

An Emerging Amphibian Infection as a Model for Teaching Phylogenetic Reconstruction

• AMANDA L. J. DUFFUS

ABSTRACT

Phylogenetic analysis and interpretation can be challenging for many students, but emerging infections can provide a rich tapestry for addressing these topics while maintaining student interest. Ranaviruses are a group of emerging infections in amphibians that have been associated with morbidity and mortality events around the globe. They have also been implicated in population declines and local extirpations of some amphibian species. Many ranaviruses have been subject to intense study by scientists as they seek to understand the impacts of these viruses on a variety of ectothermic animals. A large amount of sequence data is available on GenBank and is easily accessible for students to use to study phylogenetic relationships between different viral species, strains, and isolates. This article examines the general process of obtaining sequence data, sequence alignments, and tree building by using databases, servers, and computer programs that are freely available to all high school and undergraduate students and their instructors. Providing students with a guided framework for exploring their own questions with respect to the evolutionary relationships of ranaviruses can produce some very unique and thought-provoking results.

Keywords: Emerging infection; amphibian; Ranavirus; phylogenetic reconstruction; evolutionary relationships.

○ Introduction

Students are typically aware that amphibians are in decline globally (Stuart et al., 2004) and that a disease called chytridiomycosis, caused by the pathogenic fungus *Batrachochytrium dendrobatidis*, has been associated with many of the declines and extinction events. However, they are surprised to learn that other infectious agents, such as the ranaviruses (*Ranavirus* spp.), have also been implicated. For over a decade, I have been working with ranaviruses, a globally distributed group of emerging pathogens that infect ectothermic vertebrates, including several endangered species (Duffus et al., 2015). Ranaviruses are a great model system for studying phylogenetics because they can yield some interesting and unique results. The emergence of *Ranavirus* infections and the resultant disease has not

only caused population declines in some areas (Teacher et al., 2010), but has resulted in the entire collapse of some amphibian communities (Price et al., 2014).

Studying phylogenetics can be challenging for many students as they struggle to comprehend how sequence data are transformed into a tree and how the results are then interpreted. Emerging infectious diseases like Ebola virus disease, Middle East respiratory syndrome (MERS), and ranaviral diseases are a great way to get students interested in a variety of evolutionary questions, including how to reconstruct phylogenetic relationships. One of the most important roles that phylogenetic reconstruction may play in the context of emerging diseases is predicting whether a species, strain, or isolate is going to be virulent. If a new strain or isolate groups closely with known highly virulent strains, there is a good chance that it too might be virulent, and if it appears in a new area, it may be possible to take steps to limit or mitigate pathogen emergence.

For ranaviruses, there are a large number of partial and complete gene sequences and completely sequenced genomes available on GenBank (<http://www.ncbi.nlm.nih.gov>; Benson et al., 2013). Currently, 26 core genes have been identified in the *Ranavirus* genome (Eaton et al., 2007) and almost all of these genes can provide some fascinating and quickly generated results for students studying the evolutionary relationships between different *Ranavirus* species, and sometimes even between different *Ranavirus* strains or isolates. *Frog virus 3* (FV3), the type virus of the ranaviruses (Tan et al., 2004), is a great place to start when looking for open reading frames (ORFs). Eaton et al. (2007: tables 6 and 7) describe all 26 core genes and their corresponding ORFs in more than 10 ranaviruses. In GenBank, there are a plethora of gene sequences for the *Ambystoma tigrinum* virus (ATV), the common midwife toad virus (CMTV), FV3, and non-amphibian ranaviruses (e.g., European sheatfish iridovirus). The most commonly used gene to reconstruct *Ranavirus* phylogenetic relationships is the major capsid protein (MCP; ORF 90R in FV3). The MCP is a great place to start; however, it is not the only notable ORF. The students can potentially explore a huge variety of interesting phylogenetic questions, some examples of which can be found below:

○ Sample Questions

1. What are the relationships of different ranaviruses based on the sequence of their MCP genes?
2. Based on the MCP gene sequence, how closely related are *Ranavirus* isolates from fishes and frogs?
3. Is local adaptation seen between different strains of *Ambystoma tigrinum* virus isolated from different regions of the western United States if we use the immediate early protein ICP-4 (ORF13L in FV3)?

○ Student Learning Objectives

1. Obtain data from GenBank by using information from the literature (e.g., Eaton et al., 2007) and format this data in a manner that can be used for further analysis.
2. Use nucleotide-nucleotide BLAST (nBLAST) searches to create a data set for phylogenetic analysis.
3. Manipulate sequence data and ensure that it is in the proper format to be used in other programs (e.g., MAFFT and MEGA).
4. Use multiple free servers and/or programs for the manipulation of sequence data and the production of potentially novel phylogenetic trees.
5. Use MEGA to produce phylogenetic trees and to accurately describe the results.
6. Interpret sometimes confusing and unexpected phylogenetic trees.

○ Data Collection Methods

Obtaining Sequence Data

The most efficient way to collect data to analyze is to use the sequence of the ORF of interest in FASTA format obtained directly from GenBank and to do an nBLAST search using that sequence.

1. To obtain the sequence data, find the ORF that you are looking for; then click the link “gene” beside it.
2. This will bring up a window at the bottom of the screen; click the FASTA link beside it to bring up the FASTA formatted sequence in a new window.

Comparing Sequences

The Basic Local Alignment Search Tool (BLAST) is a web-based search engine that compares sequences and comes up with similar sequences (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). To perform a nucleotide-nucleotide BLAST, simply go to the website above and paste your sequence into the box and then click on the “BLAST” button on the bottom of the page. There is no need to change any of the settings before BLASTing your sequence. The results of your nBLAST search will have three sections: a graphic summary, descriptions, and alignments. They will all be open.

1. Scroll down to the “Description” section; this is where you can extract the data for your alignment. Depending on the

- question your students are investigating, different sequences can be selected using the “tick boxes” on the righthand side.
2. After you have selected all the sequences you are interested in working with, click on the “Download” link and a submenu will pop up; select “FASTA (aligned sequences)” and continue.
3. You will be asked to save or open the file “seqdump.txt”. Open the file in Notepad (or another text editor).
4. The next step is to organize the data in Notepad. When the data are “dumped” into the file, the result is a continuous line of text.
5. The individual sequences need to be set on their own start lines, with a line of space in between, so that they can easily be read by the alignment software.
6. Make sure that the different sequences have easily identifiable names (e.g., if you have two strains of FV3, use “FV3-1” and “FV3-2” or something similar to easily differentiate between the two). Now save the file so that the sequences can be analyzed further.

Alignment

One of the best pieces of alignment software is MAFFT (Katoah et al., 2017), which can either be downloaded to a computer or used on a server. The server that I recommend using is at <https://mafft.cbrc.jp/alignment/server/>. The interface is easy to use, and the results can be downloaded in multiple formats. Since the data set has been saved as a plain text file (.txt), it can be uploaded directly to the server.

1. Once the data set has been uploaded, click the “Submit” button just above the “Advanced Settings” heading. There is no need to adjust any of the settings or use the “Advanced Settings” further down the page to get usable results.
2. Once the server has computed the alignment, you will get a summary screen. The results can then be downloaded in several different formats and a tree can be visualized (see step 1 in the next section below).

Moving Data

It is important to teach students how to move data between different programs. The free program MEGA (Molecular Evolutionary Genetic Analysis; <http://www.megasoftware.net/>) is an excellent starting point for the estimation of nucleotide substitution models and tree visualization. Currently, if you use a Windows-based computer, MEGA 7 (Kumar et al., 2016) is available only for 64-bit versions of the operating system. Therefore, if you use an older operating system, I recommend that you download MEGA 6 (Tamura et al., 2013). To import the aligned sequences into MEGA, it is easiest if you use FASTA formatted data.

1. To obtain the FASTA formatted data, select the FASTA format link at the top of the MAFFT summary page. This will open a new tab in your web browser.
2. To select all the data, right click and then choose “Select All” in the pop-up menu.
3. Copy and paste the data into a new Notepad file.
4. This file can then be uploaded into MEGA 6 and the format changed into .meg for further analysis.

- To open the text file in MEGA, simply go to the “File” tab at the top right of the page, select “Open a File/Session,” and navigate to the Notepad file that you previously saved. This opens a window within MEGA 6 with the name of your Notepad file as the title.
- To transform the FASTA formatted text file data into .meg format, select the “Utilities” drop-down menu on the upper menu bar.
- Select “Convert to MEGA format.” A pop-up box in the middle of the screen will prompt you to name and save the file. When asked for the format of the data, make sure that you choose FASTA from the drop-down menu in the pop-up. Once you click “Save,” another box will appear letting you name and save the file as a .meg file.
- Now you must close the window with the converted data.
- To open the data in MEGA 6 so that you can work with it, go to the File menu, select “Open a File/Session,” and choose the file that you just saved as a .meg file. You will be prompted in a pop-up window to choose the type of data; select “Nucleotide Sequence” (it is the first choice) and then click on the “OK” button.
- Another pop-up will appear asking if you are looking at protein-encoding nucleotides; select “yes.”
- Now your MEGA6 window should have a box in the upper righthand corner that has a “T” and an “A” in it.

Choosing the Model

The next step in building a good phylogenetic tree is determining the appropriate model of nucleotide substitution that should be applied to the data. Follow the steps below to do this in MEGA 6.

- Go to the “Analysis” tab of the top menu and click on it to reveal a drop-down menu.
- Scroll over the first option in the list “Models” and a sub-menu will appear.
- In that submenu, select the first option: “Find the Best DNA/Protein Models (ML).” A new pop-up will appear in the middle of the screen.
- Leave the settings as they are and then select the “Compute” option.
- A pop-up with a progress bar will appear as the different nucleotide substitution models are tested for the data.
- After completion, a pop-up box with a table containing the results of the test will appear.
- Typically, the best-fitting model will be the first in the table, but always check to make sure that it is.

For a detailed description of how to build phylogenetic trees in MEGA 5, please see Newman et al. (2016). Although the versions are different, the main concepts of tree building in the program remain the same.

○ Assessment Strategies

There are several different ways that student learning can be assessed after the process has been completed. Typically, I either have students present their work in a scientific poster format or run a

mini-conference in lab or class where each student or group gives a 10–15 minute presentation. The formats of both the poster and the presentation are typical of scientific conferences, which include Introduction, Methods, Results, and Discussion sections. Presentations like this are relatively easy to evaluate. There are a plethora of rubrics available online that can be modified for your specific situation (many scientific societies have these on their websites).

Sample Assessment Questions

To ensure that students are understanding the basics of phylogenetic reconstruction and not just following the instructions blindly, questions along the following lines may be asked:

- What kind of information is available in GenBank?
- Why must the sequences be aligned before you make a phylogenetic tree from them?
- Why should you run a test to determine the best nucleotide substitution model for your data set before you use it to build a tree?
- What are the assumptions of the tree-building algorithm you used?
- What is the purpose of the bootstrap values and what do they mean?
- What isolates/strains are most closely related in your tree? Was this what you expected? Why or why not?

○ Sample Project

In our sample project, let’s examine the relationship between nine *Ranavirus* species (Table 1). Following the steps outlined above, I created a Notepad file and organized it into a readable FASTA format (Figure 1). I uploaded the file into MAFFT and the sequences aligned (Figure 2). The aligned sequences were copied and pasted into a new Notepad document, saved and uploaded into MEGA 6. The aligned sequences were converted to MEGA format (Figure 3) and saved as a .meg file. I then opened the .meg file in MEGA and ran the test for the best-fitting model for nucleotide substitution. The best fit was T92+I, which means that the Tamura 3- parameter model with evolutionary invariability is the model that needs to be chosen when building the phylogenetic tree (i.e., no gamma distribution should be applied). For the neighbor-joining and minimum evolution trees, it is not possible to select for evolutionary invariability at different sites, so the default setting (d: transitions and transversions) was used in the substitutions to include section. Next, three trees – a maximum-likelihood tree, a neighbor-joining tree, and a minimum evolution tree – were constructed (Figures 4, 5, and 6). In all three trees, consensus trees were made by eliminating all branches with < 50% support and each branch was bootstrapped 1000x. The branching pattern for all three trees is extremely similar. However, there are some notable differences in sister taxa and in the support of some of the branches. A summary of similarities and differences can be found in Table 2.

○ Challenges & Outputs

The biggest challenge with this project is getting students to ask questions that are not overwhelming. Since a large amount of

Table 1. Ranavirus isolates that were used to create the phylogenetic tree in the sample project, with abbreviations and GenBank accession numbers. The gene in question is the major capsid protein (MCP; ORF 90R in frog virus 3).

Name	Abbreviation	Accession No.
<i>Ambystoma tigrinum</i> virus	ATV-1	NC_005832.1
<i>Ambystoma tigrinum</i> virus	ATV-2	KR075877.1
<i>Andrias davidianus</i> ranavirus	CGSV-1	KF033124.1
<i>Andrias davidianus</i> ranavirus	CGSV-2	KC865735.1
Common midwife toad virus	CMTV-1	JQ231222.1
Common midwife toad virus	CMTV-2	KP056312.1
Epizootic haematopoietic necrosis virus	EHNV-1	FJ433873.1
Epizootic haematopoietic necrosis virus	EHNV-2	NC_028461.1
European catfish virus	ECV-1	KT989884.1
European catfish virus	ECV-2	KT989885.1
European sheatfish virus	ESV-1	NC_017940.1
European sheatfish virus	ESV-2	JQ724856.1
Frog virus 3	FV3-1	NC_005946.1
Frog virus 3	FV3-2	AY548484.1
Singapore grouper iridovirus	SGIV-1	NC_006549.1
Singapore grouper iridovirus	SGIV-2	AY521625.1



Figure 1. An example of how the Notepad file should be set up for import into the MAFFT server.

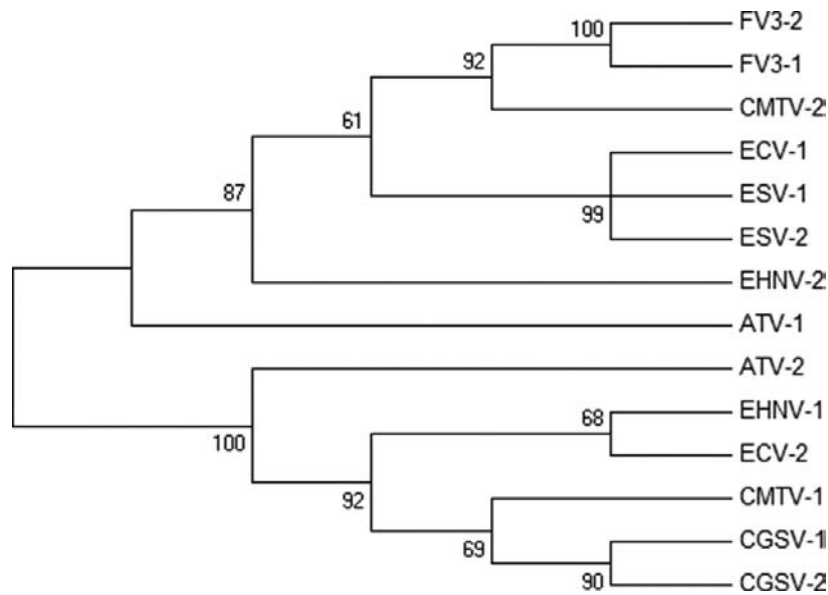


Figure 4. Maximum-likelihood tree created in MEGA 6. The tree was built using the Tamura 3-parameter model of nucleotide substitutions with transversions and transitions. Bootstrap values were calculated 1000x for each branch.

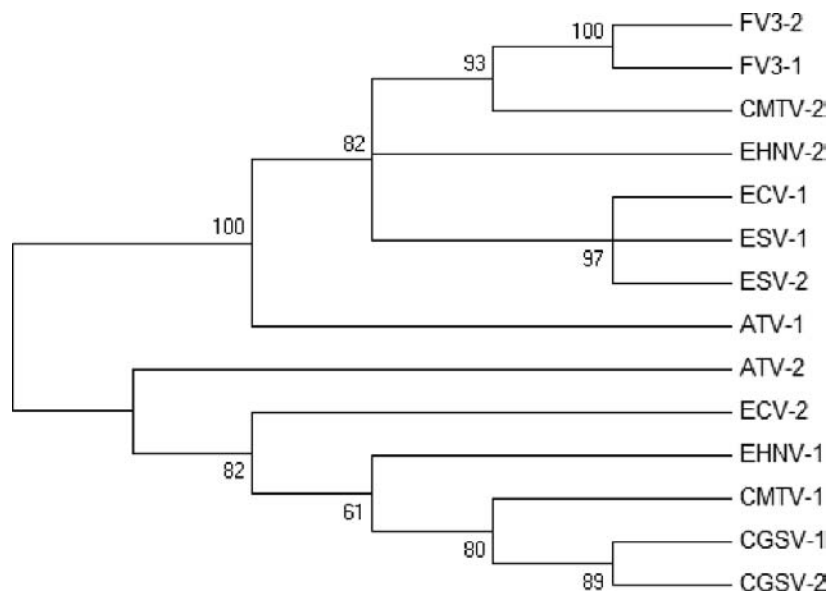


Figure 5. Neighbor-joining tree created in MEGA 6. The tree was built using the Tamura 3-parameter model of nucleotide substitutions with transversions and transitions. Bootstrap values were calculated 1000x for each branch.

sequence data for ranaviruses is available, it is easy to ask questions that use a large amount of data. Asking such questions will only end up causing frustration and lead to less-than-ideal work. Therefore, a lot of guidance in the beginning is helpful not only for the students but also for the instructor when trying to decipher the student's work. Many different open-access resources are available, including a text on ranaviruses with a chapter on phylogenetics (see Jancovich et al., 2015: chapter 3) that will be a good source of information for students and instructors.

Typically, I tell students to use two or three strains/isolates of each *Ranavirus* species that they are interested in. (However, this

might not be appropriate and depends on the question that the students are interested in answering.) This keeps the data down to a manageable size, while still letting students observe the relationships. To keep things simple, it may be easiest to ask questions such as "What is the closest relative of FV3 according to X gene?" Not all genes will necessarily give the same result; therefore, a great strategy is to make the students work in pairs or small groups and assign them each a different ORF to illustrate this. In cases like this, a mini-conference or poster session in class or lab would let the students see that all genes do not necessarily provide the same results when reconstructing a phylogeny.

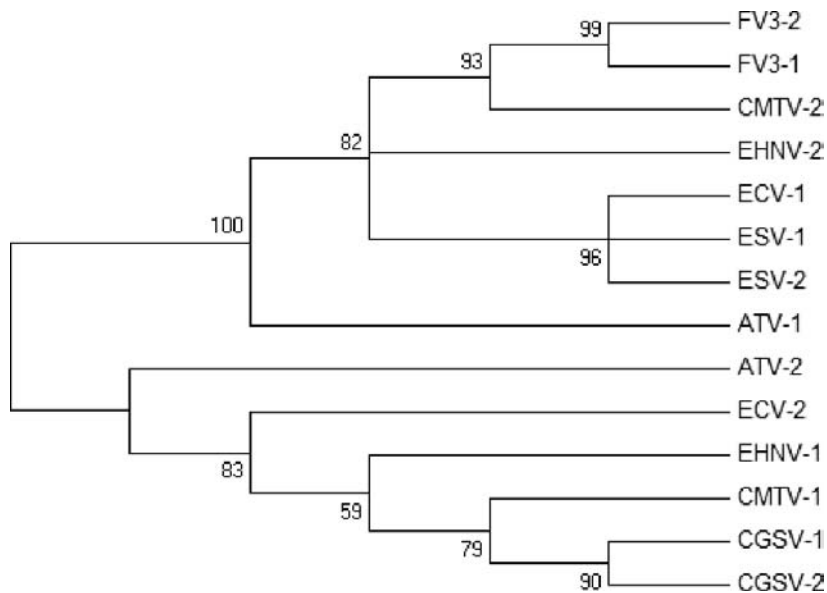


Figure 6. Minimum evolution tree created in MEGA 6. The Tamura 3-parameter model with evolutionary invariability model of nucleotide substitution was used to build the tree. Bootstrap values were calculated 1000x for each branch.

Table 2. Comparison of the three phylogenetic trees (maximum-likelihood [ML], neighbor-joining [NJ], and minimum evolution [ME]) of the observed branching patterns.

	All Three Trees	NJ & ML	NJ & ME	ML & ME
Similarities	<ul style="list-style-type: none"> • FV3s group as sister taxa with CMTV-2 • ECV-1 and ESV-1 and ESV-2 form a multifurcating branch • ATV-1 and ATV-2 are on separate clades • CGSV-1 and CGSV-2 group with CMTV-1 		<ul style="list-style-type: none"> • Trees are very similar; even the bootstrap values are comparable 	
Differences		<ul style="list-style-type: none"> • EHN-1 and ECV-2 group together in the ML, whereas in the NJ tree they are not on the same branch • EHN-2 is part of a multifurcation in the NJ tree but not in the ML tree 		<ul style="list-style-type: none"> • In the ME tree EHN-2 is in a separate clade, rather than a multifurcation as it is in the ML tree • In the ML tree, EHN-1 and ECV-2 are sister taxa, whereas in the ME tree they are not as closely related

○ Conclusions

Ranaviruses can provide a great introduction to reconstructing phylogenetic relationships. We are still learning about their origins and their relationships, which makes asking questions using them a great way for students to learn that science is continually changing as more information is gathered and further analyses

are performed. This exercise also gets students familiar with the collection and manipulation of sequence data from open-access sources. Furthermore, it promotes the use of servers and software that are freely available online for students and researchers. Projects using *Ranavirus* phylogenetics can be presented and assessed in a variety of ways, some of which will increase student engagement (e.g., mini-conference presentations).

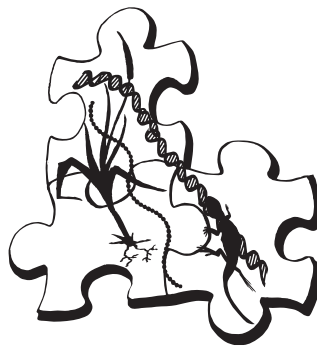
○ Acknowledgments

A big thank you goes out to Dr. Thomas Waltzek for being a great mentor and friend who helped me hone my phylogenetic analysis skills so that I could share my passion for ranaviruses with my students. I would also like to thank my undergraduate research students over the past few years for working with me as we navigate the wonderful world of *Ranavirus* phylogenetics and increasing our understanding of which of the core genes make good trees. I would also like to thank several anonymous reviewers, Dr. Sarah Rosario, and Dr. Anna Higgins-Harrell for comments on the manuscript, as they have improved this article significantly.

References

- Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Sayers, E.W. (2013). GenBank. *Nucleic Acids Research*, 41, D36–D42.
- Duffus, A.L.J., Waltzek, T.B., Stöhr, A.C., Allender, M.C., Gotesman, M., Whittington, R.J. et al. (2015). Distribution and host range of ranaviruses. In M.J. Gray & V.G. Chinchar (Eds.), *Ranaviruses: Lethal Pathogens of Ectothermic Vertebrates* (pp. 9–57). Dordrecht, The Netherlands: Springer.
- Eaton, H.E., Metcalf, J., Penny, E., Tcherepanov, V., Upton, C. & Brunetti, C.R. (2007). Comparative genomic analysis of the family Iridoviridae: re-annotating and defining the core set of iridovirus genes. *Virology Journal*, 4(11).
- Jancovich, J.K., Steckler, N.K. & Waltzek, T.B. (2015). Ranavirus taxonomy and phylogeny. In M.J. Gray & V.G. Chinchar (Eds.), *Ranaviruses: Lethal Pathogens of Ectothermic Vertebrates* (pp. 59–70). Dordrecht, The Netherlands: Springer.
- Katoah, K., Rozewicki, J. & Yamada, K.D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*. In press.
- Kumar, S., Stecher, G. & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33, 1870–1874.
- Newman, L., Duffus, A.L.J. & Lee, C. (2016). Using the free program MEGA to build phylogenetic trees from molecular data. *American Biology Teacher*, 78, 608–611.
- Price, S.J., Garner, T.W.J., Nichols, R.A., Balloux, F., Ayres, C., Mora-Cabello de Alba, A. & Bosch, J. (2014). Collapse of amphibian communities due to an introduction of *Ranavirus*. *Current Biology*, 24, 2586–2591.
- Stuart, S.N., Chanson, J.S., Cox, N.A., Young, B.E., Rodrigues, A.S., Fischman, D.L. & Waller, R.W. (2004). Status and trends of amphibian declines and extinctions worldwide. *Science*, 306, 1783–1786.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). MEGA 6: Molecular Evolutionary Genetic Analysis version 6. *Molecular Biology and Evolution*, 30, 2725–2729.
- Tan, W.G., Barkman, T.J., Chinchar, V.G. & Essani, K. (2004). Comparative genomic analysis of frog virus 3, type species of the genus *Ranavirus* (Family Iridoviridae). *Virology*, 323, 70–84.
- Teacher, A.F.G., Cunningham, A.A. & Garner, T.W.J. (2010). Assessing the long-term impacts of *Ranavirus* infection on wild common frog populations. *Animal Conservation*, 13, 514–522.

AMANDA L. J. DUFFUS is an Associate Professor in the Department of Biology and Physical Sciences at Gordon State College, Barnesville, GA 30204; email: aduffus@gordonstate.edu; webpage: www.amandaduffus.com; Twitter: @DrRanidae.



For Information:
tjarret@clemson.edu
864-656-2153

CLEMSON
BIOLOGICAL SCIENCES

Online MS in Biology

Master of Science (Non-thesis option)

Online Master's Degree in Biological Sciences
for K-12 teachers and other science educators

- All courses offered online
- Reduced tuition
- No out-of-state tuition differential
- No residency requirement
- 30 semester hours of graduate credits
- Up to 12 credits of graduate courses may transfer for the degree requirements

The courses offered in the **BIOL ONLINE** Program are fully accredited through Clemson University by the Southern Association of Colleges and Schools (SACS). CU is an equal opportunity employer