

## IMPLEMENTATION OF SOME SIMILARITY COEFFICIENTS IN CONJUNCTION WITH MULTIPLE UPGMA AND NEIGHBOR-JOINING ALGORITHMS FOR ENHANCING PHYLOGENETIC TREES

By

**Tarik S.K.M. Rabie**

Suez Canal University, Faculty of Agriculture, Animal Production  
Department, 41522 Ismailia, Egypt. e-mail: t.rabie@scuegypt.edu.eg

.Received: **18/05/2010**

Accepted: **02/06/2010**

**Abstract:** *Random Amplified Polymorphic DNA (RAPD) markers was used to analyze the genetic structure of five Indigenous Egyptian's chicken populations including Fayoumi, Dokki-4, Golden Montazah, Silver Montazah, and El-Salam, based on the taxa generated by the analysis of ten RAPD markers. The population genetic distances were estimated by using two cluster algorithms (UPGMA & NJ neighbor-joining) accompanied with ten similarity coefficients comprising Jaccard, Sørensen-Dice, Russel & Rao, Rogers & Tanimoto, Simple Matching, Pearson Phi, Lance & Williams, Mountford, Michael, and Kulchensky-1. The results demonstrated that for almost all methodologies, the Jaccard and Sørensen-Dice followed by Simple Matching coefficients revealed extremely close results, because both of them exclude negative co-occurrences. Due to the fact that there is no guarantee that the DNA regions with negative co-occurrences between two strains are indeed identical, the use of coefficients such as Jaccard and Sørensen-Dice that do not include negative co-occurrences was imperative for closely related organisms along with the NJ neighbor-joining cluster algorithm.*

### INTRODUCTION

Molecular markers are commonly used to characterize genetic diversity within or among populations or groups of individuals because they typically detect high levels of polymorphism. Furthermore, dominant markers such as RAPDs and AFLPs are efficient in allowing multiple loci to be analysed for each individual in a single gel run. In analyzing banding patterns of molecular markers, the data typically are coded as 0, and 1 vectors, 1 indicating the presence and 0 indicating the absence of a band at a specific position in the gel. With the codominant markers, such as RFLPs or SSRs, the banding patterns may be translated to homozygous or

heterozygous genotypes at each locus, and the allelic structure derived is utilized for comparison among individuals (**Peakall *et al.*, 1995; Smouse and Peakall 1999; Maguire *et al.*, 2002**). More often, however, the binary patterns obtained are used directly in comparisons of similarity of individuals. However, using random amplified polymorphic DNA (RAPD) to assess genetic variability may be slightly more complicated since the nature of the data may generate a large amount of random variation, especially at higher taxonomic levels (**Hillis *et al.*, 1996**). More seriously, many papers deal with individuals and not with allele frequencies in local populations or samples, and try to cluster or ordinate them (**Colombo *et al.*, 2000**). Although careful inspection of the results could eliminate part of this random noise such as choosing only the stronger, repeatable bands (**Dominguez, 1998**), this is usually a subjective procedure and does not necessarily avoid the problems of high dimensionality in the genetic structure itself.

Considering that the results of genotype clustering can be affected by the choice of the similarity coefficient (**Jackson *et al.*, 1989; Duarte *et al.*, 1999**), these coefficients employ several reasons of similarity or differences by total comparisons, and their values vary from 0 to 1 (**Skroch *et al.*, 1992**). The choice of an appropriate coefficient of similarity is a very important and decisive point to evaluate clustering, true genetic similarity between individuals, analyze diversity within populations and study relationship between populations, because different similarity coefficients may yield contrary outcomes (**Kosman and Leonard 2005**).

Though many coefficients are available, published studies usually do not justify their preference for any one in particular. Jackson *et al.* (1989) stated that this lack of justification can cause problems, jeopardizing the nature of the analysis. Therefore, these coefficients need to be better studied. Therefore, the objective of the study is to understand how the selection of a similarity coefficient affects resulting classifications, to determine the relatedness between the similarities coefficient, and to get the better genetic distances used for fragment parameters and biodiversity.

## **MATERIALS AND METHODS**

### ***Experimental design.***

In the present study, the data formed by **Rabie and Abdou (2010)** were used. Five Egyptian indigenous strains including Fayoumi, Dokki-4, Golden Montazah, Silver Montazah, and El-Salam compared by using 10

random amplification of polymorphic DNA (RAPD) primers. The taxa were used for further analysis.

***Interpretation of RAPD data.***

Arbitrary amplification of polymorphic DNA sequences, termed random amplification of polymorphic DNA (RAPD) analysis or Arbitrarily Primed PCR (AP-PCR) typing has been used for estimating genetic relatedness in livestock animals and avian populations. In addition, a RAPD-PCR has proved to be useful for genetic differentiation of closely related organisms and has been used to generate information on the population structure. Therefore, RAPD profiles from the pooled sample amplification were used for estimation of intra-population genetic similarity as well as genetic distance based on band frequency. The results were analyzed by comparing RAPD profiles on the basis of the presence (1) or absence (0) of each DNA band on the photographed agarose gels. A data matrix was constructed. Gel analyses were performed using GelQuest and ClusterVis software packages (SequentiX - Digital DNA Processing, Klein Raden, Germany). GelQuest software (Version 2.7.1) was used to analyze DNA fingerprint profiles and fragment patterns followed by using ClusterVis application (Version 1.4.2) that analyze binary matrix data as generated by GelQuest. For the calculation of similarities or dissimilarities (distances), the number of total matches (a), single matches (b, c) and no matches (d) are calculated out of the number of total positions ( $n=a+b+c+d$ ). Therefore, the matrices were analyzed to evaluate the differences resulting from the use of ten of the most commonly used similarity coefficients; the Jaccard, Sørensen-Dice, Russel & Rao, Rogers & Tanimoto, Simple Matching, Pearson Phi, Lance & Williams, Mountford, Michael, and Kulchensky-1 (Table 1) were compared along with different cluster algorithms (NJ-Neighbor-Joining, and UPGMA or un-weighted pair group method analysis), as well as the same coefficient with different cluster algorithms is shown in Table 1. Similarity coefficients were calculated and compared using the Spearman's correlation coefficient (**Hollander and Wolf, 1973**) by using Wessa software version 1.1.23-r6 (**Wessa, 2010**). Levels of statistical significance are not given because the analyses are derived from a single initial data matrix and therefore lack independence.

**Table 1:** Normalized similarity and correlation measures.

Coefficients	Similarity Formula	Sources
Jaccard	$\frac{a}{(n-d)}$ [= $a / (a+b+c)$ ]	Jaccard, 1901
Russel & Rao	$a/n$	Russel & Rao, 1940
Rogers & Tanimoto	$(a+d)/(a+2*(b+c)+d)$	Rogers & Tanimoto, 1960
Kulczynski-1	$a/(b+c)$	Kulczynski, 1928
Sørensen-Dice	$2*a/(2*a+b+c)$	Sørensen, 1948; Dice, 1945
Michael	$4*(a*d-b*c)/((a+d)*(a+d)+(b+c)*(b+c))$	Michael, 1920
Simple Matching	$\frac{a(a+d)}{(a+b+c+d)}$ = $(a+d)/n$	Sokal and Michener, 1958
Mountford	$2*a/(2*b*c+a*b+a*c)$	Mountford, 1962
Pearson's Phi coefficient	$((a*d)-(c*b))/\sqrt{(a+c)*(c+d)*(a+b)*(b+d)}$	Sokal and Sneath, 1963
Lance & Williams distance	$1 - D$	Lance and Williams, 1967

## RESULTS AND DISCUSSION

A presence / absence data matrix of five strains generated by ten RAPD markers was transformed into ten similarity matrices using ten different coefficients. Likewise, the dendrograms obtained from the use of two cluster algorithms. The Spearman's correlation coefficients between the ten similarity coefficients were equal or close to 1 (Table 2), making it evident that they are highly related. Levels of statistical significance are not given because the analyses are derived from a single initial data matrix and therefore lack independence. The Jaccard, Sørensen-Dice, Russel & Rao, and Lance & Williams coefficients presented correlation values equal to 1, demonstrating that there is no alteration in the ranks using any one of these coefficients, whereas, they classify the similarity among strains exactly in the same order.

Surprisingly, they were all similarity coefficients found to be equally able to classify the taxa except Mountford, and Kulczynski-1 coefficients, with equivalent dendrogram occurring for NJ-Neighbor-Joining algorithm, but the different dissimilarity values (Figure 1). Spearman's rank correlations were high between the two similarity coefficients than others (0.994) (Table 2). However, the correlations between these two classes of coefficients and the Simple Matching coefficient were lower. These results are similar to those reported by Meyer et al. (2004) while AFLP and RAPD markers were used.

Concerning UPGMA cluster algorithm, an analytical method for testing an interior branch was first used by Nei et al. (1995), and then by Li (1989) for an unrooted tree for four or five taxa, correspondingly to this

study, five taxa were used. The dendrograms showed that although the common structure of the dendrograms is highly comparable with the NJ-Neighbor-Joining algorithm, which is a widely used method for tree building which combines computational speed with uniqueness of result, while most implementations gave a single tree (**Page and Holmes, 2003**) there are minor alterations in the levels in which strains are clustered (Figure 1). The dendrogram constructed by Jaccard, Sørensen-Dice, Russel & Rao, and Lance & Williams (LW) coefficients showed some distinct differences corroborating the similarity matrices outcomes (Table 2). In addition, the rest six coefficients (Rogers & Tanimoto, Simple Matching, Pearson Phi, Mountford, Michael, and Kulchensky-1) were revealed some alterations in the grouping of the strains in other way (Figure 1). This also corroborates the differences observed in the similarity matrices where they have related values given greater than others except Simple Matching, and Michael similarity coefficients when carried out with NJ-Neighbor-Joining algorithm in contrast with the UPGMA algorithm, where the related values of Russel & Rao coefficient were greater than all the similarity coefficients except Mountford, and Kulchensky-1 (Figure 2). Although the inclusion or exclusion of negative matches,  $d$  in the binary similarity measures have been an ongoing issue (**Sokal and Sneath 1963; and Sneath, and Sokal, 1973**), the Jaccard, Tanimoto, Sørensen -Dice, Kulczynski-1, and Mountford are included in the negative match exclusive measures. **Sokal and Sneath (1963)** argued that the negative matches don't mean necessarily any similarity between two objects.

The Sørensen-Dice coefficient of similarity is frequently referred to as the measure of genetic similarity of **Nei and Li (1979)**. For a given data set, the related values of Jaccard's similarity are always greater than those of the Sørensen-Dice, RR, SM, LW, and M similarity coefficients (Figure 2). **Landry and Lapoint (1996)** suggested that the Sørensen-Dice or, and Jaccard coefficients might be preferable to the Simple Matching coefficient when RAPD analysis used to compare groups of distantly related taxa. In contrast, **Halden et al. (1994)** considered the Simple Matching coefficient to be the more appropriate measure of similarity when closely related taxa included, but **Kosman and Leonard (2005)** reported that choice should be supported with estimates of DNA sequenced entity between the taxa. A like this study, in the absence of supporting sequence identity estimates similarity values based on dominant markers data should be regarded as tentative.

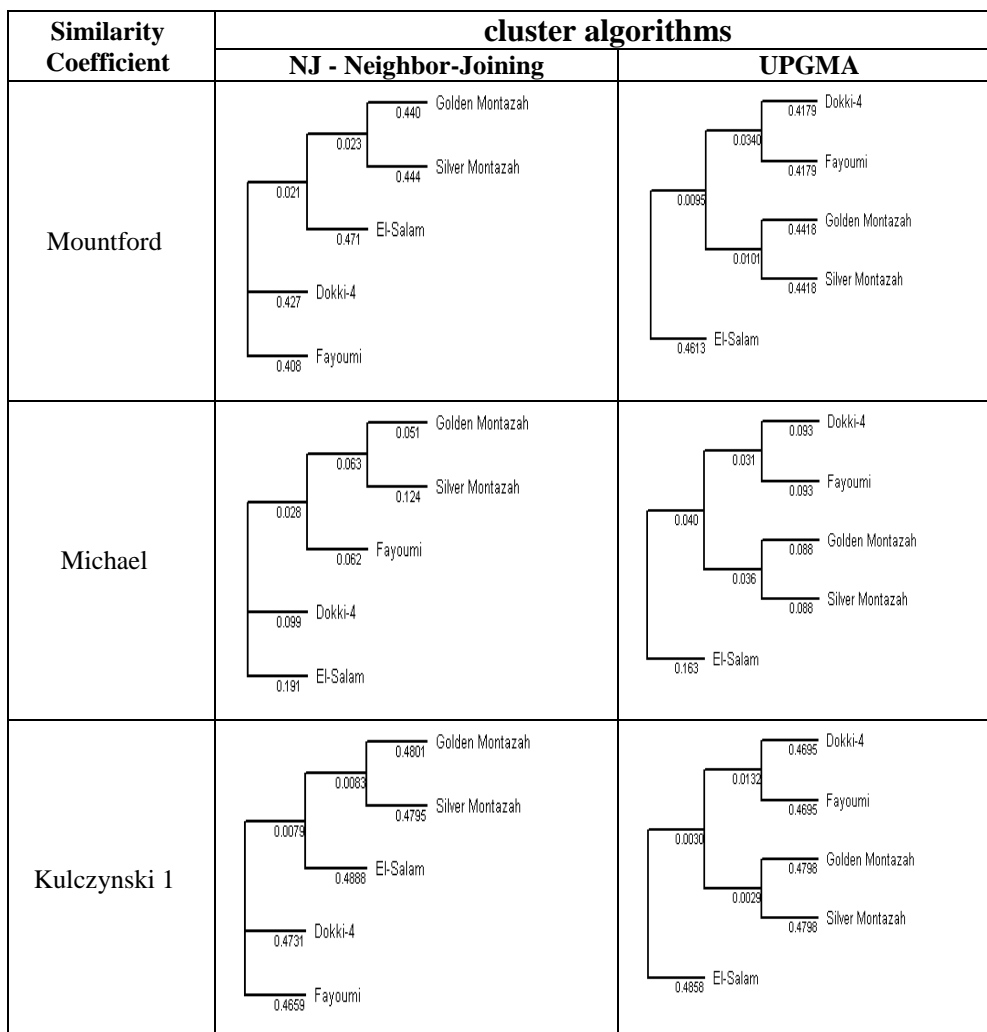
The similar appearance in Jaccard and Sørensen-Dice's coefficients-based dendrograms can be simplified by the properties of these coefficients. They are discriminated by the way in which the matrix of original data

(1=presence, and 0=absence) is employed in the similarity estimate. When two genotypes are compared, the following situations occur:  $a = 1.1; b = 1.0; c = 0.1; d = 0.0$ . Thus, Jaccard and Sørensen-Dice's coefficients are equivalent, except that double weight is given to positive co-occurrences ( $a$ ) in the Sørensen-Dice's coefficient whereas the Simple Matching coefficient includes negative co-occurrences ( $d$ ) (Duart *et al.*, 1999).

Similarity Coefficient	cluster algorithms	
	NJ - Neighbor-Joining	UPGMA
Jaccard	<p>Detailed description: NJ dendrogram for Jaccard coefficient. Golden Montazah and Silver Montazah cluster at 0.123. Fayoumi joins this pair at 0.057. Dokki-4 joins the group at 0.015. El-Salam joins the entire group at 0.108. The root is at 0.180.</p>	<p>Detailed description: UPGMA dendrogram for Jaccard coefficient. Golden Montazah and Silver Montazah cluster at 0.123. Fayoumi joins this pair at 0.089. Dokki-4 joins the group at 0.051. El-Salam joins the entire group at 0.140. The root is at 0.017.</p>
Sørensen-Dice	<p>Detailed description: NJ dendrogram for Sørensen-Dice coefficient. Golden Montazah and Silver Montazah cluster at 0.070. Fayoumi joins this pair at 0.039. Dokki-4 joins the group at 0.012. El-Salam joins the entire group at 0.060. The root is at 0.108.</p>	<p>Detailed description: UPGMA dendrogram for Sørensen-Dice coefficient. Golden Montazah and Silver Montazah cluster at 0.070. Fayoumi joins this pair at 0.049. Dokki-4 joins the group at 0.033. El-Salam joins the entire group at 0.062. The root is at 0.012.</p>
Russel & Rao	<p>Detailed description: NJ dendrogram for Russel &amp; Rao coefficient. Golden Montazah and Silver Montazah cluster at 0.2105. Fayoumi joins this pair at 0.0280. Dokki-4 joins the group at 0.0082. El-Salam joins the entire group at 0.1760. The root is at 0.2056.</p>	<p>Detailed description: UPGMA dendrogram for Russel &amp; Rao coefficient. Golden Montazah and Silver Montazah cluster at 0.1645. Fayoumi joins this pair at 0.0230. Dokki-4 joins the group at 0.0164. El-Salam joins the entire group at 0.1875. The root is at 0.0066.</p>

Cluster Analysis, UPGMA, NJ Neighbor-Joining, Similarity Coefficient.

Similarity Coefficient	cluster algorithms	
	NJ - Neighbor-Joining	UPGMA
Rogers & Tanimoto		
Simple Matching		
Pearson Phi		
Lance & Williams		



**Figure 1** - Dendrograms constructed from matrices of genetic distances obtained by distinction of the similarity coefficients by using two cluster algorithms (NJ-Neighbor-Joining, and UPGMA).

Taking into consideration the genetic basis of RAPD markers (**Williams *et al.*, 1990**), the absence of amplification of a determined band in two genotypes does not necessarily represent genetic similarity between them, which makes those coefficients that exclude these negative co occurrences from their expression of similarity (Jaccard, Sørensen-Dice, ... etc.) more adequate for use with this type of marker. **Sokal and Sneath (1963)** also stated that the simpler the coefficient the easier its interpretation; therefore, simpler coefficients should preferentially be employed. Jaccard's similarity coefficient is the simplest of its category (exclusion of *d*), and it has been widely employed with RAPD markers.

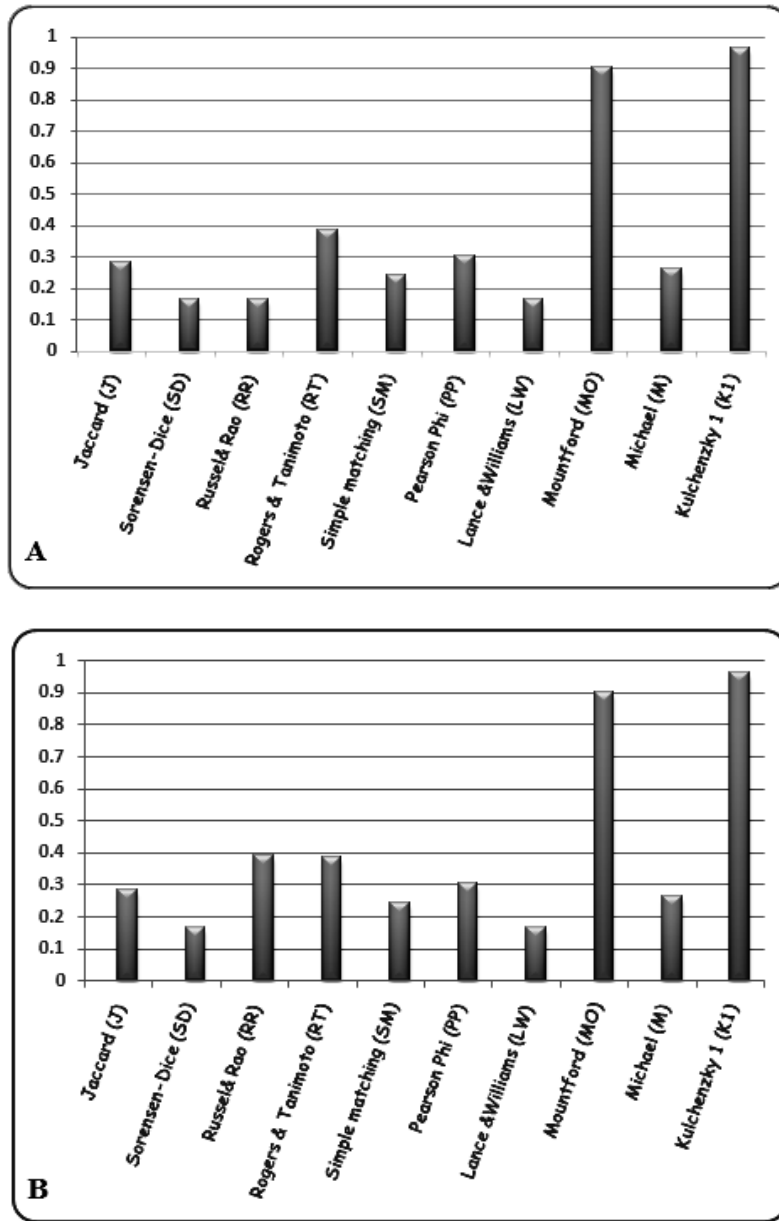
Furthermore, The Jaccard coefficient proposed at 1901 is still widely used in the various fields such as ecology and biology.

Although both of UPGMA, and NJ- neighbor-Joining techniques are designed to produce single trees (Nei, 1987; Saitou and Nei 1987), they can sometimes derive more than one dendrogram from the same data. This “Chaotic” behavior depends on the order in which data are entered (Bayer, 1985; Kovach, 1993). The results of this study agreed with that, but in contrast with Denduanghoripant *et al.* (2010) who stated that an UPGMA dendrogram revealed the same grouping found in the NJ tree. Furthermore, the authors of the NJ method think that it is more suitable for constructing phylogenetic trees than, for example, UPGMA (Saitou and Nei, 1987). Furthermore, NJ method is comparatively rapid and generally gives better results than UPGMA method. But it produces only one tree and neglects other possible trees, which might be as good as NJ trees, if not significantly better. Moreover, since errors in distance estimates are exponentially larger for longer distances, under some condition, this method will yield a biased tree (Bruno et al, 2000)

The conclusion is that although the similarity coefficients perform equally well, they are not interchangeable, as different proportions of false-positive and false-negative classifications may result. Although there are no solid arguments in favor of a particular similarity coefficient, the analyses of their behavior and properties help to guide the choice of a coefficient. However, the results suggest that based on the structures generated, the best index to use is Jaccard or Sørensen-Dice, followed by Simple Matching similarity coefficient accompanied with NJ- neighbor-Joining cluster algorithm.

**Table 2.** The Spearman's rank correlation coefficient obtained from use of ten similarity coefficients and two cluster methods, applied to a presence/absence data matrix of the five of Egyptian indigenous chicken strains.

	<b>J</b>	<b>SD</b>	<b>RR</b>	<b>RT</b>	<b>SM</b>	<b>PP</b>	<b>LW</b>	<b>MO</b>	<b>M</b>	<b>K1</b>
<b>NJ- neighbor-Joining</b>										
<b>Jaccard (J)</b>	1.000									
<b>Sørensen-Dice (SD)</b>	0.999	1.000								
<b>Russel&amp; Rao (RR)</b>	0.999	1.000	1.000							
<b>Rogers &amp; Tanimoto (RT)</b>	0.994	0.994	0.994	1.000						
<b>Simple Matching (SM)</b>	0.993	0.994	0.994	0.999	1.000					
<b>Pearson Phi (PP)</b>	0.947	0.949	0.949	0.975	0.978	1.000				
<b>Lance &amp;Williams (LW)</b>	0.999	1.000	1.000	0.994	0.994	0.949	1.000			
<b>Mountford (MO)</b>	0.972	0.963	0.963	0.976	0.964	0.935	0.963	1.000		
<b>Michael (M)</b>	0.849	0.856	0.856	0.896	0.906	0.970	0.856	0.824	1.000	
<b>Kulchenzky 1 (K1)</b>	0.969	0.960	0.960	0.963	0.950	0.899	0.960	0.994	0.768	1.000
<b>UPGMA</b>										
<b>Jaccard (J)</b>	1.000									
<b>Sørensen-Dice (SD)</b>	1.000	1.000								
<b>Russel&amp; Rao (RR)</b>	0.824	0.825	1.000							
<b>Rogers &amp; Tanimoto (RT)</b>	0.991	0.990	0.740	1.000						
<b>Simple matching (SM)</b>	0.991	0.991	0.742	1.000	1.000					
<b>Pearson Phi (PP)</b>	0.929	0.929	0.561	0.971	0.970	1.000				
<b>Lance &amp;Williams (LW)</b>	1.000	1.000	0.825	0.990	0.991	0.929	1.000			
<b>Mountford (MO)</b>	0.980	0.977	0.715	0.993	0.989	0.965	0.977	1.000		
<b>Michael (M)</b>	0.853	0.854	0.417	0.915	0.916	0.985	0.854	0.907	1.000	
<b>Kulchenzky 1 (K1)</b>	0.992	0.989	0.815	0.984	0.981	0.921	0.989	0.987	0.839	1.000



**Figure 2.** Mean of the matrices values resulted from using of all compliant similarity coefficients applied to a presence/absence data matrix of the five of Egyptian indigenous chicken strains by using of (A) NJ - Neighbor-Joining, and (B) UPGMA cluster algorithms.

## REFERENCES

- Bayer, U. (1985).** Lecture notes in earth sciences. Vol. 2. Pattern recognition problems in geology and paleontology. *Springer-Verlag, Berlin/ Heidelberg.*
- Bruno W.J.; Succi, N.D.; and Halpern, A.L. (2000)** Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Mol. Biol. Evol.*, 17, 189–197.
- Colombo C.; Second, G.; and Charrier, A. (2000)** Diversity within American cassava germ plasm based on RAPD markers. *Genet. Mol. Biol.* 23:189-200.
- Denduangboripant, J.; Setaphan, S. ; Suwanprasart, W.; and Panha, S. (2010)** *Determination of Local Tobacco Cultivars Using ISSR Molecular Marker.* *Chiang Mai J. Sci.* 37: 293-303.
- Dice, L.R. (1945)** Measures of the amount of ecologic association between species. *Ecology* 26, 297–302.
- Dominguez, P.A. (1998)** An evaluation of RAPD fragment reproductibility and nature. *Mol. Ecol.* 7:1347-1357.
- Duarte, J.M.; dos Santos, J.B.; and Melo, L.C. (1999).** Comparison of similarity coefficients based on RAPD markers in the common bean. *Genet. Mol. Biol.* 22: 427-432.
- Hallden, C.; Nilsson, N.O.; Rading, I.M.; Sall, T. (1994)** Evaluation of RFLP and RAPD markers in a comparison of *Brassica napus* breeding lines. *Theor. Appl. Gene.*, 88, 123–128.
- Hillis, D.M.; Mable, B.K.; and Moritz, C. (1996)** Applications of molecular systematics: the state of the field and a look to the future. In: Hillis DM, Moritz C, Marble BK (eds) *Molecular Systematics*. 2nd ed. Sinauer Press, Sunderland, Massachusetts, pp 515-543.
- Hollander, M.; and Wolfe, D.A. (1973)** *Nonparametric Statistical Methods: Wiley, Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York-London-Sydney, xviii+503 pp.*
- Jaccard, P. (1901).** Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull. Soc. Vaudoise Sci. Nat.* 37: 547-579.
- Jackson, A.A.; Somers, K.M.; and Harvey, H.H. (1989).** Similarity coefficients: measures for co-occurrence and association or simply measures of occurrence? *Am. Nat.* 133: 436-453.

- Kosman, E.; Leonard, K.J. (2005).** *Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploidy species. Mol. Ecol. 14: 415-424.*
- Kovach, W.L. (1993).** *MVSP Plus: multivariate statistical package, version 2.1. Kovach Computing Services, Pentraeth, Anglesey.*
- Kulczynski, S. (1928).** *Die Pflanzenassoziationen der Pieninen, Bull. Int Acad. Pol. Sci., Ser.B, Suppl. 2, p. 57.*
- Lance, G.N.; and Williams, W.T. (1967).** *Mixed-data classificatory programs.1. Agglomerative systems. Aust. Comput.J. 1:15-20.*
- Landry, P.A.; and Lapointe, F.J. (1996)** *RAPD problems in phylogenetics. Zool. Scripta, 25, 283–290.*
- Li, W-H. (1989)** *A statistical test of phylogenies estimated from sequence data. Mol. Biol. Evol. 6:424–435.*
- Maguire, T.L.; Peakall, R.; and Saenger, P. (2002)** *Comparative analysis of genetic diversity in the mangrove species Avicennia marina (Forsk.) Vierh. (Avicenniaceae) detected by AFLPs and SSRs. Theoretical and Appl. Genet., 104, 388–398.*
- Meyer, A.; Garcia, A.A.F.; Souza, A.P.; and Souza, C.L. (2004)** *Comparison of similarity coefficients used for cluster analysis with dominant markers in maize (Zea mays L).Genet. Mol. Biol. 27: 83-91.*
- Michael, E.L. (1920)** *Marine ecology and the coefficient of association: a plea in behalf of quantitative biology, Ecology 8, 54-59.*
- Mountford, M.D. (1962)** *An index of similarity and its application to classificatory problems. In: Progr. in Soil Zool. (P.W. Murphy ed.) pp. 43–50. London 1962*
- Nei, M. (1987).** *Molecular evolutionary genetics. Columbia University Press, New York.*
- Nei, M.; and Li, W.H. (1979)** *Mathematical model for studying genetic variation in terms of restriction endonucleasis. Proceedings of the National Academy of Sciences of the USA, 76, 5269–5273.*
- Nei, M.; Takezaki, N.; and Sitnikova, T. (1995).** *Assessing molecular phylogenies. Science 13, 267: 253-254*
- Page, R.D.M.; and Holmes, E.C. (2003).** *Molecular Evolution, a phylogenetic approach. Blackwell publishing company.USA. ISBN 0-86542-889-1 p.183*

- Peakall, R.; Smouse, P.E.; and Huff, D.R. (1995)** *Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloë dactyloides*. Mol. Ecol., 4, 135–147.*
- Rabie, T.S.K.M.; and Abdou, A.M. (2010)** *Genetic diversity and relationship among Egyptian indigenous chicken strains using random amplification of polymorphic DNA. Accepted for publication.*
- Rogers, D.J.; and Tanimoto, T.T. (1960)** *A computer program for classifying plants. Science 132: 1115-1118.*
- Russel, P.F.; and Rao, T.R. (1940).** *On habitat and association of species of anopheline larvae in south-eastern Madras. J. Malaria Inst. India 3: 153-178.*
- Saitou, N.; and Nei, M. (1987)** *The Neighbor-Joining Method: A New Method for Inferring Evolutionary Trees, Mol. Biol. Evol. 4, 406–425.*
- Skroch, P.; Tivang, J.; and Nienhuis, J. (1992).** *Analysis of genetic relationships using RAPD marker data. In: Applications of RAPD Technology to Plant Breeding. Joint Plant Breeding Symposia Series, Minneapolis, 1992. CCSA, ASHS, and AGA, Madison, pp. 26-30.*
- Smouse, P.E.; and Peakall, R. (1999)** *Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity, 82, 561–573.*
- Sneath, P.H.A.; and Sokal, R.R. (1973)** *Numerical Taxonomy: The Principles and Practice of Numerical Classification, W.H. Freeman and Company, San Francisco, 576 pp.*
- Sokal, R.R.; and Michener, C.D. (1958).** *A statistical method for evaluating systematic relationships. Univ. Kans. Sci. Bull. 38: 1409-1438.*
- Sokal, R.R.; and Sneath, P.H.A. (1963).** *Principles of Numeric Taxonomy. W.H. Freeman, San Francisco.*
- Sørensen, T.A. (1948)** *A method of establishing groups of equal amplitude in plant sociology based on similarity of species content, and its application to analyses of the vegetation on Danish commons. K. danske Vidensk. Selsk. Biol. Skr. 5(4), 1–34*
- Wessa, P. (2010),** *Free Statistics Software, Office for Research Development and Education, version 1.1.23-r6, URL <http://www.wessa.net/>*

Williams, J.G.K.; Kubelik, A.R.; Livak, K.J.; Rafalski, J.A.; and Tingey, S.V. (1990). *DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18: 6531-6535.*

### الملخص العربي

UPGMA تنفيذ بعض معاملات التشابه المقترنة بالخوارزميات المتعددة -

لتعزيز أشجار النشوء والتطور Neighbor-Joining و

طارق السعيد كامل محمد ربيع

جامعة قناة السويس، كلية الزراعة، قسم الإنتاج الحيواني، 41522 الإسماعيلية،

جمهورية مصر العربية.

تم استخدام لتحليل التركيب الجيني لخمس عشائر من الدجاج المحلى المصري بما في ذلك الدجاج الفيومي ، الدقي- 4 ، المنتزه الذهبي ، المنتزه فضة ، والسلام، استنادا إلى البيانات الوراثية التي تم إنشاؤها بواسطة تحليل عشر من علامات الـ RAPD. قدرت المسافات الجينية للعشائر باستخدام الخوارزميات المتعددة - UPGMA و Neighbor-Joining المقترنة RAPD مع عشر من معاملات التشابه

و التي تشمل على Jaccard, Sørensen-Dice, Russel & Rao, Rogers & Tanimoto, Simple Matching, Pearson Phi, Lance & Williams, Mountford, Michael, و Kulchenzky-1

وقد أثبتت النتائج أن جميع الطرق تقريبا و أن النتائج قريبة للغاية لكلا من Jaccard و Sørensen-Dice يليها معامل الـ Simple Matching، لأن كلا منهم استبعد المشاركة السالبة. يرجع ذلك إلى حقيقة أن ليس هناك ما يضمن أن مناطق الـ DNA مع المشاركة السالبة المشترك بين سلالتين هي في الواقع متطابقة ، استخدام معاملات مثل Jaccard و Sørensen-Dice والتي لا تتضمن المشاركة السالبة الحتمية للكائنات ذات الصلة الوثيقة بالإنضمام الى الكتلة الخوارزمية Neighbor-Joining.