



Full Length Article

Homology Modeling, Functional Annotation and Comparative Genome Analysis of *GBSS* Enzyme in Rice and Maize Genomes

Javed Iqbal Wattoo^{1,2*}, Muhammad Shahzad Iqbal³, Muhammad Arif⁴, Zafar Saleem¹, Muhammad Naveed Shahid⁵, and Muhammad Iqbal¹

¹Centre for Applied Molecular Biology (CAMB), MoST, 87-West Canal Bank Road Lahore, Pakistan

²Current address: Institute of Molecular Biology and Biotechnology (IMBB), The University of Lahore, Pakistan

³Center of Excellence in Molecular Biology (CEMB), University of Punjab, Lahore, Pakistan

⁴National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, Pakistan

⁵Institute of Molecular Biology and Biotechnology (IMBB), The University of Lahore, Pakistan

*For correspondence: javed.nibge@gmail.com

Abstract

The structure of a gene or protein is evolutionary more conserved than a DNA sequence; therefore, study of genetic inheritance at molecular level revealed by latest bioinformatics approaches has great potential to disclose the genetic basis associated with diverse phenotypes. Granule bound starch synthase (*GBSS*), a waxy gene encoded protein dictates cooking and eating quality attributes in different cereal species. In the current study, we have evaluated the sequence information of waxy gene retrieved from various data sources in two cereal genomes. We have interpreted homology modeling, functional annotation and comparative genome analysis of waxy gene among rice and maize genomes. Based on homology modeling, three dimensional (3D) structure of the gene was constructed and interpreted in both species. Several validation tests were computed to check the reliability of 3D structure. In comparative genome analysis, we found conserved domains only in rice waxy protein. These conserved domains have significant role for starch biosynthesis and inheritance of waxy gene in different rice species with variable starch contents. The study has clear implications to annotate the role of *GBSS* enzyme and linked proteins associated with diverse starch phenotypes. More insights into the structure of waxy gene will lead to annotate the role of this gene in different biological pathways in different cereal species. © 2015 Friends Science Publishers

Keywords: Bioinformatics; Computational biology; *GBSS*; Waxy gene; Rice; Maize; Comparative genomics

Introduction

Rice (*Oryza sativa* L.) is an important staple food crop and its genome is used as reference cereal to study the genetics of related grass species. The increased information of newly sequenced crop genomes has led to design the latest bioinformatics tools for detailed genetic analysis of gene families (Martinez, 2011). Quality of rice grain is mainly affected by three physicochemical traits: amylose content (AC), gel consistency (GC) and alkali spreading value or GT (ASV/GT) (Cuevas *et al.*, 2010). Traits of grain quality are mostly governed by waxy gene (*Wx*) on chromosome 6 which encodes granule bound starch synthase (*GBSS*), an enzyme which controls the inheritance of amylose content in rice (Tian *et al.*, 2005). Genome-based mapping techniques have provided useful information to dissect different genomic regions associated with grain quality traits (Stich and Buckler, 2005). But it is difficult to disclose the effect of minor genes and haplotypes due to narrow genetic base of the experimental material (Yu *et al.*, 2006). The structure of a protein or gene is evolutionary more

conserved than a DNA sequence; therefore, annotation of three dimensional structure has great potential to study the inheritance of gene at molecular level. A little or no work has been done regarding sequence analysis of *GBSS* enzyme and its allied proteins. A combination of genetic mapping techniques with bioinformatics tools have facilitated the detailed molecular dissection of rice genes and the results would also be valuable for comparative genome analysis and evolutionary studies of different cereal species. The objectives of the current study were three fold, (a) to compare sequence data of waxy gene among rice and maize genomes; (b) to study the homology modeling of waxy gene in rice and maize; (c) functional annotation and comparative genomic analysis of waxy gene. The results of current study would be supportive to analyze the *GBSS* enzyme and its protein domains in different cereal species to find the function of *GBSS* for starch contents and complex plant phenotypes. The findings would further be practical to address the myths of conserved regions in different genes and their role in different biosynthesis pathways.

Materials and Methods

Sequence Analysis

The amino acid sequences of waxy gene of rice and maize were retrieved from NCBI data base using primary accession name for each variety (<http://www.ncbi.nlm.nih.gov/>). About 15 amino acid sequences for each species were used for sequence analysis.

Homology Modeling of Waxy (Wx) Gene

A similarity-based sequence search was conducted using NCBI with non-redundant database using BLAST tool (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). The raw sequences of both species were translated to amino acid sequences using ExPASy translate tool (<http://web.expasy.org/translate>). The translated sequences were then processed with Swiss-Model tool (<http://swissmodel.expasy.org/>) to build 3D structure. All the translated sequences of both species were used to get 3D structure for further functional annotation and computational genomic analysis.

Functional Annotation

Based on sequence similarity, the waxy gene was analyzed for the presence or absence of conserved domains and orthologous and paralogous gene families. This study was conducted using three different types of bioinformatics tools; NCBI Conserved Domains Database (NCBI-CDD) (Marchler *et al.*, 2011), Protein families database (Pfam) (Aranda *et al.*, 2010) and InterProScan (Zdobnov and Apweiler, 2001). NCBI-DD is a tool for protein annotation having a huge data of multiple sequence alignment to study full length proteins. Pfam database uses Markov models to annotate multiple sequence alignment while InterProScan combines signatures of different proteins native to InterPro member along with its corresponding InterPro.

Comparative Genomics

The reliability of predicted 3D structures of all the proteins of waxy genes were checked using Ramachandran plot. Ramachandran plots were finally obtained as homology model and template for quality assessment of 3D model. 3D models of all the proteins were compared to differentiate the patterns of distribution of Alpha helices, Beta plated sheets and interconnecting loops.

Results

The present study focused to describe homology modeling, functional annotation and comparative genomic analysis of waxy gene in two cereal genomes using different bioinformatics approaches.

Homology Modeling

Protein 3D structure was constructed using Swiss model (<http://swissmodel.expasy.org/>) to study the localization and interaction of proteins in stable confirmation. In structural genomics and proteomics, homology modeling is one of the reliable method for comparative analysis. A defined sequence alignment is crucial to build a reliable 3D structure. Sequence alignment was performed using BLASTP software. Both template and query sequence were aligned using algorithms of multiple sequence alignment. The sequence homology of waxy gene of rice (Table1) ranged from 98–99% while the waxy gene from maize (Table 2) was 82–88% (Fig. 1a and Fig 2a).

Functional Annotation

For rice accessions, 3D model was constructed using Swiss model and reliability of 3D model was confirmed using Ramachandran plot (Fig. 1). The point at which charge on protein becomes zero is called an isoelectric point (PI). The ratio of PI/Mw was estimated by using PROCHECK (<http://www.ebi.ac.uk/thornton-srv/software/PROCHECK>). The same tool (PROCHECK) was used to constructed Ramachandran plot. All rice accessions (IR-64, Azucena, IRBB-57, Super-Basmati and Basmat-370) showed 99% Ramachandran reliability with PI/Mw 8.47/6609, 7.99/7764, 8.32/6509, 6.84/7609 and 7.95/8906 respectively (Table 1). Similarly, the maize accessions (Jinnuo-2, Da-M40, 20V-18, Bainianzongh and Xiaobaiyumi) revealed 100% (except 20V-18 with 99%) Ramachandran reliability with PI/Mw 9.37/6509, 8.69/8664, 9.29/5709, 7.94/8509 and 8.55/8607 respectively (Table 2).

Comparative Genome Analysis

Once the functional study of waxy gene and its proteins was completed in rice and maize, a comparative genome analysis was performed for further characterization. Both sequences of waxy gene in rice and maize (IR-64 and Jinnuo-2) were translated to proteins using Swiss model. The web-based tool Inter-pro (<http://www.ebi.ac.uk/interpro>) was used to compare the protein sequence of waxy gene in two species. The results revealed two domains (Starch synthase, catalytic domain and Glycosyl transferase, family 1) in waxy protein of rice while only one domain (Starch synthase, catalytic domain) was found in maize (Fig. 3). Starch synthase catalytic domain was found similar between rice and maize. The amino acid length in rice and maize proteins was estimated 605 and 263 respectively (Fig. 3). The results of pairwise sequence alignment using LALIGN (<http://www.ebi.ac.uk/Tools/psa/>) showed that a total of 265 amino acids overlapped between two amino acid sequences. Alignment analysis further showed 70% identity and 87% similarity between waxy proteins of maize and rice.

Table 1: Statistical simulations based on homology modeling, functional annotation and comparative sequence analysis of waxy gene in five rice accessions

Rice Accession	Type	Gene	Seq identity %	Ramachandaran %	PI/Mw	Errat %	Verify3D %
IR-64	<i>Indica</i>	<i>GBSS</i>	99.04	99.1	8.47/6609	87.29	90.22
Azucena	<i>Indica</i>	<i>GBSS</i>	99.04	99.7	7.99/7764	93.125	92.61
IRBB-57	<i>Indica</i>	<i>GBSS</i>	99.04	99.7	8.32/6509	93.125	92.61
Super Basmati	<i>Temp. Japonica</i>	<i>GBSS</i>	98.85	99.7	6.84/7609	92.917	92.42
Basmati-370	<i>Temp. Japonica</i>	<i>GBSS</i>	98.85	99.7	7.95/8906	92.917	92.42

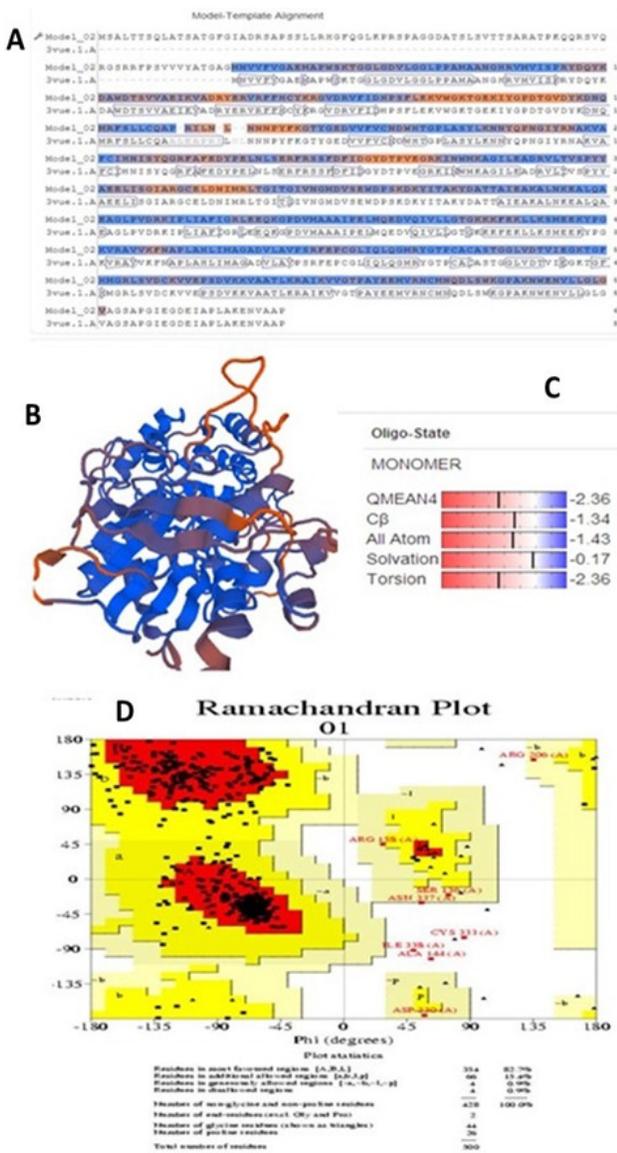


Fig. 1: Sequence and Structure analysis of Waxy gene (*Wx*) in Rice. (A) Sequence and template alignment of waxy gene prior to translation and 3D structure (B) Homology model of rice waxy gene, alpha helices are shown in blue color, beta plated sheets in orange blue and coils in brown color (C) Statistical angles between different domains (D) Ramachandran plot of waxy gene for reliability of 3D structure

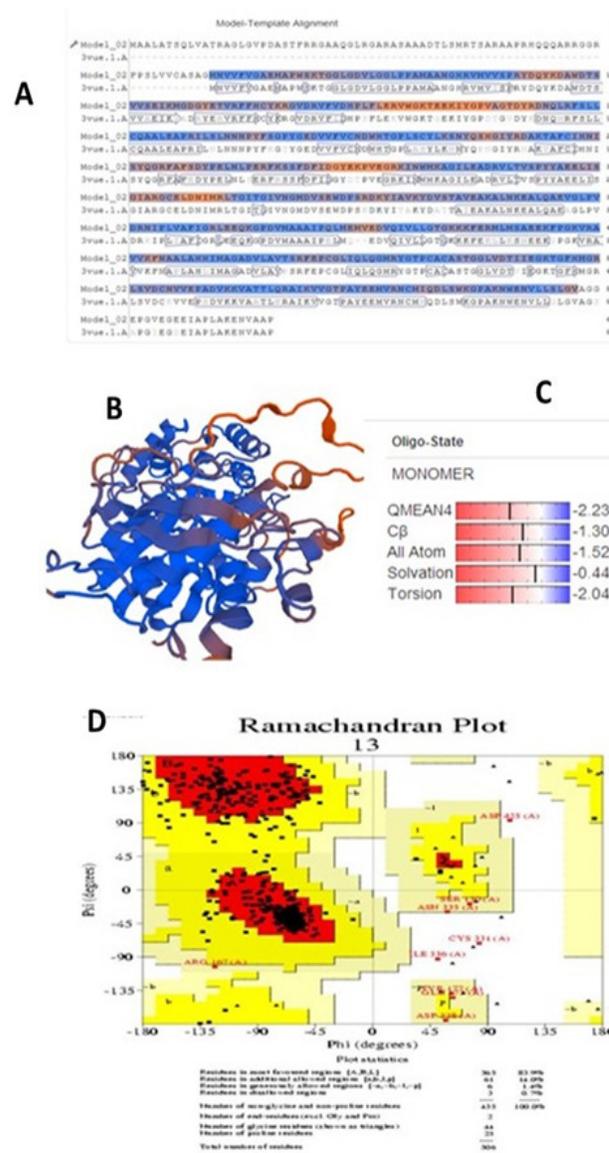
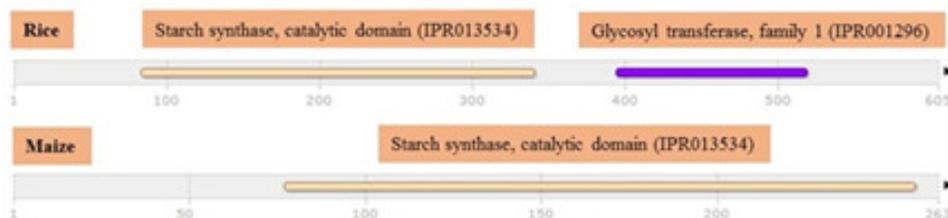


Fig. 2: Sequence and Structure analysis of Waxy gene (*Wx*) in Maize. (A) Sequence and template alignment of waxy gene prior to translation and 3D structure (B) Homology model of rice waxy gene, alpha helices are shown in blue color, beta plated sheets in orange blue and coils in brown color (C) Statistical angles between different domains (D) Ramachandran plot of waxy gene for reliability of 3D structure

Table 2: Statistical simulations based on homology modeling, functional annotation and comparative sequence analysis of waxy gene in five maize accessions

Maize Accession	Type	Gene	Seq identity %	Ramachandaran %	PI/Mw	Errat %	Verify3D %
Jinnuo-2	<i>T. aestivum</i>	<i>GBSS</i>	85	100	9.37/6509	79	83.98
Da-M40	<i>T. aestivum</i>	<i>GBSS</i>	82	100	8.69/8664	77	83.78
20V-18	<i>T. aestivum</i>	<i>GBSS</i>	88	99	9.29/5709	92	93.69
Bainianzongh	<i>T. aestivum</i>	<i>GBSS</i>	85	100	7.94/8509	79	83.98
Xiaobaiyumi	<i>T. aestivum</i>	<i>GBSS</i>	82	100	8.55/8607	76	83.7

**Fig. 3:** Comparative genomic analysis showing the domain difference in the waxy protein sequence of maize and rice

Rice grain consisted of about 90% starch, which is again divided into amylose and amylopectin with 2–35% and 65% respectively. The starch ratio in maize varies from 40–60% depending on variety type. The additional domain in rice waxy protein may be associated with increased starch content in rice grain, which ranges from 0–35%.

Discussion

The waxy protein *GBSS* primarily dictates the cooking and eating quality traits in rice and knowing the detailed structure of protein is essential to study the inheritance (origin and evolution) of waxy gene in different cereal species. Many scientists studied the genetic basis of waxy gene using different genetic mapping approaches (Lanceras *et al.*, 2000; Fan *et al.*, 2005; Tian *et al.*, 2005). But little or no work was done to study its 3D structure. The previously reported function of waxy gene and its protein (*GBSS*) to control the cooking and eating quality traits in cereals is well documented in rice (Septiningsih *et al.*, 2003; Li *et al.*, 2004). The present study describes an organized workflow using latest bioinformatics tools to describe functional annotation of waxy gene based on homology modeling and comparative genome analysis. The functional annotation was revealed using three web-based tools; NCBI-CDD (<http://www.ncbi.nlm.nih.gov/cdd/>), Swiss-model (<http://swissmodel.expasy.org/>) and Euk-mPloc 2.0 (<http://www.csbio.sjtu.edu.cn/bioinf/euk-multi-2/>). The 3D model revealed three different types of structures (Alpha helices, Beta plated sheets and Interlopes) (Fig. 1b). The protein domains present in alpha helices are mostly found in cell membranes and play significant role in transport while the domains of beta plated sheets have intracellular localization. The function of interlopes is to connect different domains and structures. The additional

domain in rice waxy protein may be associated with increased starch content in rice grain. Similarly, availability of reference genome sequence of rice can be used as prototype to identify different genes and their associations with complex phenotypic traits. The sequence information from other cereal species like wheat and maize would further be helpful for comparative genomic study and phylogenetic analysis.

Conclusion

In the current study, we have interpreted the results of homology modeling, functional annotation and comparative genome analysis of waxy gene enzyme, *GBSS* and its associated proteins in two cereal genomes using latest bioinformatics approaches. Conserved domains were found only in rice waxy protein in comparative genome analysis. These conserved domains have significant role for starch biosynthesis and inheritance of waxy gene in different rice species with variable starch contents. The study has clear implications to annotate the role of *GBSS* enzyme associated with diverse plant phenotypes. The waxy gene has the same function in maize and rice but the presence of extra domain in rice protein may be associated with different levels of starch phenotypes. However, further bioinformatics analysis focusing on many other cereal species is obligatory to explain the structural image and functions of these genes and linked proteins in detail.

Acknowledgements

The work was part of project, funded by Ministry of Science and Technology (MoST), Govt. of Pakistan. The authors wish to thank the anonymous reviewers for their valuable suggestions and thoughtful comments.

References

- Aranda, B., P. Achuthan, Y.A. Faruque, I. Armean, A. Bridge, C. Derow and H. Hermjakob, 2010. The IntAct molecular interaction database. *Nucleic Acids Res.*, 38: 525–531
- Cuevas, R.P., V.D. Daygon, H.M. Corpuz, L. Nora, R.F. Reinke, D.L. Waters and M.A. Fitzgerald, 2010. Melting the secrets of gelatinization temperature in rice. *Funct. Plant Biol.*, 37: 439–447
- Fan, C.C., X.Q. Yu, Y.Z. Xing, C.G. Xu, L.J. Luo and Q. Zhang, 2005. The main effects, epistatic effects and environmental interactions of QTLs on the cooking and eating quality of rice in a doubled haploid line population. *Theor. Appl. Genet.*, 110: 1445–1452
- Lanceras, J.C., Z.L. Huang, O. Naivikul, A. Vanavichit, V. Ruanjaichon and S. Tragoonrung, 2000. Mapping of genes for cooking and eating qualities in Thai jasmine rice (KDML105). *DNA Res.*, 7: 93–101
- Li, J., J. Xiao, S. Grandillo, L. Jiang, Y. Wan, Q. Deng, L. Yuan and S.R. McCouch, 2004. QTL detection for rice grain quality traits using an interspecific backcross population derived from cultivated Asian (*O. sativa* L.) and African (*O. glaberrima* S.) rice. *Genome*, 47: 697–704
- Marchler, B., A. Lu, S. Anderson, J.B. Chitsaz, F. Derbyshire, M.K. DeWeese and S.H. Bryant, 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.*, 39: 225–229
- Martinez, M., 2011. Plant protein-coding gene families: emerging bioinformatics approaches. *Trends Plant Sci.*, 16: 558–567
- Stich, B., A.E. Melchinger, M. Frisch, H.P. Maurer, M. Heckenberger and J.C. Reif, 2005. Linkage disequilibrium in European elite maize germplasm investigated with SSRs. *Theor. Appl. Genet.*, 111: 723–730
- Septiningsih, E.M., K.R. Trijatmiko, S. Moeljopawiro and S.R. McCouch, 2003. Identification of quantitative trait loci for grain quality in an advanced backcross population derived from the *Oryza sativa* variety IR-64 and the wild relative *O. rufipogon*. *Theor. Appl. Genet.*, 107: 1433–1441
- Tian, R., G.H. Jiang, L.H. Shen, L.Q. Wang and Y.Q. He, 2005. Mapping quantitative trait loci underlying the cooking and eating quality of rice using a DH population. *Mol. Breed.*, 15: 117–124
- Yu, J. and E.S. Buckler, 2006. Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.*, 17: 155–160
- Zdobnov, E.M. and R. Apweiler, 2001. Inter-Pro-Scan an integration platform for the signature recognition methods in Inter-Pro. *Bioinformatics*, 17: 847–848

(Received 03 October 2014; Accepted 20 January 2015)