In silico sequence analysis of Gibberellin-20 Oxidase-2, the semidwarfing protein of rice through the application of BLAST, FASTA and WU-Blast 2 programs

ASTHA DUHAN, MANOJ KUMAR SHARMA AND TAVINDERJEET KAUR

SUMMARY

Computational prediction of evolutionary relationship for genes and proteins remains a fundamental problem in biology. Multiple sequence alignment plays a very important role in the evolutionary relationship prediction. The quality measurement is the most crucial issue in the multiple sequence alignment. An appropriate scoring function, high residue identity and degree of divergence lead to more accurate and reliable multiple sequence alignments. Using two scoring matrices, Blosum and PAM series of substitution and similarity matrices, sequence analysis tools for identifying evolutionary relationships to GA20ox2, the semi-dwarf protein of rice have been selected. The quality of alignment produced by various scoring matrices features the parameters that are important for the selection of right matrix for any alignment tool. We also show that the quality of any alignment and the homologous sequences produced by BLAST has more predictive power in identifying evolutionary conserved residues for the semi-dwarf protein. This was made evident by the conserved residue pattern observed for homologous sequences scanned by BLAST (Blosum80) resulted in 18 conserved residues by multiple sequence analysis in comparison with FASTA and WU-Blast2 that showed 11 and 12 conserved residues, respectively using the scoring matrix PAM 120.

The semidwarfing gene in rice (*sd-1*) is one of the most important genes deployed in modern rice breeding. Its recessive character results in a shortened culm with improved lodging resistance and a greater harvest index, allowing for the increased use of nitrogen fertilizers. The *sd-1* gene was first identified in the Chinese variety Dee-geo-woo-gen (DGWG), and was crossed in the early 1960s with Peta (tall) to develop the semidwarf cultivar IR8, which produced record yields throughout Asia and formed the basis for the development of new high-yielding, semidwarf plant types (IRRI, 1967). Plant architecture refers to the collection of all important agronomic characters which determines the grain production. In rice it is mainly affected by plant height, tillering and panicle morphology (Wang *et al.*, 2005). In rice, short stature plants were developed by altering the plant architecture. Short stature in rice was developed by a recessive semidwarfgene sd-1, responsible for high yielding varieties (Nagano *et al.,* 2005). The sd-1 gene codes for the oxidase enzyme involved in the biosynthesis of gibberellins, Gibberellin-20 oxidase-2(GA20ox2).

GA20ox2 is tightly linked to the sd-1 locus of rice chromosome 1. Semidwarfing gene inhibits the elongation of lower internodes which makes rice resistant oto lodging (Ogi *et al.,*1993). The short stature of IR8 is due to a mutation in the plant's sd1 gene, encoding an oxidase enzyme involved in the biosynthesis of gibberellin, a plant growth hormone (Sasaki *et al.,* 2002).Ga20ox2 catalyses the conversion of GA_{53} to GA_{20}. DGWG sd-1 mutants have 383 bp deletion in genomic sequence which encode a non-functional protein in an *indica* semidwarf IR8 (Monna *et al.,* 2002). However, a substitution of Leu-266 that is highly conserved residue resulted in the loss of function in *japonica* semidwarf. Rice represents enormous gene pool for improvement of rice cultivars as well as other crops which show significant similarity to rice genome. By making the use of sequence similarity/ homology genetic relationships among cereal crops can be established (Ishii *et al.*, 1996). Comparative genomic analysis revealed that the sdw1/denso gene that controls plant height, yield and quality has located in the syntenic region of the rice semidwarf gene sd1 on chromosome 1 (Jia *et al.*, 2009).

Protein sequences contain valuable information to predict structure and function of gene product. The comparison of two sets is a fundamental task that infers about how two sequences may be related functionally and genetically. The simultaneous alignment of amino acid sequences is now a major tool in molecular biology. The advent of large genome projects led to an explosion of

Correspondence to:
ASTHA DUHAN, Department of Bioinformatics, Mata Gujri College, FATEHGARH SAHIB (PUNJAB) INDIA
Authors' affiliations:
MANOJ KUMAR SHARMA, Department of Bioinformatics, Janta Vedi College, Baraut, BAGHPAT (U.P.) INDIA
TAVINDERJEET KAUR, Department of Bioinformatics, Mata Gujri College, FATEHGARH SAHIB (PUNJAB) INDIA