[redacted]

Word Count: 1382

Environmental drivers of speciation in terrestrial and marine Pseudosuchia

ABSTRACT

Pseudosuchia is a group of reptiles that appeared over 250 million years ago and gave rise to modern crocodylians. This group underwent a significant decline in ecological and morphological diversity which has been linked to changes in environmental variables. Two important factors that have shaped the evolutionary history of Pseudosuchia include global temperature and sea level. In this report, the effect of global temperature and sea level on the speciation rate of terrestrial and marine Pseudosuchia was investigated. Wilcoxon signed-rank tests revealed that speciation is driven by high temperatures and low sea levels for both terrestrial and marine Pseudosuchia. That is likely due to the ectothermic nature of the group that relies on higher temperatures to survive, while low sea levels increase the availability of coastal habitats. Understanding of the evolutionary patterns could potentially help to predict the fate of modern crocodylians, however, human intervention might complicate the matter.

INTRODUCTION

Pseudosuchia is a very important group of reptiles that first appeared over 250 million years ago (Butler *et al*., 2011; Mannion *et al*., 2015). Today, modern crocodylomorphs (crocodiles, caimans, alligators, and gharials) include semi-aquatic predators that occupy mainly tropical regions (Mannion *et al*., 2015). This group of organisms has received much attention from the scientific community as it experienced a significant decline in biodiversity over geological time (Mannion *et al.,* 2015). Specifically, Pseudosuchia used to be a lot more ecologically and morphologically diverse and occupied habitats over a much broader latitudinal range (Carvalho *et al*., 2010; Wu, Sues and Sun, 1995). However, there are only 23 extant pseudosuchian species which live in terrestrial or freshwater environments, while in the past they occupied marine habitats too (e.g. thalattosuchians)(Pol and Gasparini, 2009). This decline in diversity is likely to have been driven by changes in environmental conditions over geological time. In fact, many authors attribute such outcomes to changes in climatic conditions, especially temperature. Fluctuations in average global temperatures have also been associated with changes in other significant environmental factors such as sea level. Both of these environmental variables impact the survival of this reptile clade by affecting the availability of prey and suitable habitats, and their ability to thermoregulate. These factors are likely to have affected the speciation of Pseudosuchia differently in marine and terrestrial habitats. While most studies focus on fluctuations of extinction rates through time, a decreasing speciation rate has been suggested to cause a decreased biodiversity more compared to increased extinction in marine Pseudosuchia (Bambach, Knoll and Wang, 2004). Understanding their unique evolutionary history could help predict and understand the fate of their living counterparts, which are currently under high risk of extinction (Mannion *et al.,* 2015). As a result, the aim of this study is to determine whether global temperature and sea level drove speciation in marine and terrestrial Pseudosuchia taxa.

RESULTS

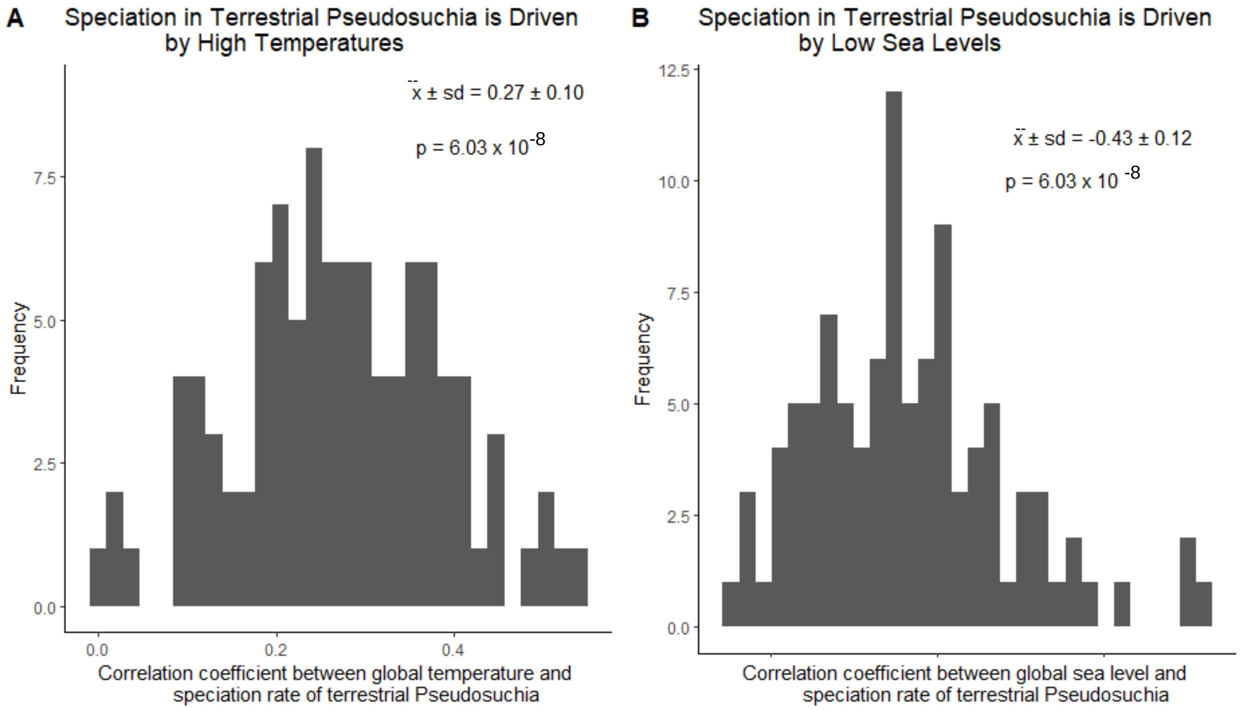
How did global temperature and sea level affect speciation rate in terrestrial Pseudosuchia?

Method and Statistical Testing

A phylogenetic tree scaled over geological time was constructed for terrestrial Pseudosuchia taxa. Speciation rates were extracted using BAMMtools package (Rabosky *et al*., 2014) and correlated with temperature proxy data and sea level data for the chosen taxa. Finally, a Wilcoxon signed-rank test was performed for each environmental variable to test the significance of the relationships found.

Result

There is a significant positive relationship between global temperature and speciation rate of terrestrial taxa (Wilcoxon signed-rank test on correlation coefficient, $\bar{x} \pm sd = 0.27 \pm 0.10$, $p = 6.03\times10^{-8}$; Fig. 1A) and a significant negative relationship between sea level and speciation rate (Wilcoxon signed-rank test, $\bar{x} \pm sd = -0.43 \pm 0.12$, $p = 6.03\times10^{-8}$; Fig. 1B). Hence, speciation of terrestrial Pseudosuchia taxa is driven by higher temperatures and low sea levels.



A Speciation in Terrestrial Pseudosuchia is Driven by High Temperatures

B Speciation in Terrestrial Pseudosuchia is Driven by Low Sea Levels

Figure 1: **Histogram A shows the frequency of the correlation coefficient between global temperature and speciation rate of terrestrial Pseudosuchia. Histogram B shows the frequency of the correlation coefficient between global sea level and speciation rate of terrestrial Pseudosuchia. Wilcoxon signed-rank tests revealed that speciation in terrestrial Pseudosuchia taxa is driven by high temperatures and low sea levels. Mean correlation coefficient ($\bar{x}$), standard deviation (sd) and the associated p-values (p) are shown.**

Interpretation

Higher temperatures drive speciation rates in terrestrial taxa, which could explain why the speciation rate has declined over the past 100 million years (Fig. 2). Following the decline in temperature during the Upper Cretaceous period and onwards (100 MYR ago; Fig .3), significant evolutionary radiation has not been observed when compared to the Triassic speciation events. The biggest speciation events, mainly during the Triassic Period, took place before this significant decline in global temperature as demonstrated in the phylogenetic tree (Fig. 4). Higher atmospheric temperature is likely to have favoured terrestrial Pseudosuchia taxa, as similarly to modern crocodylians, they were ectotherms that

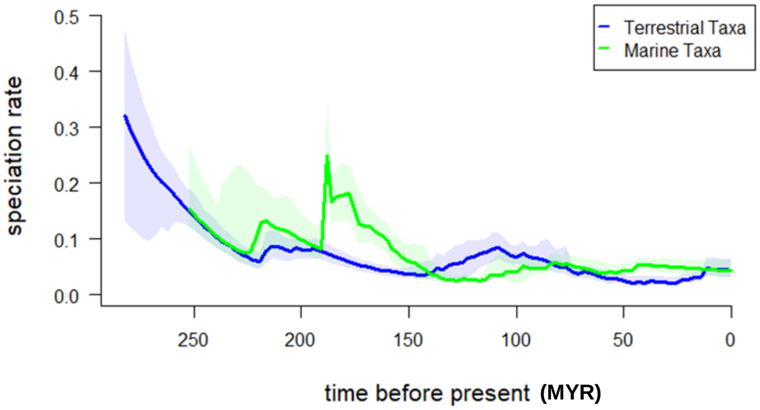


Figure 2: Figure shows the speciation rate over geological time (in millions of years; MYR) in terrestrial and marine Pseudosuchia taxa. Shading around the lines represents the confidence intervals. Figure created using BAMMtools package (Rabosky *et al.*, 2014).

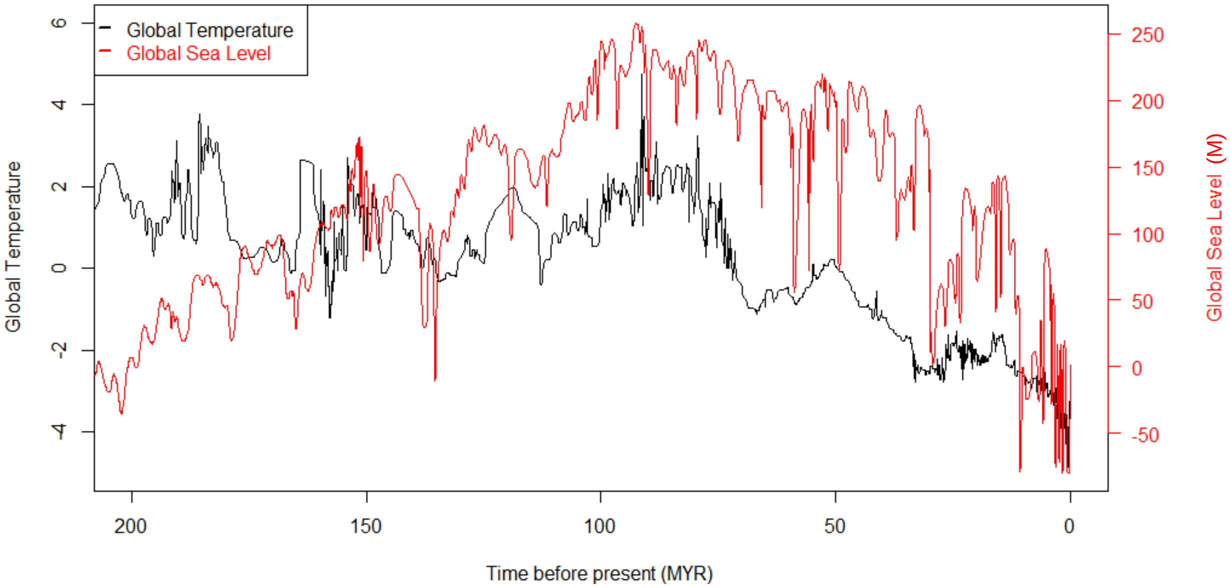


Figure 3: Fluctuations in global temperature (shown in black) and global sea level in metres (shown in red) over geological time. Proxy temperature data derived from Veizer *et al.*, (1999) and Zachos (2001). The temperature proxy is a ratio, giving the relative measure of temperature through time. Global sea level data derived from Haq, Hardenbol and Vail (1987).

relied on external heat sources to maintain their body temperature (Watanabe, 2005). However, research suggests that some pseudosuchian groups (e.g. Archosaurs) might have showed signs of endothermy (Seymour *et al*., 2004). Mannion *et al*. (2015) also correlated a significant biodiversity decline in terrestrial taxa with decreasing temperatures in northern temperate regions.

When it comes to sea levels, speciation in terrestrial taxa is favored by low sea levels (Fig. 1B). However, despite sea levels declining in the past 100 MYR (Fig. 3), speciation rate has not increased (Fig. 4). Potentially, of the two environmental variables tested, temperature plays a more significant role in determining the fate of Pseudosuchia evolution.

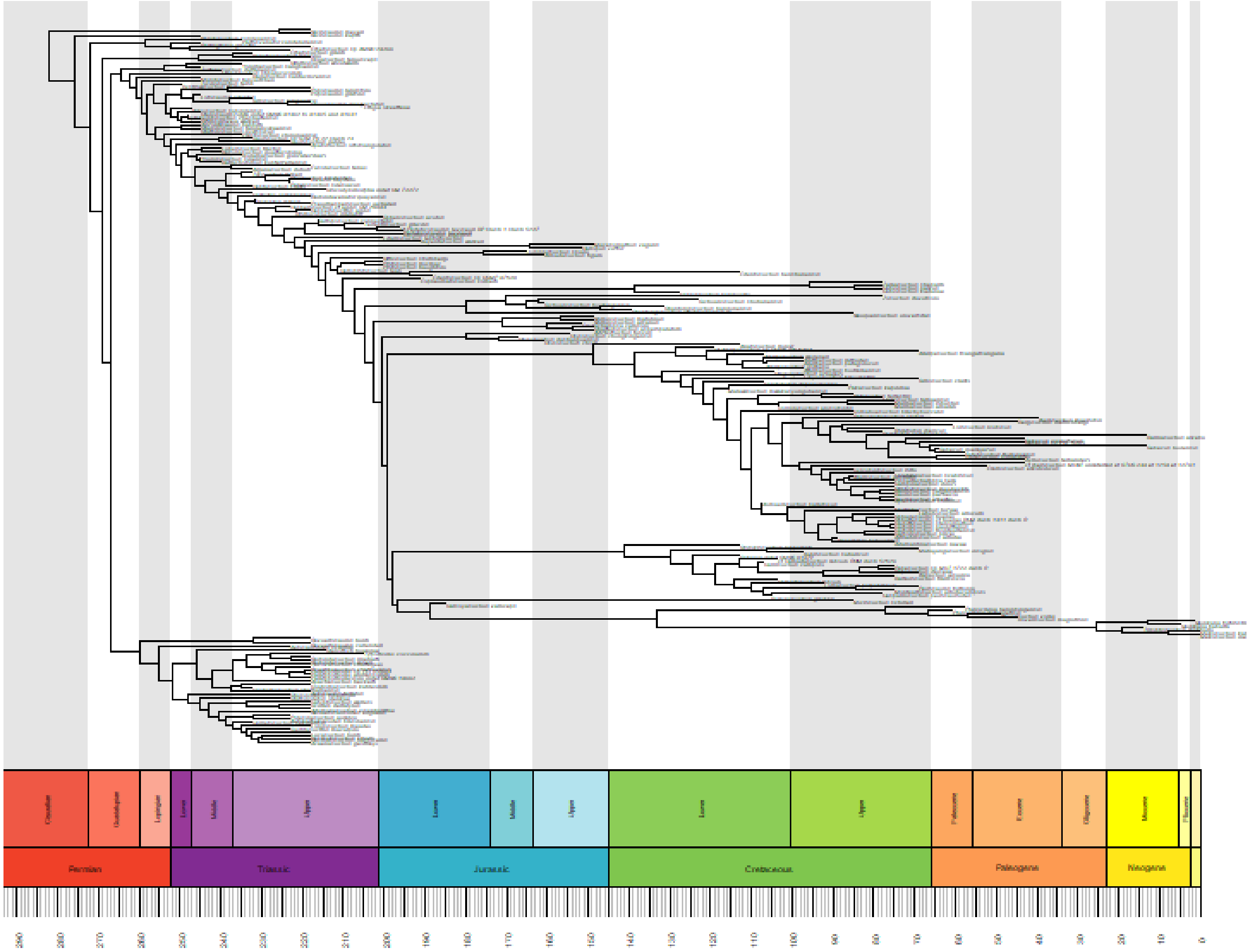


Figure 4: Phylogenetic tree of 207 terrestrial Pseudosuchia taxa scaled over geological time (MYR). Horizontal axis indicates the evolutionary history of the group over 250 million years and is divided into periods (Permian in red, Triassic in purple, Jurassic in blue, Cretaceous in green, Paleogene in orange, and Neogene in yellow) and epochs. Figure created using geoscale package (Mark, 2015) and data from Payne *et al.* (in prep).

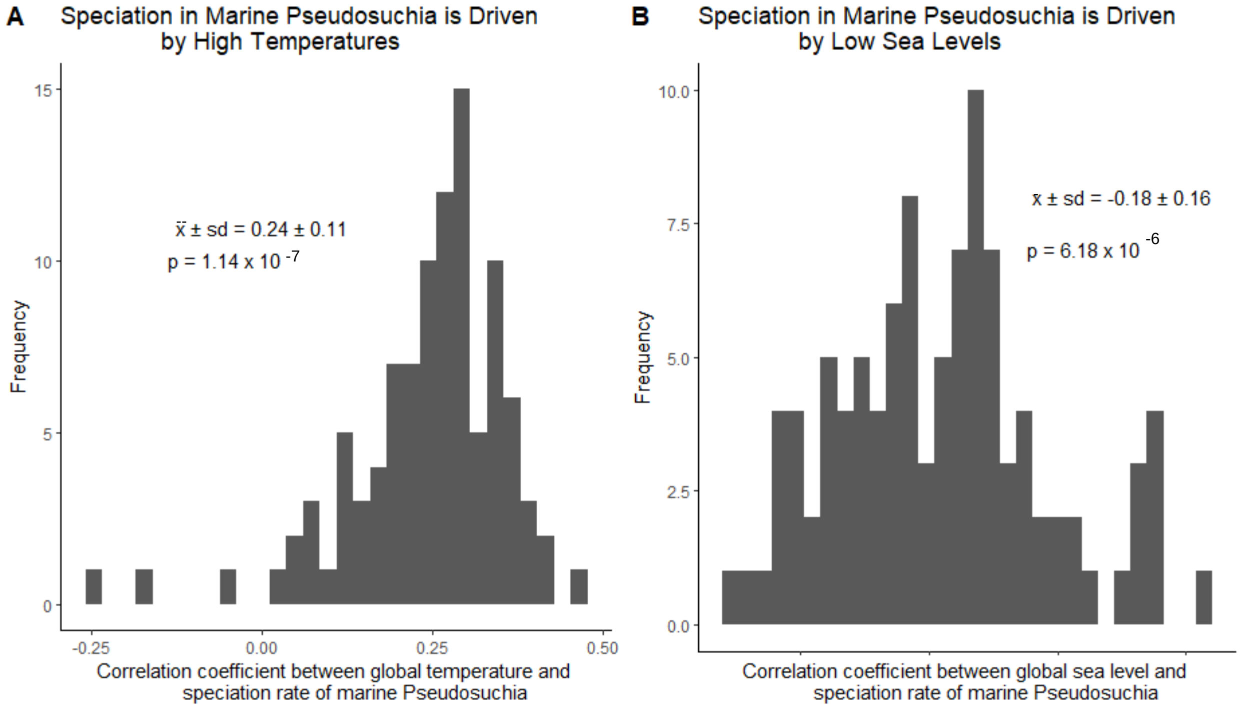
How did global temperature and sea level affect speciation rate in marine Pseudosuchia?

Method and Statistical Testing

Similarly to terrestrial taxa, a phylogenetic tree scaled over geological time was constructed for marine Pseudosuchia. Speciation rates were extracted and correlated with global sea level and temperature data. The significance of the relationships found was determined using a Wilcoxon signed-rank test.

Result

There is a statistically significant positive correlation between global temperature and speciation rate of marine taxa (Wilcoxon signed-rank test on correlation coefficient, $\bar{x} \pm sd = 0.24 \pm 0.14$, $p = 1.14\times10^{-7}$; Fig. 5A) and a significant negative relationship between sea level and speciation rate (Wilcoxon signed-rank test, $\bar{x} \pm sd = -0.18 \pm 0.16$, $p = 6.18\times10^{-6}$; Fig. 5B). Hence, speciation for marine Pseudosuchia taxa is driven by higher temperatures and low sea levels.



<u>Figure 1</u>: Histogram A shows the frequency of the correlation coefficient between global temperature and speciation rate of marine Pseudosuchia. Histogram B shows the frequency of the correlation coefficient between global sea level and speciation rate of marine Pseudosuchia. Wilcoxon signed-rank tests revealed that speciation in marine Pseudosuchia taxa is driven by high temperatures and low sea levels. Mean correlation coefficient ($\bar{x}$), standard deviation (sd) and the associated p-values (p) are shown.

Interpretation

Similar patterns to terrestrial taxa can be observed in speciation of marine Pseudosuchia. For example, the largest speciation events took place between the Lower Jurassic and Upper Cretaceous Period (Fig. 6) while temperature during those periods was high (Fig. 3). In fact, marine speciation rate peaked at 180 MYR and continued to be relatively high up until 150 MYR (Fig. 2) when both environmental conditions were at its optimum levels: temperature was high and sea level was at its lower point (Fig. 3). Our results are also supported by Martin *et al*. (2014) who found that marine crocodylomorphs diversified the most in warm temperatures. Again, that could be attributed to their ecothermic nature as in terrestrial taxa. A similar study on marine biodiversity (Mannion *et al*., 2015) found that there is no correlation between temperature and biodiversity for marine Pseudosuchia, instead they suggested that sea level changes are more important. That is because sea level changes affect coastal paleogeography through continental flooding, which not only affected marine biodiversity, but also semi-aquatic or terrestrial Pseudosuchia living near the shore.

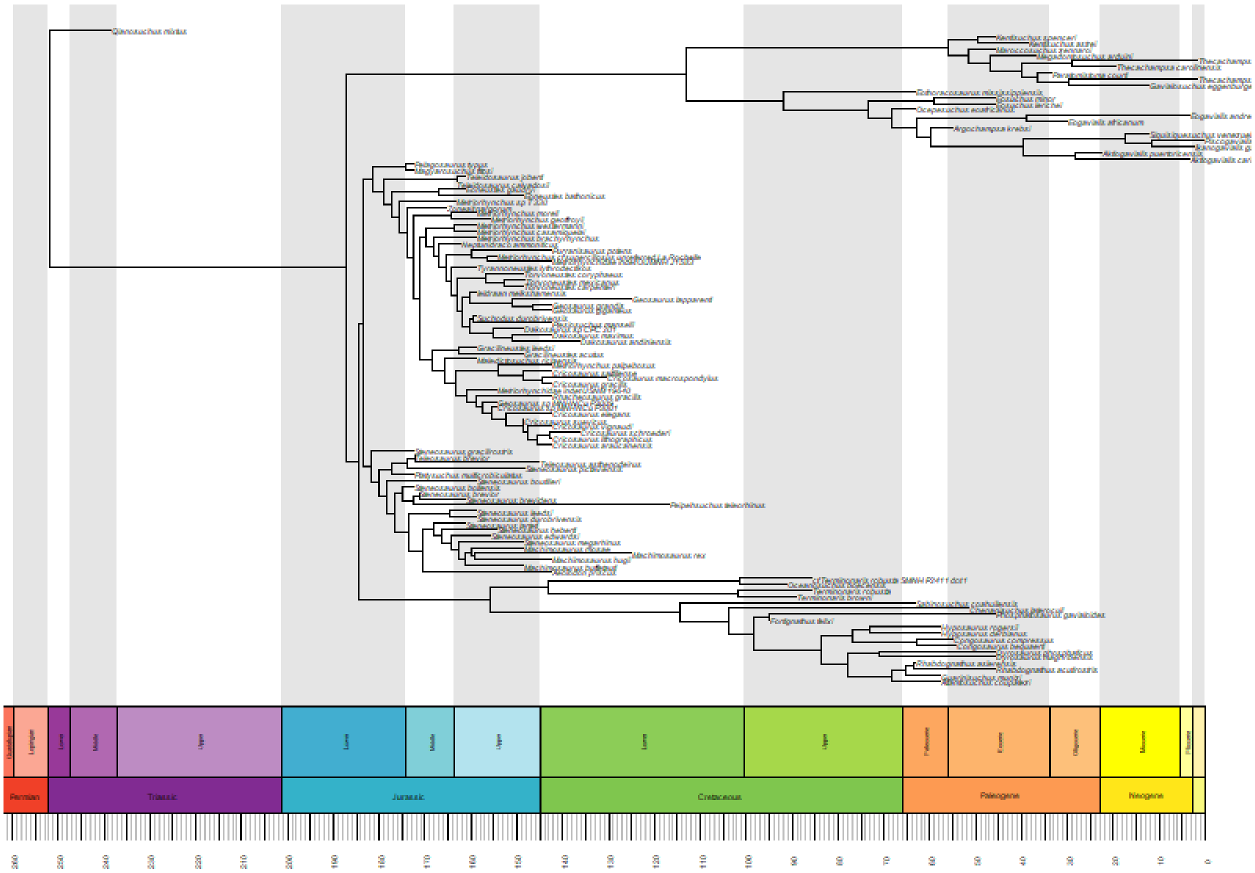


Figure 2: Phylogenetic tree of 108 marine Pseudosuchia taxa scaled over geological time (MYR). Horizontal axis indicates the evolutionary history of the group over 250 million years and is divided into periods (Permian in red, Triassic in purple, Jurassic in blue, Cretaceous in green, Paleogene in orange, and Neogene in yellow) and epochs. Figure created using geoscale package (Mark, 2015) and data from Payne *et al.* (in prep).

CONCLUSION

All things considered, both global temperature and sea level fluctuations had an important impact on the evolution of Pseudosuchia. It seems that high temperatures, however, might play a more important role in driving the speciation rate of both marine and terrestrial taxa. However, at too high temperatures aridity could become a limiting factor (Mannion *et al.*, 2015). Marine taxa could be further investigated by exploring the effect of sea temperature on speciation rate, as seen in Martin *et al*.

(2014). Concurrently, speciation should also be investigated in relation to morphological characteristics that might have shaped the survival and evolution of certain taxa, for example, body size (Slavenko *et al.*, 2016). That is because investigating abiotic factors alone is not representative of the complex interactions in the real world.

Despite our findings, it is not yet possible to make any valid predictions about the fate of extant Pseudosuchia. The evolutionary trends found, linking temperature and sea level to speciation, took millions of years to occur. Therefore, it is unlikely that major differences in crocodilian speciation will take place in the near future. However, with global temperatures and sea levels on the rise we could expect an increase in crocodilian biodiversity in the long-term and an increase in their geographic distribution. Could that also mean a potential re-introduction of crocodylians into marine environments? Apart from natural causes, however, biodiversity decline could also be attributed to human intervention through poaching, recreational hunting and habitat loss.

REFERENCES

Alboukadel Kassambara (2020) ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. https://CRAN.R-project.org/package=ggpubr

Bambach RK; Knoll AH & Wang SC (December 2004) ‘Origination, extinction, and mass depletions of marine diversity’. *Paleobiology*. **30** (4): 522–542. doi:10.1666/0094-8373(2004)030<0522:OEAMDO>2.0.CO;2. ISSN 0094-8373.

Bambach, RK; Knoll, AH. & Wang, SC. (2004) ‘Origination, extinction, and mass depletions of marine diversity’, Paleobiology, 30(4), pp. 522–542. doi: 10.1666/0094-8373(2004)030<0522:OEAMDO>2.0.CO;2.

Butler, RJ; Brusatte, SL; Reich, M; Nesbitt, SJ; Schoch, RR; & Hornung JJ (2011) ‘The sail-backed reptile Ctenosauriscus from the latest Early Triassic of Germany and the timing and biogeography of the early archosaur radiation’, *PloS one*, 6(10), p. e25693. doi: 10.1371/journal.pone.0025693.

Carvalho, IS; de Gasparini, ZB; Salgado, L; de Vasconcellos, FM; & Marinho, TS (2010) ‘Climate’s role in the distribution of the Cretaceous terrestrial Crocodyliformes throughout Gondwana’, *Palaeogeography, palaeoclimatology, palaeoecology*, 297(2), pp. 252–262. doi: 10.1016/j.palaeo.2010.08.003.

H. Wickham (2016) ‘ggplot2: Elegant Graphics for Data Analysis’. Springer-Verlag New York.

Haq BU; Hardenbol J & Vail PR (1987) ‘The chronology of fluctuating sea level since the Triassic: Science’ 235 (4793): 1156–1167. https://science.sciencemag.org/content/235/4793/1156

Mannion, PD; Benson, RBJ; Carrano, MT; Tennant, JP; Judd, J & Butler RJ (2015) ‘Climate constrains the evolutionary history and biodiversity of crocodylians’, *Nature communications*, 6, p. 8438. doi: 10.1038/ncomms9438.

Mark AB (2015) geoscale: Geological Time Scale Plotting. R package version 2.0. https://CRAN.R-project.org/package=geoscale

Mark, AB & Graeme, TL (2014) strap: Stratigraphic Tree Analysis for Palaeontology. R package version 1.4. https://CRAN.R-project.org/package=strap

Martin, JE; Amiot, R; Lécuyer, C & Benton, MJ (2014) ‘Sea surface temperature contributes to marine crocodylomorph evolution’, *Nature communications*, 5, p. 4658. doi: 10.1038/ncomms5658.

Payne ARD; Lloyd GT; Mannion PD & Davis KE. (In prep) ‘Decoupling speciation and extinction reveals both abiotic and biotic drivers shaped 250 million years of diversity of crocodile-line archosaurs’.

Pol, D. and Gasparini, Z. (2009) ‘Skull anatomy of Dakosaurus andiniensis (Thalattosuchia: Crocodylomorpha) and the phylogenetic position of Thalattosuchia’, Journal of Systematic Palaeontology, 7(2), pp. 163–197. doi: 10.1017/S1477201908002605.

Rabosky DL; Michael, CG; Carlos, JA; Pascal, OT; Jeff, JS; Joseph, WB; Huateng, H & Joanna, GL (2014) 'BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees'. Methods in Ecology and Evolution 5:701-707.

Revell, LJ (2012) phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3 217-223. doi:10.1111/j.2041-210X.2011.00169.x

Seymour, RS; Bennett-Stamper, CL; Johnston, SD; Carrier, DR & Grigg, GC (2004) 'Evidence for endothermic ancestors of crocodiles at the stem of archosaur evolution', *Physiological and biochemical zoology: PBZ*, 77(6), pp. 1051–1067. doi: 10.1086/422766.

Slavenko, A; Tallowin, OJS; Itescu, Y; Raia, P & Meiri, S (2016) 'Late Quaternary reptile extinctions: size matters, insularity dominates', *Global ecology and biogeography: a journal of macroecology*, 25(11), pp. 1308–1320. doi: 10.1111/geb.12491.

Veizer, J. *et al.* (1999) '87Sr/86Sr, δ13C and δ18O evolution of Phanerozoic seawater', *Chemical geology*, 161(1), pp. 59–88. doi: 10.1016/S0009-2541(99)00081-9.

Watanabe, ME (2005) 'Generating Heat: New Twists in the Evolution of Endothermy', *Bioscience*, 55(6), pp. 470–475. doi: 10.1641/0006-3568(2005)055[0470:GHNTIT]2.0.CO;2.

Wu, XC, Sues, HD and Sun, A (1995) 'A plant-eating crocodyliform reptile from the Cretaceous of China', *Nature*, 376(6542), pp. 678–680. doi: 10.1038/376678a0.

Zachos J (2001) 'Trends, rhythms, and aberrations in global climate 65 ma to present'. Science 292: 686-693

SUPPLEMENTARY METHODS

```
########################################################################################

#### Analysis of speciation rate and global temperature and sea level in marine and terrestrial Pseudosuchia ####

###################################################################################


# Delete all previous objects

rm(list=ls())

# Set working directory

setwd("~/BigDataBiology")

# Download all data sets from https://www-users.york.ac.uk/~kd856/WorkshopData/


########################################################################################

             #### Plot global sea level and temperature data over geological time ####

########################################################################################


# We are going to plot two graphs of environmental data and combine them into one

# Load the two data sets

# 1. Global temperature proxy data

temperature <- read.csv("~/temperatureTimeSeries.csv", header=FALSE)

# 2. Global sea level proxy data

seaLevel <- read.csv("~/seaLevelTimeSeries.csv", header=TRUE)


# Graph 1: Temperature

# Prepare temperature data

# Run the following command to remove any autocorrelation that might affect our statistical tests later and save it as a new variable.
```

```
finaltemp <- smooth(smooth(temperature$V2))
# Adjust margins
par(mar=c(4, 5, 3, 5))
# Plot global temperature against geological time (time before present)
# Adjust "xlim=c(200,0)" so that modern time (time before present = 0) is on the right of the x axis.
plot(temperature$V1,finaltemp,
    xlab='Time before present (MYR)',
    ylab='Global Temperature',
    xlim=c(200,0),
    ylim=c(-5, 6),
    type = 'l',
    col="black")

# Run this line of code to combine two graphs
par(new = TRUE)

# Graph 2: Plot global sea level against geological time (time before present)
# Again, reverse the x axis by adjusting xlim
# set the same xlim (from 200 MYR till O) for the two plots to align because the sea level dataset has
data for up to 255 MYR.
plot(seaLevel,
    xlim=c(200,0),
    type = 'l',
    col="red",
    axes=FALSE,
    xlab="",
```

```
    ylab="")

# Make room for labels of the y axis for sea level and add measurements

mtext("Global Sea Level",side=4,col="red",line=4)

axis(4, ylim=c(-70,260), col="red",col.axis="red",las=1)


## Add Legend to the graph to distinguish which line represents global temperature and which global
sea level

legend("topleft",legend=c("Global Temperature","Global Sea Level"),

      text.col=c("black","red"),col=c("black","red"), xjust = 0.3,

      border=FALSE, lty=1, lwd=2, merge=TRUE, seg.len=0.6)


# Save plot as pdf

pdf(file = '~/TemperatureSeaLevel.pdf', paper='A4')



##########################################################################################

                              #### Terrestrial and Marine Phylogenetic Trees ####

##########################################################################################


# Install and load required packages

install.packages("phytools")

install.packages("strap")

install.packages("geoscale")

library(phytools)

library(strap)

library(geoscale)
```

```
# Load phylogeny data
tree <- read.tree("~/fossilCrocPhylogeny.tre")
# Load habitat data
habitatdata <- read.csv("~/HabitatData.csv", header=T, stringsAsFactors = FALSE)

# Now we need to make two subsets of data: one for marine taxa and one for terrestrial taxa

#### Terrestrial Taxa - Phylogeny ####
# Let's explore the terrestrial habitat first
# First we need to load in the habitat data.
# Use the subset() to extract only terrestrial taxa
TerrestrialTaxa <- subset(habitatdata, habitatdata$Habitat=='Terrestrial')$Taxon
head(TerrestrialTaxa)
# create phylogenetic tree with terrestrial taxa and plot
treeT <- keep.tip(tree, TerrestrialTaxa)
plot(treeT, cex=0.2)
# Identify number of taxa in the phylogeny
treeT$tip.label
# 207 taxa
# Now we can plot this phylogenetic tree but scaled to geological time
# Save as .tre file
write.tree(treeT, file = '~/TerrestrialTaxa.tre')
# Find root for plotting first by checking ages of nodes (nodeHeights)
lengthsT <- nodeHeights(treeT)
# Now let's find the biggest number, that's our root node
```

```
root.timeT <- max(lengthsT)
print(root.timeT)
# Set root for plotting
treeT$root.timeT <- root.timeT
# Grab our OTUs (Operational Taxonomic Units = number of taxa)
all_otus <- treeT$tip.label
# Create an empty matrix containing the taxa, this is required by strap package
all_otudates <- matrix(0, nrow = length(all_otus), ncol=2)
# Turn the matrix into a data frame
all_otudates <- data.frame(all_otudates)
# Set the row names to the taxa (OTUs)
row.names(all_otudates) <- all_otus
# Set column names to FAD (First Appearance Datum) and LAD (Last Appearance Datum)
colnames(all_otudates) <- c('FAD','LAD')
# Plot the phylogenetic tree
geoscalePhylo(treeT,ages=all_otudates,
        cex.tip=0.2,
        lwd=1,
        quat.rm=T,
        units=c("Period", "Epoch"),
        boxes="Epoch")
# Save figure as pdf
pdf(file = '~/TerrestrialPhylogeny.pdf', paper='A4')


#### Marine Taxa - Phylogeny ####
```

```r
# Repeat the same steps for Marine taxa

# Now we can use subset() to extract a list of just marine taxa.

MarineTaxa <- subset(habitatdata, habitatdata$Habitat=='Marine')$Taxon

# Create a phylogeny with marine taxa and plot it

treeM <- keep.tip(tree, MarineTaxa)

plot(treeM, cex=0.2)

# Identify number of taxa in the phylogeny

treeM$tip.label

# 108 marine taxa

# Now we can plot this phylogenetic tree but scaled to geological time

# Save as .tre file

write.tree(treeM, file = '~/MarineTaxa.tre')

# Find root for plotting first by checking ages of nodes (nodeHeights)

lengthsM <- nodeHeights(treeM)

# Now let's find the biggest number, that's our root node

root.timeM <- max(lengthsM)

print(root.timeM)

# Set root for plotting

treeM$root.timeM <- root.timeM

# grab our OTUs (Operational Taxonomic Units = number of taxa)

all_otusM <- treeM$tip.label

# Create an empty matrix containing the taxa, this is required by strap

all_otudatesM <- matrix(0, nrow = length(all_otusM), ncol=2)

# Turn the matrix into a data frame

all_otudatesM <- data.frame(all_otudatesM)
```

```
# set the row names to the taxa (OTUs)

row.names(all_otudatesM) <- all_otusM

# set column names to FAD (First Appearance Datum) and LAD (Last Appearance Datum)

colnames(all_otudatesM) <- c('FAD','LAD')

# plot the data

geoscalePhylo(treeM,ages=all_otudatesM,

        cex.tip=0.3,

        lwd=1,

        quat.rm=T,

        units=c("Period", "Epoch"), boxes="Epoch")

# Save figure as pdf

pdf(file = '~/MarinePhylogeny.pdf', paper='A4')



#############################################################################################

                    #### Speciation Rates of Terrestrial and Marine Taxa ####

#############################################################################################

# Install package

install.packages("BAMMtools")

# Load installed packages

library(BAMMtools)

library(phytools)

# Load Pseudosuchia diversification data and remove the first 10% of the samples to account for the burn-in phase of the analysis using burnin=0.1

edata <- getEventData(tree, eventdata ='~/fossilCrocDiversificationData.txt', burnin=0.1)
```

```
#### Terrestrial Speciation ####
# First we are going to extract speciation rates for terrestrial taxa
# We need to make a subtree of only terrestrial taxa
streeTerrestrial <- subtreeBAMM(edata, tips=TerrestrialTaxa)
# It s phylogenetic tree with 207 tips and 206 internal nodes.
# Extract rates through time for terrestrial taxa
# rates through time (rtt) Terrestrial by using the Terrestrial subtree
rtt_T <- getRateThroughTimeMatrix(streeTerrestrial)
# Explore speciation rates (lamda)
rtt_T$lambda

#### Marine Speciation ####
# We need to make a subtree of only marine taxa to extract speciation rates
streeMarine <- subtreeBAMM(edata, tips=MarineTaxa)
streeMarine
# It is a phylogenetic tree with 108 tips and 107 internal nodes.
# Extract rates through time for marine taxa
# rates through time (rtt) Marine by using the Marine subtree
rtt_M <- getRateThroughTimeMatrix(streeMarine)
# Explore speciation rates (lamda)
rtt_M$lambda

#### Plot Speciation rates of Marine and Terrestrial Taxa ####
# Make the two plots and distinguish the graph lines by colour for marine and terrestrial taxa
# Plot speciation rate over geological time (time before present in MYR)
```

```
plotT_speciation <- plotRateThroughTime(streeTerrestrial, ratetype='speciation',
                        avgCol="blue",
                        ylim=c(0,0.5),
                        cex.axis=1,
                        intervalCol='blue',
                        intervals=c(0.05, 0.95),
                        opacity=0.1)
# Run add = TRUE to combine the two plots in one
plotM_speciation <- plotRateThroughTime(streeMarine, ratetype='speciation',
                        avgCol="green",
                        ylim=c(0,0.5),
                        cex.axis=1,
                        intervalCol='green',
                        add = TRUE,
                        intervals=c(0.05, 0.95),
                        opacity=0.1)
# Add a legend
legend('topright', legend=c("Terrestrial Taxa", "Marine Taxa"), col = c('blue', 'green'), border=FALSE, lty=1, lwd=2, merge=TRUE, seg.len=0.6)


# Save the graph to pdf
pdf("~/SpeciationRates.pdf")


##############################################################################################
                                        #### Statistical Tests ####
##############################################################################################
```

```
# Install and load packages
install.packages("ggplot2")
library (ggplot2)

# We are going to do a DCCA-based (Detrended Cross Correlation Analysis) test which calculates the correlation coefficient between two time series.
DCCA <- function(x,y,s){
  xx<-cumsum(x)
  yy<-cumsum(y)
  t<-1:length(xx)
  F2sj_xy<-runif(floor(length(xx)/s))
  F2sj_xx<-F2sj_xy
  F2sj_yy<-F2sj_xy
  for(ss in seq(1,(floor(length(xx)/s)*s),by=s)){
    F2sj_xy[(ss-1)/s+1]<-sum((summary(lm(xx[ss:(ss+s-1)]~t[ss:(ss+s-1)]))$residuals)*(summary(lm(yy[ss:(ss+s-1)]~t[ss:(ss+s-1)]))$residuals))/(s-1)
    F2sj_xx[(ss-1)/s+1]<-sum((summary(lm(xx[ss:(ss+s-1)]~t[ss:(ss+s-1)]))$residuals)*(summary(lm(xx[ss:(ss+s-1)]~t[ss:(ss+s-1)]))$residuals))/(s-1)
    F2sj_yy[(ss-1)/s+1]<-sum((summary(lm(yy[ss:(ss+s-1)]~t[ss:(ss+s-1)]))$residuals)*(summary(lm(yy[ss:(ss+s-1)]~t[ss:(ss+s-1)]))$residuals))/(s-1)
  }
  rho<-mean(F2sj_xy)/sqrt(mean(F2sj_xx)*mean(F2sj_yy))
  return(c(rho,1/sqrt(length(xx)),1-pnorm(abs(rho),mean=0,sd=1/sqrt(length(xx)))))
}

# We need the time to begin from the past instead of the future.
```

```
# This is done by substracting the maximum from each time value, and then turn it into an absolute value (remove minus sign).
times = abs(rtt_T$times-max(rtt_T$times))
times
# this reversed the time


# set a number of samples from the whole data set that will be used to extract speciation values
# we used 100
numberOfSamples = 100


# Now we will need to perform four tests for each condition to find the correlation coefficient between speciation rate in terrestrial/marine taxa and sea level/temperature.
# The four tests are:
# 1. Calculate correlation coefficients of temperature data and speciation rate of terrestrial taxa.
# 2. Calculate correlation coefficients of sea level data and speciation rate of terrestrial taxa.
# 3. Calculate correlation coefficients of temperature data and speciation rate of marine taxa.
# 4. Calculate correlation coefficients of sea level data and speciation rate of marine taxa.


                    #### 1. Terrestrial Speciation - Temperature Correlation ####
# set.seed(1) to obtain the same results if you re-run the code (optional)
set.seed(1)
# extract the time for terrestrial taxa
times = abs(rtt_T$times-max(rtt_T$times))
# determine the number of simulations by adjusting the code
# rtt_T used for Terrestrial
# lamda used for speciation
```

```
numberOfSims = length(rtt_T$lambda)/length(rtt_T$times)

samples = sample(1:numberOfSims, numberOfSamples, replace = FALSE)

# create a data set that will save the correlation coefficient (cor) of speciation (s) with temperature (temp) in Terrestrial taxa (T)

cors_temp_T <- rep(NA, numberOfSamples)

# set count to 1

count = 1

# Run the following code in one go

for (i in 1:numberOfSims ) {

  # Is it one of our samples?

  if (i %in% samples){

    # If yes, do the correlation.

    # Start by interpolating the data. We do this because the two time series are different lengths. We need them to start and end at the same times and we need the points in between to match up in order to carry out the correlation.

    # This line takes our speciation rates (lambda) and the corresponding times and interpolates the lambda onto the temperature times. We can only do a correlation for the temperature data that we have so if we have more lambda times than temperature we cannot use them

    interpdiv = approx(times, rtt_T$lambda[i,], temperature$V1, method='linear', rule=1)

    # Here we check whether there is a lambda for every temperature time, if not it's left as NA

    end = which(is.na(interpdiv$y))

    # If we have no NAs, ie. there is a time in the temperature time series for each lambda, we just use the interpolation as already calculated

    if (length(end) == 0) {

      div_rates = interpdiv$y

      ft = finaltemp

      # Otherwise, we only grab and use the times that have both lambda and temperature data

    } else {
```

```
   div_rates = interpdiv$y[-end]

   ft = finaltemp[-end]

  }

  # Now do the correlation using the interpolated data

  c = DCCA(as.numeric(unlist(div_rates)),as.numeric(unlist(ft)),length(ft)/10)

  # Store the correlation co-efficient

  cors_temp_T[count] = c[1]

  # Increase your count by 1 ready for the next correlation

  count = count+1

 }

}


# plot the correlation coefficients in a histogram

plotTST=qplot(cors_temp_T, geom="histogram", bins=30, main = "Speciation in Terrestrial Pseudosuchia
is Driven

        by High Temperatures")

plotTST

# now edit the plot by adding labels to the y and x axis, adding a main title to describe the relationship
found, and present the main findings of the wilcoxon statistical test on the plot (statistical test will
follow after this, so edit the text after performing the test)

plotTST2 = plotTST + theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
panel.background = element_blank(), axis.line = element_line(colour = "black")) +

 labs(x = "Correlation coefficient between global temperature and

    speciation rate of terrestrial Pseudosuchia", y = "Frequency") +

 annotate("text", x = 0.45, y = 9,

      label = "paste(\"x ± sd = 0.27 ± 0.10\")", parse = TRUE) +

 annotate("text", x = 0.42, y = 8,
```

```
        label = "paste(\"p = 6.03 x 10\")", parse = TRUE)
plotTST2


# Now, we will perform a signed-rank test
# First, open a .txt file in your directory where the summary statistics will be printed
sink(file="~SpeciationTerrestrialTemperatureStats.txt")
# Calculate and print to file some summary statistics, e.g., median
print(summary(cors_temp_T))
# Print the 95% confidence intervals
print(quantile(cors_temp_T, c(0.025, 0.975)))
# calculate standard deviation
sd(cors_temp_T)
# Do the statistical test
statsTST <- (wilcox.test(cors_temp_T, mu=0.0, paired = FALSE))
# extract the p value
statsTST$p.value
# Close the sink to stop writing in this .txt file.
sink()
# Now we can use these summary statistics and add them in the plot as seen previously.



                    #### 2. Terrestrial Speciation - Sea Level Correlation ####
# Repeat the exact same steps but replace temperature data with sea level.
set.seed(1)
times = abs(rtt_T$times-max(rtt_T$times))
```

```
numberOfSims = length(rtt_T$lambda)/length(rtt_T$times)

samples = sample(1:numberOfSims, numberOfSamples, replace = FALSE)

# create dataframe to store correlation coefficients

cors_seaLevel_T <- rep(NA, numberOfSamples)

count = 1


# use the same code and replace with cors_seaLevel_T and seaLevel

for (i in 1:numberOfSims ) {

  if (i %in% samples){

    interpdiv = approx(times, rtt_T$lambda[i,], seaLevel$Age, method='linear', rule=1)

    end = which(is.na(interpdiv$y))

    if (length(end) == 0) {

      div_rates = interpdiv$y

      ft = seaLevel$SL

    } else {

      div_rates = interpdiv$y[-end]

      ft = seaLevel$SL[-end]

    }

    c = DCCA(as.numeric(unlist(div_rates)),as.numeric(unlist(ft)),length(ft)/10)

    cors_seaLevel_T[count] = c[1]

    count = count+1

  }

}


# plot the correlation coefficients in a histogram and edit it
```

```
plotTSSL=qplot(cors_seaLevel_T, geom="histogram", bins=30, main = "Speciation in Terrestrial Pseudosuchia is Driven

        by Low Sea Levels")

plotTSSL

plotTSSL2 = plotTSSL + theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(), panel.background = element_blank(), axis.line = element_line(colour = "black"))+

  labs(x = "Correlation coefficient between global sea level and

      speciation rate of terrestrial Pseudosuchia", y = "Frequency") +

  annotate("text", x = -0.20, y = 11,

        label = "paste(\"x ± sd = -0.43 ± 0.12\")", parse = TRUE) +

  annotate("text", x = -0.25, y = 10,

        label = "paste(\"p = 6.03 x 10\")", parse = TRUE)

plotTSSL2


# Repeat the same steps for the statistical test

sink(file="~/SpeciationTerrestrialSeaLevelStats.txt")

print(summary(cors_seaLevel_T))

print(quantile(cors_seaLevel_T, c(0.025, 0.975)))

sd(cors_seaLevel_T)

statsTSSL <- (wilcox.test(cors_seaLevel_T, mu=0.0, paired = FALSE))

statsTSSL$p.value

sink()


#### Combine the plots for Terrestrial Taxa ####

# Now that we have created the two plots (plotTST2 and plotTSSL2) we are going to combine them and plot them side by side into one figure

# install the required package
```

```
install.packages("ggpubr")

# load the package

library(ggpubr)

# re-plot the two histograms

plotTST2

plotTSSL2

# combine them into one figure and label accordignly

ggarrange(plotTST2, plotTSSL2 + rremove("x.text"),

        labels = c("A", "B"),

        ncol = 2, nrow = 1)

# Save new figure as pdf

pdf(file = '~/CorrelationTerrestrial.pdf', paper='A4')
```

```
                              #### 3. Marine Speciation - Temperature Correlation ####

# Repeat the same steps as in the previous examples

set.seed(1)

# replace rtt_T with rtt_M as we are looking at marine taxa now

times = abs(rtt_M$times-max(rtt_M$times))

numberOfSims = length(rtt_M$lambda)/length(rtt_M$times)

samples = sample(1:numberOfSims, numberOfSamples, replace = FALSE)

# Name the data frame accordingly

cors_temp_M <- rep(NA, numberOfSamples)

count = 1
```

```
# run this code all together and replace rtt_T with rtt_M
for (i in 1:numberOfSims ) {
  if (i %in% samples){
    interpdiv = approx(times, rtt_M$lambda[i,], temperature$V1, method='linear', rule=1)
    end = which(is.na(interpdiv$y))
    if (length(end) == 0) {
      div_rates = interpdiv$y
      ft = finaltemp
    } else {
      div_rates = interpdiv$y[-end]
      ft = finaltemp[-end]
    }
    c = DCCA(as.numeric(unlist(div_rates)),as.numeric(unlist(ft)),length(ft)/10)
    cors_temp_M[count] = c[1]
    count = count+1
  }
}


# plot the correlation coefficients found in a historam
plotMST=qplot(cors_temp_M, geom="histogram", bins=30, main = "Speciation in Marine Pseudosuchia
is Driven
              by High Temperatures")
plotMST
plotMST2 = plotMST + theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  labs(x = "Correlation coefficient between global temperature and
```

```
    speciation rate of marine Pseudosuchia", y = "Frequency") +
  annotate("text", x = 0, y = 11,
      label = "paste(\"x ± sd = 0.24 ± 0.11\")", parse = TRUE) +
  annotate("text", x = -0.055, y = 10,
      label = "paste(\"p = 1.14 x 10\")", parse = TRUE)
plotMST2

# do statistical analysis
sink(file="~/SpeciationMarineTemperatureStats.txt")
print(summary(cors_temp_M))
print(quantile(cors_temp_M, c(0.025, 0.975)))
sd(cors_temp_M)
statsMST <- (wilcox.test(cors_temp_M, mu=0.0, paired = FALSE))
statsMST$p.value
sink()


                                   #### 4. Marine Speciation - Sea Level ####
# repeat the same steps
set.seed(1)
times = abs(rtt_M$times-max(rtt_M$times))
numberOfSims = length(rtt_M$lambda)/length(rtt_M$times)
samples = sample(1:numberOfSims, numberOfSamples, replace = FALSE)
# name the data frame accordingly
cors_seaLevel_M <- rep(NA, numberOfSamples)
```

```
count = 1

# run the same code but replace with sealevel data
for (i in 1:numberOfSims ) {
  if (i %in% samples){
    interpdiv = approx(times, rtt_M$lambda[i,], seaLevel$Age, method='linear', rule=1)
    end = which(is.na(interpdiv$y))
    if (length(end) == 0) {
      div_rates = interpdiv$y
      ft = seaLevel$SL
    } else {
      div_rates = interpdiv$y[-end]
      ft = seaLevel$SL[-end]
    }
    c = DCCA(as.numeric(unlist(div_rates)),as.numeric(unlist(ft)),length(ft)/10)
    cors_seaLevel_M[count] = c[1]
    count = count+1
  }
}


# plot the correlation coefficients found in a historam
plotMSSL=qplot(cors_seaLevel_M, geom="histogram", bins=30, main = "Speciation in Marine Pseudosuchia is Driven
               by Low Sea Levels")
plotMSSL
```

```
plotMSSL2 = plotMSSL + theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  labs(x = "Correlation coefficient between global sea level and
      speciation rate of marine Pseudosuchia", y = "Frequency") +
  annotate("text", x = 0.1, y = 8,
        label = "paste(\"x ± sd = -0.18 ± 0.16\")", parse = TRUE) +
  annotate("text", x = 0.04, y = 7,
        label = "paste(\"p = 6.18 x 10\")", parse = TRUE)
plotMSSL2

# Follow the same steps for the statistical test
sink(file="R Data/Crocodiles/SpeciationMarineSeaLevelStats.txt")
print(summary(cors_seaLevel_M))
print(quantile(cors_seaLevel_M, c(0.025, 0.975)))
summary(cors_seaLevel_M)
sd(cors_seaLevel_M)
mean(cors_seaLevel_M)
statsMSSL <- (wilcox.test(cors_seaLevel_M, mu=0.0, paired = FALSE))
statsMSSL$p.value
sink()
#### Combine the plots for Marine Taxa ####
# re-plot the two histograms
plotMST2
plotMSSL2
# combine them into one figure and label accordignly
ggarrange(plotMST2, plotMSSL2 + rremove("x.text"),
```

```
          labels = c("A", "B"),

          ncol = 2, nrow = 1)

# Save new figure as pdf

pdf(file = '~/CorrelationMarine.pdf', paper='A4')



########################################################################################

                                    #### Packages Used & Citations ###

########################################################################################

citation("phytools")

citation("strap")

citation("geoscale")

citation("BAMMtools")

citation("ggplot2")

citation("ggpubr")
```