Ceren Ordas
EEOB 563

Exploring the Genetic Connectivity Between Madagascar and the Indo-Pacific Through the Distribution of Bivalve Species

Introduction

Madagascar is recognized as a biodiversity hotspot and for the endemism of its marine life, but there has been limited research on bivalves. Even though freshwater mussels are known as poor dispersers across oceanic and terrestrial barriers, and their presence in Madagascar was assumed to be due to continental drift, further studies have shown that they have also arrived by transoceanic dispersal (Graf & Cummings, 2009). The majority of Madagascar's plant and animal species are endemic and have evolved from their closest relatives millions of years ago. New genetic approaches have helped with the differentiation of species that had previously been thought to be single species (Ganzhorn et al., 2014). Therefore, Madagascar can be used as a model for exploring numerous evolutionary processes, such as convergent evolution.

The behavioral range of many marine invertebrates such as dispersal range and larval development period has an influence on the diversity of these species. Genetic structure and connectivity across the range of bivalve species can therefore be caused by the interaction of the ocean currents, which cause dispersal during the larval stage (Lal et al., 2017) (Figure 1).

Previous studies have observed high levels of phenotypic variation, differences in shell morphology, and variation in physiology and behavior in Australian freshwater mussel species. It was then concluded that these species may be cryptic species, and molecular processes can improve our understanding of diversity and evolution in mussels in Madagascar as well (Baker et al., 2003). Therefore, by generating a phylogenetic tree of the species that originated in

Madagascar but also have occurrences spotted across the Indo-Pacific region, we can observe the distributions on the tree.

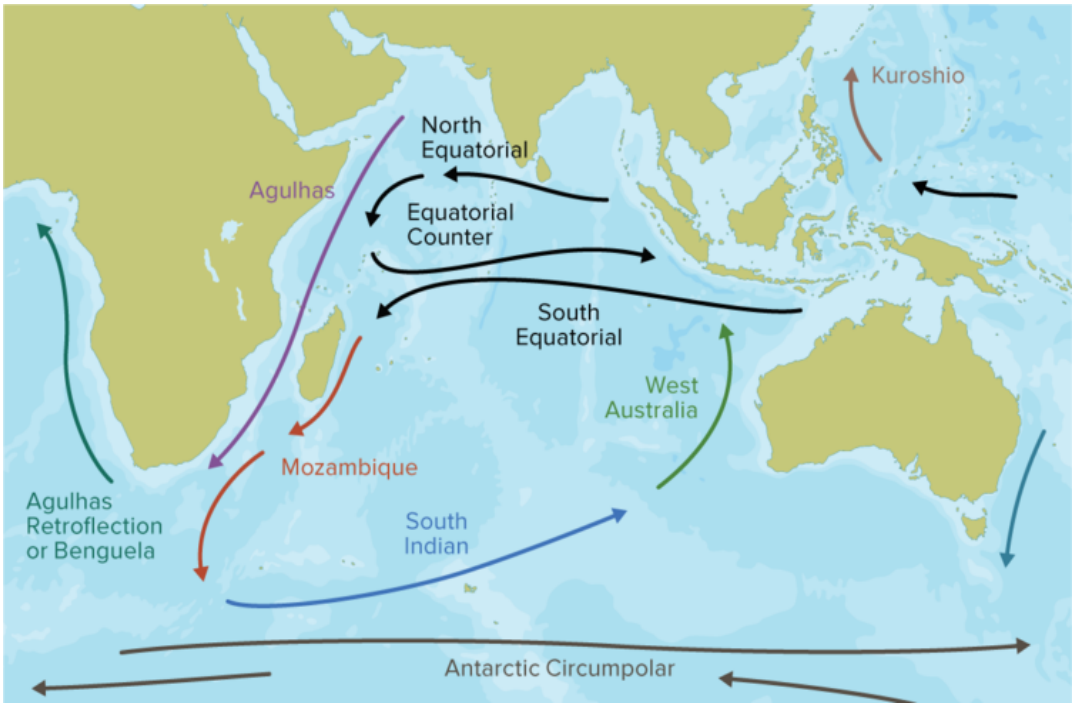

Figure 1. The ocean currents of the Indian Ocean affecting the dispersion of bivalve species during their larval stage.

The general goal of this study is to observe the wide range of endemic species of bivalves in South Madagascar and the Indo-Pacific region to assess their phylogeny for cryptic biodiversity due to their similar morphologies. Hence, we will analyze the DNA sequences of bivalve species found in Madagascar as well as the Indo-Pacific region and build a phylogenetic tree. We can then make an assumption on the dispersal type of the larvae (long vs. short disperser), and if it caused certain species to survive in different habitats and be related, or if they are different species. Therefore, we will compare and contrast molecular (DNA sequences) and morphological variation of these bivalve species.

We also hypothesize that if the species is a long-distance disperser then it will be a broad range species, but if a species in Madagascar is a short time/distance disperser and found in multiple habitats, then we would expect to observe at least two different species. If the DNA samples of the same species from different individuals are on different parts of the tree and not close to each other in distance, we predict that they are different species.

The species that will be analyzed are *Laevichlamys weberi, Laevichlamys lemniscata,* and *Mimachlamys sanguinea.* These species can vary in color and size within themselves even if they occupy the same or different regions (Figure 2). Hence, it is important to use genetic approaches, because the shell morphology is a misleading characteristic.

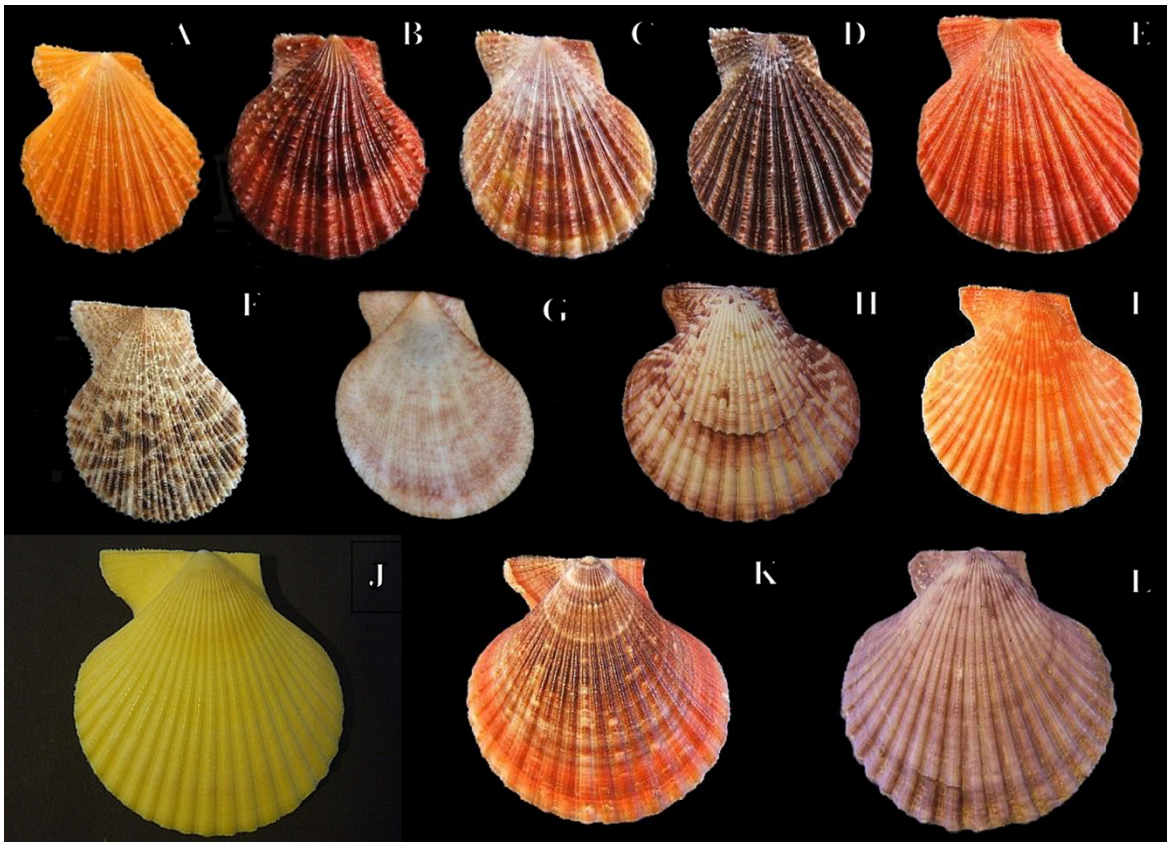

Figure 2. External shell morphologies of *Laevichlamys lemniscata* (A-E)*, Laevichlamys weberi* (F-G) and *Mimachlamys sanguinea* (H-L)*.*

Laevichlamys weberi and *Laevichlamys lemniscata* are endemic to the Madagascar region meanwhile *Mimachlamys sanguinea* is not an endemic species and seems to have a wide-range distribution across the Indo-Pacific region. Even though, *Laevichlamys lemniscata* is endemic, the World Register of Marine Species (WoRMS) documents its distribution across the Indo-Pacific as well. Meanwhile, this distribution is limited for *Laevichlamys weberi*, which is only seen in Madagascar, Kenya and Mozambique (Figure 3).

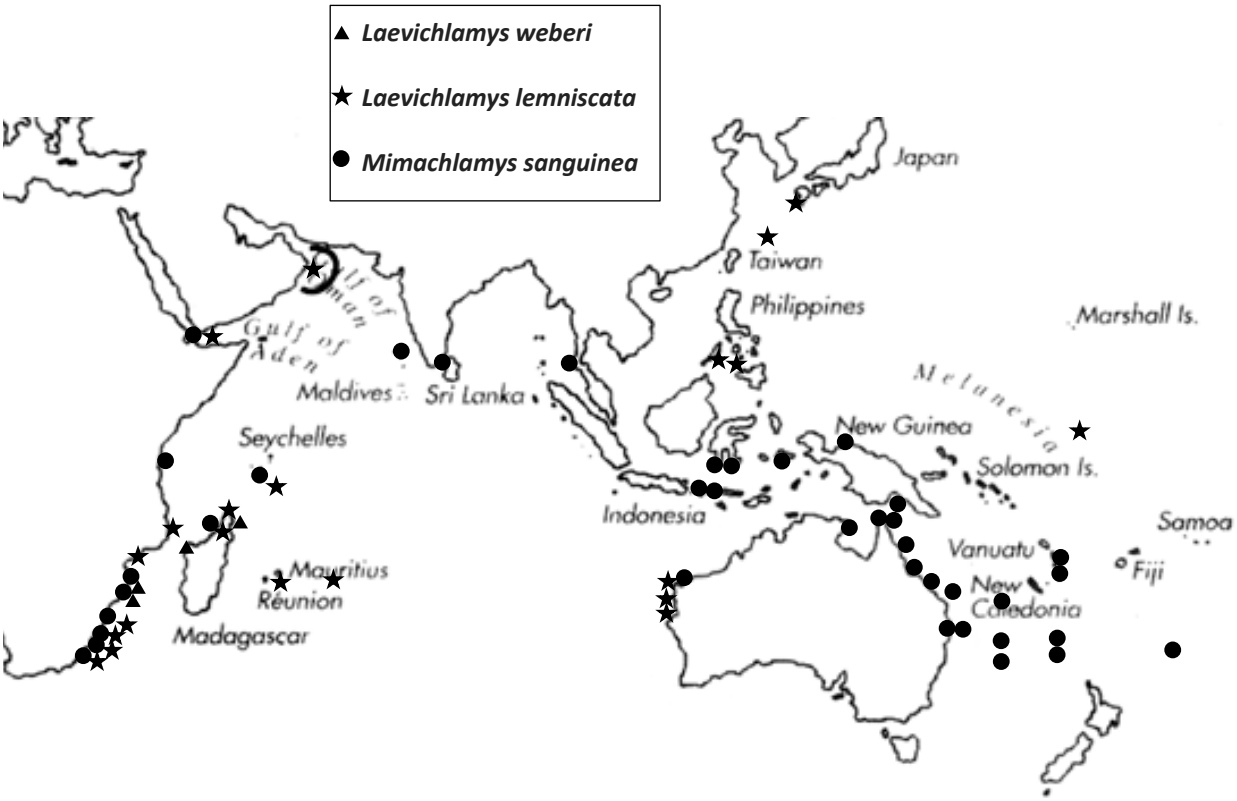

Figure 3. Localities/occurrences of previous samples examined for *Laevichlamys weberi, Laevichlamys lemniscata,* and *Mimachlamys sanguinea.*

The species *Laevichlamys weberi* and *Laevichlamys lemniscata* belong to the same tribe, Chlamydini, meanwhile *Mimachlamys sanguinea* is in the Mimachlamydini tribe but all of them are in the same subfamily. The outgroup we will be using for the phylogenetic analysis is the species *Argopecten irradians* which is in a different subfamily even though they are all Pectinidae (Table 1).

Mollusk Species	Family	Subfamily	Tribe	Genus	Location
Laevichlamys lemniscata	Pectinidae	Pedinae	Chlamydini	*Laevichlamys*	Madagascar, Mozambique
Laevichlamys weberi	Pectinidae	Pedinae	Chlamydini	*Laevichlamys*	Kenya, Madagascar, Mozambique, Indian Ocean
Mimachlamys sanguinea	Pectinidae	Pedinae	Mimachlamydini	*Mimachlamys*	Philippines, Vietnam, Madagascar, Kenya, Mozambique, Tanzania, Indian Ocean, South Pacific Ocean, Red Sea, Republic of Mauritius, Coral Sea
Argopecten irradians (outgroup)	Pectinidae	Pectininae	Aequipectinini	*Argopecten*	Eastern China Sea, South China Sea, North Atlantic Ocean, North Pacific Ocean

Table 1. Bivalve species' family, subfamily, tribe, genus, and location.

Methods

Study Species: The species we are looking into were collected at an expedition to Madagascar by The Muséum National d'Histoire Naturelle (MNHN, Paris), the Institut d'Halieutique et des Sciences Marines, University of Toliara (IH.SM), and the Wildlife Conservation Society (WCS) Madagascar Programme in April-June 2010. The species were then matched with their MNHN numbers and their locations were noted using the World Register of

Marine Species (WoRMS). We will focus on three species: *Laevichlamys weberi, Laevichlamys lemniscata,* and *Mimachlamys sanguinea.*

Tissue Sampling and DNA extraction have already been completed and the data of the DNA sequences of the species below will be provided by the Serb lab at Iowa State University. There were two DNA samples provided for each species we focused on (*Laevichlamys weberi, Laevichlamys lemniscata,* and *Mimachlamys sanguinea*) and one DNA sequence for the outgroup (*Argopecten irradians*).

Data Summary:

These species have four gene regions, representing two mtDNA and two nuclear genes: 12S rRNA, 16S rRNA, histone H3, and 28S rRNA. The raw data was not eligible; therefore, the trimmed DNA sequences for *Laevichlamys weberi, Laevichlamys lemniscata, Mimachlamys sanguinea,* and *Argopecten irradians* were extracted from the nexus file provided.

General Procedure:

The sequences were organized by gene and taxa. They were aligned to check for abnormalities/differences and the overlapping regions were trimmed. The gaps were removed. The alignments were added to the Madagascar_Mollusk file on git. To find the best model, different partition schemes were used: unpartitioned, partitioned by gene region using the gamma distribution, and partitioned by gene region using the gamma distribution for the first gene (12S) and the inverse gamma distribution for the other three (16S, 28S, and H3).

Bayesian analysis with MrBayes:

MrBayes was installed to HPC-class with the code on EEOB563 GitHub. To run the analysis, the log command was used to save all of the screen output from the analysis to a file called madagascar-part-log.txt. Next, the sequences on madagascar_data_final.nex were loaded into the program with the execute command. The outgroup taxon was set as *Argopecten irradians.*

Unpartitioned analysis:

We first performed an analysis on the unpartitioned alignment using the GTR+Γ model. The lset command was used in which the nst=6 and rates=gamma. Then, we specified priors for all of the parameters of this nucleotide substitution model using the command prset in which the rate parameter was 0.05. This way, we were able to have a single-partition analysis.

Partitioning by gene region:

The dataset we used in this study has the gene regions 12S, 16S, H3, and 28S. Then we specified the data partitions corresponding to these genes to have a substitution model for each of the partition. The charset command was used to define the subset sites belonging to each of the gene regions in which 12S=1-408, 16S=409-866, 28S=867-1713, H3=1714-2066. Then we defined our partition configuration using the partition command partition by_gene=4:12S,16S,28S,H3 and then set the partition scheme to by_gene. After that, we assumed that both genes evolved under the GTR+Γ model by lset applyto=(all) nst=6 rates=gamma just like we previously did. We also kept the same priors as before. To make each gene have a set of partition-specific parameters we used the command unlink revmat=(all) statefreq=(all) shape=(all). To have the overall substitution rate vary across the subsets of our alignment we used the command prset applyto=(all) ratepr=variable. For our

second model we kept everything from the previous setup the same except we used an inverse gamma distribution for the among site variation for the genes 16S, 28S, and H3. For the first gene 12S we kept the gamma distribution.

Running the stepping-stone sampling:

After conducting our analysis with three different models, the stepping-stone sampling was done to analyze the ratio of marginal likelihoods (ML) for the three models. For the unpartitioned and partitioned models we specified the parameters of the analysis using the ssp command in which the ngen at 1000000, diagnfreq at a 1000 and then executed the analysis by ss. The marginal likelihoods for each of the models were then noted.

Setting and running the MCMC:

The MCMC was set for the partitioned model with the inverse gamma distribution in which the analysis was performed with four chains with the temperature set to 0.2. The chains were run for 10000000 cycles. Three independent analyses were set simultaneously. The sample frequency was set to 100. The print frequency was set to 1000. The check frequency was set to 5000 and the branch length information was saved. (ngen=10000000, nruns=3, printfreq=1000, samplefreq=100, nchains=4, checkfreq=5000, temp=0.2, savebrlens=yes).

Summarizing MCMC samples:

The sumt command was used to summarize the samples of the tree topology and branch lengths and the sump command was used to summarize all of the other model parameters. Relburnin was set to yes and the burninfrac was set to 0.25. The consensus tree type was set to half-compatible.

Slurm Workload manager:

A slurm job script was created using the Slurm job script generator. The job script was then pasted in the local file. The commands for loading modules and running the programs for the partitioned model with the inverse gamma distribution was added at the bottom of the script. The script in the file was then submitted using the sbatch command.

FigTree:

The consensus tree was committed to git and added to the local repository. The Newick tree format was opened on FigTree and saved with the posterior probability scores.

Results

The stepping-stone sampling that we conducted to assess the marginal likelihood (ML) scores for of the three models helped us to analyze the ratio between each approach of partitioning. The stepping-stone sampling for the unpartitioned model gave a marginal likelihood score of -5975.87. The ML for the partitioned model with the gamma distribution had a ML score of -5858.40 and the ML for the partitioned model with the inverse gamma distribution for the among site variation for the genes 16S, 28S, and H3 and with the gamma distribution for the first gene 12S was -5856.29. Since the partitioned model with the inverse gamma distribution had the largest ML score compared to the other models, the MCMC was done and the consensus tree was constructed using this model.

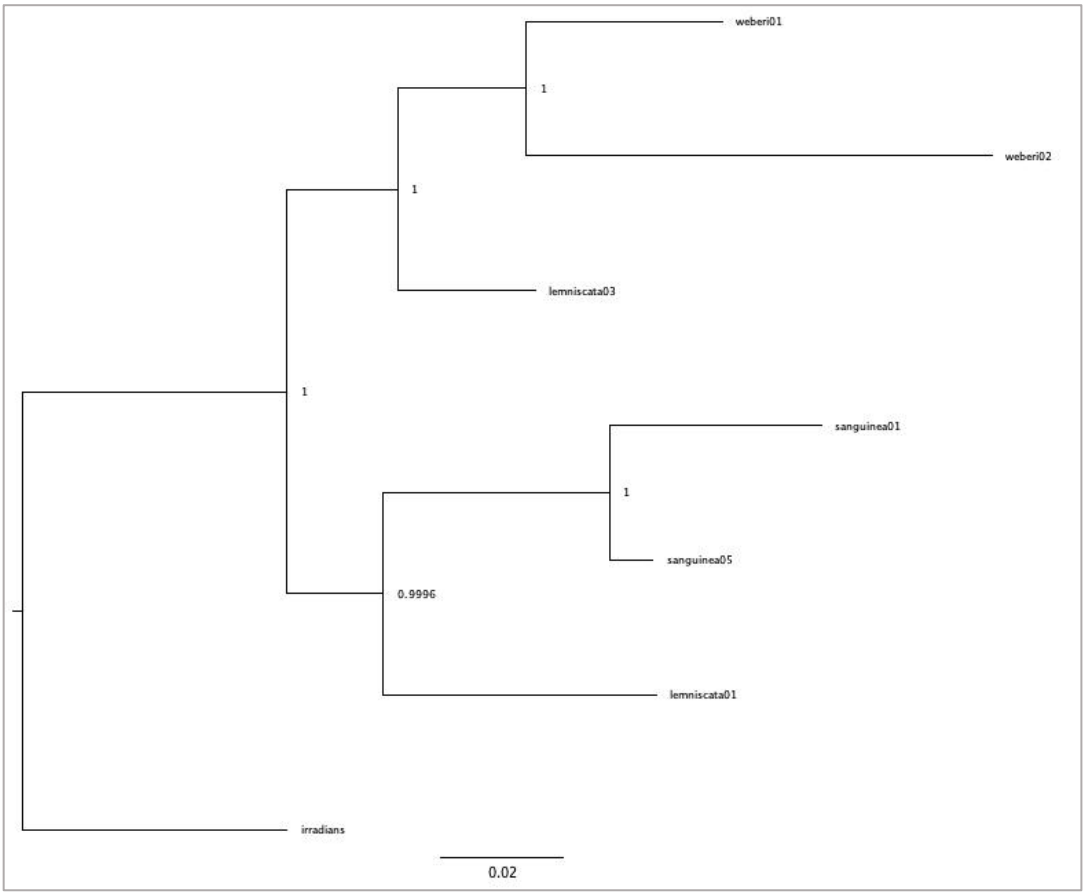

Figure 4. Bivalve species consensus tree with posterior probability scores.

When constructing the consensus tree, the outgroup was set as the *Argopecten irradians* and was rooted using that. The posterior probabilities were also high. The DNA sequences given for the two individual samples of *Laevichlamys weberi* (weberi01 and weberi02) are in the same clade with a posterior probability of 1. Their most recent common ancestor is the individual sample we had for the species *Laevichlamys lemniscata* (lemniscata03) with a posterior probability of 1. Therefore, *Laevichlamys weberi* clade seems to be more closely related to the sample we have for *Laevichlamys lemniscata* (lemniscata03).

Also, the DNA sequences for the two individual samples of *Mimachlamys sanguinea* (sanguinea01 and sanguinea05) are in the same clade with a posterior probability of 1. Their most recent common ancestor is from the sample we had for the species *Laevichlamys lemniscata* (lemniscata01) with a posterior probability of 0.9996. Therefore, *Mimachlamys sanguinea* clade seem to be more closely related to the sample we have for the other *Laevichlamys lemniscata* (lemniscata01). Meanwhile, the *Argopecten irradians* is the most distantly related to all the other species (Figure 4).

Discussion

Looking at the consensus tree, we said that the species *Laevichlamys weberi* and *Laevichlamys lemniscata* belong to the same tribe, Chlamydini, so seeing the species of *Laevichlamys weberi* (weberi01 and weberi02) being most closely related to one of the samples we had for the species *Laevichlamys lemniscata* (lemniscata03) only supports this prediction. Therefore, we were expecting both of the *Laevichlamys lemniscata* species to be in the same clade and be more closely related to *Laevichlamys weberi*, but lemniscata01 was more closely related to the *Mimachlamys sanguinea* clade, being its most recent common ancestor. Looking at the placement of the species, when comparing the *Argopecten irradians* with the others, we expected *Mimachlamys sanguinea* to be closely related to both *Laevichlamys weberi* and *Laevichlamys lemniscata*, even though they are not a part of the same tribe, since *Mimachlamys sanguinea* is a part of the Mimachlamydini tribe. But these tribes are still close in relation compared to the tribe of the *Argopecten irradians*' which is the Aequipectinini tribe. Also, the outgroup doesn't even share the same subfamily as the other species since it is in the Pectininae subfamily while the others are in the Pedinae subfamily. Therefore, the placement of

Laevichlamys weberi and *Laevichlamys lemniscata* (lemniscata03) on the tree shows that they are more closely related to each other than the other species due to them being in the same tribe, meanwhile the *Mimachlamys sanguinea* is still close in relation when we look at its placement compared to the *Argopecten irradians.*

Like we have mentioned before, the dispersal range and larval development period has a significant influence on the diversity of these species. The genetic connectivity across the range of bivalve species is due to the interaction of the ocean currents which causes the mollusks to disperse during their larval period (Lal et al., 2017).

Since previous studies have emphasized morphological differences of shells, phenotypic and behavioral variation of bivalve species that are considered to be the same can actually be cryptic species and the molecular approach that we used here with our Bayesian analysis does indicate a cryptic diversity for *Laevichlamys lemniscata* (Baker et al., 2003). Meanwhile, *Mimachlamys sanguinea* can be considered a wide range species and *Laevichlamys weberi* is just endemic to the Madagascar region, current records of it indicating its observation around Kenya, Madagascar, Mozambique. In general, it is shown to be in the Indian Ocean. Even though, its range extends to tropical East Africa, compared to *Mimachlamys sanguinea,* we can say that its range is still quite limited. When it comes to *Mimachlamys sanguinea*, the DNA samples that we have for both the individuals being in the same clade and this species dispersal ranging from Madagascar to the Indo-Pacific region underlines our prediction that they are related and it is a wide-range (long-distance disperser) species, and it is able to survive in different habitats due to its distribution on the tree. When we look at the *Laevichlamys lemniscata*'s DNA samples for both the individuals, they are not in the same clade and

lemniscata03 is more closely related to the clade of *Laevichlamys weberi* while lemniscata01 is more closely related to the *Mimachlamys sanguinea* clade. Since *Laevichlamys lemniscata* and *Mimachlamys sanguinea* are not even in the same tribe, we assume that the placement of lemniscata03 on the tree is more accurate. Therefore, the tree suggests that there is cryptic biodiversity present for lemniscata01, which might be due to its similar morphology to this particular species. The external shell morphologies of certain a species can differ within themselves but can also resemble others in terms of size, shell shape and color (for example: A, E, I, and K from Figure 1). Hence, we can make the assumption that the dispersal type of the larvae for the *Laevichlamys lemniscata* might be short-distance. The DNA sequences of the species we considered to be the same from two different individuals were on different parts of the tree and not close to each other in distance. Therefore, we seem to be observing two different species from these samples.

In our study, we ran into limitations with our sample size. Further studies should be conducted with more individuals from each species and more taxa should be added that are from the same tribes. Since the World Register of Marine Species claims that *Laevichlamys lemniscata* has a distribution across the Indo-Pacific, having more DNA sequences might still suggest that it is a wide-range species as well.

Previous research in the Pacific Ocean shows that ocean currents and geographic distances are major influences on population connectivity as well. The inland waters of Madagascar and the Indian Ocean islands have a high diversity of aquatic species and high levels of endemism. Hence, future studies will also help to emphasize the biodiversity within Madagascar and provide detailed information of behavior of bivalve larvae and its effects on

convergent evolution and genetic connectivity. By assessing these bivalve species' distributions among Madagascar and the Indo-Pacific region, we can identify biodiversity to assess conservation priorities as well.

References

1. Graf, D. L., & Cummings, K. S. Actual and Alleged Freshwater Mussels (Mollusca: Bivalvia: Unionoida) from Madagascar and the Mascarenes, with Description of a New Genus, Germainaia. *Proceedings of the Academy of Natural Sciences of Philadelphia*, *158*(1), 221–238. (2009).
2. Roland, H., & Héros, V. Description of new Muricidae (Mollusca: Gastropoda) collected during the ATIMO VATAE expedition to Madagascar "Deep South." *Zoosystema*, *35*(4), 503–523 (2013).
3. Ganzhorn, J. U., Wilme, L., & Mercier, J.-lU. C. Explaining Madagascar's biodiversity. *Conservation and Environmental Management in Madagascar*, 41–67. (2014).
4. Lal, Monal M., et al. "Swept Away: Ocean Currents and Seascape Features Influence Genetic Structure across the 18,000 Km Indo-Pacific Distribution of a Marine Invertebrate, the Black-Lip Pearl Oyster Pinctada Margaritifera." *BMC Genomics*, vol. 18, no. 1, (2017).
5. Baker, Andrew M., et al. "Cryptic Species and Morphological Plasticity in Long-Lived Bivalves (Unionoida: Hyriidae) from Inland Australia." *Molecular Ecology*, vol. 12, no. 10, pp. 2707–2717, (2003).