Phylogenetic relationships among wild and domesticated *Capsicum* species for introgression breeding

Catherine Parry, Susan Lin, Vivian Wang, Sanjeet Kumar, and Derek Barchenger

World Vegetable Center, Shanhua, Tainan, Taiwan

Introduction

The genus *Capsicum* is a member of the family *Solanaceae*, the fruits of which have been an important crop throughout human history. *Capsicum,* commonly known as pepper, originates in the temperate and tropical regions of the Americas, in what is now Peru and Bolivia, where it was first domesticated approximately 8,000 to 10,000 years ago (Carrizo Garcia et al., 2016). Prior to domestication, *Capsicum* has been an important component of the human diet (Bosland et al., 2012; Kraft et al., 2014; Pickersgill, 1997; Scaldaferro et al., 2018). The genus comprises about 35 wild species of 11 clades (complexes) (Bosland and Votava, 2012; Carrizo García et al., 2016), and includes five key domesticated species: *C*. *annuum* (L.), *C*. *baccatum* (L.), *C*. *chinense* (Jacq.), *C*. *frutescens* (L.), and *C*. *pubescens* (Ruiz & Pav.).

The first-domesticated complex, the *annuum* complex, includes the closely related *C. annuum*, *C. chinense*, and *C. frutescens* species, which are the most economically important of the *Capsicum* species today. The *annuum* complex species are thought to have arisen as a result of human selection from its wild progenitor: *C. annuum* var. *glabriusculum* (Chiltepin) (Pickersgill, 1997), which has small, erect, round, red, hot fruits. This pungent flavour, common across many *Capsicum* species, is caused by capsaicinoids, which evoke effects of heat and spice. These compounds are secondary metabolites and protect the plants from herbivores, as their activation of the Vanilloid Receptors (TRPV1) on mammalian sensory nerve endings causes discomfort (Carlo and Tewksbury, 1999). Conversely, birds are attracted to the red, fleshy fruits of *Capsicum* species, and do not possess the VR1 receptor, so are not susceptible to the effects of capsaicinoids (Jordt and Julius, 2002). Birds eat the seeds of peppers, which have improved germination on clearing the avian gut, so are thought to have a role in seed distribution in an example of ‘directed deterrence’ (Tewksbury et al., 1999).

The presence of capsaicinoids make pepper particularly desirable for humans, having considerable cultural and culinary value as a common ingredient in global cuisines. Both the pungent/spicy (hot) and non-pungent (sweet) pepper varieties are important as vegetables, spices, medicines, and colourants. They provide a good source of vitamin C, provitamin A, vitamin E, and folate, so promoting consumption of *Capsicum* may provide an opportunity to combat nutrient deficiency, particularly due to its already wide consumption (Wahyuni et al, 2013; Kantar et al, 2016). While pepper is consumed globally, and demand is continually rising (particularly in the USA), the majority (65%) of production takes place in Asia (Gandonou and Waliczek, 2013; Tayyib, 2016; Gandonou and Waliczek, 2013).

There is high diversity across the genus, especially for traits such as fruit type, colour, shape, taste, size, and capsaicinoid content, and regional-specific preferences result in further variation. The primary genepool, which is extensive, consists of members of the same species, or those that are closely related, which can be directly hybridised with the species of interest to produce fertile progeny. The similarly diverse secondary genepool contains species that are related, but the progeny produced are sterile or not vigorous. The tertiary genepool are those plants that can be hybridised, but must undergo embryo rescue in order to produce viable progeny.

For the five primary domesticated species, there are three key genepool complexes that can be accessed for breeding. These complexes are based on their degree of genetic proximity and reproductive compatibility to domesticated species. For *C. annuum*, the primary genepool consists of breeding lines, cultivars, and landraces, as well as its wild progenitor, *C. annuum* var. *glabriusculum*. The secondary genepool includes *C*. *baccatum*,

C. *chacoense*, *C*. *chinense*, *C*. *frutescens*, and *C*. *galapagoense*, while the tertiary genepool includes *C*. *cardenasii*, *C*. *eximium*, *C*. *lanceolatum*, *C*. *praetermissum*, *C*. *pubescens*, *C*. *rhomboideum*, and *C*. *tovarii*.

In contrast to other Solanaceae genera (particularly tomato and potato), the use of introgression breeding in *Capsicum* to introduce new traits of interest has been relatively underutilized (Lin et al., 2014; Hirsch et al., 2013; Mongkolporn and Taylor, 2011). This is largely due to the lack of access to germplasm, as most of the publicly available germplasm collections contain few or no accessions of wild *Capsicum* species. In addition, pre-zygotic barriers such as pollen-pistil incompatibilities inhibit fertilisation, and post-zygotic barriers result in embryo or endosperm abortion, weak hybrids, and sterility (Kamvorn et al., 2014). The immediate need for introducing additional variation through introgression breeding has also been limited by the extensive diversity already found in the primary and secondary genepools of domesticated *Capsicum*. Variation also continues to arise through random mutation, which occurs at a high frequency in *Capsicum* due to retrotransposition- 81% of the *Capsicum* genus is made up of transposable elements (Qin et al., 2014). Nevertheless, the use of wild relatives of *Capsicum* still has value in modern breeding programmes as a source of novel traits for introgression and for increasing diversity in the current genepool.

As expected, most interspecific breeding programmes so far have been focused on introducing resistance to the many diseases and pests that attack pepper. For example, the introgression of virus resistance into modern *C. annuum* cultivars from the more tropical *C. chinense* (Gonzalez and Bosland, 1991). *C. annuum* var. *glabriusculum* is also a source of disease resistance, with particular potential for tolerance to curly top virus (Bosland 2000). Recently, an accession of *C. galapagoense* was identified as a possible source of resistance to whitefly, a serious insect pest (Mohammed Rhaka, personal communication). However, no successful progeny have so far been developed, despite extensive hybridisations and embryo rescue attempts (D. Coon, personal communication; M. Rhaka, personal communication). These results are surprising, because *C. galapagoense* has been reported as part of the *C. annuum* clade, and will readily hybridise with *C. annuum* accessions (Carrizo Garcia et al., 2016; Pickersgill, 1971). One reason for unsuccessful hybridisation between *C. galapagoense* and *C. annuum* accessions may be misidentification. Several genebanks have incorrectly reported accessions identified as *C. galapagoense* which are, in fact, *C. frutescens* (P.W. Bosland, personal communication). A correctly identified *C. galapagoense* accession (confirmed not to be *C. frutescens*), was recently acquired by the World Vegetable Center (WorldVeg).

The genetic diversity and variation within wild populations of *Capsicum* has been widely studied (Aguilar-Melendez et al., 2009; Carrizo Garcia et al., 2016; Cheng et al., 2016; Loaiza-Figueroa, 1989; Oyama et al., 2006; Votava et al., 2002), and the genomic sequence of wild *C. annuum* var. *glabriusculum* is publicly available (Qin et al., 2014). However, publicly available phenotypic data for wild *Capsicum* species remains very limited. One reason for this is that *in situ* and *ex situ* populations have been found to be genetically heterogeneous, and thus, based on a manageable number of plants, are difficult to characterise (Votava et al., 2002).

It will be necessary to more clearly assess the genetic diversity among wild and domesticated species of *Capsicum* in order to more efficiently introgress novel traits of interest into domesticated species. Furthermore, there is a clear need to evaluate the accuracy of previous reports of identity, and of hybridisation ability among different genetic groups of *Capsicum*. There is potential to identify bridge crosses for improving breeding efficiency, and to characterise the barriers to hybridisation among *Capsicum* species.

The main objective of this project is to better understand the phylogenetic relationships among wild and domesticated *Capsicum* species. Through development of novel interspecific hybrids, this work also aims to identify potential candidate species for introgression of traits into *C. annuum* through bridge crosses, with the ultimate aim to support trait discovery and prebreeding.

Materials and Methods

Plant material and interspecific crosses

Originally, 48 accessions of 15 species of *Capsicum* were chosen for this experiment. These accessions were collected from diverse locations and deposited into collections at either the WorldVeg Genebank, the WorldVeg Pepper Breeding Collection in Taiwan, or the Chile Pepper Institute, New Mexico State University in USA (Table 1). Of each accession, two biological replications were used where ever possible for phenotyping and genotyping. However, a number of accessions failed to germinate (highlighted in Table 1) therefore 38 accessions were used in the final experiment, some of which lacked a replicate.

Seeds were sown on 26 April 2018 into 72-cell plastic trays of sterilised peat moss. Trays were placed in a climate-controlled greenhouse for germination at 28 ± 3 ^{o}C with a 12-hour photoperiod and ≈95% relative humidity. At the 4-6 true leaf stage, the seedlings were transplanted into pots and moved to a greenhouse without climate control. On 13 Aug., the plants were moved to a screenhouse without climate control. Plants were irrigated twice daily and regularly fertilised with Nitrophoska during the experimental period.

The accessions were morphologically characterised based on the Descriptors of *Capsicum* (*Capsicum* spp.) manual (IPGRI, 1995), with differences within and among accessions noted to be used in future research. Reciprocal hybridisations were attempted among all combinations of accessions throughout the experimental period. Ability to hybridise in reciprocal was used to confirm previous reports of relatedness, and the presence of genetic complexes; primary, secondary, and tertiary genepools; and clades (Barchenger and Bosland, 2018; Carrizo Garcia et al., 2016; van Zonneveld et al., 2015; Pickersgill, 1971). The fruits of successful hybridisations were collected upon ripening. After three days the seeds were collected from the fruits and dried for more than 1 week. Once dried, a number of crosses of interest were chosen for sowing. Selected seeds were sown from 1 Oct. into 70-cell plastic trays containing sterilised peat moss. The trays were left in a greenhouse without climate control and irrigated twice daily, and observed regularly to assess germination.

Molecular analyses

DNA was extracted from young, actively growing leaves from plants of each accession using the modified CTAB extraction method (Tanksley, 1993). A Simple Sequence Repeat (SSR) study was carried out with 27 SSR markers (Table 2). DNA was amplified by PCR, for which each well of a 96-well microtitre plate contained 2 µl of template DNA, 0.4 µl of primer (0.2 l each forward and reverse), 0.1 µl of *Taq* polymerase, 0.4 µl of deoxyribonucleotides, 1.5 µl of Gold buffer, and sterile water to a final volume of 15 µl. The reactions were carried out in the thermal cycler with an annealing temperature of 55^{o}C. The PCR products were evaluated by electrophoresis for 30 minutes at 160 Volts on 6% acrylamide gels. The results were visualised under UV light following staining with ethidium bromide (Fig. 1). Electrophoresis was repeated whenever the clarity of the bands or their exact size was uncertain.

Data were collected from images of gels for each primer pair using a binary method: each accession was scored for presence or absence of amplicons of each size. The data was processed using the Proxy package (Meyer and Buchta, 2018) to produce a dendrogram for the assessment of the relatedness between the individual accessions (Fig. 1). A distance matrix was produced using the Dice index, and for hierarchical cluster analysis by the UPGMA (unweighted pair group method with arithmetic mean) method.

Further molecular analysis included the study of the *waxy* gene region of a sub-sample of six accessions, using a single primer pair, 860F and 2R (Table 3) (Garcia et al., 2016). The accessions selected for *waxy* sequence analysis were those expected to need clarification due to possible misidentification based on SSR and morphological data, as well as representative accessions of the species. The accessions included were: VI051012 (a *C. tovarii* accession); VI051011 (*C. galapagensis*, suspected to be *C. annuum)*; VI012574 (*C. chacoense*, suspected to be *C. annuum)*; PBC1892 (*C. galapagensis)*; VI013161 (*C.*

eximium); and PBC 556 (*C. frutescens)*. The region of interest was amplified by PCR in the thermal cycler with an annealing temperature of 60^{o}C using a PCR mixture prepared as for the SSR experiments. The quality of the products were evaluated by running on a 2% agarose gel with EtB'out' at 100 Volts for 50 minutes, then visualised using a Microtex Bio-1000F gel imager. Following electrophoresis, the PCR products were sequenced by Genomics Biotechnology Co., Ltd. (Xizhi Dist., New Taipei City Taiwan) by the Sanger sequencing method. Low quality nucleotides were manually removed throughout the resulting sequence, including approximately the first and last 60 nucleotides. The sequences were first aligned using NCBI nucleotide BLAST (Altschul et al., 1990) before a consensus sequence was constructed using the CAP contig assembly programme from BioEdit (Hall, 2005). The sequences were then aligned using multiple sequence alignment tool, Clustal MAFFT (Katoh and Standley, 2013), to produce a dendrogram.

Results

The results of the SSR study indicate clear groupings of accessions according to their species (Fig. 2). In particular, all of the *C. annuum* accessions (PBC1799, AVPP9905, PBC1867, PBC196, PBC142, VI029657) were grouped together, along with VI051012 (*C. tovarii*), VI051011 (*C. galapagensis*), and VI012574 (*C. chacoense*). Closely related to this cluster is a group that contained *C. chinense* accessions (PBC 1793, PI 152225, PI 159236, VI012668, VI029446), *C. frutescens* x *chinense* derivative (PBC1820), *C. eshbaughii* (NMCA90006), *C. eximium* (VI012964, VI01316), *C. frutescens* (PBC556), and *C. galapagensis* (PBC 1892).

As expected, the *C. baccatum* accessions (VI012528, VI014924, PBC81, PBC80, VI012478) grouped closely with the *C. praetermissium* accessions (VI029697, VI029696). Interestingly, we found NMCA90030 (*C. cardenasii*) clustered with *C. chacoense* (VI012900). We also identified four additional individual groups consisting of *C. cardenasii* (NMCA90035), *C. flexuosum* (NMCA50034, NMCA50030), *C. minutifolium* (NMCA50053), and *C. rhomboideum* (NMCA50064, NMCA50017).

The results of sequencing of the *waxy* gene region clusters together PBC 556 (*C. frutescens*), VI013161 (*C. eximium*), and PBC1892 (*C. galapagoensis*), and pairs VI051011 (*C. galapagensis*) with VI012574 (*C. chacoense*), while VI051012 (*C. tovarii*) can be considered an outlier from the other accessions (Fig. 3). The morphological data collected and success of interspecific hybridisation attempts provided valuable supporting evidence to the molecular analysis (Figs. 4 and 5). Over 2,500 crosses were attempted in all possible directions involving all the accessions, and the relative success of crossing varied (Fig. 5). *Capsicum chacoense* and *C*. *annuum* var. *glabriusculum* had the highest percentage success; however, variables such as environmental conditions and relative flowering period must be considered. Of the 112 crosses of interest that were selected to be sown to test germination and investigate viability, 22 germinated after more than 5 weeks (Table 4).

Table 1 List of *Capsicum* accessions included in this study. Those shaded in grey failed to germinate, so were excluded from analysis.

Accession	Other name	Species	Source	Plant number
AVPP9905	Susan's Joy	*C. annuum*	WorldVeg[z]	2
PBC 1867	CM 334	*C. annuum*	WorldVeg	2
PBC 196	California Wonder	*C. annuum*	WorldVeg	2
PBC 1799	Bird Pepper	*C. annuum*	WorldVeg	2
VI012528		*C. baccatum*	WorldVeg	2
VI014924	AJE	*C. baccatum*	WorldVeg	2
PBC 80		*C. baccatum*	WorldVeg	2
PBC 81		*C. baccatum*	WorldVeg	2
NMCA90030	PBC1987	*C. cardenasii*	NMSU	2
NMCA90031	PBC1988	*C. cardenasii*	NMSU	-
NMCA90035	PBC1989	*C. cardenasii*	NMSU	2
PBC1563		*C. chacoense*	WorldVeg	-
PI 159236	30040	*C. chinense*	USDA	2
PI 152225	Miscucho Colorado	*C. chinense*	USDA	2
VI012668		*C. chinense*	WorldVeg	2
VI029446		*C. chinense*	WorldVeg	2
PBC1793	Scotch Bonnet	*C. chinense*	WorldVeg	2
VI013161		*C. eximium*	WorldVeg	2
VI012964		*C. eximium*	WorldVeg	2
NMCA90006	PBC1990	*C. eshbaughii*	WorldVeg	2
NMCA50030	PBC1991	*C. flexosum*	NMSU	2
NMCA50034	PBC1992	*C. flexosum*	NMSU	1
PBC 488		*C. frutescens*	WorldVeg	-
PBC1820	Bhut Jalokia	*C. chinense*	WorldVeg	2
PBC537		*C. frutescens*	WorldVeg	-
PBC 556		*C. frutescens*	WorldVeg	1
PBC1892		*C. galapagensis*	WorldVeg	1
VI051011		*C. galapagensis*	WorldVeg	2
NMCA50053	PBC1993	*C. minutifolium*	NMSU	2
PBC 1887		*C. praetermissium*	WorldVeg	2
VI029696		*C. praetermissium*	WorldVeg	2
VI029697		*C. praetermissium*	WorldVeg	2
PI593617		*C. pubescens*	USDA	-
PI58250-4		*C. pubescens*	USDA	-
PI585264		*C. pubescens*	WorldVeg	-
PBC857		*C. pubescens*	WorldVeg	-
NMCA50076		*C. pubescens*	NMSU	-
NMCA50017	PBC1995	*C. rhobodiem*	NMSU	2
NMCA50064	PBC1996	*C. rhobodiem*	NMSU	1
VI051012		*C. tovarii*	WorldVeg	2
VI050239		*C. tovarii*	WorldVeg	-
VI012478		*C. buccatum*	WorldVeg	2
PBC 142, VI059328,		*C. annuum*	WorldVeg	2
VI029657		*C. annuum*	WorldVeg	2
VI012900		*C. chacoense*	WorldVeg	2
PI 574547, PBC1969	Chile que mira p'arriba	*C. annuum var. glabriusculum*	WorldVeg	2
PI 674459, PBC1970	BG2816 selection 16-1	*C. annuum var. glabriusculum*	WorldVeg	2
VI012574		*C. chacoense*	WorldVeg	2

[z]WorldVeg-World Vegetable Center, Tainan, Taiwan; NMSU- New Mexico State University, New Mexico, USA; USDA- United States Department of Agriculture.

Table 2 Forward and reverse primer sequences for the 27 SSR markers (source) used in the study.

SSR marker	Forward primer sequence (5'-3')	Reverse primer sequence (5′-3′)
SSR_P0_493610495#	TGTGGGATTGGCGCTTTAAG	TGTATACAAAAGCGTGGCGG
SSR_P0_491740645#	CGGTGGCTAGAGAGGAAGAG	AATCCGACTCACCTTCAGCA
SSR_P0_498360802#	GCCATGCAAACAGGAAAGGA	TTGGGGTGGGTATGAAAGGT
SSR_P0_163110979#	ACGGTTGGTGGACTCTCATT	GGGATTTGCGAAAACTTGCA
SSR_P0_438581243#	AGTCGACTTACAGCTGAGGT	ACCATATAATCACGCCTCAAGA
SSR_P0_488531631#	CCTAGAATGACCCCGACTGT	GGGGCTCCATACCAGAAAGA
SSR_P0_580381951#	TCAAATGGCTTCTGTTGAGGT	AGAGTATCTGACACGCACCA
SSR_P0_495003120#	GTTGGGTTGGCAATGGACAT	CTCGACTTGTCGCTTAGCAC
SSR_P0_496193829#	TGTCCTCCTCCTTCTCTTGT	GCAGCGACAGGAGTTGAAAA
SSR_P0_495114418#	GGTTGGGCTTGATGACTGTG	TCTTATCTCTCCGCACCGAC
SSR_P0_487884506#	GCCCGTCACTAAAAGTCTCC	TCTGGAATGGCTGTACTACCA
SSR_P0_494755411#	GTGAAGTCCGGAGAGAGTGA	GGCCGAGGATATGAAGGTGA
SSR_P0_471765445#	TGGCTGTTACCGGTCATCTT	TGCGTAACAGAGGATTGCAG
SSR_P0_547445673#	TTCCACCCTTACAGCTGAGG	AAGAAAGGGGTGGGGTATGG
SSR_P0_480905753#	AACTGTGCTCCTCCCTTCTC	TCGCCATTCACTTCACTCCT
SSR_P0_449855951#	GCCGTCACCTTCGATTACAC	TGTAATCGACGGTGCTAGCA
SSR_P0_494756680#	ACGCGCTTTGTCATGTGTAA	ACGTGTTAGCCTACGGTGAA
SSR_P0_587527153#	TCCAAACTACAAGCCTGCCT	TGACACCAAGCGACAACTTT
SSR_P0_462067694#	TACATCCGCCTCTGAACTCC	TTTACTTGTGGTTTCGGAAGC
SSR_P0_512937722#	TGTTGCTCGATTAGGCTGGA	GTTTTCAAGCAGTGCCTCGA
SSR_P0_480718061#	AGGGGTGTGACATCGTTCAT	CCAGAATCACTACCCAAGGC
SSR_P0_491728482#	ACTAACTGAAACGGCTGACAC	TCGGTGTCCAATGGTAAGCT
SSR_P0_456808942#	AAGCTCAGGACTCGTTTCATT	ACGTGAATGAGCCAAGTATGT
SSR_P0_480209261#	ACAAGTATGGAGGGAGCAAATT	CTCCCGAGGCCCCATATATC
SSR_P0_164189256#	CGCACCTTTTCCGACTCTTT	AGAAAGTCACTCCTCTCCGC
SSR_P0_530599646#	TTAGGGGCCCAACAGAAGAG	CTCATTCGTGTCGGTGTTGG
AVRDC-PP173	TCTGTTCTTCCCCAAAATCC	ATAGCGCCATCGAAGAAGAT
Anthracnose SCAR	GGTATCTTATTTCATAGGGACCAGGCA	TTTGCGGTAGTGACAACAACTTTACAGCCA

Table 3 Forward and reverse primer sequences for the sequence-based markers used for amplification of the Waxy gene region

Marker	Primer sequence
860F	CATAACATTGCCTACCAAGG
2R	GTTCCATATCGCATAGCATG

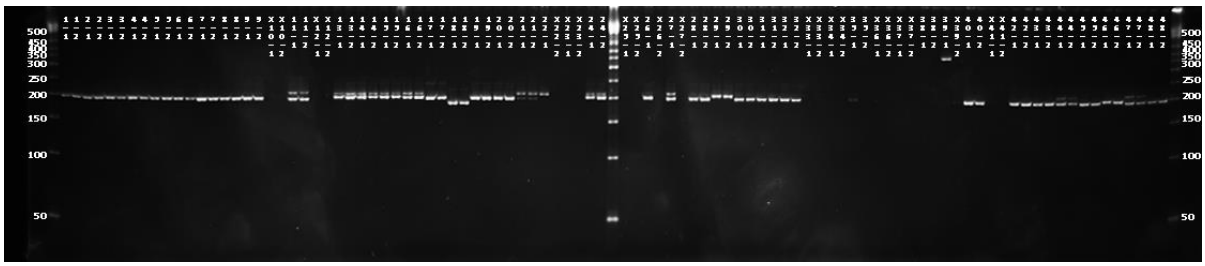

Figure 1. Example of an agarose gel visualised under UV, demonstrating the results of PCR with marker ‘SSR_P0_496193829’. The results of each marker were scored using a binary system, and the data collected used to produce the dendrogram.

Table 4 Successful germinations of the 112 crosses of interest sown and their % germination after more than 5 weeks.

Pedigree code	Parental species	Germination at 5 weeks after sowing (%)
VI012964/PBC 1867	*C. eximium/C. annuum*	80
PBC 1887/PBC1969	*C. praetermissium/C. annuum* var. *glabruisculum*	20
VI012574/VI051012	*C. chacoense/C. tovarii*	40
VI014924/PBC1969	*C. baccatum/C. annuum* var. *glabruisculum*	80
NMCA90030/AVPP9906	*C. cardensii/C. annuum*	60
NMCA90030/VI051012	*C. cardensii/C. tovarii*	100
NMCA90030/,PBC1969	*C. cardensii/C. annuum var. glabruisculum*	20
NMCA90030/VI012574	*C. cardensii/C. chacoense*	40
VI012900/VI012574	*C. chacoense/C. chacoense*	100
VI012964/PBC1820	*C. eximium/C. chinense*	100
NMCA90030/PBC 142	*C. cardensii/C. annuum*	100
VI029446/PBC1969	*C. chinense/C. annuum* var. *glabruisculum*	60
VI051011/VI012574	*C. galapagoensis/C. chacoense*	100
PBC 196/VI051011	*C. annuum/C. galapagensis*	20
NMCA90030/PBC1820	*C. baccatum/C. annuum*	100
VI012900/PBC1969	*C. chacoense/C. baccatum*	20
VI012574/PBC 142	*C. annuum/C. flexosum*	100
PI 152225/NMCA90030	*C. chinense/C. annuum* var. *glabruisculum*	80
PBC 142, VI059328	*C. chacoense/C. annuum*	100
PBC1799/VI012900	*C. chinense/C. baccatum*	100
VI012574/,PBC1970	*C. annuum/C. chacoense*	100
PBC 556/PBC1970	*C. frutescens /C. annuum*	80
VI014924/PBC1969	*C. baccatum/C. annuum* var. *glabruisculum*	80
VI014924/PBC1969	*C. baccatum/C. annuum* var. *glabruisculum*	80

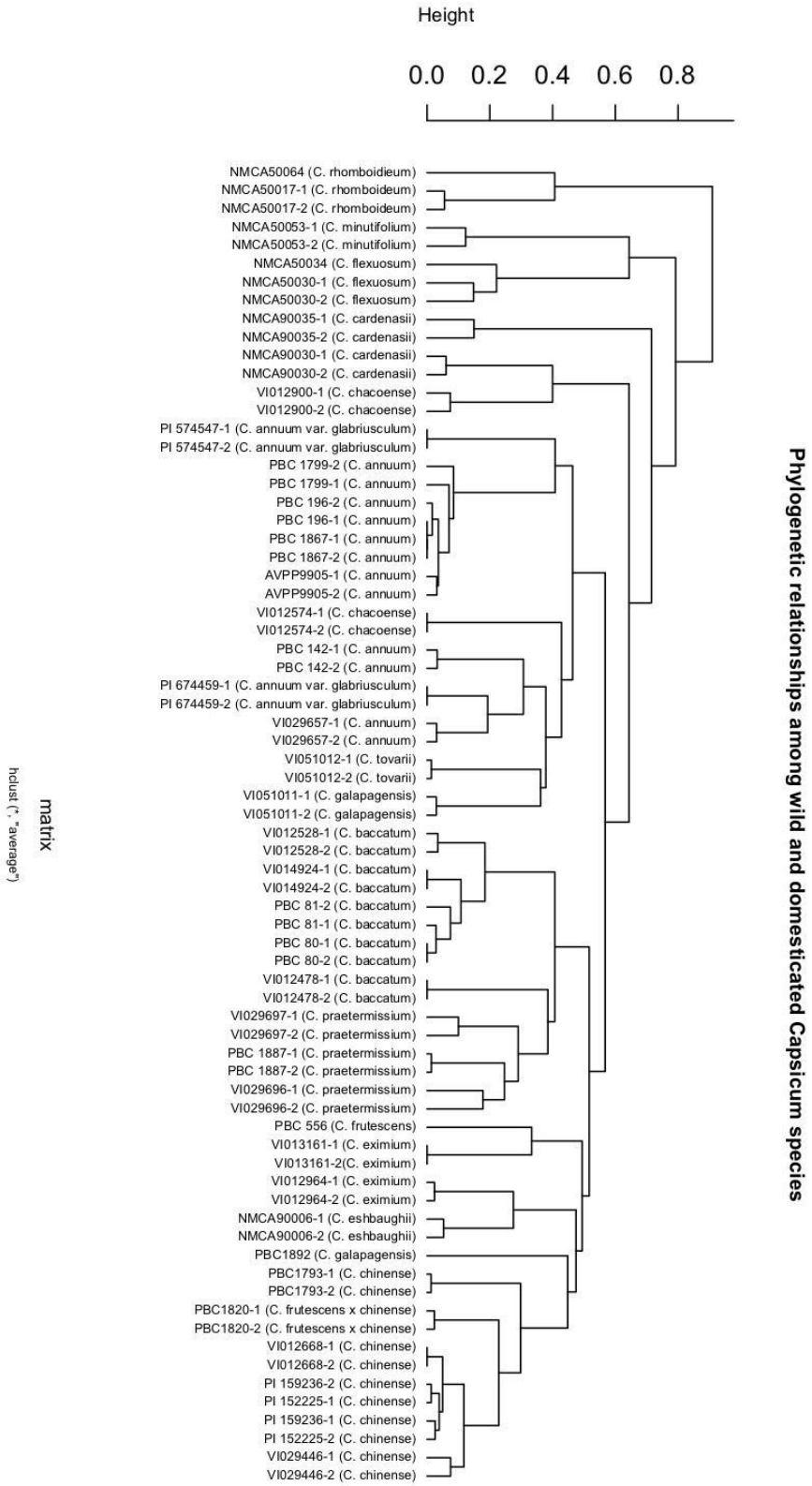

Figure 2 Dendrogram illustrating phylogenetic relationships between wild and domesticated *Capsicum* species developed using 27 SSR markers. Produced by the average linkage (UPGMA) method. 'Height' represents dissimilarity derived from the DICE matrix.

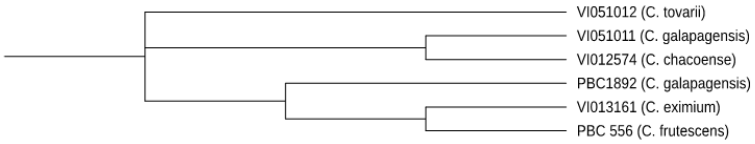

Figure 3 Dendrogram to illustrate phylogenetic relationships between six wild *Capsicum* species according to their respective *waxy* gene sequences

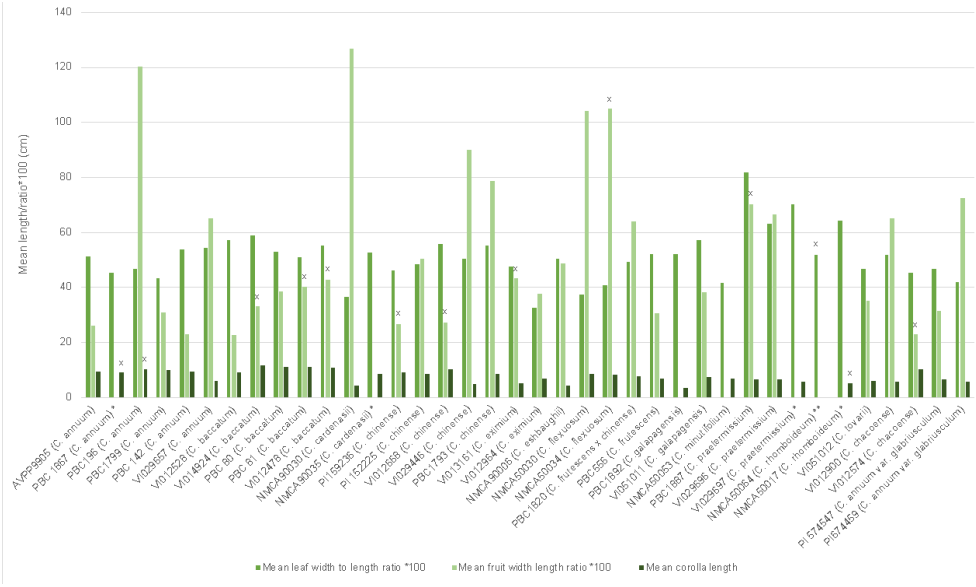

Figure 4. Mean mature leaf and mature fruit width to length ratios (x100) (cm), and mean corolla length (cm) in *Capsicum* species. Values taken as a mean of measurements of 10 or more components, unless marked by “x”. Fruit data not available for accessions marked with "*"; fruit and flower data not available for accessions marked with "**.

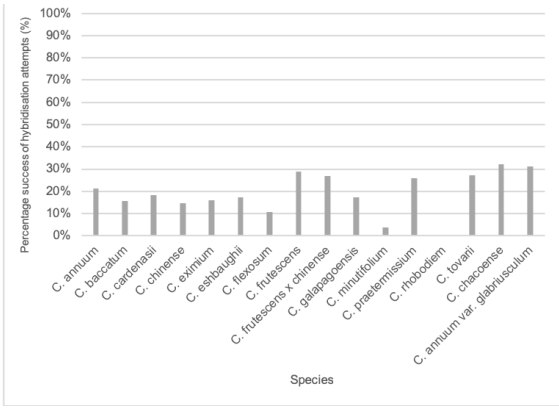

Figure 5. Percentage success of total hybridisation attempts between accessions of each *Capsicum* species

Discussion

Understanding the relatedness between accessions and the extent of their ability to hybridise is key in identifying candidates for the introgression of traits of interest into commercial varieties. Interactions between species and their compatibility can inform the design of genetic bridge strategies. This is particularly relevant in *Capsicum*, which has relatively low reported reproductive compatibility between complexes, but in which the wide diversity has great potential for crop improvement.

The results of this study are in general agreement with previous understanding of the *C. annuum* primary complex, which consists of *C. annuum*, *C. frutescens*, and *C. chinense*, along with *C. annuum* var. *glabriusculum.* The relationships between these species was reflected in the SSR analysis (Fig. 2), while a small number of crosses were achieved between each combination of these species.

Also placed within the *C. annuum* cluster are *C. chacoense* accession VI012574, and *C. tovarii* accession VI051011. The classification of *C. tovarii* remains unresolved, with some studies supporting its allocation into the *C. baccatum* complex (Tong and Bosland, 1999), while others have tentatively placed *C. tovarii* within *C. annuum* (Ibiza 2012). We successfully made 25 hybridizations between *C. tovarii* and accessions of a range of species, and the seeds of both of the sown hybridisations germinated (Table 4). Our results suggest *C*. *tovarii* might be correctly placed within the *C. annuum* complex, although further research in this area is needed.

The grouping of PBC 1892 (*C. galapagoensis*) in the *C. annuum* complex supports previous indications that *C. galapagoensis* is derived from a *C. annuum* progenitor population (Choong 1998). However, despite at least 18 hybridization attempts being made, crossed fruit between PBC1982 and any *C*. *annuum* accessions could not be obtained. Furthermore, we were unable to successfully hybridize PBC 1892 to PBC 556 (*C*. *frutescens*), which were grouped closely on the SSR-based dendrogram. The apparent disconnect between genetic similarity and morphological differences along with the low level of crossability with *C*. *galapagoensis* is noteworthy, and provides a basis for future work to better understand this unique species. The second *C. galapagoensis* accession included in the study, VI051011, was placed within the *C. annuum* cluster, and separated from its counterpart. This result is interesting given the morphology of this accession. *C. galapagoensis* is characterised by a number of features including pubescent stems and leaves, lack of anthocyanin, and small flowers and fruits (Csilléry, 2006), of which PBC1892 is a typical example. However, VI051011 has a tall growth habit, relatively large leaves and flowers, and sparse pubescence which resemble *C. annuum* characteristics. This evidence indicates misidentification of VI051011 as *C. galapagoensis*, and this accession should be assigned to *C. annuum*.

Based on both SSR and *waxy* gene sequence analyses, two *C. eximium* accessions (VI01316 and VI012964) clustered in the *C. annuum* complex, close to *C. frutescens* and *C. eshbaughii* (Figs. 2 and 3). This is an unexpected grouping, as *C. eximium* are isolated from *C. cardenasii*, despite these species being considered very closely related (Ibiza et al., 2012). It is possible that the *C. eximium* accessions included have been misidentified. Descriptions of *C. eximium* are relatively inconsistent, however, there is emphasis on its purple flowers (with previous allocation in the 'purple corolla' clade) and very small, soft fruits (Carrizo García et al., 2016; Csilléry, 2006). VI01316 and VI012964 both have white flowers, while VI012964 has relatively large fruits, both differing from what is expected in *C. eximium*. Thus, this poses questions as to the true identities of the *C. eximium* accessions.

Also adjacent to the *C. annuum* complex is a grouping that may be considered the *C. baccatum* complex, containing species in the *C. annuum* secondary genepool. This complex consists of *C. baccatum* accessions (VI012528, VI014924, PBC81, PBC80, VI012478) and *C. praetermissium* accessions (VI029697, VI029696). This confirms previous suggestions that *C. praetermissium* is in the *C. baccatum* primary genepool, and comprises a closely-related subgroup of *C. baccatum* (McLeod et al, 1981; Ibiza et al, 2012). *Capsicum baccatum* and *C. praetermissium* readily hybridised in both directions, while *C.*

praetermissium demonstrated particular tendency to hybridise with members of both *C. annuum* and *C. baccatum* complexes, which has not been previously reported. However, of the crosses between *C. praetermissium* and *C. annuum* that were sown, there was no instance of germination, with the exception of one cross between *C. praetermissium* and *C. annuum* var. *glabriusculum*, reflecting previous demonstrations that these hybrids rarely produce viable seed (Table 4).

In the wider *C. baccatum* complex are the closely related VI012900 (*C. chacoense*) and NMCA90030 (*C. cardenasii*), and slightly more distantly related NMCA90035. *C. chacoense*, although uncertain, is generally considered to be within the *C. baccatum* complex (Walsh and Hoot, 2001; Ibiza et al., 2011). *Capsicum chacoense* accessions readily hybridised with accessions of both *C. annuum* and *C. baccatum* complexes, in both directions. There was successful germination recorded in the four examples sown of crosses between *C. chacoense* and *C. annuum* complex species (Table 4), but no germination by any cross between *C. chacoense* and *C. baccatum*. This suggests that there may be a closer relationship between *C. chacoense* and the *C. annuum* complex than *C. baccatum* complex. This was also evident from clustering results based on SSR analysis.

A second *C. chacoense* accession, VI012574 was grouped away from the other *C. chacoense* (VI012900) and adjacent to the *C. annuum* complex (Fig. 2). VI012574 was selected for sequencing in order to investigate its true identity. Its morphology differs notably from VI012900, and from typical *C. chacoense* characteristics, which includes its bushy growth and thin branches, flowers with yellow anthers, and small round fruits (Csilléry, 2006), of which VI012900 is representative. In contrast, VI012574 has upright growth, elongated fruits, and relatively large flowers with blue anthers, typical features of *C. annuum*. Furthermore, VI012574 is clustered with VI051011, whose identity is also questioned. This suggests that VI012574 has been misidentified as *C. chacoense*.

More distantly related from the *C. annuum*, complex are three species making up smaller individual groups. *C. flexuosum*, *C. minutifolium*, and *C. rhomboideum*, which are known to be relatively isolated groups (Walsh and Hoot, 2001). The isolation of *C. rhomboideum* from the genus *Capsicum* is supported by its distance from *C. annuum*, the absence of successful hybridisations with any other accession, and distinct phenotype.

Many hybridisations were successful between *C. flexuosum* and a number of species, with *C. flexuosum* consistently being the male parent. These findings support the isolation of *C. flexuosum*, rather than a particular association with the *C. baccatum* and *C. annuum* clades.

Conclusions

The results of the SSR study, sequencing of the *waxy* gene sequence, morphological data, and crossability study between 48 accessions of 15 *Capsicum* species suggest that the *C. annuum* complex comprises *C. annuum*, *C. frutescens*, *C. chinense*, *C. annuum* var. *glabriusculum*, and *C. galapagoensis* (PBC1892). The *C. baccatum* complex, and *C. annuum* secondary genepool consists of *C. baccatum* and *C. praetermissium*, while *C. chacoense* is included in the *C. baccatum* secondary genepool. *C. flexuosum*, *C. minutifolium*, and *C. rhomboideum* each comprise independent clades with progressively distant relationships to the *C. annuum* complex. The validity of the identities of VI051011 as *C. galapagoensis* and VI012574 as *C. chacoense* are questioned; their morphology and clustering within the *C. annuum* complex suggests that further investigation may find them to be *C. annuum*. Confirmation of the identities of VI013161 and VI012964 as *C. eximium* is also recommended. The low level of association between apparent genetic relatedness and ability to hybridize identified in this study is novel, and sheds lights on a complex genome. This finding provides a basis for future research in the areas of pre- and post-zygotic barriers to hybridization, relatedness among domesticated *Capsicum* and wild relatives, and the usefulness of SSR markers to study diversity.

Literature Cited

Aguilar-Melendez, A., Morrell, P.L., Roose, M.L., and Kim., S.C., 2009. Genetic diversity and structure in semiwild and domesticated chiles (*Capsicum annuum*; Solanaceae) from Mexico. *Amer. J. Bot*. 96, pp. 1190-1202.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. J. Mol. Biol. 215, pp. 403-410.

Arjun, K., 2016. Mapping of QTLs in *Capsicum annuum L.* × *Capsicum galapagoense* Hunz. Thesis (MSc.), Punjab Agricultural University, Ludhiana.

Barchenger, D.W. and Bosland, P.W., 2018. Wild chile pepper (*Capsicum* spp.) of North America. In. S. Greene, K. Williams, C.K. Khoury, M.B. Kantar, and L. Marek (eds.). Cultivated Wild Relatives of North America. Springer New York, N.Y

Bosland, P.W. and Votava, E.J., 2012. Peppers: Vegetable and Spice Capsicums. 2nd ed. CAB International, Oxfordshire, UK.

Bosland, P.W., 2000. Sources of curly top virus resistance in *Capsicum*. *HortScience* 35, pp. 1321-1322.

Carlo, T.A., and Tewksbury, J.J., 2014. Directness and tempo of avian seed dispersal increases emergence of wild chiltepins in desert grasslands*. J. Ecol*. 102, pp. 248-255.

Carrizo Garcia C., Barfuss, M.H.J., Sehr, E.M., Barboza, G.E., Samuel, R., Moscone, E.A., and Ehrendorfer, F., 2016. Phylogenetic relationships, diversification, and expansion of chili peppers (*Capsicum*, Solanaceae). *Ann. Bot*. 118, pp. 35-51.

Carrizo García et al., 2016. Phylogenetic relationships, diversification and expansion of chili peppers (Capsicum, Solanaceae). *Ann. Bot.*, 118, pp. 35-51

Cheng, J., Zhao, Z., Li., B., Qin, C., Wu, Z., Trejo-Saavedra, D.I., Luo, Z., Cui, J., Rivera-Bustamante, R.F., Li, S., and Hu., J., 2016. A comprehensive characterization of simple sequence repeats in pepper genomes provides valuable resources for marker development in *Capsicum*. *Sci. Rep*. 6, pp. 18919

Choong, C.Y., 1998. *DNA Polymorphisms in the Study of Relationships and Evolution in Capsicum.* Thesis (PhD). University of Reading, UK.

Csilléry, G. 2006. Pepper taxonomy and the botanical description of the species. *Acta Agranomica Hungarica*, 54(2), pp. 151-166.

Gandonou, J.M., and Waliczek. T.M., 2013. An analysis of the recent trends in U.S. chile pepper production, consumption and imports. *J. Food Agr. Environ*. 11, pp. 361-367.

Hall, T., 2005. BioEdit (v.4.0.6) [computer program]. Available from: http://www.mbio.ncsu.edu/BioEdit/bioedit.html [Accessed Nov. 2018].

Hirsch, C.N, Hirsch, C.D., Fletcher, K., Coombs, J., Zarka, D., Van Deynze, A., De Jong, W., Velleux, R.E., Jansky, S., Bethka, P., Douches, D.S., and Buell, C.R., 2013. Retrospective view of North American potato (*Solanum tuberosum* L.) breeding in the 20th and 21st centuries. *G3*, 3(6), pp. 1003-1013.

Ibiza, V.P., Blanca, J., Cañizares, J., and Nuez, F., 2012. Taxonomy and genetic diversity of domesticated Capsicum species in the Andean region. *Genet. Resour. Crop Evol.,* 59, pp. 1077-1088.

IPGRI, AVRDC and CATIE. 1995. Descriptors for *Capsicum* (*Capsicum* spp.). International Plant Genetic Resources Institute, Rome, Italy; the Asian Vegetable Research and Development Center, Taipei, Taiwan, and the Centro Agronómico Enseñanza, Turrialba, Costa Rica.

Jordt, S.E. and Julius, D., 2002. Molecular basis for species-specific sensitivity to “hot” chili peppers. Cell 108(3), pp. 421-430.

Kamvorn, W., Techawongstein, S. Techawongstien, S., and Theerakulpisut, P., 2014. Compatibility of inter-specific crosses between *Capsicum chinense* Jacq. and *Capsicum baccatum* L. at different fertilization stages. *Sci. Hort*. 179, pp. 9-15.

Kantar, M.B., Anderson, J.E., Lucht, S.A., Mercer, K., Bernau, V., Case, K.A., Le, N.C., Frederikse, M.K., DeKeyser, H.C., Wong, Z.Z., Hastings, J.C., Baumler, D.J., 2016. Vitamin variation in Capsicum spp. provides opportunities to improve nutritional value of human diets. *PLoS One,* 11(8)

Katoh K., and Standley D.M., 2013. MAFFT multiple sequence alignment software (v.7): improvements in performance and usability. Mol. Bio. Evol. 30(4), pp. 772-780.

Kraft, K.H., Brown, C.H., Nabhan, G.P., Leudeling, E., Luna Ruiz, J.J., Eeckenbrugge, G.C., Hikmans, R.J., and Gepts, P., 2014. Multiple lines of evidence for the origin of domesticated chili pepper, *Capsicum annuum*, in Mexico. *Proc. Natl. Acad. Sci.* 111, pp. 6165-6170.

Lin, T., et al., 2014. Genomic analyses provide insights into the history of tomato breeding. *Nature*, 46(11), pp. 1220-1226.

Loaiza-Figueroa, F., Ritland, K., Laborde Cancino, J.A., and Tanksley., S.D., 1989. Patterns of genetic variation of the genus *Capsicum* (Solanaceae) in Mexico. *Plant Systematics Evol*. 165, pp. 159-188.

McLeod et al., 1981. An electrophoretic study of evolution in *Capsicum* (Solanaceae). *Evolution,* 7(3), pp. 562-574.

Meyer, D., and Buchta, C., 2018. *Proxy: Distance and Similarity Measures* (R package v.0.4-22) [computer program]. Available from: https://CRAN.R-project.org/package=proxy [Accessed Nov. 2018].

Mongkolporn, D. and Taylor, P.W.J., 2011. *Capsicum*. In: C. Kole (ed.). Wild crop relatives: Genomic and Breeding Resources. Springer, Berlin-Heidelberg.

Nicolaï, M., Canet, M., Lefebvre, V., Sage-Palloix, A.M., Palloix, A., 2013. Genotyping a large collection of pepper (*Capsicum* spp.) with SSR loci brings new evidence for the wild origin of cultivate *C. annuum* and the structuring of genetic diversity by human selection of cultivar types. *Genet. Res.Crop Evol.*, 60, pp., 2375- 2390.

OECD. 2006. Consensus Document on the biology of *Capsicum annuum* complex (chili peppers, hot peppers and sweet peppers). Series on Harmonisation of Regulatory Oversight in Biotechnology No. 36. Organisation for Economic and Co-operation and Development, Paris

Oyama, K., Hernandez-Verdugo, S., Sanchez, C., Gonzalez-Rodrigues, A., Sanchez-Pena, P., Garzon-Tiznado, J.A., and Casas. A., 2006. Genetic structure of wild and domesticated population of *Capsicum annuum* (Solanaceae) from northwestern Mexico analyzed by RAPDs. *Gen. Res. Crop Evol*. 53, pp. 553-562.

Pickersgill, B., 1971. Relationships between weedy and cultivated forms in some species of chili peppers (genus *Capsicum*). *Evolution* 25, pp. 683-691.

Pickersgill, B., 1997. Genetic resources and breeding of *Capsicum* spp. *Euphytica,* 96(1), pp. 129-133.

Qin et al., 2014. Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc. Natl. Acad. Sci.*, 111(14), pp. 5135-5140.

Scaldaferro, M.A., Barboza, G.E., and Christina Acosta, M., 2018. Evolutionary history of the chili pepper *Capsicum baccatum* L. (Solanaceae): domestication in South America

and natural diversification in the seasonally dry tropical forests. *Biol. J. Linnean Soc.*, 124(3), pp. 466-478.

Tayyib, S., 2016. Production share of chillies and peppers, by region. *Food and Agriculture Organization of the United Nations (FAO)*.

Tewksbury, J.J., Nabhan, G.P., Norman, D., Suzan, H., Tuxill, J., and Donovan, J., 1999. *In situ* conservation of wild chiles and their biotic associations. *Conservation Biol.,* 13, pp. 98-107.

Tong, N., and Bosland, P.W., 1999. *Capsicum tovarii*, a new member of the *Capsicum* baccatum complex. *Euphytica* 109, pp. 71–77.

van Zonneveld, M., Ramirez, M., Williams, D.E., Petz, M., Meckelmann, S., Avila, T., Bejarano, C., Rios, L., Pena, K., Jager, M., 2015. Screening genetic resources of Capsicum peppers in their primary center of diversity in Bolivia and Peru. PLoS One 10(9)

Votava, E.J., Nabhan, G.P., and Bosland., P.W., 2002. Genetic diversity and similarity revealed via molecular analysis among and within an *in situ* population and *ex situ* accessions of chiltepin (*Capsicum annuum* var. *glabriusculum*). *Conservation Gen*. 3, pp. 123-129.

Wahyuni, Y., Ballester, A.R., Sundarmonowati, E., Bino, R.J., Bovy, A.G., 2013. Secondary metabolites of *Capsicum* species and their importance in the human diet. *J. Nat. Products*, 76, pp. 783-793.

Walsh, B.M., Hoot, S.B., 2001. Phylogenetic relationships of *Capsicum* (Solanaceae) using DNA sequences from two noncoding regions: The chloroplast atpB-rbcL spacer region and nuclear waxy introns. *Internatl. J. Plant Sci.*, 162(6), pp. 1409-1418.