RESEARCH NOTE

BMC Research Notes

Open Access

Population structure in diverse pepper (*Capsicum* spp.) accessions



Jack McCoy¹, Natalia Martínez-Ainsworth², Vivian Bernau³, Hannah Scheppler¹, Grant Hedblom⁴, Achuyt Adhikari⁵, Anna McCormick⁵, Michael Kantar^{5*}, Leah McHale¹, Lev Jardón-Barbolla², Kristin L. Mercer¹ and David Baumler^{4*}

Abstract

Background Peppers, bell and chile, are a culturally and economically important worldwide. Domesticated *Capsicum spp.* are distributed globally and represent a complex of valuable genetic resources.

Objectives Explore population structure and diversity in a collection of 467 peppers representing eight species, spanning the spectrum from highly domesticated to wild using 22,916 SNP markers distributed across the twelve chromosomes of pepper.

Results These species contained varied levels of genetic diversity, which also varied across chromosomes; the species also differ in the size of genetic bottlenecks they have experienced. We found that levels of diversity negatively correlate to levels of domestication, with the more diverse being the least domesticated.

Keywords Species complex, Admixture, Sweet, Pungent

Objective

Domestication and globalization have significantly affected crop diversity. Domesticated plants have long been of interest to those exploring natural selection [1]. Crop populations are subject to the same evolutionary

*Correspondence:

Michael Kantar mbkantar@hawaii.edu

David Davialar

David Baumler dbaumler@yahoo.com

1 Department of Lortic

¹ Department of Horticulture and Crop Science, The Ohio State University, Columbus, OH, USA

² Centro de Investigaciones Interdisciplinarias en Ciencias Y

Humanidades, Universidad Nacional Autónoma de México, Mexico City, Mexico

³ North Central Region Plant Introduction Station, Agriculture Research Service, United States, Department of Agriculture and Department

of Agronomy, Iowa State University, Ames, IA, USA

⁴ Department of Food Science and Nutrition, University of Minnesota-Twin Cities, St. Paul, MN, USA

⁵ Department of Tropical Plant & Soil Sciences, University of Hawaii at Manoa, Honolulu, HI, USA forces that impact genetic diversity in the wild: gene flow, drift, selection, mutation, and assortative mating [2]. However, in cultivated populations, human management can influence these factors, resulting in a combination of natural and human-mediated evolutionary change. Landraces, domesticates that have not undergone improvement by modern plant breeding methods, present ideal populations in which to explore diversification and genetic structure [3]. Traditional landrace varieties often retain a higher genetic diversity than elite lines [4]. The degree of domestication can also impact genetic diversity, and this is difficult to measure precisely as it is often a continuous rather than discrete process. For instance, in chile peppers (Capsicum spp. L.), four levels of domestication are typically identified: wild, semi-wild, landrace, and commercial peppers [5]. Nevertheless, while genome wide levels of diversity are higher amongst landraces than commercial peppers, there can be stronger (e.g., Chile de Agua) or weaker (e.g., Mirasol) fixation for all the classic domestication syndrome traits within specific lineages



© The Author(s) 2023. **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/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

of landraces and the degree to which they hybridize with wild peppers also varies [6].

Chile pepper (Capsicum spp.) is a culturally and economically valuable vegetable, spice, and medicine worldwide [7]. Peppers are one of the most popular vegetables across the world, and they are consumed raw, cooked, and dried for use as a spice by nearly 25% of the world's population [8, 9]. Peppers are known to have very rich vitamin content, as well as producing a notable heat [10, 11]. There are five domesticated species in genus Capsi*cum* that form a species complex which make up the crop we call pepper [12]. These species are Capsicum annuum, C. chinense, C. frutescens, C. baccatum, and C. pubescens. The chile pepper species complex is cultivated worldwide [12], with C. annuum and C. chinense considered to have been subject to the greatest degree of selection and demographic pressure [13]. The spread of pepper around the world has expanded its traditional uses, leading to new phenotypes and population genetic structure relative to landraces in the center of origin [14]. The present study sampled named chile peppers collected from all over the world representing eight different *Capsicum* species-the five domesticates plus C. chacoense, C. eximium, and C. praetermissum-to assess a broad range of cultivated and wild genetic resources. These peppers sample a wide range of phenotypes, uses, and cultivation histories. Thus, the objective of this study was to explore the population structure and genetic diversity of 467 chile pepper accessions from diverse genetic backgrounds.

Data description

Plant material, culture, and phenotypic evaluation

Pepper accessions were sourced from various seed producers across North America (Additional file 2: Table S1). In total, this study included 467 accessions spanning eight species [*C. annuum* (n=294), *C. baccatum* (n=33), *C. chacoense* (n=2), *C. chinense* (n=119), *C. eximium* (n=3), *C. frutescens* (n=12), *C. praetermissum* (n=1), *C. pubescens* (n=3)]. Two replicate plants of each accession were grown in in an RCBD in greenhouses at the University of Minnesota in summer 2018 in five-gallon containers using a standard potting mix under

fluorescent lights followed by metal halide lighting. For each accession, healthy young leaf tissue was harvested from the healthiest plant for DNA extraction. The soil was fertilized with Pure Blend Pro Grow (3-2-4), Pure Blend Pro Bloom (2-3-5), and Cal-Mag Plus (all from Botanicare, AZ, USA) for a ten-week period following the manufacturer's recommendations.

Sequencing and SNP calling

DNA was extracted from the 467 accessions grown at the University of Minnesota representing eight Capsicum species using Qiagen DNeasy kit (Qiagen Ltd, Germantown, MD, USA) following the manufacturer's instructions. All samples were sent to the University of Minnesota Genomics Center where libraries for doubledigest genotyping-by-sequencing (GBS) [15] were constructed using Apek1 and Btg1 restriction enzymes. The libraries were sequenced using Illumina Novaseq 6000. Fastq files were demultiplexed using Illumina bcl2fastq software. Trimmomatic was used to remove the first 12-bases (adapter sequences) from the beginning of each read [16]. Cleaned reads were aligned to the *C. annuum* reference genome (UCD-10X-F1; a cross between Criollos de Morelos 334 landrace and a non-pungent blocky pepper-breeding line; [17]) using BWA-mem [18]. Variants were called using Freebayes software to jointly call variants across all samples [19]. The initial VCF file was filtered using VCFtools to remove variants with minor allele frequency <1%, variants with genotype rates <95%, and samples with genotype rates < 10%. This generated a total of 22,916 SNPs across the 12 chromosomes (Additional file 1: Fig. S1). All raw sequence data can be found in the NCBI SRA under the project name PRJNA876725 (Table 1).

Population genetics

Principal component analysis (PCA) was conducted with R package SNPRelate using 22,916 markers to visualize the overall distribution of genetic diversity and discern population structure related to species [22]. In addition, a neighbor joining tree was constructed using the r-package fastreeR. Diversity was explored in four common

 Table 1
 Data associated with this manuscript

Label Name of data file File type D	Data repository and identifier
Data set 1 pepper collection .fastq N	NCBI Sequence Read Archive: https://identifiers.org/ bioproject:PRJDB3 [20]
Data set 2 pep_12_chrom.vcf .vcf Z	Zenodo: https://doi.org/10.5281/zenodo.7268809 [21]
Data set 3 Filtered10percentminorallele_2966markershmp Z 77genotypes.hmp	Zenodo: https://doi.org/10.5281/zenodo.7268809 [21]
Data set 4 Vitamin_phenotype.txt .txt Z	Zenodo: https://doi.org/10.5281/zenodo.7268809 [21]

cultivated species (*C. annuum* (n=294), *C. baccatum* (n=33), *C. chinense* (n=119), *C. frutescens* (n=12)) by calculating nucleotide diversity (π) with TASSEL [23] using a sliding window of 100 markers sliding by every individual marker, but not for the four rarer species. Differences between species π related to patterns of demography (i.e., species) were determined by conducting a Dunn multiple comparison test.

Population structure

Results from double-digest GBS of 467 pepper accessions yielded 22,916 polymorphic SNPs. These SNPs were distributed throughout the euchromatic and pericentromeric regions, providing even genome-wide coverage (Additional file 1: Fig. S1). PCA projections revealed clear population structure within and between species. The first two principal components explained a combined 17.75% of the genetic variance (Fig. 1; PC1, 9.29% and PC2, 8.46%) and three major groups were identified.

The first principal component separated *C. baccatum* accessions (Fig. 1-bottom right) from the other species. The second principal component separated *C. annuum* (Fig. 1-bottom left) from *C. chinense* and *C. frutescens* (Fig. 1-top left). Further, the combined PC1 and PC2 shows potential evidence of interspecific hybridization between the commonly cultivated species s (*C. annuum*, *C. baccatum*, *C. chinense*, *C. frutescens*), as seen from individuals located in genetic space between the major cultivated species clusters; however, it is possible that there are other drivers of this pattern. Chiefly, *C. annuum* and *C. baccatum* each formed separate groups, while *C. chinense* and *C. frutescens* formed a distinct group together. Additionally, *C. pubescens* and *C. eximium* formed a small group near *C. baccatum* accessions.

Nucleotide diversity

There were differences in nucleotide diversity (π) between the four major crop species (*C. annuum*, *C. chinense*, *C. baccatum*, and *C. frutescens*, p<0.001 Dunn test). We



Fig. 1 Principal component analysis of 467 accessions representing eight Capsicum species. There are three distinct clusters dominated by C. chinense (top left), C. annuum (bottom left), and C. baccatum (bottom right)

estimated π for *C. annuum* to be 0.0404 ± 0.0000387 (se), *C. chinense* to be 0.0432 ± 0.0000470 (se), *C. baccatum* to be 0.102 ± 0.0000977 (se), and *C. frutescens* to be 0.0745 ± 0.0000756 (se) (Fig. 2). In addition to genome-wide differences (Fig. 2a), there are large chromosome-wide differences in π values between species for individual chromosomes, as well, especially for chromosome 2 (Fig. 2b). Interestingly, sliding window estimates

of π in *C. baccatum* diverge from those of other species most frequently, with π increasing in *C. baccatum* in regions for which π decreases in other species (e.g., Fig. 2c; chromosomes three and six).

Phylogenetic analysis of plant and fruit phenotypes

This study afforded an opportunity for phylogenetic analysis of 467 pepper accessions sourced from all over



Fig. 2 Capsicum nucleotide diversity (π) calculated **A** genome wide by species and **B** chromosome-wide by species. **C** Scans of nucleotide diversity (π) across each chromosome of four species. Box plots show the median, box edges represent the first and third quartiles, and the whiskers extend to farthest data points within the 1.5 × interquartile range outside box edges



Fig. 3 Neighbor joining tree of accessions explored in this manuscript. Colors represent different species assignment based on passport data

the world (Fig. 3). In general, common phenotypes appear to cluster together regardless of where the varieties were developed. For example, accessions with purple color cluster together (Purple Nurple, Fluorescent Purple, Royal Black, Purple Glow in the Dark, and the Black Pearl-Fig. 3). Varieties with variegated leaves clustered together (e.g., Tricolor Variegata, Jigsaw, Rainforest Trifoliage, Calico Hybrid Ornamental, and Var. Black Pearl). The clade containing Shishitou and Hot Shishitou also contain Pepperonia Italian and Pepperoncini Greek accessions, showing relationships between plants with shared phenotypes that were developed in very different geographies. There was also a clade that included many long-fruited varieties (e.g., Big Jim, Monster Chili and Joe E. Parker) indicating that these may also harbor a locus for producing the long fruit phenotype. Further there was a cluster of varieties producing large width and length fruits (e.g., Giant Aconcagua, Cubanelle, Big Sweet Red, and Big Bertha), indicating that while developed in different cultures there may be a similar genetic basis for these commonly selected phenotypes. There are also numerous occurrences of clades that contain clusters of varieties with colored fruits (black, brown, purple, white, yellow, and orange) other than the most common green and red.

Species clustering was consistent with previous population studies in *Capsicum* [24–26]. Nucleotide diversity differed across domesticated species, with evidence of reduced diversity in the more widely cultivated species (*C. annuum* and *C. chinense*). Species

groupings in PCA aligned with expectations based on previous research [14, 27–29], with notable admixture suggestions between *C. annuum*, *C. chinense*, and *C. frutescens* (Fig. 1). However, since species were determined based on historic records, species misidentification, as well as actual interspecific hybridization, may contribute to this pattern. Nevertheless, gene flow between species would be consistent with the interspecific hybridization long used in pepper breeding [30]. For example, *C. chinense*, which readily crosses with *C. annuum*, has been used as a source for novel traits [30].

Like previous work, diversity was found to be lower in the more widely cultivated species, C. annuum and *C. chinense*, than in *C. baccatum* [13]. The reduction in nucleotide diversity for C. annuum and C. chinense may be tied to regions under selection during domestication and improvement. The reduction may also be due to limited numbers of parents being the basis of improvement as peppers became common in new geographies across the world [29]. Species genome-wide level differences in nucleotide diversity did not appear to be driven by specific chromosomes; however, there were notable reductions in diversity in C. annuum and C. chinense on chromosomes with known domestication loci that have previously been identified (e.g., chromosomes two, six, and ten [31]). The different domestic Capsicum species have different patterns of diversity but show a potential signal of interspecific hybridization and show some clustering by shared phenotype.

Limitations

- There is limited information about geographic origin of species
- Limited phenotypic information on samples
- Limited information on origin of cultivar names

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13104-023-06293-3.

Additional file 1: Figure S1. Distribution of 22,916 SNPs across 12 chromosomes from genotyping by sequencing of 467 Capsicum accessions. The numbers 0-19 represent the number of SNPs which fall into each 1,000,000 bp bin across each chromosome.

Additional file 2: Table S1. Accessions and Species, name from seed provider, and species number used in this study.

Acknowledgements

Not applicable.

Author contributions

JM: analysis, writing/revision, NM: writing/revision, VB: writing/revision, HS: writing/revision, GH: data generation, AA: data generation, AM: data generation, MK: analysis, writing/revision, LM: writing/revision, LJ: writing/revision, KM: writing/revision, DB: Analysis, conceptualization, drafting, analysis, writing/revision. All authors read and approved the final manuscript.

Funding

We would like to thank the USDA-AFRI, Physiology of Agricultural Plants section for support under grant number 2017-06351, "Genetic structure and mechanisms of drought adaptation in Capsicum". Additional funding provided by The College of Tropical Agriculture and Human Resources, University of Hawai'i at Mānoa; USDA Cooperative State Research, Education and Extension (CSREES), Grant/Award Number: HAW08039-H. N.M.-A. thanks her postdoctoral fellowship granted by CONACyT, Modalidad 1. 11200/224/2021.

Availability of data and materials

NCBI Sequence Read Archive: https://www.ncbi.nlm.nih.gov/bioproject/ PRJNA876725 [20]. Zenodo: https://doi.org/10.5281/zenodo.7268809 [21]

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests

No competing interests.

Received: 28 November 2022 Accepted: 20 February 2023 Published online: 25 February 2023

References

- Darwin C. The variation of animals and plants under domestication. J. murray. 1868. 2
- 2. Loveless MD, Hamrick JL. Ecological determinants of genetic structure in plant populations. Annu Rev Ecol Syst. 1984;15(1):65–95.

- Zohary D, Hopf M. Domestication of pulses in the old world: legumes were companions of wheat and barley when agriculture began in the Near East. Science. 1973;182(4115):887–94.
- 4. Liu A, Burke JM. Patterns of nucleotide diversity in wild and cultivated sunflower. Genetics. 2006;173:321–30.
- Taitano N, Bernau V, Jardón-Barbolla L, Leckie B, Mazourek M, Mercer K, et al. Genome-wide genotyping of a novel Mexican chile pepper collection illuminates the history of landrace differentiation after capsicum Annuum L. domestication. Evolut Appl. 2019;12(1):78–92.
- Pérez-Martínez AL, Eguiarte LE, Mercer KL, Martínez-Ainsworth NE, McHale L, van der Knaap E, Jardón-Barbolla L. Genetic diversity, gene flow, and differentiation among wild, semiwild, and landrace chile pepper (*Capsicum annuum*) populations in Oaxaca. Mexico Am J Bot. 2022;109(7):1157–76.
- Aguilar-Meléndez A, Vásquez-Dávila MA, Manzanero-Medina GI, Katz E. Chile (Capsicum spp.) as food-medicine continuum in multiethnic Mexico. Foods. 2021;10(10):2502.
- Sherman PW, Billing J. Darwinian gastronomy: why we use spices spices taste good because they are good for us. Bioscience. 1999;49:453–63.
- Kraft KH, Brown CH, Nabhan GP, Luedeling E, Ruiz JDJL, d'Eeckenbrugge GC, Gepts P. Multiple lines of evidence for the origin of domesticated chili pepper, capsicum annuum, in Mexico. Proc Natl Acad Sci. 2014;111(17):6165–70.
- Phillips KM, Ruggio DM, Ashraf-Khorassani M, Haytowitz DB. Difference in folate content of green and red sweet peppers (*Capsicum annuum*) determined by liquid chromatography-mass spectrometry. J Agric Food Chem. 2006;54:9998–10002.
- Wahyuni Y, Ballester AR, Sudarmonowati E, Bino RJ, Bovy AG. Secondary metabolites of capsicum species and their importance in the human diet. J Nat Prod. 2013;76:783–93.
- 12. Bosland PW, Votava EJ, Votava EM. Peppers: vegetable and spice capsicums, vol. 22. Cabi: Wallingford; 2012.
- Pickersgill B, Heiser CB, McNeill J. Numerical taxonomic studies on variation and domestication in some species of Capsicum (No. BOOK). 1979
- Tripodi P, Rabanus-Wallace MT, Barchi L, Kale S, Esposito S, Acquadro A, et al. Global range expansion history of pepper (Capsicum spp.) revealed by over 10,000 genebank accessions. Proc Natl Acad Sci. 2021;118(34):e2104315118.
- Poland JA, Brown PJ, Sorrells ME, Jannink JL. Development of highdensity genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE. 2012;7(2):e32253.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
- Hulse-Kemp AM, Maheshwari S, Stoffel K, Hill TA, Jaffe D, Williams SR, et al. Reference quality assembly of the 35-Gb genome of capsicum annuum from a single linked-read library. Hortic Res. 2018;5(1):1–13.
- Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN]. 2013
- Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907.2012.
- Kantar et al. University of Hawaii. Elucidating the population structure and diversity of publicly available pepper accessions 2022; NCBI Sequence Read Archive. https://identifiers.org/bioproject:PRJDB3.
- McCoy J, Martinez N, Bernau V, Scheppler H, Hedblom G, Adhikari A, Halpin-McCormick A, Kantar M, McHale L, Jardon L, Mercer K, Baumler D. Population structure in diverse pepper (Capsicum spp.) accessions. 2022. Zenodo. https://doi.org/10.5281/zenodo.7268809.
- 22. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012;28(24):3326–8.
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23(19):2633–5.
- Giuliano G, Cao Y, Zhang K, Yu H, Xu D, Chen S, et al. Pepper variome reveals the history and key loci associated with fruit domestication and diversification. Mol Plant. 2022. https://doi.org/10.1016/j.molp.2022.09. 021.
- Chiaiese P, Corrado G, Minutolo M, Barone A, Errico A. Transcriptional regulation of ascorbic acid during fruit ripening in pepper (*Capsicum annuum*) varieties with low and high antioxidants content. Plants. 2019;8(7):206.

- Taitano N, Bernau V, Jardón-Barbolla L, Leckie B, Mazourek M, Mercer K, van der Knaap E. Genome-wide genotyping of a novel mexican chile pepper collection illuminates the history of landrace differentiation after capsicum Annuum L. domestication. Evolut Appl. 2019;12(1):78–92.
- Nicolaï M, Cantet M, Lefebvre V, Sage-Palloix AM, Palloix A. Genotyping a large collection of pepper (Capsicum spp.) with SSR loci brings new evidence for the wild origin of cultivated C. annuum and the structuring of genetic diversity by human selection of cultivar types. Genetic Resour Crop Evolut. 2013;60(8):2375–90.
- Colonna V, D'Agostino N, Garrison E, Albrechtsen A, Meisner J, Facchiano A, Tripodi P. Genomic diversity and novel genome-wide association with fruit morphology in Capsicum, from 746k polymorphic sites. Sci Reports. 2019;9(1):1–14.
- Pereira-Dias L, Vilanova S, Fita A, Prohens J, Rodríguez-Burruezo A. Genetic diversity, population structure, and relationships in a collection of pepper (Capsicum spp.) landraces from the Spanish centre of diversity revealed by genotyping-by-sequencing (GBS). Hortic Res. 2019. https://doi.org/10. 1038/s41438-019-0132-8.
- Pickersgill B. Relationships between weedy and cultivated forms in some species of chili peppers (genus Capsicum). Evolution. 1971;25(4):683–91.
- Paran I, Van Der Knaap E. Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. J Exp Bot. 2007;58(14):3841–52.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

