Genetic Diversity among Wheat Accessions from the USDA National Small Grains Collection

J. Michael Bonman,* Ebrahiem M. Babiker, Alfonso Cuesta-Marcos, Kathy Esvelt-Klos,
Gina Brown-Guedira, Shiaoman Chao, Deven See, Jianli Chen, Eduard Akhunov, Junli Zhang,
Harold E. Bockelman, and Tyler C. Gordon

ABSTRACT

Accessions of common wheat (*Triticum aestivum* L. subsp. *aestivum*) from the USDA–ARS National Small Grains Collection (NSGC) are a resource for wheat scientists worldwide. The genetic diversity of the wheat core subset, representing approximately 10% of the collection's 42,138 *T. aestivum* accessions, was examined using 390 diversity arrays technology (DArT) markers, 4941 single nucleotide polymorphisms (SNPs), and descriptor data. The marker profiles revealed duplicates, which were excluded to form an informative core (iCore) of 3230 accessions. The iCore population structure and diversity within various subgroups were examined with analysis of molecular variance, principal coordinate analysis, cluster analysis, and by ranking the contribution of individual accessions to overall diversity. Accession groups based on molecular marker data corresponded well to their geographic origin, and population structure was accounted for primarily by differences between Iranian landrace accessions and the rest of the accessions. Accessions classified as breeding lines were overrepresented among those ranked as most diverse based on SNP data, whereas Iranian landraces were underrepresented. Although less diverse as a group, Iranian landrace accessions had a higher frequency of resistance to bunt diseases and Russian wheat aphid compared with the iCore as a whole. The present study provides support for establishing core subsets based on geographic origin of accessions and will be a basis for further study of diversity among NSGC wheats.

J.M. Bonman, E.M. Babiker, H.E. Bockelman, K. Esvelt-Klos, and T.C. Gordon, USDA–ARS, Small Grains and Potato Germplasm Research Unit, 1691 South 2700 West, Aberdeen, ID 83210; A. Cuesta-Marcos, Dep. of Crop and Soil Science, Oregon State Univ., Corvallis, OR 97331; G. Brown-Guedira, Plant Science Research Unit, 3411 Gardner Hall, North Carolina State Univ., Raleigh, NC 27695; S. Chao, USDA–ARS Genotyping Lab., Biosciences Research Lab., Fargo, ND; D. See, USDA–ARS, Wheat Genetics, Quality, Physiology & Disease Research Unit, Washington State Univ., Pullman WA; J. Chen, Univ. of Idaho Aberdeen Research & Extension Center, Aberdeen ID; E. Akhunov, Dep. of Plant Pathology, Kansas State Univ., Manhattan, KS; J. Zhang, Dep. of Plant Sciences, Univ. of California, Davis, CA. Received 15 Sept. 2014. Accepted 26 Jan. 2015. *Corresponding author (mike.bonman@ars.usda.gov).

Abbreviations: AMOVA, analysis of molecular variance; DArT, diversity arrays technology; GRIN, Germplasm Resources Information Network; iCore, informative core; NSGC, National Small Grains Collection; PCoA, principal coordinate analysis; PIC, polymorphism information content; RWA, Russian wheat aphid; SNP, single nucleotide polymorphism.

THE USDA–ARS NSGC is a resource for plant scientists worldwide, especially plant breeders, and it currently contains 58,392 accessions of wheat and wild relatives of wheat, among which are 42138 accessions of common wheat. Since 1897, these materials have been acquired by plant explorers and plant scientists and most recently have been maintained as part of the USDA's National Plant Germplasm System. For approximately the past 30 yr, the collection has been systematically evaluated for various traits of interest to plant breeders and geneticists, including agronomic characteristics such as plant height and maturity

and resistance to important diseases and insect pests. These data are available online through the Germplasm Resources Information Network (GRIN).

In 1995, a core subset of the NSGC was established following the principles outlined by Brown (1989), where 10% of the accessions were selected using stratified random sampling by country of origin. The number of accessions chosen from each country was proportional to the log of the number of accessions per country. Thus, in the core subset, a broad representation by geographic origin was attained and countries with many accessions tended to be underrepresented on a percentage basis and vice versa. For example, accessions from Turkey represent 6.7% of the total *T. aestivum* accessions and constitute only 1.5% of the core subset, whereas accessions from the Philippines represent 0.02% of the total yet constitute 0.2% of the core subset.

In 1997, a group of 6484 common wheat landraces originating from Iran were received by the NSGC through the work of C.O. Qualset at the University of California, Davis. These accessions were originally collected in 1935 from throughout Iran and maintained by scientists at the University of Tehran (Dworkin, 2009). When these wheats were brought into the collection, they were thought to represent a significant source of new genetic diversity and 10.5% of these materials, or 682 accessions, were added to the core subset. The new accessions from Iran showed a relatively high proportion of resistance to common and dwarf bunt diseases (Bonman et al., 2006). In adult plant field screening tests in Kenya, some of the Iranian accessions also proved resistant to new and highly virulent races of the wheat stem rust pathogen (Newcomb et al., 2013). Salinity tolerance was present among these accessions (Jafari–Shabestari et al., 1995), as was significant variation in two quality traits: starch pasting properties and dough color (Bhattacharya et al., 1997, 1999).

Molecular markers can be used as a tool to evaluate crop plant germplasm and have been used to better understand the structure and diversity of germplasm collections (Hao et al., 2006, 2008; Dreisigacker et al., 2005; Dreisigacker et al., 2005; Huang et al., 2002; Cavanagh et al., 2013). The purpose of the present study was to use molecular markers to (i) evaluate the genetic diversity of the *T. aestivum* core subset accessions with respect to geographic origin, improvement status, time period when accessions were obtained, and plant growth habit and (ii) rank accessions based on their contribution to genetic diversity and then compare the most and least diverse groups of accessions using GRIN descriptor data.

MATERIALS AND METHODS
Plant Materials and Marker Data

The NSGC *T. aestivum* core subset consists of 4007 accessions. From this set, single plant selections were derived resulting in 3836 accessions for further study. Two sets of molecular markers were used to genotype the 3836 accessions, including DArT markers (Akbari et al., 2006) and SNPs (Cavanagh et al., 2013). The DNA used in the analyses was extracted from a single seedling from each accession using a hexadecyltrimethyl ammonium bromide (CTAB) protocol (Stewart and Via, 1993). Diversity arrays technology marker data are available online through the GRIN database. Single nucleotide polymorphism markers used in the present study were generated by an Illumina Infinium 9K iSelect platform (Cavanagh et al., 2013) through the Triticeae Coordinated Agricultural Project. During the analysis of data, a few differences between the SNP and DArT results were observed that could be traced back to particular 96–well plates. Since these were likely sample handling errors, 56 accessions showing these anomalies were excluded from further study. Marker data are available at http://wheat.pw.usda.gov/ggpages/maps/NSGCwheat/.

Marker and Genotype Filtering

Data were analyzed using JMP Genomics 6.1 (SAS Institute, 2012). A total of 699 accessions within the core subset showed 99.5% similarity to at least one other accession based on marker profiles. Of these, 470 were excluded and the data for the remaining accessions were filtered to remove markers with minor allele frequencies of 5% or less and accessions with more than 10% missing data. These criteria resulted in 4941 SNP markers and 3230 accessions (Supplemental Table S1). For purposes of analyzing global diversity, this group of accessions was designated the informative core (iCore) after the designation coined by Muñoz-Amatriaín et al. (2014). Diverse arrays technology data for 407 markers were available for iCore and, after filtering, 390 of these were used as a comparison data set for the SNP results in an analysis of molecular variance (AMOVA) and in ranking accessions for their contribution to genetic variation.

Informative Core Population Structure

The population structure within the iCore was examined by conducting an AMOVA, principal coordinate analysis (PCoA), cluster analysis, and by calculating the inflation factor (λ) (Devlin and Roeder, 1999) for the population and subsets of the population.

For the AMOVA, data from GRIN were used to classify accessions by geographical origin, improvement status, growth habit, and the time period when they were acquired by the NSGC. The geographic origin groups were eastern Asia, south-central Asia, western Asia, eastern Europe, southern Europe, western and northern Europe, North and Central America, South America, east and South Africa, northern Africa, Iran, and other. Because the core subset has many accessions from Iran, these accessions were placed into a separate geographic group and landrace accessions from Iran were classified as a separate improvement status group. Accessions where the origin could not be classified into one of the geographic origin groups or where too few accessions were present to form a separate group for analysis were classified as other. These accessions originated primarily from Europe, the former Soviet Union, Oceania, and western Africa. Improvement status groups were designated as improved (cultivar and breeding lines), landrace, landrace from Iran, or uncertain status. Date of acquisition groups were designated so that approximately equal numbers of accessions were

included in most of the groups as follows: 1900 to 1949 ($n = 511$), 1950 to 1959 ($n = 474$), 1960 to 1969 ($n = 588$), 1970 to 1979 ($n = 514$), 1980 to 1996 ($n = 558$), 1997 ($n = 489$), and 1998 to 2006 ($n = 99$). To test the genetic variation accounted for by the various ways of grouping accessions, AMOVA was conducted separately with SNP and DArT markers using the R package pegas with 1000 permutation and Euclidean distance as genetic distance (Paradis, 2010; R Development Core Team, 2013). Accessions with other geographic origin were excluded from the AMOVA, as were accessions of uncertain improvement status.

Principal coordinate analysis of a genetic similarity matrix generated from the SNP data was conducted to visualize population structure. Genetic similarity between accessions was evaluated by genetic distance using GenAlex (Peakall and Smouse, 2006; Peakall and Smouse, 2012), in which the number of alleles at which two lines differ is summed across all markers. In all cases, heterozygote genotype calls were treated as missing data.

Cluster analysis of the iCore was done using SNP data by first generating a similarity matrix using the relationship matrix function in JMP Genomics based on allele sharing for each accession pair. Accessions were then grouped based on the fast ward hierarchical clustering procedure.

The inflation factor (λ) was calculated using the genomic control method of Devlin and Roeder (Devlin and Roeder, 1999) in SNP and Variation Suite v7 (Golden Helix, 2010). Calculations of λ were made for the iCore, for the iCore without the Iranian landraces, and for the Iranian landraces. In these analyses, a λ close to 1 indicates a lack of population structure.

Phenotypic Data

Data for various descriptors, including agronomic traits and disease and pest resistance ratings, were obtained from the GRIN for accessions within the iCore. The GRIN descriptor days to anthesis ($n = 2196$) was the number of days from 1 January when 50% of the spikes are fully extended from the boot. Days to flowering ($n = 3063$) was the value recorded for days to anthesis expressed as a ratio to days to anthesis for a standard check. Flowering data was collected based on plantings in late November at Maricopa, AZ. The descriptor plant height ($n = 2297$) was the average height of plants at maturity measured in centimeters from the ground to top of the spike, excluding awns. The plant height ratio ($n = 3061$) was the value recorded for plant height expressed as a ratio to plant height for a standard check. Plant growth habit groups were spring, winter, or facultative based on spring planting at Aberdeen, ID.

The GRIN data for disease and insect resistance were generated over the past several decades by NSGC cooperators and details of trials can be found online (USDA–ARS, 2014). Russian wheat aphid (RWA) and leaf rust resistance were tested on greenhouse-grown seedlings, and resistance to stripe rust, stem rust, and bunt was evaluated on adult plants in field trials. Data for stem rust and leaf rust were recorded on a 0-to-9 scale with lower values meaning greater resistance. Data for stripe rust severity were recorded on a 0-to-8 scale, and both reaction type and severity were rated separately, with lower values being more resistant. Bunt was rated as incidence relative to a susceptible check. For analysis in the present study, accessions were classified as either resistant or susceptible using the following

criteria: accessions with RWA (Biotype 1) severity scores of ≤ 3 were classified as resistant; accessions with bunt incidence of less than 5% relative to susceptible checks within each trial were classified as resistant, combining dwarf and common bunt resistance and considering an accession susceptible if it scored >5% in any single trial (Bonman et al., 2006); accessions with stem rust rating of ≤ 2 were classified as resistant. Many accessions had been tested for stripe rust resistance at two locations in Washington State and only accessions with the following combinations of results were classified: those with infection type ratings of ≤ 2 in both trials were considered resistant and those with a score of 8 in at least one trial were classified as susceptible. Stripe rust severity of $\leq 5\%$ in two trials were classified as resistant, whereas those with scores of $\geq 20\%$ in at least one trial were considered susceptible. For leaf rust, ratings of ≤ 2 were classified as resistant and those >2 were considered susceptible.

Genetic Relationships within and Among Groupings

Polymorphism information content (PIC) values from SNP data were used as a measure of genetic diversity within clusters, geographic regions, growth habit categories, the various categories of improvement status, and for the various time periods when accessions were acquired. The PIC values for each group were calculated with the marker properties function in JMP Genomics. Within the Iranian accessions, specific origin was available for 476 accessions from Clusters 6, 8, and 10. With the graphing function in JMP Genomics 6.1, these data were used to map the locations where accessions from these clusters originated within the country.

To examine the relationship between accessions within the iCore grouped by improvement status, genetic matrices for each group were generated using the population measures feature in JMP Genomic. Rogers index was used as a measure of genetic distance between the groups and for clustering the groups with the fast ward hierarchical procedure.

Comparing Iranian Landraces with Other Landrace Groups

Initial analysis of the iCore showed that the Iranian landrace accessions appeared to be less diverse than all other landraces in general. To better assess the relative diversity of the Iranian landraces compared with bordering countries, PIC values were also generated for landrace accessions originating from Afghanistan, Armenia, Azerbaijan, Iraq, Pakistan, Turkey, and Turkmenistan.

To compare phenotypic variation among landrace accessions from Iran with variation among landrace accessions from bordering countries, data from the NSGC as a whole, rather than only the core subset, were used. The descriptors days to flowering and plant height ratio were the most complete with values available for more than 12,000 accessions in GRIN. Accessions were grouped by plant habit and the variances of the two descriptors for landraces from Afghanistan, Iraq, Pakistan, and Turkey were compared with variances for accessions from Iran using the O'Brien test for unequal variances (O'Brien, 1979). Landraces from Armenia and Azerbaijan were included and grouped together for the analysis, since these two countries border one another and each had relatively few accessions in the NSGC.

Ranking Accessions for Contribution to Genetic Diversity

Accessions were ranked for their contribution to the genetic diversity of the iCore with the method previously used with NSGC barley accessions by Muñoz-Amatriaín et al. (2014). The R package geneticsubsetter developed by Ryan Graebner for this purpose is available at http://barleyworld.org/breeding-genetics/analysis. The program is used to sort accessions according to their general contribution to the average PIC of the markers. Using rankings generated from both SNP and DArT datasets, accessions were split into the most diverse half (top 50%) and the least diverse half (bottom 50%) and compared using GRIN descriptor data for (i) mean and standard deviation in plant height and maturity date; (ii) resistance to RWA, dwarf and common bunt, stripe rust, and leaf rust; (iii) percentage of each improvement status group; and (iv) percentage facultative, spring, and winter types. These traits were chosen because sufficient data were present in GRIN to enable analysis. The variation in height and maturity was compared using O'Brien's test for unequal variances (O'Brien, 1979) and means were compared using Welch means testing allowing for unequal standard deviations (Welch, 1951). Fisher's exact test was used to compare the occurrence of resistances among accessions classified as either most or least diverse based on marker results. The χ^2 test was used to compare the occurrence of various improvement status and growth habit categories among accessions classified as either most or least diverse based on marker results.

RESULTS
Duplication in the NSGC Core Subset

Most of the 699 accessions showing 99.5% similarity were in fact nearly 100% identical to at least one other accession in the core subset. These accessions generally grouped into sets of two or three nearly identical accessions based on the SNP data, except for accessions from Iran, which also had groups from four to 23 identical or very nearly identical accessions. Although accessions from Iran represented 18.3% of the core subset, they comprised 30.3% of the accessions that were identical or nearly identical to at least one other accession. Data from DArT analysis generally agreed with the SNP assessment of duplication.

Informative Core Population Structure

The AMOVA showed that the grouping of accessions by origin, improvement status, acquisition date, and growth habit did not account for much of the variation within the iCore, and, consequently, most of the variation could be accounted for by variation among accessions within groups (Table 1). Geographic origin accounted for the highest proportion of the remaining variation, 13.2% based on SNP data and 9.8% based on DArT data, followed by improvement status groups, date of acquisition groups, and finally by plant growth habit groups, which accounted for only 2.5 and 3.1% of the total variation with DArT and SNP markers, respectively. The SNP and DArT results were

Table 1. Analysis of molecular variation of 3230 common wheat accessions grouped by geographic origin, improvement status, date acquired, and habit using two different marker datasets.

Source of variation[†]	df	Marker type[‡]	Variance components (F)	Variation (%)
Among geographic origin	10	SNP	121.32**	13.2
		DArT	7.49**	9.8
Among improvement status	3	SNP	380.78**	10.3
		DArT	5.96**	7.7
Among date acquired	7	SNP	63.37**	6.9
		DArT	3.68**	4.8
Among habit	2	SNP	28.02**	3.1
		DArT	1.94**	2.5

** Significant at the 0.01 probability level.

† Source of variation includes two parts: among groups and among individuals within groups; the total variation is 100%. Only the variation among groups is listed.

‡ DArT, diverse arrays technology; SNP, single nucleotide polymorphism.

comparable with more variation being accounted for by the groups using SNP results. In a previous study using mostly modern cultivars, 7.8% of the differentiation among genotypes was attributable to plant growth habit (Cavanagh et al., 2013). When only cultivars and breeding lines from the present study were used in the AMOVA and landrace, landraces from Iran, and uncertain status accessions were removed, the variation accounted for by plant growth habit was more similar (5.8%) to that found in the previous study.

In the PCoA plot of genetic similarity among lines, the first principal coordinate separated the majority of Iranian landraces from the majority of other lines (Fig. 1). The cluster analysis of the iCore population (Fig. 2) showed the groupings corresponded with the geographic origin of the accessions, with half or more of the accessions from a specific region falling within one to three clusters (Table 2). The relationships among broader geographic groupings were also consistent. For example, many accessions from Europe (eastern Europe, southern Europe, western and northern Europe) grouped together in Cluster 4. Accessions from the various Asian regions also grouped together. Relationships were consistent with historical connections between regions as, for example, southern Europe accessions frequently occurring with accessions from northern Africa and South America.

Where accessions grouped based on the marker, results were related to when the accessions were acquired by the NSGC (Table 3). For instance, many accessions from Cluster 9, mostly coming from East Asia and south-central Asia, were obtained during the first half of the 20th century. Most Iranian accessions were acquired in 1997, and these grouped into three clusters (6, 8, and 10). Plant growth habit was also clearly delineated based on the marker data, with winter types predominating in Clusters 4, 5, and 10. Similarly, improvement status categories were also separated based on the marker results, with breeding lines and
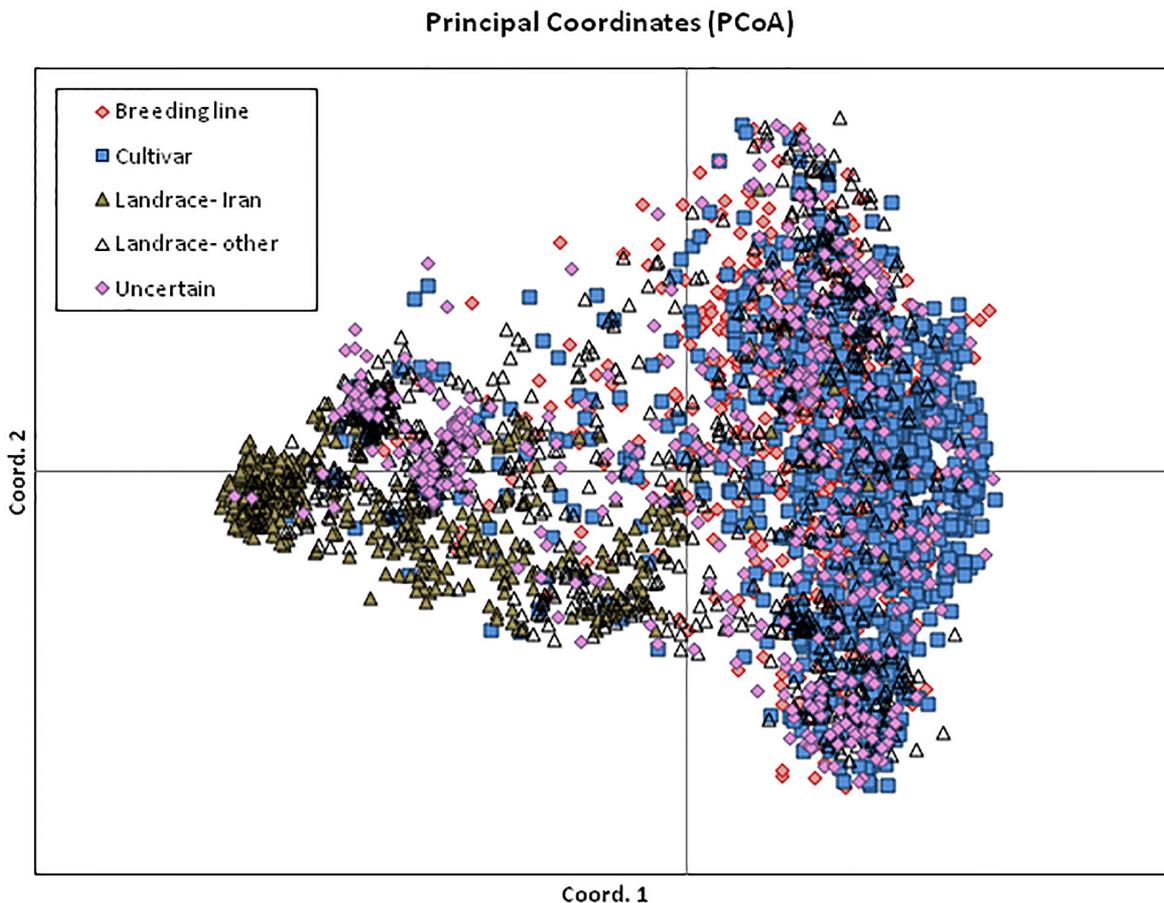
Principal Coordinates (PCoA)

Figure 1. Population structure of 3230 accessions of *Triticum aestivum* subsp. *aestivum* visualized by plotting principal coordinate (PCoA) of a genetic similarity matrix generated from single nucleotide polymorphism data. Colored markers indicate the improvement status of accessions.

cultivars (improved) most common in some clusters and landraces being most common in others (Table 3).

The inflation factor (λ = 1.91) generated from the iCore was consistent with the hypothesis of a structured population. Calculated using the genomic control method of Devlin and Roeder (Devlin and Roeder, 1999), the inflation factor is expected to approach 1 in a sample without discernable population structure. The inflation factors from stratification into groups containing Iranian landraces or all other lines, λ = 1.00 and 1.02 respectively, indicate little structure within these two groups. These results support the hypothesis that much of the population structure within the iCore was due to genetic differences between the Iranian landraces and the remainder of the iCore.

Genetic Relationships within and Among Groupings

Genetic diversity within various groupings as measured by PIC values varied little except that winter types were somewhat less diverse than spring types and the Iranian accessions were much less diverse than other groupings (Table 3). This lower diversity within the Iranian accessions was also evident among accessions in Cluster 10 (86.2% from Iran), where the high similarity within the cluster was illustrated

by the red color within the matrix and the dendrogram branch compression towards the figure axis (Fig. 2).

Among accessions from Iran, 94.2% fell within Clusters 6, 8, and 10 (Table 2). Cluster 6 also contained a majority of the accessions from western Asia. In contrast, Cluster 10 had practically no accessions from western Asia, but did have accessions from south-central Asia. Cluster 8 had many accessions from both western Asia and south-central Asia. When the specific geographic origin of 476 accessions from Iran were mapped by cluster group, the geographic distribution of the three clusters within Iran fit the broader regional geographic distribution of clusters (Fig. 3). Cluster 6 accessions were more frequent in western Iran adjacent the western Asia region, Cluster 10 accessions were more common in eastern Iran adjacent the south-central Asia region, and Cluster 8 accessions were found in both eastern and western Iran (Fig. 3).

The relationship among various improvement status groups was examined by generating a measure of genetic distance (Rogers index) and clustering the groups based on the index values. As would be expected, improved accessions consisting of breeding lines and cultivars grouped together (Fig. 4). Landrace accessions grouped with accessions of unknown origin and landraces originating from
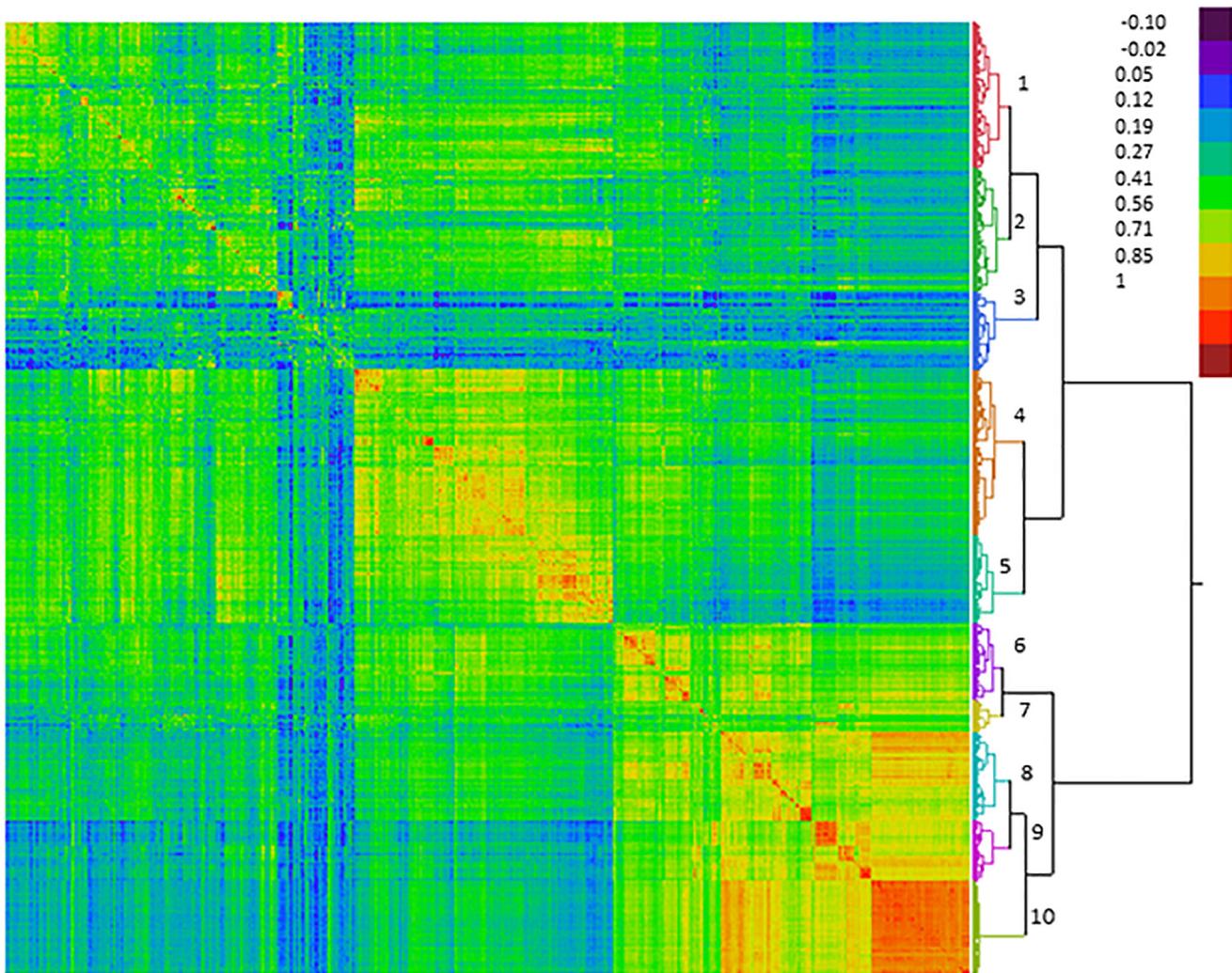
Figure 2. Population structure of 3230 accessions of *Triticum aestivum* subsp. *aestivum* visualized by hierarchical clustering with Ward's method using a genetic similarity matrix generated from single nucleotide polymorphism data. The heat plot was constructed based on allele -sharing coefficients calculated in the relationship matrix function in JMP Genomics v.6.1. The heat plot represents a pairwise marker-to-marker allele sharing between two accessions with red indicating the highest proportion of shared alleles and blue indicating the lowest. Ten clusters (subpopulations) were grouped by different colors in the population. Values in each cluster represent subpopulation number.

Table 2. Number of accessions by geographic region and percentage of accessions within a region placed into various hierarchical clusters generated from an allele-sharing matrix from analysis of 4941 SNP markers in 3230 common wheat accessions from the USDA–ARS National Small Grains Collection core subset.

Geographic region	Total (*n*)	Hierarchical cluster[†]									
		1	2	3	4	5	6	7	8	9	10
Eastern Asia	118	5.9	*17.8*	2.5	13.6	3.4	1.7	16.9	1.7	*36.4*	0.0
South-central Asia	259	4.2	3.9	2.3	8.1	0.8	3.5	5.0	*18.5*	*41.3*	12.4
Iran	514	0.4	1.4	1.0	0.0	0.0	18.9	2.7	20.6	0.4	*54.7*
Western Asia	280	6.8	12.5	9.3	6.1	0.7	*38.6*	3.2	*17.5*	3.6	1.8
Eastern Europe	315	12.4	7.9	0.3	*62.2*	11.4	2.2	0.3	1.0	1.3	1.0
Western and northern Europe	360	10.3	5.3	1.4	*30.3*	*49.4*	0.6	0.0	1.4	0.8	0.6
Southern Europe	278	*34.2*	19.8	8.3	*26.6*	5.8	4.7	0.0	0.0	0.7	0.0
Northern Africa	132	*34.1*	12.9	8.3	2.3	0.0	4.5	9.8	*24.2*	3.8	0.0
East and South Africa	199	21.1	22.1	30.2	7.0	0.0	1.0	9.0	5.5	4.0	0.0
South America	317	*34.1*	*30.3*	14.5	5.4	9.1	2.5	1.9	0.3	1.6	0.3
North and Central America	294	18.0	*25.5*	*21.1*	22.8	9.2	0.7	0.7	1.0	1.0	0.0
Other[‡]	164	*22.6*	7.9	7.9	18.9	2.4	3.7	3.7	*27.4*	4.3	1.2

[†] Sum of figures in italics represents 50% or greater of the total for the region.

[‡] Includes accessions originating primarily from Europe (where specific country of origin is unknown), the former Soviet Union, Oceania, and western Africa.

Table 3. Number of accessions and polymorphism information content (PIC) values by date period acquired, habit, and improvement status, and the percentage of each subcategory occurring within 10 hierarchical clusters generated from an allele-sharing matrix from analysis of 4941 SNP markers in 3230 common wheat accessions from the USDA–ARS National Small Grains Collection core subset.

Accession category	Subcategory	n	PIC	Hierarchical cluster†									
				1	2	3	4	5	6	7	8	9	10
Date period acquired	1900–1949	511	0.29	23.0	6.9	8.0	16.1	10.1	27.9	21.6	12.8	*40.2*	3.7
	1950–1959	474	0.29	*31.5*	14.6	16.3	14.5	14.1	6.5	*25.5*	4.9	9.5	4.0
	1960–1969	588	0.29	18.8	*24.4*	21.3	24.2	*39.3*	8.0	9.8	7.5	10.1	2.8
	1970–1979	511	0.29	9.9	16.5	14.6	*25.5*	24.2	5.0	19.6	16.4	25.6	1.5
	1980–1996	558	0.30	13.5	29.9	*33.1*	14.5	8.7	17.2	7.8	25.2	13.6	4.3
	1997	489	0.19	0.4	1.4	0.8	0.2	0.3	35.5	13.7	*31.8*	1.0	*83.7*
	1998–2006	99	0.29	2.8	6.0	5.3	5.0	3.4	0.0	2.0	1.3	0.0	0.0
Habit	Spring	1845	0.30	*76.0*	*78.0*	*93.5*	34.9	7.0	*67.2*	*72.5*	*64.6*	*68.3*	30.4
	Winter	1223	0.27	16.6	18.7	3.8	*62.3*	*90.6*	25.2	26.5	27.5	25.6	*62.3*
	Facultative	162	0.28	7.5	3.3	2.7	2.8	2.3	7.6	1.0	7.9	6.0	7.4
Improvement status	Improved	1440	0.29	*50.1*	*73.9*	*70.7*	*56.6*	*85.2*	6.9	*54.9*	5.6	12.6	2.5
	Landrace, Iran	512	0.19	0.4	1.7	1.1	0.0	0.0	37.0	13.7	34.8	1.0	*86.2*
	Landrace, other	856	0.28	33.9	12.7	16.4	26.7	5.0	*45.8*	17.6	*40.7*	*66.3*	10.4
	Uncertain	422	0.29	15.6	11.8	11.8	16.6	9.7	10.3	13.7	19.0	20.1	0.9
Total		3230		495	417	261	565	298	262	102	305	199	326

† Figures in italics are highest values within a hierarchical cluster by subcategory group.
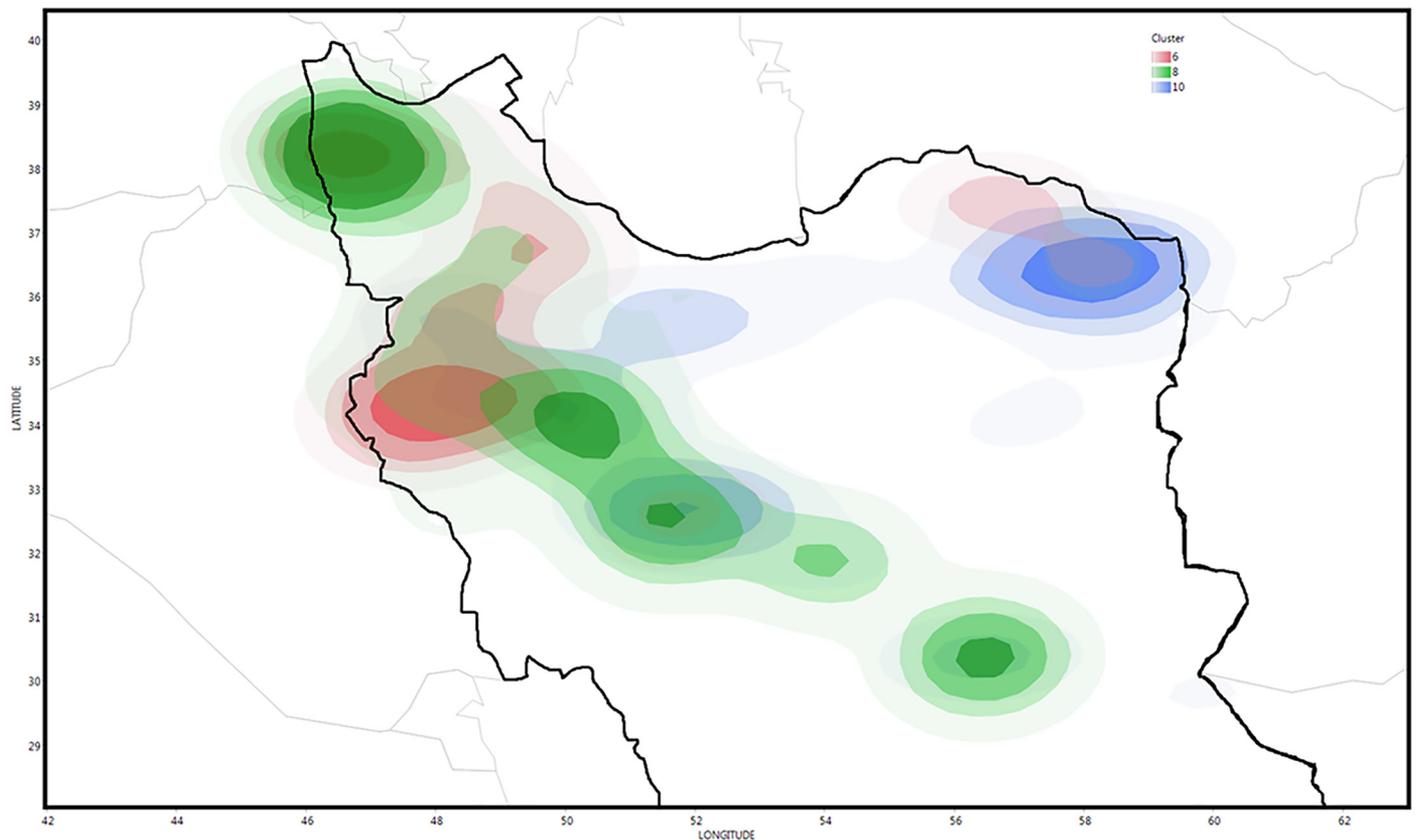

Figure 3. Geographic origin for 476 Iranian accessions of *Triticum aestivum* subsp. *aestivum* grouped into three clusters (6, 8, and 10) based on single nucleotide polymorphism markers. Color intensity is proportional to accession density using the contour plot function in JMP Genomics graph builder.
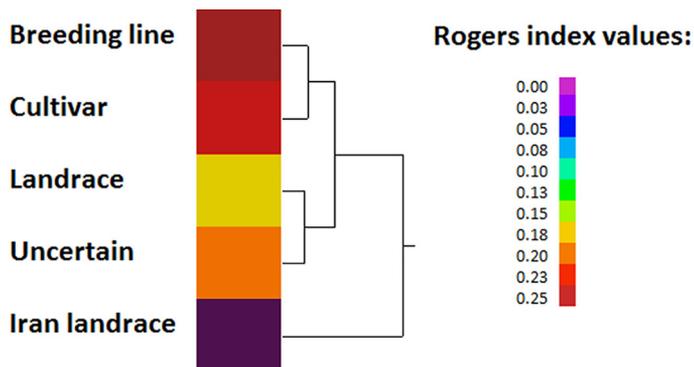
Figure 4. Hierarchical clustering of accessions within the informative core by improvement status based on Rogers index of genetic distance.

Iran grouped separately from all other categories. This result parallels that of the overall population structure, where Iranian materials tended to group separately from most other NSGC accessions.

Comparing Iranian Landraces with Other Landrace Groups

When compared with landraces originating from adjacent countries, the PIC value for Iran (0.19, $n = 512$) was lower than those for Afghanistan (0.22, $n = 53$), Armenia (0.23, $n = 21$), Azerbaijan (0.25, $n = 10$), Iraq (0.22, $n = 22$), Syria (0.29, $n = 8$), Turkey (0.24, $n = 42$), and Turkmenistan (0.23, $n = 5$). Only landraces from Pakistan (0.19, $n = 35$) had a value similar to that of Iran. Similarly, when compared using all of the phenotypic data in GRIN for days to flowering and plant height ratio, the accessions from bordering countries generally showed significantly greater variance for these traits versus the Iranian landraces (Table 4).

Ranking Accessions for Contribution to Genetic Diversity

Based on both SNP and DArT rankings, accessions in the most diverse group had significantly shorter plant stature and earlier maturity compared with those ranked as the least diverse (Table 5). Also, the variation in these traits was significantly greater in nearly all instances for accessions ranked in the most diverse group. For example, the standard deviation for the descriptor days to anthesis was 41.8 for accession classified as most diverse based on SNP data, but only 31.9 for those classified as least diverse (Table 5).

With respect to disease and insect resistance descriptors, resistance to the rust diseases was significantly more frequent among the most diverse category compared with the least diverse category, except for stripe rust based on DArT ranking of diversity (Table 6). The opposite result was found for resistance to the bunt diseases and RWA, since the least diverse group showed a higher frequency of resistance based on both the SNP and DArT diversity categories (Table 6).

Landraces from Iran and winter habit accessions were underrepresented in the most genetically diverse accessions based on both marker types, whereas breeding lines and spring habit accessions were overrepresented (Table 6).

DISCUSSION

Crop germplasm collections, such as the NSGC, are valuable resources for plant breeders and other scientists working in the area of crop improvement. Evaluation of the genetic resources maintained in these collections is a prerequisite for the effective use of these materials (Hawkes, 1981). Traditionally, phenotypic assessments have been the basis of germplasm evaluations, but the use of molecular markers and other new molecular biological tools is increasing our understanding of the natural genetic variation present in crop germplasm collections. A deeper understanding of this genetic variation should enable more efficient management of our germplasm resources and more effective use of them in practical breeding work. Such understanding is particularly important for wheat, since yield increases from 1980 to 2010 have declined to 0.8% per year (Graybosch et al., 2014), a level where production gains will not keep pace with demand (Reynolds et al., 2012).

In the present study, many accessions within the *T. aestivum* core subset were closely related or even identical based on marker profile. Passport data sometimes gave clues as to how the duplication arose. For example, the landraces Nulsu Misu (PI87116) and Ejuiea (PI87117) were both collected before 1930 by the same scientist, P. Dorsett from the USDA, in the same location in Korea. It is likely a single Korean landrace was known by two names, and this genotype was collected as two accessions. Another example is PI191303, listed as the cultivar Alba from Belgium and having the same marker profile as CItr13256, a cultivar from Netherlands also named Alba. These types of duplication probably account for most of the cases where two or three accessions have matching marker data.

Results from near duplication based on SNP data also revealed relationships among accessions that were not evident based on accession origin, but were consistent based on passport data. For example, PI266822 was released as the cultivar Tainui in New Zealand in 1939, yet except for two of 5644 SNP's, it shares the identical marker profile as two accessions from Portugal that were received by the National Plant Germplasm System in the first half of the 20th century. Information recorded in GRIN indicates that Tainui was a selection from Portugal 19, thus establishing that the marker results parallel the information on origin of the accession.

To maintain a high degree of genetic diversity within a manageable number and size of samples (Porceddu and Damania, 1991; Hawkes, 1981), it is important to avoid duplication of accessions. Because the marker data are based on single plant selections from each accession, accessions showing absolute identity based on the marker data warrant further study to determine if they also appear to be identical based on

Table 4. Comparison of phenotypic variation within Iranian landrace accessions to that of landrace accessions from countries bordering Iran. Analysis used all available GRIN *Triticum aestivum* data for the descriptors days to flowering and plant height ratio, grouped by plant habit.

Habit	Country of origin	Days to flowering			Plant height ratio		
		n	Mean	Relative standard deviation[†]	*n*	Mean	Relative standard deviation[†]
Facultative	Iran	539	1.51	1.00	538	1.76	1.00
	Turkey	499	1.48	1.09	499	1.85	1.17*
	Afghanistan	76	1.55	1.28**	76	1.83	1.15
	Pakistan	208	1.41	1.69**	208	1.67	1.31**
Spring	Iran	2792	1.40	1.00	2792	1.76	1.00
	Turkey	1505	1.44	1.11**	1506	1.86	1.07*
	Afghanistan	907	1.42	1.24**	906	1.8	1.13**
	Pakistan	542	1.30	1.41**	542	1.79	1.25**
	Armenia and Azerbaijan	33	1.48	1.05	33	1.89	0.96
	Iraq	73	1.25	0.93	73	1.81	1.26**
Winter	Iran	3764	1.54	1.00	3766	1.77	1.00
	Turkey	381	1.53	1.15**	381	1.84	1.11
	Afghanistan	532	1.54	1.37**	533	1.86	1.24**
	Pakistan	62	1.41	1.19	62	1.72	1.39**
	Armenia and Azerbaijan	32	1.56	1.14	32	2.06	1.48**

* Significantly different from the standard deviation of accessions from Iran by the O'Brien test for unequal variances ($P < 0.05$).

** Significantly different from the standard deviation of accessions from Iran by the O'Brien test for unequal variances ($P < 0.01$).

[†] Standard deviation relative to that of accessions from Iran.

Table 5. Mean and standard deviation of four GRIN descriptors among common wheat accessions from the USDA–ARS National Small Grains Collection core subset grouped by genetic diversity as assessed with molecular markers.

GRIN descriptor[†]	Statistic	SNP diversity category			DArT diversity category		
		Prob > F[‡]	Most diverse	Least diverse	Prob > F	Most diverse	Least diverse
Days to anthesis	Mean	0.001	101.2	106.4	ns	102.1	105.3
	Standard deviation	<0.0001	41.9	31.8	0.001	40.2	35.4
Days to flowering	Mean	<0.0001	1.3	1.5	<0.0001	1.3	1.4
	Standard deviation	0.04	0.19	0.18	ns	0.2	0.2
Plant height	Mean	<0.0001	112.1	121.5	<0.0001	114.5	118.0
	Standard deviation	<0.0001	20.7	18.2	0.0004	21.2	19.0
Plant height ratio	Mean	<0.0001	1.7	1.8	<0.0001	1.8	1.8
	Standard deviation	<0.0001	0.33	0.28	<0.0001	0.33	0.29

[†] Days to anthesis ($n = 2196$) was the number of days from 1 January when 50% of the spikes are fully exerted from the boot; days to flowering ($n = 3063$) was the value recorded for days to anthesis expressed as a ratio to days to anthesis for a standard check; plant height ($n = 3061$) was the average height of plants at maturity, measured in centimeters from the ground to top of the spike, excluding awns; plant height ratio ($n = 2297$) was the value recorded for plant height expressed as a ratio to plant height for a standard check.

[‡] Obrien's test for unequal variances for standard deviations; Welch ANOVA test for means equal allowing for standard deviations not equal.

important phenotypic traits. Such information will be useful for efficiently managing the collection in the future.

The various analyses in the present study support the hypothesis that most of the population structure within the iCore is due to differences between the Iranian landrace accessions acquired in the late 1990s and the rest of the iCore. This result has implications for studies using the iCore for association analysis, where population structure can lead to false positive associations (Viljálmsson and Nordborg, 2013). Work is currently underway to obtain phenotypic data from the single plant selections from the NSGC wheat core accessions used in the present study. Fortuitously, there are a large number of Iranian landraces included within the NSGC core, and, therefore, it may be most advantageous to analyze for marker–trait associations within these landraces

separately from the rest of the core subset. Such an approach would avoid potential loss of information by overcorrecting for the effects of population structure when analyzing the whole core subset as a single dataset.

Not only are the Iranian landrace accessions different, they are also less diverse as a group than the other *T. aestivum* accessions in the iCore. The Iranian landrace accessions had a lower PIC value compared with other groups of accessions (Table 3), were less diverse compared with landraces from all adjacent countries except Pakistan, and were more commonly found in the less diverse half of the iCore based on the contribution to diversity ranking. Even though 655 SNPs included in the Infinium 9K iSelect platform were "…discovered in a diverse panel of wheat landraces…" (Cavanagh et al., 2013), it is possible that the

Table 6. Percentage resistance, improvement status category, and habit among common accessions from the USDA–ARS National Small Grains Collection core subset grouped by genetic diversity as assessed with molecular markers.

Descriptor category[†]	Trait	*n*	SNP diversity category			DArT diversity category		
			P[‡]	Most diverse	Least diverse	*P*	Most diverse	Least diverse
Resistance	Russian wheat aphid	2350	<0.0001	0.5	2.9	0.03	1.0	2.1
	Bunt	2862	0.002	0.9	2.4	0.04	1.2	2.1
	Stem rust	1274	<0.0001	25.5	8.8	<0.0001	26.4	11.6
	Stripe rust severity	2326	0.0001	30.4	23.3	ns[§]	27.3	26.3
	Stripe rust infection type	1865	<0.0001	38.5	27.0	ns	33.7	31.3
	Leaf rust	2362	<0.0001	7.7	3.2	<0.0001	7.7	3.7
Improvement status	Breeding line	666	<0.0001	76.4	23.6	<0.0001	66.3	33.7
	Cultivar	774	ns	50.5	49.5	ns	46.1	53.9
	Landrace	856	ns	48.0	52.0	ns	50.5	49.5
	Landrace-Iran	512	<0.0001	11.9	88.1	<0.0001	36.5	63.5
	Uncertain	422	<0.0001	57.6	42.4	ns	46.8	53.2
Habit	Facultative	162	ns	45.3	54.7	<0.04	41.7	58.3
	Spring	1833	<0.0001	62.8	37.	<0.0001	58.5	41.5
	Winter	1223	<0.0001	31.5	68.5	<0.0001	38.5	61.5

[†] Data from the Germplasm Resources Information Network and classified as follows: Russian wheat aphid seedling score for Biotype 1 greenhouse trials of 3 or less = resistant; relative bunt incidence from field trials of less than 5% = resistant, combining dwarf and common bunt resistance and considering an accession susceptible if >5% in any single trial; stem rust adult plant field rating of 2 or less = resistant; stripe rust adult plant rating of infection type of 2 or less in two trials = resistant, a rating of 8 in at least one trial = susceptible; stripe rust adult plant severity rating of 5% or less in two trials = resistant, susceptible if severity of 20% or more in at least one trial; leaf rust seedling score from greenhouse trials of 2 or less = resistant.

[‡] *P* is the probability of independence of the two variables, trait and diversity category, based on Fisher's exact test for the descriptor category resistance and the χ^2 test for descriptor categories improvement status and habit.

[§] ns, not significant.

genetic diversity within the Iranian landraces was under-represented resulting in ascertainment bias in the SNP discovery process. However, using the largest available *T. aestivum* datasets in GRIN, the Iranian landraces showed less variation in days to flowering and plant height ratio compared with the landraces from bordering countries (Table 4). This result supports the findings from the iCore SNP analysis; the landrace accessions from Iran were less diverse than other landraces within the collection. Further work using a genotyping by sequencing approach will be needed to determine the degree to which ascertainment bias might possibly influence the assessment of genetic diversity among various groups of NSGC accessions.

In spite of the fact that the Iranian landraces appeared to be less diverse as a group, these materials contributed genetic variation to the collection that was not well represented otherwise, as illustrated by Cluster 10 (Fig. 2), which is made up of accessions of all growth habit types mostly from Iran and is distinct from the rest of the iCore. Similarly, these accessions were distinct when compared with other improvement status groups based on other analyses including PCoA and Rogers genetic similarity index. Although nearly 90% of the Iranian landraces were found within the least diverse group based on diversity ranking using SNP data, this group had more frequent resistance to bunt disease and RWA compared with the most diverse group. Russian wheat aphid and bunt resistance are relatively rare in the collection as a whole (Bonman et al., 2006; Porter et al., 1993). If the core subset were assembled based only on the diversity of molecular marker loci (e.g., using contribution to diversity ranking), the frequency of these rare but agronomically important alleles would have been reduced. This result fits with the findings of Reeves et al. (2012), who concluded that generating core collections based on neutral marker data would not necessarily result in greater allelic variation for important agronomic loci that have been under selection. Thus, establishing the core subset based on geographic origin was a reasonable approach as it captured the rare resistances present within the Iranian landraces.

In contrast to the results with the rare resistances to RWA and bunt, resistance to the rust diseases was significantly more frequent among the more genetically diverse iCore accessions. These more diverse accessions also tended to have more spring habit accessions and breeding lines. The greater rust resistance among this group is likely related to breeding activity, and indeed rust resistance was more frequent among breeding lines (data not shown). The fact that accessions classified as breeding lines were overrepresented among the most diverse half of the iCore shows that plant breeding activity has enhanced the genetic diversity of wheat within the NSGC.

In summary, the application of molecular marker data in the present study provides a description of genetic diversity within the NSGC core subset that was previously impossible to obtain using only phenotypic data from GRIN. Apparent duplication within the core subset was more frequent than previously thought. This realization provides new avenues for further study that could lead to

much more efficient means of managing the collection. Also, it is now clear that the Iranian landrace accessions that were acquired by the NSGC in 1997, although relatively less diverse as a group, contributed genetic variability to that collection that was previously not well represented. Finally, results from this work will provide a basis for work beyond the core subset as advances in molecular technology make it feasible to obtain and analyze extensive DNA sequence information for the collection as a whole.

Acknowledgments

References

Akbari, M., P. Wenzl, V. Caig, J. Carling, L. Xia, S. Yang, G. Uszynski, V. Mohler, A. Lehmensiek, H. Kuchel, M.J. Hayden, N. Howes, P. Sharp, P. Vaughan, B. Rathmell, E. Huttner, and A. Kilian. 2006. Diversity array technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor. Appl. Genet. 113:1409–1420. doi:10.1007/s00122-006-0365-4

Bhattacharya, M., J. Jafari-Shabestari, C.O. Qualset, and H. Corke. 1997. Diversity of starch pasting properties in Iranian hexaploid wheat landraces. Cereal Chem. 74:417–423. doi:10.1094/CCHEM.1997.74.4.417

Bhattacharya, M., Q. Luo, and H. Corke. 1999. Time-dependent changes in dough color in hexaploid wheat landraces differing in polyphenol oxidase activity. J. Agric. Food Chem. 47:3579–3585. doi:10.1021/jf990041i

Bonman, J.M., H.E. Bockelman, B.J. Goates, D.E. Obert, P.E. McGuire, C.O. Qualset, and R. Hijmans. 2006. Geographic distribution of common and dwarf bunt resistance in landraces of *Triticum aestivum* subsp. *aestivum*. Crop Sci. 46:1622–1629. doi:10.2135/cropsci2005.12-0463

Brown, A.H.D. 1989. Core collections: A practical approach to genetic resources management. Genome 31:818–834. doi:10.1139/g89-144

Cavanagh, C.R., S. Chao, S. Wang, B.E. Huang, S. Stephen, S. Kiani, K. Forrest, C. Saintenac, G.L. Brown-Guedira, A. Akunova, D. See, G. Bai, M. Pumphrey, L. Tomar, D. Wong, S. Kong, M. Reynolds, M. Lopez da Silva, H. Bockelman, L. Talbert, J.A. Anderson, S. Dreisigacker, S. Baenziger, A. Carter, V. Korzun, P.L. Morrell, J. Dubcovsky, M.K. Morrell, M.E. Sorrells, M.J. Hayden, and E. Akhunov. 2013. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc. Natl. Acad. Sci. USA 110:8057–8062. doi:10.1073/pnas.1217133110

Devlin, B., and K. Roeder. 1999. Genomic control for association studies. Biometrics 55:997–1004. doi:10.1111/j.0006-341X.1999.00997.x

Dreisigacker, S., P. Zhang, M.L. Warburton, B. Skovmand, D. Hoisington, and A.E. Melchinger. 2005. Genetic diversity among and within CIMMYT wheat landrace accessions investigated with SSRs and implications for plant genetic resources management. Crop Sci. 45:653–661. doi:10.2135/cropsci2005.0653

Dworkin, S. 2009. The Viking in the wheat field: A scientist's struggle to preserve the world's harvest. Walker & Company, New York.

Golden Helix. 2010. SNP & variation suite v7.1. Golden Helix, Inc., Bozeman, MT.

Graybosch, R., H.E. Bockelman, K.A. Garland-Campbell, D.F. Garvin, and T. Regassa. 2014. Wheat. In: S. Smith, B. Diers, J. Specht, and B. Carver, editors, Yield gains in major U.S. field crops.

CSSA Spec. Publ. 33. ASA, CSSA, SSSA, Madison, WI. p. 459–487.

Hao, C.Y., Y. Dong, L.F. Wang, G.X. You, H.N. Zhang, H.M. Ge, J.Z. Jia, and X.Y. Zhang. 2008. Genetic diversity and construction of core collection in Chinese wheat genetic resources. Chin. Sci. Bull. 53:1518–1526. doi:10.1007/s11434-008-0212-x

Hao, C.Y., X.Y. Zhang, L.F. Wang, Y.S. Dong, X.W. Shang, and J.Z. Jia. 2006. Genetic diversity and core collection evaluations in common wheat germplasm from the Northwestern Spring Wheat Region in China. Mol. Breed. 17:69–77. doi:10.1007/s11032-005-2453-6

Hawkes, J.G. 1981. Germplasm collection, preservation, and use. In: K.J. Frey, editor, Plant breeding II. Iowa State University, Ames. p. 57–83.

Huang, X., A. Börner, R. Röder, and M. Ganal. 2002. Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor. Appl. Genet. 105:699–707. doi:10.1007/s00122-002-0959-4

Jafari-Shabestari, J., H. Corke, and C.O. Qualset. 1995. Field evaluation of tolerance to salinity stress in Iranian haxaploid wheat landrace accessions. Genet. Resour. Crop Evol. 42:147–156. doi:10.1007/BF02539518

Muñoz-Amatriaín, M., A. Cuestra-Marcos, J.B. Endelman, J. Comadran, J.M. Bonman, H. Bockelman, S. Chao, J. Russell, R. Waugh, P.M. Hays, and G.J. Muehlbauer. 2014. The USDA barley core collection: Genetic diversity, population structure, and potential for genome-wide association studies. PLoS ONE 9:E94688. doi:10.1371/journal.pone.0094688

Newcomb, M., M. Acevedo, H.E. Bockelman, G. Brown-Guedira, B.J. Goates, E.W. Jackson, Y. Jin, P. Njau, M.N. Rouse, D. Singh, R. Wanyera, and J.M. Bonman. 2013. Field resistance to the Ug99 race group of the stem rust pathogen in spring wheat landraces. Plant Dis. 97:882–890. doi:10.1094/PDIS-02-12-0200-RE

O'Brien, R.G. 1979. A general ANOVA method for robust tests of additive models for variances. J. Am. Stat. Assoc. 74:877–880. doi:10.1080/01621459.1979.10481047

Paradis, E. 2010. Pegas: An R package for population genetics with an integrated-modular approach. Bioinformatics 26:419–420. doi:10.1093/bioinformatics/btp696

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

Peakall, R., and P.E. Smouse. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research: An update. Bioinformatics. doi:10.1093/bioinformatics/bts460

Porceddu, E., and A.B. Damania. 1991. Sampling strategies for conserving variability of genetic resources of seed crops. ICARDA Tech. Manual No. 17. ICARDA, Aleppo, Syria.

Porter, D.R., J.A. Webster, and C.A. Baker. 1993. Detection of resistance to the Russian wheat aphid in hexploid wheat. Plant Breed. 110:157–160. doi:10.1111/j.1439-0523.1993.tb01228.x

R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Stat. Comput., Vienna, Austria.

Reeves, P.A., L.W. Panella, and C.M. Richards. 2012. Retention of agronomically important variation in germplasm core collections: Implications for allele mining. Theor. Appl. Genet. 124:1155–1171. doi:10.1007/s00122-011-1776-4

Reynolds, M., J. Foulkes, R. Furbank, S. Griffiths, J. King, E. Murchie, M. Parry, and G. Slafer. 2012. Achieving yield gains in wheat. Plant Cell Environ. 35:1799–1823. doi:10.1111/j.1365-3040.2012.02588.x

SAS Institute. 2012. JMP Genomics, release 6.1. SAS Inst. Inc., Cary, NC.

Stewart, N.C., and L.E. Via. 1993. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques 14:748–749.

USDA–ARS. 2014. National Genetic Resources Program. Germplasm Resources Information Network (GRIN). National Germplasm Resources Laboratory, Beltsville, MD. http://www.ars-grin.gov/npgs/ (accessed 4 Sept. 2014).

Viljálmsson, B.J., and M. Nordborg. 2013. The nature of confounding in genome-wide association studies. Nat. Rev. Genet. 14:1–2. doi:10.1038/nrg3382

Welch, B.L. 1951. On comparison of several mean values: An alternative approach. Biometrika 38:330–336. doi:10.1093/biomet/38.3-4.330