

ARTICLE ID: 85

Rice Genome Project: Salient Features, Applications

Sujit S. Jadhav¹ and Vinayak B. Ingale²

- 1. M.Sc. (Ag) Research Scholar, Department of Genetics and Plant Breeding, School of Agriculture, ITM University, Gwalior.*
- 2. M.Sc. (Ag) Research Scholar, Department of Genetics and Plant Breeding, School of Agriculture, ITM University, Gwalior*

Introduction

The Rice Genome Project, also known as the International Rice Genome Sequencing Project (IRGSP), was a collaborative effort involving multiple organizations and countries. It was jointly undertaken by institutions including the National Institute of Aerobiological Sciences (NIAS), the Society for Techno-innovation in Agriculture (STAFF), and the Ministry of Agriculture, Forestry, and Fisheries (MAFF), with the National Institute of Aerobiological Sciences (NIAS) playing a pivotal role.

The sequencing of the rice genome follows a similar methodology as many other extensive genome sequencing projects. Key sequencing templates are generated from large insert genomic libraries, specifically constructed within bacterial artificial chromosomes (BACs) or P1-derived artificial chromosomes (PACs). The predominant approach to sequencing the rice genome involves the utilization of genomic BAC or PAC libraries derived from the Nipponbare variety, chosen as the common template across the International Rice Genome Sequencing Project (IRGSP). It's worth noting that China, while contributing to the sequencing of chromosome 4 as an IRGSP member, employs a different rice variety, indica Guang Lu Ai 4 (Sasaki and Burr 2000), Budiman (1999) In a report available on the CUGI website, there is a comprehensive account of the preparation process for two high-coverage rice BAC libraries (providing 25-fold genome coverage), which were utilized by the IRGSP.

The project took its initial steps in September 1997, during a workshop held in conjunction with the International Symposium on Plant Molecular Biology in Singapore. During this



workshop, scientists from across the globe gathered and made a commitment to collaborate on the ambitious task of sequencing the rice genome. This commitment led to a subsequent meeting held six months later in Tsukuba, Japan, where representatives from several countries, including Japan, Korea, China, the United Kingdom, and the United States, came together to establish the guidelines for the project.

One of the key principles agreed upon by the participants was the sharing of resources and the timely release of physical maps and annotated DNA sequences into public databases. Over time, the IRGSP expanded to include participation from a total of 11 countries. The development of rules and standards for completing the genome sequencing was overseen by the IRGSP Working Group, which comprised a representative from each member country.

In essence, the Rice Genome Project was a remarkable international collaboration in the field of genomics, with the goal of sequencing the entire rice genome to advance agricultural practices and scientific knowledge.

Rice: Overview

Rice, scientifically known as (*Oryza sativa*), stands as a crucial global crop. When considering food production, rice, wheat, and maize collectively contribute to approximately 50% of the world's total output. Remarkably, rice alone serves as the primary staple for half of the world's population. (Sasaki and Burr, 2000). Additionally, it shares a significant degree of co-linearity with these other cereal crops. Rice holds a position of immense importance in global diets, serving as a dietary staple for approximately half of the world's population, particularly in East and Southeast Asia, where nearly 95 percent of the world's rice production is consumed. This versatile grain can be prepared in various ways, such as boiling or grinding into flour. The rice plant thrives in unique agricultural landscapes known as paddies, often referred to as rice paddies. These are small, flat, and deliberately flooded fields predominantly found in southern and eastern Asia. The cultivation of rice dates back to as early as 3500 BC, and by 2000 BC, it had spread across a wide range of agricultural regions, particularly in deltas, floodplains, coastal plains, and even some terraced valley slopes.

Rice: A Model Genome

The decision to conduct the first whole genome sequencing of a cereal crop naturally led to rice being the prime candidate. This is due to the well-established mapping and comprehensive characterization of the rice genome, which happens to be the most compact among major cereal crop genomes, estimated at 400 to 430 Mb. For comparison, the sorghum genome, another significant cereal crop, is substantially larger at 750 to 770 Mb, while the wheat genome dwarfs the rice genome by a staggering ~ 37 times, clocking in at nearly 16,000 Mb as per (Arumuganathan and Earle, 1991). The presence of a significant degree of synteny among grass genomes, encompassing species such as rice, wheat, maize, barley, rye, and sorghum, establishes rice as an outstanding model cereal crop. (Gale and Devos, 1998)

Remarkably, when we compare genetic complexity, rice outpaces even humans. While the human genome is estimated to encompass somewhere between 30,000 and 40,000 genes, indica rice possesses an impressive range of 45,000 to 56,000 genes, and japonica rice may potentially host up to 63,000 genes, as indicated by these two genome blueprints. This genetic richness underscores the significance of rice in genomic research and its potential impact on agriculture and biology.

Objectives of Rice Genome Project

The objectives of the Rice Genome Project can be summarized as follows:

1. Achieve a comprehensive understanding of the function of each gene within the rice genome by the year 2020.
2. Explore the functional diversity of alleles related to agriculturally important genes found within the primary gene pool of rice.
3. Apply the insights gained from functional genomics research to enhance the genetic improvement of rice.

Tools of the Project

Genome sequencing is the process of determining the precise sequence of DNA nucleotides, including the order of adenine (A), cytosine (C), guanine (G), and thymine (T), which constitute an organism's genetic material. The sequencing of chromosomes

within genomes has been assigned to various countries and organizations as part of collaborative efforts.

To facilitate genome sequencing, researchers typically create large insert genomic libraries, where genomic fragments are inserted into vectors like bacterial artificial chromosomes (BACs) or P1-derived artificial chromosomes (PACs). These libraries serve as the primary templates for sequencing.

In the case of the rice genome, genomic BAC or PAC libraries are predominantly used, with the Nipponbare variety being chosen as the common template for the International Rice Genome Sequencing Project (IRGSP). It's important to note that China, a member of the IRGSP, stands out as the sole country employing a different rice variety, indica, for the sequencing of chromosome 4, specifically using Guang Lu Ai 4 as the reference variety for this task. This approach allows for a more comprehensive understanding of the genetic diversity within the rice species and its implications for agriculture and research.

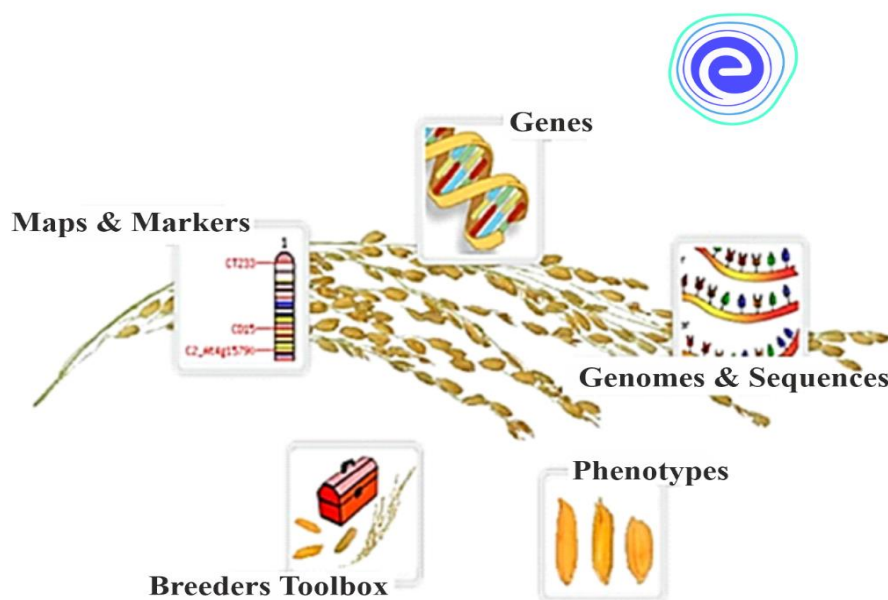


Fig: Rice Genome Project

Salient Features of Rice Genome Project

The salient features of the rice genome project are as follows:

1. The genome size of rice measures 389 megabytes.
2. Researchers have identified a total of 37,544 protein-coding sequences linked to non-transposable elements. Among these, there are approximately 2,859 genes shared between rice and other cereals, with some potentially serving as distinguishing factors between monocot and dicot lineages.
3. Rice carries a transposon load of at least 35% and contains representatives from all known transposon superfamilies.
4. The availability of a comprehensive and high-quality map-based sequencing approach has enabled scientists to explore the structure and evolution of the rice genome. Most notably, they have determined the identity and order of 37,544 rice genes.
5. The sequence contains the essential elements required for functional genomics and molecular breeding programs aimed at deciphering intricate cellular processes and enhancing rice yield.

Applications of Rice Genome Project

1. A substantial variety of intentionally created rice mutants, spanning a wide range of types and quantities, have proven beneficial in advancing molecular product improvement.
2. Rice Genomics stands as an open-access journal exclusively committed to the field of rice genome research.

Progress reports

The meeting kicked off with updates from various members of the IRGSP. Takuji Sasaki gave a detailed account of the advancements made by the RGP (Japan) concerning the



short arm of chromosome 1. As of the present moment, they have successfully completed and shared with the DNA Data Bank of Japan a total of 98 PACs/BACs, encompassing approximately 12 million base pairs of genetic sequence. Significantly, no less than 53 of these PACs/BACs have undergone the annotation process. Furthermore, the RGP is actively engaged in the development of an Expressed Sequence Tag (EST) map. In pursuit of this, they have generated approximately 5000 EST markers, all of which are slated for incorporation into the rice database. (Integrated Rice Genome Explorer, or INE; [Sakata et al., 2000](#)); You can access this data on the RGP website. (see also [Yamamoto and Sasaki, 1997](#)). Robin Buell from TIGR and Rod Wing from CUGI provided updates on the sequencing progress for chromosomes 10 and 3. CUGI has made available 36 million base pairs of rice Sequence Tagged Connectors (STCs), while the CCW has contributed 6 million base pairs of sequence data from chromosomes 3 and 10 to GenBank. Additionally, the CCW has published a recent report focused on data mining of rice sequence information. ([Mao et al., 2000](#)). STCs, or Sequence Tagged Connectors, consist of sequences obtained from random sequencing of BAC (Bacterial Artificial Chromosome) ends. These sequences are then employed to construct minimal tiling paths, which serve as the foundational blueprint for the comprehensive sequencing of a given region or chromosome. ([Venter et al., 1996](#); [Mahairas et al., 1999](#)). The TIGR rice website has undergone significant updates, now featuring nearly 7 million base pairs (out of an estimated 9.6 to 11.5 million base pairs) of ongoing sequence work from the lower arm of chromosome 10. TIGR has recently initiated the sequencing of chromosome 3, with roughly 4.8 million base pairs completed out of an estimated 28 million currently in production. It's important to note that the collective efforts of U.S. groups (CCW, TIGR, and PGIR) have led to the completion of approximately 3.2 million base pairs of rice genome sequencing, a substantial portion of which has been annotated. The remainder is in various stages of Phase I and II unfinished sequence. TIGR has introduced a new database and website display that facilitates gene name searches and provides preliminary annotations for unfinished BACs. The group is actively working on establishing a rice gene index, which will be linked to other plant gene indices on the website. Moreover, there are plans to develop an orthologous gene alignment database for the identification of potential orthologs and paralogs, featuring graphic displays of alignments between different species. One challenge being faced is the lack of functional genomics data. Many of the sequences bear homology to



hypothetical genes in Arabidopsis, with no available information regarding expression or potential functions. For updates on the progress of sequencing by various IRGSP members and links to other IRGSP websites, you can visit the RGP website. (<http://rgp.dna.affrc.go.jp/Seqcollab.html>).

Conclusion

Sequencing the rice genome is an immense undertaking. As of now, approximately 3.5% of the genome has been sequenced (equivalent to 15 out of 430 million base pairs), and an additional 3 to 5% is currently in progress. Despite this, the data already made available have proven to be invaluable for understanding the structure and organization of the genome (see, e.g., Mao et al. in 2000)., Much of this knowledge is transferable to other cereal crops and monocots in general. A significant portion of the nuclear genomes in most plants, and indeed many eukaryotes, is composed of repetitive DNA elements. Repetitive DNA is estimated to make up at least 50% of the rice genome and potentially as much as 70% of the maize genome, as indicated by (Nagano et al. in 1999). The complete sequencing of the rice genome will provide critical insights into the impact of repetitive elements on genome structure and evolution in the plant kingdom. Furthermore, the IRGSP serves as a testbed for sequencing and finishing methods applicable to complex genomes. These advancements will serve as invaluable resources for future projects involving the sequencing of genomes in other eukaryotic organisms.

REFERENCES

- Arumuganathan, K., and Earle, E.D. (1991). Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* 3, 208–218.
- Budiman, M.A. (1999). Construction and characterization of a BAC library of *Oryza sativa* L. ssp. japonica cv. Nipponbare for genomic studies.
- Gale, M.D., and Devos, K.M. (1998). Comparative genetics in the grasses. *Proc. Natl. Acad. Sci. USA* 95, 1971–1974.



Image-<https://d2cyt36b7wnvt9.cloudfront.net/exams/wp-content/uploads/2022/02/18031606/14.png>

Mao, L., Wood, T.C., Yu, Y., Budiman, M.A., Woo, S.S., Sasinowski, M., Goff, S., Dean, R.A., and Wing, R.A. (2000). Rice transposable elements: A survey of 73,000 sequence-tagged-connectors (STCs). *Genome Res.* 10, 982–990.

Sakata, K., Antonio, B.A., Mukai, Y., Nagasaki, H., Sakai, Y., Makino, K., and Sasaki, T. (2000). INE: A rice genome database with an integrated map view. *Nucleic Acids Res.* 28, 97–101.

Sasaki, T., and Burr, B. (2000). International Rice Genome Sequencing Project: The effort to completely sequence the rice genome. *Curr. Opin. Plant Biol.* 3, 138–141.

Venter, J.C., Smith, H.O., and Hood, L. (1996). A new strategy for genome sequencing. *Nature* 381, 364–366.

Yamamoto, K., and Sasaki, T. (1997). Large-scale EST sequencing in rice. *Plant Mol. Biol.* 35, 135–144.

