

Global diversity analysis identifies ample space for using CIMMYT and ICARDA’s wheat germplasm collections in breeding programs

Carolina Sansaloni¹*, Jorge Franco², Bruno Santos³, Mario Caccamo³, Sukhwinder Singh¹, Cesar Petroli¹, Thomas Payne¹, David Marshall⁴, Benjamin Kilian⁵, Carolina Saint Pierre¹, José Crosa¹, Andrzej Kilian⁶, Peter Wenzl⁷, Mohamed Fawzy Nawar⁸, Kai Sonder¹, Jaime Campos¹, Prashant Vikram¹, Kevin Pixley¹

¹ CIMMYT- International Wheat and Maize Improvement Center, Mexico; ²Universidad de la República del Uruguay, Uruguay; ³ NIAB- National Institute of Agricultural Botany, UK; ⁴ James Hutton Institute, UK; ⁵ The Crop Trust, Germany; ⁶ Diversity Arrays Technology, Australia; ⁷ CIAT- International Center for Tropical Agriculture; ⁸ ICARDA- International Center for Agricultural Research in the Dry Areas.
* c.sansaloni@cgiar.org

ABSTRACT

Climate change and the growing human population trigger the demand to explore and use genetic resources. Wheat’s undomesticated wild species, crop wild relatives, and landraces have survived extreme environmental challenges and continuous cycles of natural selection, and therefore offer a wealth of variation for cultivar improvement.

CIMMYT’s Seeds of Discovery (Seed)ⁱ initiative characterized more than 100,000 CIMMYT and ICARDA germplasm bank accessions with 130,000-450,000 molecular markers. Diversity analyses successfully grouped 4,206 wild relatives according to species and sub-species. Elite breeder lines and cultivars were mostly tightly clustered and were much less diverse than landraces and unimproved accessions both among 20,000 tetraploid (AB genome), and 60,000 hexaploid (ABD genome) accessions.

This detailed analysis enabled the formation of core sub-sets that encompass the total diversity of the germplasm bank collections while greatly reducing the numbers of materials that researchers must evaluate in search of needed diversity. This work is likely the largest crop genotyping effort to date, generating resources to underpin the breakthroughs necessary to develop the crops for the future.

RESULTS

Diversity analyses:

- **Wild relatives:** 3D representation of the multidimensional scaling (using DArTseq SNPs) clearly identified (clustered) species and even subspecies (Fig. 1A).
- **Tetraploids:** although breeding programs visibly reduced diversity, a few elite breeding lines appear to maintain a wide range of the diversity found among landrace materials (Fig. 2A).
- **Hexaploids:** Landraces, elite breeding lines and synthetic derivatives formed genetically distinct groups (Fig. 2C).

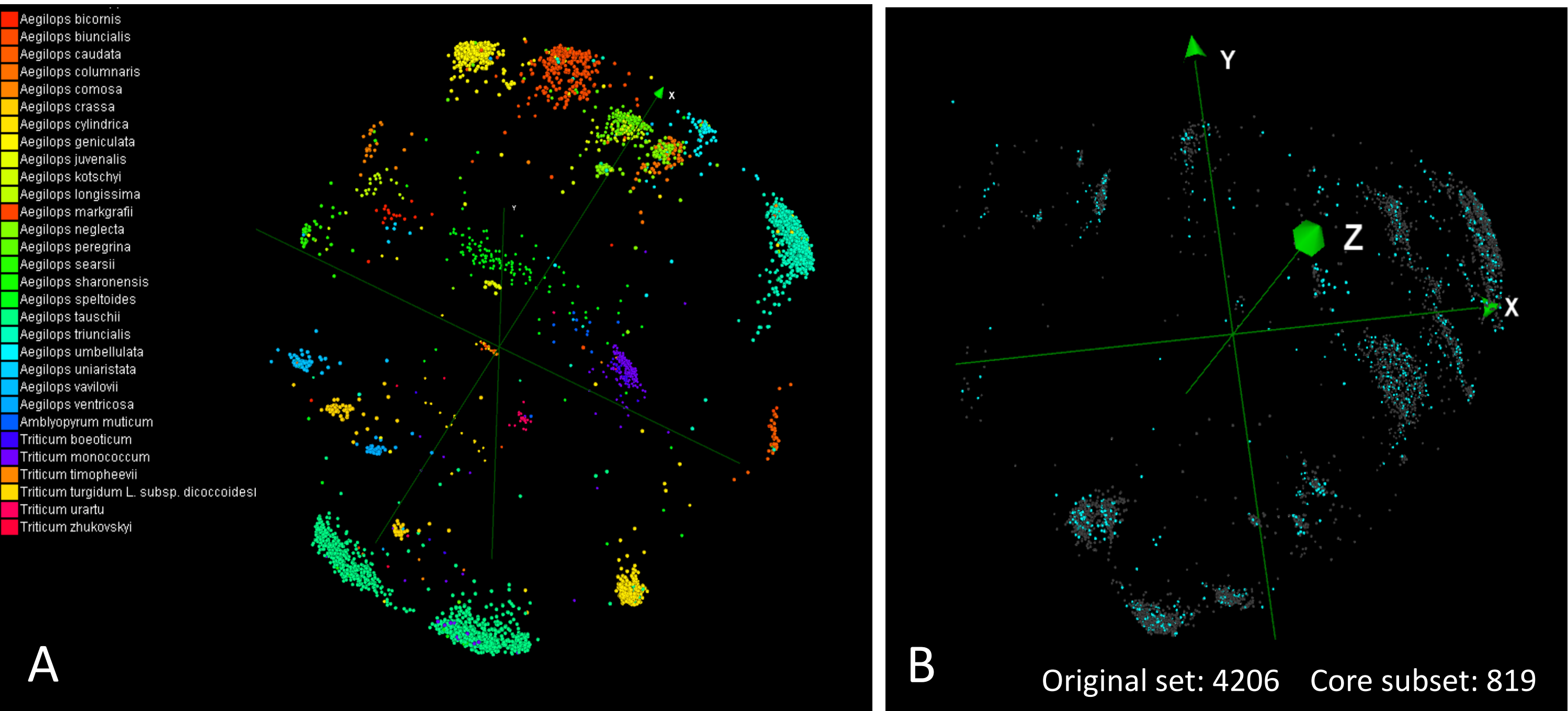


Figure 1 . Multidimensional scaling representation of A) 4,206 Wild relatives wheat accessions; B) Core subset with 819 accessions.

Core germplasm sub-sets:

Understanding the global diversity is key to developing core sub-sets that capture most of the allele richness and genetic diversity of the overall germplasm groups. Core sub-sets are a valuable starting point for researchers searching for novel diversity, donor germplasm for pre-breeding programs (Fig. 1B, 2B and 2D).

The use of genetic resources stored in germplasm banks requires an extensive, long-term, scientific and financial commitment, involving (Fig. 3):

- Studying the global genetic diversity of the germplasm bank accessions, as reported herein,
- Identifying novel diversity for traits of economic importance, and
- Introgressing useful, novel diversity into elite cultivars via germplasm enhancement programs.

These two last points are pursued by Seed’s pre-breeding program.

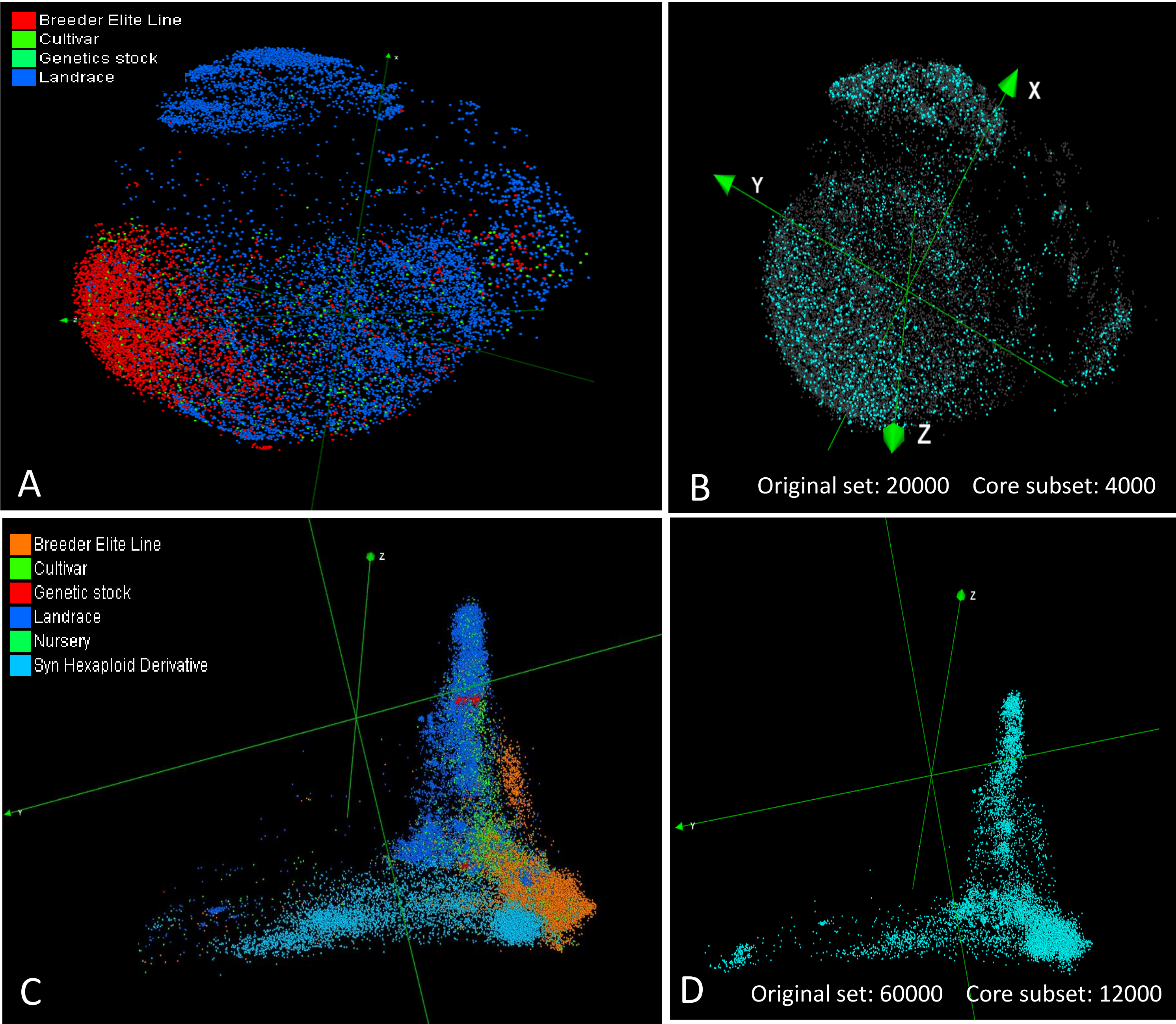


Figure 2 . Diversity study representation of A) 20,000 tetraploid cultivated accessions with B) the core subset collection, and C) 60,000 hexaploid cultivated accessions with D) the core subset collection.

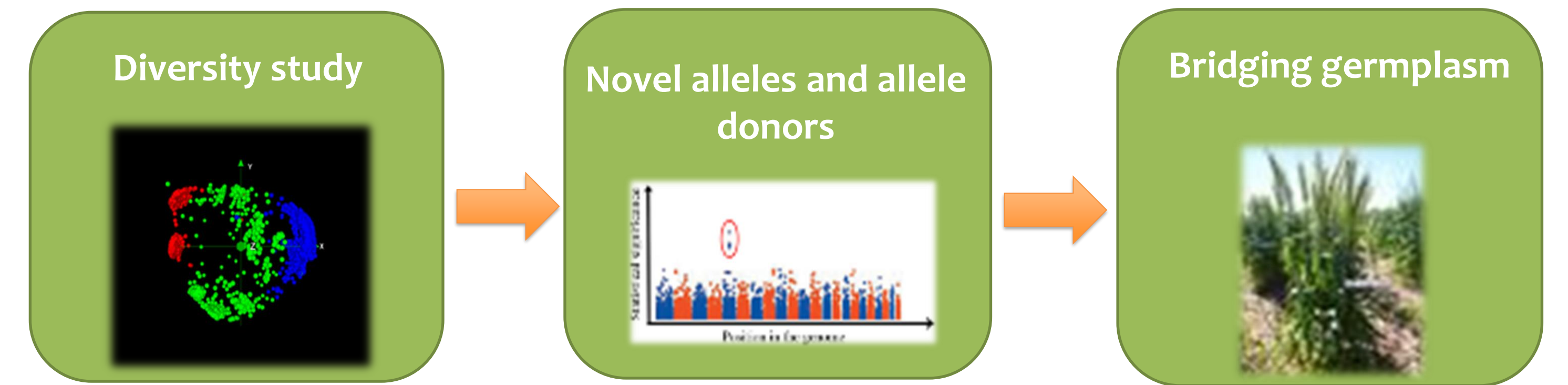


Figure 3 . Scheme showing the utilization of genetic resource strategy

CONCLUSION

Diversity analyses revealed that tetraploid, and especially hexaploid elite wheat breeding lines and cultivars, are genetically narrow and encompass a small portion of the total genetic diversity relative to their respective landraces and unimproved germplasm. This suggests ample potential to explore germplasm bank accessions for useful diversity to enrich the elite germplasm pools, particularly to address new challenges to production.

The curated data, as well as links to passport and characterization/evaluation data, are or will be available (www.seedsofdiscovery.org) for use by the larger community. We are keen to partner with interested researchers and genebanks to further leverage the value of existing data, and to generate new data, knowledge and value for breeding programs.

METHODS

- The CIMMYT and ICARDA wheat germplasm bank collections were divided into three groups for diversity analyses:
 - Wild relatives (WR): 4,206 accessions from 30 species with 20 genomes: A, B, D, C, M, S, U, N, T, UC, DN, GA, AB, UM, DM, DC, MU, SU, UMX, DMS, DMU, GAA.
 - Tetraploids (domesticated species): 20,000 accessions from 10 species; most of them *Triticum turgidum* subsp *durum*, but also *Triticum turgidum* subsp *turgidum*, *dicocon*, *carthlicum*, *turanicum*, *polonicum*, *paleocolchicum* and *Triticum aethiopicum*
 - Hexaploids (domesticated species): 60,000 accessions from 7 species; most of them are *Triticum aestivum* L. *aestivum*, but also, *T. aestivum* L. *spelta*, *compactum sphaerococcum*, *macha*, *Triticoaegilops* and *Triticosecale*.
- DArTseqTM technology revealed 450,000, 130,000 and 200,000 high-quality SNP and SilicoDArT (presence/absence variations) markers for WR, tetraploid and hexaploid sets respectively.
- All markers were aligned to the IWGSC v0.4 reference genomeⁱⁱ and against the DArT consensus mapⁱⁱⁱ.
- Core sub-sets containing 20% of the accessions in each germplasm group were formed based exclusively on genetic information.