

# How can Sequence Alignment Method Affect the Phylogenetic Relationship Results?

Hossam F. Abou-Shaara

Damanhour University, Faculty of Agriculture, Department of Plant Protection, Damanhour, 22516, Egypt

---

## Abstract

Sequence alignment is a fundamental step before analysing the phylogenetic relationships between organisms. The effect of the alignment method on the phylogenetic relationship results has not been well studied. Three methods for sequence alignment are available in Geneious software: (1) Geneious alignment, (2) ClustalW, and (3) MUSCLE. The study aimed to investigate the impacts of these three alignment methods on the phylogenetic relationships between test organisms. Therefore, the phylogenetic tree was established for the same organisms (honey bee species and subspecies) after using the three alignment methods, and then the results were compared. The constructed tree after using ClustalW differed than Geneious alignment and MUSCLE without a clear effect on relationship between the closely related organisms. Based on this study, it is better for researchers to test different alignment methods before constructing the final phylogenetic tree to select the most suitable one.

**Keywords:** bioinformatics; software; Geneious; NCBI; genetics.

---

## 1. Introduction

Bioinformatics is essential interdisciplinary field of science to understand relationships between organisms based on their genetic materials. Also, this field of science helps in understanding the genetic structure of organisms including the effects of hybridization on them [1]. Bioinformatics is developed rapidly in line with the development of computer programs and mathematical models. There are various computer programs and online resources available to perform bioinformatics studies.

The phylogenetic relationships between species and subspecies of many organisms have been studied [1-6]. Sequence alignment is the basic step before constructing the phylogenetic tree between organisms. The alignment can be performed using different methods including: ClustalW [7], MUSCLE [8], and Geneious [9]. The effects of

the sequence alignment method on the constructed phylogenetic tree have not been widely investigated. There are many programs currently available to study the phylogenetic relationships between organisms including Geneious software [9]. This software allows the use of different alignment methods. Therefore, the study aimed to investigate the effects of three alignment methods available at Geneious (Geneious alignment, ClustalW, and MUSCLE) on the results of the phylogenetic relationships between some honey bee species and subspecies (test organisms). The outcomes of this study are essential for researchers using bioinformatics.

## 2. Materials and methods

### Sequences

Sequences for some honey bee subspecies and species were selected to perform the analysis. The sequences were downloaded from the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/>) (NCBI) for *Apis mellifera lamarckii* (Accession no. KY464958.1,

---

\* Corresponding author: Hossam Abou-Shaara, Tel, 0453209743 Fax, 0453316535 Email, [hossam.farag@agr.dmu.edu.eg](mailto:hossam.farag@agr.dmu.edu.eg)

16589 bp, [3]), *Apis mellifera sahariensis* (Accession no. NC\_035883.1, 16569 bp, [10]), *Apis mellifera scutellata* (Accession no. KJ601784.1, 16411 bp, [11]), *Apis mellifera capensis* (Accession no. KX870183.1, 16470 bp, [2]), and *Apis mellifera syriaca* (Accession no. KP163643.1, 15428 bp, [12]). These subspecies are highly close to each other and belong to the same species (*mellifera*). Also, sequences to some close bee species were used in the analysis, *Apis nuluensis* (Accession no. NC\_036235.1, 15843 bp, [13]), *Apis cerana* (Accession no. GQ162109.1, 15895 bp, [14]), *Apis koschevnikovi* (Accession no. AP017643.1, 15278 bp, [15]), *Apis laboriosa* (Accession no. NC\_036155.1, 15058 bp, [5]), and *Apis florea* (Accession no. JX982136.1, 17694 bp, [16]).

### Sequence alignment methods

The analysis was performed using Geneious prime 2019.1 (www.geneious.com). Three alignment methods were tested: 1) Geneious alignment (type: global alignment with free end gaps, and

cost matrix: 65% similarity), 2) MUSCLE (3.8.425 by Robert C. Edgar with maximum number of iterations of 8), and 3) ClustalW (2.1, cost matrix: IUB, gap open cost=15, gap extend cost= 6.66, with free end gaps). Then, the phylogenetic trees were built using geneious tree builder (genetic distance model: Tamura-Nei, tree build method: Neighbor-joining, and without outgroup) for the aligned sequences by the three methods.

### 3. Results and Discussion

The results varied according to the alignment method (Figure 1 and 2). The subspecies *A. m. sahariensis* had the lowest identity percentage according to the three alignment methods. The identity percentage of *A. m. sahariensis* ranged from 53.19 to 55.82%, 43.87 to 48.29%, and 48.50 to 49.45% for Geneious alignment, MUSCLE, and ClustalW, respectively.

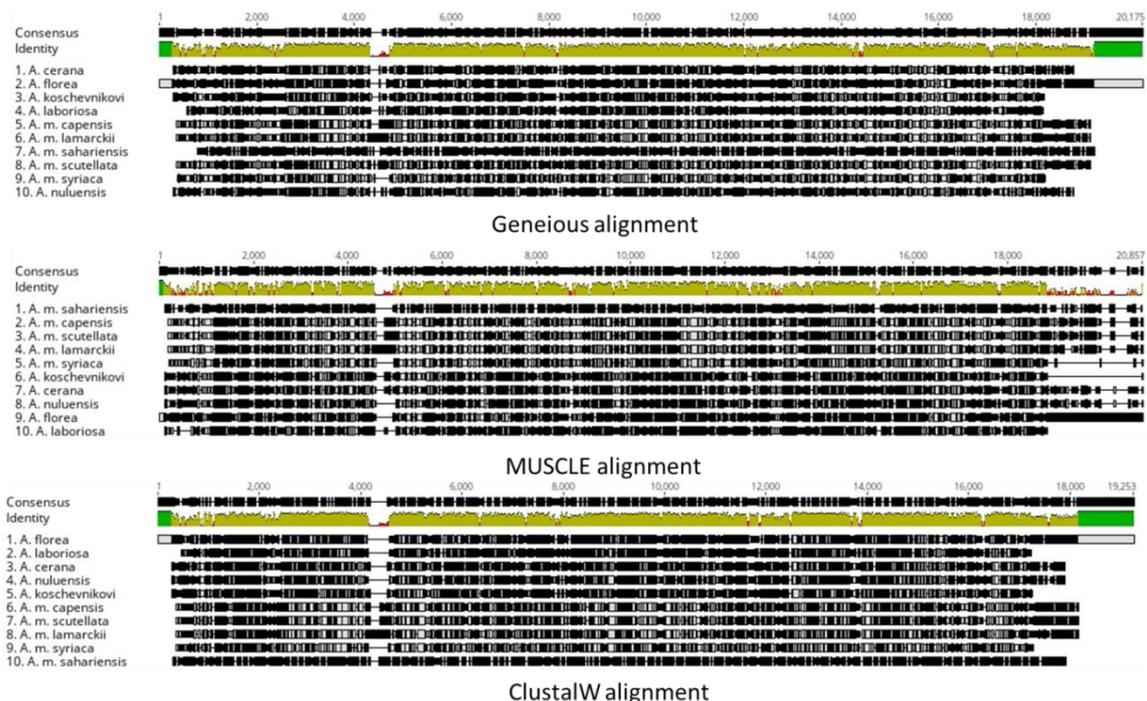


Figure 1. Sequence alignment using three methods

	A. cerana	A. florea	A. koschevnik...	A. laboriosa	A. m. capensis	A. m. lamarkii	A. m. saharie...	A. m. scutell...	A. m. syriaca	A. nuluensis
A. cerana		78.96%	84.73%	80.95%	81.08%	79.83%	54.20%	81.21%	81.73%	90.73%
A. florea	78.96%		80.03%	81.64%	78.27%	77.08%	53.19%	78.34%	80.01%	78.85%
A. koschevnikovi	84.73%	80.03%		81.54%	82.30%	81.30%	54.05%	82.46%	82.37%	84.40%
A. laboriosa	80.95%	81.64%	81.54%		80.37%	79.51%	53.52%	80.53%	80.97%	81.21%
A. m. capensis	81.08%	78.27%	82.30%	80.37%		95.76%	55.79%	98.24%	95.91%	80.80%
A. m. lamarkii	79.83%	77.08%	81.30%	79.51%	95.76%		54.98%	95.08%	95.70%	79.46%
A. m. sahariensis	54.20%	53.19%	54.05%	53.52%	55.79%	54.98%		55.82%	55.79%	53.92%
A. m. scutellata	81.21%	78.34%	82.46%	80.53%	98.24%	95.08%	55.82%		96.30%	81.30%
A. m. syriaca	81.73%	80.01%	82.37%	80.97%	95.91%	95.70%	55.79%	96.30%		82.36%
A. nuluensis	90.73%	78.85%	84.40%	81.21%	80.80%	79.46%	53.92%	81.30%	82.38%	

**Geneious alignment**

	A. m. saharie...	A. m. capensis	A. m. scutell...	A. m. lamarkii	A. m. syriaca	A. koschevnik...	A. cerana	A. nuluensis	A. florea	A. laboriosa
A. m. sahariensis		47.94%	47.96%	47.29%	46.44%	45.49%	48.29%	47.88%	43.87%	47.19%
A. m. capensis	47.94%		99.16%	96.32%	91.83%	77.82%	79.92%	79.52%	73.03%	79.88%
A. m. scutellata	47.96%	99.16%		96.03%	92.09%	78.06%	80.16%	79.77%	73.16%	80.09%
A. m. lamarkii	47.29%	96.32%	96.03%		92.23%	77.37%	78.91%	78.53%	71.95%	78.97%
A. m. syriaca	46.44%	91.83%	92.09%	92.23%		82.34%	79.57%	79.62%	70.54%	80.68%
A. koschevnikovi	45.49%	77.82%	78.06%	77.37%	82.34%		82.68%	82.13%	70.77%	80.78%
A. cerana	48.29%	79.92%	80.16%	78.91%	79.57%	82.68%		92.50%	73.14%	80.72%
A. nuluensis	47.88%	79.52%	79.77%	78.53%	79.62%	82.13%	92.50%		72.81%	80.83%
A. florea	43.87%	73.03%	73.16%	71.95%	70.54%	70.77%	73.14%	72.81%		80.01%
A. laboriosa	47.19%	79.88%	80.09%	78.97%	80.68%	80.78%	80.72%	80.83%	80.01%	

**MUSCLE alignment**

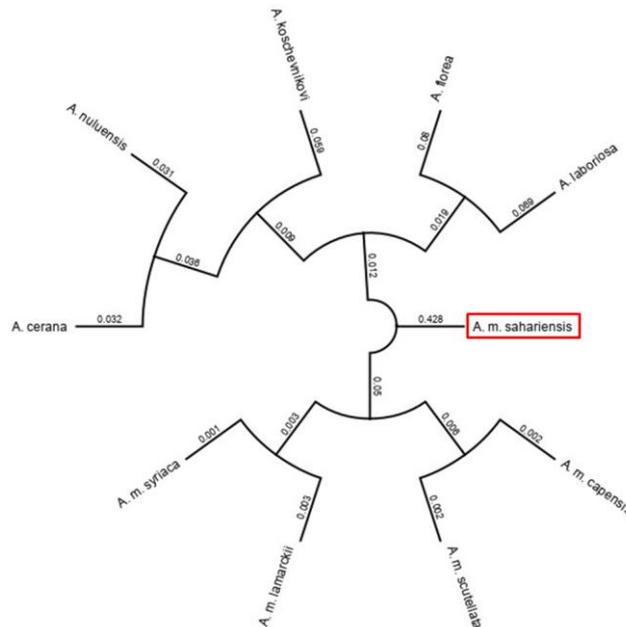
	A. florea	A. laboriosa	A. cerana	A. nuluensis	A. koschevnik...	A. m. capensis	A. m. scutell...	A. m. lamarkii	A. m. syriaca	A. m. saharie...
A. florea		83.28%	78.75%	78.77%	80.04%	77.42%	77.63%	76.31%	79.57%	48.50%
A. laboriosa	83.28%		80.66%	80.92%	80.87%	79.76%	80.00%	78.88%	80.68%	48.51%
A. cerana	78.75%	80.66%		93.03%	86.09%	80.78%	81.15%	80.09%	82.43%	49.01%
A. nuluensis	78.77%	80.92%	93.03%		85.82%	80.63%	81.01%	79.96%	82.49%	48.92%
A. koschevnikovi	80.04%	80.87%	86.09%	85.82%		82.14%	82.39%	81.14%	82.82%	48.95%
A. m. capensis	77.42%	79.76%	80.78%	80.63%	82.14%		99.16%	96.18%	96.87%	49.30%
A. m. scutellata	77.63%	80.00%	81.15%	81.01%	82.39%	99.16%		95.91%	97.11%	49.45%
A. m. lamarkii	76.31%	78.88%	80.09%	79.96%	81.14%	96.18%	95.91%		96.79%	48.78%
A. m. syriaca	79.57%	80.68%	82.43%	82.49%	82.82%	96.87%	97.11%	96.79%		49.07%
A. m. sahariensis	48.50%	48.51%	49.01%	48.92%	48.95%	49.30%	49.45%	48.78%	49.07%	

**ClustalW alignment**

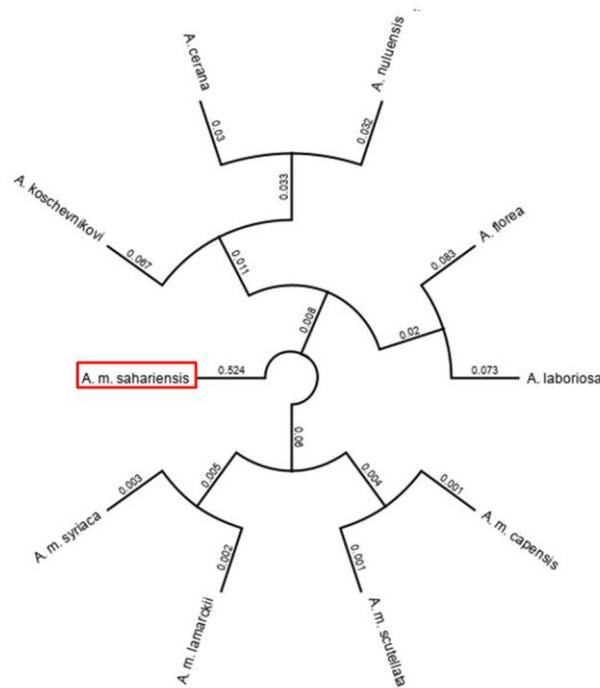
**Figure 2.** Heat-map and identity% resulted from the three methods. The lowest identity percentage is to *A. m. sahariensis* (marked by a rectangle)

The phylogenetic trees constructed after Geneious alignment or MUSCLE were highly similar (Figure 3 and 4) and correctly separate subspecies

than species while *A. m. sahariensis* was placed in a separated branch.



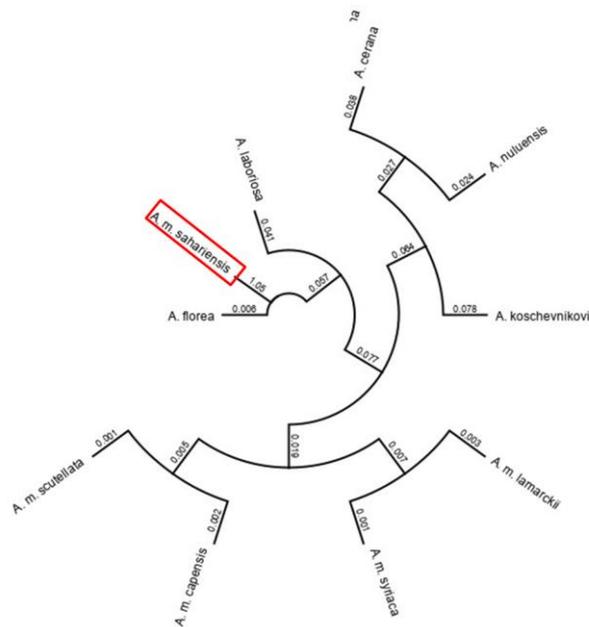
**Figure 3.** Phylogenetic trees constructed using Geneious alignment. The lowest identity percentage is to *A. m. sahariensis* (marked by a rectangle)



**Figure 4.** Phylogenetic trees constructed using MUSCLE (B). The lowest identity percentage is to *A. m. sahariensis* (marked by a rectangle)

The tree constructed after using ClustalW (Figure 5) was partially different, and incorrectly placed a subspecies (*A. m. sahariensis*) close to species (*A.*

*florea*). This can be explained by the differences in the mathematical models used in each sequence alignment method.



**Figure 5.** Phylogenetic trees constructed using ClustalW (C). The lowest identity percentage is to *A. m. sahariensis* (marked by a rectangle)

Most relationships between the studied subspecies except *A. m. sahariensis* were in line with

previous studies. For example, the close relationship between the Egyptian honey bees (*A.*

*m. lamarckii*) and the Syrian honey bees (*A. m. syriaca*) was previously found based on protein coding genes and rRNA [3]. Also, the Egyptian bees were placed away from other African bees including *A. m. capensis* and *A. m. scutellata* [6], and this supports the findings of the constructed phylogenetic trees in this study using different alignment methods. Moreover, the constructed trees successfully placed two wild bees: *A. florea* and *A. laboriosa* close to each other and away from cavity-nesting bees and this in accordance with [17], except when ClustalW was used (Fig. 5).

#### 4. Conclusion

It can be concluded that sequence alignment method can affect the phylogenetic relationship results. Based on this study, it is better to test different sequence alignment methods to select the most appropriate one for the analysis.

#### References

1. Abou-Shaara, H.F., Bayoumi, S.R. Using mitochondrial DNA similarity percentages to analyze the maternal source of hybrid bees from two honey bee subspecies. *Scientific Papers: Animal Science and Biotechnologies*, 2019, 52, 32-40.
2. Eimanifar, A., T. Kimball, R., L. Braun, E., Ellis, J. D., The complete mitochondrial genome of the Cape honey bee, *Apis mellifera capensis* Esch. (Insecta: Hymenoptera: Apidae). *Mitochondrial DNA Part B*, 2016, 1, 817-819. <https://doi.org/10.1080/23802359.2016.1241682>
3. Eimanifar, A., T. Kimball, R., L. Braun, E., M. Moustafa, D., Haddad, N., Fuchs, S., Grunewald, B., Ellis, J. D., The complete mitochondrial genome of the Egyptian honey bee, *Apis mellifera lamarckii* (Insecta: Hymenoptera: Apidae). *Mitochondrial DNA Part B*, 2017a, 2, 270-272. <https://doi.org/10.1080/23802359.2017.1325343>
4. Abou-Shaara, H.F. A scientific note on the evolutionary relationships between honey bees and their enemies. *Agronomical Research in Moldavia*, 2018, 51, 101-107. <https://doi.org/10.2478/cerce-2018-0009>
5. Takahashi, J. I., Rai, J., Wakamiya, T., Okuyama, H., Characterization of the complete mitochondrial genome of the giant black Himalayan honeybee (*Apis laboriosa*) from Nepal. *Conservation Genetic Resources*, 2018, 10, 59-63. <https://doi.org/10.1007/s12686-017-0765-6>
6. Abou-Shaara, H.F. Utilizing bioinformatics to detect genetic similarities between African honey bee subspecies. *Journal of Genetics*, 2019, 98, 96. <https://doi.org/10.1007/s12041-019-1145-7>
7. Thompson, J. D., Gibson, T. J., Higgins, D. G., Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics*, 2003, 1, 2-3. <https://doi.org/10.1002/0471250953.bi0203s00>
8. Edgar, R. C., MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 2004, 32, 1792-1797. <https://doi.org/10.1093/nar/gkh340>
9. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., Drummond, A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 2012, 28, 1647-1649. <https://doi.org/10.1093/bioinformatics/bts>
10. Haddad, N., Adjlane, N., Loucif-Ayad, W., Dash, A., Rajashekar, B., Al-Nakeeb, K., Sicheritz-Ponten, T., Mitochondrial genome of the North African Sahara Honeybee, *Apis mellifera sahariensis* (Hymenoptera: Apidae). *Mitochondrial DNA Part B*, 2017, 2, 548-549. <https://doi.org/10.1080/23802359.2017.1365647>
11. Gibson, J. D., Hunt, G. J., The complete mitochondrial genome of the invasive Africanized honey bee, *Apis mellifera scutellata* (Insecta: Hymenoptera: Apidae). *Mitochondrial DNA Part A*, 2016, 27, 561-562. <https://doi.org/10.3109/19401736.2014.905858>
12. Haddad, N. J., Mitochondrial genome of the Levant Region honeybee, *Apis mellifera syriaca* (Hymenoptera: Apidae). *Mitochondrial DNA Part A*, 2016, 27, 4067-4068. <https://doi.org/10.3109/19401736.2014.1003846>
13. Eimanifar, A., Kimball, R. T., Braun, E. L., Fuchs, S., Grunewald, B., Ellis, J. D., The complete mitochondrial genome and phylogenetic placement of *Apis nigrocincta* Smith (Insecta: Hymenoptera: Apidae), an Asian, cavity-nesting honey bee. *Mitochondrial DNA Part B*, 2017b, 2, 249-250. <https://doi.org/10.1080/23802359.2017.1318683>
14. Tan, H. W., Liu, G. H., Dong, X., Lin, R. Q., Song, H. Q., Huang, S. Y., Yuan, Z.G., Zhao, G.H., Zhu, X. Q., The complete mitochondrial genome of the Asiatic cavity-nesting honeybee *Apis cerana* (Hymenoptera: Apidae). *Plos one*, 2011, 6, e23008. <https://doi.org/10.1371/journal.pone.0023008>
15. Wakamiya, T., Tingek, S., Okuyama, H., Kiyoshi, T., Takahashi, J. I., The complete mitochondrial genome of the cavity-nesting honeybee, *Apis koschevnikovi* (Insecta: Hymenoptera: Apidae). *Mitochondrial DNA Part B*, 2017, 2, 24-25. <https://doi.org/10.1080/23802359.2016.1275847>
16. Wang, A. R., Kim, M. J., Park, J. S., Choi, Y. S., Thapa, R., Lee, K. Y., Kim, I., Complete mitochondrial genome of the dwarf honeybee, *Apis florea*

(Hymenoptera: Apidae). Mitochondrial DNA, 2013, 24, 208-210. <https://doi.org/10.3109/19401736.2012.744986>  
17. Arias, M.C., Sheppard, W.S. Phylogenetic relationships of honey bees (Hymenoptera: Apinae:

Apini) inferred from nuclear and mitochondrial DNA sequence data. Molecular Phylogenetic and Evolution, 2005, 37, 25-35. <https://doi.org/10.1016/j.ympev.2005.02.017>