

Genetic Sourcing of Wild Relatives for Crop Improvement in the Super Pangenomic Era



Association Studies of Super-pangenome, Gene Expression and Trait (ASSET)



PLANT
GENOMICS



CROP
IMPROVEMENT

SUMMARY

- A series of recent advances have created a new paradigm for mining the genetic diversity from large crop wild relative (CWR) populations.
 - I. Progress made in long-read sequencing and whole-genome mapping is revealing the hidden genetic diversity in CWRs through the constructions of high-quality super-pangenomes that capture the full spectrum of genetic variants (SNPs + indels + SVs) across large CWR populations.
 - II. Tools have been developed to accurately associate super-pangenomic data with gene expression and phenotyping data in order to identify specific genetic variants that can be used for accelerated crop improvement of modern cultivars.
 - III. Molecular tools are being developed to allow for accelerated and more precise trait/genetic variants introgression, making mining efforts in the more distant CWRs more valuable for plant breeders.
- Super-pangenome association studies promise to significantly accelerate the improvement of under-studied crops.
- BTI scientists have demonstrated their ability to build complex super-pangenomes and perform genomic/transcriptomic/phenotypic data association studies (ASSET) and propose to use this expertise for the discovery of novel genes and genetic variants for the rapid improvement of agronomically important crops that do not usually benefit from such advanced research tools

Background Information

Crop Wild Relatives

Crop Wild Relatives are wild species closely related to a domesticated crop that have remained adapted to various ecological environments, harboring broad genetic diversity which is not present in cultivated varieties. Hence, these wild species represent a rich and (mostly) untapped source of genes underlying biotic and abiotic stress tolerance, as well as increased nutritional quality.

Super-pangenomes provide a complete genomic variation repertoire of a genus, making previously hidden genetic variants accessible to genetic studies and breeding. Super-pangenome analyses can identify the dispensable genome which consists of genetic components only present in certain CWRs and responsible for adaptation and survival in different environments. Many agronomically important genes in plant species are found in the dispensable genome, which therefore represents a promising avenue for introducing beneficial genetic variants into important crop species.

Structural Variants (SVs) are defined as genomic variations larger than 30–50 bp and include changes caused by deletions, insertions, copy number variations, inversions, translocations or fusions. They have been shown to have strong influence on gene structure and expression, and are considered to be playing a critical role in the genetic control of agronomically important traits. In tomato for instance, the largest proportion of sequence differences among cultivars is ascribed to SVs and they often serve as the sole arbiters of changes in phenotypic expression among diverse accessions.

Super-pangenomes

Structural Variants

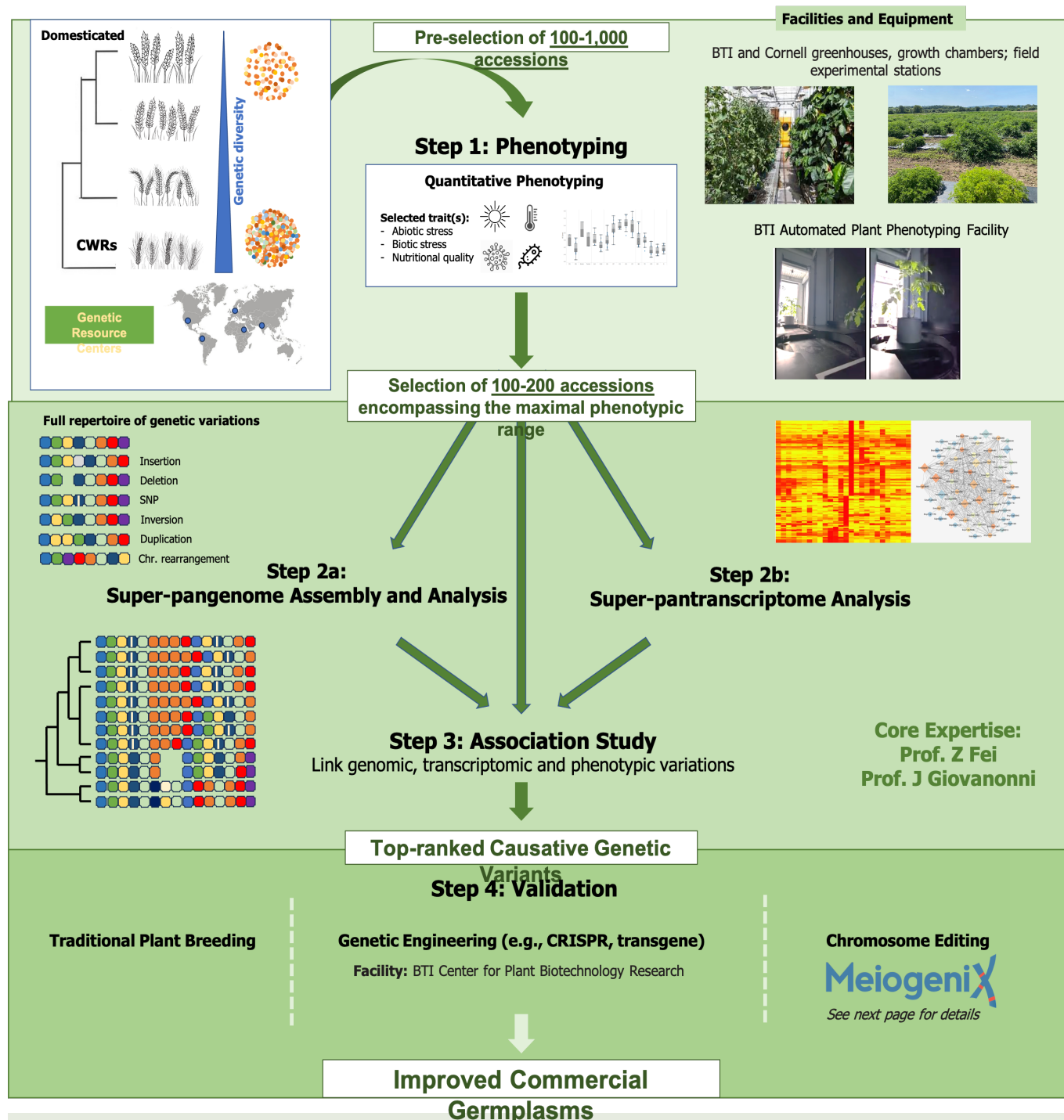


Figure Legend:

Super-pangenome, expression and phenotypic association studies can rapidly accelerate germplasm improvement by tapping into the genetic diversity of CWRs in 4 steps:

Step 1, consisting of a phenotypic screen of hundreds of accessions for a given trait (e.g., resistance to water stress) is followed in **Step 2a** by the assembly and analysis of the super-pangenome for a group of 100–200 accessions selected based on their phenotypic variation for the studied trait (e.g., cluster of accessions capturing the full range of resistance/susceptibility phenotype). Gene expression data (**Step 2b**) is generated in parallel by RNA-seq for the selected accessions (to compare expression profiles in control and water-stressed plants).

Data from Step 1, 2a and 2b is then processed in **Step 3** for an association study that will identify and rank candidate genetic variants.

Step 4 consists of an *in vivo* validation of selected candidates. Traditional plant breeding can be challenging when valuable variants from distant CWRs are identified and/or if the identified genetic region has a strong linkage drag. Gene editing or chromosome editing represent powerful alternatives to traditional plant breeding.

Note: Elements of this figure are based on Khan et al., (Trends Plant Sci 25:148–158, 2020) and Jayakodi et al., (DNA Res. 28:1–9, 2021).

Advantages Over Traditional Breeding Approaches

I. Capturing the Full Scope of Genetic Diversity Across Broader CWR Populations

Traditionally, SNPs identified relative to a single reference genome have been used for genomic selection. In comparison, super-pangenomes offer a comprehensive understanding of the entire genetic diversity within crop populations, as it captures the entire spectrum of genetic variants across the studied CWR accessions and reveals novel alleles and gene structures that are missed by traditional approaches.

II. Precise Mining of Causative Genetic Variants

The ASSET super-pangenome approach can identify any type of causative genetic variant across a large set of CWRs with great precision. In comparison, traditional approaches have a limited accuracy and are usually not able to achieve gene level resolution: mapping-based approaches identify larger genomic regions and often fail to pinpoint the specific genetic variant causing the desired phenotype.

III. Accelerated and High Precision Crop Improvement

By directly identifying causative genetic variants, the ASSET super-pangenome approach allows for precise and targeted breeding strategies while reducing significantly the time and resources needed for multiple rounds of conventional breeding (mapping crosses...). This approach can guide precise gene editing effort to modify or (re)create structural variations in cultivated germplasms to optimize gene function and trait expression. It also facilitates the introgression of multiple traits simultaneously in a single breeding line (stacking for the development of superior cultivars), and the application of the discovery to other crops.

IV. Discovery of New Underlying Mechanisms

The ASSET super-pangenome approach does not require prior knowledge to identify causative genetic variants, and can therefore lead to the discovery of novel underlying mechanisms.

Partnership with MeioGenix, Inc.



Opportunity:

Although CWRs have already provided breeders with several “game-changing” traits, incorporation of genes from CWRs into the cultivated gene pool is not straightforward. Breeders are reluctant to use CWRs in commercial breeding programs owing to the challenges presented by hybridization barriers and linkage drag causing poor agronomic performance and a wide range of phenotyping challenges. Recent advances in genome editing technologies, such as groundbreaking chromosome editing technologies developed by the Cornell-based start-up MeioGenix, are creating new opportunities for overcoming several of these problems.

BTI and MeioGenix scientists are partnering to develop a pipeline that can efficiently discover and introgress valuable genetic variants from distant CWRs into cultivated germplasms, thus offering unprecedented opportunities for crop improvement.

[Click here to visit the MeioGenix website.](#)

Key References – work involving BTI scientists is highlighted in bold

Research Articles

Alonge et al., *Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato*. *Cell* 182, 145–161 (2020).

Gao et al., *The Tomato Pan-Genome Uncovers New Genes and a Rare Allele Regulating Fruit Flavor*. *Nat Genet* 51, 1044–1051 (2019).

Li et al., *Graph-Based Pan-Genome Reveals Structural and Sequence Variations Related to Agronomic Traits and Domestication in Cucumber*. *Nat Commun* 13, 682 (2022)

Li et al., *Super-Pangenome Analyses Highlight Genomic Diversity and Structural Variation Across Wild and Cultivated Tomato Species*. *Nat Genet* 55, 852–860 (2023).

Wang et al., *Genome of Solanum pimpinellifolium Provides Insights Into Structural Variants During Tomato Breeding*. *Nat Commun* 11, 5817 (2020).

Wu et al., *A Citrullus Genus Super-Pangenome Reveals Extensive Variations in Wild and Cultivated Watermelons and Sheds Light on Watermelon Evolution and Domestication*. *Plant Biotechnol J* doi: 10.1111/pbi.14120 (2023).

Zhou et al., *Graph Pangenome Captures Missing Heritability and Empowers Tomato Breeding*. *Nature* 60, 527–534 (2022).

Reviews

Bayer et al., *Plant Pan-Genomes are the New Reference*. *Nat Plants* 6, 914–920 (2020).

Della Coletta et al., *How the Pan-Genome is Changing Crop Genomics and Improvement*. *Genome Biol* 22, 3 (2021).

Khan et al., *Super-Pangenome by Integrating the Wild Side of a Species for Accelerated Crop Improvement*. *Trends Plant Sci* 25:148–158 (2020).

Video

Profs. Zhangjun Fei and Jim Giovannoni discuss their creation and study of the tomato pan-genome. Thanks to the pan-genome, “lost” genes encoding traits like flavor, disease resistance and stress tolerance can be discovered and reintroduced into tomatoes (watch on Youtube)

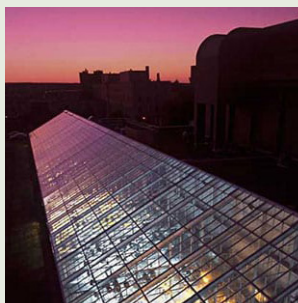
Contact:

Julien Fey, Director of the BTI Translational Program
jpf23@cornell.edu

Web:

btiscience.org/our-research/

Boyce Thompson Institute for Plant Research
533 Tower Rd.
Ithaca, NY 14853



Inspired by how plants work for humanity.

BTI's mission: To advance and communicate scientific discovery in plant biology to improve agriculture, protect the environment, and enhance human health.