***Original Research Article***

# *In-Silico* Structural Annotation of Hypothetical Protein Zm00014a_031618 (*Zea mays*)


## ABSTRACT

**Aims:** The aim of this study is to use *In-Silico* studies for structural annotation of hypothetical Protein Zm00014a_031618 (*Zea mays*) retrieved from NCBI.

**Study design:** The use of bioinformatics *In-Silico* studies for the structural annotation of a hypothetical protein.

**Place and Duration of Study:** The research was conducted at the Bioinformatics Laboratory, Chevron Biotechnology Centre, Modibbo Adama University of Yola, Nigeria. Between June 2018 to July 2018.

**Methodology:** The hypothetical protein was retrieved from NCBI, physical and chemical parameters was calculated using expasy's ProtParam server, the server SOPMA was used for secondary structure analysis (helix, sheets, and coils) and I-Tasser l was use to obtain the 3D structure.

**Results:** Expasy's ProtParam server computated the various physical and chemical parameters such as molecular weight (MW) 30117.97, total number of positively (+R) 27, negatively charged residues (-R) 30, theoretical isoelectric point (pI) 5.96, aliphatic index (AI) 103.67 and grand average hydropathy (GRAVY) 0.293. The SOPMA server was used for calculating the secondary structural features of protein sequences as Alpha helix 39.16%, 310 helix 0.00%, Pi helix 0.00%, Beta bridge 0.00%, Extended strand 14.69%, Beta turn 6.64%, Bend region 0.00%, Random coil 39.51%, Ambiguous states 0.00% and Other states 0.00%. I-Tasser was used for predicting the 3D structure 2qttA from PDB as template.

**Conclusion:** This study has help in understanding the structural analysis of the hypothetical Protein Zm00014a_031618 (Zea mays)


*Keywords: [Zea mays, Hypothetical protein, structural annotation and Intraspecific }*

## 1. INTRODUCTION

*Zea mays* for more than hundreds years has been a subject of genetics studies (1). It is one of the most extensively studied plant species in genetics and it is also been use as research model for genome evolution and genetic diversity (2,3). The genome is made up of 10 chromosomes with its size approximately 2.3 to 2.7 Gb and the genome is also diploid genetically (4,5,6,7,). Just like other larger genome of plant species, the *Z. mays* genome is typically made up of nongenic, low-copy DNA that harbor single genes or repetitive fraction punctuated by islands of unique. The repetitive elements is highly responsible for the wide range of diversity within the species which includes ribosomal DNA (rDNA), transposable elements (TEs) and high-copy short-tandem repeats mostly present at the centromeres, telomeres, and heterochromatin knobs (8,9,10,11). *Z. mays* plant has an extraordinary levels of genomic diversity, phenotypic (12) and transcriptomic (13,14,15). Looking at the genomic level *Z. mays* exhibits a high level of INDEL Polymorphisms (16) and (17) and Single Nucleotide Polymorphisms (18). Averagely the frequency of single nucleotide polymorphism (SNP) between two maize inbreds is said to be approximately 1 substitution per 100 bases (19,20). Recent studies using sequencing data have shown that maize genome exhibits rather variable levels of naturally occurring genetic diversity which depends on the lines involved in the comparison (21,22).

Intraspecific genome variation has been long attributed to changes in size of heterochromatic DNA outside coding sequences that expanded and contracted the chromosomes (98). Intraspecific variations which are approximately 38.8% from the average of 5.5 pg/2n nucleus have been reported in Zea mays (23,24,25,26,27,28). *Z. mays* is known to have large amount of intraspecific sequence variation (29,30) in form of deletion/insertion and single nucleotide polymorphisms. The main mechanism which have effect in the generation of intraspecific genome diversity and in the evolution of the maize genome, segmental duplications and whole genome duplications (polyploidization), retrotransposition and DNA transposition, expansion/contraction of simple sequence repeats (SSRs) and single base mutations and translocation of genes or gene segments by transposons and capture (31,32). Intraspecific allelic variation is mostly as a result of qualitative changes that change the nature of the gene products and quantitative changes it also alter the amount of the gene product produced. Quantitative changes in gene expression can be as a result of cis- or trans variations in gene regulation (33).
The present study focused on the *In Silico* Structural Annotation of Hypothetical Protein Zm00014a_031618 (*Zea mays*)

## 2. MATERIAL AND METHODS

### 2.1 Sequence Retrieval

Amino acid sequence of hypothetical Protein Zm00014a_031618 (*Zea mays*) was retrieved from NCBI database (www.ncbi.nlm.nih.gov/protein/1394916517).

### 2.2 Physiochemical Analysis of the Protein

The physiochemical properties of the hypothetical protein such as molecular weight, atomic composition, amino acid composition, theoretical pI, instability index, and grand average of hydrophobicity (GRAVY) using ProtParam tool (web.expasy.org/cgi-bin/protparam/protparam) (34).

### 2.3 Secondary Structure Analysis

The server SOPMA was use for secondary structure analysis (helix, sheets, and coils) of the hypothetical protein using SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) (35).

### 2.4 Sequence Analysis

BLASTP is mostly used for the identification of a query amino acid sequence and to find similar sequences in protein databases. The FASTA sequence of the Hypothetical Protein Zm00014a_031618 (*Zea mays*) was the query sequence and similar proteins in different databases were searched for using the BLASTP program (36).

### 2.5 3D Structural Prediction

The 3D structural was predicted with the use of I-Tasser (37)

## 3. RESULTS AND DISCUSSION

The present study focused on the *In Silico* Structural Annotation of a hypothetical Protein Zm00014a_031618 (*Zea mays*) from NCBI database.
The presented in table 1 showed the physicochemical characterization of Hypothetical Protein Zm00014a_031618 (*Zea mays*) with 286 amino acid sequence using Expasy's ProtParam server where the computation of various physical and chemical parameters was performed. Molecular weight (MW), total number of positively (+R), negatively charged residues (-R), theoretical isoelectric point (pI), extinction coefficient (EC), aliphatic index (AI) and grand average hydropathy (GRAVY).

**Table 1: Physiochemical Analysis of the Hypothetical Protein**

| Molecular weight (Da) | pI | -R | +R | EC | II | AI | GRAVY |
|---|---|---|---|---|---|---|---|
| 30117.97 | 5.96 | 30 | 27 | 13200 | 23.10 | 103.67 | 0.293 |

The results are presented in Table 2 showed the SOPMA which was used for calculating the secondary structural features of protein sequences such as Alpha helix, 310 helix, Pi helix, Beta bridge, Extended strand, Beta turn, Bend region, Random coil, Ambiguous states and Other states.

**Table 2. SOPMA of the Hypothetical Protein**

| Parameter | % content | Parameter | % content |
|---|---|---|---|
| Alpha helix | 39.16% | Beta turn | 6.64% |
| 310 helix | 0.00% | Bend region | 0.00% |
| Pi helix | 0.00% | Random coil | 39.51% |
| Beta bridge | 0.00% | Ambiguous states | 0.00% |
| Extended strand | 14.69% | Other states | 0.00% |

```
>PWZ58979.1 hypothetical protein Zm00014a_031618 [Zea mays]
MAAEAGPISKVLIVVGNPTPCCSLRPKALLCSVSRFAYSVGIGLCSGLDAAMQTEAMPLVHKFKLVEAPA
HESTFPKGAPWVRYHGNYKGLHIDLVLPGKDAVLGVDSVGTVSAALLTSFSIQTLKPDLIINAGTAGGFK
AKGASIGDVFLASDVSFHDRRIPIPVFDMYGIGARKTSAVPNILKELNLKIGKLSTGDSLDMSPQDEKVI
LSNDATVKDMEGAAVAYVADMFSTPAIFVKAVTDIVDGEKPTSEEFLQNLIAVTAALDLAVTKVVDFISG
KRISDL
```

**Fig. 1. Faster sequence of the hypothetical Protein Zm00014a_031618 (*Zea mays*)**

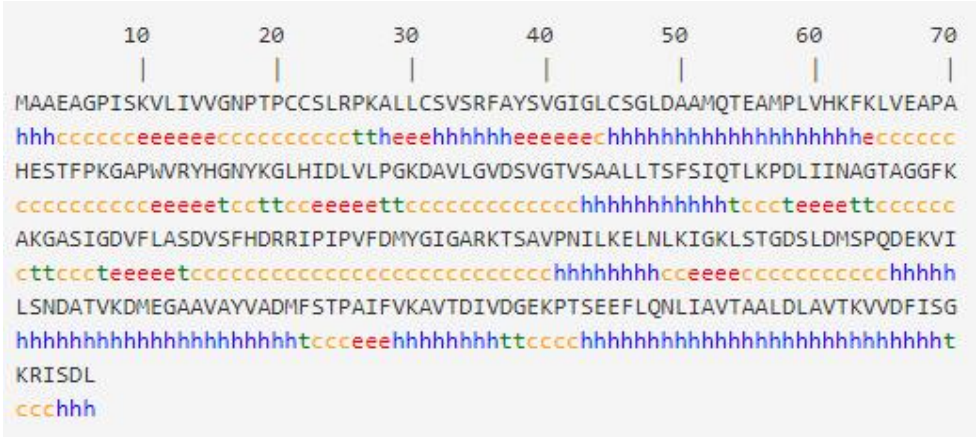


**Fig. 2. SOPMA of structural prediction of the hypothetical Protein**

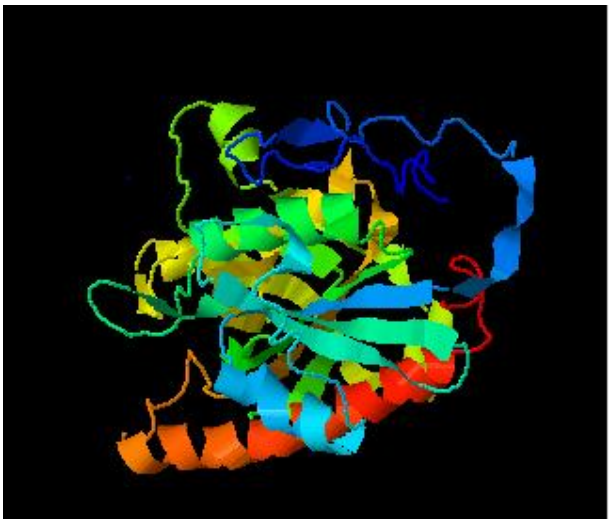

**Fig. 3. The prediction 3D structural using I-Tasser of the Hypothetical Protein the temple used is 2qttA from PDB**

The instability index (II) was computed to be 23.10 which make the hypothetical protein classified as a stable protein because a protein whose instability index is less than 40 is said to be a stable protein (38). The protein was predicted to contain several helices which is consistent with the ProtParam results present in Figure 1 this makes the protein more flexible for folding which is likely to increase the protein interaction. The sequence of the hypothetical protein was found to be rich in alanine. The proteins with very high AI may show stability in a wide temperature range where lower AI proteins are not thermal stable and show more flexibility. The amino acid sequences which had most in number are alanine (32) followed by leucine and valine (28), glycine and serine (23) and while the least is Tryptophan (1). The hypothetical protein had a total number of 30 negatively charged residues (Asp + Glu) and total number of 27 positively charged residues (Arg + Lys). The molecular formula of the protein was found as $C_{1354}H_{2181}N_{349}O_{401}S_{11}$. The GRAVY was shown to be 0.293GRAVY which shows a better interaction of protein and water is occurring in low GRAVY. (Ikai AJ. 1980). The secondary structure of the hypothetical protein was predicted by SOPMA serve showed the random coil was the most predominant (39.51%), followed by alpha helix (39.16%), followed extended strand (14.69%) and beta turn (6.64%) was the least. I-Tasser modeling server was used to get the 3D structure of the model with a C-score of 1.03 which is the highest score compare to the remaining four other 3D structure.

## 4. CONCLUSION

*In silico* studies helps in the annotation of protein structure with reduction in the cost of material and time. This study has help in understanding the structural analysis of the hypothetical Protein Zm00014a_031618 (*Zea mays*).

## CONSENT

Not Applicable

## ETHICAL APPROVAL

Not Applicable

## REFERENCES

1. COE E, EAST E. THE BIRTH OF MAIZE GENETICS. IN MAIZE HANDBOOK.VOLUME II: GENETICS AND GENOMICS. EDITED BY BENNETZEN JL, HAKE S. NEW YORK,USA: SPRINGER; 2009:3–15.

2. HE E. “THE ORIGINS OF MAIZE GENETICS,” NATURE REVIEWS GENETICS, L. 2001:2(11);898–905.

3. LB J. “THE FUTURE OFMAIZE,” IN HANDBOOK OFMAIZE: GENETICS AND GENOMICS, J. L. BENNETZEN AND S. HAKE, EDS., SPRINGER, BERLIN, GERMANY, 2009. 771–779.

4. BENNETZEN JL. “MAIZE GENOME STRUCTURE AND EVOLUTION,” IN HANDBOOK OF MAIZE: GENETICS AND GENOMICS, J. L. BENNETZEN AND S. HAKE, EDS., SPRINGER, BERLIN, GERMANY. 2009:179–199

5. SCHNABLE PS, WARE D, FULTON RS, *et al*., “THE B73 MAIZE GENOME: COMPLEXITY, DIVERSITY, AND DYNAMICS,” SCIENCE. 2009;326(5956):1112–1115.

6. RAYBURN AL, BIRADAR DB. BULLOCK DG, MCMURPHY LM. “NUCLEAR DNA CONTENT IN F1 HYBRIDS OF MAIZE,” HEREDITY. 1993;70:294–300.

7. ZHOU S, WEI F, NGUYEN J. *et al*., “A SINGLE MOLECULE SCAFFOLD FOR THE MAIZE GENOME,” PLOS GENETICS. 2009;5(11).

8. MCCLINTOCK B. “THE ORDER OF GENES C, SH, AND WX IN ZEA MAYS WITH REFERENCE TO A CYTOLOGICAL KNOWN POINT ON THE CHROMOSOME,” PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA. 1931;17(8)485–491.

9. PEACOCK WJ, DENNIS ES, RHOADES, PRYOR AJ. “HIGHLY REPEATED DNA SEQUENCE LIMITED TO KNOB HETEROCHROMATIN IN MAIZE,” PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, 1981;78(7):4490–4494.

10. ANANIEV EV, PHILLIPS RL, RINES HW, “CHROMOSOMESPECIFIC MOLECULAR ORGANIZATION OF MAIZE (ZEA MAYS L.) CENTROMERIC REGIONS,” PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA. 1998;95(22)13073–13078.

11. MORGANTE M., “PLANT GENOME ORGANISATION AND DIVERSITY: THE YEAR OF THE JUNK!,” CURRENT OPINION IN BIOTECHNOLOGY. 2006;17(2)168–173.

12. BUCKLER ES, GAUT BS, MCMULLEN MD. MOLECULAR AND FUNCTIONAL DIVERSITY OF MAIZE. CURRENT OPINION IN PLANT BIOLOGY. 2006;9(2):172–6.

13. FLINT-GARCIA SA, THUILLET AC, YU J, PRESSOIR G, ROMERO S M, MITCHELL SE, DOEBLEY J, et al. MAIZE ASSOCIATION POPULATION: A HIGH-RESOLUTION PLATFORM FOR QUANTITATIVE TRAIT LOCUS DISSECTION. THE PLANT JOURNAL : FOR CELL AND MOLECULAR BIOLOGY. 2005;44(6):1054–64.

14. STUPAR RM, SPRINGER NM. CIS-TRANSCRIPTIONAL VARIATION IN MAIZE INBRED LINES B73 AND MO17 LEADS TO ADDITIVE EXPRESSION PATTERNS IN THE F1 HYBRID. GENETICS. 2006;173(4):2199–210.

15. SWANSON-WAGNER R. A, JIA Y, DECOOK R, BORSUK LA, NETTLETON D, SCHNABLE PS. ALL POSSIBLE MODES OF GENE ACTION ARE OBSERVED IN A GLOBAL COMPARISON OF GENE EXPRESSION IN A MAIZE F1 HYBRID AND ITS INBRED PARENTS. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA. 2006;103(18):6805–10.

16. FU Y, WEN TJ, RONIN YI, CHEN HD, GUO L, MESTER DI, YANG Y, *et al*. GENETIC DISSECTION OF INTERMATED RECOMBINANT INBRED LINES USING A NEW GENETIC MAP OF MAIZE. GENETICS. 2006;174(3):1671–83.

17. MESSING J, DOONER HK. ORGANIZATION AND VARIABILITY OF THE MAIZE GENOME. CURRENT OPINION IN PLANT BIOLOGY. 2006;9(2):157–63.

18. VROH BI, MCMULLEN I, SANCHEZ-VILLEDA MD, SCHROEDER H, GARDINER S, POLACCO J, SODERLUND M, et al. SINGLE NUCLEOTIDE POLYMORPHISMS AND INSERTION–DELETIONS FOR GENETIC MARKERS AND ANCHORING THE MAIZE FINGERPRINT CONTIG PHYSICAL MAP. CROP SCIENCE. 2006;46(1):12.

19. TENAILLON M. I., M. C. SAWKINS, A. D. LONG, R. L. GAUT, J. F. DOEBLEY, AND B. S. GAUT, "PATTERNS OF DNA SEQUENCE POLYMORPHISM ALONG CHROMOSOME 1 OF MAIZE (ZEA MAYS SSP. MAYS L.)," PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA. 2001;98(16)9161–9166.

20. CHING A, CALDWELL KS, JUNG M. *et al.,* "SNP FREQUENCY, HAPLOTYPE STRUCTURE AND LINKAGE DISEQUILIBRIUM IN ELITEMAIZE INBRED LINES," BMC GENETICS. 2002;3,(19)

21. A. RAFALSKI AND E. ANANIEV, "GENETIC DIVERSITY, LINKAGE DISEQUILIBRIUM AND ASSOCIATION MAPPING," IN HANDBOOK OF MAIZE: GENETICS AND GENOMICS, J. L. BENNETZEN AND S. HAKE, EDS., PP. 201–219, SPRINGER, BERLIN, GERMANY, 2009

22. RAFALSKI A. MORGANTE M, "CORN AND HUMANS: RECOMBINATION AND LINKAGE DISEQUILIBRIUM IN TWO GENOMES OF SIMILAR SIZE," TRENDS IN GENETICS. 2004;20(2)103–111,

23. RAYBURN AL, PRICE HJ, SMITH JD, GOLD JR. "C-BAND HETEROCHROMATIN AND DNA CONTENT IN ZEA MAYS," AMERICAN JOURNAL OF BOTANY. 1985;72(10):1610–1617,.

24. RAYBURN AL, "FLOW CYTOMETRIC ASSESSMENT OF NUCLEOTIDE VARIABILITY AND ITS EVOLUTIONARY IMPLICATIONS," IN CLASSICAL AND MOLECULAR CYTOGENETIC ANALYSIS, W.J. RAUP AND B.S. GILL, EDS, KANSAS AGRICULTURAL EXPERIMENTAL STATION, MANHATTAN, KAN, USA, 1994;110–115

25. LEE JH, ARUMUGANATHAN K, KAEPPLER S. M. *ET AL*., "VARIABILITY OF CHROMOSOMAL DNA CONTENTS IN MAIZE (ZEA MAYS L.) INBRED AND HYBRID LINES," PLANTA, 2002;215(4)666–671,

26. LAURIE DA, BENNET MD. "NUCLEAR DNA CONTENT IN THE GENERA ZEA AND SORGHUM. INTERGENERIC, INTERSPECIFIC AND INTRASPECIFIC VARIATION," HEREDITY. 1985;55(3);307–313.

27. BIRADAR DP, RAYBURN AL. "HETEROSIS AND NUCLEAR DNA CONTENT IN MAIZE," HEREDITY. 1993;71(3)300–304,.

28. A. RAFALSKI AND E. ANANIEV, "GENETIC DIVERSITY, LINKAGE DISEQUILIBRIUM AND ASSOCIATION MAPPING," IN HANDBOOK OF MAIZE: GENETICS AND GENOMICS, J. L. BENNETZEN AND S. HAKE, EDS., SPRINGER, BERLIN, GERMANY, 2009;201–219

29. TENAILLON MI, SAWKINS MC, LONG AD, GAUT RL, DOEBLEY JF. *et al*., PATTERNS OF DNA SEQUENCE POLYMORPHISM ALONG CHROMOSOME 1 OF MAIZE (*ZEA MAYS SSP. MAYS L*). PROC. NATL. ACAD. SCI. 2001;98:9161–9166.

30. VROH BI, MCMULLEN M. D., SANCHEZ-VILLEDA H, SCHROEDER S, GARDINER J. et al., SINGLE NUCLEOTIDE POLYMORPHISMS AND INSERTION–DELETIONS FOR GENETIC MARKERS AND ANCHORING THE MAIZE FINGERPRINT CONTIG PHYSICAL MAP. CROP SCI. 2005;46:12–21

31. RAFALSKI A., ANANIEV E, "GENETIC DIVERSITY, LINKAGE DISEQUILIBRIUM AND ASSOCIATION MAPPING," IN HANDBOOK OF MAIZE: GENETICS AND GENOMICS, J. L. BENNETZEN AND S. HAKE, EDS., SPRINGER, BERLIN, GERMANY, 2009;201–219,

32. DOEBLEY J. "MOLECULAR EVIDENCE FOR GENE FLOW AMONG ZEA SPECIES," BIOSCIENCE. 1990;40(6) 443–448.

33. WITTKOPP PJ, HAERUM BK, CLARK AG. EVOLUTIONARY CHANGES IN CIS AND TRANS GENE REGULATION. NATURE. 2004;430:85–88.

34. GASTEIGER E, HOOGLAND C, GATTIKER A, DUVAUD S, WILKINS MR, et al. PROTEIN IDENTIFICATION AND ANALYSIS TOOLS ON THE EXPASY SERVER. THE PROTEOMICS PROTOCOLS HANDBOOK. 2005;571-607

35. GEOURJON C, DELEAGE G SOPMA: SIGNIFICANT IMPROVEMENTS IN PROTEIN SECONDARY STRUCTURE PREDICTION BY PREDICTION FROM MULTIPLE ALIGNMENTS. COMPUT APPLIC BIOCI. 1995;11:681-684.

36. ALTSCHUL SF, GISH W, MILLER W, MYERS EW, LIPMAN DJ. BASICLOCAL ALIGNMENT SEARCH TOOL. J MOL BIOL. 1990;215:403—10

37. YANG Z. I-TASSER: FULLY AUTOMATED PROTEIN STRUCTURE PREDICTION IN CASP8. PROTEINS. 2009;77(9):100-113.

GURUPRASAD K. REDDY BV, PANDIT MW. PROTEIN ENG. 1990;4:155—61

38. IKAI AJ. THERMO STABILITY AND ALIPHATIC INDEX OF GLOBULAR PROTEINS. JOURNAL OF BIOCHEMISTRY. 1980;88:1895-1898