Published online December 14, 2017

ORIGINAL RESEARCH

Genetic Diversity, Population Structure, and Genetic Correlation with Climatic Variation in Chickpea (*Cicer arietinum*) Landraces from Pakistan

Syed Gul Abbas Shah Sani, Peter L. Chang, Asif Zubair, Noelia Carrasquilla-Garcia, Matilde Cordeiro, Ramachandra Varma Penmetsa, M. Farooq H. Munis, Sergey V. Nuzhdin, Douglas R. Cook,* and Eric J. von Wettberg*

ABSTRACT

Chickpea (*Cicer arietinum* L.) production in arid regions, such as those predominant in Pakistan, faces immense challenges of drought and heat stress. Addressing these challenges is made more difficult by the lack of genetic and phenotypic characterization of available cultivated varieties and breeding materials. Genotyping-by-sequencing offers a rapid and cost-effective means to identify genome-wide nucleotide variation in crop germplasm. When combined with extended crop phenotypes deduced from climatic variation at sites of collection, the data can predict which portions of genetic variation might have roles in climate resilience. Here we use 8113 single nucleotide polymorphism markers to determine genetic variation and compare population structure within a previously uncharacterized collection of 77 landraces and 5 elite cultivars, currently grown in situ on farms throughout the chickpea growing regions of Pakistan. The compiled landraces span a striking aridity gradient into the Thal Desert of the Punjab. Despite low levels of variation across the collection and limited genetic structure, we found some differentiation between accessions from arid, semiarid, irrigated, and coastal areas. In a subset of 232 markers, we found evidence of differentiation along gradients of elevation and isothermality. Our results highlight the utility of exploring large germplasm collections for nucleotide variation associated with environmental extremes, and the use of such data to nominate germplasm accessions with the potential to improve crop drought tolerance and other environmental traits.

Core Ideas

- Cultivated landraces may harbor "genomic gems" for crop improvement.
- Covariation between genomes and environments can prioritize crop accessions for study.
- Genomics, modeling, and genetics can unlock the potential of landraces.
- Pakistan's historical landraces are a source of variation for limited water and high heat.

CULTIVATED CHICKPEA is a small herbaceous annual plant that displays significant phenotypic variation across growing regions on six continents. It was among the founder crops domesticated approximately 10,000 yr ago in Southwest Asia (Zohary, 1976). In prehistoric times, cultivation spread from its origin in Southeastern Anatolia, initially throughout the Fertile Crescent, to South Asia by 6000 yr ago, and to East Africa, notably Ethiopia, by ~3000 yr ago (Maliro et al., 2008). The Spanish introduced chickpeas into the New World, particularly to Mexico,

S.G.A.S. Sani, M.F.H. Munis, Dep. of Plant Sciences, Quaid-i-Azam Univ., Islamabad, Pakistan 45320; S.G.A.S. Sani, P.L. Chang, N. Carrasquilla-Garcia, R.V. Penmetsa, D.R. Cook, Dep. of Plant Pathology, Univ. of California, Davis, CA 95616; P.L. Chang, A. Zubair, M. Cordeiro, S.V. Nuzhdin, Dep. of Biological Sciences, Univ. of Southern California, Los Angeles, CA 90007; S.V. Nuzhdin, Dep. of Applied Mathematics, Peter the Great St. Petersburg Polytechnic Univ., St. Petersburg, Russia 195251; E.J. von Wettberg, Dep. of Biological Sciences and International Center for Tropical Botany, Florida International Univ., Miami, FL 33199; E.J. von Wettberg (current address), Dep. of Plant and Soil Sciences, Univ. of Vermont, Burlington, VT 05405. Received 2 Aug. 2017. Accepted 27 Oct. 2017. *Corresponding author (ebishopv@uvm.edu, drcook@ucdavis.edu).

Abbreviations: Fst, fixation index; SNP, single nucleotide polymorphism.

Plant Genome 11:170067
doi: 10.3835/plantgenome2017.08.0067

Argentina, and Chile. Within the past century, the crop was introduced to Australia and North America. Thus chickpea has shifted geographically to a variety of distinct agro-ecological zones: from its point of origin in Turkish Mesopotamia and throughout the historical Fertile Crescent, where the spring growing season is moderate and the crop matures as temperature increases and moisture decreases in early summer; to tropical and subtropical East Africa, South Asia, and Australia, where chickpea is grown in the post-monsoonal fall and winter months; to North America, where chickpea is cultivated during late spring and summer. Across most of its range, chickpea is grown on residual soil moisture following the rainy season and on marginal lands prone to end-of-season water deficit and other environmental extremes. These environmental challenges, coupled with associated biotic challenges such as fungal disease, are the primary factors that limit chickpea yield. Similar variation in geography and climate are shared across most staple crops, including other legumes [lentil (*Lens culinaris* Medik.), peas (*Pisum sativum* L.)] and cereals [wheat (*Triticum aestivum* L.), barley (Hordeum vulgare L.)] that were also domesticated in the Fertile Crescent (Zohary 1976).

Pakistan is the second leading grower of chickpea after India, accounting for ~7% of global hectares but less than 3% of global production (FAOSTAT, 2015). Chickpea in Pakistan is produced at a rate of only 25% that realized in more intensive agricultural systems (FAOSTAT, 2015), highlighting the need to improve both crop genetics and management. In Pakistan, like the rest of South Asia, there are two popular market classes of chickpea: 'Desi' and 'Kabuli'. The majority (80%) of production of both market classes is in the Punjab province, with important but smaller contributions from the Sindh and Khyber Pakhtunkhaw provinces. Pakistan's Punjab is characterized by limited soil moisture, especially in the excessively hot and dry Thal region, which accounts for ~90% of the Punjab's chickpea production. More generally, throughout Pakistan, chickpea production spans a strong aridity gradient, from relatively wet coastal and irrigated areas to extremely arid, rainfed conditions in the Thal Desert.

Smallholder chickpea farmers in Pakistan still depend heavily on landraces, including mixed seed obtained from a semiformal seed distribution system. By contrast, modern breeding efforts in Pakistan increasingly focus on lineages that originate outside of Pakistan, emphasizing improved yields under well-managed, productive conditions. In addition to the potential for genetic erosion, the trend away from use of local genetic diversity may miss adaptations inherent to local landraces. These traditional genotypes have been grown for millennia on farms, often in marginal areas with low inputs. Early agriculturalists probably selected genotypes that retained some degree of production under such conditions; if so, then landraces from these areas might contain tolerance to drought, heat, and other co-occurring factors. To the extent that landraces harbor useful alleles, their characterization may have utility for crop improvement and their preservation may be a buffer against future genetic erosion.

In recent years (i.e., since 2013–2016), the area devoted to chickpea production in Pakistan has remained relatively constant, although year-to-year production rates in the same regions have fluctuated greatly (Pakistan Bureau of Statistics, 2011). Such yield fluctuations are driven by local environmental conditions, which is typical of dryland agriculture and is a situation that is predicted to become worse as climates warm and precipitation becomes more variable and less predictable. The molecular, genetic and phenotypic characterization of traditional varieties from extreme climates, especially the identification of genes conferring climate resilience, would facilitate the introgression of alleles that are well suited to local conditions in Pakistan into otherwise modern high-yielding backgrounds.

The emergence of genomic-reduction next-generation sequencing approaches (e.g., RAD-SEQ, Baird et al., 2008) coupled with high-density global climate datasets (e.g., Hijmans et al., 2005) provide the power to test hypotheses about climatic factors and the distribution of crop genetic diversity. Traditional statistical approaches such as the Mantel test lack the power to effectively find associations between markers and climate (e.g., Legendre and Fortin, 2010; Guillot and Rousset, 2013; Legendre et al., 2015). More powerful approaches such as BEDASSLE can remove auto-correlation among factors and associate a reduced set of single nucleotide polymorphisms (SNPs) with climatic or other environmental variables (Bradburd and Bradburd, 2013; Bradburd et al., 2013). Similar approaches have been used in crops such as barley, revealing extensive genetic variation that correlates with climatic variation (e.g, Russell et al., 2016). In crops with limited genetic variation, including chickpea, such approaches can serve to prioritize otherwise narrow germplasm for further study and ultimately use in breeding programs.

In the present study, we focus on the assessment of genetic relationships in a panel of chickpea accessions that includes previously uncharacterized landraces from the Thal Desert and more mesic sites across Pakistan, along with selected breeding material and elite cultivars from Pakistan, Iran, and India. We identify and use genome-wide SNPs to examine genetic differentiation among these genotypes and to determine the degree to which segregating variation is correlated with climatic variation at the sites of landrace collection.

Materials and Methods

Germplasm Collection

Chickpea landraces and elite cultivars used in this study were obtained from the Plant Genetic Resource Institute National Agricultural Research Centre, Pakistan; the ICRISAT, India; and the USDA National Genetic Resources Program (accession metadata are listed in Supplemental Table S1; geographic distribution is summarized by climate group and district in Fig.1). Sites of

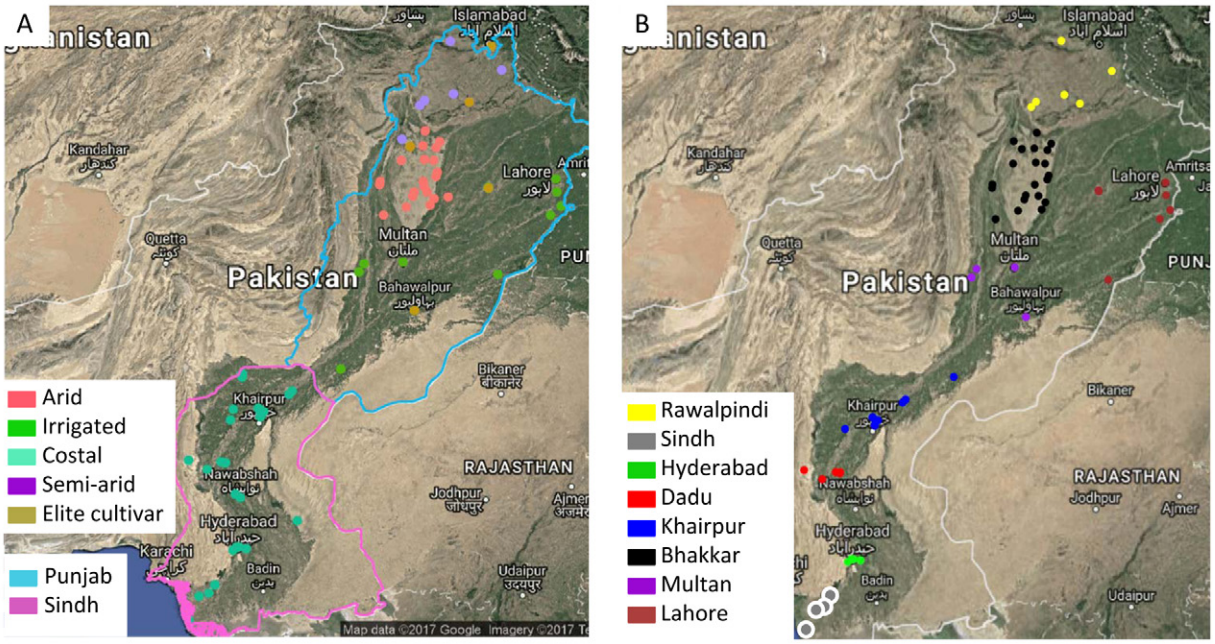

Fig. 1. Geographical distribution of Pakistani landraces labeled according to (A) the assigned climate of origin and (B) according to district of origin.

landrace collection span the major chickpea growing areas of Pakistan, in addition to reference accessions from neighboring India and Iran. Rainfall and topography at the sites of collection, which are major axes of environmental differentiation in Pakistan, were used to assign landraces to one of four groups, whereas modern breeding lines from Pakistan, Iran, and India were assigned arbitrarily (i.e., without any relationship to the environment) to Groups 5 to 7, respectively. Group 1 contains accessions from the Thal Desert, where chickpea is grown as a fall to winter crop under arid conditions on residual soil moisture and often only in years with sufficient rainfall. Group 2 contains accessions with semiarid origins, where cultivation is without irrigation in the Northern Sindh and Punjab (Pothohar region) provinces. Group 3 includes accessions grown with irrigation, typically as a secondary crop in rotation with wheat. Group 4 includes accessions originating in or near coastal areas of the Sindh province, where chickpea is grown under rainfed conditions (~15–18 cm per year) during June to September following the rice (*Oryza sativa* L.) crop. These four groups represent the diversity of cultivation approaches, seasonalities, and growing regions of Pakistan. Accessions from India and Iran were selected from the minicore collection of Upadhyaya and Ortiz (2001) and correspond to the low admixed central Asian and Indian subcontinent clusters described and characterized by Penmetsa et al. (2016).

DNA Extraction, Library Construction, and Sequencing

DNA was extracted from fresh leaf tissues using a DNeasy 96 (Qiagen, Hilden, Germany) plant kit following the manufacturer's protocol. Extractions were performed from a single plant. DNA samples were quantified using Picogreen. Each sample of DNA was digested with 20 units (U) of *Hind*III restriction enzyme, for 60 min at 37°C in a 20-µL reaction, then heat-inactivated at 65°C for 20 min. Digested DNAs were ligated with barcoded adapters (5′-ACACTCTTTCCCTACACGACGCTCTTC-CGATCTxxxxxx and 5′-AGCTyyyyyyyyAGATCG-GAAGAGCGTCGTGTAGGGAAAGAGTGT, where "xxxxxx" and "yyyyyyy" denote the barcode and barcode complementary sequences) with T4 buffer and T4 ligase in a total reaction volume of 50 µL at 20°C for 1 h and then heat-inactivated at 65°C for 10 min. Five µL of each sample was pooled and washed using a PEG8000 cleanup protocol. Pooled samples were digested with *Nla*III at 37°C for 1 h, heat-inactivated at 65°C for 20 min, and then ligated and heat inactivated as above with a common adaptor (5'-AGATCGGAAGAGCGGTTCAG-CAGGAATGCCGAG and 5'-CTCGGCATTCCTGCT-GAACCGCTCTTCCGATCTCATG). Samples were washed with the AMPure XP Bead clean-up reagent (Beckman Coulter, Brea, CA) and enriched by polymerase chain reaction using a biotinylated primer. Biotinylated amplicons were captured with Dynal M Streptavidin beads (Dynabeads, ThermoFisher Scientific, Waltham, MA) following the manufacturer's protocol. An enrichment polymerase chain reaction step (10 cycles) was performed with nonbiotinylated polymerase chain reaction primers. A final fragment size selection clean-up with AMPure XP Beads was performed to minimize low molecular weight primer dimers. The quality, quantity, and reproducibility of libraries were assessed on a Bioanalyzer DNA High Sensitivity chip (Agilent Technologies, Santa Clara, CA). Fragments were sequenced as 100 base reads on an HiSeq 4000 platform (Illumina, San Diego, CA) at the UC Davis Genome Center. All Illumina

data are available from the NCBI under the BioProjects PRJNA353637 and PRJNA396092.

Illumina reads were mapped using the published *C. arietinum* Crop Development Center Frontier genome as a reference (Varshney et al., 2013). Alignment was carried out with Burrows–Wheeler aligner MEM under default mapping parameters (Li and Durbin, 2009). The Genome Analysis Tool Kit pipeline was used to call polymorphisms, following Genome Analysis Tool Kit best practices (McKenna et al., 2010). This pipeline considers indel realignment and base quality score recalibration. The Haplotype Caller program in Genome Analysis Tool Kit calls variants across all the samples simultaneously. Variants were filtered with standard hard filtering parameters (Auwera et al., 2013; DePristo et al., 2011): mapping quality > 37, quality by depth > 24, mapping quality rank sum test < 2. Using these parameters, we identified 8113 SNPs that were further filtered to 1775 SNPs that were called in at least 90% of genotypes for diversity analyses.

Genotype Data Analysis

The STRUCTURE algorithm (Pritchard et al., 2003) was used to assign genetic groups on the basis of allele frequencies. Elite cultivars and landrace accessions from Iran and India were included to provide geographic context around Pakistan. As a complementary approach, principal coordinate analysis was used to further understand relationships among groups. Two accessions ('PK1865D' and 'ICC7819') were removed from further analysis because of their low coverage. Phylogenetic relationships among accessions were determined with MUSCLE for alignment and RAxML for tree construction via maximum likelihood (Stamatakis, 2014).

Genetic diversity within groups on the basis of climatic differences at collection sites (Groups 1–4, as above) was determined with Genalex version 6.53 (Peakall and Smouse, 2006; Smouse and Peakall, 2012) to estimate observed heterozygosity, expected heterozygosity, fixation index (Fst), and the percentage of polymorphism. In a separate analysis, the collection was grouped by market classes (i.e., Desi and Kabuli), which are distinguished on the basis of seed coat color (condensed tannins) and flower color (anthocyanins), with Desi having dark seed coats and colored flowers, with generally smaller seed, and Kabuli having lighter colored seed and white flowers, with generally larger seed. These market classes reflect convergence through breeding (Penmetsa et al., 2016), rather than singular origins of the associated traits. On the basis of these categories, we hierarchically analyzed variation with an analysis of molecular variance, implemented in Genalex version 6.53.

Correlating Genomic Diversity with Geographic Distance and Climatic Variation

Isolation by distance was assessed with the Mantel test performed on a correlation matrix of geographic distance and genetic distance. Geographic distance was the Euclidean distance determined from the collection point of landraces. In total, 9999 permutations were performed with Genalex version 6.53. Because simple correlations of genetic distance and geographic distance are not well suited to estimating the impact of factors that only partially vary with distance [e.g., climate or other ecological features; Diniz-Filho et al., (2013)], we also analyzed the relationship of genetic variation to climatic factors with BEDASSLE (Bradburd and Bradburd, 2013; Bradburd et al., 2013). BEDASSLE can separate the effect of geographic distance on genetic differentiation from underlying environmental differences. We used the β-binomial implementation of the BEDASSLE model, which accounts for overdispersion in the dataset, thus offering a substantially better fit of the model. Given the statistical model for differentiation, BEDASSLE uses the Bayesian Markov chain Monte Carlo (Hastings, 1970) approach to make inferences on model parameters. BEDASSLE has several benefits over Mantel tests in that it can (i) account for autocorrelation of factors, both in molecular markers and climatic factors; (ii) account for nonlinearity in spatial variation in climatic and environmental factors; and (iii) directly measure the relationship between genetic variation and climatic or environmental factors rather than their partial correlation with genetic distance.

BEDASSLE was performed using a subset of 232 noncorrelated SNPs to find patterns potentially obscured in the larger dataset. Reducing the number of SNPs is a necessary step in BEDASSLE analyses, because linkage among markers would otherwise confound the analysis. Noncorrelated SNP were identified following the guidelines in BEDASSLE (Bradburd et al., 2013) and as recommended by Coop et al. (2010), using the pattern of covariance in allele frequencies among populations as a null model to test the correlation among individual SNPs.

Climate data were obtained separately for each landrace collection site in an attempt to correct for the possibility that grouping landraces into large classes (arid, semiarid, irrigated, and coastal) would obscure spatial signal (Supplemental Fig. S1). Information about ecological variables (altitude, temperature, and precipitation levels) was analyzed using R's (R Development Core Team, 2013) raster package (Hijmans et al., 2016). This package uses data from the WorldClim database (Hijmans et al., 2005; http://worldclim.org/version1, accessed 8 Nov. 2017), which has average monthly climate data for minimum, mean, and maximum temperature and for temperature from 1960 to 1990. For altitude, the direct measure was used. For temperature and precipitation, we computed the mean temperature during the chickpea vegetative growing season (October–January) and also the 19 bioclimatic variables with R's dismo (Hijmans et al., 2017) package. For the computed variables, we then chose mean temperature, isothermality (BIO3, mean diurnal temperature range over annual temperature range), and the CV of seasonal precipitation (BIO15)(Supplemental Fig. S2) because these showed higher intergroup variation and lower intragroup variation. We used the highest resolution datasets available (30 arc-seconds, ~1 km). For all computed variables, we

Table 1. Genetic diversity in Pakistani landraces.

Pop		*N*	Na†	Ne	I	Ho	He	uHe	F	%P
Arid	Mean	20.889	1.104	1.028	0.031	0.009	0.019	0.019	0.351	10.37%
	SE	0.004	0.003	0.001	0.001	0.000	0.001	0.001	0.005	
Semiarid	Mean	6.949	1.058	1.029	0.028	0.005	0.018	0.019	0.717	5.77%
	SE	0.003	0.003	0.001	0.001	0.000	0.001	0.001	0.005	
Irrigated	Mean	13.938	1.092	1.037	0.037	0.008	0.023	0.024	0.581	9.18%
	SE	0.003	0.003	0.002	0.001	0.000	0.001	0.001	0.005	
Coastal	Mean	33.789	1.121	1.028	0.031	0.006	0.019	0.019	0.416	12.10%
	SE	0.007	0.004	0.001	0.001	0.000	0.001	0.001	0.005	
Pakistani elite	Mean	5.987	1.060	1.038	0.032	0.006	0.022	0.024	0.615	6.00%
	SE	0.001	0.003	0.002	0.001	0.001	0.001	0.001	0.006	
India	Mean	12.943	1.249	1.055	0.068	0.021	0.038	0.040	0.207	24.81%
	SE	0.003	0.005	0.002	0.002	0.001	0.001	0.001	0.005	
Iran	Mean	8.961	1.233	1.060	0.072	0.026	0.042	0.044	0.167	23.21%
	SE	0.003	0.005	0.002	0.002	0.001	0.001	0.001	0.005	
Desi	Mean	42.747	1.145	1.030	0.033	0.007	0.019	0.020	0.324	14.48%
	SE	0.008	0.004	0.001	0.001	0.000	0.001	0.001	0.005	
Kabuli	Mean	32.817	1.102	1.034	0.033	0.007	0.021	0.021	0.501	10.19%
	SE	0.007	0.003	0.002	0.001	0.000	0.001	0.001	0.005	
Grand mean and SE over loci and groups										
Total	Mean	14.779	1.131	1.039	0.043	0.012	0.026	0.027	0.337	13.06%
	SE	0.038	0.001	0.001	0.001	0.000	0.000	0.000	0.002	2.96%

† Na, number of different alleles; Ne, number of effective alleles = $1 \div (\Sigma \pi^2)$; I, Shannon's information index = $-1 \times \Sigma[\pi \times \log(\pi)]$; Ho, observed heterozygosity = number of heterozygotes ÷ *N*; He, expected heterozygosity = $1 - \Sigma\pi^2$; uHe, unbiased expected heterozygosity = $[2N \div (2N - 1)] \times$ He; F, fixation index = (He − Ho) ÷ He = 1 − (Ho ÷ He); %P, percent of loci that are polymorphic.

restricted ourselves to the months from October to January, which is the vegetative growing season for chickpea in Pakistan.

Results

Marker Polymorphism and Diversity Analysis

Based on sequencing of restriction site-associated DNA tags, we detected 8113 SNPs in the Pakistani collection. For diversity analyses, the set was thinned to 1775 well represented SNPs called in at least 90% of genotypes. Accessions from Iran and India had higher gene diversity (~0.04 vs. 0.02) and percentage of polymorphism (~24 vs. 5–12%) than the Pakistani landraces (Table 1). Within the Pakistani landraces, we found limited differentiation between agro-ecological zones (arid, semiarid, irrigated, and coastal, as well among the two market classes) (Table 2). Gene diversity ranged from 0.018 in the semiarid accessions to 0.23 in the accessions from irrigated areas, with overlapping variances, and percentage of polymorphic loci varied from 5.77% in the semiarid accessions to 12.1% in the coastal accessions. Desi and Kabuli market classes had essentially identical gene diversity (Desi, 0.019; Kabuli, 0.02), although Desi had a slightly higher

Table 2. Differences in fixation index (Fst) among agro-ecological groups and market classes. Fst was calculated following Weir and Hill (2002) implemented in Genalex version 6.503 (Smouse and Peakall, 2012).

	Arid	Semiarid	Irrigated	Coastal	Elite breeding	India	Iran
Arid	0.000	–	–	–	–	–	–
Semiarid	0.089	0.000	–	–	–	–	–
Irrigated	0.012	0.043	0.000	–	–	–	–
Coastal	0.027	0.114	0.019	0.000	–	–	–
Elite breeding	0.070	0.059	0.025	0.100	0.000	–	–
India	0.076	0.072	0.033	0.073	0.035	0.000	–
Iran	0.132	0.100	0.096	0.124	0.093	0.064	0.000
		Desi				Kabuli	
Desi		0.000				–	
Kabuli		0.049				0.000	

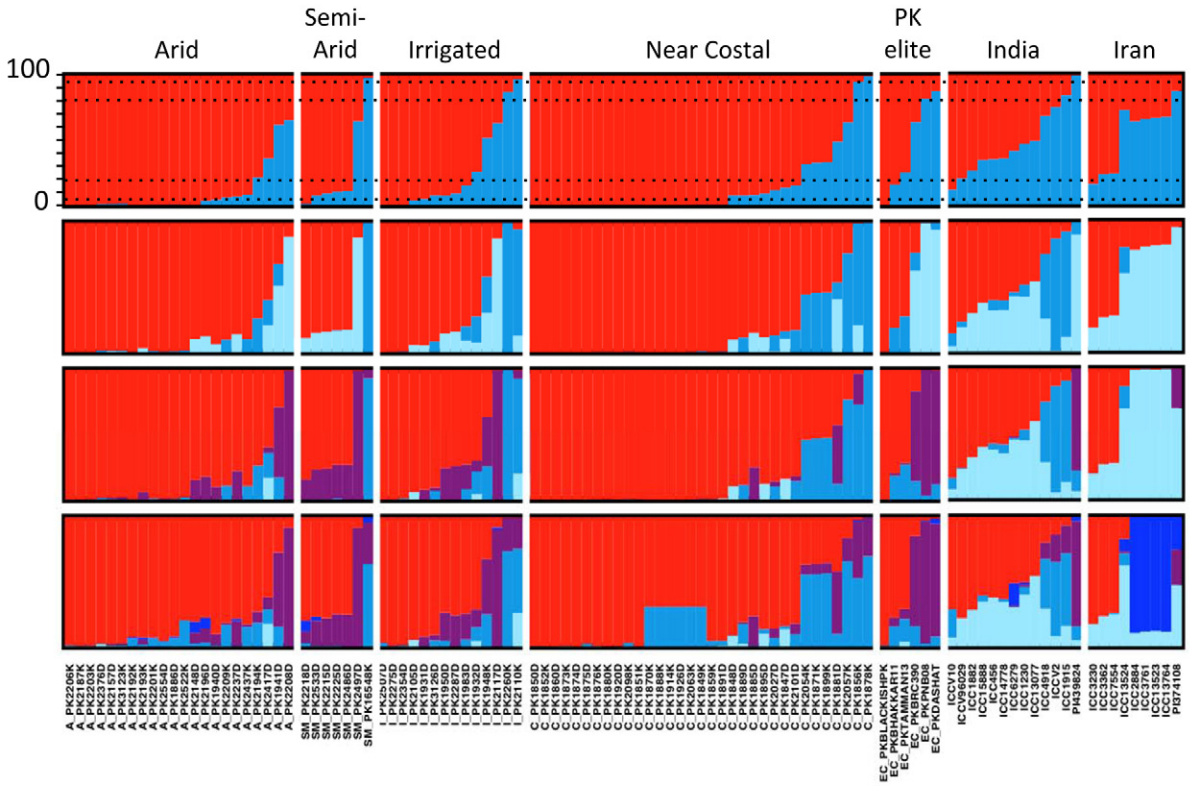

Fig. 2. STRUCTURE plot assessing the relationship among germplasm accessions. The 21 Iranian and Indian lines were selected from the chickpea core collection to represent diversity in regions neighboring Pakistan. *K* = 2 best describes the variation among the accessions in both sets of accession (neighbors), but we also show *K* = 3 to *K* = 5.

rate of polymorphic loci than Kabuli (14.5 vs. 10.2%; Table 2). In general, diversity was lower across the entire collection than diversity in global landrace collections measured by other means (e.g., Varshney et al., 2013; Penmetsa et al., 2016) or in the wild ancestor of chickpea, where the same marker type can have percent polymorphism above 80% and gene diversity estimates of 0.12 (von Wettberg et al., unpublished data, 2017).

We used several complementary approaches to visualize patterns of variation among landraces and found low levels of subdivision. With STRUCTURE, a *K* of 2 best fitted the data (Evanno et al., 2005)(Fig. 2 and Supplemental Fig. S3). At higher *K* values, the group of Pakistani landraces shown in blue in Fig. 2 becomes increasingly subdivided, largely on the basis of their inferred shared ancestry with accessions from Iran, India, and modern Pakistani elite cultivars. One interpretation is that there is considerable shared introgression among Pakistan, Iran, and India, but only for the group shown in light blue in Fig. 2, although the group shown in red in Fig. 2 remains largely intact as a genetic unit unique to Pakistan. In the STRUCTURE analysis, we found limited differentiation among Pakistani versus Iranian and Indian material or among agro-ecological zones within Pakistan, and very low differentiation among the Desi and Kabuli market classes, with the exception that Group 1 individuals (shown in red in Fig. 2) were unique to the Pakistani material. A principal coordinate analysis revealed similar patterns, with no divergence among agro-ecological zones or market classes (Fig. 3a,b). As expected, the Principal Component 1 axis resolved the genetic groups inferred by STRUCTURE (Fig. 3c). Phylogenetic analysis revealed numerous well-supported clades that contain the majority of landraces (Fig. 4). These clades effectively subdivide groups assigned by STURCTURE into largely geographically coherent subsets. Lack of support in deep branches of the tree precludes any inference of the relationships among clades. Analysis of molecular variation attributed 93% of variation to that occurring within agro-ecological zones, and only 7% between them. Moreover, we found very low estimates of Fst among different market classes and among the four broad agro-ecological or climatic zones (Table 2).

Isolation by Distance and Correlation of Genetic Variation with Climatic Variation

Despite the absence of measurable differentiation among climate groups or market types, we did detect a weak but significant pattern of isolation by distance within South Asian chickpea landraces when we included the Iranian and Indian accessions (Fig. 5). This pattern derives largely from the distinct genotypes seen in the Indian and Iranian genotypes at higher levels of *K* in the structure analysis (Fig. 2). Although the slope of the regression was significantly different from zero ($p = 0.01$), the R^2 value was only 0.079. Without the Indian and Iranian landraces, we found an even lower R^2 value of only 0.003, which was not significantly different from zero and is consistent with the highly homogenous nature of the

Pakistani landrace collection. Consequently, geographic isolation may not be the most important axis of genetic differentiation within the whole of Pakistan across a large number of genome-wide SNPs.

To further explore genetic differentiation among these groups, we plotted the heatmap of the Fst estimator (Supplemental Fig. S4), computed according to Weir and Hill (2002). The Fst heatmap shows a clear clustering based on collection site districts: the populations from Punjab province districts cluster separately from the Sindh districts populations, whereas the Rawalpindi population is closer to the Punjab cluster but has the highest Fst distance when compared pairwise to the other populations. These relationships could derive from the extremes of altitude and temperature that are observed among collection sites.

To further investigate the relative impact of ecological factors compared to geographic distance in these populations, we ran BEDASSLE analysis with 40 million generations, using a burn-in of 45% and sampling every 2500th generation. As expected, altitude was negatively correlated with mean temperature ($r^2 = -0.827$) (Supplemental Fig. S1). However, the correlation varies when broken down by population. Rawalpindi has the highest average altitude and consequently also has the lowest mean temperature. The step size for the parameters of the effect size of geographic distance (aD), the effect size(s) of ecological distance(s) (aE), the parameter controlling the shape of the decay in covariance with distance (a2), the Watterson estimator for describing the genetic diversity in a population (θ), the per-generation mutation rate (μ), and a measure of population differentiation related to Fst (ϕ) were adjusted to have an acceptance rate between 20 and 60% as recommended by Bradburd et al. (2013). To test the goodness of the inferred model, we plotted the pairwise Fst from the posterior predictive samples alongside the observed pairwise Fst for the populations (Supplemental Fig. S5). The output from the model tracked well with the observed Fst distribution (Supplemental Fig. S4). Finally, we computed the mean and SE for the ratio of the coefficients of each ecological factor compared with the coefficient for geographic distance via R's coda package (Plummer et al., 2006) (Supplemental Fig. S6). The strongest ratio was seen for altitude (elevation) and the weakest effect was caused by seasonal precipitation. This followed from our observation that precipitation was the least variable ecological factor among the populations, whereas altitude was the most variable factor.

Discussion

We found generally low levels of polymorphism in the collection of Pakistani landraces, consistent with many previous studies that have found low genetic variation in chickpea germplasm (Kazan et al., 1993; Abbo et al., 2003; Roorkiwal et al., 2014; Penmetsa et al., 2016). We also found limited differentiation between the two market classes of chickpea within Pakistan, in agreement with recent work showing that the polyphyletic origin

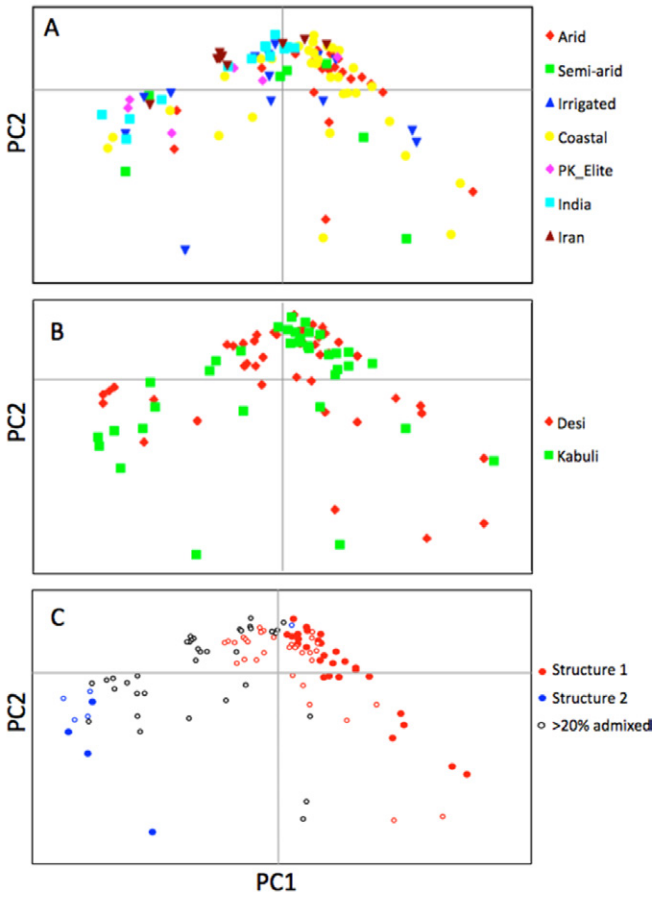

Fig. 3. Principal coordinate analysis (PCoA) depicting relationships among the accessions. (A) PCoA by agro-ecological origin of germplasm. (B) PCoA separated by market class (Desi or Kabuli). (C) PCoA separated by group membership in STRUCTURE at $K = 2$. Filled circles have >95% membership; open colored circles have 80 to 95% membership; black open circles are admixed at rates of >20 to <80%.

of alleles of a single locus responsible for distinguishing the two market classes, and we found a recent history of selective breeding that creates genome-wide differentiation not reflective of common ancestry (Penmetsa et al., 2016). Similarly, we also found limited differentiation among climatic zones within Pakistan. We did, however, find genetic differentiation among genotype groups based on region, a result supported both by phylogenetic analysis and Fst values among regionally grouped accessions. Moreover, in a subset of 232 loci, we found greater cosegregation of genetic variation with isothermality, a factor that is one aspect of the aridity gradient that occurs across chickpea production areas in Pakistan. Thus although the genome-wide signal of differentiation was low, we identified significant correlations between genetic, geographic, and climatic factors.

Potential to Uncover Drought-Adapted Traits from Pakistani Landraces

Although chickpea is a crop with low genetic diversity, it is hardly monomorphic. The observation of a significant

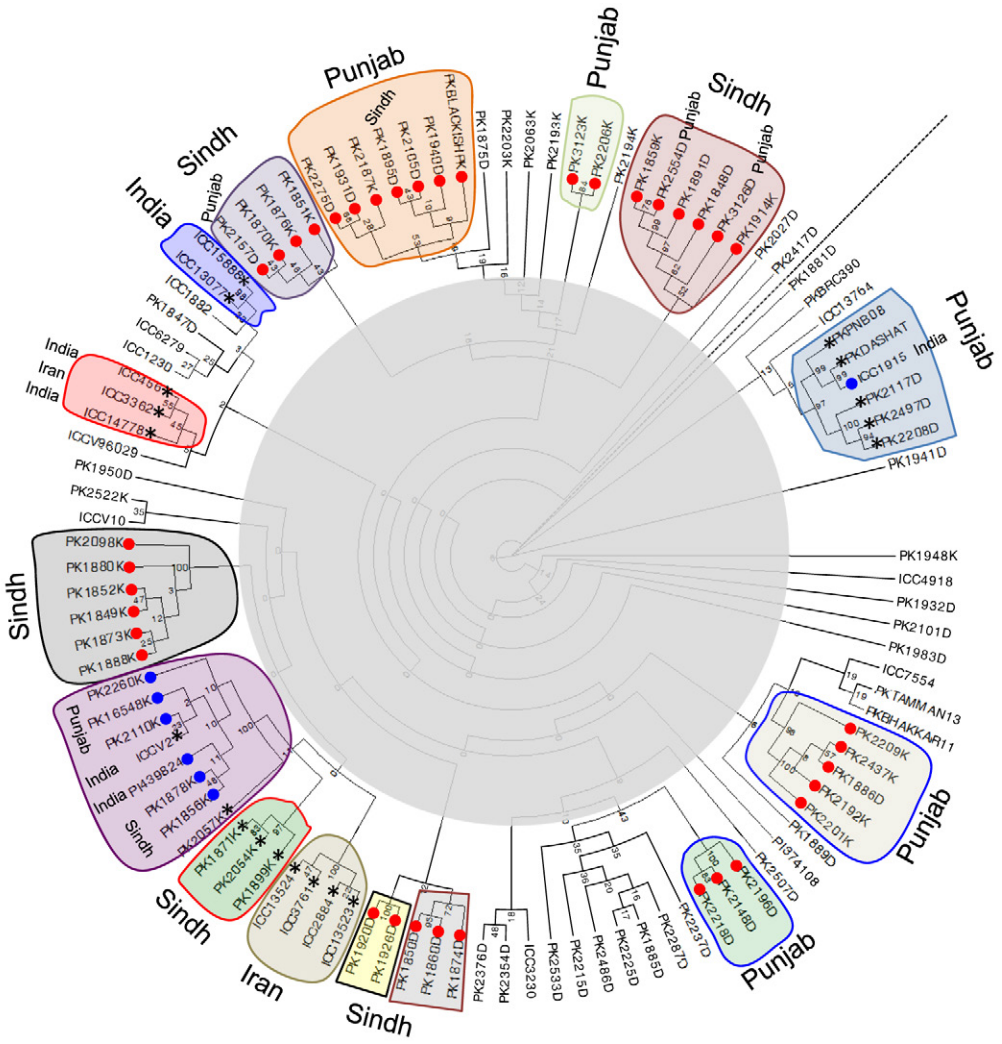

Fig. 4. RAxML phylogenetic tree depicting relationships among accessions. Most accessions occur in clades with marginal to strong support, grouped by colored shapes. Gray shading masks the unsupported deep roots of the tree. Country or province of origin is indicated. Accessions that occur in supported clades are marked according to STRUCTURE group: Group 1 (red), Group 2 (blue), or admixtures of >20 to <80%.

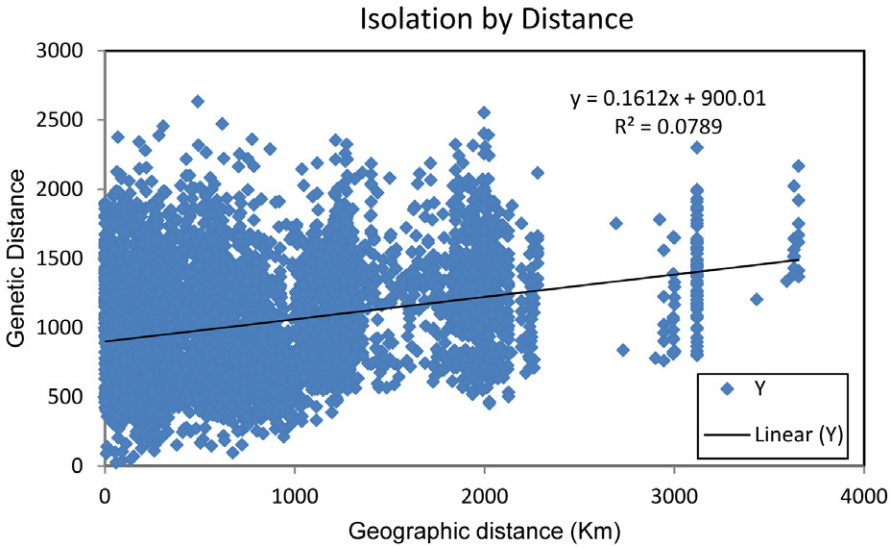

Fig. 5. Isolation by geographic distance in single nucleotide polymorphisms (SNPs) across the sampled agro-ecological zones in Pakistan. $r = 0.28$; $P = 0.10$; $R^2 = 0.08$.

relationship between allelic content and key components of climate, altitude, and isothermality indicates that the strong altitudinal and aridity gradient across Pakistani chickpea production zones may structure the existing genetic variation, despite that fact that the overall level of genetic variation is low. Breeders in search of material adapted to drought and heat may find landraces from these locations to be useful sources of tolerance traits. Our findings complement approaches such as the Focused Identification of Germplasm Strategy, which use climate correlations in the absence of genetic information to mine for sources of adaptive variation in underused germplasm (Khazaei et al., 2013). Tying these approaches to assessments of genetic variation and formally testing them with emerging tests such as BEDASSLE will strengthen our pursuit of traits that may increase or stabilize yields in marginal settings.

Despite such tantalizing correlations, more work is needed to clarify whether the climatic–phenotypic correlations in germplasm collections are both consistent and meaningful. For example, although Focused Identification of Germplasm Strategy may be widely used to find accessions from more arid regions for breeding for drought tolerance (e.g., Khazaei et al., 2013), whether such correlations have any functional relevance and which genes or genome regions confer useful traits is often unknown. Moreover, one must test the possibility of tradeoffs; for example, accessions with desirable qualities in more arid regions may not be desirable when grown under irrigated conditions, or if shifts in phenology move growth into more mesic growing periods during cooler, rainier seasons. A combination of genomic data and more detailed phenotyping would improve these methods, uncovering stronger climatic–phenotypic correlations. Furthermore, for the subset of accessions from the most arid regions that we have identified here, whole-genome analysis coupled with detailed phenotyping is warranted to understand the genetic basis of responses to water shortage in these genotypes.

Our analysis also underscores the observation that temperature and precipitation averages alone may obscure important aspects of variation. Isothermality, a measure that incorporates diurnal and yearly temperature variation, can be a key component of aridity and elevational gradients. As chickpea in Pakistan is grown primarily during the post-monsoon season, its capacity to handle daily extremes may be more important. A lack of both heat and cold tolerance can constrain chickpea production, and breeding for the Thal Desert of Pakistan, where temperature extremes can be pronounced, needs to prioritize both factors.

The Utility of Genomic Reduction: Hunting for Genomic Gems that Correspond to Climatic Gradients

In chickpea, the low genetic diversity of cultivated germplasm limits our capacity to breed for abiotic stress tolerance and pest and pathogen resistance. It took the creation of a wild-cultivated cross to create the first molecular map for the crop (Kazan et al., 1993). Likely multiple population bottlenecks in the wild taxa and the crop, coupled with a predominantly selfing mating system have led to this low diversity [one scenario is proposed by Abbo et al. (2003)]. This is distinct from other Southwestern Asian crops, such as the wind-pollinated cereals barley and wheat, where relatively high levels of diversity segregate in cultivated material, despite a domestication bottleneck (e.g., Poets et al., 2015). Genomic techniques such as genotyping-by-sequencing are essential for uncovering what diversity is segregated in international chickpea germplasm collections. Despite the generally low sequence diversity, chickpea germplasm harbors variation in important agronomic traits, from abiotic stress tolerance and disease resistance, to nutritional qualities such as vitamin and mineral content, to the two phenotypically distinct market classes (Desi and Kabuli). Genomic reduction approaches provide a desirable combination of high power and low cost, and are especially effective for population genetic and phylogenetic analyses that attempt to correlate genetic structure with traits. Genetic variation within landraces, which is significantly reduced in elite germplasm (Chang et al., unpublished data, 2017), may yet prove to be the sort of "genomic gems" that can increase disease resistance or abiotic stress tolerance.

The power of approaches such as BEDASSLE to uncover relationships between genetic diversity and environmental factors depends on having a sufficient number of noncorrelated markers. In selfing crops of low genetic diversity like chickpea, genetic linkage can extend to several kilobases, which increases correlation among markers, thereby reducing the marker set available for analysis as well as the precision of deduced associations. The inherently low genetic diversity of the Pakistani landraces analyzed here is an extreme example of this limitation. Thus among 8113 SNPs, we filtered to 232 noncorrelated SNPs. This number of SNPs was sufficient to test correlation of genetic variation with climatic factors and to nominate germplasm accessions for further analysis, but did not have sufficient precision to confidently identify associated genome intervals.

Breeding for climate resilient chickpea in Pakistan is likely to benefit from our improved understanding of the genetic diversity inherent to Pakistani germplasm compared with worldwide collections of this important legume crop. Despite low genetic diversity, we find association with geographical regions and climatic factors, raising the possibility of selection for regionally specific climate traits. Thus, although our findings are in agreement with findings by Roorkiwal et al. (2014), who reported low level of genetic diversity within *C. arietinum* compared with the wild species with primarily diversity array technology markers, our findings also raise the possibility that reassortment of even low amounts of genetic variation may be an important factor in agro-climatic adaptation.

Molecular genetic and genomic studies, such as the one conducted here, must be combined with precision phenotyping to determine if locally specific gene pools harbor useful traits, even if genetic diversity is low. Our observation that a small fraction of generally low genetic variation associated with climate provides a molecular correlation with this argument. Landraces of chickpea have longstanding cultivation histories in Pakistan, under some of the most adverse (heat and drought) environmental conditions under which chickpea is cultivated. However, because landraces are a product of human history, recent human migration events cannot be entirely discounted. Human movement of seeds prior to collection of landraces may obscure patterns of gene–environment correlations and local environmental adaptations. It is also the case that the precise boundaries of landrace distribution are unknown and thus point location estimates of ecological factors may not be representative. Thus the value of methods like those used here will only be understood in the light of subsequent verification through phenotyping and breeding. Determining whether such materials harbor undiscovered traits or extreme trait values to combat the challenges of modern agriculture—in this case, climate resilience—is a critical endeavor.

Conclusion

Since the pioneering work of Vavilov (e.g., Vavilov, 1926), national germplasm collections have been a key agricultural resource. Although many of the larger national and international collections are widely available, many smaller national collections are not in wide circulation or available to the international community. When these collections contain material from ecologically diverse regions, or from agro-ecosystems with unique characteristics, they are likely to harbor unique genetic variants and adaptations. There is a need for continued efforts to conserve and characterize this material with emerging methods. Along with assessments of this germplasm, the institutions themselves need to be strengthened to ensure the long-term protection and utility of this germplasm.

Supplemental Information

Supplemental Table S1 List of chickpea accessions. PGRI, Plant Genetic Resource Institute National Agricultural Research Centre, Pakistan; ICRISAT, Institute for Crops Research in the Semi-Arid Tropics; USDA NPGS, United States Department of Agriculture, National Plant Germplasm System.

Supplemental Fig. S1. Pairwise scatterplots, correlations and distributions of ecological factors including mean temperature, isothermality, CV of seasonal precipitation (ppSeasonCV), and altitude for different field sites grouped by population. The fifth column and row show the distribution of the climatic variables as a boxplot and histogram respectively. The upper diagonal shows pairwise correlation values for different environmental variables, and the lower diagonal shows pairwise scatterplots. Density plots broken up by populations are on the diagonal. Sample sizes for each population are shown in lowermost-rightmost cell.

Supplemental Fig. S2. Boxplot for distribution of median values of ecological factors across populations. ppSeasonCV, CV of seasonal precipitation.

Supplemental Fig. S3. Mean of the estimated LN probability of observing data at different estimates of population number (K), performed in STRUCTURE Harvester, following Evanno et al., 2005.

Supplemental Fig. S4. Heatmap for pair-wise Fst across populations.

Supplemental Fig. S5. Posterior predictive sampling with 5000 simulated datasets. Black dots show simulated pairwise Fst. Red dots represent the observed pair-wise Fst.

Supplemental Fig. S6. Mean and SE of the ratio of coefficients for each ecological factor (aE).

Conflict of Interest Disclosure

The authors declare that there is no conflict of interest.

Acknowledgments

The authors thank Emily Warschefsky, Ezgi Ogutcen, and Nina Noujdina for helpful discussions, and the Pakistan Genetic Resource Institute National Agricultural Research Center, Islamabad, Pakistan, for provision of biological materials. This work was supported by grants from the US National Science Foundation Plant Genome Program NSF-PGRP 1339346 to D.R.C., E.J.B.v.W., and R.V.P; by a cooperative agreement from the United States Agency for International Development under the Feed the Future Program AID-OAA-A-14-00008 to D.R.C., E.J.B.v.W., S.VN., and R.V.P.; by the Higher Education Commission of Pakistan; an International Research Support Initiative Program fellowship to S.G.A.S.S.; and by the Russian Science Foundation, project No. 16-16-00007 to S.V.N (bioclimatic analysis).

References

Abbo, S., J. Berger, and N.C. Turner. 2003. Viewpoint: Evolution of cultivated chickpea: Four bottlenecks limit diversity and constrain adaptation. Funct. Plant Biol. 30:1081–1087. doi:10.1071/FP03084

Auwera, G.A., M.O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, et al. 2013. From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Curr. Protoc. Bioinformatics 43:11.10.1–11.10.33. doi:10.1002/0471250953.bi1110s43

Baird, N.A., P.D. Etter, T.S. Atwood, M.C. Currey, A.L. Shiver, Z.A. Lewis, et al. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3(10):e3376.

Bradburd, G., and M.G. Bradburd. 2013. Package 'BEDASSLE'. Comprehensive R Archive Network. R Foundation for Statistical Computing. https://cran.r-project.org/ (accessed 9 Nov. 2017).

Bradburd, G.S., P.L. Ralph, and G.M. Coop. 2013. Disentangling the effects of geographic and ecological isolation on genetic differentiation. Evolution 67:3258–3273. doi:10.1111/evo.12193

Coop, G., D. Witonsky, A. Di Rienzo, and J.K. Pritchard. 2010. Using environmental correlations to identify loci underlying local adaptation. Genetics 185(4):1411–1423. doi:10.1534/genetics.110.114819

DePristo, M.A., E. Banks, R. Poplin, K.V. Garimella, J.R. Maguire, C. Hartl, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43:491–498. doi:10.1038/ng.806

Diniz-Filho, J.A.F., T.N. Soares, J.S. Lima, R. Dobrovolski, V.L. Landeiro, M.P.D.C. Telles, et al. 2013. Mantel test in population genetics. Genet. Mol. Biol. 36(4):475–485. doi:10.1590/S1415-47572013000400002

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A

simulation study. Mol. Ecol. 14:2611–2620. doi:10.1111/j.1365-294X.2005.02553.x
FAOSTAT. 2015. FAO statistical databases. Food and Agricultural Organization. http://faostat.fao.org/ (accessed 9 Nov. 2017).
Guillot, G., and F. Rousset. 2013. Dismantling the Mantel tests. Methods Ecol. Evol. 4(4):336–344. doi:10.1111/2041-210x.12018
Hastings, W. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109. doi:10.1093/biomet/57.1.97
Hijmans, R.J., S.E. Cameron, and J.L. Parra. 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 35:1965–1978.
Hijmans, R.J., J. van Etten, and J. Cheng. 2016. Package "raster". R Foundation for Statistical Computing. https://cran.r-project.org/package=raster (accessed 9 Nov. 2017).
Hijmans, R.J., S. Phillips, J. Leathwick, and J. Elith. 2017. Package 'dismo'. R Foundation for Statistical Computing. http://r.adu.org.za/web/packages/dismo/dismo.pdf (accessed 9 Nov. 2017).
Kazan, K.M.F.J., F.J. Muehlbauer, N.E. Weeden, and G. Ladizinsky. 1993. Inheritance and linkage relationships of morphological and isozyme loci in chickpea (*Cicer arietinum* L.). Theor. Appl. Genet. 86:417–426. doi:10.1007/BF00838556
Khazaei, H., K. Street, A. Bari, M. Mackay, and F.L. Stoddard. 2013. The FIGS (Focused Identification of Germplasm Strategy) approach identifies traits related to drought adaptation in *Vicia faba* genetic resources. PLoS One 8:e63107. doi:10.1371/journal.pone.0063107
Legendre, P. and M.J. Fortin. 2010. Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Mol. Ecol. Resour. 10(5):831–844. doi:10.1111/j.1755-0998.2010.02866.x
Legendre, P., M.J. Fortin, and D. Borcard. 2015. Should the Mantel test be used in spatial analysis? Methods Ecol. Evol. 6(11):1239–1247. doi:10.1111/2041-210X.12425
Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760. doi:10.1093/bioinformatics/btp324
Maliro, M.F., D. McNeil, B. Redden, J.F. Kollmorgen, and C. Pittock. 2008. Sampling strategies and screening of chickpea (*Cicer arietinum* L.) germplasm for salt tolerance. Genet. Resour. Crop Evol. 55:53–63. doi:10.1007/s10722-007-9214-9
McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303. doi:10.1101/gr.107524.110
Pakistan Bureau of Statistics. 2011. Agriculture Statistics of Pakistan 2010–11. Pakistan Bureau of Statistics. http://www.pbs.gov.pk/content/agriculture-statistics-pakistan-2010-11 (accessed 17 Nov. 2017).
Peakall, R., and P.E. Smouse. 2006. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6:288–295. doi:10.1111/j.1471-8286.2005.01155.x
Penmetsa, R.V., N. Carrasquilla-Garcia, E.M. Bergmann, L. Vance, B. Castro, M.T. Kassa, et al. 2016. Allelic variation at a domestication-related transcription factor and the multiple origins of the kabuli chickpea (*Cicer arietinum*). New Phytol. 211:1440–1451.
Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: Convergence diagnosis and output analysis for MCMC. R News 6:7–11.
Poets, A.M., Z. Fang, M.T. Clegg, and P.L. Morrell. 2015. Barley landraces are characterized by geographically heterogeneous genomic origins. Genome Biol. 16(1):173. doi:10.1186/s13059-015-0712-3
Pritchard, J. K., Wen, W., and Falush, D. 2003. Documentation for STRUCTURE software: Version 2. Stanford University. https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4/structure_doc.pdf (accessed 10 Nov. 2017).
R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing. http://www.R-project.org (accessed 9 Nov. 2017).
Roorkiwal, M., E.J. von Wettberg, H.D. Upadhyaya, E.J. Warschefsky, A. Rathore, and R.V. Varshney. 2014. Exploring germplasm diversity to understand the domestication process in *Cicer* spp. using SNP and DArT markers. PLoS One 9:E102016. doi:10.1371/journal.pone.0102016
Russell, J., M. Mascher, I.K. Dawson, S. Kyriakidis, C. Calixto, F. Freund, et al. 2016. Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. Nat. Genet. 48(9):1024–1030.
Smouse, P.E., and R. Peakall. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28:2537–2539. doi:10.1093/bioinformatics/bts460
Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. doi:10.1093/bioinformatics/btu033
Upadhyaya, H.D., and R. Ortiz. 2001. A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. Theor. Appl. Genet. 102:1292–1298. doi:10.1007/s00122-001-0556-y
Varshney, R.K., C. Song, R.K. Saxena, S. Azam, S. Yu, A.G. Sharpe, et al. 2013. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. Nat. Biotechnol. 31:240–246.
Vavilov, N.I. 1926. The origin of the cultivation of 'primary' crops, in particular cultivated hemp. In: Studies on the origin of cultivated plants, Soviet Academy of Sciences, Leningrad, Union of Soviet Socialist Republics. English translation by D. Love, 1992. Cambridge Univ. Press, Cambridge, UK. p. 221–233.
Weir, B.S., and W.G. Hill. 2002. Estimating F-statistics. Annu. Rev. Genet. 36:721–750. doi:10.1146/annurev.genet.36.050802.093940
Zohary, D. 1976. Lentil. In: N.W. Simmonds, editor, Evolution of crop plants. Longman Publishers, London, UK. p. 163–164.