

Review of congruence between global crop wild relative hotspots and centres of crop origin/diversity

Maxted, Nigel; Vincent, Holly

DOI:

[10.1007/s10722-021-01114-7](https://doi.org/10.1007/s10722-021-01114-7)

License:

Creative Commons: Attribution (CC BY)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Maxted, N & Vincent, H 2021, 'Review of congruence between global crop wild relative hotspots and centres of crop origin/diversity', *Genetic Resources and Crop Evolution*, vol. 68, no. 4, pp. 1283-1297.
<https://doi.org/10.1007/s10722-021-01114-7>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.



REVIEW

Review of congruence between global crop wild relative hotspots and centres of crop origin/diversity

Nigel Maxted · Holly Vincent

Received: 22 October 2020 / Accepted: 11 January 2021 / Published online: 1 February 2021
© The Author(s) 2021

Abstract Crop wild relatives (CWR) are important sources of adaptive diversity for plant breeding programmes. This paper aims to investigate the extent to which the centres of crop origin/diversity are congruent with areas of high CWR diversity. We established the predicted potential CWR distributions for 1,425 CWR species related to 167 crops using 334,527 known distribution locations and generated a global CWR hotspot map. This was then compared to the centres of origin/diversity proposed by Vavilov (amended by Hawkes); Zeven and Zhukovsky's mega gene centres, Harlan's centres and non-centres of crop domestication; and crop domestication areas identified using current archaeological evidence proposed by Purugganan and Fuller. Greatest congruence between the global CWR hotspots and other concepts was found with the concept proposed by Vavilov and amended by Hawkes, but there remained significant differences between the CWR hotspots and Vavilov's concept. This paper concludes that all four centre concepts reviewed have some overlap with CWR diversity but that Vavilov's original concept has the closest geographic coincidence with CWR hotspots. With the benefit of significant additional global datasets to those used by Vavilov, we were able to suggest amendments to his concept, adding further

centres based on CWR hotspots in west and east USA, West Africa, South-east Brazil and Australia. As a result of this study more precise targeting of CWR and crop landrace can be implemented in future, aiding global food and nutritional security.

Keywords Agrobiodiversity · Centre of origin · Conservation · Crop wild relative · Genetic diversity · N.I. vavilov

Introduction

The transition from hunter-gatherer communities to agriculture first occurred some 10,000 years ago in several independent localities globally (Diamond 2002). Since the mid-nineteenth century scholars have attempted to pinpoint the locations where crops were domesticated, however, the exact whereabouts have proven elusive. Alphonse de Candolle was one of the first crop geographers who wrote extensively on the geography and origins of individual cultivated plant species, using historical data, presence of related wild species, variation patterns and archaeological information to determine broad areas of plant domestication (Candolle 1855). He combined his research on individual crop species to determine three separate centres of plant crop domestication, the Fertile Crescent, Mesoamerica and South East Asia, however

N. Maxted (✉) · H. Vincent
School of Biosciences, University of Birmingham,
Birmingham B15 2TT, UK
e-mail: n.maxted@bham.ac.uk

innovative his approach, the data available to him undoubtedly restricted his conclusions. Charles Darwin also investigated the variability of domesticated species versus their wild counterparts, although he focused primarily on the processes of inheritance and selection of traits under artificial human direction (Darwin 1868), he concluded it would be extremely difficult to discover the exact centres of domestication.

The person most often associated with centres of crop origin or diversity today is Nikolai Ivanovich Vavilov, the Russian geneticist and father of plant genetic resources conservation and utilisation. Inspired by the work of de Candolle and Darwin, Vavilov attempted to define global crop domestication areas more precisely and published his theory on the centres of crop origins/diversity based upon a study of crop and wild relative literature, geography and nomenclature (Vavilov 1926). Vavilov initially identified five centres of plant crop origination using a so called differential phytogeographical approach which involved the following steps: (a) delimitation of plants into Linnean species and morphological units; (b) determination of the geographical distribution of these plants in the past; (c) determination of the inherited variation of characteristics of each plant species; (d) identification of geographical areas where there is a wide range of inherited variation in crop varieties and multiple wild relative species presence (Vavilov 1992). The five centres he initially identified were geographically broad and encompassed the Mediterranean, Central and South America, the Far East and South-western Asia. The identification of these centres of crop origins/diversity formed the foundation and rationale for many of Vavilov's collecting missions.

The accumulation of information from a wide study of global plant diversity, collecting missions and an increase in archaeological findings helped Vavilov to refine his centres of crop origin and diversity theory, and increase the number of centres from five to eight, including several sub-regions, each with a wealth of crop landrace and CWR diversity (Vavilov 1935). These areas included: the Chinese centre; the Indian centre; the Indo-Malayan sub-centre; the Inner Asiatic centre; the Asia Minor centre; the Mediterranean centre; the Ethiopian centre; the Central American centre; The Peruvian-Ecuadorian-Bolivian centre with sub-centres in both Chiloe, Chile and around the Brazil-Paraguay border. Vavilov once again modified his theory in 1940 by combining the Inner Asia and

Asia Minor centres, whilst introducing a new sub-centre around Bogota, Colombia (Vavilov 1940); the Brazil-Paraguay sub-centre was omitted from this publication for unknown reasons, although Hawkes (1983) suggests it was overlooked accidentally during a period of severe personal and professional struggle for Vavilov and should be reinstated. Further, Loskutov (2020) recently pointed out a nuance that had been widely missed by those reading Vavilov in translation, that in his 1927 paper for the Fifth International Genetic Congress in Berlin, “Geographische Zentren unserer Kulturpflanzen” (geographical centres of our cultivated plants) Vavilov distinguishes between traditional agriculture stocks found in the centres of genetic diversity and wild or weedy relatives of crop species found in gene centres or centres of origin (Vavilov 1928).

After Vavilov's premature death, his colleagues continued to develop his centres of crop origin/diversity concept and even today, scientists are still investigating the originations of individual crops. P.M. Zhukovsky, a colleague of Vavilov's, sought to delimit areas of crop diversity and areas of wild species diversity separately (Zhukovsky 1965). He defined 12 broad areas termed megagene centres which contained a wealth of domesticated plant diversity (Zeven and Zhukovsky 1975). The megacentres were based upon Vavilov's centres of origin theory and showed areas of high crop and wild relative diversity. Zhukovsky enlarged centres such as the Ethiopian centre to encompass the whole of Africa and included the European and Siberian region along with the whole of Australia. Zhukovsky also described over 100 micro gene centres within the megagene centres which exhibited exceptional local diversity and richness of wild species related to cultivated crops. While Zeven and de Wet (1982) provide a useful discussion of the history of the centres they prefer to refer to regions of diversity, possibly because their concept covered such extensive areas that they could no longer be referred to as centres. They recognized the same 12 broad regions of diversity as Zeven and Zhukovsky (1975).

Harlan (1971) sought to advance the theories proposed by Vavilov and Zhukovsky, he proposed centres and non-centres of agricultural origin using a combination of methods rather than relying heavily on phytogeography techniques as Vavilov did. Harlan also had better data to work with as a great deal of

archaeological and plant genetic work had been accumulated since Vavilov's death. Harlan described three main centres in which agriculture developed independently and then spread, the Near East, the North Chinese and the Mesoamerican centres. Each main centre had a corresponding non-centre to which the ideas of crop domestication spread and were widely utilised leading to a more diffuse spread of domestication and a great variety of forms of crops (Harlan 1971). The corresponding non-centres are defined as the African, Southeast Asian and South Pacific and the South American non-centres.

More recently, Purugganan and Fuller (2009) using current archaeological information alone propose 24 areas of interest grouped into 13 centres of domestication for the major global crops. The centres are largely coincident with Zeven and Zhukovsky (1975), but distinguish primary and secondary centres, are more tightly drawn and exclude Australia as an area of interest. Current knowledge of the centres of crop domestication is being greatly enhanced by the significant ease of availability of ecogeographic and archaeological data and the developments of GIS and genomic scientific methods.

PGRFA, particularly crop landraces and crop wild relatives (CWR), contain the breadth of trait diversity required to sustain crop improvement. Crop landrace are the traditional varieties maintained by farmers over millennia through numerous cycles of planting, harvesting and seed selection, although they have passed through the domestication bottleneck they have not been bred by scientific breeders and so have retain significant trait diversity which is routinely used by plant breeders (Camacho Villa et al. 2005). While CWR are the wild and weedy relations of domesticated crops, including their progenitors. They have been increasingly and successfully used as gene donors for improving crops for over the last 60 years (Haijar and Hodgkin 2007; Maxted and Kell 2009). CWR and landraces vary in many ways, but one key difference is the availability of ecogeographic data, there is now a large global data sets available for CWR taxa (Castañeda-Álvarez et al. 2016a), whereas landraces datasets are comparatively rarely available and are not comprehensive (Maxted et al. 2009). The availability of large global CWR data sets meant it was for the first time possible to undertake global level *ex situ* (Castañeda-Álvarez et al. 2016b) and *in situ* (Vincent et al. 2019) gap analysis to identify priorities

for global conservation action. This data being available not because the taxa are CWR, but because they are wild, often common weedy species and have been studied by botanists for centuries. As the progenitors and congeneric taxa of domesticated crops, CWR closely related to crops are often assumed to exist near centres of crop origin/diversity (Vavilov 1926). As such the CWR conservation community regularly describe the eight Vavilov centres of origin as being areas of particular CWR richness worldwide, even suggesting conservation efforts should be concentrated there (Rubenstein et al. 2005; Maxted and Kell 2009; Stolten et al. 2010; Ford-Lloyd et al. 2011; Kell et al. 2015). However, the spatial relationship between CWR and centres of origin/diversity has yet to be tested and the availability of a large CWR global data sets (Castañeda-Álvarez et al. 2016a) means this is now possible.

In this paper we aim to investigate the extent to which the four centres of crop origin/diversity concepts are congruent with recently proposed high CWR diversity hotspots. Further whether there are any significant CWR hotspots found outside the proposed centres of crop diversity and whether a revision of centres of origin is appropriate. Whether the term centre of origin or diversity is more appropriate to describe the concept and the implications for results in terms of CWR conservation.

Methods

Crops important for food security and farmer income generation were identified for inclusion in the analysis by consulting the Harlan and de Wet CWR Inventory (Vincent et al. 2013) and the GRIN Global CWR Portal (GRIN Global 2015), which in turn used gene pool (GP) 1B or GP2; or, taxon group (TG) 1B-3 concepts to define inclusion (Maxted et al. 2006). To model the distributions of target CWR species, occurrence records were downloaded from the georeferenced CWR data repository (Castañeda-Álvarez et al. 2016a). The dataset was edited to remove cultivated taxa and occurrences, occurrences outside of taxon native range, non-target taxa and occurrences with no coordinates or inaccurate coordinates (where coordinates have greater than 10 km potential inaccuracy). Target CWR were recorded at the species level due to identification inconsistencies at the

subordinate taxa level and the poor number of occurrence records for many sub-taxa. Species nomenclature was revised to match that of the Germplasm Resources Information Network (GRIN) (GRIN Global 2015) and species native ranges were obtained from the Harlan and de Wet Inventory (Vincent et al. 2013).

To maximise the use of small sets of occurrence records per species and overcome the effects of uneven, unrepresentative sampling across species native ranges, species distribution modelling was used to predict potential CWR distributions. MaxEnt software (version 3.3.3a) was used to model species potential distributions due to it being considered the best algorithm for producing accurate predictive distribution maps with presence only data (Elith and Leathwick 2009). Only species with 10 or more unique occurrence records were modelled using MaxEnt (Ramírez-Villegas et al. 2010).

To measure suitability of habitat for modelled species and produce quality predictions, MaxEnt requires environmental variables, occurrence points and background points from the species native area (Phillips et al. 2006). Twenty-seven variables were chosen as potential inputs for MaxEnt modelling and were subjected to stepwise variance inflation factor (VIF) analysis to remove collinear variables based upon variable values obtained from occurrences of each CWR (Merow et al. 2013). Variables with a collinear threshold value of 10 or more were removed from the total set. Variables included bioclimatic variables obtained from the Worldclim database (Hijmans et al. 2005), altitude and seven major soil factors from the ISRIC World Soil Information database (Hengl et al. 2014).

Modelling was performed at the 2.5 arc minutes resolution (~ 5 km at the equator), using 10,000 random background points from each species native range to train models. MaxEnt was performed using a five-fold cross validation technique and models were projected onto the native range of the species. The criteria proposed by Ramírez-Villegas et al. (2010) were applied to each predictive CWR distribution model to test their validity; (a) the five-fold average of the test Area Under the Test of the Receiver Operating Characteristic (ROC) curve (ATAUC) was greater than 0.7, (b) the standard deviation of the ATAUC for the individual five individual folds was lower than 0.15, and, (c) the proportion of the predicted

distribution where the standard deviation was greater than 0.15 was less than 10% of the total predicted distribution. For valid models, presence/absence maps were created per species by applying the maximum training sensitivity plus specificity (MAXTRSS) logistic threshold to the logistic values generated across the native range (Liu et al. 2013). For CWR that produced invalid models or had fewer than 10 unique records, potential distributions were approximated using a 50 km circular buffer around each species occurrence (Hijmans et al. 2001).

To assess the congruence between global CWR hotspots and centres of plant origin, CWR geographical hotspots needed to be identified. A CWR species richness map at the 5 arc minutes (~ 10 km at the equator) resolution was produced by overlaying the potential CWR distributions, created using MaxEnt and the 50 km circular method, and counting the number of unique CWR per grid cell. The gridded CWR richness map was used as input in the Hotspot Analysis tool in ArcGis 10.2 to find geographical areas that were deemed significant high CWR richness hotspots. The Getis-Ord G_i^* statistic (Getis and Ord 1992) underpins the Hotspot Analysis tool and is calculated for every input feature. The statistic assumes a null hypothesis of Complete Spatial Randomness (CSR) for the features being assessed, meaning that the grid cells with high CWR richness counts are hypothetically assumed to be randomly spatially distributed about the study area. The observed sum of the grid cell values combined with its neighbours is then calculated and compared to the expected sum of these values under the assumed null hypothesis. If the observed sum of values differs greatly from the expected sum of values then a significant Z score is produced, meaning that the observed sum greatly varies from the mean under a normal distribution. The Z scores are produced in this manner for every cell in the study area and given a corresponding p value to define the confidence with which the null hypothesis can be rejected for that cell. For example a cell with a Z score of > 2.58 has a corresponding p value of 0.01, therefore with a confidence level of 99% it can be assumed that the results are not the product of a random distribution and the null hypothesis can be rejected. For each cell in the species richness grid, neighbours were set as those sharing a boundary edge or corner. Grid cells that had a positive Z score with significance level of 1% or less

($p < 0.01$), indicating substantial clustering, were selected as being spatially significant CWR rich hotspots.

To assess the congruence of the four concepts for the centres of crop domestication and CWR hotspots, shapefiles for the four concepts were digitised for use in GIS using the GIS software QGIS, version 'Essen' and validated to ensure polygons were not self-intersecting or unclosed. The individual centre shapefiles from each concept were then rasterised to a grid of 5 arc minutes resolution (~ 10 km at the equator) to enable comparison with CWR hotspots and CWR distributions, which were represented by rasters of the same resolution.

To assess how well the four concepts for the centres of crop origin and CWR richness matched, a global raster defining CWR hotspots was obtained from Vincent et al. (Vincent et al. 2019). The CWR hotspots were delineated using a species richness map as input for the Getis-Ord G_i^* statistic which measures whether the observed values in cells (in this case, number of unique CWR species) are significantly different from expected values under the null hypothesis of complete spatial randomness (CSR) (Vincent et al. 2019). Area of overlap and percentage coverage between CWR hotspots and individual centres within concepts on crop domestication and diversity, as well as each concept as a whole, was calculated by overlaying rasters in R (R Core Team 2015). Additionally, numbers of CWR potential distributions located within individual centres and total concepts were obtained. Concepts were ranked based upon three criteria to identify those best representing CWR hotspots. The rankings were: (a) Highest to lowest percentage of CWR hotspots area overlapping concepts, (b) highest to lowest percentage of concept area overlapping CWR hotspots, and (c) highest to lowest number of unique CWR potential distributions within concepts.

Results

A total of 1,425 CWR species were identified as being closely related to or important for crop breeding for 167 crops that significantly contribute to global human food security and farmer income provision (Vincent et al. 2013). The 1,425 CWR comprised 236 GP1B species; 675 GP2 species; 30 TG1B species; 154 TG3

species (Harlan and de Wet 1971); 103 species with confirmed breeding use, and 67 species with potential use in crop breeding. The downloaded CWR occurrence dataset was edited from over 5 million total records to 334,527 records of suitable quality for the analysis. Occurrence records with good quality coordinate data were unavailable for 164 of the target CWR species; therefore only 1,261 priority CWR could be included in the analysis. Furthermore, 470 species were poorly represented in the dataset with fewer than 10 unique records each. In total 791 CWR species potential distributions were successfully modelled using MaxEnt, with the remaining 470 CWR modelled using the 50 km circular buffer method (Vincent et al. 2019).

CWR hotspots were defined using the Getis-Ord G_i^* statistic in ArcMap 10.2. Figure 1 shows CWR hotspots are found throughout temperate, sub-tropical and tropical areas. Particular hotspots of species richness are concentrated around the Mediterranean basin and Europe, and including the Fertile Crescent, the Caucasus and Central Asia (including northern India). In North America, a large cluster of CWR hotspots is located on the eastern coast of the USA through towards Kansas, and another is located on the west coast of the USA from California to Washington State. In Central America, the CWR hotspots start from central Mexico and northern Nicaragua. In South America patches of CWR hotspots appear along the Andes, beginning in western Peru and Bolivia, linking to northern Argentina. Further clusters in the same region are found in eastern Paraguay and the border around Paraguay and Argentina. In Brazil, the major CWR hotspots are centred in Goiás, São Paulo and Minas Gerais, with smaller clusters of hotspots spreading south from these states to Rio Grande do Sul. In mainland Africa small, fragmented hotspots can be found in the west, along the coastal areas of Côte d'Ivoire, Nigeria and Cameroon. Similar small hotspots can be found in southwest Ethiopia and around the coastal borders of Tanzania and Kenya. Further CWR hotspots are found in southern India and Sri Lanka, northern Australia, southern China, Indonesia, Malaysia, the Philippines, Myanmar, Thailand, Laos and Cambodia. The global CWR hotspots (Fig. 1) constitute 7.44% of the terrestrial world area and cover 1,019 target CWR species from 160 key crops. Crops (and their CWR) found outside CWR hotspots are horseradish (*Amoracia rusticana* G.

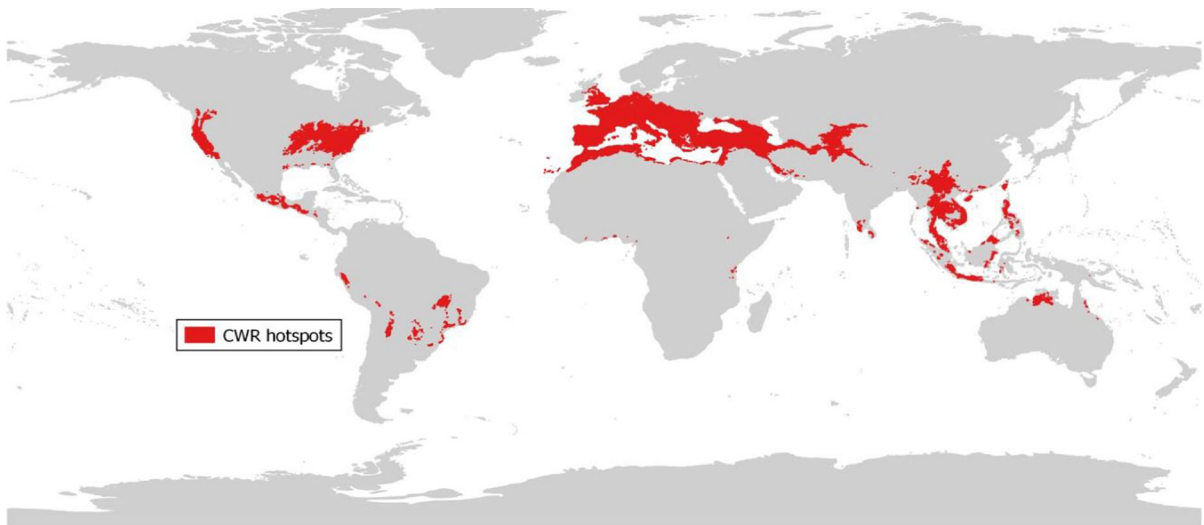


Fig. 1 Global CWR hotspots at 5 arc minutes resolution

Gaertn. et al.), brazilnut (*Bertholletia excelsa* Humb. & Bonpl.), aji (*Capsicum baccatum* L. var. pendulum (Willd.) Eshbaugh), mandarins (*Citrus reticulata* Blanco), cardamom (*Elettaria cardamomum* (L.) Maton), amur grape (*Vitis amurensis* Rupr.) and yautia (*Xanthosoma violaceum* Schott).

The GIS polygon representations of the four concepts for the centres of crop domestication were drawn in QGIS, version ‘Essen’, and are shown in Fig. 2. The four concepts on centres of crop domestication and diversity differ, both in terms of numbers of centres, their exact location and size. Indeed, the most encompassing concept, suggested by Zeven and

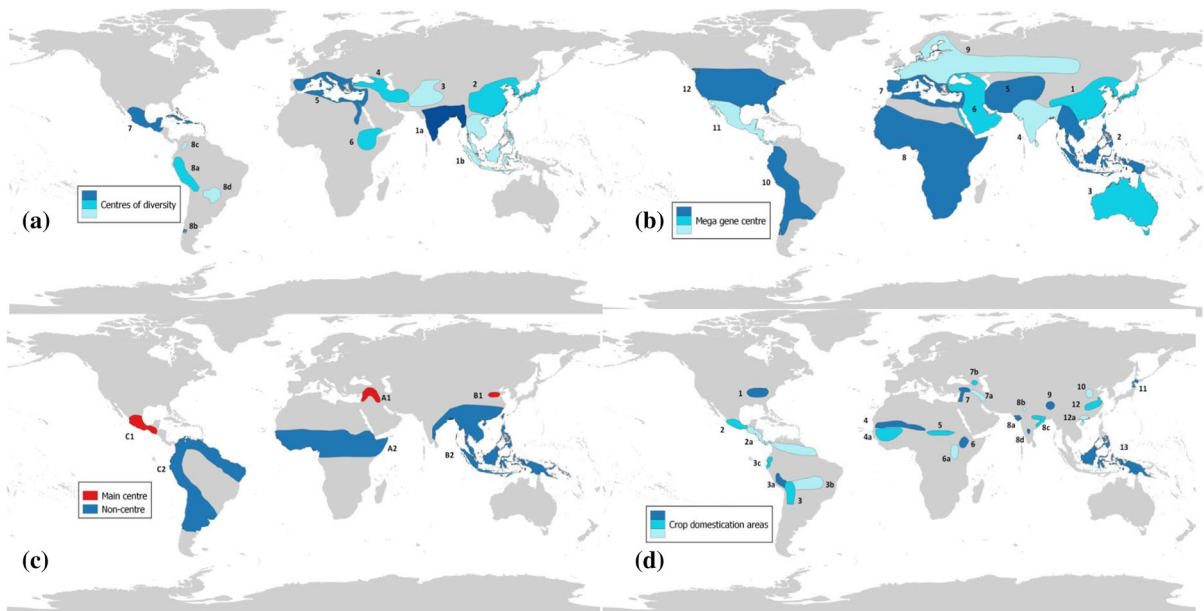


Fig. 2 Centres of crop domestication and diversity: **a** Vavilov centres of crop diversity (Vavilov 1935, 1940; Hawkes 1983); **b** Mega gene centres of cultivated plants (Zeven and Zhukovsky

1975); **c** Centres and non-centres of agricultural origin (Harlan 1971); and **d** Current knowledge on crop domestication areas (Purugganan and Fuller 2009)

Zhukovsky (1975), covers 1,184,377 grid cells and is 561.49% larger than the concept with the smallest area—that of Purugganan and Fuller (2009). Purugganan and Fuller (2009) have delimited the most individual centres with 24, followed by Vavilov (1935, 1940) amended by Hawkes (1983) and Zeven and Zhukovsky (1975) with 12 each, and lastly, Harlan (1971) has described the fewest with six.

To investigate the congruence percentage overlap of each concept with CWR hotspots and total number of CWR (Fig. 3). In terms of percentage coverage of CWR hotspots, the Zeven and Zhukovsky megagene centres contain the highest proportion of hotspots with 92.71%, followed by Vavilov centres of diversity with 49.02%, Harlan centres and non-centres with 20.25% and lastly, Purugganan and Fuller crop domestication areas with 11.99% CWR hotspot coverage. For the percentage of total CWR species within centres, Zeven and Zhukovsky megagene centres again contain the highest proportion with 1161 species (92.07% of total CWR) from 164 crops, followed by Vavilov centres of diversity with 910 species (72.16% of total CWR) from 162 crops, Harlan centres and non-centres with 789 species (62.57% of total CWR) from 150 crops and finally, Purugganan and Fuller crop domestication areas with 771 species (61.14% of total CWR) from 155 crops.

The CWR hotspot area, CWR species and related crop coverage of every individual centre in each concept of domestication and diversity is shown in Fig. 4. The Vavilov centres with greatest overlap with CWR hotspots were centres four (the Mediterranean) and five (Inner Asia), with 75.00 and 61.65%

respectively. The centres with no or poor overlap were 6, 8b, 8c and 8d. Vavilov centre 5 has the greatest number of total CWR with 279, followed by centre 4 with 219, and 1b with 184 CWR. Centre 8b contained the fewest CWR species with only two. Centre 3 had the greatest diversity of related crops with 79, closely followed by centres 5 and 4 with 78 and 77 CWR taxa respectively.

The Zeven and Zhukovsky megagene centres exhibit similar levels of CWR species representation to Vavilov centres; however, contain greater numbers of related crops. The centre with the greatest area overlaps with CWR hotspots was centre seven (the Mediterranean) with 69.25%, followed by centre two (East Asia) with 29.85% coverage. Centres with the least overlap were centre eight (Africa) and centre three (Australia). Centres containing the most CWR species were: seven with 279 CWR, six with 240 and two with 228 CWR. The fewest were contained in centre three, with 43 CWR. The greatest number of related crops was found in centre six with 85, closely followed by centre five with 84 crops. The Harlan centres and non-centres have poor intersection with CWR hotspots, except for the centre in the Fertile Crescent. They also contain low numbers of CWR species and related crops in comparison to Vavilov and Zeven and Zhukovsky centres. The Harlan centre with the greatest overlap with CWR hotspots was A1 (Fertile Crescent) with 63.65%, followed by C1 (Mesoamerica) with 24.95%. Centres B1 and A2 had no or very poor coincidence with CWR hotspots. The centre with the largest number of CWR species was B2 with 260 CWR, followed by A1 with 205 CWR;

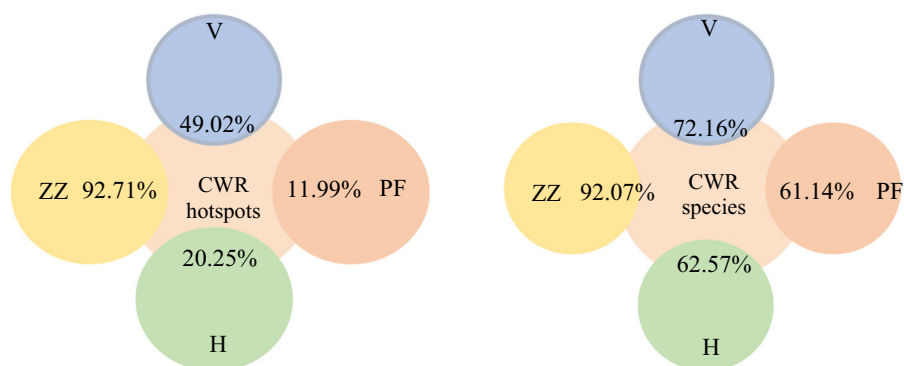


Fig. 3 Percentage of CWR hotspots area coincident with each concept of crop domestication and diversity and percentage of total CWR within each concept. V Vavilov centres of origin;

PF Puruggnan and Fuller crop domestication areas; H Harlan centres and non-centres of crop domestication; ZZ Zeven and Zhukovsky Mega gene centres

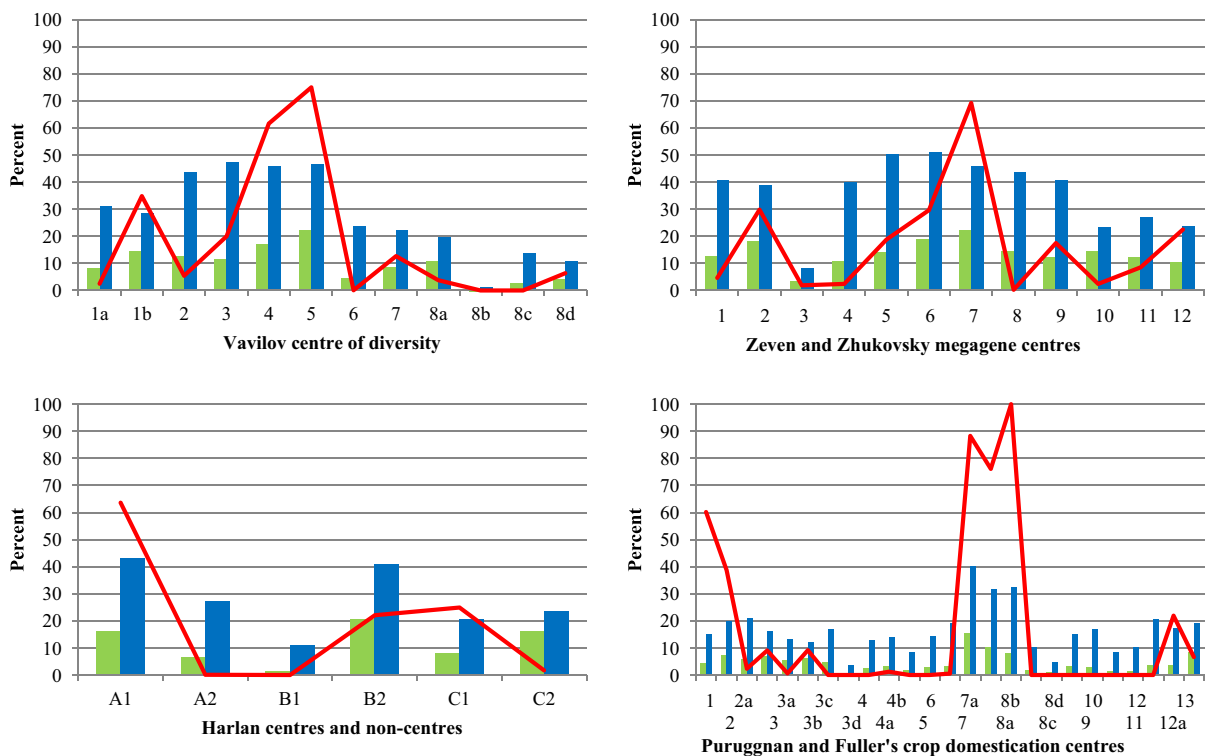


Fig. 4 Percentage overlap with CWR hotspots, percentage of total CWR species and percentage of total related crops per crop origin centre. The green bar = percentage of total CWR, the

blue bar = percentage of total related crops and red line = congruence with CWR hotspots (%)

centres B1 and A2 had the fewest CWR with 19 and 81 respectively. Centre A1 has the greatest diversity of related crops with 72, followed by B2 with 68; B1 had the fewest related crops with 18.

Puruggnan and Fuller centres of crop domestication have greater overlap with CWR hotspots than the centres in the three other concepts; however, the number of CWR species in each individual centre is much lower than all of the other concepts. In terms of number of related crops, the Puruggnan and Fuller centres are at a similar level to the Harlan centres and non-centres. Centre 7b has 100% overlap with CWR hotspots, followed by centre seven with 88.26% and 7a with 76.13%; however, 12 further centres had no overlap at all with CWR hotspots. Centre seven had the greatest number of CWR species with 194, followed by centre 7a with 127 and centre 13 with 103 CWR. Centre four had the fewest CWR with eight. Centre seven also had the greatest diversity of related crops with 67, followed by 7b with 54 and 7a with 53; the centres with the fewest crops were centre

four and centre 8b, with six and eight crops, respectively.

The Vavilov centres were the first concept for crop centres proposed and still today they remain the most widely quoted. Harlan (1992) criticises Vavilov's approach for being based on 'intuitive geographic patterns' and it is true that despite his extensive personal travel and collections, the evidence base available to Vavilov was limited. Subsequently much greater information was available to Harlan and later authors. Notably the generation of the Harlan and de Wet Inventory (Vincent et al. 2013) and the collation of the geo-referenced CWR data repository (Castañeda-Álvarez et al. 2016a), along with the associated global ex situ (Castañeda-Álvarez et al. 2016b) and in situ CWR (Vincent et al. 2019) gap analysis enabled a more evidence-based approach to current CWR conservation. However, the analysis presented above does demonstrate just how perceptive Vavilov's 'intuition' was for his time and our study confirms Vavilov's concept is still the best geographic representation of global PGRFA diversity available.

Although inevitably some additional areas, unrecognized by Vavilov, but identified by recent CWR ecogeographic data analysis (Vincent et al. 2013, 2019; Castañeda-Álvarez et al. 2016b) can be used to amend his concept and bring it in line with current knowledge. The additions being central Coastal and Minas Gerais Brazil, Western coastal USA, Eastern coastal to mid-West USA, Coastal West African, East African and Northern Australian (Fig. 5 and Table 1).

Vavilov produced several different versions of his concept, each with slight changes, so the version used for comparison here is Vavilov (1951), but following the suggestion by Hawkes (1983) of reinstating the Brazil-Paraguay sub-centre, its inclusion is justified by the analysis presented. However, one inclusion is unjustified by the analysis presented here is the Vavilov centre 8a—the island of Chiloe and surrounding area of Chile—Vavilov (1926) comments that the Chiloe centre is recognized because of the presence of long day potato landraces not found in Peru, Bolivia or Ecuador and not because of CWR presence which is used elsewhere to define centres. As Vavilov (1992) essential uses CWR and crop landrace richness (along with crop origin) to define centres and landrace richness was not assessed here, this sub-centre is retained in the amended concept.

Discussion

Use of the full breadth of plant genetic resources, in particular CWR genetic diversity, provides a vast underutilised source of genetic diversity that has the potential to greatly enhance plant breeding; to make more nutritious, higher yielding, increased abiotic and biotic tolerant and climate change resilient crop varieties (Hawkes et al. 2000; Warschefsky et al. 2014; Redden 2015; Maxted et al. 2020). However, CWR are under-conserved thus preventing breeders from accessing a wider range of genetic resources (Tanksey and McCouch 1997; McCouch et al. 2013) and threatened in the wild, jeopardising the future of CWR genetic diversity and its potential use in breeding (Kell et al. 2012). Throughout the CWR conservation literature, it is often remarked that Vavilov centres origin/diversity are coincident with areas of high CWR diversity (Rubenstein et al. 2005; Maxted and Kell 2009; Stolten et al. 2010; Ford-Lloyd et al. 2011; Hummer and Hancock 2015; Kell et al. 2015). In this paper we examined the validity of this hypothesis by examining whether the hotspots and distributions of 1261 CWR related to 167 crops were indeed coincident with Vavilov centres of origin and other theories on areas of crop domestication and diversity.

The results indicate that unsurprisingly, all four concepts have some overlap with CWR hotspots and individual distributions. Zeven and Zhukovsky

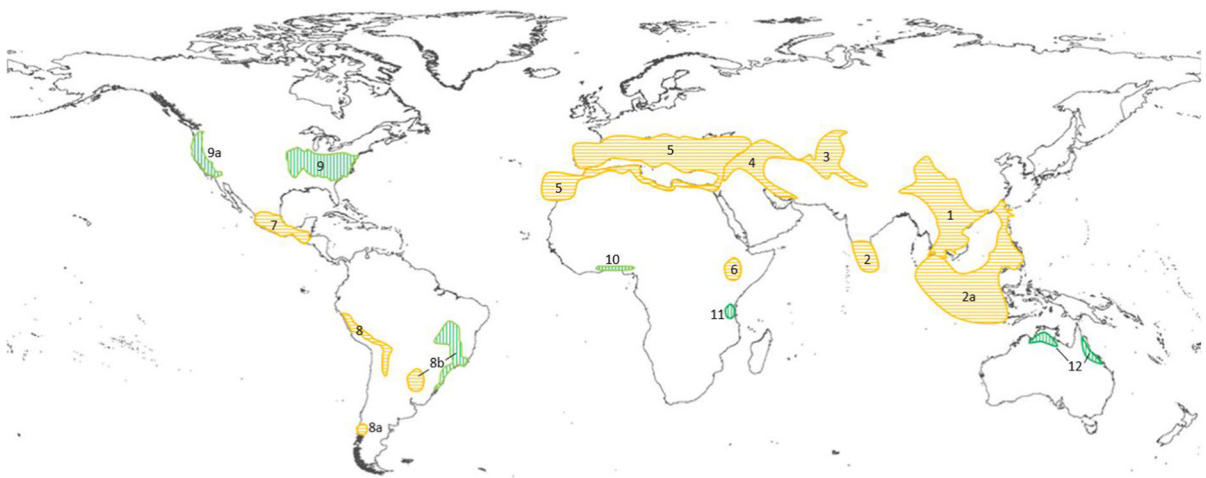


Fig. 5 Amended version of Vavilov centres of crop diversity concept (Orange horizontal hatched areas are original Vavilov centre and green vertical hatched areas are additional centres;

Numbers refer to Vavilov Centre standard notation with additional number added for additional centres identified)

Table 1 Crop gene pools associated with amended version of Vavilov centres of crop diversity concept

Gene Centres			Typical crop gene pools
Number	Name	Countries included	
1	Chinese	China, Vietnam, Laos and Cambodia	<i>Panicum miliaceum</i> L.—Chinese millet; <i>Sesamum indicum</i> L.—sesame; <i>Avena nuda</i> L.—naked oat; <i>Fagopyrum esculentum</i> Moench—buckwheat; <i>Glycine max</i> (L.) Merr.—soybean; <i>Vigna angularis</i> (Willd.) Ohwi & Ohashi—adzuki bean; <i>Raphanus sativus</i> L.—radish; <i>Brassica</i> L. species; <i>Colocasia esculenta</i> (L.) Schott—taro yam; <i>Allium</i> L. species; <i>Cucurbita moschata</i> Duchesne—butternut squash; <i>Phyllostachys</i> Siebold & Zucc. spp.—small bamboos; many temperate (<i>Pyrus</i> L., <i>Malus</i> Mill., <i>Prunus</i> L., <i>Chaenomeles</i> Lindl. spp.) and tropical fruit trees (<i>Citrus</i> L. spp.); <i>Camellia sinensis</i> (L.) Kuntze—Chinese tea; as well as plants producing oils, spices, medicines and fibres
2	Indian	India and Sri Lanka	<i>Oryza sativa</i> L.—rice; <i>Eleusine coracana</i> (L.) Gaertn.—finger millet; <i>Cicer arietinum</i> L.—chickpea; <i>Cajanus cajan</i> (L.) Huth—pigeon pea; <i>Vigna aconitifolia</i> (Jacq.) Maréchal—moth bean, <i>Vigna umbellata</i> (Thunb.) Ohwi & Ohashi—rice bean; <i>Vigna unguiculata</i> (L.) Walp. subsp. <i>sequepedalis</i> (L.) Verdc.—asparagus bean; <i>V. unguiculata</i> (L.) Walp. subsp. <i>cylindrica</i> (L.) Verdc.—catjang; <i>Trigonella foenum-graecum</i> L.—fenugreek; <i>Solanum melongena</i> L.—eggplant; several <i>Amaranthus</i> L. species; <i>Colocasia esculenta</i> (L.) Schott—taro yam; <i>Dioscorea alata</i> L.—yam; tropical fruits (<i>Citrus</i> L. spp., <i>Musa</i> L. spp., <i>Mangifera</i> L. spp.); oil producing species, fibres (<i>Corchorus olitorius</i> L.—jute), spices (<i>Piper nigrum</i> L.—pepper), stimulants and dye plants; sugar plants such as <i>Saccharum officinarum</i> L.—sugarcane
2a	Indo-Malayan	Thailand, Malaysia, Indonesia and Philippines	<i>Dioscorea</i> L. spp.—yams; <i>Citrus maxima</i> (Burm.) Merr.—pomelo; <i>Musa</i> L. spp.—banana; <i>Cocos nucifera</i> L.—coconut
3	Central Asian	Afghanistan, Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan	<i>Triticum aestivum</i> L.—wheat; <i>Pisum sativum</i> L.—garden pea; <i>Lens culinaris</i> Medik.—lentil; <i>Brassica</i> L., <i>Eruca</i> Mill. and <i>Lepidium</i> L. species; <i>Linum</i> L., <i>Sesamum</i> L. and <i>Coriandrum</i> L. (one of their centres); <i>Carthamus tinctorius</i> L.; <i>Cannabis sativa</i> L. subsp. <i>indica</i> (Lam.) E. Small & Cronquist; <i>Gossypium herbaceum</i> L.; various vegetables and melon species, spice crops, etc.; fruit and nut trees in the genera <i>Malus</i> Mill., <i>Pyrus</i> L., <i>Prunus</i> L., <i>Pistacia</i> L., <i>Juglans</i> L., <i>Corylus</i> L., etc
4	Neat Eastern	Turkey, Transcaucasia, Turkmenistan and Iran	<i>Triticum monococcum</i> L., <i>T. turgidum</i> L. <i>durum</i> (Desf.) Husn., <i>T. turgidum</i> L. and <i>T. aestivum</i> L.—wheats; <i>Secale cereale</i> L.—rye; <i>Avena sativa</i> L.—oat; <i>Cicer arietinum</i> L.—chickpea; <i>Lens culinaris</i> Medik.—lentil; <i>Vicia ervilia</i> (L.) Willd.—bitter vetch; <i>Pisum sativum</i> L.—garden pea; forages (<i>Medicago sativa</i> L.—lucerne, <i>Trifolium resupinatum</i> L.—strawberry clover, <i>Trigonella foenum-graecum</i> L.—fenugreek, <i>Onobrychis viciifolia</i> Scop.—sainfoin, <i>Lathyrus cicera</i> L.—chickling vetch and <i>Vicia sativa</i> L.—common vetch); oil-producing plants (<i>Sesamum</i> L., <i>Linum</i> L., <i>Brassica</i> L., <i>Camelina</i> Crantz, <i>Eruca</i> Mill. spp.); melons (<i>Cucumis</i> L. and <i>Cucurbita</i> L. spp.); vegetables (<i>Lepidium</i> L., <i>Brassica</i> L., <i>Daucus</i> L., <i>Eruca</i> Mill., <i>Allium</i> L., <i>Petroselinum</i> Hill, <i>Lactuca</i> L. and <i>Portulaca</i> L. spp.); fruit crops (<i>Malus</i> Mill., <i>Pyrus</i> L., <i>Punica</i> L., <i>Ficus</i> L., <i>Cydonia</i> Mill., <i>Vitis</i> L., <i>Pistacia</i> L. spp.); dye plants (<i>Crocus sativus</i> L. and <i>Rubia tinctorum</i> L.)
5	Mediterranean	Countries bordering the Mediterranean Sea	<i>Vicia faba</i> L.—fababean, <i>Lathyrus ochrus</i> (L.) DC.—Cyprus vetch, <i>Vicia sativa</i> L.—common vetch, large-seeded <i>Cicer arietinum</i> L.—chickpea, <i>Hedysarum coronarium</i> L.—Italian sainfoin, <i>Onobrychis viciifolia</i> Scop.—sainfoin; various oil-producing plants and spices; <i>Olea europaea</i> L.—olive and <i>Ceratonia siliqua</i> L.—carob; <i>Beta vulgaris</i> L.—beets, <i>Brassica oleracea</i> L.—cabbages, <i>Portulaca oleracea</i> L.—purslane, <i>Allium</i> L. spp.—onions, <i>Asparagus officinalis</i> L.—asparagus, <i>Lactuca sativa</i> L.—lettuce, <i>Pastinaca sativa</i> L.—parsnip, <i>Tragopogon porrifolius</i> L.—salsify; ethereal oil species and spices
6	Abyssinian	Ethiopia	<i>Triticum aestivum</i> L.—wheats, <i>Hordeum vulgare</i> L.—barley, <i>Sorghum bicolor</i> (L.) Moench—sorghum, <i>Cicer arietinum</i> L.—chickpea; <i>Lens culinaris</i> Medik.—lentil; <i>Vicia ervilia</i> (L.) Willd.—bitter vetch; <i>Pisum sativum</i> L.—garden pea, <i>Trigonella foenum-graecum</i> L.—fenugreek, <i>Brassica oleracea</i> L.—cabbages, <i>Allium</i> L. spp.—onions, <i>Lepidium sativum</i> L.—peppergrass, <i>Vigna unguiculata</i> (L.) Walp. subsp. <i>unguiculata</i> —cowpea, <i>Lupinus</i> L. spp.—lupins, <i>Linum usitatissimum</i> L.—flax; plus indigenous cereal <i>Eragrostis tef</i> (Zuccagni) Trotter—teff and <i>Eleusine coracana</i> (L.) Gaertn.—African millet; oil-bearing <i>Guizotia abyssinica</i> (L. f.) Cass.—Niger; <i>Coffea arabica</i> L.—coffee, <i>Catha edulis</i> (Vahl) Forssk. Ex Endl.—khat and <i>Ensete ventricosum</i> (Welw.) Cheesman—Abyssinian banana

Table 1 continued

Gene Centres			Typical crop gene pools
Number	Name	Countries included	
7	Mesoamerican	Mexico and Guatemala	<i>Zea mays</i> L.—corn / maize; <i>Phaseolus vulgaris</i> L.—common bean, <i>P. coccineus</i> L.—runner bean, <i>P. acutifolius</i> A. Gray—tepary bean; <i>Chenopodium berlandieri</i> Moq.—hauzontle and <i>Amaranthus cruentus</i> L.—purple amaranth; <i>Cucurbita</i> L., <i>Sechium</i> P. Browne and <i>Capsicum</i> L. spp. (<i>C. annuum</i> L.) bell and mostly mild hot pepper; <i>Pachyrhizus tuberosa</i> (Lam.) Spreng.—yam bean, <i>Ipomaea batatas</i> (L.) Lam.—sweet potato and <i>Maranta arundinacea</i> L.—arrowroot; <i>Gossypium hirsutum</i> L.—cotton; many tropical and temperate fruits; <i>Nicotiana tabacum</i> L.—tobacco, <i>Bixa orellana</i> L.—annatto and <i>Theobroma cacao</i> L.—cocoa
8	South American	Peru, Ecuador, Bolivia	<i>Solanum tuberosum</i> L.—potato, <i>Oxalis tuberosa</i> Molina—oca, <i>Tropaeolum tuberosum</i> Ruiz & Pav.—anu, <i>Ullucus tuberosus</i> Caldas—ulluco; <i>Solanum lycopersicum</i> L.—tomato, <i>Solanum muricatum</i> Aiton—Peruvian pepino, <i>Cyclanthera pedata</i> (L.) Schrad.—achocha or caigua, <i>Physalis peruviana</i> L.—Cape gooseberry and <i>Cucurbita maxima</i> Duchesne—pumpkin; <i>Phaseolus vulgaris</i> L.—common bean, <i>P. lunatus</i> L.—Lima bean, <i>Lupinus mutabilis</i> Sweet—pearl lupin, <i>Capsicum</i> L. spp. (<i>C. baccatum</i> L., <i>C. chinense</i> Jacq. and <i>C. pubescens</i> Ruiz & Pav.)—hot peppers, <i>Gossypium barbadense</i> L.—cotton <i>Chenopodium quinoa</i> Willd.—quinoa, <i>Chenopodium pallidicaule</i> Aellen—kañiwa, <i>Amaranthus caudatus</i> L.—foxtail amaranth, <i>Erythroxylum coca</i> Lam.—coca and <i>Lepidium meyei</i> Walp.—maca
8a		Chiloe, Chile	<i>Solanum</i> L. spp.; <i>Solanum tuberosum</i> L.—potato
8b		Brazil and Paraguayan	<i>Manihot esculenta</i> Crantz—manioc, <i>Arachis hypogaea</i> L.—peanut, <i>Theobroma cacao</i> L.—cocoa, <i>Hevea brasiliensis</i> (willd. ex A. Juss.) Müll. Arg.—rubber plant and <i>Ilex paraguariensis</i> A. St.-Hil.—mate
9	Western and Eastern USA	United States of America	<i>Allium sativum</i> L.—garlic, <i>Asparagus officinalis</i> L.—asparagus, <i>Avena sativa</i> L.—oats, <i>Brassica</i> L. spp.—cabbages, <i>Chenopodium quinoa</i> Willd.—quinoa, <i>Cucumis melo</i> L.—melon, <i>Digitaria exilis</i> (Kippist) Stapf—Fonio millet/White fonio, <i>Fragaria vesca</i> L.—strawberry, <i>Helianthus annuus</i> L.—sunflower, <i>Ipomaea batatas</i> (L.) Lam.—sweet potato, <i>Lactuca sativa</i> L.—lettuce, <i>Malus domestica</i> Borkh.—apple, <i>Manihot esculenta</i> Crantz—cassava, <i>Panicum miliaceum</i> L.—common millet, <i>Prunus avium</i> (L.) L.—cherry, <i>Pyrus communis</i> L.—pear, <i>Tripsacum</i> L. and <i>Zea</i> L. spp.—maize, and <i>Vitis vinifera</i> L.—grape vine
10	Coastal West African	Ghana, Togo, Benin, Nigeria and Cameroon	<i>Pennisetum glaucum</i> (L.) R. Br.—Pearl millet, <i>Vigna unguiculata</i> (L.) Walp. subsp. <i>unguiculata</i> —Cowpea, <i>Digitaria exilis</i> (Kippist) Stapf—Fonio millet/White fonio, <i>Oryza glaberrima</i> Steud.—African rice, <i>Dioscorea rotundata</i> Poir.—Yam, <i>Elaeis guineensis</i> Jacq.—African Oil Palm
11	East African	Tanzania and Kenya	<i>Coffea arabica</i> L.—Coffee, <i>Coffea canephora</i> Pierre ex A. Froehner—Robusta coffee, <i>Cucumis melo</i> L.—Melon, <i>Cucumis sativus</i> L.—Cucumber, <i>Digitaria exilis</i> (Kippist) Stapf—Fonio millet/White fonio, <i>Dioscorea cayenensis</i> Lam.—Lagos yam, <i>Echinochloa frumentacea</i> Link—White millet/Siberian millet, <i>Elaeis guineensis</i> Jacq.—African Oil Palm, <i>Eleusine coracana</i> (L.) Gaertn.—Finger millet, <i>Ensete ventricosum</i> (Welw.) Cheesman—Abyssinian banana, <i>Gossypium arboreum</i> L.—Tree cotton, <i>Gossypium barbadense</i> L.—Sea Island cotton, <i>Gossypium herbaceum</i> L.—Short-staple cotton, <i>Gossypium hirsutum</i> L.—Cotton, <i>Lablab purpureus</i> (L.) Sweet—Hyacinth bean, <i>Oryza glaberrima</i> Steud.—African rice, <i>Pennisetum glaucum</i> (L.) R. Br.—Pearl millet, <i>Saccharum officinarum</i> L.—Sugarcane, <i>Sorghum bicolor</i> (L.) Moench—sorghum, <i>Vigna subterranea</i> (L.) Verdc.—Bambara groundnut, <i>Vigna unguiculata</i> (L.) Walp. subsp. <i>unguiculata</i> —Cowpea
12	Northern Australian	Australia	<i>Glycine max</i> L.—soybean, <i>Gossypium</i> L. spp.—cultivated cotton crops, <i>Olea europaea</i> L.—olive, <i>Oryza glaberrima</i> Steud. / <i>Oryza sativa</i> L.—African rice and rice

megagene centres have the greatest overlap with CWR diversity, most likely due to the fact they cover such a high proportion of the terrestrial world compared to the other concepts. In fact, the combined total area of the other three concepts is still less than the area

covered by Zeven and Zhukovsky. This leads us to believe that these centres are too extensive and imprecise to correctly determine centres or hotspots. Harlan centres and non-centres appear to have the poorest fit with CWR hotspots and distributions, with

no centre located around the Mediterranean basin, Europe, or Central Asia where there is strong CWR hotspots presence. In this study, Purugganan and Fuller offer the most recent information on areas of crop domestication; therefore, they are more compact, well-defined, and specific in comparison to the relatively large areas defined in the other concepts. Although there is significant overlap with CWR hotspots for some centres identified by Purugganan and Fuller, such as those in the Caucasus and Fertile Crescent, 12 of their other centres indicate no overlap with CWR hotspots and six of those contained fewer than 20 CWR species. Furthermore, the major CWR hotspots in the Mediterranean and Europe are not represented.

Vavilov centres of origin/diversity are consistently placed in the top two highest rankings for CWR hotspot and distribution congruence, making them the best fitting concept in relation to CWR presence. However, Vavilov centres are not coincident with all CWR hotspots. Important areas such as eastern USA, the west coast of USA, west Africa, eastern south America and northern Australia were overlooked by Vavilov but contain a great deal of CWR diversity with sunflower (*Helianthus* L. spp.), grape (*Vitis* L. spp.), currants and fruit tree CWR in the USA; *Sesamum indicum* L., *Eleusine coracana* (L.) Gaertn., *Coffea* L. spp., various *Dioscorea* L. spp. and *Vigna* Savi spp. beans and *Psophocarpus* Neck. ex DC. CWR from west and east Africa, *Ilex* L. spp., *Manihot* Mill. spp., *Arachis* L. spp., *Gossypium* L. spp. and *Passiflora* L. spp. found in eastern south America; and *Oryza australiensis* Domin, *O. meridionalis* Ng, *O. rufipogon* Griff., *Gossypium* L. spp. and *Glycine* Willd. spp. found in northern Australia. It is unclear why Vavilov would have deliberately discounted these regions, but it seems most likely that despite his wide travels and several visits to the USA (Loskutov 1999), he did not have the distributional knowledge of CWR and crop landrace richness that has become so easily available to later authors (Harlan 1992; Purugganan and Fuller 2009; Jain and Kharkwal 2012; Vincent et al. 2013, 2019; Castañeda-Álvarez et al. 2016b). Although it should be noted that Vavilov's work in this area was continued by Vavilov Institute staff. Bakhareva (1988) applying the botanical and geographical methods developed by Vavilov established the existence of two additional independent gene centers of the origin and diversity of cultivated plants

and their wild relatives—West African and Central African. While Vavilov (1997) himself expresses an interest in greater understanding of Australian diversity, it was Zhukovsky (1965) who first recognised eastern and west USA and Australia as additional gene centres, proposals corroborate by this study's findings.

The hypothesis tested in this study was the coincidence between CWR hotspots and concepts of crop origin / diversity. Yet Vavilov based his centres on CWR and crop landrace richness and crop origin (Vavilov 1992) and we tested the hypothesis using CWR data alone. Although it should be acknowledged that crop landrace richness diversity was poorly understood in Vavilov's day and remains so today, only now being studied in more detail to aid their conservation planning and implementation (Veteläinen et al. 2009; Porfiri et al. 2009). Vavilov even in the 1920s would have had a more detailed knowledge of CWR phytogeography than for crop landrace and it is likely that CWR presence would have been the predominant factor the location of his centres. It is often assumed that hotspots for CWR and landrace richness are themselves correlated, as is indicated by discussion among PGR collectors (Maxted Pers. Comm.), but this has yet to be formally tested due to continued the lack of information on landrace phytogeography. The fact that at least in Europe, farmers are to be rewarded for the public good of CWR and crop landrace retention (Maxted et al. 2016), means data should soon be available to finally test this assumption and answer the question concerning CWR and crop landrace coincidence.

The other factor used by Vavilov to define centres was crop origin. Vavilov (1926) was aware of the literature on crop domestication available at the time, as can be seen from the extensive citation list included in his 'Centres of origin of cultivated plants' which he dedicated to Alfonse de Candolle for his work on phytogeography and the origin of crops. Vavilov incorporate the available knowledge into the identifications of his centres, but again subsequently there has been significant additions to the literature on crop domestication post-Vavilov's arrest in 1940, including the discussion provided by Purugganan and Fuller (2009). However, the question should be asked: whether there is a necessary correlation between CWR and landrace richness and sites of domestication? Our answer would be that there is likely to be a partial rather than full correlation, areas rich in GP1b

and GP2 CWR are likely to be sites of domestication, but there are also sites of crop domestication outside of CWR hotspots, as noted in our results above for horseradish, brazilnut, aji, mandarins, cardamom, amur grape and yautia. This suggested partial correlation is corroborated by the results of the matching of CWR hotspots to the Purugganan and Fuller (2009), whose 24 centres of domestication for major global crops showed a 11.99% correlation with CWR hotspots and 61.14 correlation with CWR species. Also, in this context it is worth considering what would an amended Vavilov concept be used for, it has and is likely to be used in focusing conservation action, therefore it is expedient to disassociate CWR and landrace richness from sites of domestication. Centres based purely on CWR and landrace richness are likely to be more appropriate targets of conservation action, than those thought to be related to historic domestication. Such centre based on, at least initially relative CWR concentration (till better landrace distributional data becomes available), might be better termed gene centres, using Vavilov (1928) own nomenclature, rather than origin which implies an alternative focus and one less a priority for contemporary conservation action.

This analysis highlights that all of the four centre of crop diversity concepts have some overlap with CWR diversity, however the most up to date knowledge on crop domestication areas indicated they are much more narrowly defined than CWR hotspots and centres of agrobiodiversity, which range beyond areas of domestication. Conversely the megagene centres or regions identified by Zeven and Zhukovsky are too all encompassing and non-specific, so whilst covering much CWR diversity it would be difficult to focus CWR conservation on such large areas. Harlan centres and non-centres poorly represent CWR diversity globally due to a lack of centres, particularly in Central Asia, the Mediterranean basin and Europe. Vavilov centres of diversity present a reasonable proxy for CWR species diversity; however, areas such as eastern South America, the United States of America, pockets of West and East African and Northern Australian were original unrecognized by Vavilov and should be added to his concept. There remains currently a lacuna in terms of available global landrace distributional data, as discussed above, however the increased use of environmental stewardship payment for public good conservation may mean

such data becomes increasingly available in the future and Vavilov's amended concept will possibly require further revision. These amended Vavilov centres of diversity should be given priority when planning global CWR conservation activities to maximise impact per resource availability for underpinning global food and nutritional security.

Author contributions Conceptualized N.M. and H.V.; investigation N.M. and H.V.; writing original draft H.V. and N.M.; reviewing and editing NM.; funding acquisition NM.

Funding This work was undertaken as part of the project 'Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives', which was supported by the Government of Norway. The project is managed by the Global Crop Diversity Trust and the Millennium Seed Bank of the Royal Botanic Gardens, Kew, and implemented in partnership with national and international gene banks and plant breeding institutes around the world. Further funding was provided by the FAO Commission on Genetic Resources for Food and Agriculture and the University of Birmingham.

Data availability Occurrence data used for this analysis (Castañeda-Álvarez et al. 2016a) are available at [http:// www.cwrdiversity.org/checklist/cwr-occurrences.php](http://www.cwrdiversity.org/checklist/cwr-occurrences.php).

Code availability The authors do not provide any code.

Compliance with ethical standards

Conflict of interest The authors are aware of no Conflicts of interest/Competing interests.

Ethical approval The authors are aware of no need for ethical approval in undertaking this research.

Consent to participate The authors are aware of no need for consent to participate.

Consent for publication The authors are aware of no need for consent to publish.

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

from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bakhareva SN (1988) Rastitel'nye resursy Zapadnoi i Tzentrāl'noi Afriki. [Plant Resources of West and Central Africa]. VIR, Leningrad (in Russian)
- Camacho Villa TC, Maxted N, Scholten MA, Ford-Lloyd BV (2005) Defining and identifying crop landraces. *Plant Genet Resour Charact Eval* 3(3):373–384
- Castañeda-Álvarez NP, Khoury CK, Achicanoy HA, Bernau V, Dempewolf H, Eastwood RJ, Guarino L, Harker RH, Jarvis A, Maxted N, Müller JV, Ramirez-Villegas J, Sosa CC, Struik PC, Vincent H, Toll J (2016) Global conservation priorities for crop wild relatives. *Nat Plants* 2:16022
- Castañeda-Álvarez NP, Khoury CK, Sosa CC, Sotelo S, Vane-gas I, Calderón SS, Vincent H, Harker RH, Bernau V, Eastwood RJ, Ramirez-Villegas J, Dempewolf H, Guarino L, Toll J, Jarvis A, Müller J, the Global Consortium of Crop Wild Relative Occurrence Data Providers and Maxted N (2016a) A global occurrence dataset for crop wild relatives. Available at: <http://www.cwrdiversity.org/checklist/cwr-occurrences.php>
- Darwin C (1868) *The variation of animals and plants under domestication*. Orange Judd & Co, London
- De Candolle ALPP (1855) *Géographie botanique raisonnée*. Victor Masson, Paris
- Diamond J (2002) Evolutionary consequences and future of plant and animal domestication. *Nature* 418:700–707
- Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. *Annu Rev Ecol Syst* 40:677–697
- Ford-Lloyd BV, Schmidt M, Armstrong SJ, Barazani O, Engels J, Hadas R, Hammer K, Kell SP, Kang DM, Khoshbakht K, Li Y, Long C, Lu B-R, Ma K, Nguyen VT, Qiu L, Ge S, Wei W, Zhang Z, Maxted N (2011) Crop wild relatives—undervalued, underutilized and under threat? *BioSci* 61:559–565
- Getis A, Ord JK (1992) The analysis of spatial association by use of distance statistics. *Geogr Anal*. <https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>
- GRIN Global (2015) US National Plant Germplasm System: Query Crop Relatives in GRIN Available at: <https://npgsweb.ars-grin.gov/gringlobal/taxon/taxonomysearchcwr.aspx> [Accessed 10th December 2015].
- Hajjar R, Hodgkin T (2007) The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica* 156:1–13
- Harlan J (1971) Agricultural origins centers and noncenters. *Science* 174:468–474
- Harlan JR, de Wet JMJ (1971) Towards a rational classification of cultivated plants. *Taxon* 20:509–517
- Harlan JR (1992) *Crops and man*, 2nd edn. American Society of Agronomy, Madison WI, USA
- Hawkes JG (1983) *The diversity of crop plants*. Harvard University Press, Cambridge Mass
- Hawkes JG, Maxted N, Ford-Lloyd BV (2000) *The ex situ conservation of plant genetic resources*. Kluwer Dordrecht, Netherlands
- Heng IT, de Jesus JM, MacMillan RA, Batjes NH, Heuvelink GB, Ribeiro E, Samuel-Rosa A, Kempen B, Leenaars JG, Walsh MG Gonzalez MR (2014) Soil Grids 1km—global soil information based on automated mapping. *PLoS One* 9(8) 105992.
- Hijmans RJ, Cameron SE, Para JL, JarvisA JPG (2005) Very high-resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25:1965–1978
- Hijmans RJ, Guarino L, Rojas CM, E. (2001) Computer tools for spatial analysis of plant genetic resources data: 1, DIVA-GIS. *Plant Genet Resour Newsl* 127:15–19
- Hummer K, Hancock J (2015) Vavilovian centers of plant diversity: implications and impacts. *HortScience* 50:780–783
- Jain H, K Kharkwal M (eds.) (2012) *Plant breeding: mendelian to molecular approaches*. Springer Science & Business Media
- Kell SP, Maxted N, Bilz M (2012) European crop wild relative threat assessment: knowledge gained and lessons learnt. In: Maxted N, Dulloo ME, Ford-Lloyd BV, Frese L, Iriondo JM, Pinheiro de Carvalho MAA (Eds.) *Agrobiodiversity Conservation: Securing the Diversity of Crop Wild Relatives and Landraces*. CAB International, Wallingford
- Kell S, Qin H, Chen B, Ford-Lloyd B, Wei W, Maxted N (2015) China's crop wild relatives: Diversity for agriculture and food security. *Agric Ecosyst Environ*. <https://doi.org/10.1016/j.agee.2015.02.012>
- Liu C, Newell WM, G. (2013) Selecting thresholds for the prediction of species occurrence with presence-only data. *J Biogeogr* 40:778–789
- Loskutov IG (1999) *Vavilov and His Institute: A History of the World Collection of Plant Genetic Resources in Russia*. IPGRI, Rome, Italy
- Loskutov IG (2020) Vavilov Institute (VIR): historical aspects of international cooperation for plant genetic resources. *Genet Resour Crop Evol* 67:2237–2253
- Maxted N, Kell S (2009) Establishment of a network for the *in situ* conservation of crop wild relatives: status and needs. Commission on Genetic Resources for Food and Agriculture, Food and Agriculture Organization of the United Nations, Rome, Italy. 211 pp. <http://www.fao.org/docrep/013/i1500e/i1500e18a.pdf>
- Maxted N, Ford-Lloyd BV, Jury SL, Kell SP, Scholten MA (2006) Towards a definition of a crop wild relative. *Bio-divers Conserv* 152:673–2685
- Maxted N, Amri A, Castañeda-Álvarez NP, Dias S, Dulloo ME, Fielder H, Ford-Lloyd BV, Iriondo J, Magos Brehm J, Nilsen L-B, Thormann I, Vincent H, Kell SP (2016) Joining up the dots: a systematic perspective of crop wild relative conservation and use. In: Maxted N, Ehsan Dulloo M, Ford-Lloyd BV (eds) *Enhancing crop genepool use: capturing wild relative and landrace diversity for crop improvement*, CAB International, Wallingford, UK, pp 87–124
- Maxted N, Hunter D, Ortiz Rios RO (2020) *Plant genetic conservation*. Cambridge University Press, Cambridge, UK
- Maxted N, Veteläinen M, Negri V (2009) Landrace inventories: needs and methodologies. In: Veteläinen M, Negri V,

- Maxted N (eds) European landraces: on-farm conservation management and use. Bioersivity Technical Bulletin Bioersivity International, Rome
- McCouch S, Baute GJ, Bradeen J, Bramel P, Bretting PK, Buckler E, Burke JM, Charest D, Cloutier S, ColeG DH, Dingkuhn M, Feuillet C, Gepts P, Grattapaglia D, Guarino L, Jackson S, Knapp S, Langridge P, Lawton-Rauh A, Lijua Q, Lusty C, Michael T, Myles S, Naito K, Nelson RL, Pontarollo R, Richards CM, Rieseberg L, Ross-Ibarra J, Rounsley S, Sackville Hamilton RS, Schurr U, Stein N, Tomooka N, van der Knaap E, van Tasse ID, Toll J, Valls J, Varshney RK, Ward J, Waugh R, Wenz IP, Zamir D, (2013) Agriculture: feeding the future. *Nature* 499(7456):23–24
- Merow C, Smith MJ, Silander JA (2013) A practical guide to MaxEnt for modeling species' distributions: what it does and why inputs and settings matter. *Ecography* 36(10):1058–1069
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modeling* 190:231–259
- Porfiri O, Costanza MT, Negri V (2009) Landrace inventories in Italy and the Lazio Region case study. In: European Landraces: On-farm conservation management and use (eds.) Veteläinen M, Negri V, Maxted N) Bioersivity Technical Bulletin 15. Pp.117–123. Rome, Italy: Bioersivity International
- Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. *Nature* 457:843–848
- R Core Team (2015) R: a language and environment for statistical computing. Foundation for statistical computing. Vienna Austria. ISBN 3–900051–07–0URL <http://www.Rproject.org/>
- Ramirez-Villegas J, Khoury C, Jarvis A, Guarino DDG, L, (2010) A gap analysis methodology for collecting crop gene-pools: a case study with *Phaseolus* beans. *PLoS ONE* 5(10):13497. <https://doi.org/10.1371/journal.pone.0013497>
- Redden R (2015) Wild relatives for the crop improvement challenges of climate change: the adaptation range of crops. In: Redden R, Yadav SS, Maxted N (eds) *Crop Wild Relatives and Climate Change*. Dulloo ME, Guarino L Smith P) Wiley-Blackwell Hoboken New Jersey, pp 61–79
- Rubenstein KD, Heisey P, Shoemaker R, Sullivan J Frisvold G (2005) *Crop genetic resources: an economic appraisal*. Economic Information Bulletin No (EIB2). United States Department of Agriculture, Beltsville, Maryland, USA.
- Stolten S, Maxted N, Ford-Lloyd B, Kell SP, Dudley N (2010) Food stores: protected areas conserving crop wild relatives and securing future food stocks. *Multiple Benefits for Conservation and Use*. Earthscan, London and Washington DC, Arguments for Protected Areas, pp 53–69
- Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277:1063–1066
- Vavilov NI (1926) The centres of origin of cultivated plants. *Works of Applied Botany and Plant Breeding* 16(2):48
- Vavilov NI (1928) Geographische Genzentren unserer Kulturpflanzen. *Verhandlungen des V Internationalen Kongresses für Vererbungswissenschaft*. Berlin 1927. 1:342–369.
- Vavilov NI (1935) The phytogeographical basis for plant breeding. *Theoretical Basis for Plant Breeding* 117–75.
- Vavilov NI (1940) The theory of the origin of cultivated plants after Darwin. *Nauka (Science)* 255–75
- Vavilov NI (1951) The origin, variation, immunity and breeding of cultivated plants. Transl. by KS Chester Ronald Press, NY
- Vavilov NI (1992) Origin and geography of cultivated plants. Cambridge University Press, Cambridge, UK, Transl. by D. Love
- Vavilov NI (1997) Five continents. Translated into English by Doris Lve. IPGRI/VIR, Italy, Rome
- Veteläinen M, Negri V, Maxted N (eds) (2009) European landraces: on-farm conservation management and use. Bioersivity Technical Bulletin Bioersivity International, Rome
- Vincent H, Amri A, Castañeda-Álvarez NP, Dempewolf H, Dulloo ME, Guarino L, Hole D, Mba C, Toledo A, Maxted N (2019) Modeling of crop wild relative species identifies areas globally for *in situ* conservation. *Communications Biology*. <https://doi.org/10.1038/s42003-019-0372-z>
- Vincent H, Wiersema J, Kell S, Fielder H, Dobbie S, Castañeda-Álvarez NP, Guarino L, Eastwood R, León B, Maxted N (2013) A prioritized crop wild relative inventory to help underpin global food security. *Biol Cons* 167:265–275
- Warschefsky E, Penmetsa RV, Cook DR, von Wettberg EJ (2014) Back to the wilds: tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. *Am J Bot* 101(10):1791–1800
- Zeven AC, de Wet MJM (1975) *Dictionary of cultivated plants and their regions of diversity*. Pudoc, Wageningen, the Netherlands
- Zeven AC, Zhukovsky PM (1982) *Dictionary of cultivated plants and their centres of diversity*. Pudoc, Wageningen, the Netherlands
- Zhukovsky PM (1965) Main gene centres of cultivated plants and their wild relatives within the territory of the USSR. *Euphytica* 14:177–188

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