УДК 57.2788

Альнаддафа Л.
Доктор университета, Альбаас университет,
факультет «Сельского хозяйства»
Россия, г. Москва
Lina M. Alnaddafa
Biotechnology and molecular biology, faculty of Agricultural
Al-Baath University Homs, Russia, Moscow

APPLICATION OF BIOINFORMATICS IN THE STUDYING OF GENETIC RELATIONSHIPS AMONG WHEAT AND AEGILOPS

Annotation: *Studying genetic relationships among Wheat and Aegilops species using the program NEB Cutter to determine the positions cuts enzymes for the sequences of the internal transcribed spacer (ITS) of them. The sequences of the ITS region were inserted into the program. Nine restriction enzymes (AvaI, DrdI, EcoRV, HaeII, Hin1I, SmaI, XmaI, XmaIII, HindIII) were utilized. The results indicated that there was a high level of polymorphism in ITS loci for Aegilops and Triticum. The phylogenetic relationships among species were reconstructed using the unweighted pair group mean arithmetic average (UPGMA). There were mainly three clades in this tree. Reflecting close relationships among genomes of studied species that possess M, D, C, U, A, N and S genomes.*

Keywords: *Bioinformatics, internal transcribed spacer, ribosomal DNA, NEB Cutter, restriction enzymes, genetic relationships, Triticum L, Aegilops L.*

ПРИМЕНЕНИЕ БИОИНФОРМАТИКИ ПРИ ИЗУЧЕНИИ ГЕНЕТИЧЕСКИХ ВЗАИМООТНОШЕНИЙ ПШЕНИЦЫ И ЭГИЛОПСА

Аннотация: *Изучение генетических связей между видами пшеницы и Эгилопса с помощью программы NEB Cutter для определения положений срезов ферментов по последовательностям внутреннего транскрибируемого спейсера (ИТС) из них. Последовательности ее области были включены в программу. Были использованы девять рестрикционных ферментов (AvaI, DrdI, EcoRV, HaeII, Hin1I, SmaI, XmaI, XmaIII, HindIII). Полученные результаты свидетельствуют о высоком уровне полиморфизма в его локусах для Aegilops и Triticum. Филогенетические отношения между видами были реконструированы с использованием невзвешенного парного группового среднего арифметического (UPGMA). На этом дереве было в основном три клады. Отражающие тесные связи между геномами изучаемых видов, обладающих геномами M, D, C, U, A, N и S.*

Ключевые слова: *Биоинформатика, внутренний транскрибируемый спейсер, рибосомная ДНК, NEB Резак, рестрикционные ферменты, генетические отношения, Пшениц л, Эгилоп л.*

1. Introduction:

In 1970 Paulien Hogeweg and Ben Hesper coined Term Bioinformatics as the study of informatic processes in biotic systems. Bioinformatics interacts with computational management and biological information analysis (medical information, proteins, genes, genomes, ecological systems, cells, artificial intelligence and robots). Bioinformatics was defined by the National Center for Biotechnology Information (NCBI 2001), as the field of science in which biology, information technology and computer science merge into a single discipline [11,c.3]. The application of various bioinformatics tools in biological research allows results to be stored, retrieved, analyzed, annotated and visualized, and promotes a better understanding of the biological system in its fullness. Bioinformatics

software tools range from simple command-line tools to more complex graphics programs and standalone web services available from different bioinformatics companies or public institutions [11,c.3/10,c.3688]. The NCBI provides a popular web-based implementation searching for their databases [10,c.3688]. Phylogenetics is defined as the study of genetic relationships among individuals or groups of organisms. Taxonomists use various anatomical methods to find the genetic relationships that take too much time [11,c.3]. Phylogenetic trees are constructed utilizing bioinformatics several methods, based on the sequence alignment. Different algorithmic methods for the construction of phylogenetic trees are developed which are used depending on the different evolutionary lineages [10,c.3688]. Thus the essential purpose of bioinformatics is to integrate large-scale data to understand the molecular mechanism involved in the different development processes.

The *Aegilops* and *Triticum* genera, belonging to the Triticeae tribe within the Pooideae subfamily of the Poaceae grass family, are important in the wheat germplasm due to their evolutionary relationship to the major agricultural crop T. aestivum L. [2,c.108]. *Aegilops* is the closest genus to *Triticum* which includes cultivated wheat, a crop that is of global importance and has a limited pool of genes for modern breeding. *Aegilops* species are a potential future resource for traits for wheat breeding, such as adaptation to different ecological conditions and resistance to pests and diseases. *Aegilops* contains species belonging to the secondary gene pool of wheat, meaning that they have a genome homologous with wheat and that conventional crossing may be used to transfer genes to wheat (*Ae. tauschii* genome D and *Ae. speltoides* genome B) [4,c.37]. Wheat species are economically important and present to large parts of the human population as the staple food. bread wheat *T. aestivum* (BBAADD, 2n = 6x = 42) accounts for 95% of worldwide wheat production. whereas durum wheat *T. durum* (or *T. turgidum* (2n = 4x = 28) symbolizes the rest 5%. Species of the genus *Triticum* L. exist as diploid, tetraploid and hexaploid chromosome numbers with a basic chromosome number n=7 [3,c.270].

Internal Transcribed Spacer (ITS) sequences are located in eukaryotic rDNA genes between the 18S and 5.8S rDNA coding regions (ITS1) and between the 5.8S and 26S

rDNA coding regions (ITS2). These spacer sequences have a high rate of evolution and are present in all known eukaryote nuclear rDNA genes [9,c.699]. Several studies have successfully revealed how useful ITS sequences are in studying plant phylogenetic and genomic relationships at lower taxonomic levels. One of the most remarkable properties of rDNA genes is that it may appear that individual copies evolve more or less in unison. That is, all repeat copies within an array (or genome) may share the same set of mutations together instead of each gene copy acquiring unique sequence variation due to the evolutionary accumulation of mutations. This uniformity arises from one or more processes of homogenization of intergenic sequences collectively referred to as concerted evolution. Thus the ITS region an interesting topic for evolutionary/ phylogenetic investigations [2,c.108]. Restriction enzymes of type II are amongst the most valuable tools available to molecular biology researchers. These enzymes recognize short sequences of DNA (4–8 nucleotides) and cleave at their recognition sites, or near them. Now DNA constructs can quickly be sequenced and tools to locate restriction enzyme sites within these constructs are especially valuable [5,c.1]. In this paper, it described a tool of the program NEB cutter which analyzes rDNA sequences for the ITS region of The *Aegilops* L. and *Triticum* L. genera to study genetic relationships among them using restriction enzyme sites.

2. Experimental

2.1.Plant materials:

The plant material consisted of (23) accessions representing (5) *Triticum* L. species: *T. urartu, T. durum, T. turgidum, T. dicoccon, T. aestivum* and (18) accessions *Aegilops* L. species *Ae. tauschii, Ae. Speltoides, Ae. crassa, Ae. umbellulata, Ae. triuncialis, Ae. comosa, Ae. peregrine, Ae. caudata, Ae. biuncialis, Ae. ovata, Ae. neglecta, Ae. ventricosa, Ae. searsii, Ae. cylindrica, Ae. kotschyi, Ae. longissima, Ae. bicornis, Ae. Sharonensis*.

2.2. Methods and Statistical analysis:

The sequences of Internal Transcribed Spacer (ITS) for these different species were obtained from GenBank for 23 species. Then the obtained sequences were inserted using the NEBcutter V2.0 program (http://nc2.neb.com/NEBcutter2/) which detects restriction enzyme recognition sites, cuts DNA with restriction enzymes and are freely available on the web [10,c.3688]. The program calculates the positions of all restriction enzyme sites in addition noting those that might potentially be blocked by overlapping methylation and finds the open reading frames (ORFs) in the sequence. This algorithm favors longer ORFs to shorter ones. It then displays a schematic diagram of the sequence based on the rules described in the Methods and all restriction enzymes that cut it just once [5,c.1]. Thereafter all enzymes that cut the sequence are shown, all bases that form parts of a restriction enzyme recognition sequence are highlighted. In this study 9 restriction enzymes (AvaI, DrdI, EcoRV, HaeII, Hin1I, SmaI, XmaI, XmaIII, HindIII) were utilized using the NEBcutter program. Then this program was shown the table for each enzyme that indicated the positions cuts of the sequences [10,c.3688]. Consequently, analyzed by the presence(1) /absence (0) of restrictions. Thereafter, the phylogenetic relationships among species were reconstructed using the Unweighted Pair Group Mean Arithmetic average (UPGMA).

3. Results and Discussion:

Plant breeders are interested in wild relatives and wheat progenitors and great efforts have been made to transfer their genetic variation to genotypes that are domesticated. These species have a wide repertoire of key alleles that can be used in programs to improve wheat [2,c.108]. In the present study restriction enzymes showed different levels of variation among species analyzed: Hin1I enzyme produced 36 cuts positions in all species, AvaI enzyme generated 28 cuts positions, EcoRV enzyme created 13 cuts positions and XmaIII, HaeII, DrdI, SmaI, HindIII, XmaI produced 10, 8, 5, 5, 3, 2 cuts positions respectively (Fig. 1).

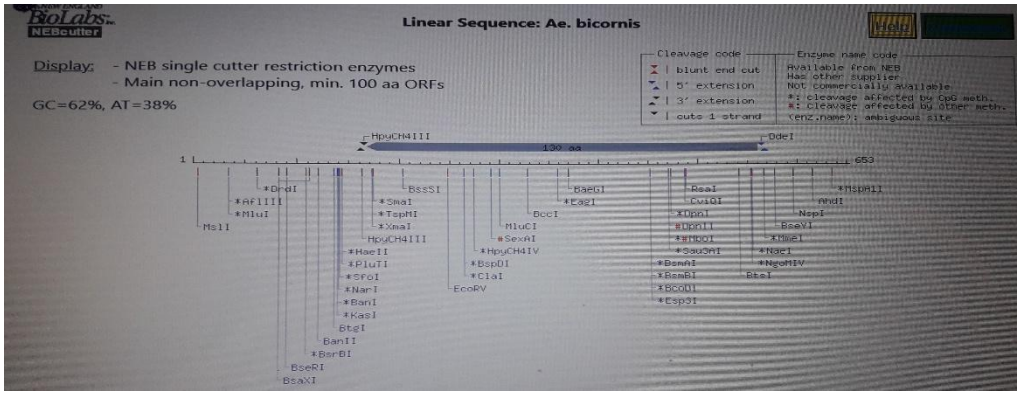

Figure. 1: Positions cuts of some enzymes for *Ae. Bicornis* using the NEBcutter program.

The phylogenetic tree of 23 *Aegilops* L. and *Triticum* L. species generated using (UPGMA) as shown in (Fig. 2). There were mainly three clades in this tree. The first clade included *Ae. searsii*, *Ae. tauschii*, *Ae. peregrine*, *Ae. biuncialis*, *Ae. neglecta*, *Ae. caudata*, *Ae. ventricosa*, *T. aestivum*, *Ae. Bicornis*. The second clade consisted of *Ae. umbellulata*, *Ae. comosa*, *T. dicoccon*, *Ae. Speltoides*,T. *durum*, *Ae. Sharonensis*, *Ae. crassa*, *Ae. cylindrica*, *Ae. kotschyi*, *Ae. triuncialis*, *T. turgidum*. The third clade contained of *T. urartu*, *Ae. ovata*, *Ae. Longissima*. Many researchers studied the relationships among *Aegilops* L. and *Triticum* L. species depending on several methods such as sequence variations of the ITS regions [2,c.108/9,c.699], development and application of the first high-throughput genotyping platform specifically designed for screening wheat relative species [8,c.1993], identifying orthologous chromosomal regions in the U and M genomes relative to wheat (D genome) [6,c.1], the applicability of SSR and ISSR markers in evaluating the genetic relationships [7,c.81] Cleaved Amplified Polymorphic Sequence [3,c.270], characterize the allelic variation of glutenin and gliadin [1,c.510].

The studied species are belonging to five sections of *Aegilops*. Clade A and C constitute sections (*Aegilops*, *Siptosis*, *Cylindropyrum* and *Vertebrata*) respectively. While clade B constitutes five sections (*Aegilops*, *Comopyrum*, *Siptosis*, *Vertebrata* and *Cylindropyrum*). In the current study, the level of polymorphism in ITS loci was higher than those reported previously in the same loci using different restriction enzymes for

Aegilops and *Triticum* species [2,c.108/4,c.37]. While my results agreement with [9,c.699]. Regarding The clade A the position of *Ae. peregrine, Ae. biuncialis, Ae. neglecta* in tree confirms its close relationship between them that belong to the same section. Like was *Ae. searsii, Ae. bicornis* and *Ae. tauschii, Ae. ventricosa* which gathered into *Siptosis* and *Vertebrata* sections respectively. [13,c.14570] reported that *Ae. tauschii* is the maternal parent of three tetraploids, *Ae. cylindrica*, *Ae. crassa* and *Ae. ventricosa* [12,c.828] and *Ae. tauschii* could serve as the donor of the D genome for wheat [9,c.699]. Based on these conclusions the current results indicated a close relationship between (*Ae. cylindrica, Ae. crassa*) observed which have D genome. In addition to located the two genotypes *Ae. tauschii, T. aestivum* in the same clade A. Similarity *Ae. triuncialis* (genome UC) and *Ae. kotschyi* (genome US) grouped to gather with *Ae. umbellulata* (genome U) in The clade B which considered a parent to them [13,c.14570]. likewise *T. dicoccon, Ae. speltoides, T. durum, Ae. sharonensis* and *T. turgidum* grouped in the same clade. These species Shared a B genome. The origin of the B genome is still a matter of discussion. Recent studies showed that the species of section *Siptosis* were the main contributor of the B genome to polyploid wheats. In addition close relationships between *Ae. umbellulata* and *Ae. comosa* (genome M). This agreement with [13,c.14570] observed the plasmon of *Ae. umbellulata* is closely related to the plasmons of *Ae. comosa*. The tree showed close relationships among studied species that possess M, D, C, U, A, N and S genomes and was in agreement with the results of earlier studies. [6,c.1] studied the Syntenic relationships between the U and M Genomes of *Aegilops* and wheat using conserved orthologous set (COS) markers for the U and M genome chromosomes. Their findings suggest that the genome M is closer to the genome D of wheat than the genome U.

[1,c.510] indicated that the di- and tetraploid wild relatives possessing U, C, D and A genomes harbor wide variation in the glutenin and gliadin subunits indicating large genetic diversity within and between species.

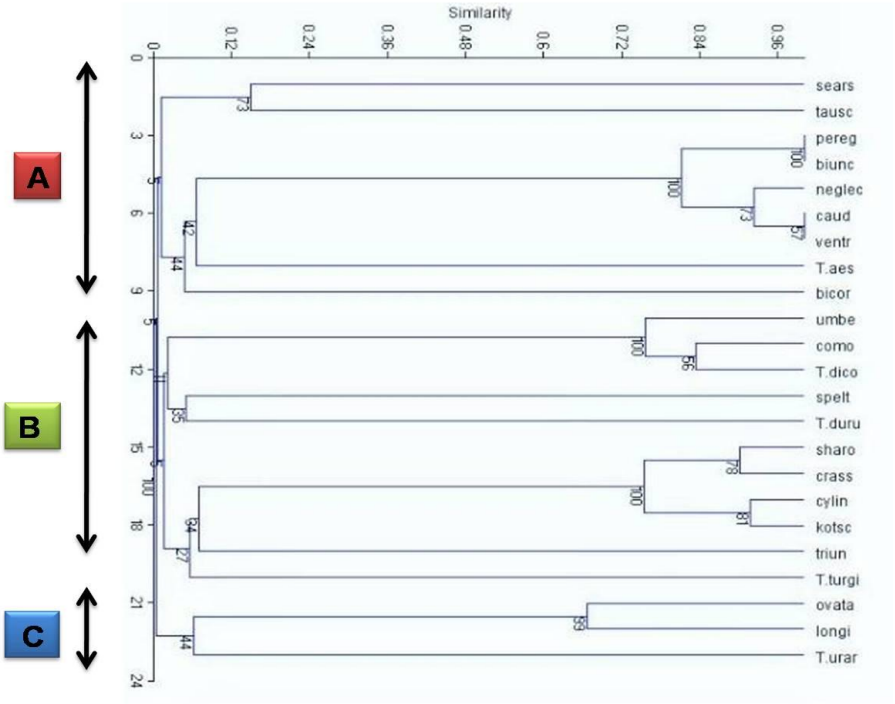

Figure. 2: Phenogram of UPGMA cluster analysis in *Aegilops* L. and *Triticum* L. species.

Consequently, it is suggested that the discovery of this highly diverse gene pool should encourage researchers to explore valuable and new alleles for the improvement of new varieties that are adapted to new uses.

In conclusion, restriction enzymes are a vital tool for analyzing DNA. This study showed that bioinformatics data can be modeled and used computationally for sequence analysis and classification. Thus this approach can be considered as an alternative method compared to restriction fragment length polymorphism (RFLP) methods.

References:

1. Ahmadi. J, Pour-aboughadareh. A, FAbriki-ourang. S and Mehrabi A.A. Molecular detection of glutenin and gliadin genes in the domesticated and wild relatives of wheat using allele-specific markers // Cereal Research Communications. – 2018. – N(46). – C. 510 - 520.

2. Alnaddaf. L.M, Moualla. M.Y. and Haider. N. Genetic relationships among *Aegilops* L. and *Triticum* L. species based on the internal transcribed spacer sequences of nrDNA (ITS) // Asian Journal of Agricultural Sciences. – 2013. –N(5). – C. 108 - 117.

3. Alnaddaf. L. M, Moualla. M. Y. and Haider. N. Resolving Genetic Relationships among Aegilops L. and Triticum L. species using analysis of Chloroplast DNA by Cleaved Amplified Polymorphic Sequence (CAPS) //Asian Journal of Agricultural Sciences.- 2012. – N (4). – C. 270 -279.

4. Alnaddaf. L.M, Moualla. M.Y. Using of restriction enzymes in analysis of the genetic variability for some *Aegilops* L. Species in the (ITS) genomic region // Alley-science. ur. – 2019. –N(10). – C. 37 – 47.

5. Maurya. G. K. Restriction enzyme // Encyclopedia of Animal Cognition and Behavior. – 2019. –N(1254). – C. 1 – 4.

6. Molna´r. I, Sˇimkova. H, Leverington-Waite. M, Goram. R, Cseh. A, et al. Syntenic relationships between the U and M genomes of *Aegilops*, wheat and the model species brachypodium and rice as revealed by COS markers // Plos One. – 2013. –N(8). – C. 1 – 12.

7. Moradkhani. H, Mehrabi. A. A, Etminan. A, Pour-Aboughadareh. A. Molecular diversity and phylogeny of *Triticum-Aegilops* species possessing D genome revealed by SSR and ISSR markers // Plant Breeding And Seed Science. – 2015. –N(71). – C. 81 – 95.

8. Przewieslik-Allen. A. M, Burridge. A. J, Wilkinson. P. A, Winfield. M. O, Shaw. D. S, McAusland. L, King. J, King. I. P, Edwards. K. J and Barker. G. L. A. Developing a High-Throughput SNP-Based Marker System to Facilitate the Introgression of Traits From *Aegilops* Species Into Bread Wheat (*Triticum aestivum*) // Frontiers in Plant Science. – 2019. –N(9). – C. 1993 – 2005.

9. Safari. Z and Mehrabi A. A. Molecular phylogeny of *Aegilops* L. and *Triticum* L. species revealed by internal transcribed spacers of ribosomal genes // Journal of Agricultural *Science* and *Technology*. – 2019. –N(21). – C. 699 - 714.

10. Vincze. T, Posfai. J and Roberts. R. J. NEBcutter: a program to cleave DNA with restriction enzymes // Nucleic Acids Research. – 2003. –N(31). – C. 3688 – 3691.

11. Wani. M. Y, Ganie. N. A, Rani. S, Mehraj. S, Mir. M. R, Baqual. M. F, Sahaf. K. A, Malik. F. A and Dar. K. A. Advances and applications of bioinformatics in various fields of life // International Journal of Fauna and Biological Studies. – 2018. –N(5). – C. 3 – 10.

12. Wan. Y, D. Wang. P.R. Shewry and N.G. Halford. Isolation and characterization of five novel high molecular weight subunit genes from *Triticum timopheevi* and *Aegilops cylindrica*. // TAG. – 2002. –N(104). – C. 828 – 839.

13. Wang. G. Z, N. T. Miyashita and K. Tsunewaki. Plasmon analysis of Triticum (wheat) and Aegilops: PCR–single-strand conformational polymorphism (PCR-SSCP) analyses of organellar DNAs. // Proc. Natl. Acad. Sci. USA. – 1997. –N(94). – C. 14570 – 14577.