Gene cluster identification for secondary metabolite production in *Oryza sativa japonica*

Himanshu Singh, Vikas Kaushik

Department of Biotechnology, School of Bioengineering and Biosciences, Lovely Professional University, Phagwara, Punjab.

Abstract:

Plants are able to produce various kinds of secondary metabolites. These secondary metabolites are a rich source of bioactive compounds that can be utilized as medicinal, therapeutic and agrochemical agents. Rapid developments in computational biology pave the way to find out the mechanism for the production of plant metabolites. The recent development in this field provides information that genes encoded some of these metabolites is arranged in an operon like gene cluster. These clusters are governed by the same regulatory element. Advancement in molecular biology, genome mining, and analysis techniques provides us the opportunity to modify these genes for large scale production of these specialized metabolites. Rice which is a major staple food for the human population was utilized in the present study to find out the gene clusters able to synthesize specialized metabolites. In this present study, we have used plantismash tool to find out the gene clusters in Oryza sativa group japonica and indica. Plantismash tool is able to provide information about the annotation and expression analysis of plant gene clusters. We have found out the 39 gene clusters on 12 chromosomes of Oryza sativa group japonica.

Keywords: Plant Metabolites, BLAST, Gene clusters, Plantismash, *Oryza Sativa*, Secondary Metabolites

Introduction

Plant metabolites are specialized chemicals that provide important ecological functions, protection against various diseases and stresses. Plant secondary metabolites are used for various purposes by humans from ancient times. Bioactive compounds of these metabolites are used for various medicinal and therapeutic purposes. Recent advancement in molecular biological techniques and computational biology provides the opportunity to find the new metabolic pathways which are responsible for the production of these specialized chemicals [1]. It has recently found out that genes that are responsible for the formation of these metabolites are arranged in clusters and governed by the same regulatory element. Gene clusters are a group of two or more genes that are found in the DNA of any organism that collectively works in encoding similar proteins or enzymes [2]. Gene clusters work in various ways and they usually have the same promoter region by which all of them get expressed. It is believed till recent, that genes producing these secondary metabolites are scattered in the genome and express exclusive mutual. Secondary metabolic pathways consist of genes for ― signature enzymes that make the scaffold of the secondary metabolite, along with genes for― tailoring enzymes that carry out subsequent modifications to this scaffold [3]. Examples of plant signature enzymes are terpene synthases (for terpenes), chalcone synthases (for flavonoids), and CYP79 family enzymes for cyanogenic glucosides [3,4]. Examples of tailoring enzymes include oxidoreductases, methyltransferases, acyltransferases, and glycosyltransferases. A good candidate gene cluster will contain genes encoding a signature enzyme and tailoring enzymes. Given that these genes contribute to a common pathway, they may be expected to be tightly coexpressed, although this is not always the case (5).

Rice is the staple food of more than 3 billion people worldwide. A significant quantity of recommended niacin and zinc are provided by rice. The digestibility of rice protein is very high (88 per.) and thus is considered as biologically richest protein. After wheat, rice is considered second in the most important crops of the world [6]. Rice (*Oryza sativa*) is of family Poaceae and genus Oryza with over 20 cultivated wild species. *Oryza sativa* and *Oryza glaberrima* being the most cultivated rice species. Oryza sativa is grown worldwide while *Oryza glaberrima* has been cultivated for about the last 3500 years in West Africa. Rice contains a basic chromosome number of n = 12. The species' can be diploid as well as triploid. Both *Oryza sativa* and *Oryza glaberrima L*. being diploid species (2n = 24). Asian cultivated rice is the first fully sequenced crop genome [7].

Oryza sativa japonica rice varieties having a short plant. Grains of this group are short and round. These are having low amylase content. *Oryza sativa indica* group having tall light green leaves. Grains of this group are long and flat. They are having high amylose content.

Experimentation:

Collection of data:

The rice genome consists of 12 chromosomes. The genome of *Oryza sativa japonica* downloaded from Genbank, NCBI [8].

Plantismash:

It allows the fast far-reaching recognizable proof, explanation and examination of optional metabolite biosynthesis quality bunches over the plant kingdom. It is a particular augmentation of the generally utilized anti-SMASH webserver, customized particularly to target plant genomes. It first takes an input sequence and then searches for gene clusters in it taking standard sequences in the inbuilt database as a reference and shows results when and if gene clusters are found in the input sequence. In this software, we first give the email to which we want our results to be sent. In the second input box, we upload our file of genomic or nucleotide sequence in GenBank or EMBL format. The file as a whole is quite big in size. To avoid and cut off time consumption we can use the third input box. In this, we upload the NCBI accession no. of our desired file. Another option would be to segment the whole genome into their chromosomes and upload each individually. This helps in reducing time consumption significantly. Once the results are obtained we can analyze them for finding gene clusters, their size, location, and core domains [9].

BLAST:

Basic Local alignment search tool is an algorithms utilized for checking the sequence similarity in primary sequence of amino acids and nucleotides of genetic material i.e. DNA or RNA. Sequence similarity search was used to annotate the putative gene clusters present in chromosomes of *Oryza sativa*. [10]

Results and Discussion:

Chromosomes of *Oryza sativa japonica* run on plantismash webserver and the following results were obtained as per Table 1 and Table 2:

Table 1: Gene clusters of *Oryza sativa japonica* (chr-1 to chr-12)

Chromosome	Sr. No.	Gene cluster	Size (kb)	Core Domains
Chromosome 1	1.	Saccharide	71.44	2OG-FeII_Oxy, DIOX_N, Peptidase_S10, UDPGT_2
	2.	Lignan-Polyketide	70.90	Chal_sti_synt_C, Chal_sti_synt_N, Dirigent, p450
	3.	Saccharide	82.51	Aminotran_1_2, UDPGT_2
	4.	Saccharide	72.22	UDPGT_2, p450
	5.	Alkaloid	33.28	Bet_v_1, Epimerase, Methyltransf_11
Chromosome 2	1.	Saccharide	139.97	Glycos_transf_1, p450
	2.	Saccharide-Polyketide	211.17	Chal_sti_synt_C, UDPGT_2, p450
	3.	Terpene	369.98	COesterase, Terpene_synth, Terpene_synth_C, p450
Chromosome 3	1.	Lignan-Saccharide	97.55	Cellulose_synt, Dirigent, Methyltransf_11, UDPGT_2
	2.	Saccharide	64.12	Amino_oxidase, UDPGT_2, adh_short

Chromosome 4	1.	Terpene	212.71	Terpene_synth, Terpene_synth_C, adh_short_C2, p450
	2.	Saccharide-Alkaloid	360.51	Cu_amine_oxid, UDPGT_2, adh_short
	3.	Saccharide	169.20	Peptidase_S10, UDPGT_2
	4.	Terpene	334.35	2OG-FeII_Oxy, Terpene_synth, Terpene_synth_C
	5.	Saccharide	42.28	Peptidase_S10, UDPGT_2
	6.	Terpene	61.50	Terpene_synth, Terpene_synth_C, Transferase
	7.	Lignan	82.15	2OG-FeII_Oxy, DIOX_N, Dirigent, Methyltransf_7
Chromosome 5	1.	Saccharide	207.12	2OG-FeII_Oxy, DIOX_N, Transferase, UDPGT_2
Chromosome 6	1.	Putative	71.58	2OG-FeII_Oxy, DIOX_N
	2.	Putative	105.71	Peptidase_S10, Transferase, adh_short_C2
	3.	Saccharide	165.15	Transferase, UDPGT_2
	4.	Polyketide	133.31	Chal_sti_synt_C, p450
Chromosome 7	1.	Lignan	86.46	Aminotran_1_2, Dirigent
	2.	Lignan-Saccharide	88.18	Aminotran_1_2, Dirigent, Glycos_transf_1
	3.	Lignan	86.37	COesterase, Dirigent, p450
Chromosome 8	1.	Saccharide-Terpene	127.02	Methyltransf_2, Terpene_synth, Terpene_synth_C, UDPGT_2
	2.	Lignan-Alkaloid	132.28	Bet_v_1, Dirigent, Epimerase
	3.	Putative	83.82	COesterase, adh_short
Chromosome 9	1.	Saccharide	99.62	AMP-binding, UDPGT_2, p450
	2.	Putative	150.15	COesterase, Peptidase_S10, adh_short
Chromosome 10	1.	Saccharide	141.74	Transferase, UDPGT_2, p450
	2.	Lignan-Saccharide	432.20	Dirigent, UDPGT_2, p450
	3.	Polyketide	141.94	Acetyltransf_1, COesterase, Chal_sti_synt_C, Epimerase
	4.	Polyketide	139.12	Amino_oxidase, Chal_sti_synt_C, GMC_oxred_C, GMC_oxred_N
Chromosome 11	1.	Alkaloid	41.98	HMGL-like, Str_synth, p450
	2.	Lignan	130.12	Dirigent, Peptidase_S10
	3.	Saccharide	468.44	2OG-FeII_Oxy, UDPGT_2, adh_short, adh_short_C2
Chromosome 12	1.	Lignan	323.86	Dirigent, Methyltransf_2, p450
	2.	Saccharide	67.79	Glycos_transf_1, p450

- Results obtained from *Oryza sativa japonica* have the most occurrence of saccharide and putative gene clusters in chromosomes (1,2,3,4,5,6,8,9,10,11,12) and (8,9,6) respectively.

Analysis: We analyzed results obtained for putative clusters to find similarities in order to identify and predict their functions via Blast.

Japonica group Putatives:

- Chromosome 6, Cluster 1: Most similar sequence that was obtained was Proteosome subunit alpha-type. (Query coverage: 9%. Similarity: 100%.)
- Chromosome 6, Cluster 2: Most similar sequence that was obtained was Houba Copia like Retrotransposon. (Query coverage: 21%. Similarity: 100%.)
- Chromosome 8, Cluster 3: Most similar sequence that was obtained was Carboxylesterase. (Query coverage: 11%. Similarity: 97%.)
- Chromosome 9, Cluster 2: Most similar sequence that was obtained was Glycosyltransferase. (Query coverage: 17%. Similarity: 99%.)

Conclusion:

The gene clusters obtained via plantismash of rice chromosomes 1 to 12 of japonica group. The predominant cluster that was obtained was saccharide while Putative has second-most occurrence. Core domains that have the most occurrences are 2OG-FeII_Oxy, transferases, and peptidases.

References:

1.) Nützmann HW, Huang A, Osbourn A. , *New Phytol*, 2016, 211, 771-89.
2.) Yi, G., Sze, S.-H., & Thon, M. R., *Bioinformatics (Oxford, England)*, 2007, 23, 1053–1060.
3.) Osbourn, A., *Trends in Genetics*, 2010, 26, 449–457.
4.) D'Auria, J. C., & Gershenzon, J., *Current Opinion in Plant Biology*, 2005, 8, 308–316.
5.) Takos AM, Knudsen C, Lai D, Kannangara R, Mikkelsen L, Motawia MS, Olsen CE, Sato S, Tabata S, Jorgensen K, et al., *Plant Journal*, 2011, 68, 273–286.
6.) Ghosh, B., Ali, M.N. & Gantait, S. , *J. Res. Rice*, 2016, 4, 2–9.
7.) C-H Wang, X-M Zheng, Q Xu, X-P Yuan, L Huang, H-F Zhou, X-H Wei, S Ge, *Heredity*, 2014, 112, 489–496.
8.) Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW, Nucleic Acids Res., 2013, 41, 36-42.
9.) Kautsar SA, Suarez Duran HG, Blin K, Osbourn A, Medema MH, Nucleic Acids Res., 2017, 45, 55-63.
10.) Stephen Altschul, Warren Gish, Webb Miller, Eugene Myers, David J. Lipman, Journal of Molecular Biology, 1990, 215, 403–410.