American-Eurasian J. Agric. & Environ. Sci., 16 (6): 1176-1180, 2016
ISSN 1818-6769

DOI: 10.5829/idosi.aejaes.2016.16.6.12975

Mutational Modeling of Barley Yellow Dwarf Virus

[1]Muhammad Jamil, [2]Hameed Ramzan, [1]Noman Latif, [1]Muhammad Mansoor,
[1]Muhammad Ehsan Elahi, [1]Fawad Anwar, [1]Nazir Hussain and [1]Amanullah

[1]Arid Zone Research Centre (AZRC, PARC), Dera Ismail Khan - 29050, Pakistan
[2]COMSATS Institute of Information Technology Islamabad, Pakistan

Abstract: Barley yellow dwarf virus (BYDV) is a positive RNA virus which has five different serotypes, has tendency to mutate with time and changes its genome sequence, it has very high mutation rates with one nucleotide per cycle. Mutational model for whole genome sequences of BYDV was constructed that determined a limit where spontaneous mutations in true genome of barley yellow dwarf virus convert the genome into random one. Entropy associated to each n-mer of true and random genome and relative deviation was calculated. The results showed higher deviation for 9-mer. Markovian entropy was found for all n-mers of true as well as random genome. Relative deviation was also calculated for markovian entropy of true genome and random genome which also gave higher value for 9-mer. This shows that at 9-mer information loss is maximum. At entropy value 16.56339512 true genome converts into random one, which is the limit point where true genome will be converted into random genome.

Key words: BYDV · Genetic Underpinning · N-Mers · Mutation · Random Genome · True Genome

NTRODUCTION

Barley Yellow Disease (BYD) considered to be an important viral disease in terms of reducing plant yield worldwide. Barley Yellow Dwarf Virus (BYDV) is the casual organism for this disease in many cereals which include barley, wheat, oat and many others [1]. The mechanism by which this virus infects the plants is not clear but many proteins which are involved in the disease include coat proteins, movement proteins and fusion proteins. The main proteins involved in the interaction with membrane of host are movement proteins [2]. In grasslands it plays role in shifting of community composition [3]. The constraints of a virus on host infection have a genetic underpinning. So, it is important to acquire knowledge about genetic diversity in pathogen populations for understanding their role in ecological processes. To understand the process of genetic diversity it is necessary to have knowledge about what evolutionary forces are acting on the viral genome [4]. At present there are limited numbers of evidences present about genetic diversity. The main reason behind this is lack of studies that investigate variation in non-genomic structures and the way how these variations affect the gene function. Through an analysis of diversity, it is possible to separate out the effects of gene function and phylogenetic history in creating the pattern of diversity [5]. In some cases, variations a virus strain may lead to severe disease symptoms, less productivity and overall less growth of the plant. Different viruses have different effects on the plants therefore it is necessary to understand the process of genetic variation in the viruses [6]. As BYDV is one of the most prominent positive RNA virus of the cereal crops and it resulted in a great loss in the productivity of crop. It has high mutation rate i.e. one nucleotide changes cycle replication [7]. So it is necessary to study the mutation process in this virus.

The evidences about biology, mode of spread, genome organization, role of 3’ sites, ribosomal frameshifting, variation among strains, countries and genetic diversity have been investigated. However little s known about whole genome analysis and mutation modeling of this virus. The main aim of the present study is to create phylogenetic trees for whole genome sequences. Analyze different genes which include coat protein gene, movement protein gene, RNA dependent RNA polymerase gene, fusion protein gene and also that of whole genome.

Corresponding Author: Muhammad Jamil, Arid Zone Research Centre (AZRC, PARC), Dera Ismail Khan - 29050, Pakistan.
Hameed Ramzan COMSATS Institute of Information and Technology Islamabad, Pakistan.

The entropy of true and random genome and also their relative deviation are estimated. In addition explore the mutation model for whole genome of BYDV.

MATERIALS AND METHODS

For mutational modeling and for finding entropies, sequences were obtained for whole genome sequences of BYDV. 45 whole genome sequences of all five serotypes of BYDV were taken from NCBI [8]. The length of each sequence was found which is denoted by Mn. All of the n-mers for each type of sequences were found using software named R. These n-mers (n) were found for n=1, 2, 3, 4, 5, 6, 7, 8 and 9. After this n-mers were counted in each type of sequences. The differences between sequences were found which are denoted by Ks. Then probability of occurrence of each type of n-mer was found. The probability of occurrence is denoted by Ps.

For finding entropy, values of differences between sequences (Ks) (Eq 2) and probability of occurrence of n-mers (Ps) (Eq 3) are needed. As these values can be calculated from Eq 2 and Eq 3 so entropy Hn(G) can be calculated by using these values.

For finding length of genome (Mn), Ks, Ps and Hn(G) different formulas are used which are;

Equation 1: Mn = M – n +1
Equation 2: Ks = Total differences/Total sequences
Equation 3: Ps = Ks/Mn
Equation 4: Hn(G) = $\sum_{s=1}^{4n} Ps \log\log_2 Ps$

An algorithm was created using all these formulas which calculates all Ks, Ps, Hn(G) values. Average entropy values for all n-mers were found.

```
Stream.writer outputfile:
String genomic sequence = Read up End (Genomic sequence)
String n-mer = Readline (n-mer file)
int count = 0
for(i=0;i<genomic sequence.length)
{n-mer = Readline(n-mer)
int length = n-mer.length
String s = genomic sequence.n-mer(i, length)
If(s==n-mer)
{Count++;}
Mn = M – n +1
KS= total number of differences / total number of sequences in the file
PS= Ks/Mn
Hn = $\sum_{s=1}^{4n} Ps \log\log_2 Ps$
Outputfile.write(“Mn:”+Mn+””+”KS:”+KS+””+”PS:”+PS+””+”Hn:”+Hn+””) i=length}
```

Entropy values were obtained for both true genome and random genome sequences. Random genome includes the sequences which have same GC contents as the true genome.

After obtaining the entropy values of true and random genomes comparison of average entropies of n-mers was carried out. Comparison among the average entropies of different n-mers was carried out with help of graph that help to indicate the n-mer which carried the most relevant informations. The n-mer with most relevant information was determined by finding deviation between true genome and random genome entropy. The n-mer with most deviation is considered to be the one which carry most relevant information and it also indicate that here deviation is maximum.

For analyzing mutation process a computational model was made that simulates genome replication. As mutations induced in the sequence the entropy associated to the genome increases and sequence becoming start random. Then markovian entropy was calculated for both true and random genome and their relative deviation was found.

Equation 5: Hn = (n - α)Hβ – (n - β)Hα
Here α represent n-mers and $\beta = \alpha - 1$

Then mutations were induced in the true genome and a limit was determined were a true genome of barley was converted into random one.

RESULTS

The entropy values associated to all n-mers of true and random genome were calculated and are shown in the Table 1.

The relative deviation of all n-mers of random and true genome gave larger value for 9-mer.

So it is clear that 9-mer carries the most relevant information and also here information loss is maximum.

A computational model was constructed that can replicate the true genome of BYDV with induced mutations. 1000 time replications were run for all n-mers and the results of these replications show that entropy associated to the strains increases as the mutations are induced in the strains.

Table 1: Entropy values for random and true genome and their relative deviation

n-mer	Random Genome	True Genome	Relative Deviation
1	1.99682775992	1.9917845195	0.00504
2	3.83958319236	3.83624918585	0.00333
3	5.8293768621	5.81532220601	0.01405
4	7.84233496416	7.82036635443	0.02197
5	9.52167985551	9.46287989064	0.0588
6	11.3461724511	11.1951892267	0.15098
7	13.3020157035	12.9373518231	0.3646638804
8	14.2097592714	13.6299725948	0.5797866766
9	14.7030646647	14.0253218413	0.6777428234

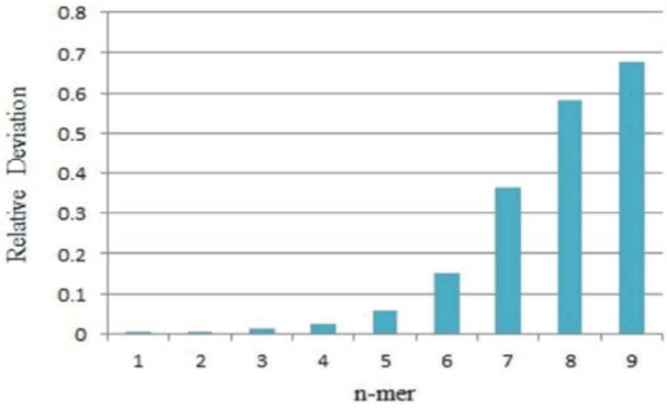

Fig. 1: Relative Deviation of random and true genome entropy

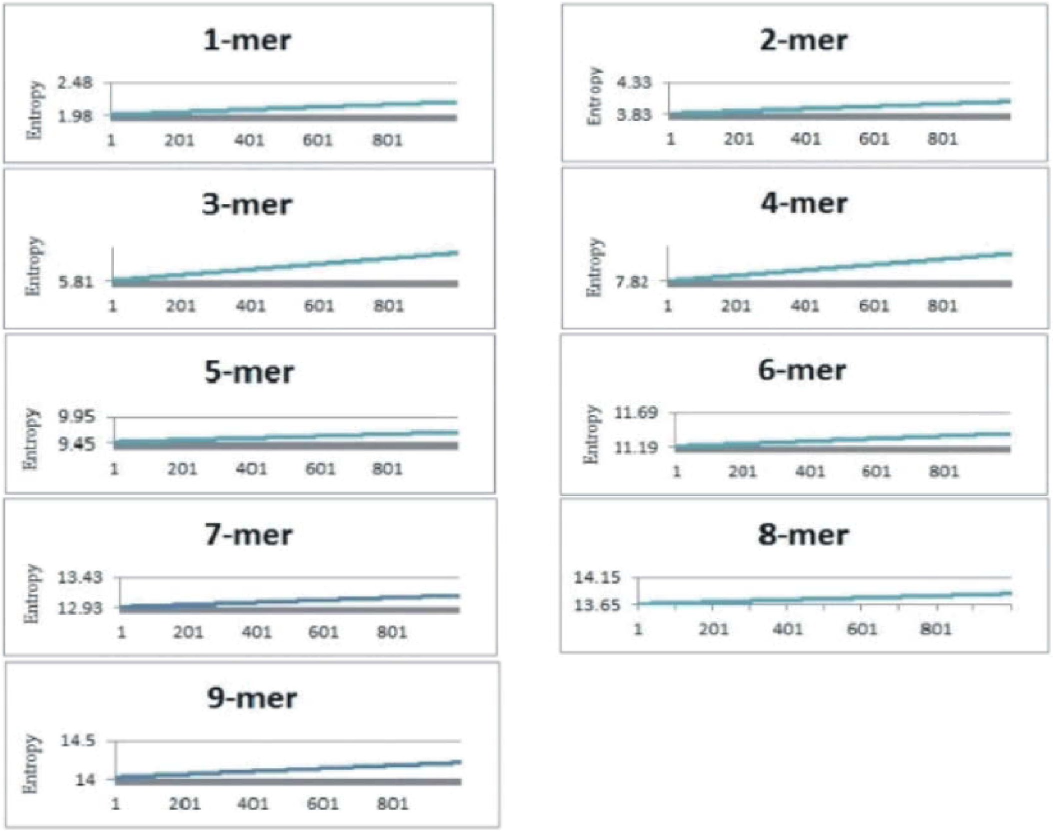

Fig. 2: Replications with induced mutations in true genome of BYDV

Table 2: Markovian entropy of true and random genome and their relative deviation

n-mer	Random Genome	True genome	Relative Deviation
2	3.839583	3.836249185	.003333815
3	5.755857655	5.748017000	.007840655
4	7.728866173	7.713301000	.015565173
5	9.638446075	9.607800000	.030646075
6	11.48496177	11.40395000	.08101177
7	13.34603821	13.20640711	.1396311
8	15.08447380	14.88751704	.20093034
9	16.56339512	16.18468582	.3787093

It is clear that true genome deviates from random genome by a value of 0.3787093. So replications of true genome with induced mutations were carried out until true genome become random

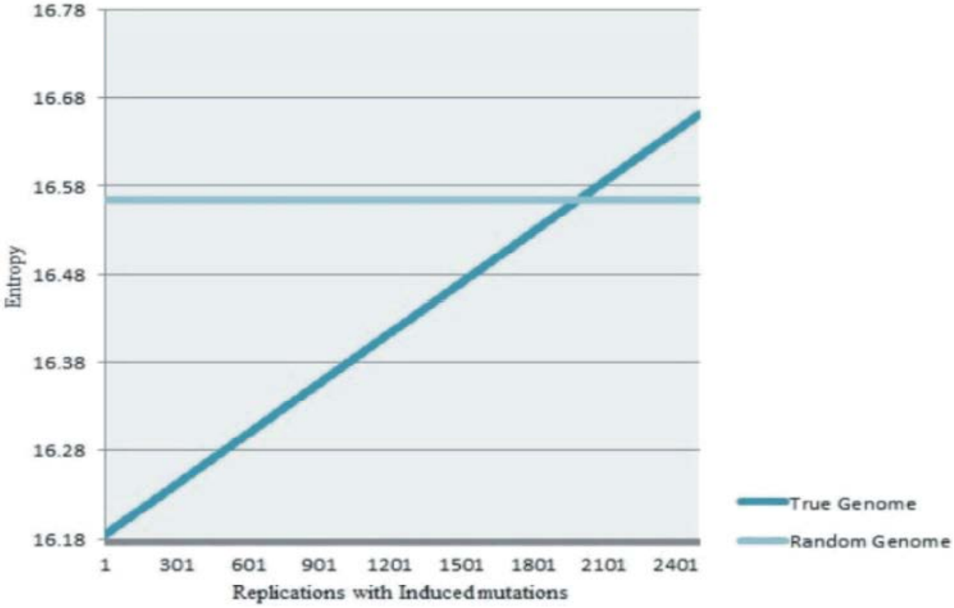

Fig. 3: Replications with induced mutations in true genome and a point where genome become random one

The variation rate in BYDV was measured by calculating relative deviation of true genome and random genome markovian entropy. The relative deviation for 9-mer was found to be 0.3787093 and is greater value.

DISCUSSION

Previous studies showed no evidences about genetic diversity of BYDV. This could be due to lack of informations about the variation in genome and the way how these variations affect the genetic structures [5]. Although the process of genetic diversity is studied in BYDV but it is limited to serotypes PAV and PAS that are that are named on the basis of vector relationship and their origin. Both are transmitted by vector *Rhopalosiphum padi* and *Sitobion avenae* [9].

The main aim of this study is to study mutation process in BYDV. Sequences were obtained from NCBI for whole genome sequences for all five serotypes of BYDV. First of all entropy associated to every n-mer was found and finally average entropy was calculated. For calculating entropy different values were required which include length of genome (Mn) (Eq1), differences between sequences (Ks) (Eq 2), probability of occurrence of n-mer (Ps) (Eq 3). As entropy refers to loss of information so information loss was predicted at every n-mer. In this way entropy loss at every n-mer of true genome was calculated. In the similar manner the average entropies of a random genomes n-mers were calculated (Table 1). It can be easily predicted that entropy increase as size of n-mers increases. Then relative deviation of true and random genome entropy was calculated and the n-mer was predicted where loss of information was maximum. At 9-mer the entropy loss is maximum so, this n-mer carries most relevant information’s and here information loss is maximum (Fig. 1).

Mutation process in BYDV was studied by computational model and replications with induced mutations per cycle were obtained. As BYDV is positive RNA virus so here mutation rate is higher one nucleotide change per replication [7]. All n-mers were 1000 time replicated with induced mutations. The entropy of all n-mers was found to be increased which clearly show that increase in mutation results increase in the entropy (Fig. 2). Then markovian entropy for true genome as well as random genome was calculated (Eq 5). The relative deviation between true and markovian entropy was calculated and again this show the higher entropy value for 9-mer (Table 2). So it is clear that in whole genome of BYDV. the highest information loss occurs at 9-mer and it can be considered the point which carries the maximum information. In order to convert a true genome into random one, true genome was replicated with induced mutations until the limit is obtained where true genome converts into random one. It can be easily predict about the structure changes in virus genome and these will prove helpful in creating vaccines against viruses which change their structure with time. From the present study the limit point can be obtained where a true genome become random one and also that how much information loss has occurred.

CONCLUSION

It can be concluded that a mutation model was developed for measuring information contained in the genome. To determine the limit where a true genome converts into random one. In addition at 9-mer the information loss is maximum and also this mer contains the maximum information's.

REFERENCES

1. Lister, R.M. and R. Ranieri, 1995. Distribution and economic importance of barley yellow dwarf. In Barley Yellow Dwarf: 40 Years of Progress. American Phytopathological Society, pp: 29-53.
2. Krichevsky, A., S.V. Kozlovsky, Y. Gafni and V. Citovsky, 2006. Nuclear import and export of plant virus proteins and genomes. Molecular Plant Pathology, 7: 131-46.
3. Malmstrom, C.M., C C. Hughes, L.A. Newton and C.J. Stoner, 2005. Virus infection in remnant native bunchgrasses from invaded California grasslands. New Phytol., 168: 217-230.
4. Wang, X., S. Chang, Z. Jin, L. Li and G. Zhou, 2001. Nucleotide sequences of the coat protein and read-through protein genes of the Chinese GAV isolate of barley yellow dwarf virus. Acta Virologica, 45: 249-252.
5. Moury, B., 2004. Differential selection of genes of cucumber mosaic virus subgroups. Molecular Biology and Evolution, 21: 1602-1611.
6. Mayo, M.A., 2002. ICTV at the Paris ICV: results of the plenary session and the binomial ballot. Virology Division News, 147: 2254-2260.
7. Drake, J.W. and J.J. Holland, 1999. Mutation rates among RNA viruses. PNAS, 96(24): 13910 -13913.
8. http://www.ncbi.nlm.nih.gov/
9. Hall, G., 2006. Selective constraint and genetic differentiation in geographically distant barley yellow dwarf virus populations. Journal of General Virology, 87: 3067-3075.