*Tested Studies for Laboratory Teaching*
*Proceedings of the Association for Biology Laboratory Education*
Volume 39, Article 24, 2018

# Teaching Phylogeny and Direction of Viral Transmission using a Real HIV Criminal Case

**Kuei-Chiu Chen and Dalia Zakaria**

Weill Cornell Medicine-Qatar, Department of Premedical Education, Education City, Qatar Foundation, P.O. Box 24144, Doha, Qatar
**(kuc2005@qatar-med.cornell.edu; dez2003@qatar-med.cornell.edu)**

Teaching phylogenetics often faces the challenge in making this topic interesting and relevant. This exercise uses a real criminal case described in the original PNAS publication about how a physician in Louisiana deliberately injected his ex-girlfriend with HIV from a patient under his care. During teaching sequences of HIV genes obtained from the patient, the ex-girlfriend, and HIV-positive individuals not related to the case are analyzed using a free version of CLC Main Workbench software to analyze phylogeny of the viral sequences. The results not only indicated the close relationship in the viral gene sequences obtained from the ex-girlfriend and the patient, it also suggested the direction of transmission by using the concept of basal, derived, sister taxa, monophyly and paraphyly in phylogeny interpretation. To present additional relevance, instructors may mention that this case has established a precedent in using phylogenetics to convict a person in the US criminal court.

**Keywords**: molecular phylogeny, HIV, monophyly, paraphyly, criminal case, case study teaching, reverse transcriptase, *pol* gene, *env* gene

## Introduction

Rarely does phylogenetic analysis strikes students as interesting or relevant. Even with the use of molecular data and the increasing importance in applied science such as conservation, it is often viewed as static. For instructors who are passionate about phylogenetics, finding an exciting example for use in teaching is often challenging. Inspired by a recent case study teaching publication on an HIV transmission legal case (Cheeptham and Shuster 2013) and the original publication from authors who handled the samples from the individuals involved (Metzker et al. 2002), this exercise brings viral genetics and molecular phylogenetics into US legal system. Using a story-telling format, the activity gradually unfolds a criminal case that employs molecular biology, viral genetics and finally the use of phylogenetics to suggest the direction of viral transmission. This teaching activity culminates in the conviction of the defendant, a physician who has been determined by the court that he deliberately infected his ex-girlfriend with HIV obtained from his HIV-positive patient and injected the virus to his ex-girlfriend in the guise of routine vitamin B12 supplement. This seminal legal case and the science that supported the conviction of the defendant have established the precedent of using phylogenetics in the criminal justice. Since then more legal cases have introduced phylogeny as evidence. As a result, a new application of phylogenetic analysis, forensic phylogenetics, has been established. With a real-world application and the intrinsically exciting nature of crime stories, the value of phylogenetic analysis may be more accepted and enjoyed by students.

### Planning of Activities

The entirety flow of activities usually requires one and a half to two hours, covering the following topics

- Story background and circumstantial evidence
- Reconstruction of molecular phylogeny using parsimony
- Bootstrap character resampling method
- CLC Main Workbench phylogeny reconstruction using real case sequences
- Discussion of two phylogenetic trees using distance data from the original publication by Metzker and coauthors (2002)

- Final conclusion of the court conviction and establishment of legal precedent of using phylogenetics in criminal justice

The story begins with the argument between a Lafayette, Louisiana based gastroenterologist and his girlfriend, to the injection of vitamin B12 by the physician to the girlfriend who had been HIV-negative all along to her being tested positive for HIV within a span of five months after the injection. With circumstantial evidence, the prosecutor built a case against the defendant that he deliberately infected the victim with the HIV obtained from an AIDS patient under his care. Students then follow the instructions of how to reconstruct a phylogeny using parsimony to the use of distance method and bootstrap resampling of characters that demonstrate the level of support for each internal branch of the tree. Students later use a free version of CLC Main Workbench (2017) to reconstruct and interpret the phylogeny produced using Neighbor-Joining method from sequences of reverse transcriptase gene: two sequences from the victim (the girlfriend), six from the patient under the care of the physician (the defendant), and eight sequences isolated from HIV-positive individuals in the Lafayette area. The final discussion involves using the two phylogenetic trees from the original publication by Metzker et al. (2002). In this discussion, the first tree, based on the sequences from the fast-mutating membrane protein gene *env*, shows the clustering of sequences from the victim as sister taxa to those from the patient under the physician's care. This suggests the closer relationship between the two sets of sequences than either group to the local control sequences. Despite the close relationships between the groups of sequences, no suggestion could be made with respect to the direction of viral transmission. The critical information came from the phylogeny based on the sequences of reverse transcriptase gene *pol*, a gene with slightly slower mutation rate. In the phylogenetic tree, the sequences obtained from the victim form a component of the monophyletic trees made up of the gene sequences collected from the patient's HIV samples. In other words, if we exclude the sequences from the victim, the phylogenetic tree will be paraphyletic, suggesting the most recent common ancestor of the victim's sequences originated from some of the HIV sequences found in the patient. This observation suggests that the viral sequences in the victim originated from the physician's patient.

# Student Outline
# Revealing the Direction of HIV Transmission:
# A Case Study of Molecular Evolution and Phylogeny

## Synopsis

This activity is adapted from a case study teaching and its original research paper that describes a real criminal case of a woman infected with HIV through injection by her ex-boyfriend. Through this activity you will learn the basic biology of HIV and how to search for nucleotide sequences that are similar to one another using online databases. In addition, you will use downloadable software to reconstruct and assess phylogenetic trees using the viral gene sequences from this case. Finally, by analyzing the branching patterns in the phylogenetic trees you are able to determine the direction of viral transmission between the individuals in this case. The worksheet at the end of this chapter should be completed and turned in by the due date announced by your instructor.

### Laboratory Objectives

*Conceptual*

At the end of this laboratory and the associated laboratory lecture you should be able to…

1. Describe the basic life history of the human immunodeficiency virus (HIV).

2. Describe what deficiency in HIV that causes high variation in nucleotide sequence of genes of the virus, for example, reverse transcriptase.

3. Describe the difficulty in treating HIV infections by connecting the concepts learned from Immunology lab and the molecular biology of the virus.

4. Reconstruct a phylogenetic tree based on parsimony principle using provided character dataset and assumptions and interpret the results.

5. Understand the principles of bootstrap resampling and interpret the results.

6. Interpret basal and derived characters on a phylogenetic tree and determine the direction of HIV transmission based on the branching pattern of phylogenetic tree based on viral sequences.

*Procedural*

At the end of this laboratory and the associated laboratory lecture you should be able to…

1. Use CLC Main Workbench to reconstruct phylogenies and conduct bootstrap resampling.

2. Conduct BLAST search for sequences and molecular phylogeny using sequence accession numbers.

## Part I – Story Background and Circumstantial Evidence

In 1995, a nurse accused her married gastroenterologist ex-boyfriend of deliberately injecting her with HIV infected blood. The story began in July 1994 when the nurse broke off her relationship with the American physician, Dr. Richard Schmidt. About a month later, he injected his ex-girlfriend with what he claimed to be vitamin B-12. The nurse used to have B-12 injections in the past, so this injection was taken as routine. A few months later (January 1995) she tested Positive for HIV and suspected that the vitamin B-12 injection she received from Dr. Schmidt was infected with HIV. The nurse had several HIV tests prior to that time due to the nature of her job and all of them tested negative. Consequently, Dr. Schmidt was accused of a second-degree murder and was brought to trial.

*Questions*

1. If you were a member of Dr. Schmidt's defense team, what other ways you may think of (other than the vitamin B-12 injection) which could be the source of the nurse's infection with HIV?

2. What kinds of tests or information could be required to deny Dr. Schmidt's defense's alternative hypothetical scenarios?

Other possible sources of infection were also considered such as the nature of Dr. Schmidt's former girlfriend's profession as a nurse as well as her previous sexual relationships. However, the nurse was previously routinely tested several times and none of her employment records showed any accidental or occupational exposures or any reported needle sticks at work. Furthermore, all of her seven ex-boyfriends, including Dr. Schmidt tested HIV negative.

During the investigation process, a paper work of blood withdrawn from one of Dr. Schmidt's HIV-positive patients at his office on August 4, 1994 appeared to have been deliberately hidden in a storage room with other older records.

To support the detected circumstantial evidences against Dr. Schmidt, the HIV reverse transcriptase (RT) sequences obtained from the victim (the nurse) and from Dr. Schmidt's HIV positive patient were analyzed. Because of the high mutation rate of reverse transcriptase *(pol)* gene, it could be used as a tool to investigate the relationships between different strains of HIV. Phylogenetic analysis was performed to determine whether Dr. Schmidt's HIV positive patient is the source of the nurse's HIV infection or not.

*Questions*

3 HIV-1 mutates rapidly. Based on this, would you expect to find:

a. A single, identical RT sequence in the victim and the patient
b. A single RT sequence in the victim, which is non-identical to the RT sequence in the patient
c. A set of RT related sequences that share a common ancestor in either of the victim or the patient

Explain your answer.

4. If the physician's HIV positive patient is the source of infection according to the nurse's allegation, would you expect the RT sequences in both the victim and patient to be:

a. Identical
b. Related
c. Completely different

Explain your answer.

With the circumstantial evidence, the prosecutor was able to argue for admitting evidence based on HIV gene sequences collected from the nurse (Victim) and from Dr. Schmidt's HIV-positive patient (the Patient herein). This is based on established reputation on the reliability and repeatability of a set of molecular genetics techniques, including genomic DNA isolation, polymerase chain reaction, DNA sequencing and phylogenetic analysis of DNA sequences.

*Questions*

5. Is there a need to include control gene sequences if you are comparing the HIV gene sequences obtained from the Patient and the Victim? If so, what type of HIV sources should be considered?

## Part II – Phylogenetic Analysis Using CLC Main Workbench

The following exercise is a phylogenetic analysis based on of the reverse transcriptase gene (*pol*) sequences from the Victim, the Patient, and local controls. The local controls are sequences collected from HIV-positive individuals based in Lafayette, Louisiana area and these individuals are not connected to either the Patient or the Victim. The numbers of sequences included are two from the victim, six from the patient, and eight from local controls, two sequences of which obtained from the same HIV-positive individual.

1. Download the most recent CLC Main Workbench trial version. Use Limited Mode features if your version has expired.

2. Download the zip folder of HIV sequences from the course management system Canvas.

3. Launch CLC Main Workbench. Open limited mode if applicable.

4. On top of the panel click New, and then choose Folder. Name your folder with a relevant title.

5. Click Import and locate your HIV sequences folder. Click Next. The 16 sequences will show up within the folder.

6. Highlight all 16 sequences. On the command bar locate Toolbox and choose Alignments and Tree and then Alignments.

7. You will see the sequences in the window that says "Selected elements (16)". Click Next.

8. In the next window choose **Very accurate (slow)** for Alignment section. Leave other settings unchanged. Click Next.

9. In the next window on Result handling choose Open, and then click Finish.

10. You will see sequences aligned with the nucleotides color-coded for easy viewing.

Notice that most of the nucleotides aligned well, with gaps in samples labeled with LA followed by a number. These samples were local controls, which were collected from HIV-positive individuals in Lafayette, Louisiana but are not connected to the victim or the patient.

Also notice a few nucleotide substitutions of a particular position. You may occasionally see a nucleotide that is not A, T, C, or G. When the instrument was determining the nucleotide at a particular position, it may not always be certain which of the four nucleotides there was at that position. If it was completely uncertain, the nucleotide will be noted as N. If the instrument determines that the nucleotide was either an A or G (purine), then it was coded as R, which stands for pu**R**ine. For C or T, the code would be Y, representing p**Y**rimidine. There are also a few other codes for the remaining combinations of nucleotides but we will not discuss them here.

11. Now save the alignment by Choosing File → Save to save the alignment in the same sequence folder.

12. You will see an alignment file show up in the list. Highlight it.

13. Go to the Toolbox again and choose Alignments and Trees and choose Create Trees. The Alignment file should be in the window under the Selected elements (1) heading. Click Next.

14. Keep the setting of the next window unchanged. In the following window asking for Results handling, choose Open, and then click Finish. You will see a phylogram as the default tree layout.

15. At the right-hand side there is a Tree Settings panel that allows you to make changes in the output of the tree. You may change the Tree layout for cladogram, circular cladogram, circular phylogram, or radial tree. In addition, you should explore other settings at the bottom of this panel to see how the trees may show up differently. Those options are available for view if you click the small rectangle to the right of the name of the option to expand the list. Finally, make sure you explore bootstrap options. Choose to show the bootstrap values. What is the bootstrap value for the branch that unites the sequences from the victim and the patient?

*Questions*

1. After aligning the RT sequences obtained from the Victim, the Patient and the local controls, what did you learn from the phylogenetic tree that you have generated using CLC workbench? Specifically focusing on a) if the two sequences from the Victim are mostly closely related; b) are the two sequences from the Victim more basal or more derived from the Patient's sequences; c) are either the Patient's or Victim's sequences more closely related to one another or to those from the local controls.

2. Using the phylogenetic tree that you have generated and the additional information provided by your instructor, are you able to infer the more likely direction of the HIV transmission, from the patient to the victim or vice versa? Why?

**Part III. Phylogenetic Analysis from the Original Publication**

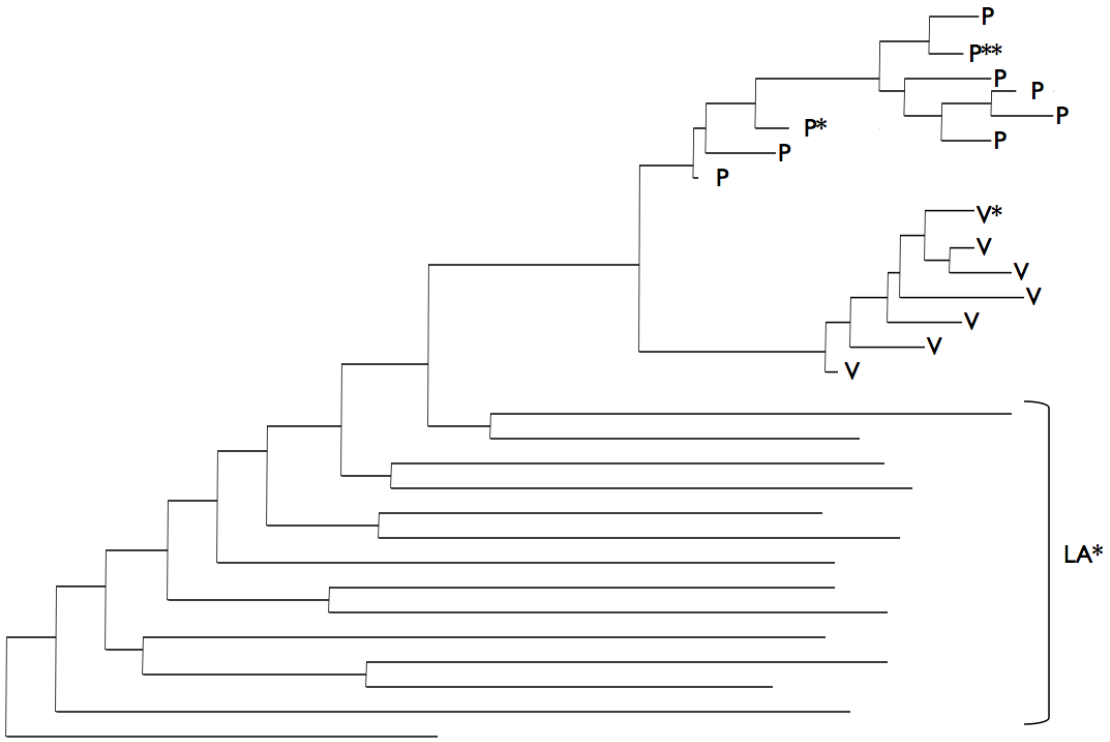



**Figure 1.** The phylogenetic tree based on the sequences of the envelope gene (*env*) of the patient (P), victim (V) and the local controls (LA) adapted from the original Metzker et al. (2002) paper. P* represents 19 closely clustered patient sequences; P** represents another 25 closely clustered patient sequences; V* represents 45 closely clustered victim sequences and LA* is a schematic representation of the 26 sequences from the local controls. The bootstrap value that joined the patient and the victim sequences was 100%.

*Questions*

1. What general relationship among the sequences of the victim, patient and local controls is shown in the tree in Figure 1?

2. Are you able to infer the direction of viral transmission from the tree in Figure 1?

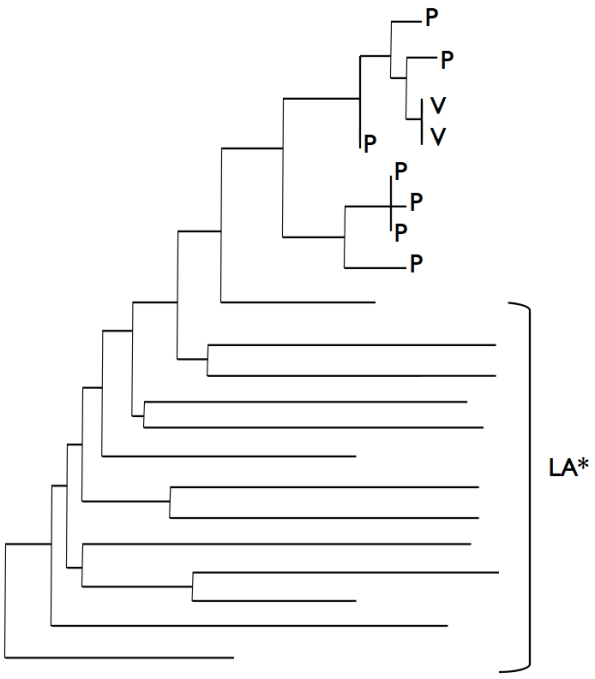



**Figure 2**. The phylogenetic tree adapted from Metzker et al. (2002) based on the sequences from the RT gene. P and V represent patient and victim respectively. LA* represents a schematic topology of local controls. 100% of bootstrap replicates place victim sequences with those of the patient.

*Questions*

3. Are you able to determine the direction of viral transmission based on the branching pattern between the patient sequences and the victim sequences? Explain

4. Given the circumstantial evidence and the phylogenetic evidence, what do you think the verdict was in this case?

## Cited References

Cheeptham N, Regassa L, Shuster M. 2013. Murder by HIV? Undergraduate edition. National Center for Case Study Teaching in Science. Retrieved from http://sciencecases.lib.buffalo.edu/cs/collection/detail.asp?case_id=673&id=673

Michael L. Metzker, David P. Mindell, Xiao-Mei Liu, Roger G. Ptak, Richard A. Gibbs, David M. Hillis. 2002. Molecular evidence of HIV-1 transmission in a criminal case. Proceedings of the National Academy of Sciences of the United States of America 99(22):14292-7.

## MOLECULAR PHYLOGENY WORKSHEET

1. The data matrix below was obtained from a partial gene sequence of six species and an outgroup.

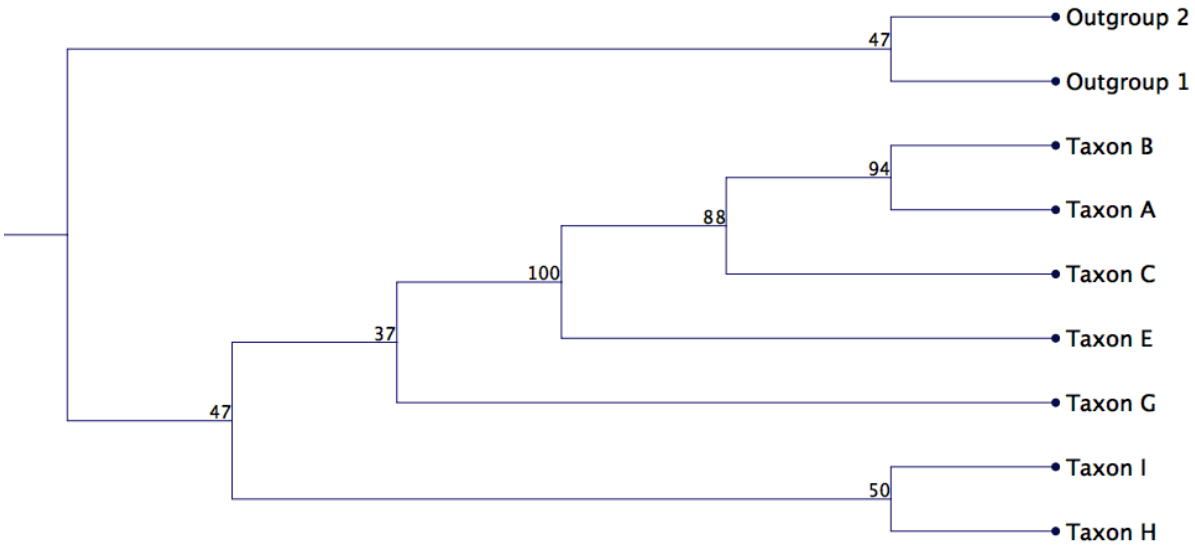
| | NUCLEOTIDE POSITION | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TAXON | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| A | A | T | C | A | T | T | G | A | C | C | G | A |
| B | G | T | A | C | C | G | A | G | C | T | G | A |
| C | A | T | C | A | T | T | G | A | C | C | G | A |
| D | G | T | A | C | C | G | A | G | C | A | G | A |
| E | T | T | C | A | T | T | A | A | C | C | G | A |
| F | T | T | A | A | T | C | A | A | C | C | G | A |
| Outgroup | T | C | A | G | C | C | A | G | T | T | A | A |

a) Following the principle of using **shared derived characters**, lay out the complete nested sets in the space below. For instance, you should provide an answer similar to the following example:

$$[(C(A, B)_{7,9,11})_{1,3}\ (F(D,E)_{8,9})_{12}]_{2,4,5,6,10}$$

b) Reconstruct the phylogeny and present it as a cladogram, including changes in characters and character states, in the space below.

2. The cladogram below shows the bootstrap values as percentage of support for the corresponding branches.




Draw a modified cladogram with polytomy if you set the bootstrap threshold at 60% for the tree above. (Practice by using CLC Main Workbench to see how the setting of bootstrap threshold changes the topology of the trees.)

## Materials

Students should visit CLC Main Workbench and download a trial version of the software https://www.qiagenbioinformatics.com/products/clc-main-workbench/. After the trial period expires the Limited Mode is still sufficient for the purpose of this exercise. The sequences are accessible from NCBI with the accession numbers below and CLC Main Workbench can import the folder in which the sequences are stored:

AY156806 (victim sequence 1)
AY156807 (victim sequence 2)
AY156797 (patient sequence 1)
AY156800 (patient sequence 2)
AY156801 (patient sequence 3)
AY156802 (patient sequence 4)
AY156803 (patient sequence 5)
AY156788 (local control 25)
AY156789 (local control 26)
AY156790 (local control 27)
AY156791 (local control 28)
AY156792 (local control 29)
AY156787 (local control 26-2)
AY156794 (local control 31)
AY156795 (local control 32)

## Notes for the Instructor

In the actual exercise the authors use a series of slides to guide students how to reconstruct the phylogeny using the parsimony approach. This is reflected in some of the questions in the worksheet (see Student Outline). To discuss further on the different resolution of the phylogeny determined by the ENV gene and the RT gene it may be worthwhile to discuss the higher mutation rate in the ENV gene compared to the RT gene. The accumulated mutations in ENV sequences although allowing the determination of sister taxa relationship between the sequences from the victim and the patient, it was not able to determine the direction of viral transmission. If the samples were to have obtained much later the phylogenetic information would have lost due to the rapid mutation in the viral genes. To compare the magnitude of mutation rate difference, it is worth mentioning that the intrahost variation of either gene sequence is higher in the patient than in the victim. This is consistent with a longer infection history in the patient than in the victim.

To close the discussion of the criminal case it may be worthwhile to read a short except from the original publication (Metzker et al., 2002), p.14297 that describes how this case has established the use of phylogenetic analysis to support or reject criminal viral transmission cases. Before closing the story quite often students are curious what has become of the victim, therefore any updates that maybe obtained by the instructor will always be of interest to students as a closing of the story.

## Cited References

Cheeptham N, Regassa L, Shuster M. 2013. Murder by HIV? Undergraduate edition. National Center for Case Study Teaching in Science. Retrieved from http://sciencecases.lib.buffalo.edu/cs/collection/detail.asp?case_id=673&id=673

Michael L. Metzker, David P. Mindell, Xiao-Mei Liu, Roger G. Ptak, Richard A. Gibbs, David M. Hillis. 2002. Molecular evidence of HIV-1 transmission in a criminal case. Proceedings of the National Academy of Sciences of the United States of America 99(22):14292-7.

## Acknowledgments

The authors would like to thank the Class 2021 and 2022 for their valuable feedback at WCM-Q.

## About the Authors

Kuei-Chiu Chen is an Associate Professor at Weill Cornell Medicine-Qatar (WCM-Q), where she teaches premedical students biology since 2012. She also leads research projects involving premedical students on biodiversity, population genetics and conservation using molecular markers. She has received The Best Research Mentor Award and Excellence in Teaching Award in her current post.

Having a Ph.D. in microbiology and postdoctoral experience in immunology, Dalia Zakaria joined WCM-Q in 2016 as a Teaching Specialist in the Premedical Education Department where she teaches Introductory Biology and Immunology. She has been participating in student-centered research projects most recently focusing on human microbiota and molecular taxonomy. She has recently received WCM-Q Excellence in Teaching Award.

## Mission, Review Process & Disclaimer

The Association for Biology Laboratory Education (ABLE) was founded in 1979 to promote information exchange among university and college educators actively concerned with teaching biology in a laboratory setting. The focus of ABLE is to improve the undergraduate biology laboratory experience by promoting the development and dissemination of interesting, innovative, and reliable laboratory exercises. For more information about ABLE, please visit **http://www.ableweb.org/.**

Papers published in *Tested Studies for Laboratory Teaching: Peer-Reviewed Proceedings of the Conference of the Association for Biology Laboratory Education* are evaluated and selected by a committee prior to presentation at the conference, peer-reviewed by participants at the conference, and edited by members of the ABLE Editorial Board.

## Citing This Article

Chen K-C, Zakaria D. 2018. Teaching phylogeny and direction of viral transmission using a real HIV criminal case. Article 24 In: McMahon K, editor. Tested studies for laboratory teaching. Volume 39. Proceedings of the 39th Conference of the Association for Biology Laboratory Education (ABLE). **http://www.ableweb.org/volumes/vol-39/?art=24**