International Journal of Pharmacy and Biological Sciences
ISSN: 2321-3272 (Print), ISSN: 2230-7605 (Online)
IJPBS™ | Volume 8 | Issue 3 | JUL-SEPT | 2018 | 959-965
Research Article | Biological Sciences | Open Access | MCI Approved| ज्ञान-विज्ञान विमुक्तये
|UGC Approved Journal |

QUALITY PROTEIN MAIZE: A STUDY ON MICRO EVOLUTION AND MOLECULAR BASIS OF EVOLUTIONARY SUB FUNCTIONALIZATION

Vyatikram Ghatak[1], Soumen Saha[1,2], Sinchan Adhikari[1] & Parthadeb Ghosh[1]*

[1]Cytogenetics & Plant Biotechnology Research Unit, Department of Botany, University of Kalyani, Kalyani, Nadia, West Bengal, India
[2]Cytogenetics & Plant Biotechnology Research Unit, Department of Sericulture, Raiganj University, Raiganj-733134, Uttar Dinajpur, West Bengal, India

*Corresponding Author Email: pdggene@rediffmail.com

ABSTRACT

Genes Opaque-2 and ohp-1, 2 determine the amount of lysine and tryptophan in maize kernels. Sequence analysis at these loci reveals the evolutionary pattern and chronology of different phylogenetic events in the evolutionary history of Z. mays. ssp. mays. Various techniques of bioinformatics has been used which suggests that maize ohp-1 and ohp-2 (Opaque-2 Heterodimerization Protein 1 and 2) sequences are orthologues and both of them are paralogous to Opaque-2. The molecular basis of evolutionary sub functionalization in ohp-1, 2 is marked to be a leucine to valine substitution in residue 290 due to which zipper formation is hindered. However, all the residues in the bZIP region are 100% identical in these two proteins suggesting that evolutionary selection has constrained sequence divergence in the bZIP domain. Tajima's D value suggests non-neutrality and operation of selection in ohp1, 2 but no selection in Opaque-2 locus. UPGMA analysis has been used to calculate the evolutionary rates at these loci and to mark the time of separation of Sorghum, maize and teosintes.

KEY WORDS

MSA, PSI-BLAST, Maximum Likelihood, Neighbor Joining, UPGMA, Tajima's D value, MEME.

INTRODUCTION

Opaque-2 gene encodes a transcription factor involved in the regulation of zein storage proteins. In conventional maize with wild type *Opaque*-2 gene nitrogen is diverted for zein synthesis and hence endosperm lack essential amino acids lysine and tryptophan [1, 2]. *Opaque-2* heterodimerizing protein 1 (*ohp-1*) is a bZIP class of transcriptional regulatory protein with a duplicate gene, *ohp-2*. Both of them can bind to the promoter only when they are complexed with the *Opaque-2* gene product. The present study enquires the molecular basis of this function loss in *ohp-1* and *ohp-2* [3, 4]. Sequence polymorphism of proteins encoded by these genes has been subjected to the techniques of microevolutionary analysis [5, 6] to reveal the pattern and chronology of phylogenetic events [7, 8, 9] in these loci which controls the protein quality in quality protein maize.

MATERIALS AND METHODS

The protein sequences of interest are chosen from NCBI database at www.ncbi.nlm.nih. Scoring matrices used in the present study are mainly PAM250 and BLOSUM62. Clustal X 2.0 and BioEdit [10] software's are used for multiple sequence alignment. Homology has been searched primarily by using BLAST-the online tool available at NCBI. Using the PSI-BLAST [11] programme, patterns of similarity in same position of the query is searched against protein sequence database and then the matches of that initial pattern are extended by iterated search and gradual refinement at each cycle. Bit score (S′) is derived from the raw alignment score S in which the statistical properties of the scoring system used are considered [12]. E value (number of hits one can “expect” to see by chance when searching a database of a particular size) below 0.05 is considered to be significant. The E values are interpreted in the present study according to the following guideline: E<0.02 implies homology of the sequence with the query; E=0.02–0.05 indicates possible homology and E>0.05 suggests chance similarity [13, 14]. Two different types of distance analysis, namely, Neighbour Joining [15] and UPGMA [16] (Unweighted Pair Group Method with Arithmatic Mean) has been done using MEGA 4 [17]. Tajima’s neutrality test and Tajima’s relative rate test is done to know the nature of evolution. Using Tajima’s test of Neutrality, the number of segregating sites (S) and the average pair wise differences (π) are compared to detect the nature of selection in sequences under study [18]. The difference between these two estimates has been depicted as Tajima’s D value [19]. Under neutral expectations, both π and S gives similar estimates. So, when Tajima’s D is approximately zero, neutrality of the locus can be predicted. A positive D value suggests balancing selection and a negative D value depicts the operation of purifying selection in the loci under study [20]. Neighbor-joining trees are drawn using distances between pairs of taxa. This method allows different rate of evolution along the different branches of the tree, it has limitations when a molecular clock is operative in the sequences under study (as known from Tajimas relative rate test). The evolutionary rates and divergence times are calculated using UPGMA method which does not consider differential rates of evolution along the different branches of the evolutionary tree. Distance from nodes is calculated by arithmetic mean enabling the study of evolutionary rates and divergence times. For the calibration of UPGMA trees in this work, a divergence time of 50 million years between *Oryza* and the *Sorghum*-maize lineage has been used as calculated by Wolfe et al., [21]. Common motifs has been identified by using the MEME algorithm at www.meme.nbcr.net.

RESULTS AND DISCUSSION

NCBI sequence database has been used to retrieve the *Opaque*-2 sequence of *Zea mays* ssp. *mays* (gi|162462284| ref|NM_001111951.1|). The protein sequence, consisting of 458 residues, is subjected to Tajima's Neutrality Test along with *Opaque*-2 protein sequences of teosintes. D value of 0.26 with a standard error of 0.931219 suggests lack of operation of selection in the maize *Opaque*-2 loci. Results of pairwise distance estimate using 1000 bootstrap replicates for standard error calculation show close significant affinities between *Opaque*-2 proteins of maize and Mexican teosinte (Z. *mays* ssp. *mexicana* and Z. *mays* ssp. *parviglumis*). Tajima’s relative rate test between the *Opaque*-2 proteins of Z. *mays* ssp. *mays* and Z. *mays* ssp. *parviglumis* using diploid perennial Z. *diploperennis* as out group shows chi-square = 3.00 at 1df., in a probability level of 0.05 suggesting the null hypothesis of this test to be true. That is, evolution at this locus follows a molecular clock. Maximum likelihood analysis shows that maize (*Zea mays* ssp. *mays*) *Opaque*-2 protein has close relationship with the *Opaque*-2 protein of *Z. mays* ssp. *parviglumis* and these two genes are orthologous. As suggested by the relative rate test using the *Opaque*-2 proteins of maize and teosinte, the rate of evolution of O*paque*-2 gene in the lineage under study follows a molecular clock. Owing to this constant rate of evolution among the different branches of the tree **(Figure 1a),** UPGMA method is adopted assuming that divergence of maize and teosinte dates back to 5000B.C [3, 4, 5]. This node (node 7) in the UPGMA tree **(Figure 1)** shows a bootstrap support of 51%. Calibration of UPGMA tree with time interval of 7000 yrs (5000B.C.) in node 7 (split of maize and teosinte), gives an estimate of divergence times among the *Opaque*-2 genes of various members of the genus *Zea*. The UPGMA tree also provides information about the evolutionary rate at the *Opaque*-2 locus in the *Zea* lineage which is calculated over the branch with the highest bootstrap support value [22]. This rate is found to be 0.00087.

Both maize *ohp-1* and *ohp-2* (*Opaque*-2 heterodimerizing protein 1 and 2) show sequence similarity with *Opaque*-2 proteins of maize (*Z. mays* ssp. *mays*), teosinte (*Z. mays* ssp. *mexicana*, *Z. mays* ssp. *parviglumis* and *Z. diploperennis*) and *Sorghum*. BLASTp indicates the presence of bZIP domain in all of these proteins. The positions of the bZIP domain in these sequences vary among maize and teosintes suggesting the occurrence of rearrangements in O*paque*-2 and *ohp-1,* 2 genes since maize splitted from the teosintes. Multiple sequence alignment between these proteins reveals a pattern of divergence between maize *ohp-1,* 2 and maize and teosinte *Opaque*-2 proteins **(Figure 2a).** The NJ tree suggests the paralogous nature of O*paque-2* and *ohp-1,* 2 genes in maize and the nodes are supported by high bootstrap values **(Figure 3a)**. Maximum Likelihood tree **(Figure 3b)**, having highly significant branching patterns, suggests that *Opaque-2* and *ohp*-1, *2* separated earlier than the separation of rice and maize line. This indicates an ancient duplication event in the O*paque-2* locus which occurred earlier than the separation of the tribes of Poaceae after which one copy underwent subfunctionalization resulting in the fact that *ohp-1* and *ohp-2* can't bind DNA as homodimer, they need heterodimer formation with *Opaque-2* protein for functioning. This subfunctionalization may be viewed as the first step of neofunctionlisation. In order to find out the molecular basis of this evolutionary subfunctionalization, a multiple sequence alignment approach of these three maize proteins (*Opaque-2*, *ohp-1* and *ohp-2*) has been effectively adopted (**Figure 2b**). In position 290, *Opaque*-2 protein has a leucine residue which takes part in zipper formation with the leucine residue present in position 283. *ohp-1, 2* has a valine substituted in place of leucine in this position within the bZIP domain (residue 290). Thus, though being present, the leucine 283 can't form a zipper owing to the absence of its partner seven residues apart in *ohp-1,* 2 and so they are dependent on their paralogue (*Opaque-2*) for effective interaction with DNA. There is no earlier report on the molecular biology of subfunctionalization in the *Opaque-2* locus that has been revealed in the present study. Thus, in the present study, paralogy has been established between maize *Opaque*-2 and *ohp*-1, 2 proteins. These genes are the products of a duplication event which occurred long before the separation of rice and maize-*Sorghum* lineage. Meme analysis reveals the presence of two small motifs in all the proteins under study. Patterns of distribution of these motifs differ in maize, teosinte and *Sorghum Opaque*-2 when compared to their distribution in maize *ohp-1* and 2 and *Oryza* bZIP proteins (**Figure 4a**). To pinpoint this pattern difference, sequence logos of these two motifs has been analyzed. The leucine residue present in position 7 of the first motif seems to be information rich and universally present. The second motif seems to be very important in the present context. It has two information rich leucine residues in position 2 and 9 which, being seven residues apart, is capable of leucine zipper formation. Three copies of motif 2 is present in maize *Opaque-2*.

A copy of this motif in the bZIP domain is absent in maize *ohp-1* and *ohp-2* (**Figure 4a**). All the residues in the bZIP region of maize *ohp-1* and *ohp-2* are 100% identical in the two proteins **(Figure 4b)** suggesting that evolutionary selection has limited the sequence divergence in this region and the conserved sequence pattern is highly essential for the functional role of the proteins. Distance matrix **(Figure 5)** reveals that distance between maize *ohp-1* and *Sorghum Opaque-2* is 0.018 (Standard error=0.011). However the *Sorghum* protein has a greater distance (0.027, SE=0.016) with maize *ohp-2*. Thus, *ohp-1* and *ohp-2* in maize seems to originate from two different ancestors one of which is more closely related to *Sorghum* than the other. ML tree suggests that allotetraploidization between two ancestors (presently extinct, one of which is more closely related to *Sorghum* than the other) has given rise to modern corn, that is, two copies of maize *Opaque-2 heterodimerising protein* (*ohp-1* and *ohp-2*) are orthologous. Tajima's relative rate test between *ohp-1* and *ohp-2* using *Oryza* bZIP protein as outer group gives a chi-square value of 3.19 at 1 degree of freedom suggesting uniformity of substitution rates in this lineage. Tajima's D value is calculated to be 1.14 which suggests non-neutrality and operation of selection in this locus. UPGMA tree with 1000 bootstrap replicates is calculated and calibrated using the divergence time of 50 million years between rice and the ancestor of maize-*Sorghum* as proposed by Wolfe *et. al*., [21]. The time of divergence between the two lineages (one giving rise to the maize ancestor contributing *ohp-2* and the other radiating into *Sorghum Opaque-2* and maize *ohp-1*) is 16.2 million years. Under the assumption of molecular clock (as supported by Tajima's relative rate test), evolutionary rate of this locus is 0.0009 **(Figure 1b).**

Table 1: Pair-wise distances of the sequences under study along with SE values within parenthesis. The closest value between *maize ohp-1* and *Sorghum opaque2* is shown in highlight along with its standard error.

	maize-ohp1	***Sorghum opaque2***	***maize-ohp2***	***Oryza bZIP***	***Hordeum***	***Triticum***	***Arabidopsis***
maize-ohp1		*0.026*	*0.020*	*0.030*	*0.047*	*0.122*	*0.121*
Sorghum opaque2	*0.018*		*0.016*	*0.033*	*0.056*	*0.137*	*0.134*
maize-ohp2	*0.045*	*0.027*		*0.036*	*0.056*	*0.113*	*0.127*
Oryza bZIP	*0.097*	*0.106*	*0.135*		*0.045*	*0.118*	*0.126*
Hordeum	*0.267*	*0.324*	*0.326*	*0.246*		*0.106*	*0.092*
Triticum	*0.698*	*0.737*	*0.644*	*0.651*	*0.597*		*0.119*
Arabidopsis	*0.715*	*0.755*	*0.746*	*0.730*	*0.563*	*0.703*	

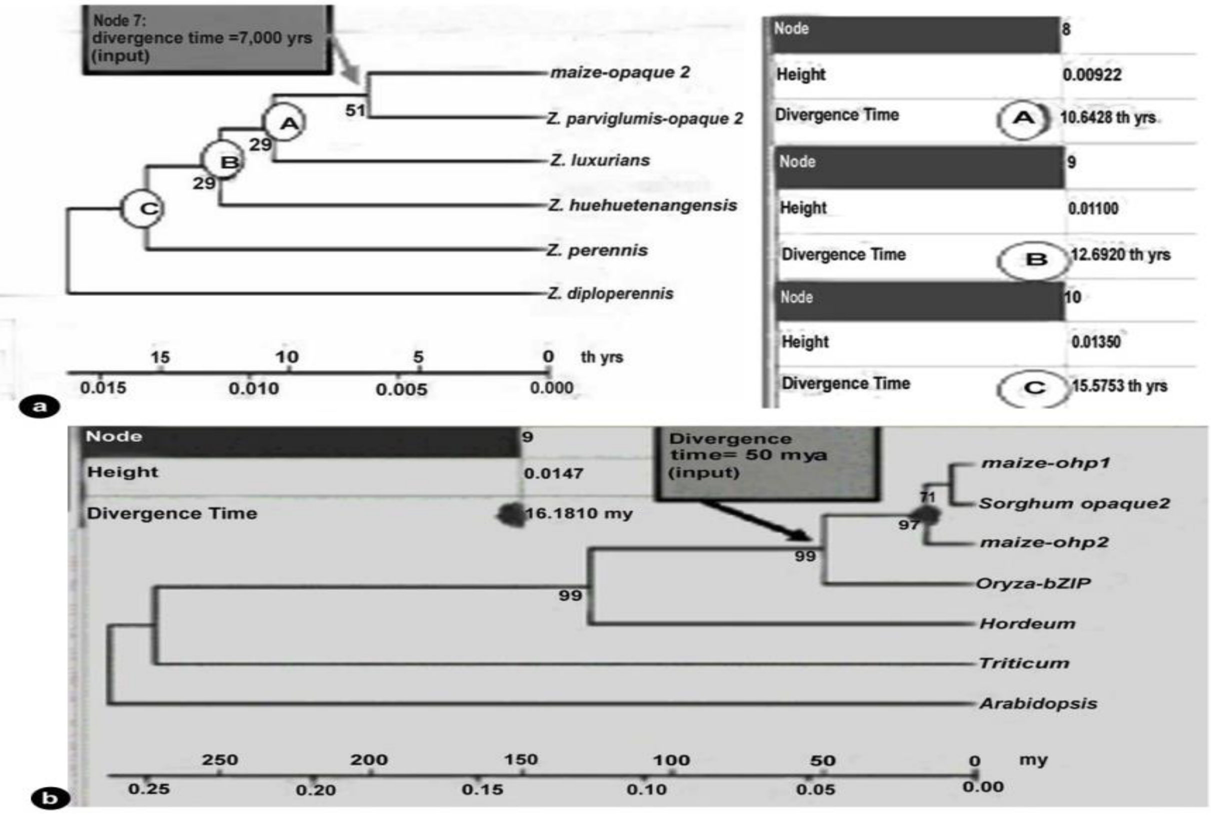

Fig. 1 UPGMA trees showing divergence times calculated based on evolutionary rates at maize (a) *Opaque-2* (b) *ohp-1* and *ohp-2* loci. (The times used for tree calibration are shown in grey boxes. Values on nodes represent bootspot support in percentage).

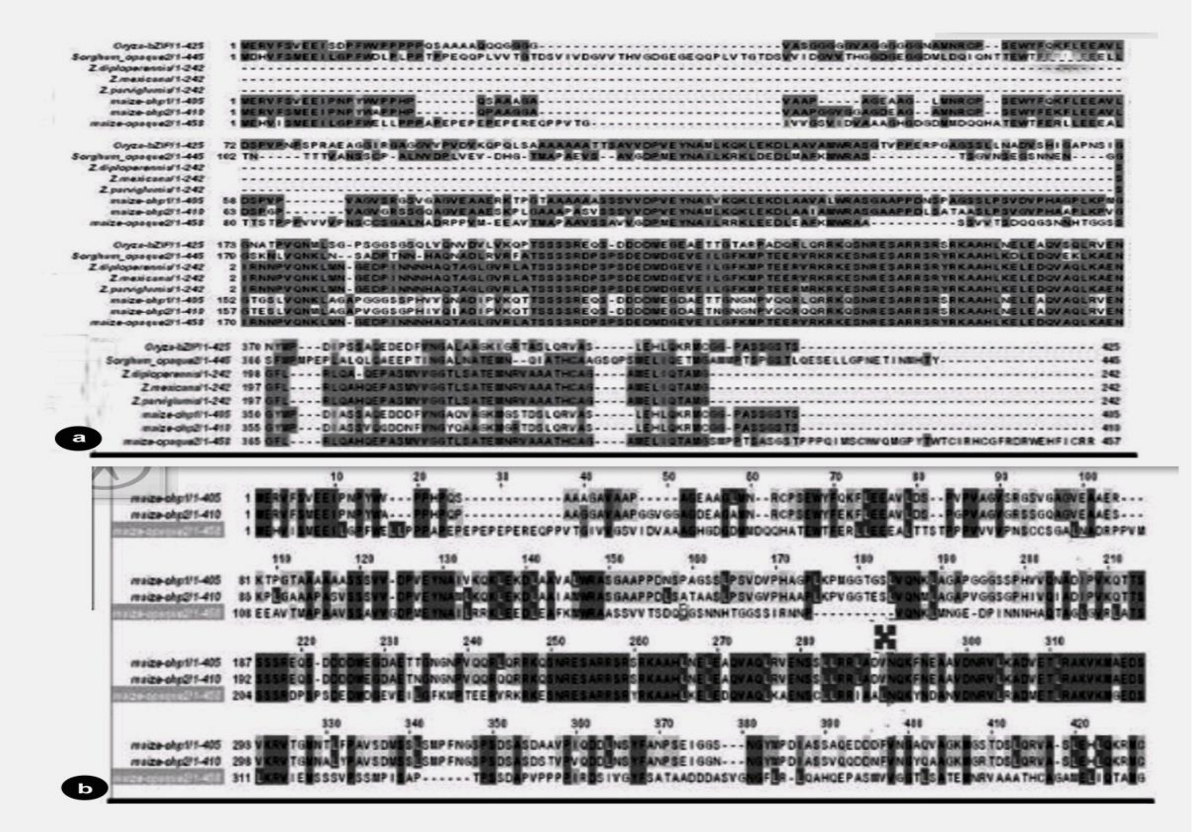

Fig. 2 (a) Multiple sequence alignment showing distinct pattern of residue distribution between maize-teosinte *Opaque-2* and maize *ohp-1*, *ohp-2*. (b) Molecular basis of sub functionalization of *ohp-1*, *ohp-2* (highlighted leucine residues) in contrast to the *Opaque-2* sequence.

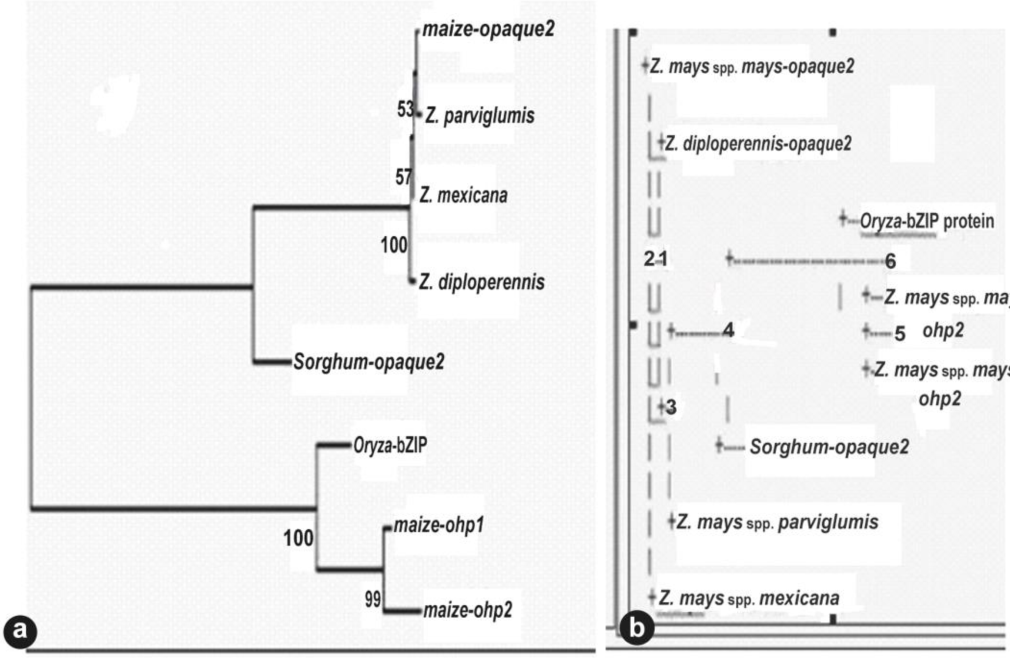

Fig. 3. (a) Neighbor-Joining tree (b) Maximum Likelihood tree showing distinct clustering between maize-teosinte *Opaque-2* and maize *ohp-1*, *ohp-2* sequences.

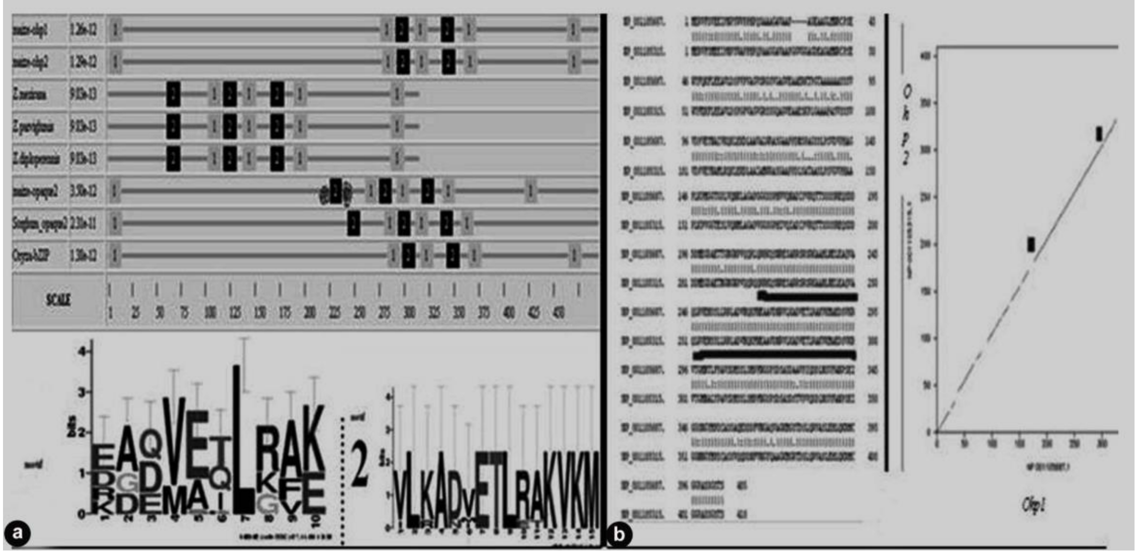

Fig. 4. (a) MEME showing comparative account of motif distribution along with the sequence logo of the two motifs found an extra copy of motif 2 in maize *Opaque-2* (third sequence from below) is to be noted. (b) Pairwise alignment and dot matrix comparison showing 100 percent sequence similarity in the bZIP region of maize *ohp-1* (upper sequence along x axis) and *ohp-2* (lower sequence along y axis).

ACKNOWLEDGMENTS

Financial assistance provided by University Grant Commission (UGC) New Delhi, is gratefully acknowledged. We also thank to Department of Science and Technology (DST-FIST programme), Govt. of India, Department of Botany, University of Kalyani for Instrumental facilities.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

1. A. M. Henry, D. Manicacci, M. Falque, C. Damerval. Molecular evolution of the *Opaque-2* Gene in *Zea mays* L.J Mol Evol 2005; 61:550-558.
2. J. F. Crow, J. Kermicle. Oliver Nelson and Quality Protein Maize. Genetics. 2002; 160: 819-821.
3. F. Gavazzi, B. Lazzari, P. Ciceri, E. Gianazza, A. Viotti. Wild type Opaque2 and Defective opaque2 Polypeptides Form Complexes in Maize Endosperm Cells and Bind the Opaque2-Zein Target Site $^{1[OA]}$. Plant Physiol. 2007; 145: 933-945.
4. S. Guindon, F. Delsuc, J.F. Dufayard, O. Gascuel. Estimating Maximum Likelihood Phylogenies with PhyML. Bioinformatics for DNA Sequence Analysis. (Methods in Molecular Biology. 2009; 537:113-137.
5. B. S. Gaut, J. F. Doebley. DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci 1997; 94:6809-6814.
6. P. Goloubinoff, S. Paabo, A. C. Wilson. Evolution of maize inferred from sequence diversity of the *Adh2* gene segment from archaeological specimens. Proc Natl Acad Sci USA. 1993; 90: 1997-2001.
7. W. C. Galinat. The origin of maize. Annu. Rev. Genet.1971; 5:447-478.
8. G.W. Beadle. The ancestry of corn. Sci. Am. 1980; 242: 112-119.
9. J. Doebley. The genetics of maize evolution. Annu. Rev. Genet. 2004; 38: 37-59.
10. R. R. Hudson, M. Kreitman, M. Aguadé. A Test of Neutral Molecular Evolution Based on Nucleotide Data. Genetics1987; 116: 153-159.
11. S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman. Basic local alignment search tool. J. Mol. Biol. 1990; 215: 403-410.
12. M. Kimura. Evolutionary rate at the molecular level. Nature 1968; 217: 624-626.
13. J. Felsenstein. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol.1981; 17:368-376.
14. D. M. Hillis, M.W. Allard, M. M. Miyamoto. Analysis of DNA sequence data: phylogenetic inference. Method Enzymol. 1993; 224: 456-87.
15. M. Nei, T. Gojobori. Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions. Mol Biol Evol. 1986; 3: 418-426.

16. 17. S. Kumar, K. Tamura, M. Nei. MEGA4: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief. Bioinform. 2004; 5: 150-163.
17. L. Bromham, D. Penny. The Modern Molecular Clock. Nat. Rev. Genet. 2003; 4: 216-224.
18. F. Tajima. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. Genetics 1989; 123: 585-595.
19. K. Tamura, J. Dudley, M. Nei, S. Kumar. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol. Biol. Evol. 2007; 24: 1596-1599.
20. K.H. Wolfe, W.H. Li, P.M. Sharp. (1989). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA. 1987; 5:9054-9058.
21. M. R. Dietrich. The Origins of the Neutral Theory of Molecular Evolution. J. Hist. Biol. 1994; 27: 21-59.

Received:05.05.18, Accepted: 07.06.18, Published:01.07.2018

***Corresponding Author:**
Parthadeb Ghosh*
Email: pdggene@rediffmail.com