Agro-Morphological, Yield, and Genotyping-by-Sequencing Data of Selected Wheat (*Triticum aestivum*) Germplasm From Pakistan

Madiha Islam [1], Abdullah [2], Bibi Zubaida [1], Nageena Amin [3], Rashid Iqbal Khan [4], Noshin Shafqat [5], Rabia Masood [6], Shahid Waseem [7], Jibran Tahir [8], Ibrar Ahmed [7]*, Muhammad Naeem [9]* and Habib Ahmad [1]*†

[1] Department of Biotechnology and Genetic Engineering, Hazara University, Mansehra, Pakistan, [2] Department of Biochemistry, Quaid-i-Azam University, Islamabad, Pakistan, [3] Institute of Molecular Biology and Biotechnology, Bahauddin Zakariya University, Multan, Pakistan, [4] Institute of Horticultural Sciences, University of Agriculture, Faisalabad, Pakistan, [5] Department of Agriculture, Hazara University, Mansehra, Pakistan, [6] Department of Botany, Hazara University, Mansehra, Pakistan, [7] Alpha Genomics Private Limited, Islamabad, Pakistan, [8] Terrestrial Bioscience New Zealand Limited, Auckland, New Zealand, [9] Federal Seed Certification and Registration Department, Islamabad, Pakistan

INTRODUCTION

Wheat (*Triticum aestivum* L.) is the staple food crop for about 30% of the world's population and contributes over 20% of calories from food (Shewry and Hey, 2015). Current global wheat yield should be doubled to feed a projected human population of 9 Billion by 2050 (Ray et al., 2013). Major challenges that hamper the target of significantly increasing yield include climatic changes, reduction in arable land availability, changes in socio-economic conditions of people in developing countries, loss of biodiversity, and biotic and abiotic stresses (Godfray et al., 2010). The target of yield increase can be achieved by investigating and utilizing the genetic diversity in available wheat germplasm, improving cultivar genetics and crop management practices (Godfray et al., 2010; Philipp et al., 2019).

Genetic diversity provides a foundation for crop improvement (Govindaraj et al., 2015) in order to develop varieties that have a better yield as well as resistance to biotic and abiotic stresses (Khan et al., 2015). Assessment of genetic diversity also helps to understand genomic composition, identify genes for vital traits, conserve and classify genetic variation in plant germplasm, and develop techniques for plant propagation (Khan et al., 2015). Since frequent use of few parents or less diverse genotypes leads to genetic erosion by producing progenies with low heterozygosity and/or inbreeding depression, it is critical to determine genetic diversity in the intended parental lines before starting a breeding program (Tar'an et al., 2005). The progenies of parents with low genetic diversity may quickly become prone to biotic and abiotic stresses (Govindaraj et al., 2015; Joukhadar et al., 2017). Conversely, using diverse parental lines or genotypes can produce progenies of desirable genetic makeup that have the tolerance to biotic and abiotic stresses, and that produce higher grain yields (Tar'an et al., 2005).

Agronomic and morphological data have been widely used to screen wheat varieties that are tolerant to stress, including drought (Ali et al., 2013), rust (Singh et al., 2005; Afzal et al., 2008; Luo et al., 2009; Chen et al., 2014), salinity (Zafar et al., 2015), and spot blotch (Jamil et al., 2018). Molecular markers were extensively used to evaluate the genetic diversity and population structure of wheat germplasm (Du et al., 2002; Khan et al., 2005; Ahmed et al., 2010; Akhunov et al., 2010; Sobia et al., 2010; van Poecke et al., 2013; Manickavelu et al., 2014; Zeshan et al., 2016). Studies using randomly amplified polymorphic DNA (RAPD) markers demonstrate narrow

genetic backgrounds in most varieties introduced by the same research institutes (Mukhtar et al., 2002; Ahmed et al., 2010). RAPD markers, however, can be problematic in terms of reproducibility and reliability, which can lead to inconsistent and/or weakly supported inferences (Penner et al., 1993). Single nucleotide polymorphisms (SNPs) are the most abundant polymorphism that exist in plant genomes (Batley and Edwards, 2007). SNPs are appropriate for investigating marker-trait association, analyzing genetic polymorphism, mapping quantitative trait loci (QTLs), studying population structure, and genomic selection. However, many SNPs are required to cover a significant part of the genome (Kumar et al., 2012). Recent advancements in high-throughput sequencing coupled with the introduction of the genotyping-by-sequencing (GBS) technique has made it possible to identify genome-wide SNPs in a cost effective manner. These SNPs are useful in crop breeding, DNA fingerprinting, tagging of resistance genes for biotic and abiotic factors, and analyzing genetic diversity (Elshire et al., 2011; Edae et al., 2014; He et al., 2014; Perea et al., 2016; Jamil et al., 2018, 2019). For genomic DNA digestion, the restriction endonucleases utilized in GBS reduce genomic complexity, thereby enabling easier analyses of large and complex genomes such as wheat. Wheat is an allohexaploid with 42 chromosomes and has a genome size up to 17 GB (Clavijo et al., 2017). Breeders can benefit from these cost-effective informative markers during the selection of desirable wheat offspring (Alipour et al., 2017).

Among top wheat-producing countries, Pakistan ranks 4th in Asia and 11th in the world (Saeed et al., 2012). To the best of our knowledge, genetic diversity in Pakistani wheat cultivars, advance lines, and landraces has not been evaluated using GBS markers. Here we report agro-morphological and yield data, along with GBS data in wheat germplasm from Pakistan. A schematic workflow of the overall study is given in **Figure 1A**. This data will be useful for inferring genetic diversity, population genetics, marker-assisted selection in breeding, genome-wide association studies (GWAS), mapping of rust and drought-resistant genes and other desirable quantitative trait loci (QTL) as well as for planning effective crop breeding programs in the future.

METHODS

Collection of Genotypes and Field Trial

A total of 104 wheat cultivars (CVs), landraces (LRs), and advance lines (ALs) were collected from different research institutes, breeders, and original collectors of landraces in Pakistan. An additional seven cultivars were collected from separate research institutes to be included as duplicate controls in agro-morphological data. A wild relative, *Triticum monococcum* (genotype ID: 209), was obtained from the Wide Hybridization Department, National Agriculture Research Center, Islamabad, and included in this study. **Supplementary Table 1** gives a list of all 112 genotypes (including seven duplicates) for which agro-morphological and yield data were recorded. This table also gives the NCBI sample accession numbers of a subset of 52 genotypes, which were used to generate GBS data. Among the 112 genotypes mentioned in this table, 55 cultivars were also reported in an online Wheat Atlas (http://wheatatlas.

org/country/varieties/PAK/0?AspxAutoDetectCookieSupport= 1; Accessed on 1st August 2019). **Supplementary Table 2** provides the detailed information about the year of release, pedigree and selection details for these cultivars, presence of the semi-dwarf (Rht) gene, and the area for which the cultivar was developed for these 55 common cultivars, as provided in the online Wheat Atlas. The field trial was conducted in a plain field in Mandra, a town located 45 km south of Islamabad, in the Potohar region (arid zone). The geographical coordinates for the site are 33°38′N, 73°26′E. Before sowing, the field was plowed, fertilizer was homogeneously mixed in the soil, and the soil was leveled. Seeds of the genotypes were sown from 15th November 2015 to 20th November 2015. Each genotype was sown in one square meter block, comprising 25 plants (5 rows × 5 columns) except for four genotypes for which <25 seeds per genotype were available (identified in **Supplementary Table 1**). The sixth row for all blocks comprised a rust spreader cultivar, called Morocco. The genotypes were sown in triplicate, in randomized blocks. **Figure 1B** gives a snapshot of the field trial.

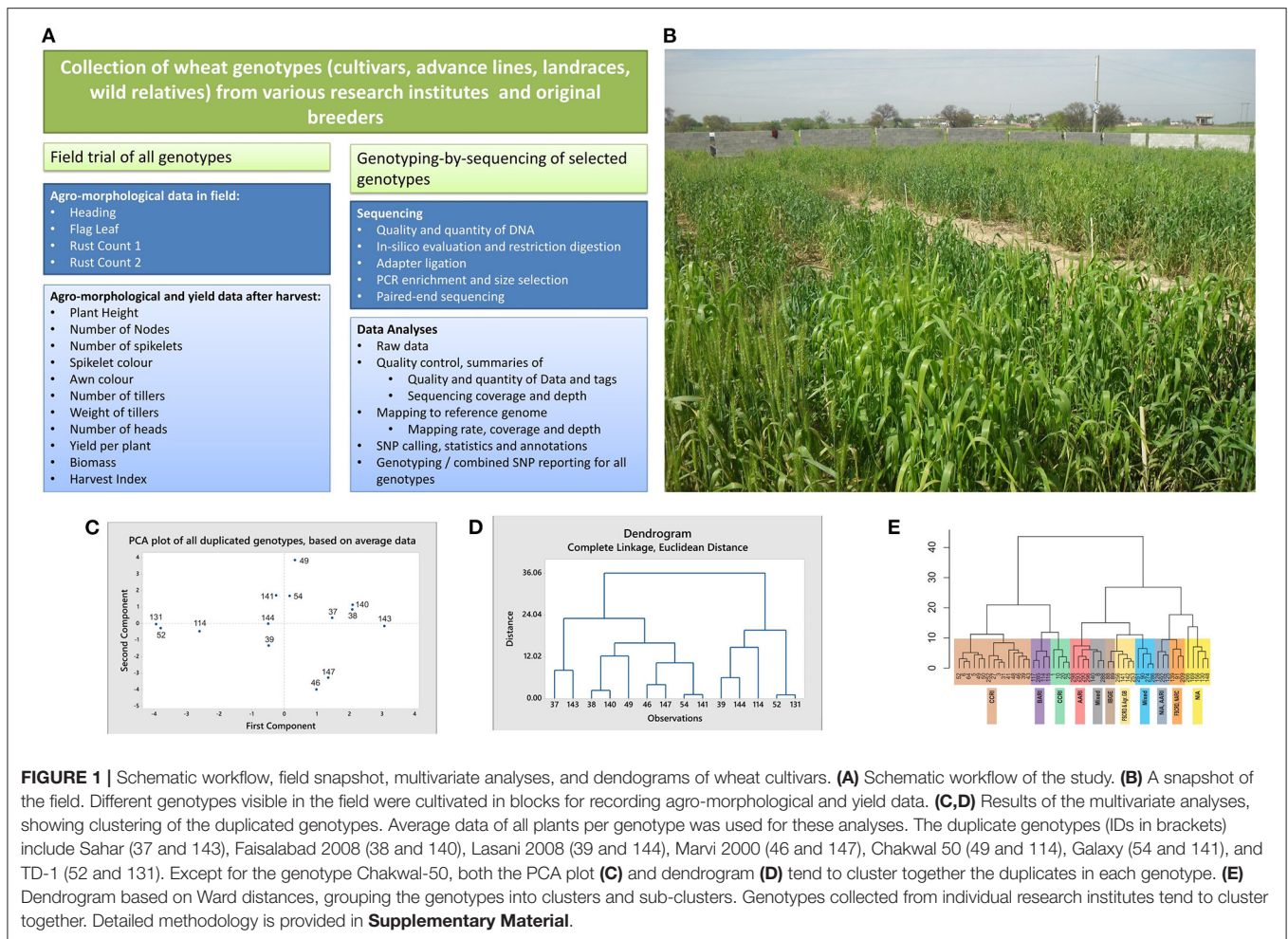## Agro-Morphological and Yield Data

Data were recorded in the field as well as after harvest. The field data consists of four qualitative variables. This data was based on the observation and scoring of data of entire blocks; individual plants were given the same score as that of the block for these four variables. At maturity, five plants per block were uprooted from the soil and labeled individually. The plant labeling after the harvest followed the EnRnPn scheme, where "E" showed "Entry" number (1–300 unique genotype IDs among 112 genotypes, as given in **Supplementary Table 1**), "R" represented replicate number (1–3), and "P" indicated plant number (1–5). For example, E1R2P5 represents entry (genotype ID) number 1, replicate number 2, and plant number 5. This labeling scheme ensured keeping identity of the plants while recording the subsequent qualitative and quantitative data. With few exceptions mentioned below, agro-morphological and yield data were recorded for 15 individual plants (five plants per replicate, in triplicates) per genotype.

Data Recorded in the Field

The traits or agro-morphological variables for which qualitative data were recorded in the field included heading (H), flag leaves (FL), rust count 1 (RC1), and rust count 2 (RC2). Heading data were recorded at the booting stage for most of the plants in the field, and all data were recorded in a single field visit. The data were scored as 1–8, based on the presence or absence of heads on most of the plants in the entire block. Flag leaf status was recorded as drooping to erect for the entire block and given scores as 1–4. Stripe rust was scored on a scale from 0–9, as reported elsewhere (Dinglasan et al., 2016). Stripe rust was scored twice; first count (RC1) was recorded 29th March 2016 and the second count (RC2) was recorded on 15th April 2016.

Data Recorded After Harvesting Plants at Maturity

After maturation, harvesting of the plants started on 30th April 2016 and continued till 15th May 2016. Most of the genotypes

FIGURE 1 | Schematic workflow, field snapshot, multivariate analyses, and dendograms of wheat cultivars. **(A)** Schematic workflow of the study. **(B)** A snapshot of the field. Different genotypes visible in the field were cultivated in blocks for recording agro-morphological and yield data. **(C,D)** Results of the multivariate analyses, showing clustering of the duplicated genotypes. Average data of all plants per genotype was used for these analyses. The duplicate genotypes (IDs in brackets) include Sahar (37 and 143), Faisalabad 2008 (38 and 140), Lasani 2008 (39 and 144), Marvi 2000 (46 and 147), Chakwal 50 (49 and 114), Galaxy (54 and 141), and TD-1 (52 and 131). Except for the genotype Chakwal-50, both the PCA plot **(C)** and dendrogram **(D)** tend to cluster together the duplicates in each genotype. **(E)** Dendrogram based on Ward distances, grouping the genotypes into clusters and sub-clusters. Genotypes collected from individual research institutes tend to cluster together. Detailed methodology is provided in **Supplementary Material**.

(CVs, ALs, and some LRs) were ready to harvest by the end of April; many LRs and some CVs were late in maturity and were harvested in the first and second week of May. Cold adapted LRs from the temperate region of Gilgit in northern Pakistan were the last to reach maturity. Fewer than five plants per block could be collected at maturity for these genotypes (sample IDs: 253, 255, and 256), leading to missing post-harvest data for the rest of the plants. Remaining plants for these genotypes did not reach maturity till the end of May 2016 (one month after the start of harvest) and were abandoned in the field. The following qualitative and quantitative data were recorded after the harvest:

Qualitative data were recorded for Spikelet color (SC) and Awn color (AC). The colors were scored either 1 (red to brown) or 2 (white to amber), as reported by elsewhere (Ormoli et al., 2015).

Quantitative data were recorded for nine variables, including Plant height (PH), Number of nodes (NN), Number of spikelets (NS), Number of tillers (NT), Weight of tillers (WT), Number of heads (NH), Yield per plant (YP), Biomass (B), and Harvest Index (HI). A brief description of each of the quantitative data recorded is given in **Supplementary Material**.

Genotyping by Sequencing

Based on economic importance, a sub-set of 52 genotypes (**Supplementary Table 3**) was selected to generate genotyping-by-sequencing (GBS) data. The varietal evenness for these 52 genotypes was based on the agro-morphological and yield data. Seeds were grown at room temperature in plastic trays (12 inches width × 24 inches length × 2.5 inches depth; 4 × 8 cells) using autoclaved soil and sand mixed 2:1. After 14 days of sowing, leaf tissues from 10 seedlings per sample were harvested and pooled for DNA extraction using the GeneJET Plant Genomic DNA kit (Catalog No. K0791, ThermoFisher Scientific USA). The quality and quantity of DNA were confirmed with 1% agarose gel electrophoresis and uDrop Plate of Multiskan GO (ThermoScientific, USA). DNA samples were lyophilized and shipped to Novogene Inc. Hong Kong for sequencing.

At Novogene, the purity and integrity of DNA were determined with agarose gel, and Qubit® 2.0 fluorometer was used for accurate quantification of DNA concentration. For library construction, all samples contained at least 1.5 ug DNA. MseI and NlaIII restriction endonucleases were selected after *in silico* evaluation to generate >400,000 tags per sample and were employed for digestion of DNA (0.3–0.6 ug). Adapters

were ligated to DNA along with a unique barcode for each wheat genotype. All libraries were pooled and subjected to a polymerase chain reaction (PCR) for the enrichment of sequence data. The qualified libraries were sequenced using Illumina high-throughput sequencing with 144 bp paired-end run. Average insert size of 303 bp was determined for each genotype, using Bioanalyzer.

The sequencing data was generated on a HiSeq 2500 instrument. Adapters were trimmed from the ends. Those reads which were either contaminated with library adapters, 10% unknown bases (N) or 50% low-quality bases were not used in downstream analysis. The quality of short reads was assessed using FastQC version 0.11.6 (Andrews et al., 2020) using default parameters. *Triticum aestivum* TGACv1 (Clavijo et al., 2017) was used as a reference genome for mapping the short reads using Burrows-Wheeler Alignment (BWA) version 0.7.1 (Li and Durbin, 2009) with default parameters. The reference genome was downloaded from Ensembl (ftp://ftp.ensemblgenomes. org/pub/release33/plants/fasta/triticum_aestivum/dna; File: Triticum_aestivum.TGACv1.dna.toplevel.fa.gz; date accessed 22nd March 2018).

All variants were filtered using SAMtools version 1.6 (Li et al., 2009) using parameters "−q = 1, −C = 50, −m = 2, −F = 0.002, −d = 1,000." PICARD version 2.18.0 (Broad Institute, 2018) was used to remove duplicates. To further reduce the error rate in substitutions calling, only those SNPs were selected that had coverage depth higher than 4x and mapping quality higher than 20. ANNOVAR (Wang et al., 2010) was used for the functional annotation of each substitution.

Data Records

The agro-morphological and yield data are presented in **Supplementary Table 3**. The table also provides information about the qualitative and quantitative data for 15 plants per genotype (five plants per plot, triplicates), along with the keys used for the qualitative data. **Supplementary Figure 1** is a Box-plot representation of the dispersion in the data for all 15 variables studied. Minitab version 18 was used to generate this figure.

All GBS sequencing data and associated BAM files have been submitted in Sequence Read Archive (SRA) of the NCBI repository (NCBI BioProject, 2019) and assigned SRA project number SRP179096. Individual Fastq files were given accession numbers SRR8441393 through SRR8441444; BAM files were given accession numbers SRR8467619 through SRR8467670. In total, 89.036 GB of clean data were produced; per sample data ranged from 1.01 to 2.5 GB. The lowest Phred score value for Q30 was 89.41%. The values of GC content in individual samples ranged from 42.14 to 44.17%. Information about individual samples, quantity, and quality of generated data are provided in **Supplementary Table 4** along with details of each wheat variety, numbers of bases generated per sample and their respective quality values. Reference genome mapping information is given in **Supplementary Table 5**. This table provides a summary statistic of the mapping of short reads to the wheat reference genome.

Supplementary Table 6 gives summary statistics about the variants called (SNPs) for individual genotypes. This table also gives functional attributes of the SNPs and gives the number of transition and transversion mutations. The average number of SNPs per genotype was 364,074 ± 54,479. When SNPs for all genotypes were merged, the total number of SNPs reached 2 Million. These combined SNPs, with exact nucleotide positions on the wheat reference genome, are given in the file "Genotyping and SNPs data" (Islam et al., 2020), available on Figshare. This file contains a complete record of SNPs. The data in each column can be read from left to right–#Chromosome: Chromosome position along the small arm and long arm of the chromosome, #Position: The coordinate position of nucleotide base which showed substitution, #Reference: The nucleotide present in the reference genome, #Allele: The type of substitution in the reference genome showing first the allele present in the reference genome and then the allele present in the sample sequence in the current study, #Gene: The name of the gene in which the substitution exists, #Annopos: Type of substitution according to the location, such as intergenic, genic, intronic, UTR, synonyms and non-synonyms. The next column shows the substitution present in each sample in a diploid form such that GG represents the homozygous condition and AG represents the heterozygous condition.

GBS data generated for various crops including wheat has been used to study genetic diversity, population genetics, phylogenetics (Lateef, 2015; Li et al., 2015; Chung et al., 2017; Elbasyoni et al., 2018), association mapping and genome-wide association studies (Bastien et al., 2014; Arruda et al., 2016; Muqaddasi et al., 2017; Yu et al., 2017; Jamil et al., 2018), linkage map and quantitative trait loci (QTL) mapping (Bielenberg et al., 2015; Verma et al., 2015; Balsalobre et al., 2017; Hussain et al., 2017; Scheben et al., 2017), marker-assisted and genomic selection (He et al., 2014; Scheben et al., 2017). Together with agro-morphological and yield data, GBS data generated for wheat genotypes in this study will be extremely useful in future crop breeding programs. The data will be helpful in the breeding of elite wheat cultivars having high yield and resistance to biotic and abiotic stresses to feed the growing human population.

Technical Validation

Seven cultivars were included as duplicate controls in the current study for technical validation. The dupicates were collected form different research institutes. Identity of the duplicates, their sources of collection, and description about their comparison in agro-morphological and yield data, as well as GBS data is provided **Figures 1C–E** and **Table 1**. Details about the technical validation for varietal evenness is provided in **Supplementary Material**.

Availability of the Wheat Genotypes

Sources of the wheat genotypes collection have been listed in **Supplementary Table 1**. The source institutes are expected to annually refresh and retain the propagating material, which is essential the viability of germplasm over the years. As per the Plant Breeders' Rights Act 2016 in Pakistan, original breeders

TABLE 1 | Pearson's correlations among alleles of the Chakwal-50 replicates (IDs 49 and 114) using GBS data.

Comparison among alleles	Pearson's correlation
Genotype 49, Allele 1, 2	0.997
Genotype 114, Allele 1, 2	0.996
Genotype 49, Allele 1; Genotype 114 Allele 1	0.493

of the cultivars and advance lines retain the property rights of their breeding material. In line with this Act, the authors are not authorized to share and disseminate the genotypes covered by the Act. The authors welcome queries from other researchers and potential breeders about the availability and sharing of the genotypes which are not protected by the Act. Where applicable, the respective laws of donor and recipient countries will govern the transfer of the propagating/living material to other countries outside Pakistan.

CODE AVAILABILITY

Except CASAVA, all software tools used are free to use and publicly available.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

MI: study design, sample collection, data recording in the field and after the harvest, data analyses, DNA extractions, and manuscript writing. Abdullah: data collection in the field and after the harvest, data analyses including GBS data, DNA extractions, and manuscript writing. BZ: sample collection and data analyses. NA: data collection after the harvest. RK and SW: data collection in the field and after the harvest. NS and RM: sample collection and data collection after the harvest. JT: GBS data analysis. IA: study design, field work, data collection, data analyses including GBS data, manuscript editing, and study co-supervision. MN: study design, sample collection, field work, data collection, data analyses, and study co-supervision. HA: study design and overall project supervision. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.617772/full#supplementary-material

REFERENCES

Afzal, S. N., Haque, M. I., Ahmedani, M. S., Rauf, A., Munir, M., Firdous, S. S., et al. (2008). Impact of stripe rust on kernel weight of wheat varieties sown in rainfed areas of Pakistan. *Pak. J. Bot.* 40, 923–929.

Ahmed, M. F., Iqbal, M., Masood, M. S., Rabbani, M. A., and Munir, M. (2010). Assessment of genetic diversity among Pakistani wheat (*Triticum aestivum* L.) advanced breeding lines using RAPD and SDS-PAGE. *Electron. J. Biotechnol.* 13, 1–10. doi: 10.2225/vol13-issue3-fulltext-2

Akhunov, E. D., Akhunova, A. R., Anderson, O. D., Anderson, J. A., Blake, N., Clegg, M. T., et al. (2010). Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC Genomics* 11:702. doi: 10.1186/1471-2164-11-702

Ali, A., Ali, N., Ullah, N., Ullah, F., Adnan, M., and Swati, Z. A. (2013). Effect of drought stress on the physiology and yield of the Pakistani wheat germplasms. *Int. J. Adv. Res. Technol.* 2, 419–430.

Alipour, H., Bihamta, M. R., Mohammadi, V., Peyghambari, S. A., Bai, G., and Zhang, G. (2017). Genotyping-by-Sequencing (GBS) revealed molecular genetic diversity of Iranian wheat landraces and cultivars. *Front. Plant Sci.* 8:1293. doi: 10.3389/fpls.2017.01293

Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., and Wingett, S. (2020). *FastQC. Babraham Bioinforma.* Available online at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed April 01, 2021).

Arruda, M. P., Brown, P., Brown-Guedira, G., Krill, A. M., Thurber, C., Merrill, K. R., et al. (2016). Genome-wide association mapping of *Fusarium* head blight resistance in wheat using genotyping-by-sequencing. *Plant Genome* 9, 1–14. doi: 10.3835/plantgenome2015.04.0028

Balsalobre, T. W. A., da Silva Pereira, G., Margarido, G. R. A., Gazaffi, R., Barreto, F. Z., Anoni, C. O., et al. (2017). GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. *BMC Genomics* 18:72. doi: 10.1186/s12864-016-3383-x

Bastien, M., Sonah, H., and Belzile, F. (2014). Genome wide association mapping of *Sclerotinia sclerotiorum* resistance in soybean with a genotyping-by-sequencing approach. *BMC Genomics* 7, 1–13. doi: 10.3835/plantgenome2013.10.0030

Batley, J., and Edwards, D. (2007). "SNP applications in plants," in *Association Mapping in Plants*, eds N. C. Oraguzie, E. H. A. Rikkerink, S. E. Gardiner, and H. N. De Silva (New York, NY: Springer), 95–102.

Bielenberg, D. G., Rauh, B., Fan, S., Gasic, K., Abbott, A. G., Reighard, G. L., et al. (2015). Genotyping by sequencing for SNP-based linkage map construction and QTL analysis of chilling requirement and bloom date in peach [*Prunus persica* (L.) Batsch]. *PLoS ONE* 10:e0139406. doi: 10.1371/journal.pone.0139406

Broad Institute (2018). *Picard toolkit. GitHub Repos.* Available online at: https://broadinstitute.github.io/picard/ (accessed April 01, 2021).

Chen, W., Wellings, C., Chen, X., Kang, Z., and Liu, T. (2014). Wheat stripe (yellow) rust caused by *Puccinia* striiformis f. sp. tritici. *Mol. Plant Pathol.* 15, 433–446. doi: 10.1111/mpp.12116

Chung, Y. S., Choi, S. C., Jun, T. H., and Kim, C. (2017). Genotyping-by-sequencing: a promising tool for plant genetics research and breeding. *Hortic. Environ. Biotechnol.* 58, 425–431. doi: 10.1007/s13580-017-0297-8

Clavijo, B. J., Venturini, L., Schudoma, C., Accinelli, G. G., Kaithakottil, G., Wright, J., et al. (2017). An improved assembly and annotation of the allohexaploid

wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Res.* 27, 885–896. doi: 10.1101/gr.217117.116

Dinglasan, E., Godwin, I. D., Mortlock, M. Y., and Hickey, L. T. (2016). Resistance to yellow spot in wheat grown under accelerated growth conditions. *Euphytica* 209, 693–707. doi: 10.1007/s10681-016-1660-z

Du, J.-K., Yao, Y.-Y., Ni, Z.-F., Peng, H.-R., and Sun, Q.-X. (2002). Genetic diversity revealed by ISSR molecular marker in common wheat, spelt, compactum, and progeny of recurrent selection. *Acta Genet. Sin.* 29, 445–452.

Edae, E. A., Byrne, P. F., Haley, S. D., Lopes, M. S., and Reynolds, M. P. (2014). Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. *Theor. Appl. Genet.* 127, 791–807. doi: 10.1007/s00122-013-2257-8

Elbasyoni, I. S., Lorenz, A. J., Guttieri, M., Frels, K., Baenziger, P. S., Poland, J., et al. (2018). A comparison between genotyping-by-sequencing and array-based scoring of SNPs for genomic prediction accuracy in winter wheat. *Plant Sci.* 270, 123–130. doi: 10.1016/j.plantsci.2018.02.019

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379. doi: 10.1371/journal.pone.0019379

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., et al. (2010). Food secrity: the challenge of feeding 9 billion people. *Science* 327, 812–819. doi: 10.1126/science.1185383

Govindaraj, M., Vetriventhan, M., and Srinivasan, M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genet. Res. Int.* 2015:431487. doi: 10.1155/2015/431487

He, J., Zhao, X., Laroche, A., Lu, Z.-X., Liu, H., and Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front. Plant Sci.* 5, 1–8. doi: 10.3389/fpls.2014.00484

Hussain, W., Baenziger, P. S., Belamkar, V., Guttieri, M. J., Venegas, J. P., Easterly, A., et al. (2017). Genotyping-by-sequencing derived high-density linkage map and its application to QTL mapping of flag leaf traits in bread wheat. *Sci. Rep.* 7:16394. doi: 10.1038/s41598-017-16006-z

Islam, M., Abdullah, Zubaida, B., Shafqat, N., Masood, R., Khan, U., et al. (2020). *Data From: Agro-Morphological, Yield, and Genotyping-by-Sequencing Data of Selected Wheat Germplasm*. Available online at: https://figshare.com/s/971830c580b2a324bc8d (accessed April 01, 2021).

Jamil, M., Ali, A., Gul, A., Ghafoor, A., Ibrahim, A. M. H., and Mujeeb-Kazi, A. (2018). Genome-wide association studies for spot blotch (*Cochliobolus sativus*) resistance in bread wheat using genotyping-by-sequencing. *Phytopathology* 108, 1307–1314. doi: 10.1094/PHYTO-02-18-0047-R

Jamil, M., Ali, A., Gul, A., Ghafoor, A., Napar, A. A., Ibrahim, A. M. H., et al. (2019). Genome-wide association studies of seven agronomic traits under two sowing conditions in bread wheat. *BMC Plant Biol.* 19:149. doi: 10.1186/s12870-019-1754-6

Joukhadar, R., Daetwyler, H. D., Bansal, U. K., and Gendall, A. R. (2017). Genetic diversity, population structure, and ancestral origin of Australian wheat. *Front. Plant Sci.* 12:2115. doi: 10.3389/fpls.2017.02115

Khan, I. A., Awan, F. S., Ahmad, A., and Fu, Y. (2005). Genetic diversity of Pakistan wheat germplasm as revealed by RAPD markers. *Genet. Resour. Crop Evol.* 52, 239–244. doi: 10.1007/s10722-004-5730-z

Khan, M. K., Pandey, A., Thomas, G., Akkaya, M. S., Kayis, S. A., Ozsensoy, Y., et al. (2015). Genetic diversity and population structure of wheat in India and Turkey. *AoB Plants* 7:plv083. doi: 10.1093/aobpla/plv083

Kumar, S., Banks, T. W., and Cloutier, S. (2012). SNP Discovery through next-generation sequencing and its applications. *Int. J. Plant Genomics* 2012:831460. doi: 10.1155/2012/831460

Lateef, D. D. (2015). DNA marker technologies in plants and applications for crop improvements. *J. Biosci. Med.* 03, 7–18. doi: 10.4236/jbm.2015.35002

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, H., Vikram, P., Singh, R. P., Kilian, A., Carling, J., Song, J., et al. (2015). A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. *BMC Genomics* 16:216. doi: 10.1186/s12864-015-1424-5

Luo, P., Hu, X., Zhang, H., and Ren, Z. (2009). Genes for resistance to stripe rust on chromosome 2B and their application in wheat breeding. *Prog. Nat. Sci.* 19, 9–15. doi: 10.1016/j.pnsc.2008.02.017

Manickavelu, A., Jighly, A., and Ban, T. (2014). Molecular evaluation of orphan Afghan common wheat (*Triticum aestivum* L.) landraces collected by Dr. Kihara using single nucleotide polymorphic markers. *BMC Plant Biol.* 14:320. doi: 10.1186/s12870-014-0320-5

Mukhtar, M. S., Rahman, M., and Zafar, Y. (2002). Assessment of genetic diversity among wheat (*Triticum aestivum* L.) cultivars from a range of localities across Pakistan using random amplified polymorphic DNA (RAPD) analysis. *Euphytica* 128, 417–425. doi: 10.1023/A:1021261811454

Muqaddasi, Q. H., Reif, J. C., Li, Z., Basnet, B. R., Dreisigacker, S., and Röder, M. S. (2017). Genome-wide association mapping and genome-wide prediction of anther extrusion in CIMMYT spring wheat. *Euphytica* 213:73. doi: 10.1007/s10681-017-1863-y

NCBI BioProject (2019). *Genotyping by Sequencing of Wheat Germplasm From Pakistan to Elucidate Genetic Diversity*. Available online at: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA514955/ (accessed April 01, 2021).

Ormoli, L., Costa, C., Negri, S., Perenzin, M., and Vaccino, P. (2015). Diversity trends in bread wheat in Italy during the 20th century assessed by traditional and multivariate approaches. *Sci. Rep.* 5:8574. doi: 10.1038/srep08574

Penner, G. A., Bush, A., Wise, R., Kim, W., Domier, L., Kasha, K., et al. (1993). Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. *Genome Res.* 2, 341–345. doi: 10.1101/gr.2.4.341

Perea, C., De La Hoz, J. F., Cruz, D. F., Lobaton, J. D., Izquierdo, P., Quintero, J. C., et al. (2016). Bioinformatic analysis of genotype by sequencing (GBS) data with NGSEP. *BMC Genomics* 17:498. doi: 10.1186/s12864-016-2827-7

Philipp, N., Weise, S., Oppermann, M., Börner, A., Keilwagen, J., Kilian, B., et al. (2019). Historical phenotypic data from seven decades of seed regeneration in a wheat ex situ collection. *Sci. Data* 6:137. doi: 10.1038/s41597-019-0146-y

Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS ONE* 8:e66428. doi: 10.1371/journal.pone.0066428

Saeed, B., Gul, H., Shah, P., Khan, A., Anwar, S., and Ali, S. (2012). Yield of wheat varieties under solid and skip row geometries. *ARPN J. Agric. Biol. Sci.* 7, 591–594.

Scheben, A., Batley, J., and Edwards, D. (2017). Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnol. J.* 15, 149–161. doi: 10.1111/pbi.12645

Shewry, P. R., and Hey, S. J. (2015). The contribution of wheat to human diet and health. *Food Energy Secur.* 4, 178–202. doi: 10.1002/fes3.64

Singh, R. P., Huerta-Espino, J., and William, H. M. (2005). Genetics and breeding of durable resistance to leaf and stripe rusts in wheat. *Turkish J. Agric. For.* 29, 121–127.

Sobia, T., Muhammad, A., and Chen, X. (2010). Evaluation of Pakistan wheat germplasms for stripe rust resistance using molecular markers. *Sci. China Life Sci.* 53, 1123–1134. doi: 10.1007/s11427-010-4052-y

Tar'an, B., Zhang, C., Warkentin, T., Tullu, A., and Vandenberg, A. (2005). Genetic diversity among varieties and wild species accessions of pea (*Pisum sativum* L.) based on molecular markers, and morphological and physiological characters. *Genome* 48, 257–272. doi: 10.1139/g04-114

van Poecke, R. M. P., Maccaferri, M., Tang, J., Truong, H. T., Janssen, A., van Orsouw, N. J., et al. (2013). Sequence-based SNP genotyping in durum wheat. *Plant Biotechnol. J.* 11, 809–817. doi: 10.1111/pbi.12072

Verma, S., Gupta, S., Bandhiwal, N., Kumar, T., Bharadwaj, C., and Bhatia, S. (2015). High-density linkage map construction and mapping of seed trait QTLs in chickpea (*Cicer arietinum* L.) using Genotyping-by-Sequencing (GBS). *Sci. Rep.* 5:17512. doi: 10.1038/srep17512

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38:e164. doi: 10.1093/nar/gkq603

Yu, L.-X., Zheng, P., Zhang, T., Rodringuez, J., and Main, D. (2017). Genotyping-by-sequencing-based genome-wide association studies on Verticillium wilt resistance in autotetraploid alfalfa (*Medicago sativa* L.). *Mol. Plant Pathol.* 18, 187–194. doi: 10.1111/mpp.12389

Zafar, S., Ashraf, M. Y., Niaz, M., Kausar, A., and Hussain, J. (2015). Evaluation of wheat genotypes for salinity tolerance using physiological indices as screening tool. *Pakistan J. Bot.* 47, 397–405.

Zeshan, A., Afzal, M., Alghamdi, S. S., Kettener, K., Ali, M., Mubushar, M., et al. (2016). Evaluation of genetic diversity among the Pakistani wheat (*Triticum aestivum* L.) lines through random molecular markers. *Brazilian Arch. Biol. Technol.* 59:e16160282. doi: 10.1590/1678-4324-2016160282