Al-Farabi Kazakh National University (KazNU)

Faculty of Biology and Biotechnology

DISCIPLINE: «Modern Problems of Plant Genetics»

Lecture 2

Challenging Features of Plant Genomes.

Amirova Aigul Kuzembayevna

Associated professor,
Candidate of Biological Science

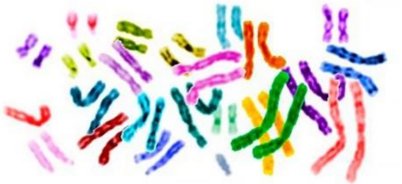
aigul_amir@mail.ru

***Purpose of the lesson*:**
Acquaintance with the major achievements and problems in plant genomics.

Main questions:

1. **Plant pan-genomics: recent advances, new challenges, and roads ahead.**
2. **Characterizing genetic diversity and variation.**
3. **Advancing crop improvement and breeding.**
4. **Challenges for sustainable agriculture production.**
5. **Major achievements in plant pan-genomics.**

MODERN PROBLEMS OF PLANT GENETICS



- The plant kingdom has amazing diversity and importantly provides a variety of resources and food energy intake for humans (Food and Agriculture Organization of the United Nations, 1995).
- The estimated number of land plant species is approximately 391,000, and their genomes are unusually diverse and complicated with genome sizes that vary dramatically from approximately 60 Mb to 150 Gb.

- Polyploidization events and variations in the amounts of repetitive DNA have played important roles in influencing the different sizes of plant genomes, which are vital to plant speciation and evolution.

- The dynamics of transposable elements, along with self-incompatibility, have long been recognized as significant evolutionary forces that contribute to plant genome changes.

- **MODERN PROBLEMS OF PLANT GENETICS**

- **All of these properties, high repetitive DNA content, high degree of heterozygosity, and polyploidy, make it technically challenging and time-consuming to generate high-quality plant genome assemblies.**

- **High-quality reference genome sequences are the prerequisite and basis for promoting fundamental and applied research in plants and animals. Triggered by developments in computing power, sequencing technologies, and assembly methods, the genomes of more than 700 plants species, from non-vascular to flowering, have been released in the past 20 years (Sun et al., 2021)**

- **Plant pan-genomics: recent advances, new challenges, and roads ahead**

- **High-quality reference genome sequences are the prerequisite and basis for promoting fundamental and applied research in plants and animals. Triggered by developments in computing power, sequencing technologies, and assembly methods, the genomes of more than 700 plants species, from non-vascular to flowering, have been released in the past 20 years (Sun et al., 2021).**

- **Third-generation sequencing technologies, such as those that use the PacBio and Oxford Nanopore platforms, can generate reads with significantly increased lengths and they have been widely applied along with well-established assembly algorithms to construct large and complicated plant genomes at unprecedented high resolution (Koren et al., 2017; Cheng et al., 2021; Niu et al., 2022).**

- Plant pan-genomics: recent advances, new challenges, and roads ahead

- The construction of high-quality assemblies of these very large and complicated plant genomes indicates the significant progress that has been made in giga-genome assembly.

- Furthermore, the emergence of high-fidelity (HiFi) sequencing technologies along with haplotype-resolved assembly software have greatly facilitated the exploration of polyploid and highly heterozygous plant genomes (Cheng et al., 2021).

- Allele-aware autopolyploid and heterozygous genomes of cultivated alfalfa, potato, sugarcane, and tea have been constructed by integrating HiFi and Hi-C data (Zhang et al., 2018; Chen et al., 2020; Zhang et al., 2021b; Sun et al., 2022).

MODERN PROBLEMS OF PLANT GENETICS

- The advances in computing power and sequencing and assembly technologies have promoted the construction of almost complete genomes, even gap-free genomes, which has provided a solid foundation for comparative genomics analysis among different plant accessions and helped to minimize the negative effects caused by incomplete genome assembly.
- The dynamics of plant genomes and processes such as the amplification of transposable elements, gene tandem duplication, genome rearrangements, and mutations can lead to a continuum of changes from single-nucleotide polymorphisms (SNPs), gene presence/absence variations (PAVs), to structural variations (SVs) that provide the raw material for natural selection, phenotypic diversity, and adaptation (McClintock, 1956; Gabur et al., 2019; Tao et al., 2019).

- <u>**Plant pan-genomics: recent advances, new challenges, and roads ahead**</u>

- **The availability of high-quality genomes of more and more species has led to the realization that a single genome may not be enough to reflect the landscape of a species because of the large numbers of variations between accessions. Therefore, the “pan-genome” concept was conceived to represent all the genetic information of a species, including core genes that are present in all strains and dispensable genes that are present only in a subset of strains (Tettelin et al., 2005) (Fig. 1A).**
- **The definitions and objectives of pan-genome were then modified and developed since it was proposed (Rasko et al., 2008; Snipen et al., 2009; Alcaraz et al., 2010; Plissonneau et al., 2018), and the pan-genome can be either sequence-based or gene-based (Golicz et al., 2020).**

- **Plant pan-genomics: recent advances, new challenges, and roads ahead**

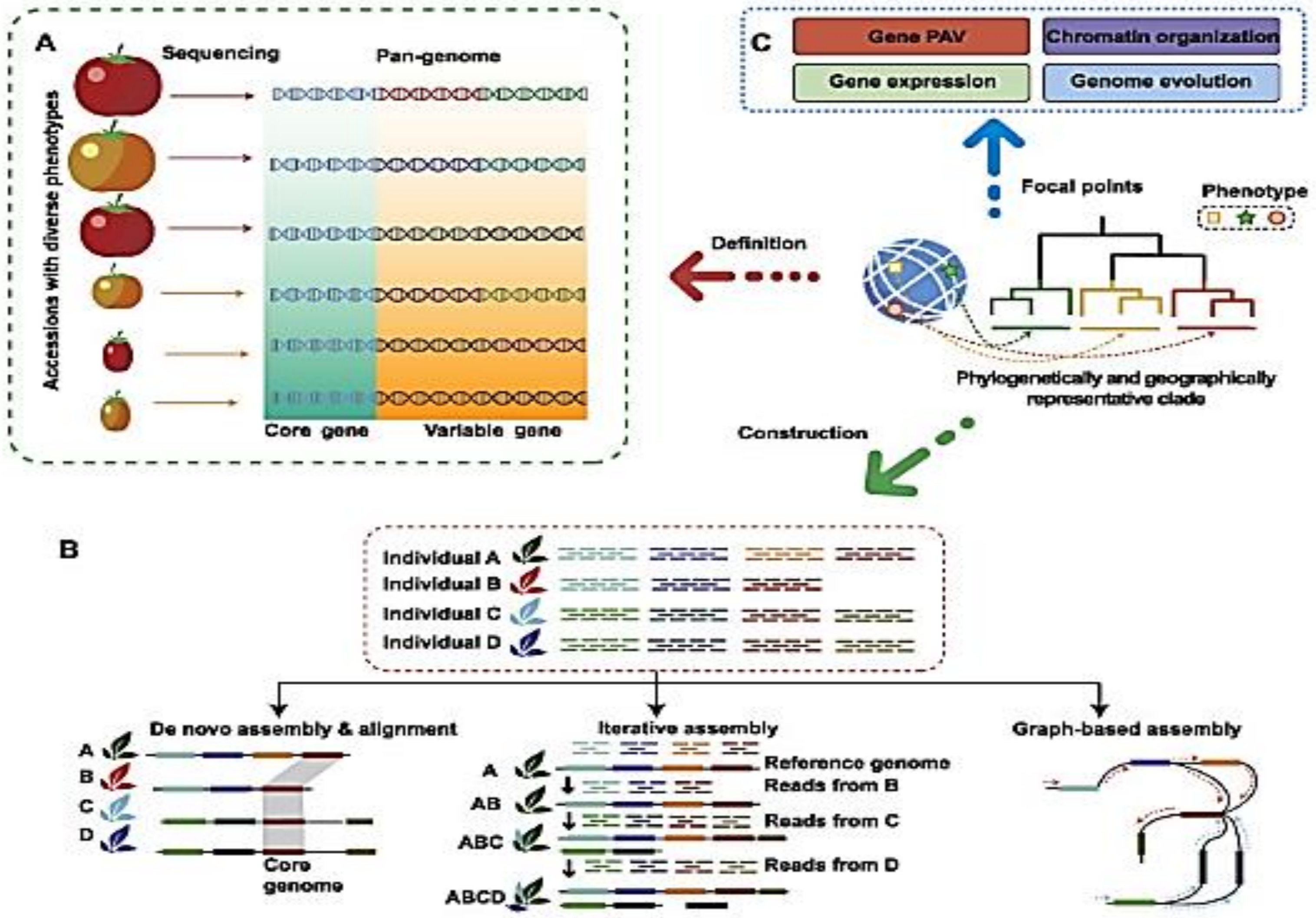


Fig. 1. Schematic diagram showing the concept, construction methods, and research focus of pan-genomics. A: Definition and components of a representative pan-genome. The phylogenetically or geographically representative accessions are selected to construct the pan-genome. The pan-genome can be broken down into a “core gene” that contains genes present in all individuals, and a “variable gene” that contains genes present in a subset of strains. B: Pan-genome assembly approaches, including de novo assembly, iterative assembly, and graph-based assembly. The sequencing reads from different samples are used for the assembly of pan-genome. Reads from shared genomic segments are indicated with the same colors. For de novo assembly, the core regions are highlighted using grey color. For iterative approach, individual A's genome is used as a reference and then the sequencing reads from other individuals are sequentially mapped to the reference genome. The unmapped are then assembled to construct non-redundant pan-genome. A pan-genome is constructed from graph assembly and the relationships between each node can be traced by following the paths of the graph. C: Focus of current pan-genome studies.

- **Plant pan-genomics: recent advances, new challenges, and roads ahead**

- **Compared with the gene-based pan-genome, a sequence-based pan-genome could captures genic as well as nongenic sequences, such as TEs and noncoding RNAs (ncRNAs), which play fundamental roles in the structural organization and function in plant genomes (Tahir Ul Qamar et al., 2020). Until now, pan-genomics studies in plants, including rice, soybean, maize, wheat, cucumber, chickpea, and tomato, have focused mainly on crop breeding, adaptation, and evolution (Hirsch et al., 2014; Li et al., 2014; Schatz et al., 2014; Montenegro et al., 2017; Wang et al., 2018; Zhao et al., 2018; Gao et al., 2019; Alonge et al., 2020; Liu et al., 2020; Walkowiak et al., 2020; Hufford et al., 2021; Li et al., 2022a).**

- **In this review, we briefly summarize the recent major achievements and pipelines for pan-genome research, and discuss potential challenges and perspectives for future pan-genomics studies to provide a basis for applications related to crop improvement.**

- **Plant pan-genomics: recent advances, new challenges, and roads ahead**

- **Approaches for pan-genome construction**
- **Recent advances in sequencing technologies have enabled the assembly of high-quality reference genomes for a large number of plants concurrently.**

- **However, how to integrate multiple genomes from a subset of accessions and make the integrated genetic information easily accessible to biologists remain challenging (Li et al., 2020).**

- **Methods that have been used to construct pan-genomes include de novo assembly, iterative assembly, and graph-based assembly (Li et al., 2014; Schatz et al., 2014; Golicz et al., 2016a; Danilevicz et al., 2020; Liu and Tian, 2020; Qin et al., 2021) (Fig. 1B).**

- **Plant pan-genomics: recent advances, new challenges, and roads ahead**

- **The most straightforward way to construct a pan-genome is by de novo assembly of the genomes of multiple samples, followed by comparative analyses to detect all variant types and characterize the identified genes as core or dispensable (Mahmoud et al., 2019).**

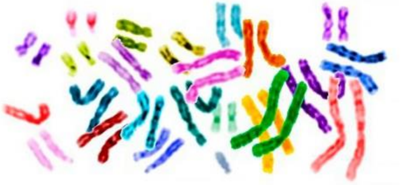
- **The progress in long-read sequencing technologies and complementary approaches such as the construction of Hi-C and BioNano maps have made it feasible to obtain high-quality plant genomes at the chromosome level, including telomere-to-telomere genome assemblies (Miga et al., 2020).**
- **The de novo assembly strategy copes well with repeat regions, but it requires a high depth of sequencing reads to build highly contiguous and accurate genome assemblies, which is costly for large plant genomes and hundreds of reference genomes for one species (Hurgobin and Edwards, 2017).**

- **Characterizing genetic diversity and variation**

- **Uncovering a vast spectrum of variation:** Pan-genome studies have moved beyond single reference genomes to create a more complete picture of a species' genetic makeup, revealing a wide range of structural variations (SVs) such as PAVs, copy-number variations, and large-scale rearrangements.

- **Identifying core and dispensable genomes:** Researchers can now differentiate between genes that are common to all individuals (core genes) and those that vary across the population (dispensable genes).

- **Unveiling novel genes:** Pan-genome analyses have identified numerous genes that were not present in earlier, single-reference genomes, providing a richer pool of genetic resources.



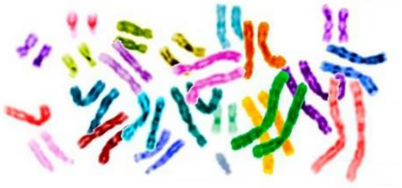
- **Advancing crop improvement and breeding**

- **Enabling 'breeding by design':** By providing a more thorough understanding of a species' genetic makeup, pan-genomics allows breeders to more effectively select and combine desirable traits.
- **Improving stress resilience:** Pan-genomic data is crucial for identifying genetic variations that confer resistance to stresses like drought or disease, which is vital for adapting crops to climate change.
- **Enhancing breeding efficiency:** The ability to explore the genetic diversity of wild relatives and landraces through pan-genomics makes it possible to accelerate and improve crop breeding programs.
- **Providing new insights into domestication:** Studies, like those in rice, have used pan-genomes to investigate the evolutionary history of crops and understand how gene flow shaped modern varieties.

- **Agriculture and Plant Genomes**

- We grow about 200 crops for food, feed, or fiber, and almost all were introduced to the U.S.; they were modified genetically to be adapted to climate and consumer desires.
- Clearly, genomics can help in issues related to food safety, food quality, and food diversity.

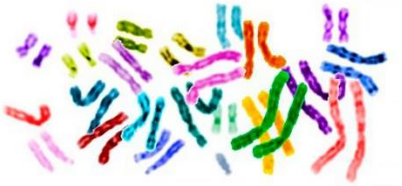
- Genomics provides objectivity in breeding as never before possible; it allows hypothesis testing of quantitative genetics applications in plant improvement.



- **Agriculture and Plant Genomes**

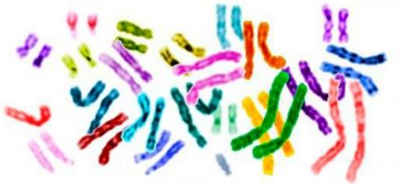
- **Among these are the applications of (i) DNA markers into cultivar identification, seed purity analysis, germplasm resource evaluation, heterosis prediction, genetic mapping, cloning and breeding; and (ii) gene expression data in supporting the description of crop phenology, the analytic comparison of crop growth under stress versus non-stress conditions, or the study of fertilizer effects. Besides, various purposes of using transgenic technologies in agriculture, such as generating cultivars with better product quality, better tolerance to biotic or abiotic stress.**

- CHALLENGES FOR SUSTAINABLE AGRICULTURE PRODUCTION.



- To entail a comprehensive system-centered technology that integrates innovative farming approaches, long-term sustainable agronomic practices, and value-added climate-resilient crops, genomic-based technologies offer, for this task, solid foundational tools and genetic tools insights for shaping the future agriculture.

- **Major achievements in plant pan-genomics.**



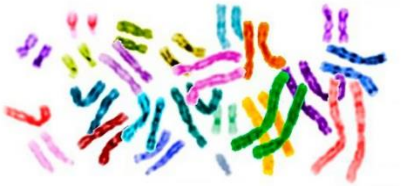
- **A PANGENOME** is information about the complete set of genes of a taxon, among which one can distinguish a set of universal genes common to all representatives of the taxon and variable genes that are partially or completely specific to its representatives.

- Pan-genomes for major crops, such as maize, rice, wheat, and soybean, have been constructed based on high-quality genomes of multiple samples, which has led to great progress in studies into the evolution of plant genomes and the identification of key genes associated with important agronomic traits.

- **MAJOR ACHIEVEMENTS IN PLANT PAN-GENOMICS**
- **These studies have shown that the construction of a pan-genome can eliminate deviations from a single reference genome as much as possible and can present a nearly full view of the diversity within a species.**

- **Grain species**
- **Major grain species include rice, wheat, maize, soybean, millet, barley, oats, and sorghum, which are indispensable sources of energy in the human diet (Bansal et al., 2016).**

- **MAJOR ACHIEVEMENTS IN PLANT PAN-GENOMICS**



- **Because of their importance, a major focus of plant pan-genome research has been to obtain a full view of the genetic variations within each of these grain species. The first plant pan-genome based on high-quality genomes was released in 2014 for wild soybean, which provided a potentially rich resource for improving the genetic diversity of cultivated soybean that was lost during domestication**

- **Questions for monitoring the material being studied:**

- **1. Plant pan-genomics: recent advances, new challenges, and roads ahead.**
- **2. Characterizing genetic diversity and variation.**
- **3. Advancing crop improvement and breeding.**
- **4. Challenges for sustainable agriculture production.**
- **5. Major achievements in plant pan-genomics.**

Recommended list of references:

- 1. Шулембаева К.К., Токубаева А.А. Реконструкция генома мягкой пшеницы на основе хромосомной инженерии и отделенной гибридизации: монография. КазНУ им. аль-Фараби. - Алматы : Қазақ ун-ті, 2019. - 240 с.
- 2. Огурцов А.Н., Близнюк О.Н., Масалитина Н.Ю. Основы генной инженерии и биоинженерии. Учебное пособие. Часть 1.: Молекулярные основы генных технологий. Харьков: НТУ "ХПИ", 2018. - 288 с.
- 3. Нефедова Л.Н., Применение молекулярных методов исследования в генетике: Учебное пособие. - М.: НИЦ Инфра-М, 2012. - 104 с.
- 4. Муминов Т.А., Куандыков Е.У. Основы молекулярной биологии : курс лекций. - Алматы : ССК, 2017. – 222 с.
- 5.Varshney Rajeev K. Plant Genetics and Molecular Biology. - London: Springer, 2018. - 298 p.
- 6. Halford Nigel G. Crop Biotechnology: Genetic Modification And Genome Editing. - London: World Scientific, 2018. - 218 p.
- 7. Glick Bernard R. Molecular biotechnology: principles and applications of recombinant DNA. - 4th ed. - Washington, 2010. - 1200 p.

GOOD LUCK!