Submission ID #: 68323
Scriptwriter Name: Debopriya Sadhukhan
Project Page Link: https://review.jove.com/account/file-uploader?src=20838563

Title: A Concoction Pipeline for Generating Molecular Operational Taxonomic Units (MOTUs) Among Riparian and Aquatic Beetles

Authors and Affiliations:

Emmanuel D. Delocado[1,2], Voltaire Rafael H. Banzon, Jr.[1], Enrico Gerard S. Sanchez[1,3]

[1]Ateneo Biodiversity Research Laboratory, Department of Biology, School of Science and Engineering, Higher Education Cluster, Ateneo de Manila University
[2]Ateneo Institute of Sustainability, Ateneo de Manila University
[3]Ateneo School of Medicine and Public Health, Higher Education Cluster, Ateneo de Manila University, Don Eugenio Lopez Sr. Medical Complex

Corresponding Authors:

Emmanuel D. Delocado (edelocado@ateneo.edu)

Email Addresses for All Authors:

Emmanuel D. Delocado (edelocado@ateneo.edu)
Voltaire Rafael H. Banzon, Jr. (voltaire.banzon@student.ateneo.edu)
Enrico Gerard S. Sanchez (enrico.sanchez@student.ateneo.edu)

Author Questionnaire

1. We have marked your project as author-provided footage, meaning you film the video yourself and provide JoVE with the footage to edit. JoVE will not send the videographer. Please confirm that this is correct.

√ Correct

2. Microscopy: Does your protocol require the use of a dissecting or stereomicroscope for performing a complex dissection, microinjection technique, or something similar? **No**

3. Software: Does the part of your protocol being filmed include step-by-step descriptions of software usage? **Yes**

4. Proposed filming date: To help JoVE process and publish your video in a timely manner, please indicate the proposed date that your group will film here: **06/30/2025**

When you are ready to submit your video files, please contact our Content Manager, Utkarsh Khare.

Current Protocol Length

Number of Steps: 30
Number of Shots: 49

Introduction

REQUIRED:

1.1. **Emmanuel D. Delocado:** Seeking to speed up species discovery in the megadiverse tropics, this protocol aims to generate species clustering hypotheses in aquatic and riparian beetles using a concoction of computer-based analyses.

1.1.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B-roll: 2.6.1.*

What significant findings have you established in your field?

1.2. **Enrico Gerard S. Sanchez:** This concoction pipeline, which utilized COI DNA sequence, has led to greater evidence for species delimitation for the hydrophilid beetle from the genus *Anacaena,* which was previously erected only using morphological data.

1.2.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B-roll: LAB MEDIA: Figure 10.*

What research gap are you addressing with your protocol?

1.3. **Voltaire Rafael H. Banzon, Jr.:** This protocol seeks to address the taxonomic impediment that species discovery through morphology-based taxonomy, which is considered the gold standard in insect systematics, is often time-consuming and confusing for understudied yet megadiverse and highly inconspicuous aquatic beetles.

1.3.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera.

What advantage does your protocol offer compared to other techniques?

1.4. **Enrico Gerard S. Sanchez:** This pipeline generates molecular clustering hypotheses called MOTUs, regardless of whether the sequences correspond to known species with reference sequences or whether the species are undiscovered and undescribed.

1.4.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B-roll: 5.7.2, 5.8.1., 5.8.2.*

How will your findings advance research in your field?

1.5. **Voltaire Rafael H. Banzon, Jr.:** With this, rather than dissecting and comparing hundreds of beetle genital structures all at once, the pipeline provides preliminary clustering on which specimens can be scrutinized for conspecificity.

1.5.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera.

Protocol

2. Sequence Alignment as a Preparatory Step for Threshold-Based Approaches

Demonstrator: Enrico Gerard S. Sanchez

2.1. To begin, launch MEGA X *(Mega Ex)* and load the DNA sequences for alignment **[1]**. Open the alignment interface by selecting **Align**, then clicking on **Edit/Build Alignment** *(Edit or Build Alignment)*, and choosing **Create a new alignment**. Click **OK** to confirm the selection, and select **DNA** as the datatype **[2]**.

2.1.1. WIDE: Talent seated at a computer desktop launching MEGA X software and loading the DNA sequences.

2.1.2. SCREEN: 68323_screenshot_1.mp4 00:09-00:25.

2.2. Hover the mouse over the **Edit** tab, click on **Insert Sequence from File**, navigate to the appropriate directory, and select the sequence files to load into MEGA **[1]**.

2.2.1. SCREEN: 68323_screenshot_2.mp4 00:05-00:20.

2.3. Now, click on **Alignment**, then select **Align by ClustalW** *(Clustal-W)* to align the sequences using default settings, and click **OK** to proceed **[1]**.

2.3.1. SCREEN: 68323_screenshot_3.mp4 00:02-00:24.

2.4. For manual editing of the sequences, delete any insertions by clicking on the inserted bases or positions and pressing the **Delete** key on the keyboard **[1]**. Next, correct deletions by clicking on the dash that represents a missing base, removing it, and typing the intended base **[2]**. Then, find the earliest position where all sequences contain a character, click on the blank box in the row header immediately to the left of this position, and drag to select all excess starting positions. Press **Delete** to trim them **[3]**.

2.4.1. SCREEN: 68323_screenshot_4.mp4 00:03-00:10.

2.4.2. SCREEN: 68323_screenshot_4.mp4 00:11-00:23.

2.4.3. SCREEN: 68323_screenshot_4.mp4 00:24-00:41.

2.5. Then, find the last aligned position and click on the box to the right of this point. Press **Delete** again to trim the end **[1]**.

2.5.1. SCREEN: 68323_screenshot_5.mp4 00:02-00:16.

2.6. Now, select all sequences, and click on the **Translated Protein Sequences** tab. When prompted, verify the genetic code as **Invertebrate Mitochondrial [1]**. If the genetic code is different, click **No**, and a menu will appear allowing one to tick the box for Invertebrate Mitochondrial genetic code **[2]**.

2.6.1. SCREEN: 68323_screenshot_6.mp4 00:02-00:09.

2.6.2. SCREEN: 68323_screenshot_6.mp4 00:10-00:23.

2.7. If stop codons, marked by asterisks in the alignment, appear across an entire column, click on **DNA Sequences** and delete the first position for all sequences **[1]**.

2.7.1. SCREEN: 68323_screenshot_7.mp4 00:02-00:26.

2.8. Again, click on **DNA Sequences**, then save the alignment in .mas/x *(dot mas or dot mas-ex)* format **[1-TXT]**.

2.8.1. SCREEN: 68323_screenshot_8.mp4 00:02-00:22. **TXT: Export the sequence into other pertinent file types, such as .meg file and .fasta file**

3. Delimitation by TaxonDNA (Species Identifier 1.8) and Kimura 2-Parameter (K2P) in MEGA X

Demonstrator: Voltaire Rafael H. Banzon, Jr.

3.1. To start delimitation using the TaxonDNA *(Taxon-D-N-A)* module, open the TaxonDNA software. Click on **Import**, then select **FASTA** *(Fasta)*, and upload the alignment file in FASTA format **[1]**.

3.1.1. SCREEN: 68323_screenshot_9.mp4 00:06-00:28.

3.2. Now, click on **Modules**, then select **Cluster**. Set the threshold value to 3 percent and check the box labeled **Generate individual information on every cluster [1]**.

3.2.1. SCREEN: 68323_screenshot_10.mp4 00:02-00:12.

3.3. Click on **Make clusters now!** to start clustering **[1]** and save the results by taking a screenshot **[2]**.

3.3.1. SCREEN: 68323_screenshot_11.mp4 00:02-00:08.

3.3.2. SCREEN: 68323_screenshot_11.mp4 00:09-00:18.

3.4. For delimitation using the Kimura 2-parameter method, open MEGA and click on **File**, then select **Open A File/Session** *(Open a File or Session)* **[1]**. Click on **Distance**, select **Compute Pairwise Distances**, and confirm the .meg *(Meg)* file for delimitation **[2]**.

3.4.1. SCREEN: 68323_screenshot_12.mp4 00:04-00:18.

3.4.2. SCREEN: 68323_screenshot_12.mp4 00:19-00:27.

3.5. Hover over the box next to **Model/Method** *(Model or Method)*, and click the arrow to expand the dropdown menu. Select the **Kimura 2-parameter** model and click **OK** to run the program **[1]**. Open the data output window to view the results **[2]**.

3.5.1. SCREEN: 68323_screenshot_13.mp4 00:02-00:11.

3.5.2. SCREEN: 68323_screenshot_13.mp4 00:12-00:24.

3.6. Click on **File**, then select **Export/Print Distances** *(Export and print distances)* to save the result **[1]**.

3.6.1. SCREEN: 68323_screenshot_14.mp4 00:02-00:17.

4. Delimitation by ASAP and the Tree Estimation Step for Coalescent-Based Approaches

Demonstrator: Enrico Gerard S. Sanchez

4.1. Open the Assemble Species by Automatic Partitioning or ASAP *(asap)* web server **[1]**. Click on the orange box labeled **Choose a file**, and upload the .fasta *(Fasta)* file **[2]**. Scroll down and click **Go [3]**.

4.1.1. SCREEN: 68323_screenshot_15.mp4 00:01-00:03.

4.1.2. SCREEN: 68323_screenshot_15.mp4 00:04-00:15.

4.1.3. SCREEN: 68323_screenshot_15.mp4 00:16-00:24.

4.2. To download the clustering result, click on **list** for the row with the lowest ASAP score and the highest p-value rank **[1]**.

4.2.1. SCREEN: 68323_screenshot_16.mp4 00:02-00:16.

4.3. Next, to start the tree estimation, open MEGA and click on **Data**, then select **Open A File/Session** to load the .meg file **[1]**. Click on **Models**, then select **Find Best DNA/Protein Models (ML)** *(Best DNA and Protein Models)* to determine the best-fit substitution model **[2]**.

4.3.1. SCREEN: 68323_screenshot_17.mp4 00:02-00:15.

4.3.2. SCREEN: 68323_screenshot_17.mp4 00:16-00:26.

4.4. Click **OK** on the **Analysis Preferences** menu **[1]**.

4.4.1. SCREEN: 68323_screenshot_18.mp4 00:02-00:11.

4.5. Now, to generate the tree, click on **Phylogeny**, then select **Construct/Test Maximum Likelihood Tree** *(Construct and Test Maximum Likelihood Tree)* **[1]**.

4.5.1. SCREEN: 68323_screenshot_19.mp4 00:03-00:10.

4.6. Use the best-fit substitution model by clicking on the box beside **Model/Method**, and selecting the appropriate model from the dropdown menu **[1]**.

4.6.1. SCREEN: 68323_screenshot_20.mp4 00:02-00:17.

4.7. Under **Phylogeny Test**, click on the box beside **Test of Phylogeny**, then choose **Bootstrap method** from the dropdown menu **[1]**. Click on the box beside **No. of**

Bootstrap Replications *(Number of Bootstrap Replications)*, type 1000, and click **OK** to run the analysis **[2]**.

4.7.1. SCREEN: 68323_screenshot_21.mp4 00:02-00:12.

4.7.2. SCREEN: 68323_screenshot_21.mp4 00:14-00:27.

4.8. Once the window containing the resulting tree opens, save the session as an .mts/x *(dot M-T-S or dot M-T-S-X)* file **[1]**. Also, save the output as a .nwk *(newick)* file and export the tree as a .png *(P-N-G)* image **[2]**.

4.8.1. SCREEN: 68323_screenshot_22.mp4 00:02-00:12.

4.8.2. SCREEN: 68323_screenshot_22.mp4 00:13-00:47.

5. Delimitation by Poisson Tree Processes (PTP) and Multi-Rate Poisson Tree Processes (mPTP)

Demonstrator: Voltaire Rafael H. Banzon, Jr.

5.1. For the PTP-based delimitation, visit the PTP web server **[1]**. Click on **Choose file**, and upload the tree in Newick format **[2]**. Under **My tree is**, select **Rooted**. In the box under **Outgroup taxa names**, input the outgroup by typing the name of the taxon tip **[3]**.

5.1.1. SCREEN: 68323_screenshot_23.mp4 00:02-00:04.

5.1.2. SCREEN: 68323_screenshot_23.mp4 00:05-00:09.

5.1.3. SCREEN: 68323_screenshot_23.mp4 00:10-00:34. *Video Editor: Blur the email id or remove the part where the talent is entering the email id.*

5.2. Next, under **Maximum likelihood solution**, click on **Download delimitation results** to save the PTP-ML *(P-T-P-M-L)* output **[1]**.

5.2.1. SCREEN: 68323_screenshot_24.mp4 00:01-00:08.

5.3. Similarly, under **Highest Bayesian-supported solution**, click on **Download delimitation results** to save the PTP-BI *(P-T-P-B-I)* output **[1]**.

5.3.1. SCREEN: 68323_screenshot_25.mp4 00:03-00:08.

5.4. For the mPTP *(M-P-T-P)*-based delimitation by mPTP, visit the mPTP web server **[1]**.

5.4.1. SCREEN: 68323_screenshot_26.mp4 00:02-00:04.

5.5. Upload the Newick file by dragging it onto the gray square or by clicking on the square **[1]**. Once the data is loaded, click on **Proceed to outgroup selection [2]**. On the Outgroup specification page, select the outgroup by clicking the checkbox next to the taxa labels of outgroup specimens **[3]**. Then, click on **Model selection,** select **MPTP**, and click on **Visualization Options**. On the **Visualization options** page, accept the default settings **[4]**.

5.5.1. SCREEN: 68323_screenshot_27.mp4 00:01-00:07.

5.5.2. SCREEN: 68323_screenshot_27.mp4 00:08-00:10.

5.5.3. SCREEN: 68323_screenshot_27.mp4 00:11-00:16.

5.5.4. SCREEN: 68323_screenshot_27.mp4 00:17-00:28.

5.6. Click on **Submit**. Right-click the files under **Downloadable Files** and choose **Save as** to save the results **[1]**.

5.6.1. SCREEN: 68323_screenshot_28.mp4 00:06-00:37.

5.7. Now, to generate the Molecular Operational Taxonomic Units or MOTU *(mow-tu)*, open the tree using a photo editing program or PowerPoint **[1]**. Create a bar to represent the results of each molecular species delimitation approach **[2]**.

5.7.1. SCREEN: 68323_screenshot_29.mp4 00:02-00:15.

5.7.2. SCREEN: 68323_screenshot_29.mp4 00:55-01:26, 02:13-02:16.

5.8. If all the approaches yield identical results for a given molecular cluster, designate the cluster as a MOTU by consensus **[1-TXT]**. NOTE: The VO has been edited.

5.8.1. SCREEN: 68323_screenshot_30.mp4 00:02-00:22. **TXT: Designate a cluster as a MOTU if the majority of methods yield the same result** NOTE: The information of 5.8.2 has been added here concisely as onscreen text as the authors didn't provide any video for this.

~~5.8.2. SCREEN: To be provided by authors: Highlighting a cluster with mostly matching results and labeling it as a MOTU (majority).~~

Results

6. Results

6.1. This figure presents the maximum likelihood gene tree and molecular clustering results for *Byrrhinus* beetles based on COI-3' *(C-O-I three-prime)* sequences and six delimitation methods **[1]**.

6.1.1. LAB MEDIA: Figure 9. *Video editor: Highlight the left image (tree-like structure) when the VO says "COI-3' sequences" and the colored bars (**expect** the black bars) on the right image when the VO says "and six delimitation methods".*

6.2. All methods consistently identified four MOTUs *(mow-tus)*: *Byrrhinus negrosensis*, *Byrrhinus villarini*, *Byrrhinus* A, and *Byrrhinus* B **[1]**, with identical clustering across methods **[2]**.

6.2.1. LAB MEDIA: Figure 9. *Video editor: Highlight the four uniform black bars under the "MOTU" label.*

6.2.2. LAB MEDIA: Figure 9. *Video editor: Highlight all the Byrrhinus negrosensis, Byrrhinus villarini, Byrrhinus sp A, and Byrrhinus sp B labels on the tree-like structure.*

6.3. *Byrrhinus negrosensis* and *Byrrhinus villarini* were clearly separated despite originating from the same locations **[1]**.

6.3.1. LAB MEDIA: Figure 9. *Video editor: Highlight all the Byrrhinus negrosensis and Byrrhinus villarini labels in the lower portion of the tree and the colored bars (**except black**) in the 3rd and 4th rows from the top in the right image.*

6.4. Sequences from four provinces were grouped into *Byrrhinus* A, showing no strong geographic structuring **[1]**.

6.4.1. LAB MEDIA: Figure 9. Video editor: *Highlight all the Byrrhinus sp A labels at the top portion of the tree.*

6.5. This figure shows clustering of *Anacaena* beetles into four MOTUs, including *Anacaena angatbuhay*, *Anacaena auxilium*, *Anacaena* A, and *Anacaena* B, with full agreement across methods **[1]**.

6.5.1. LAB MEDIA: Figure 10.

6.6. *Anacaena auxilium* overlapped geographically with *Anacaena* A and *Anacaena* B, indicating that geography alone did not explain molecular divergence **[1]**.

6.6.1. LAB MEDIA: Figure 10. *Video editor: Highlight the Anacaena auxilium, Anacaena sp A, and Anacaena sp B labels.*