage and harvesting. While these errors can occur, it is imperative that genebanks have stringent quality management systems (QMS) and quality assurance (QA) practices in place and that are regularly reviewed to minimize mistakes from occurring. The genebank regeneration and sample handling errors identified in this study are important, however the majority (90.7%) occurred prior to 2014 (Table 3), during regeneration events in a precursor genebank that was amalgamated into the AGG. While the predecessors that sourced, regenerated and maintained this material should be commended for their foresight and diligence, QMS and QA practices historically employed to ensure sample integrity, were not as advanced as they are today. Since 2014, and the amalgamation into the AGG, all sampling handling practices have been reviewed and more stringent QMS and QA practices implemented as preventative measures to minimize similar errors occurring into the future.

While 19.1% of taxonomic errors in our study were due to genebank regeneration practices, 80.9% of errors were traced back to the original seed that was deposited into the AGG

*Brassica* long-term collection (Table 2). These original donor seed taxonomic errors may be due to misidentification of the seed from the depositor and/or donor, lack of information in exchange between depositors and the collection curators, mislabelling or mishandling on entering the donor genebank or poor information tracking systems. An additional source of error could be due to mislabelling or mishandling during registration into the recipient genebank, however the likelihood of this type of error is very low. As these errors occurred outside the AGG it can be difficult to determine the true cause of a particular error. It is likely that the difficulty in correctly classifying the species due to the reliance on subjective morphological traits (Pradhan *et al.*, 2011) and the existence of several morphotypes both within and between species (Guo *et al.*, 2014), poor pollination control of outcrossing *Brassica* species during regeneration, the lack of information and history of some of the material, together with the age of the material in the collection (acquired from 1987 onwards) contributed to this level of taxonomic error (online Supplementary Table S1).

All the errors for *B. carinata* and *B. oleracea* labelled accessions, and 92.9 and 90.0% respectively for *B. juncea* and *B. rapa* were due to original seed donor source (Table 2). Within the misclassified *B. rapa* labelled accessions (108) that were traced back to original seed source, 36 (33.3%) of these were identified, by single seed analysis, as being a mixture of *B. rapa* and another *Brassica* species. This high level of species mixture within *B. rapa* as compared to the other species in this study, may be due, in part, to the self-incompatible nature of the *B. rapa* species. Self-incompatibility, or the inability of self-fertilization, is seen as a frequent problem in *B. rapa* accessions (Takayama and Isogai, 2005) and is one of the major causes of heterozygosity and mixed phenotypes in *B. rapa* breeding programs through cross pollination by other species when pollination control is not applied. Guo *et al.* (2014) identified 46.8% of *B. rapa* accessions as self-incompatible, with almost no seed set when protected from cross-pollination, whereas self-compatible types had a few pods under self-pollinating conditions. This high level of species mixture within *B. rapa* shown in our study is likely a reflection of the contaminated seed lots which arose from the lack of stringent protocols during seed regeneration or multiplication practices from the donor source or genebank. The high level of self-incompatibility of the *B. rapa* species and the high level of mixtures identified in this study show that great care needs to be taken and strict pollination control procedures, particularly in open-pollinated cultivars, need to be followed during seed regeneration and multiplication.

Some interesting observations from our study were the reclassification of accessions by source genebanks from the correct to an incorrect species, and naming contradictions (Labelled Taxonomy and Primary Accession Name) in the passport data of some error accessions. For example, the AGG passport data showed two *B. rapa* labelled accessions (AGG93685 and AGG90310), identified by MPCR as *B. juncea*, were originally correctly classified as *B. juncea* but had been reclassified as *B. rapa* by the USDA in 1992 (online Supplementary Table S1). Similarly, two accessions (AGG90304 and AGG93687) labelled as *B. nigra* that were identified to be *B. juncea* by MPCR in this study, had originally been classified correctly as *B. juncea* and were reclassified to *B. nigra* by the USDA in 1989 (online Supplementary Table S1). In addition, accession AGG92971 was labelled as *B. napus* but found to be *B. nigra* by MPCR, had the Primary Accession Name of 'Brassica Nigra' in the genebank passport data (online Supplementary Table S1), while one accession (AGG93214) labelled as *B. juncea*,

had 'Carinata' as the Primary Accession Name and was tested to be *B. carinata* by MPCR (online Supplementary Table S1). Furthermore, two accessions (AGG92013 and AGG97702) labelled as *B. rapa* and one accession (AGG92102) labelled as *B. napus*, produced no amplification products when tested with MPCR, with the accession AGG97702 having 'SV Raphanus' as the Primary Accession Name (online Supplementary Table S1). The MPCR marker assay (Koh *et al.*, 2017) targets specific regions on the chromosomes (A, B and C) of the six *Brassica* species in U's triangle, with relatives of the *Brassica* species (such as *Raphanus* spp.) not possessing these A, B or C chromosomes, and will hence lack amplification product by MPCR. The absence of amplification product for accessions, AGG92013, AGG97702 and AGG92102, suggests that these accessions are not one of the six main *Brassica* species in U's triangle (Koh *et al.*, 2017). On visual inspection of the seeds from these accessions, they appeared to have a vastly different seed morphology as compared to seeds from the *Brassicaceae* species evaluated in this study (online Supplementary Fig. S1). Further genetic and morphological analysis would need to be performed to confirm species identity for AGG92013, AGG92102 and AGG97702. These observations highlight that there are difficulties and confusion with the taxonomic identification of *Brassica* species (OECD, 2016) and that caution needs to be taken in the taxonomic identification, reclassifying and naming of *Brassica* accessions. These observations also emphasize the importance of database curation by genebanks to ensure correct and complete passport data is sourced from the donor upon material acquisition (Girma *et al.*, 2012), together with the adoption of appropriate information systems for the recording, tracing and collating of accession data. Accurate and effective data management is vital for genebank curators to manage their collections (Weise *et al.*, 2020), and that poor data curation, at least in the past (Anastasio *et al.*, 2011; Ellis *et al.*, 2018; Hay *et al.*, 2021), may have contributed to the taxonomic errors arising in global genebank collections.

In this study, the MPCR assay proved to be a reliable, robust and cost-effective method to determine the correct taxonomic identity of *Brassica* accessions in the AGG long-term collection. MPCR was also able to resolve species constituents in sample mixtures, without the need for expensive and sophisticated equipment or software. However, a limitation of this assay is that, unlike the more advanced SNP based methods, such as genotyping-by-sequencing (GBS) (Khedikar *et al.*, 2020; Abdel-Haleem *et al.*, 2022) and array (Xiao *et al.*, 2021) technologies, MPCR does not provide sequence information for assessment of genetic diversity. Subsequently, the MPCR assay cannot differentiate individuals arising from intraspecific hybridization (Mason *et al.*, 2015) and lacks the capability to resolve taxonomy to a subspecies level, for example, it cannot distinguish *B. rapa* spp. *silvestris* individuals from *B. rapa* spp. *oleifera* individuals. Whilst the assessment of genetic diversity and resolution of subspecies members of the AGG *Brassica* germplasm was outside the scope of this study, this is planned for future investigation using a SNP based assay to further evaluate the genetic integrity of the *Brassica* collection.

The AGG has used the misclassification and contamination data identified in our study to develop more stringent curation and distribution protocols to best manage the misclassified and contaminated accessions. All accessions that have been identified as erroneous are currently being archived and removed from distribution. The AGG is in the process of developing a schedule to reimport and rebuild the *Brassica* accessions where the donor is

the source of the error, and through this process will advise the source genebanks of the identified inaccuracies related to previously supplied germplasm seed lots. Germplasm that has been contaminated through internal genebank practices will be purified where possible, with contaminated seed lots either discarded or archived if previously distributed. QMS and QA practices, together with seed handling and regeneration protocols, have been developed by the AGG since 2014, with significant improvements around pollination control for open-pollinated *Brassica* species. These new protocols aim to eliminate the occurrence of contamination, with taxonomic validation now undertaken as a routine practice to immediately identify it if it does occur within the genebank program.

In concluding, the taxonomic identity of *Brassica* germplasm within the AGG collection has been quantified (93.8% as taxonomically correct), and taxonomic errors (6.2%) assessed to identify the cause of error as either a donor source error, or through previous poor genebank practices. The conventional taxonomic classification of the *Brassica* species is very highly correlated to the MPCR analytical results (93.8%), however, the 6.2% of incorrect accessions serves as a reminder to genebanks that despite adhering to standards and guidelines for long-term conservation and maintenance of genetic resources, mistakes can and do occur. Ensuring germplasm maintains its true genetic identity is critical for genebanks, therefore the continual improvement in QMS and QA practices is essential to minimize the occurrence of mistakes, and to identify and rectify any mistakes when they do occur. For the AGG, the identification of such errors has enabled more stringent QMS and QA curation and regeneration protocols, especially around pollination control, to be developed to best manage the *Brassica* collection, and to limit impact on downstream research and breeding programs. All new *Brassica* accessions received into the genebank, and following internal regeneration events, are now screened using the MPCR protocol as a rapid taxonomy screen prior to being accessioned into the collection, until a cost-effective *Brassica* species specific SNP based technology is available. All historical recipients of misclassified or contaminated accessions/seed lots will be notified, as will the original donors of the germplasm. Future directions to interrogate the *Brassica* collection further will use technologies that enable taxonomic verification, genetic diversity, genetic integrity and trait identification to be determined. This will further increase the genetic information of this collection and facilitate the utilization and uptake of genebank germplasm and accelerate the development of new varieties suited to future growing conditions.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S1479262123000035>.

**Acknowledgements.** This work was funded by Agriculture Victoria and the Grains Research and Development Corporation, grant number 9176106 to SLN, through the AGG project – Phase 3, 2017–2022. EMS would like to thank Virginia McQueen, Joshua Koh and Narelle Nancarrow for advice and assistance during this research.

## References

- Abbo S, Berger J and Turner NC (2003) Evolution of cultivated chickpea: four bottlenecks limit diversity and constrain adaptation. *Functional Plant Biology* **30**, 1081–1087.
- Abdel-Haleem H, Luo Z and Szczepanek A (2022) Genetic diversity and population structure of the USDA collection of *Brassica juncea* L. *Industrial Crop and Products* **187**, 115379.
- Aguilar J, Gramig GG, Hendrickson JR, Archer DW, Forcella F and Liebigh MA (2015) Crop species diversity changes in the United States: 1978–2012. *PLoS ONE* **10**, e0136580.
- Anastasio AE, Platt A, Horton M, Grotewold E, Scholl R, Borevitz JO, Nordborg M and Bergelson J (2011) Source verification of mis-identified *Arabidopsis thaliana* accessions. *Plant Journal* **67**, 554–566.
- Annisa, Chen S and Cowling WA (2013) Global genetic diversity in oilseed *Brassica rapa*. *Crop and Pasture Science* **64**, 993–1007.
- Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Corréa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier MC, Fan G, Renault V, Bayer PE, Golicz AA, Manoli S, Lee TH, Thi VV, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom CH, Wang X, Canaguier A, Chauveau A, Bernard A, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Wang X, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury JM, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou Y, Hua W, Sharpe AG, Paterson AH, Guan C and Winkler P (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **345**, 950–953.
- Chen S, Wan Z, Nelson MN, Chauhan JS, Redden R, Burton WA, Lin P, Salisbury PA, Fu T and Cowling WA (2013) Evidence from genome-wide simple sequence repeat markers for a polyphyletic origin and secondary centers of genetic diversity of *Brassica juncea* in China and India. *Journal of Heredity* **104**, 416–427.
- Crop Trust (2021) Available at <https://www.croptrust.org/> Website visited on 20 January 2021.
- Czajkowska BI, Oliveira HR and Brown TA (2019) A discriminatory test for the wheat B and G genomes reveals misclassified accessions of *Triticum timopheevii* and *Triticum turgidum*. *PLoS ONE* **14**, e0215175.
- Dempewolf H, Eastwood RJ, Guarino L, Khoury CK, Müller JV and Toll J (2014) Adapting agriculture to climate change: a global initiative to collect, conserve and use crop wild relatives. *Agroecology and Sustainable Food Systems* **38**, 369–377.
- Ellis D, Oswald C, Coombs J, Soto J, Gomez R, Douches D, Panta A, Silvestre R and Anglin NI (2018) Genetic identity in genebanks: application of the SolCAP 12 K SNP array in fingerprinting and diversity analysis in the global in trust potato collection. *Genome* **61**, 523–537.
- FAO (2021) ([fao.org/faostat/en/#data](http://fao.org/faostat/en/#data)). Website visited on 26 April 2021.
- FitzJohn RG, Armstrong TT, Newstrom-Lloyd LE, Wilton AD and Cochrane M (2007) Hybridisation within *Brassica* and allied genera: evaluation of potential for transgene escape. *Euphytica* **158**, 209–230.
- Girma G, Korie S, Dumet D and Franco J (2012) Improvement of accession distinctiveness as an added value to the global worth of the Yam (*Dioscorea* spp) Genebank. *International Journal of Conservation Science* **3**, 199–206.
- Gizlice Z, Carter TE Jr and Burton JW (1994) Genetic base for North American public soybean cultivars released between 1947 and 1988. *Crop Science* **34**, 1143–1151.
- Guo Y, Chen S, Li Z and Cowling WA (2014) Center of origin and centers of diversity in an ancient crop, *Brassica rapa* (Turnip rape). *Journal of Heredity* **105**, 555–565.
- Hay FR, Whitehouse KJ, Ellis RH, Sackville Hamilton NR, Lusty C, Ndjondjop MN, Tia D, Wenzl P, Santos LG, Yazbek M, Azevedo VCR, Peerzada OH, Abberton M, Oyatomi O, de Guzman F, Capilit G, Muchugi A and Kinyanjui Z (2021) CGIAR genebank viability data reveal inconsistencies in seed collection management. *Global Food Security* **30**, 100557.
- Hyten DL, Song Q, Zhu Y, Choi I-K, Nelson RL, Costa JM, Specht JE, Shoemaker RC and Cregan PB (2006) Impacts of genetic bottlenecks on soybean genome diversity. *Proceeding of the National Academy of Sciences of the United States of America* **103**, 16666–16671.
- IBPGR (1990) *Descriptors for Brassica and Raphanus*. Rome: International Board for Plant Genetic Resources, IBPGR.
- Khedikar Y, Clarke WE, Chen L, Higgins EE, Kagale S, Koh CS, Bennett R and Parkin IAP (2020) Narrow genetic base shapes population structure



- and linkage disequilibrium in an industrial oilseed crop. *Brassica carinata* A. Braun. *Scientific Reports* **10**, 12629.
- Koh JCO, Barbulescu DM, Norton S, Redden S, Salisbury PA, Kaur S, Cogan N and Slater AT (2017) A multiplex PCR for rapid identification of *Brassica* species in the triangle of U. *Plant Methods* **13**, 49.
- Kumar A, Singh BK, Singh VV and Chauhan JS (2013) Cytomorphological and molecular evidence of synthesis of interspecific hybrids between *Brassica rapa* and *B. fruticulosa* through sexual hybridization. *Australian Journal Crop Science* **7**, 849–854.
- Lefol E, Fleury A and Darmency H (1996) Gene dispersal from transgenic crops. II. Hybridization between oilseed rape and the wild hoary mustard. *Sexual Plant Reproduction* **9**, 189–196.
- Lefol E, Séguin-Swartz G and Downey RK (1997) Sexual hybridisation in crosses of cultivated *Brassica* species with the crucifers *Erucastrum gallicum* and *Raphanus raphanistrum* – potential for gene introgression. *Euphytica* **95**, 127–139.
- Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP, Zhao M, Ma J, Yu J, Huang S, Wang X, Wang J, Lu K, Fang Z, Bancroft I, Yang TJ, Hu Q, Yue Z, Li H, Yang L, Wu J, Zhou Q, Wang W, King GJ, Pires JC, Lu C, Wu Z, Sampath P, Wang Z, Guo H, Pan S, Min J, Zhang D, Jin D, Li W, Belcram H, Tu J, Guan M, Qi C, Du D, Li J, Jiang L, Batley J, Sharpe AG, Park BS, Ruperao P, Cheng F, Waminal NE, Huang Y, Dong C, Wang L, Hu Z, Zhuang M, Huang J, Shi J, Mei D, Liu J, Lee TH, Jin H, Li Z, Li X, Zhang J, Xiao L, Zhou Y, Liu Z, Liu X, Qin R, Tang X, Liu W, Wang Y, Zhang Y, Lee J, Kim HH, Denoed F, Xu X, Liang X, Hua W, Chalhoub B and Paterson AH (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nature Communications* **5**, 3930.
- Mason AS, Zhang J, Tollenaere R, Vasquez Teuber P, Dalton-Morgan J, Hu L, Yan G, Edwards D, Redden R and Batley J (2015) High-throughput genotyping for species identification and diversity assessment in germplasm collections. *Molecular Ecology Resources* **15**, 1091–1101.
- Motilal LA, Zhang D, Umaharan P, Mischke S, Pinney S and Meinhardt LW (2011) Microsatellite fingerprinting in the International Cocoa Genebank, Trinidad: accession and plot homogeneity information for germplasm management. *Plant Genetic Resources: Characterization and Utilization* **9**, 430–438.
- Ndjiondjop MN, Semagn K, Zhang J, Gouda AC, Kpekin SB, Goungoulou A, Wambugu P, Drame KN, Bimpong IK and Zhao D (2018) Development of species diagnostic SNP markers for quality control genotyping in four rice (*Oryza* L.) species. *Molecular Breeding* **38**, 131.
- OECD (2016) *Safety Assessment of Transgenic Organisms in the Environment, Volume 5: Chapter 3 Brassica Crops (Brassica spp.)*. Paris: OECD Consensus Documents, Harmonisation of Regulatory Oversight in Biotechnology, OECD Publishing.
- Olodo KF, Barnaud A, Kane NA, Mariac C, Faye A, Couderc M, Zekraoui L, Dequincey A, Diouf D, Vigouroux Y and Berthouly-Salazar C (2020) Abandonment of pearl millet cropping and homogenization of its diversity over a 40 year period in Senegal. *PLoS ONE* **15**, e0239123.
- Palanivel H and Shah S (2021) Unlocking the inherent potential of plant genetic resources: food security and climate adaptation strategy in Fiji and the Pacific. *Environment, Development and Sustainability* **23**, 14264–14323.
- Pilling D, Bélanger J and Hoffmann I (2020) Declining biodiversity for food and agriculture needs urgent global action. *Nature Food* **1**, 144–147.
- Pradhan A, Nelson MN, Plummer JA, Cowling WA and Yan G (2011) Characterisation of *Brassica nigra* collections using simple sequence repeat markers reveals distinct groups associated with geographical location, and frequent mislabelling of species identity. *Genome* **54**, 50–63.
- Shen X, Zhang Y, Lei Z, Lin Y, Cao M, Hang Y and Sun X (2018) Genetic relationships among 11 nonheading Chinese cabbage accessions and the SCAR marker specific for ‘Suzhouqing’. *HortScience* **53**, 1288–1293.
- Takayama S and Isogai A (2005) Self-incompatibility in plants. *Annual Review of Plant Biology* **56**, 467–489.
- U N (1935) Genome-analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japanese Journal of Botany* **7**, 389–452.
- Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, Huang S, Li X, Hua W, Wang J, Wang X, Freeling M, Pires JC, Paterson AH, Chalhoub B, Wang B, Hayward A, Sharpe AG, Park BS, Weisshaar B, Liu B, Li B, Liu B, Tong C, Song C, Duran C, Peng C, Geng C, Koh C, Lin C, Edwards D, Mu D, Shen D, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King GJ, Bonnema G, Tang H, Wang H, Belcram H, Zhou H, Hirakawa H, Abe H, Guo H, Wang H, Jin H, Parkin IA, Batley J, Kim JS, Just J, Li J, Xu J, Deng J, Kim JA, Li J, Yu J, Meng J, Wang J, Min J, Poulain J, Wang J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links MG, Zhao M, Jin M, Ramchiary N, Drou N, Barkman PJ, Cai Q, Huang Q, Li R, Tabata S, Cheng S, Zhang S, Zhang S, Huang S, Sato S, Sun S, Kwon SJ, Choi SR, Lee TH, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li Y, Du Y, Liao Y, Lim Y, Narusaka Y, Wang Y, Wang Z, Li Z, Wang Z, Xiong Z and Zhang Z (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics* **43**, 1035–1039.
- Warwick SI, Simard M-J, Légère A, Beckie HJ, Braun L, Zhu B, Mason PG, Séguin-Swartz GT and Stewart Jr CN (2003) Hybridization between transgenic *Brassica napus* L. and its wild relatives: *Brassica rapa* L., *Raphanus raphanistrum* L., *Sinapis arvensis* L., and *Erucastrum gallicum* (Willd.) OE Schulz. *Theoretical and Applied Genetics* **107**, 528–539.
- Weise S, Lohwasser U and Oppermann M (2020) Document or lose it-on the importance of information management for genetic resources conservation in Genebanks. *Plants* **9**, 1050.
- Wong MML, Gujaria-Verma N, Ramsay L, Yuan HY, Caron C, Diapari M, Vandenberg A and Bett KE (2015) Classification and characterization of species within the genus *Lens* using genotyping-by-sequencing (GBS). *PLoS ONE* **10**, e0122025.
- Xiao Q, Wang H, Song N, Yu Z, Imran K, Xie W, Qiu S, Zhou F, Wen J, Dai C, Ma C, Tu J, Shen J, Fu T and Yi B (2021) The Bnapus50 K array: a quick and versatile genotyping tool for *Brassica napus* genomic breeding and research. *G3 (Bethesda, Md.)* **11**, jkab241.
- Zhan Z, Nwafor CC, Hou Z, Gong J, Zhu B, Jiang Y, Zhou Y, Wu J, Piao Z, Tong Y, Liu C and Zhang C (2017) Cytological and morphological analysis of hybrids between *Brassicoraphanus*, and *Brassica napus* for introgression of clubroot resistant trait into *Brassica napus* L. *PLoS ONE* **12**, e0177470.