

A fuzzy weighted least squares approach to construct phylogenetic network among subfamilies of grass species

Rinku Mathur¹ and Neeru Adlakha²

Abstract

Phylogenetic networks are considered as the structures that are used to understand the evolutionary pathways among the different organisms. Evolutionary relations are due to the preservice of mutations in sequences, occurred due to non-tree like events like horizontal gene transfer, Homoplasy, sexual hybridization and recombination, etc. The effective and efficient reconstruction of the networks for these events is an challenging task in computational biology. In this article, a Fuzzy Weighted Least Squares (FWLS) approach is developed and employed to detect these events in commonly known species of grasses. The results obtained by the proposed method predicts the possibility of hybridization or recombination among the inter cluster species i.e., *Oryza* and *Triticum* and intra cluster species i.e. *Bentgrass* and *Brachypodium*. Results also provide the optimized values of Q

¹ Department of Applied Mathematics and Humanities, S. V. National Institute of Technology, Surat, Gujarat, India, e-mail: rinkumathur56@gmail.com

² Department of Applied Mathematics and Humanities, S. V. National Institute of Technology, Surat, Gujarat, India, e-mail: neeru.adlakha21@gmail.com

in comparison to the other available least squares method and thus error level is also minimized.

Mathematics Subject Classification: 92Bxx, 92B05

Keywords: Phylogenetic Networks, Fuzzy Set, Subfamilies, Grass Species, Hybridization, Reticulation distance

1 Introduction

Evolution of human civilization has depended mainly on agriculture. It has not only given man the freedom from the risks of hunting and collecting food from forest but has also given assurance of continuous availability of food. Agriculture plays an important role in providing basic necessities of mankind in three ways – food, clothes and shelter. In response to food, genomic efforts have largely focused on three important subfamilies of grasses, the Ehrhartoideae (rice or oryza), the Panicoideae (maize, sorghum and sugarcane) and the Pooideae (wheat, barley, bentgrass and brachypodium) [1] that provides the basis of human nutrition, and also dominates many economical, ecological and agricultural systems.

In previous decades, the evolution of grass species has long been assumed to be a branching process that could be represented only by a tree structure. In such a structure, species can solely be linked to their closest ancestor and direct interspecies relationships (connection branches) are not allowed. Such well – known evolutionary mechanisms like horizontal gene transfer and hybridization or allopolyploidy cannot, however, be appropriately represented by means of tree topology or structure [2-6]. Reticulate patterns of relationships have been found in number of phylogenetic situations like bacterial evolution in which lateral gene transfer is the mechanism allowing bacteria to ex-change genes across species [7]. In case of plant evolution, allopolyploidy or hybridization leads to the

appearance of new species by combining the chromosome complements of the two parent species [3,5,8] and an hybridization event is shown in Figure 1 [9].

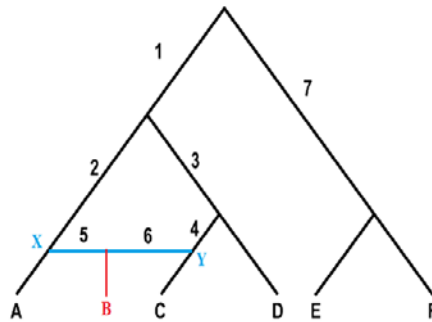


Figure 1: A Phylogenetic network with a single hybrid species B [9].

Biologists are thus confronted while dealing with such type of complex events occurred during evolution of species. They have, therefore, been looking for mathematical algorithms or techniques for deriving trees from character data, and then networks. Therefore, mathematicians and mathematically inclined biologists have, over the past 15 years, been developing algorithms for phylogenetic networks, which are the most obvious extensions of a tree structure for evolution. These networks allow reticulations among the branches on the diagram, rather than a tree – like or bifurcating structure [10].

In this paper, a Fuzzy Weighted Least Squares (FWLS) Approach has been developed to find out the reticulate evolutionary relationships or to construct the Phylogenetic network among the eight important crop plants of grasses by means of addition of new branches (reticulations) to the tree. The proposed method gives the more optimized and improved results than the least squares approach [2] and also suggests that the new hybridized species can produce the higher yield and have more strength to exist in adverse atmospheric conditions.

2 Basic Concepts and Definitions

2.1 Dissimilarities and trees

Let X be a set with n elements. A dissimilarity or distance matrix of order $n \times n$ on X is a non-negative real function d on $X \times X$ satisfying the following two conditions [11, 12]:

- (i) for all $x, y \in X$, $d(x, y) = d(y, x)$;
- (ii) for all $x, y \in X$, $d(x, y) \geq d(x, x) = 0$.

The dissimilarity d is a metric if it satisfies the classical metric triangular inequality:

$$\text{for all } x, y, z \in X, \quad d(x, z) \leq d(x, y) + d(y, z).$$

2.2 Graph

A graph G is a combination of vertices and edges and is denoted as (V, E) , where V is the node (or vertex) set, each node is either a taxon belonging to a set X or an intermediate node belonging to $V - X$ and E is a set of unordered pairs of distinct elements of V . The degree $\partial(v)$ of a vertex v is the number of edges $e \in E$ such that $v \in e$. A leaf is a vertex of degree one [12].

2.3 Tree

Any connected graph with a unique path between any two distinct vertices u and v is called a tree and is denoted as $T(uv)$. A tree T has exactly $|V| - 1$ edges. Related to the set X , the X -trees are associated by two properties:

- (i) the set of leaves of T is X ;

- (ii) for any $v \in V - X$, $\partial(v) \geq 3$.

An X tree with n leaves has at most $n-2$ internal vertices, $2n-2$ vertices and thus $2n-3$ edges and any given Phylogenetic tree representing the evolutionary history of taxon or organisms can be transformed into a binary tree by adding links of length zero wherever it is necessary [8, 12].

A dissimilarity d on X is said to be ultra-metric if it satisfies the stronger inequality:

$$\text{for all } x, y, z \in X, d(x, z) \leq \max(d(x, y), d(y, z)).$$

A dissimilarity d on X is said to be a tree metric if it satisfies the four-point condition, i.e.,

$$\text{for all } x, y, z, w \in X,$$

$$d(x, y) + d(z, w) \leq \max\{d(x, z) + d(y, w), d(x, w) + d(y, z)\}.$$

According to Buneman [13], the two conditions are equivalent for a dissimilarity matrix d on X .

- (i) d is a tree metric.
(ii) d satisfies the four-point condition.

Moreover, a tree metric admits a unique tree representation [11].

2.4 Phylogenetic network

A weighted graph defined as a triplet (V, E, l) , where l is a function of edge lengths assigning real nonnegative numbers to the edges is called a Phylogenetic or reticulated network denoted by R .

A Phylogenetic network is connected if for every pair of nodes u and v , there exists at least one path from u to v . It is called undirected if there is no direction associated with the branches. Given a connected and undirected phylogenetic network R , the minimum-path-length distance between nodes u and v in any weighted graph is defined as:

$$d(u, v) = \min\{l_p(u, v) : p \text{ is a path from } u \text{ to } v\}.$$

A set of reticulation distances can be associated with the set of pairwise distances among the taxa in X . They are the minimum-path-length distances among taxa whose relationships are represented by a phylogenetic or reticulated network [8,12].

2.5 Some properties of a reticulation distance

A reticulation distance is no longer a tree distance. Figure 2 depicts that the four-point condition characterizing phylogenetic trees does not satisfy a reticulation distance. This condition holds for the tree distance τ associated with tree T (Figure 2 left) but not for the distance γ associated with the network $T + xz$ (Figure 2 right) [14]:

$$\begin{array}{ll} \tau(x, y) + \tau(z, w) = 4; & \gamma(x, y) + \gamma(z, w) = 4; \\ \tau(x, z) + \tau(y, w) = 6; & \gamma(x, z) + \gamma(y, w) = 3.5; \\ \tau(x, w) + \tau(y, z) = 6; & \gamma(x, w) + \gamma(y, z) = 5. \end{array}$$

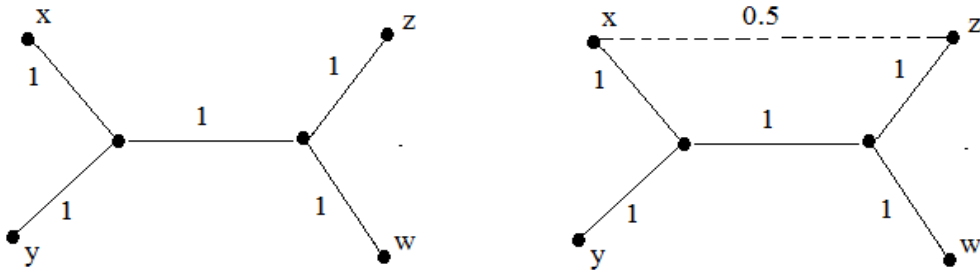


Figure 2: A phylogenetic tree T and phylogenetic network $T + xz$ [14].

2.6 Fuzzy set theory

The term fuzzy was first proposed by Zadeh in 1965 [15, 16], when he

published the famous paper on the Fuzzy Sets which is the extension of classical set theory. The fuzzy set theory was developed to improve the oversimplified model, thereby developing a more robust and flexible model in order to solve the real world complex problems involving human aspects. In this approach, an element can belong to a set to a degree k ($0 \leq k \leq 1$) in contrast to the classical set theory where an element must definitely belong or not belong to a set. For instance, one can be definitely good or bad in the classical set theory whereas in the fuzzy set theory, we can say that someone is 60 percent good or bad, i.e. with the degree of 0.6 [16]. The characteristic function thus allows various degrees of membership for the elements of a given set. The fuzzy set, thus determines the vagueness in describing the planning goals and other uncertainties involved in the desired values. In the last two decades, the fuzzy set theory has received a wide attention in the field of environmental planning, management and also in bioinformatics with the great pace [16].

2.7 Fuzzy set and fuzzy relation set

If X is a collection of n objects generally denoted by x , then a Fuzzy set A in X is a set of ordered pairs [17].

$$A = \{ (x, \mu_A(x)) / x \in X \}$$

where $\mu_A(x)$ is the value of the membership function or grade of membership of x in A .

Also $A = \{ ((x, y), \mu_A(x, y)) / x, y \in X \}$ is called the fuzzy relation set.

2.8 Fuzzy membership function

A function defined on the set X to the membership space ranges from the 0

to1 is called a fuzzy membership function and is denoted as $\mu_A(x)$. i.e.,

$$\mu_A(x): X \rightarrow [0,1]$$

Also $\mu_A(x, y): X \times X \rightarrow [0,1]$ is called fuzzy relation membership function [17].

3 Materials and Methods

3.1 Data material

The accession numbers representing the important crop plants of grass species used in this study are collected from the BMC research note of Esteban Bortiri, et al. [18] and our previous paper [2] in which the DNA sequences are retrieved from National Centre for Biotechnology Information (NCBI) site [19]. The detailed information of the species including accession numbers is given in Appendix.

3.2 Data analysis

The distance (dissimilarity) matrix of the DNA sequences among the eight crop plants of grasses was availed from the literature of our previous work [2, 20] given in Table 1. In addition, the tree distance matrix given in Table 2 and the corresponding Phylogenetic tree is shown in Figure 3. The phylogenetic tree considered as the cornerstone for the detection of phylogenetic network, was constructed by the most applicable Neighbor Joining Method [21]. On the basis of the distance matrix, a fuzzy relational membership matrix among these plants is generated in this work by defining the function:

$\mu_f : X \times X \rightarrow [0,1]$ which is defined as:

$$\mu_f(i, j) = \frac{1}{1 + d(i, j)}$$

Table 3 shows the fuzzy relational matrix W obtained by above function among the plant species.

At last, the phylogenetic tree matrix and fuzzy matrix W is analysed analytically for the Fuzzy Weighted Least Squares Criterion to get the desired results for the construction of phylogenetic network.

3.3 Fuzzy weighted least squares approach to infer phylogenetic network

In this paper, the Fuzzy Weighted Least Squares (FWLS) Method is developed to construct the evolutionary network of the concerned problem of the eight crop plants of grass species. This method is the extension of the Least Squares Method that is explored in our previous work given by Makarenkov [2, 8]. The distance matrix (Table 1) obtained from the DNA sequences is used to construct the binary Phylogenetic tree T which is considered as the fundamental structure for the reticulated network to be reconstructed among the plant species. The original tree may be rooted or not; this does not matter when constructing undirected reticulated networks.

To add a new branch to a Phylogenetic tree, all the possible pairs of nodes are operated out that are not already linked by a branch and select the one that reduces the value of the fuzzy weighted least-squares criterion the most. Let us consider a binary Phylogenetic tree T deduced from a distance function d and a pair of nodes x and y in T not linked by a branch (Figure 4a). Now we look for an optimum value l , according to the fuzzy weighted least-squares loss function, for a potential new branch xy which may be added to the tree T , while keeping fixed the lengths of all pre - existing tree branches (Figure 4b).

To find out the optimum value of the length of the first reticulation branch, firstly, a set $A(xy)$ interpreting the distances between pairs of taxa that are

susceptible of changing if a new reticulation branch $(x y)$ is added. Let τ be a function of the distances in T between pairs of nodes. The set $A(xy)$ includes all pairs of taxa ij of X such that

$$\text{Min}\{\tau(i, x) + \tau(j, y) ; \tau(j, x) + \tau(i, y)\} < \tau(i, j). \quad (1)$$

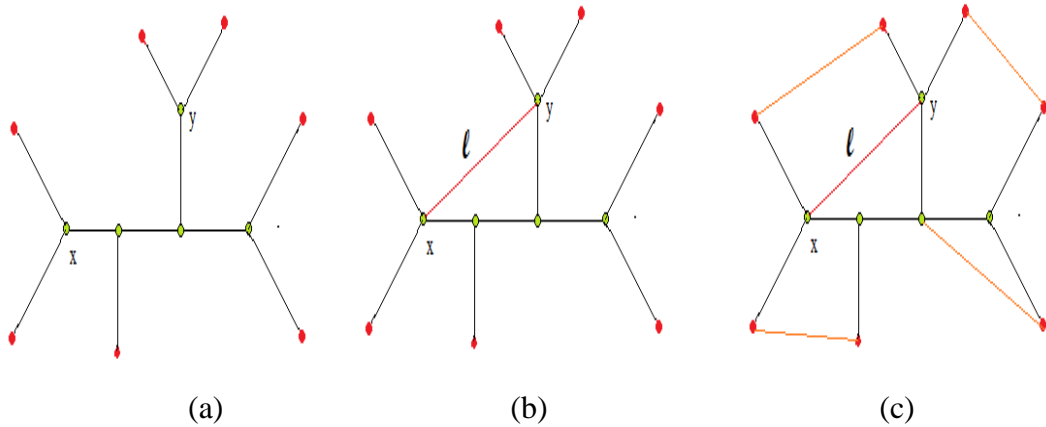


Figure 4: Description of method for deducing phylogenetic networks. (a) A binary phylogenetic tree T is considered. (b) New branch of length l can be added to T to link nodes x and y . (c) Reticulate network deduced from T by addition of reticulation branches [8].

To determine the optimum value l of a new potential branch (xy) , we have to subdivide $A(xy)$ into the m following subsets [2, 8]:

$$\begin{aligned} A_1 &= \{ij\} \text{ such that: } \tau(i, j) - \text{Min}\{\tau(i, x) + \tau(j, y) ; \tau(j, x) + \tau(i, y)\} \\ &= \text{Min}_{\{ij \in A(xy)\}} \{\tau(i, j) - \text{Min}\{\tau(i, x) + \tau(j, y) ; \tau(j, x) + \tau(i, y)\}\} = l_1 ; \\ A_k &= \{ij\} \text{ such that: } \tau(i, j) - \text{Min}\{\tau(i, x) + \tau(j, y) ; \tau(j, x) + \tau(i, y)\} \\ &= l_k > l_{k-1} \text{ (for } k = 2, \dots, m-1), \\ A_m &= \{ij\} \text{ such that } \tau(i, j) - \text{Min}\{\tau(i, x) + \tau(j, y) ; \tau(j, x) + \tau(i, y)\} \\ &= \text{Max}_{\{ij \in A(xy)\}} \{\tau(i, j) - \text{Min}\{\tau(i, x) + \tau(j, y) ; \tau(j, x) + \tau(i, y)\}\} = l_m = \tau(x, y) > l_{m-1} , \end{aligned}$$

where $A(xy) = \{A_1 \cup A_2 \cup \dots \cup A_m\}$ and m is the number of subsets of distinct values that the quantity $\tau(i, j) - \{\tau(i, x) + \tau(j, y); \tau(j, x) + \tau(i, y)\}$ can take over the set $A(xy)$.

Each subset A_k is associated with an interval of possible length values l of the branch xy for which a particular optimization problem, i.e., a quadratic function has to be minimized, subject to a corresponding interval of length values of xy .

Suppose that $l_k \leq l \leq l_{k+1}$, where $k=0, \dots, m-1$. The constraint means that only the distances, i.e., the minimum-path-lengths $\tau(i, j)$, that are such that $ij \in \{A_m \cup A_{m-1} \cup \dots \cup A_{k+1}\}$ will change lengths.

The problem formulated to compute the optimum length value l of a potential new branch xy on the fixed interval $l_k \leq l \leq l_{k+1}$ is as [2]:

$$Q^*(xy, k) = \sum_{p=k+1}^m \sum_{ij \in A_p} \mu_f(i, j) (\text{Min}\{\tau(i, x) + \tau(j, y); \tau(j, x) + \tau(i, y)\} + l - d(i, j))^2 \rightarrow \text{min.} \quad (2)$$

where $\mu_f(i, j)$ is the fuzzy membership function showing the degree of similarity (or dissimilarity) among the different taxa.

Minimizing of $Q^*(xy, k)$ means to minimize the fuzzy weighted quadratic sum of differences between the values of the given evolutionary distance d and the associated reticulation estimates. A non trivial solution $l^*(xy, k)$ to this problem is the following:

$$l^*(xy, k) = \frac{\sum_{p=k+1}^m \sum_{ij \in A_p} \mu_f(i, j) (d(i, j) - \text{Min}\{\tau(i, x) + \tau(j, y); \tau(j, x) + \tau(i, y)\})}{\sum_{p=k+1}^m \sum_{ij \in A_p} \mu_f(i, j)} \quad (3)$$

If this quantity does not meet the constraint, the optimal solution $l^*(xy, k)$ has to be selected from the boundary values l_k and l_{k+1} . Now, the only task is to calculate the value of the desired fuzzy weighted least-squares objective

function corresponding to this particular solution on the interval $l_k \leq l \leq l_{k+1}$ of length values of xy .

$$Q = \sum_{i \in X} \sum_{j \in X} \mu_f(i, j) [\tau(i, j) - d(i, j)]^2 \rightarrow \min \quad (4)$$

These computations are repeated over all intervals of branch lengths established for the given pair of nodes xy not linked by a branch. The global optimum value of Q as well as l over the set of defined intervals $l_k \leq l \leq l_{k+1}$, for $k = 0, \dots, m-1$, are obtained recursively. To incur the optimum value of Q over the set of all possible new branches, these computations should be repeated for all pairs of tree nodes that are not linked by a branch. Once the first new branch has been added to the evolutionary network, the best second, third, and the following reticulation branches may be placed into it in the similar fashion (Figure 4c).

3.4 Criteria to add reticulation branches in the network

The three possible goodness-of-fit criteria are used in this article to determine when to stop adding new branches to an evolutionary network. The total number of vertices in an unrooted binary phylogenetic tree with n leaves is $2n-2$.

Hence, the maximum number of branches to be added in a reticulated network, deduced from a binary phylogenetic tree with n leaves is $(2n-2)(2n-3)/2$. Nevertheless, any metric distance can be interpreted by a complete graph with $n(n-1)/2$ branches. Thus, any of the two limits $(2n-2)(2n-3)/2$ or $n(n-1)/2$ can be considered as the maximum possible number of branches in a reticulated network. If the latter limit is considered, the number of degrees of freedom of a reticulated network with N branches is

$n(n-1)/2 - N$ [2, 8].

Hence, the first goodness-of-fit function that we consider is the following:

$$Q_1 = \frac{\sqrt{\sum_{i \in X} \sum_{j \in X} \mu_f(i, j) [\tau(i, j) - d(i, j)]^2}}{n(n-1)/2 - N} = \frac{\sqrt{Q}}{n(n-1)/2 - N} \quad (5)$$

The minimum value of Q_1 defines a stopping rule for addition of new branches to the reticulate phylogeny. Now, a slightly modified criterion, denoted by Q_2 , which usually adds more reticulation branches to the network than criterion Q_1 is [2]:

$$Q_2 = \frac{\sum_{i \in X} \sum_{j \in X} \mu_f(i, j) (\tau(i, j) - d(i, j))^2}{n(n-1)/2 - N} = \frac{Q}{n(n-1)/2 - N} \quad (6)$$

Third one is the Akaike information criterion (AIC) which is a useful and selects the model that minimizes the following quantity [8]:

$$AIC = \frac{Q}{(2n-2)(2n-3)/2 - 2N} \quad (7)$$

4 Results and Discussion

It was observed in our previous work that while constructing the Phylogenetic tree from the distance matrix (Table 1) brings some changes in original evolutionary distances of the grass species. Out of total 28 pairs of inter species combinations, the distances among thirteen pairs of species are increased from their original distances representing that they are going away from their common ancestry. Other thirteen pairs show decrease in their distances representing, they are coming closer to their ancestry and other two pairs remains unchanged [2].

Table 1: Genetic distance matrix among the eight plants of grass species based on K2P distance formula [20]

	Oryza	Brachypodium	Bentgrass	Hordeum	Zea	Saccharum	Sorghum	Triticum
Oryza	0.000000	0.062395	0.065619	0.065354	0.056720	0.054645	0.055303	0.057951
Brachypodium		0.000000	0.041923	0.043900	0.067502	0.065314	0.066117	0.047344
Bentgrass			0.000000	0.037143	0.066557	0.064195	0.064361	0.039504
Hordeum				0.000000	0.066442	0.064265	0.064480	0.017205
Zea					0.000000	0.007709	0.008981	0.067182
Saccharum						0.000000	0.004000	0.065045
Sorghum							0.000000	0.065504
Triticum								0.000000

Figure 3 shows that the Phylogenetic tree of eight crop plants of grass species exhibits three main clusters; first one belongs to Pooideae subfamily that is composed of Brachypodium, Triticum, Hordeum and Bentgrass species ; the Zea, Sorghum and Saccharum species belongs to the second cluster of Panicoideae subfamily and third one is the Ehrhartoideae subfamily consists of only rice species .

The results obtained after analytical calculations for the construction of the phylogenetic network shows the branch lengths between all the possible pairs of taxa that are not directly connected to each other.

It was observed that the value of least square criterion (Q) obtained after fitting the phylogenetic tree (Table 2) to the evolutionary distance matrix by using the NJ method of T-Rex [22] Package was 0.00005889. The value of Q was then minimized to 0.00005575 by Fuzzy Weighted Least Squares criterion over the same phylogenetic tree.

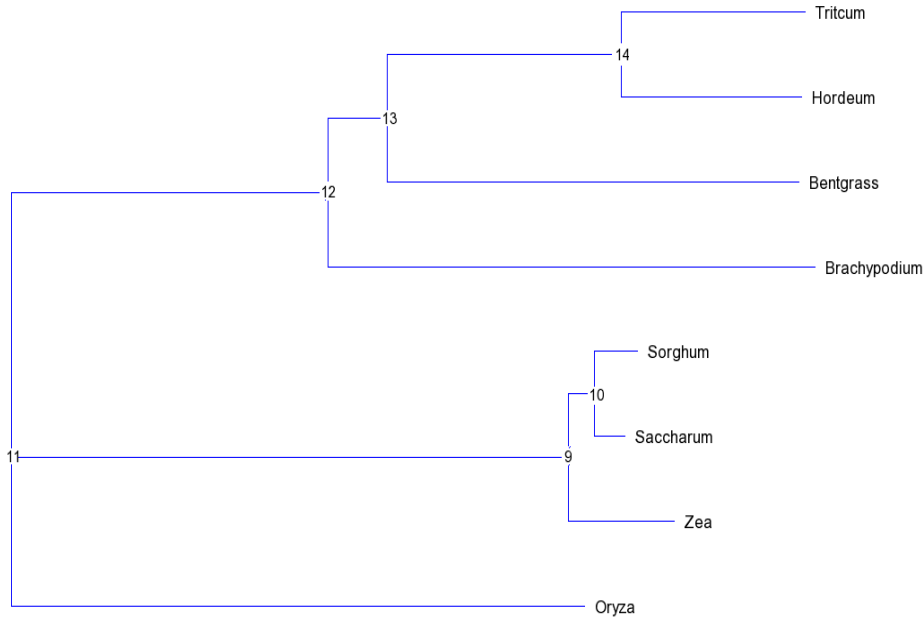


Figure 3: Phylogenetic tree of eight crop plants of grass species obtained by NJ method [21].

After fitting the reticulation branches to the Phylogenetic tree by using the new technique proposed in this article, the value of Q reduces from 0.00005575 to 0.00002501 and the corresponding reticulation distance matrix among the eight plants of grass species is shown in Table 4.

While constructing the phylogenetic network (Figure 5), the first reticulated branch between Oryza and Triticum is added to the tree that decreases the value of Q from 0.00005575 to 0.00003321 and the second branch between Bentgrass and Brachypodium is added that decreased Q from 0.00003321 to 0.00002803.

The rules used to stop the addition of reticulation branches to the tree suggested the different solutions while building the evolutionary network. The number of branches to be added to the tree according to the three criteria is shown in Figure 6.

Table 2: Phylogenetic tree metric distances by NJ method

	oryza	Brachypodium	Bentgrass	Hordeum	Zea	Saccharum	Sorghum	Triticum
oryza	0.000000	0.063269	0.062540	0.062676	0.056861	0.054606	0.055201	0.062834
Brachypodium		0.000000	0.044245	0.044382	0.067325	0.065069	0.065665	0.044540
Bentgrass			0.000000	0.038245	0.066595	0.064340	0.064935	0.038402
Hordeum				0.000000	0.066732	0.064477	0.065072	0.017205
Zea					0.000000	0.008047	0.008643	0.066890
Saccharum						0.000000	0.004000	0.064634
Sorghum							0.000000	0.065230
Triticum								0.000000

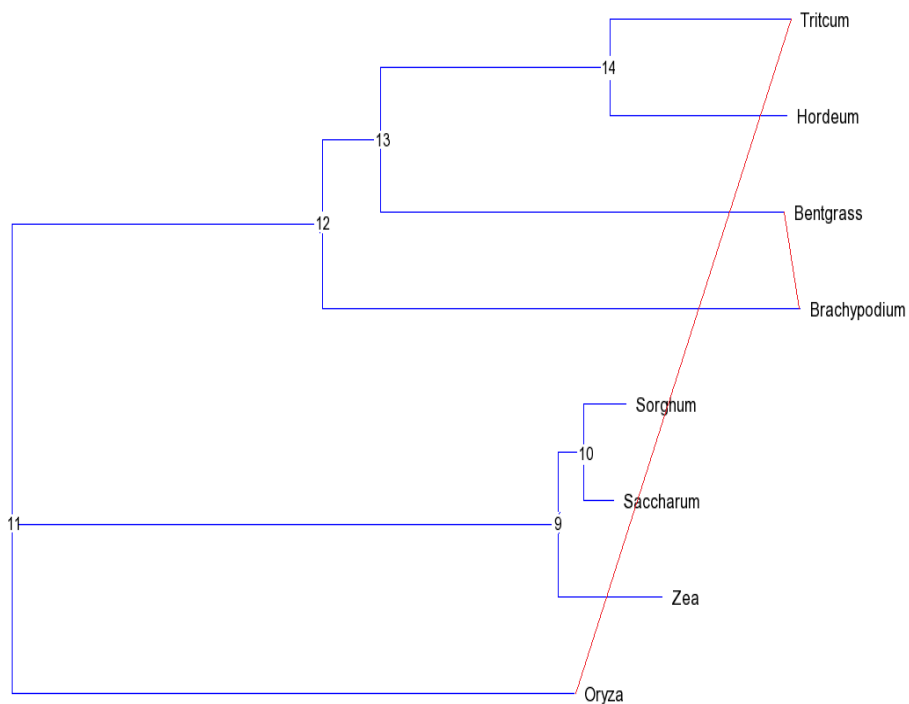


Figure 5: Phylogenetic network among crop plants of grass species obtained by FWLS method

Figure 6 depicts that the values of FWLSC decreases continuously with the addition of new branches and the minimum values of Q_1 and Q_2 was observed at the second reticulation and hence only two new branches can be added according to first two criterions. For the AIC criterion, the minimum value is at the fourth reticulation was founded.

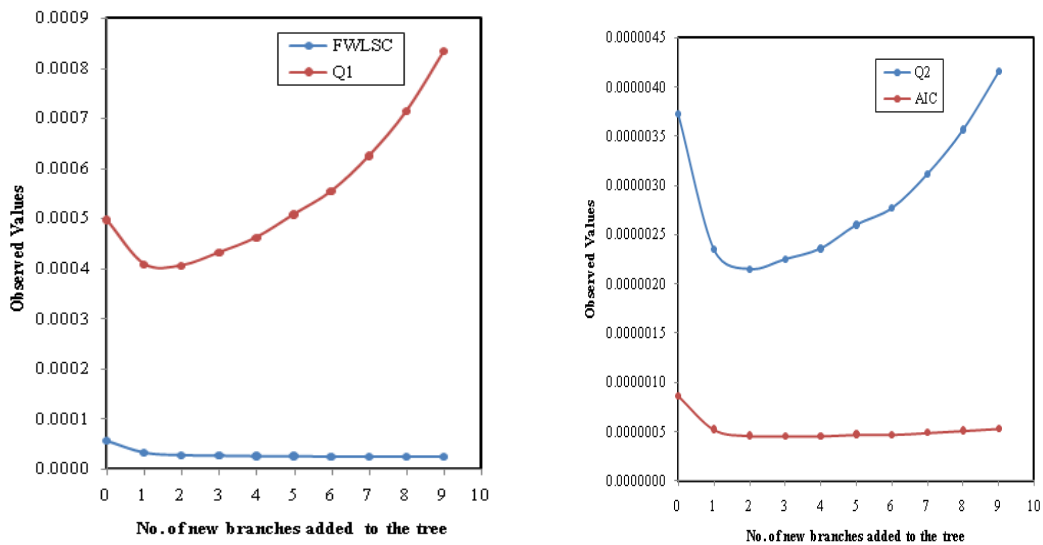


Figure 6: Nature of different rules to add the number of reticulation branches to the Phylogenetic tree

Finally, the results obtained in this article optimizes the results of our previous work and are in full agreement with our results and the experimental results obtained by Esteban Bortiri, et al. [18] using the shotgun sequencing technique and it is depicted that pairs (Oryza, Triticum) and (Bentgrass, Brachypodium) are closely related i.e. reticulation is possible between them. Figure 7 depicts the comparison between our previous results obtained by LSC of T-Rex package [22] and this new technique (FWLSC) which shows that the value of Q is minimized continuously with the addition of new branches in the tree to make the phylogenetic network.

Table 3: Fuzzy relational membership matrix W among plants of grass species.

	Oryza	Brachypodium	Bentgrass	Hordeum	Zea	Saccharum	Sorghum	Triticum
Oryza	1.000000	0.941269	0.938421	0.938655	0.946324	0.948186	0.947595	0.945223
Brachypodium		1.000000	0.959763	0.957946	0.936766	0.938690	0.937983	0.954796
Bentgrass			1.000000	0.964187	0.937596	0.939677	0.939530	0.961997
Hordeum				1.000000	0.937697	0.939615	0.939425	0.983086
Zea					1.000000	0.992349	0.991098	0.937047
Saccharum						1.000000	0.996015	0.938927
Sorghum							1.000000	0.938522
Triticum								1.000000

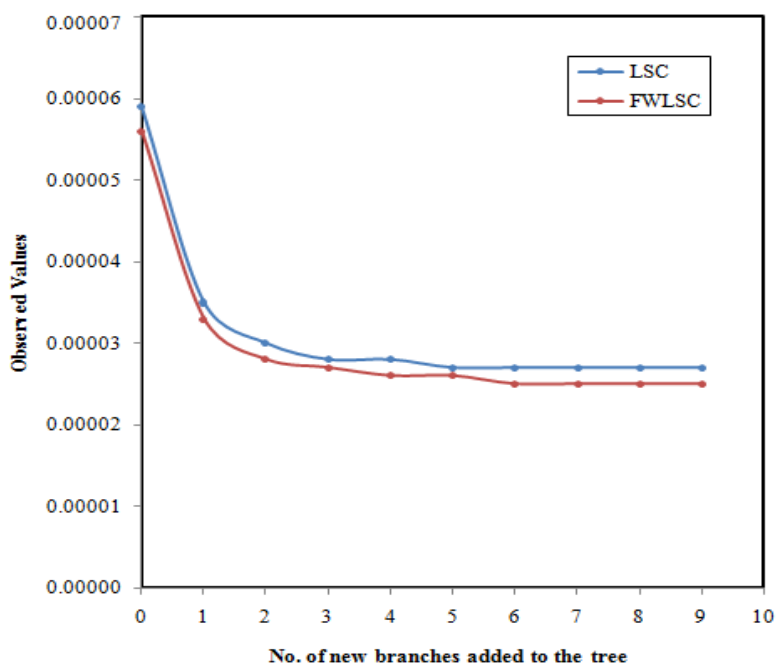


Figure 7: Graph showing the difference between the values at different reticulations by LSC and FWLSC

Table 4: Reticulation distance Matrix among plants of grass species by FWLS method

	Oryza	Brachypodium	Bentgrass	Hordeum	Zea	Saccharum	Sorghum	Triticum
Oryza	0.000000	0.062395	0.062540	0.062676	0.056720	0.054606	0.055201	0.057951
Brachypodium		0.000000	0.041923	0.043900	0.067325	0.065069	0.065665	0.044540
Bentgrass			0.000000	0.037143	0.066557	0.064195	0.064361	0.038402
Hordeum				0.000000	0.066442	0.064265	0.064480	0.017205
Zea					0.000000	0.007709	0.008643	0.066890
Saccharum						0.000000	0.004000	0.064634
Sorghum							0.000000	0.065230
Triticum								0.000000

5 Conclusion

A phylogenetic network among the eight subfamilies of grass species is predicted. It is also concluded that inter – species and intra – species pair of plants can be hybridized to produce the new crop plant. Inter – species pair contains Oryza and Triticum and the intra – species pair contains Bentgrass and Brachypodium. The error level and the value of Q is also optimized to 0.00005575 by using this method of FWLS. The results obtained by this method are compared graphically with the least square method in the Figure 7. The new species produced can have more strength to cope up with adverse weather conditions and yields more grains than the previous ones.

Thus, the new technique proposed in this article could be of great use and helpful to the community of agricultural and evolutionary biology researchers to generate their information mathematically and computationally for developing the new hybridized species instead of going directly to the wet lab.

Appendix

Crop Plant Species	GenBank Acc. No.	GI No.	Base Pairs
Oryza Sativa Japonica Group	X15901	11957	134525
Brachypodium distachyon Cultivar	EU325680	193075536	135199
Agrostis Stolonifera Cultivar	EF115543	118201189	136584
Hordeum Vulare Cultivar	EF115541	118201020	136462
Zea mays	X86563	11990232	140384
Saccharum hybrid Cultivar	AP006714	49659489	141182
Sorghum bicolor Cultivar	EF115542	118201104	140754
Triticum aestivum	AB042240	13928184	134545

References

- [1] J.P. Vogel, Genome sequencing and analysis of the model grass *Brachypodium distachyon*, *Nature*, **463**, (2010), 763-768.
- [2] R. Mathur and N. Adlakha, A least squares method to determine reticulation in eight grass Plastomes, *Online J. of Bioinformatics*, **12**(2), (2011), 230 -242.
- [3] U. Chouhan and K.R. Pardasani, A Maximum Parsimony Model to Reconstruct Phylogenetic Network in Honey Bee Evolution, *International Journal of Biological and Life Sciences*, **3** (3), (2007), 220-224.
- [4] U. Chouhan and K.R. Pardasani, A Linear Programming Approach to Study Phylogenetic networks in Honeybee, *Online J. of Bioinformatics*, **11**(1), (2010), 72-82.
- [5] D. Gusfield, D. Hickerson and S. Eddhu, An efficiently computed lower bound on the number of recombinations in phylogenetic networks: Theory

- and empirical study, *Discrete Applied Mathematics*, **155**(6-7), (2007), 806-830.
- [6] R. Mathur and N. Adlakha, Evolutionary network to predict the reassortment of avian-human A/H5N1 influenza virus in India, *Trends in Bioinformatics*, (2012) (published online first).
- [7] W. F. Doolittle, Phylogenetic classification and the universal tree, *Science*, **284** (5423), (1999), 2124-2128.
- [8] V. Makarenkov and P. Legendre, From a Phylogenetic Tree to a Reticulated Network, *Journal of Computational Biology*, **11** (1), (2004), 195-212.
- [9] C.R. Linder and L.H. Rieseberg, Reconstructing patterns of reticulate evolution in plants, *American Journal of Botany*, **91**(10), (2004), 1700-1708.
- [10] D.A. Morrison, Networks in phylogenetic analysis: new tools for population biology, *International Journal of Parasitology*, **35** (5), (2005), 567 – 582.
- [11] V. Makarenkov and B. Leclerc, Circular orders of tree metrics, and their uses for the reconstruction and fitting of phylogenetic trees, Mathematical hierarchies and Biology, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, *American Mathematical Society Providence*, **RI 37**, (1997), 183-208.
- [12] C. Semple and M. Steel, *Phylogenetics*, Oxford University Press, 2003.
- [13] P. Buneman, The Recovery of Trees from Measures of Dissimilarity, *Mathematics in Archaeological and Historical Sciences*, Edinburgh University Press, (1971), 387-395.
- [14] P. Legendre and V. Makarenkov, Reconstruction of Biogeographic and Evolutionary Networks Using Reticulograms, *Systematic Biology*, **51**(2), (2002), 199-216.
- [15] L.A. Zadeh, Fuzzy Sets, *Information and Control*, **8**(3), (1965), 338-353.
- [16] S.A. Mohaddes and M.G. Mohayidin, Application of the Fuzzy Approach for Agricultural Production Planning in a Watershed, a Case Study of the Atrak

- Watershed, Iran, *American-Eurasian Journal of Agricultural & Environmental Sciences*, **3**(4), (2008), 636-648.
- [17] H. J. Zimmermann, *Fuzzy set theory and its applications*, – 4th ed. Springer Private Limited, 2006.
- [18] E. Bortiri, D.D. Coleman, G.R. Lazo, et al., The complete chloroplast genome sequence of *Brachypodium distachyon*: sequence comparison and phylogenetic analysis of eight grass Plastomes, *BMC Research Notes*, **1**, (2008), 61.
- [19] <http://www.ncbi.nlm.nih.gov/nucleotide/>
- [20] J. Felsenstein, *PHYMLIP (Phylogeny Inference Package)* version 3.67: dnadist. Distributed by the author, Department of Genetics, University of Washington, Seattle, 1993.
- [21] N. Saitou and M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Molecular Biology and Evolution*, **4**(4), (1987), 406-425.
- [22] V. Makarenkov, T-Rex: reconstructing and visualizing phylogenetic trees and reticulation networks, *Bioinformatics*, **17**(7), (2001), 664-668.