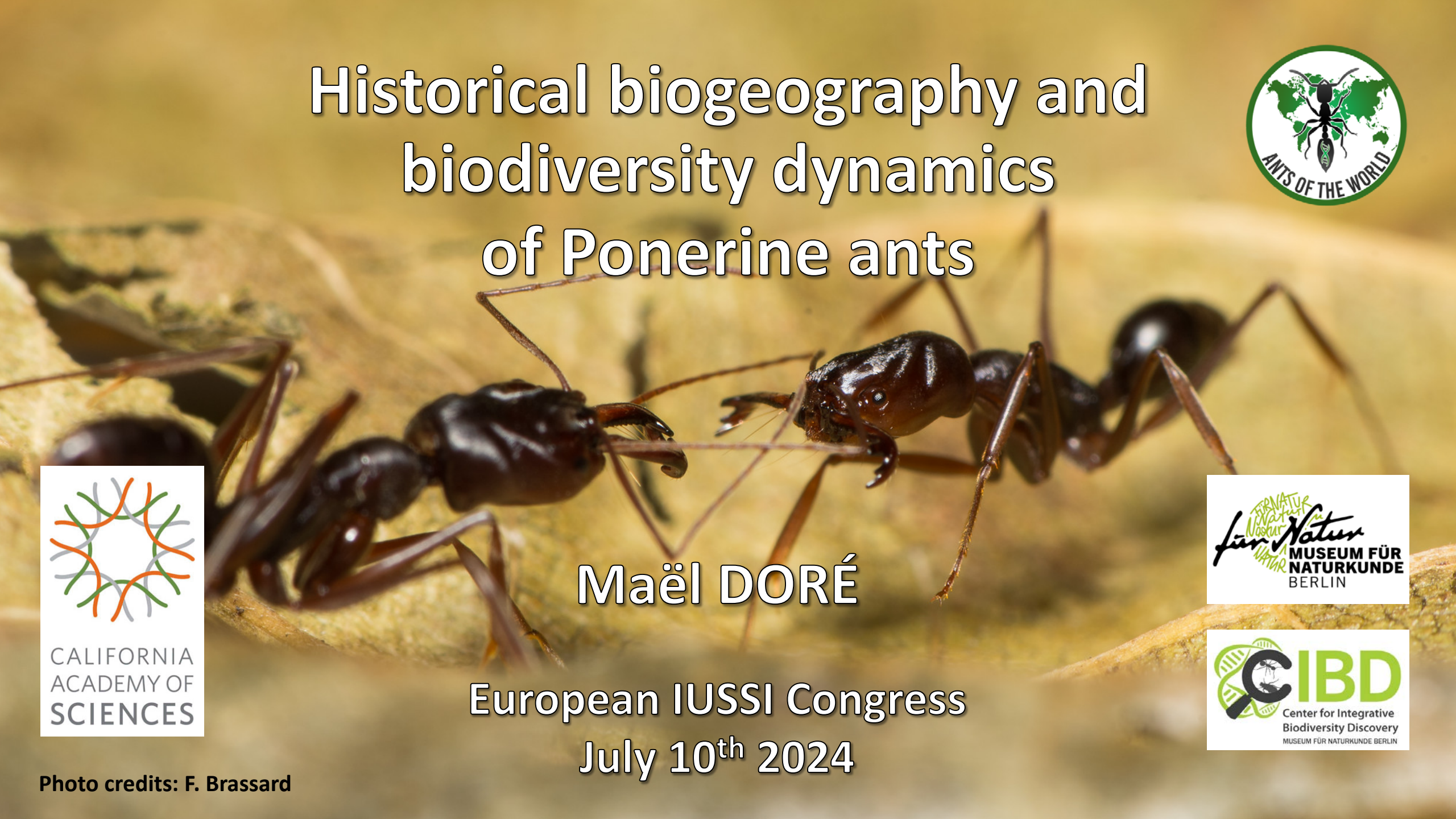
# Historical biogeography and biodiversity dynamics of Ponerine ants



Maël DORÉ

European IUSSI Congress
July 10th 2024

Photo credits: F. Brassard

# Ponerinae Subfamily

Top 3 ant subfamily for **diversity**
with 50 genera and ca. **1,400 species**

**Predatory** ants: include the **trap-jaw** ants (Anochetus, Odontomachus)

High diversity of **mandible shapes** related to **diet specialization**

Plesiomorphic robust morphology

Include the **largest ant species**: *Dinoponera gigantea*


| Formicinae | Myrmicinae | **Ponerinae** |
|---|---|---|
| 3,300 sp. | 7,200 sp. | **1,400 sp.** |
| 22% | 48% | **10%** |

# Ponerinae Subfamily

Top 3 ant subfamily for **diversity**
with 50 genera and ca. **1,400 species**

**Predatory** ants: include the **trap-jaw** ants
(Anochetus, Odontomachus)

High diversity of **mandible shapes**
related to **diet specialization**

Plesiomorphic robust morphology

Include the **largest ant species**:
*Dinoponera gigantea*

Photo Credits: Magdalena Sorger

# Ponerinae Subfamily

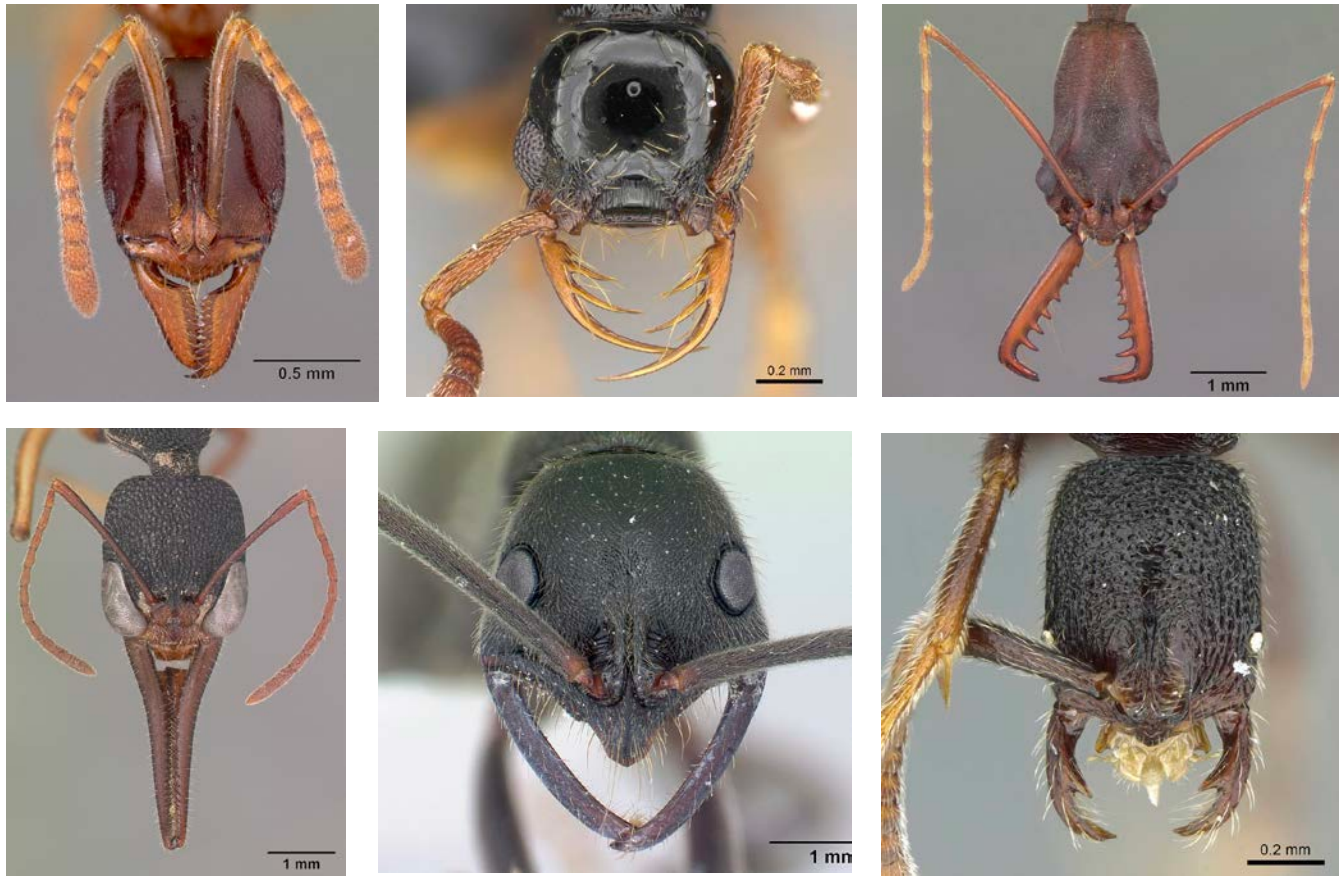
Top 3 ant subfamily for **diversity**
with 50 genera and ca. **1,400 species**

**Predatory** ants: include the **trap-jaw** ants
(Anochetus, Odontomachus)

High diversity of **mandible shapes**
related to **diet specialization**

Plesiomorphic robust morphology

Include the **largest ant species**:
*Dinoponera gigantea*




Source: AntWeb

# Ponerinae Subfamily

Top 3 ant subfamily for **diversity**
with 50 genera and ca. **1,400 species**

**Predatory** ants: include the **trap-jaw** ants
(Anochetus, Odontomachus)

High diversity of **mandible shapes**
related to **diet specialization**

Plesiomorphic robust morphology

Include the **largest ant species**:
*Dinoponera gigantea*

Photo Credits: Tatyana Kolesnikova

*Dinoponera gigantea* in Belem, Brazil
Up to 40 mm

# Current diversity

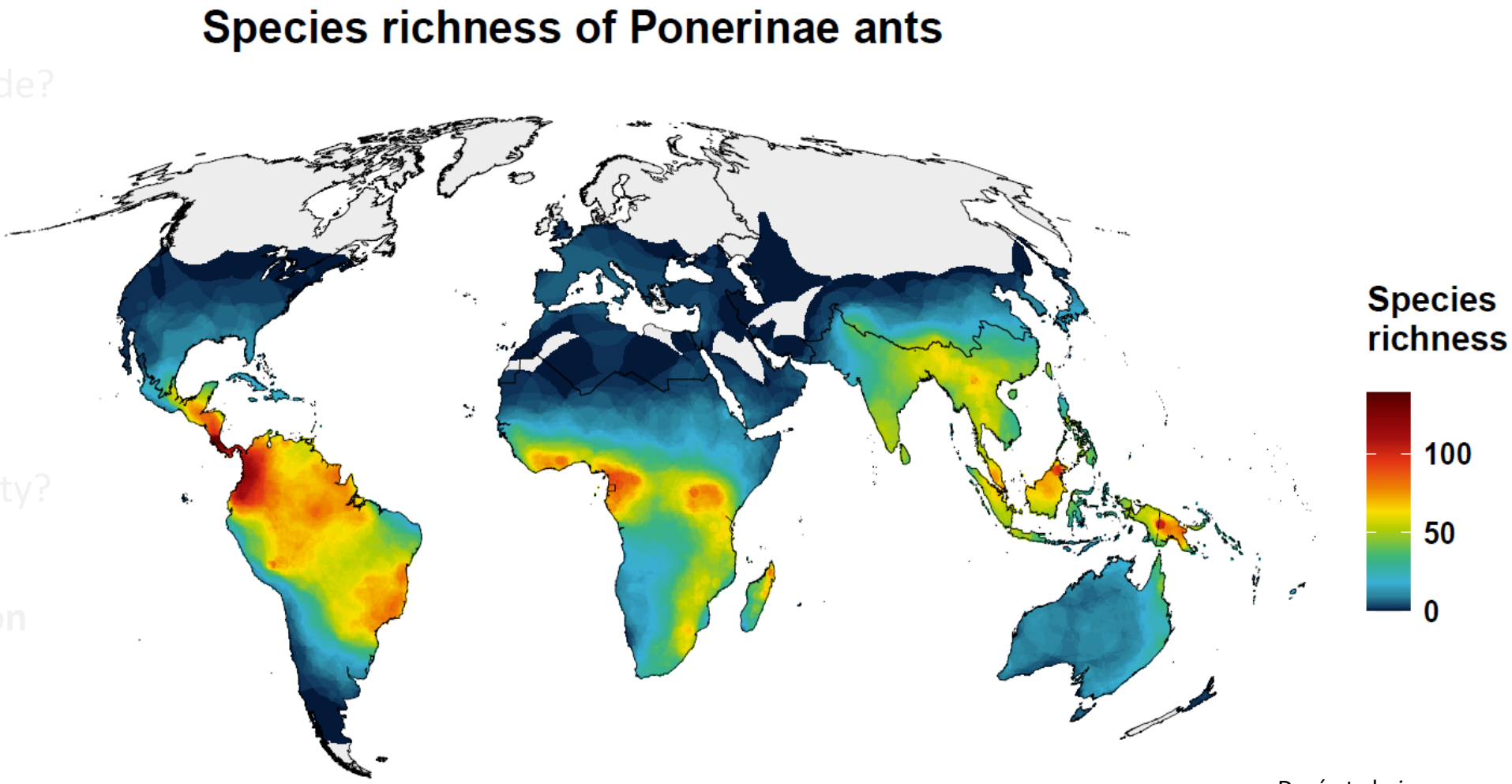
1/ **Origin** of the clade?

2/ Source of diversity?

***In situ*** **diversification**

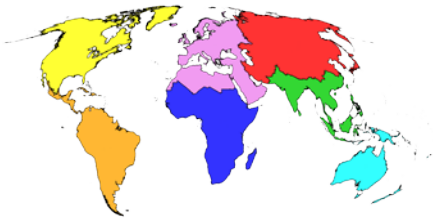
*vs.*

**Dispersal**




Doré et al., *in prep.*

# Current diversity



## Species richness of Ponerinae ants

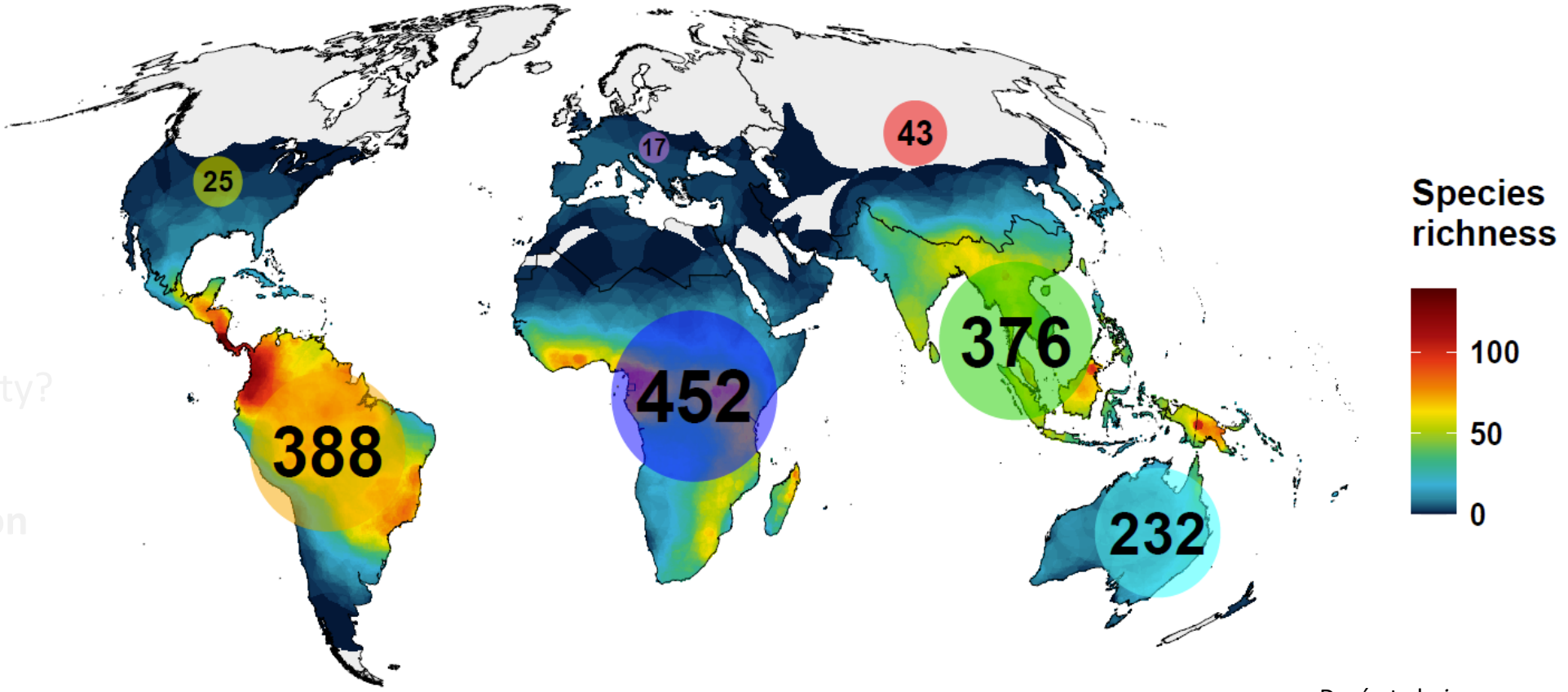
1/ **Origin** of the clade?

2/ Source of diversity?

***In situ*** **diversification**

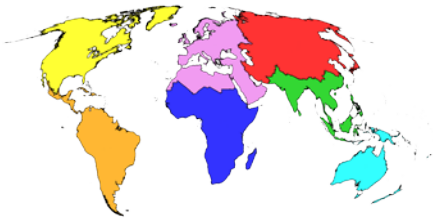
*vs.*

**Dispersal**




Doré et al., *in prep.*

# Current diversity



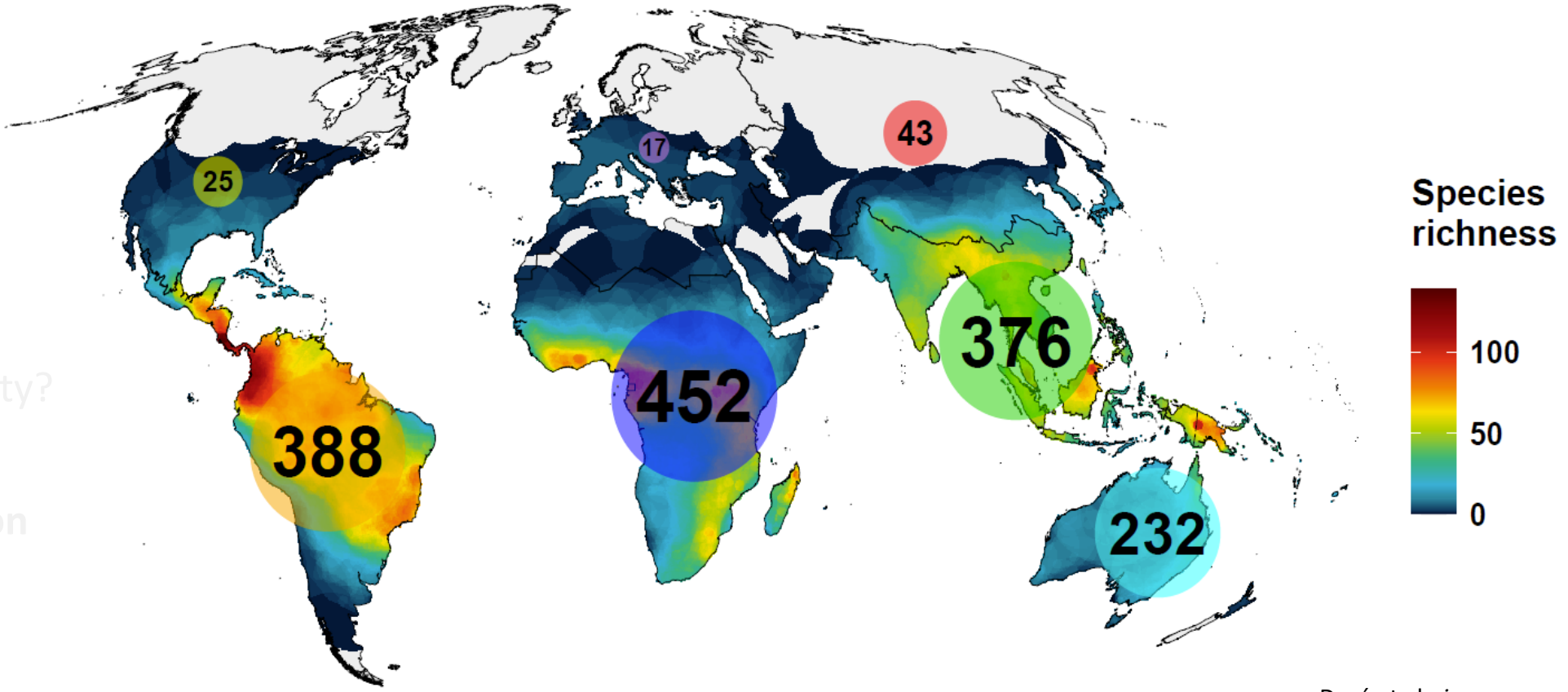
1/ **Origin** of the clade?

2/ Source of diversity?

*In situ* diversification

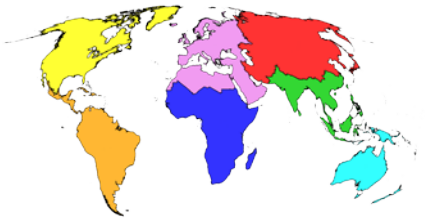
*vs.*

Dispersal




Doré et al., *in prep.*

# Current diversity



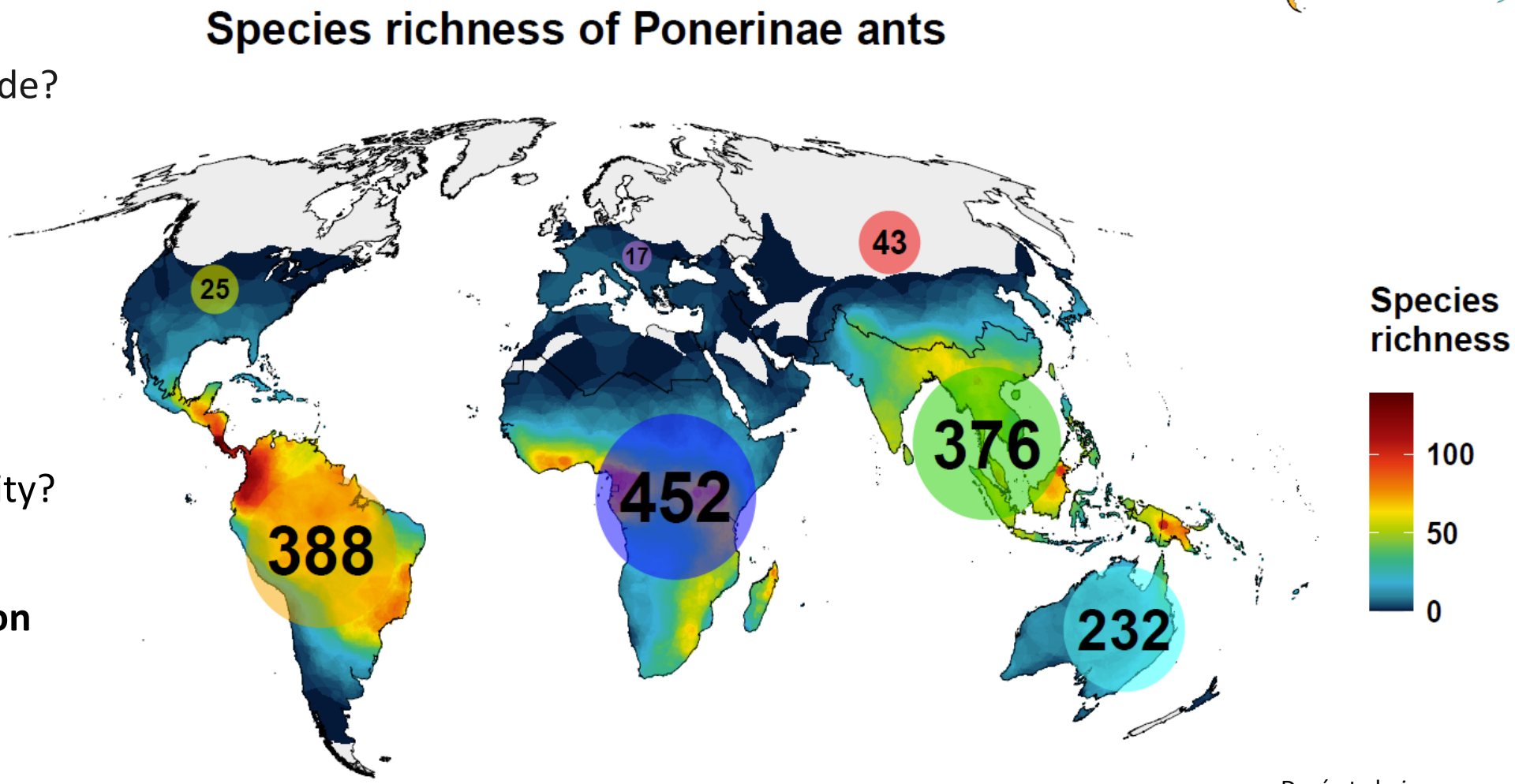
1/ **Origin** of the clade?

2/ Source of diversity?

***In situ* diversification**

*vs.*

**Dispersal**




Doré et al., *in prep.*

# Phylogenomic methods

**Taxa sampling**: 789 species (51.4%)
among all 50 Genera (100%)

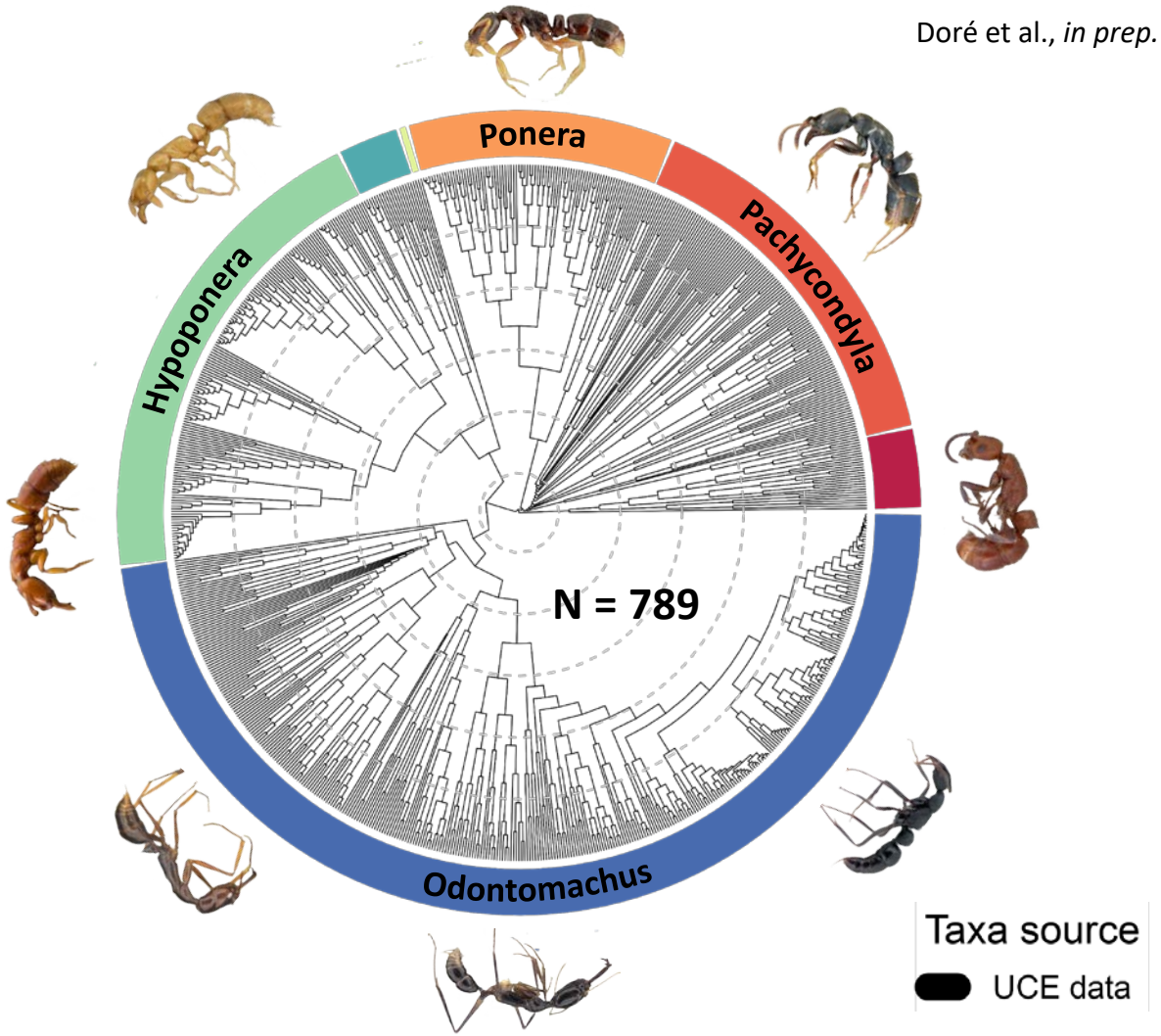
Molecular data: 2,365 **UCE loci** = 1.05 M sites

Phylogenetic inferences: ML in **IQ-Tree**
with **SWSC partitioning**

**Sensitivity analyses**:
Inference method x Loci filtering x
Partitioning scheme

**Grafting** of missing species:
Impute all 745 valid species with no UCE data using
available taxonomic and geographic information

Doré et al., *in prep.*




**New species-level phylogeny for Ponerinae**

# Phylogenomic methods

**Taxa sampling**: 789 species (51.4%)
among all 50 Genera (100%)

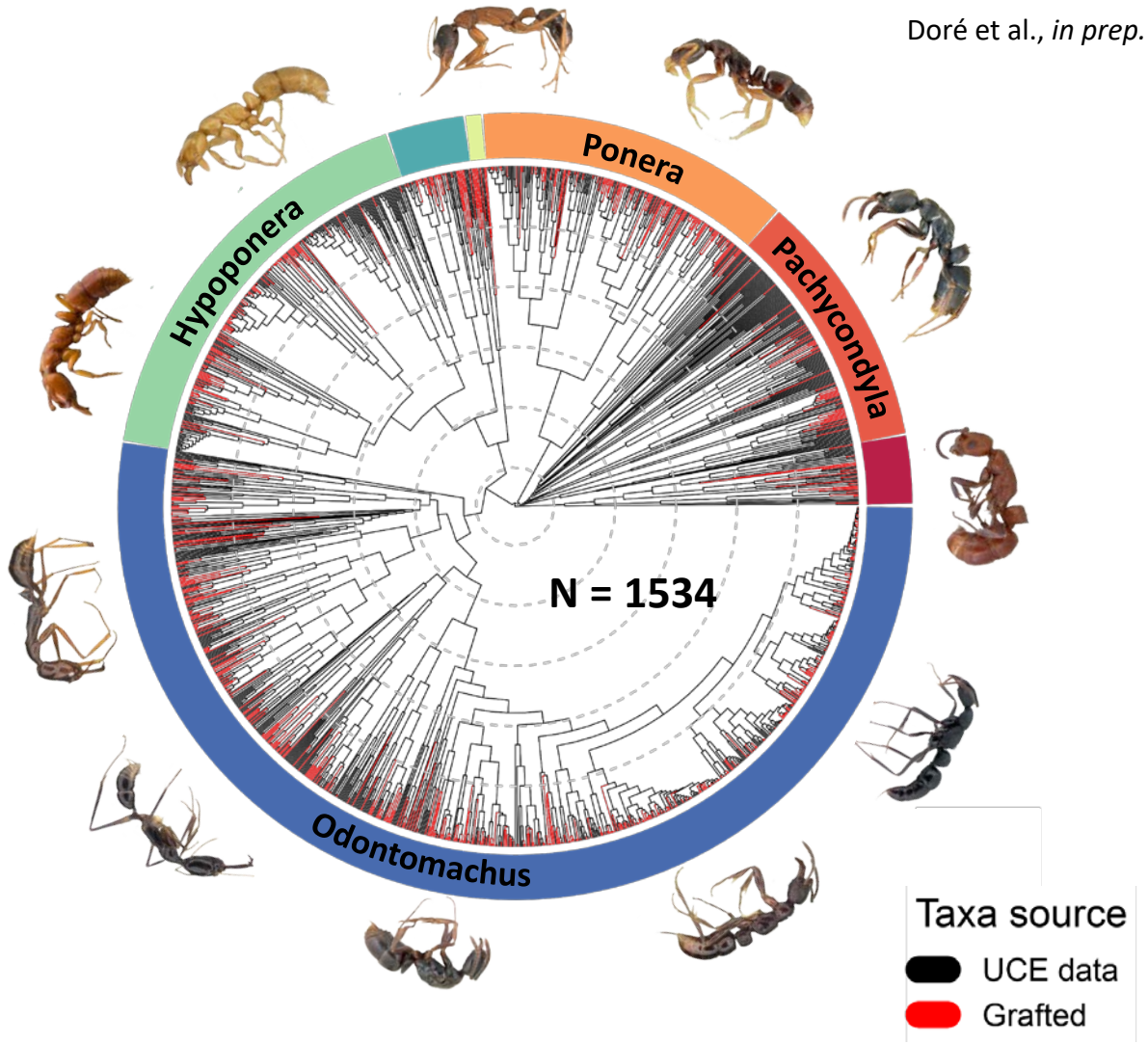
Molecular data: 2,365 **UCE loci** = 1.05 M sites

Phylogenetic inferences: ML in **IQ-Tree**
with **SWSC partitioning**

**Sensitivity analyses**:
Inference method x Loci filtering x
Partitioning scheme

**Grafting** of missing species:
Impute all 745 valid species with no UCE data using available taxonomic and geographic information

Doré et al., *in prep.*




**New species-level phylogeny for Ponerinae**

# Biogeographic data

**Occurrence sources:**

- AntWeb (62,301 records)
- GABI/AntMap (93,557 records)

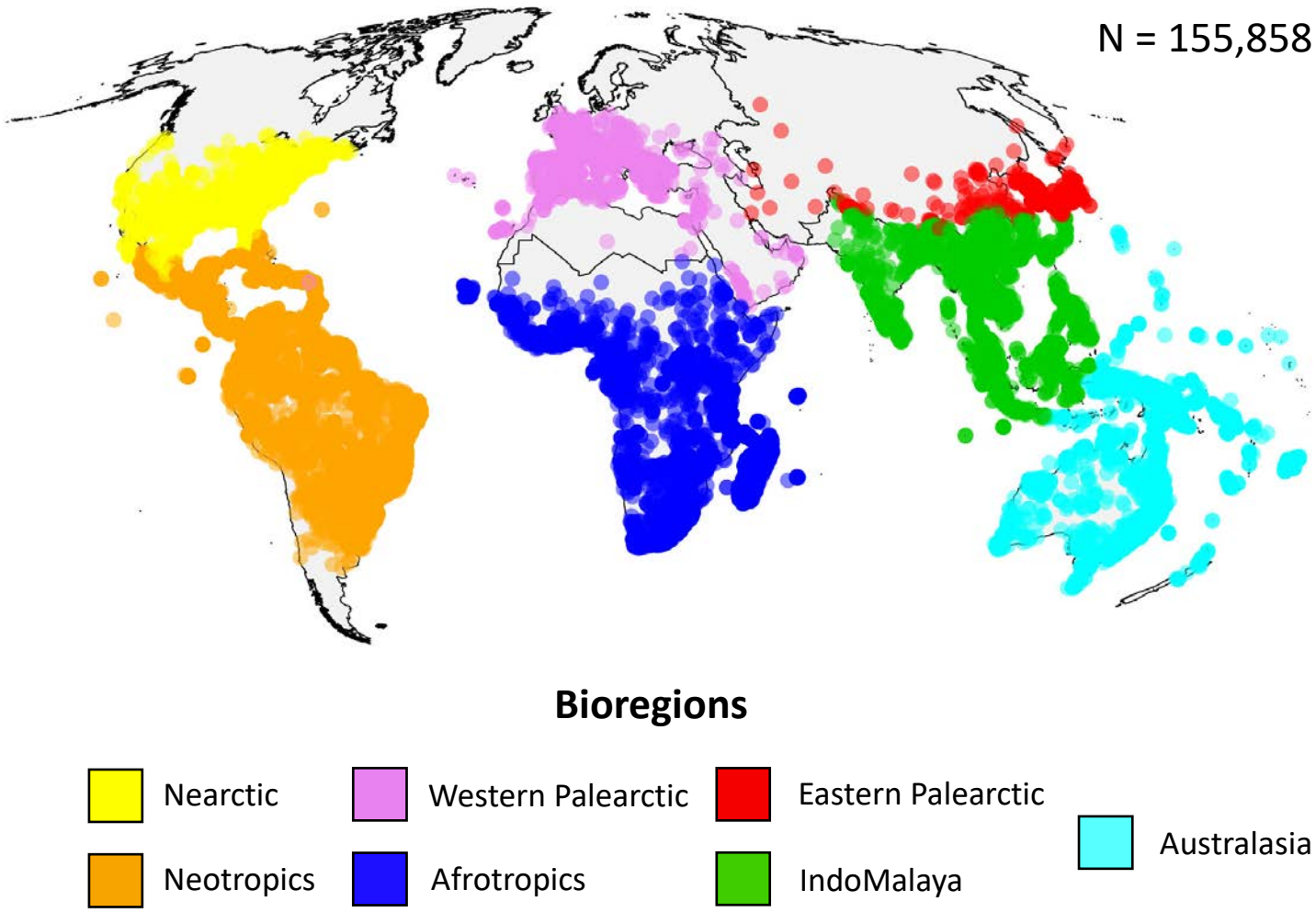
**Bioregions** as defined in AntWeb

**Data curation:**

- Automatic and visual checks
  - *CoordinateCleaner* (Zizka et al., 2019)
- Geogazeeting
  - *geoBoundaries* (Runfola et al., 2020)
- Accuracy estimates

**Final database:** 155,858 records

- 105,574 records from **GPS**
- 50,284 records from **locality**



# Biogeographic data

**Occurrence sources:**

- AntWeb (62,301 records)
- GABI/AntMap (93,557 records)

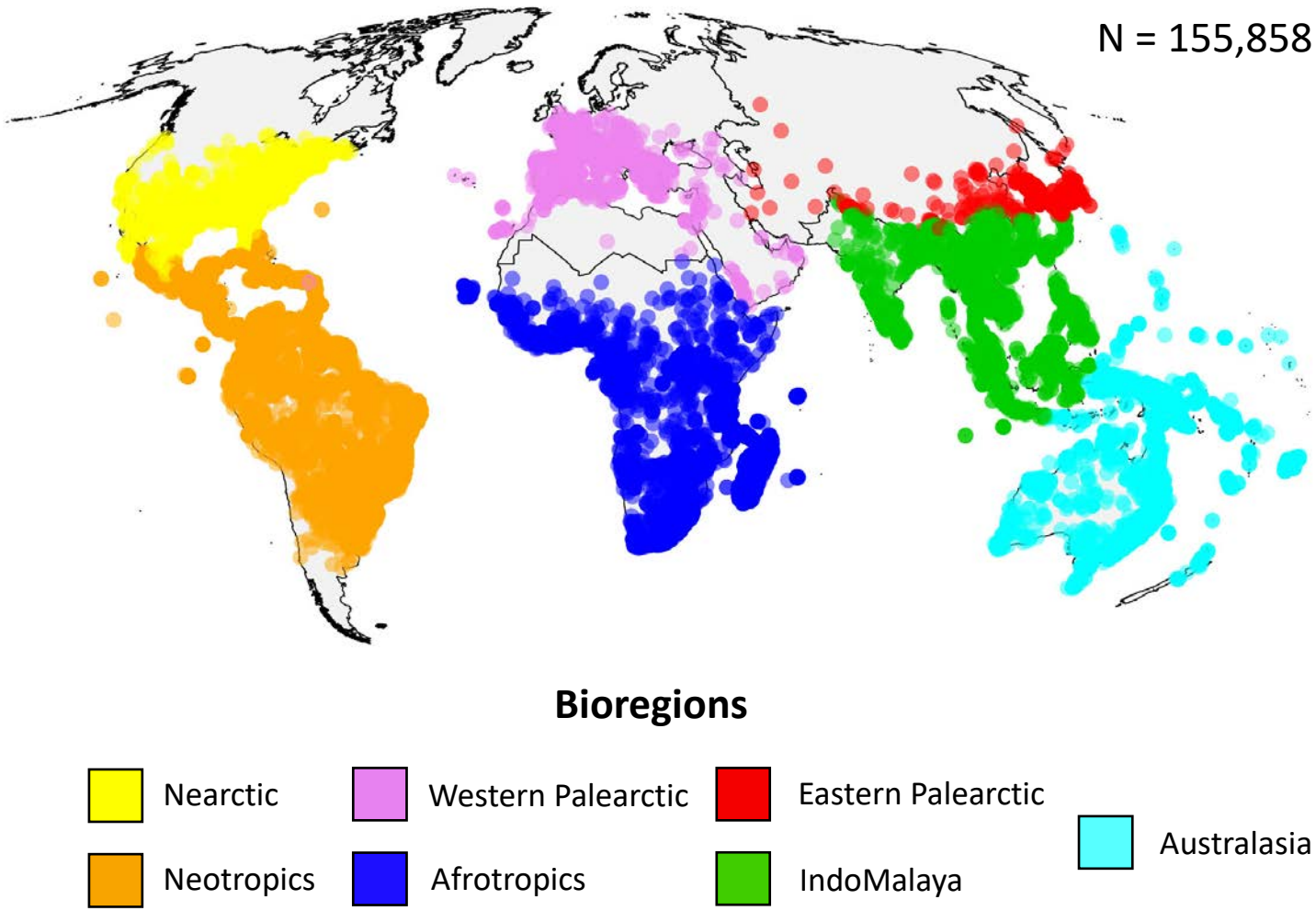
**Bioregions** as defined in AntWeb

**Data curation:**

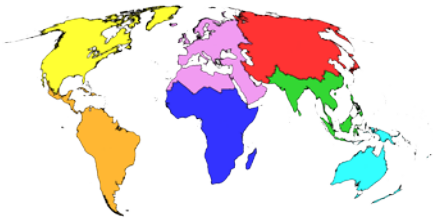
- Automatic and visual checks
  - *CoordinateCleaner* (Zizka et al., 2019)
- Geogazeeting
  - *geoBoundaries* (Runfola et al., 2020)
- Accuracy estimates

**Final database:** 155,858 records

- 105,574 records from **GPS**
- 50,284 records from **locality**



# Biogeographic methods



**BioGeoBEARS** framework:

- DIVALIKE, DEC, DIVALIKE + J, DEC + J
- Model selection based on AICc
- Best model = **Time-stratified DEC + J**

Model of **continental connectivity**

**Ancestral Range Estimates (ARE):**

- Marginal likelihoods of node ranges

**Biogeographic Stochastic Mapping (BSM):**

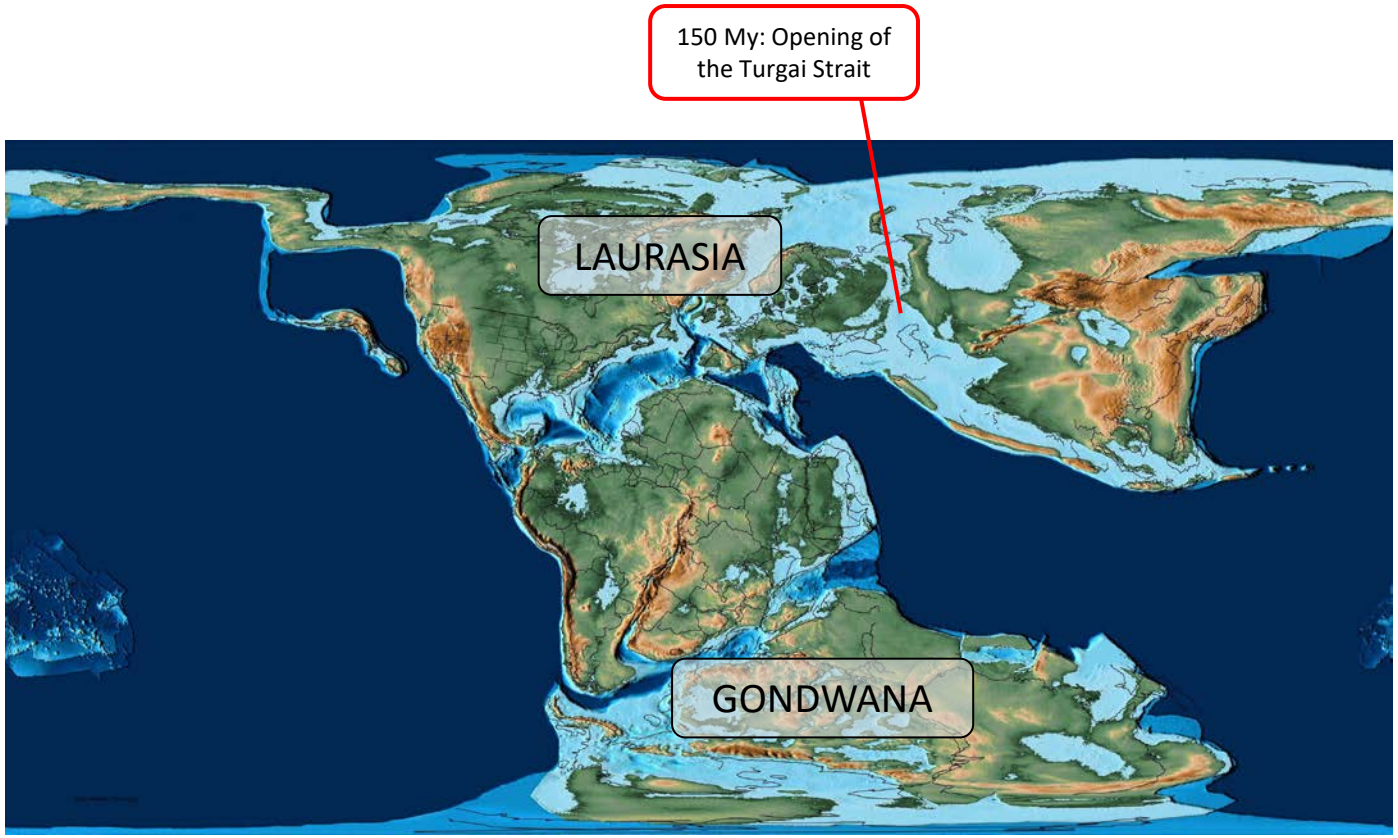
- Simulation of **biogeographic histories** compatible with ancestral ranges estimates
- Summary statistics across 1000 BSM

**Model comparison**

| | DEC+J | DIVALIKE+J | DIVALIKE | DEC |
|---|---|---|---|---|
| **logLk** | -2154 | -2205 | -4391 | -4467 |
| **# params** | 3 | 3 | 2 | 2 |
| **AICc** | 4313 | 4416 | 8787 | 8939 |
| **Δ AICc** | **0** | 103 | 4474 | 4626 |
| **Akaike weights** | **100 %** | 0 % | 0 % | 0 % |
| **d** | 0.007 | 0.008 | 0.030 | 0.024 |
| **e** | 0.000 | 0.000 | 0.019 | 0.020 |
| **y** | 0.994 | 0.991 | 1.000 | 1.000 |
| **v** | 0.994 | 0.991 | 1.000 | 1.000 |
| **s** | 0.994 | - | - | 1.000 |
| **j** | 0.019 | 0.018 | - | - |

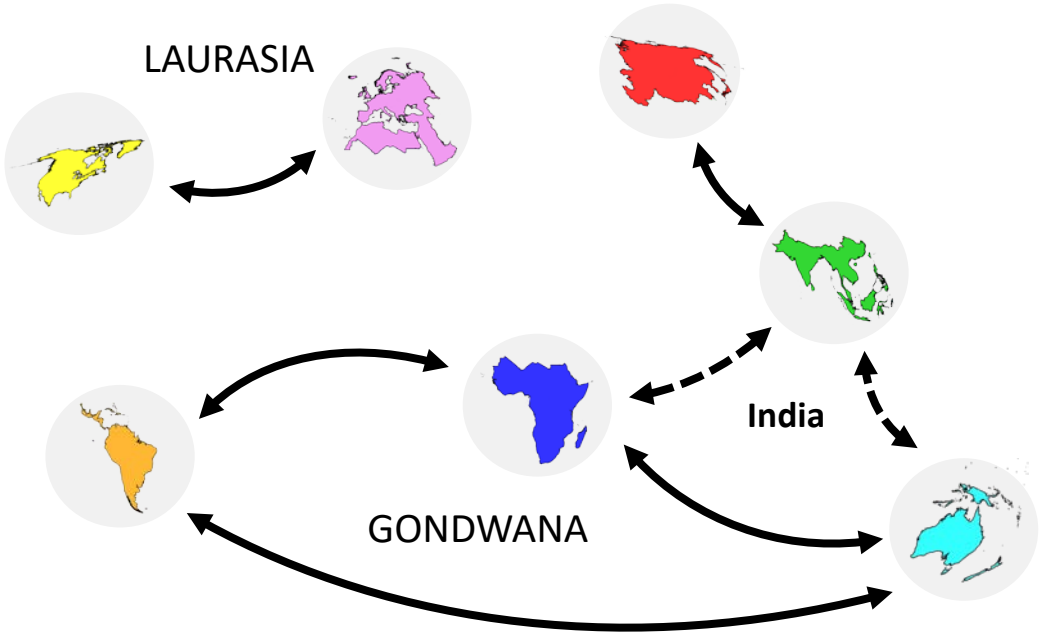
Doré et al., *in prep.*

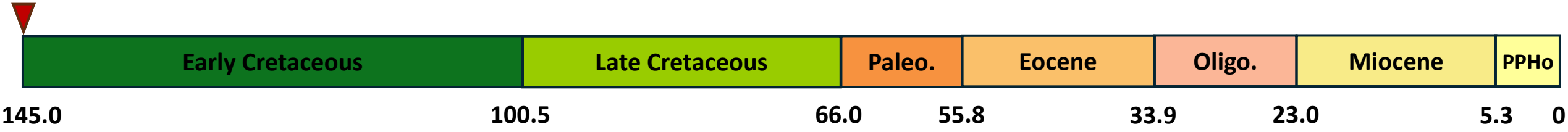
# Biogeographic methods




Source: PALEOMAP Project by C. Scotese

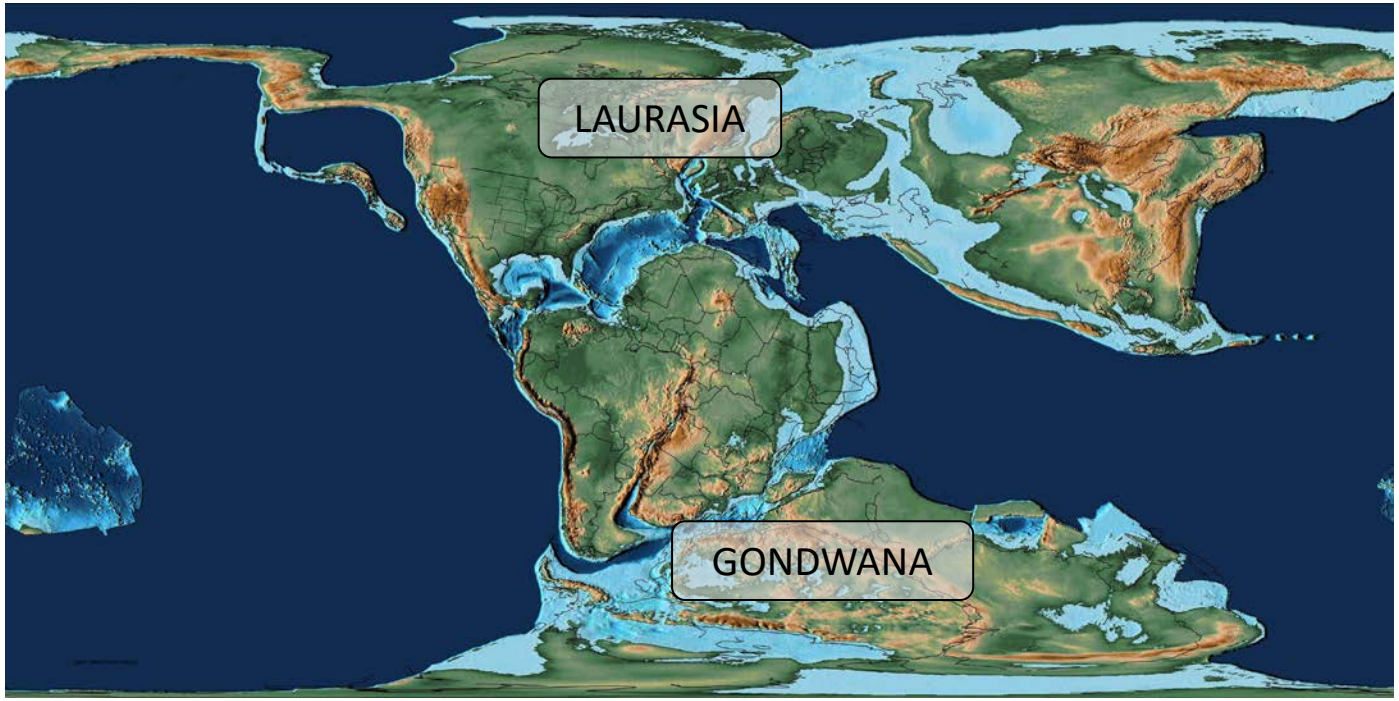
Early Cretaceous
145.0 My

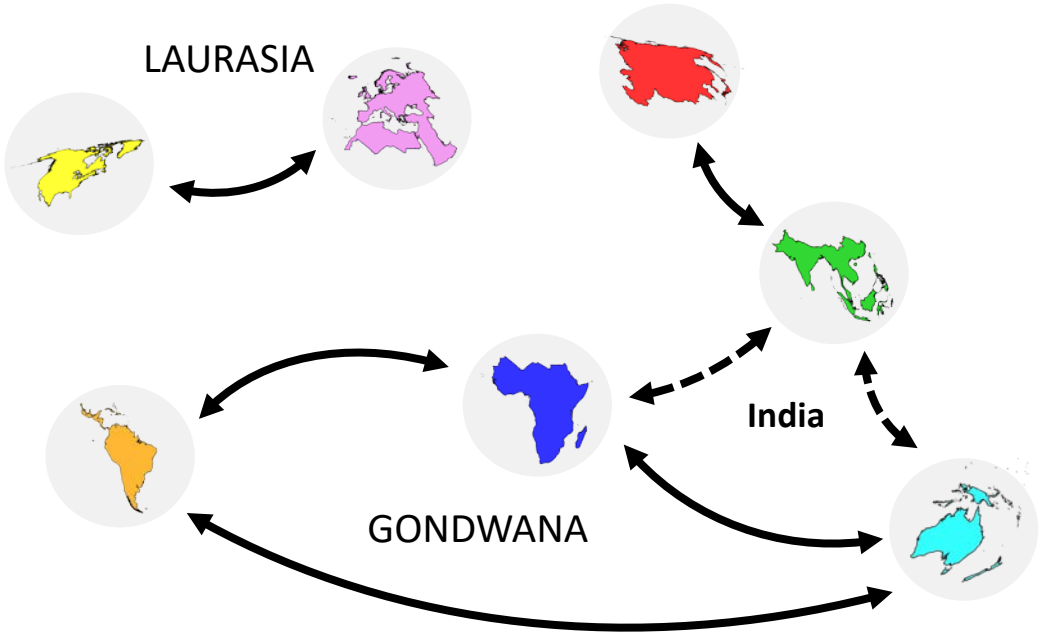


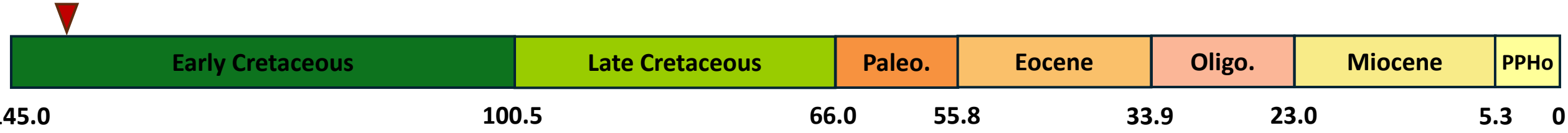
# Biogeographic methods

Early Cretaceous
140.0 My



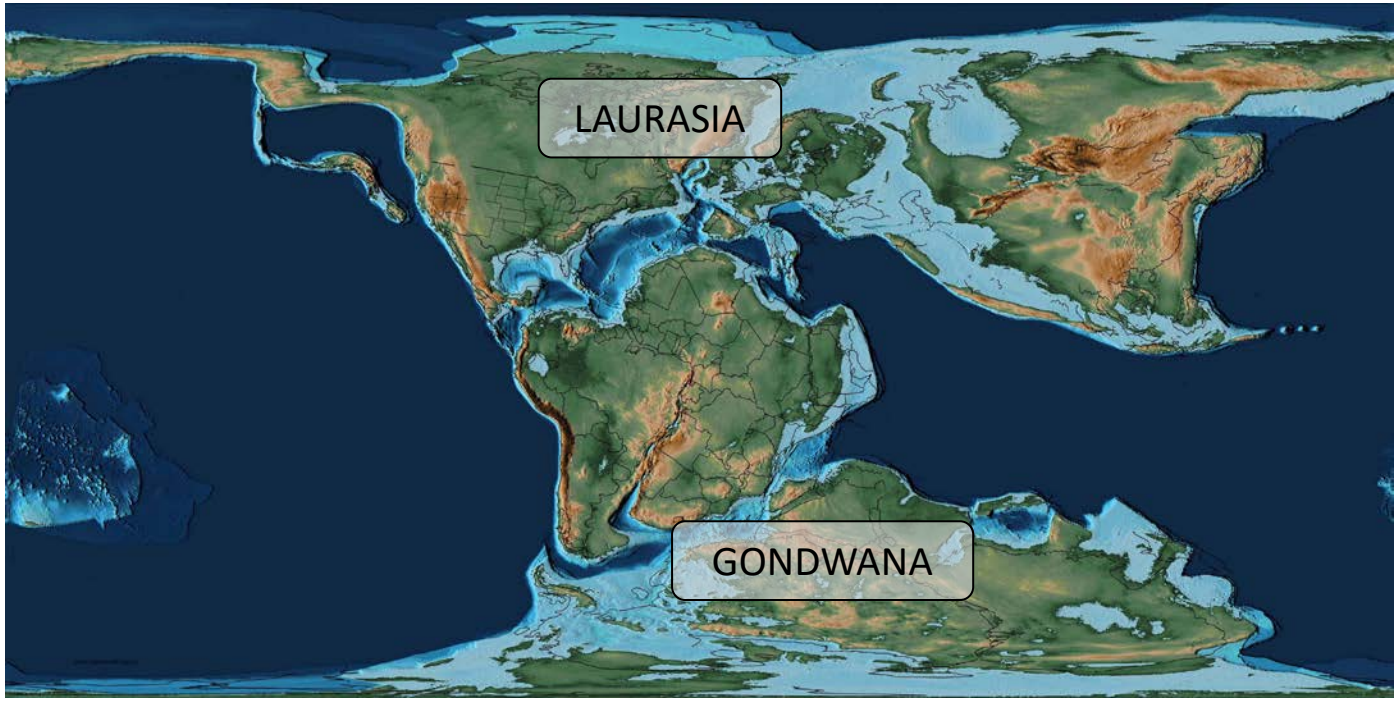



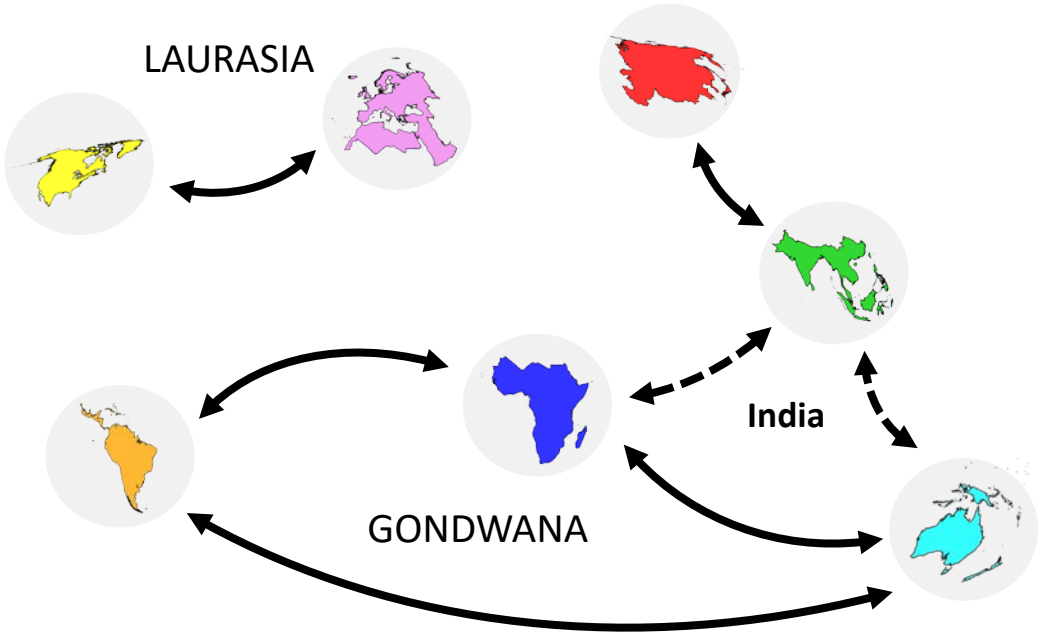

Source: PALEOMAP Project by C. Scotese



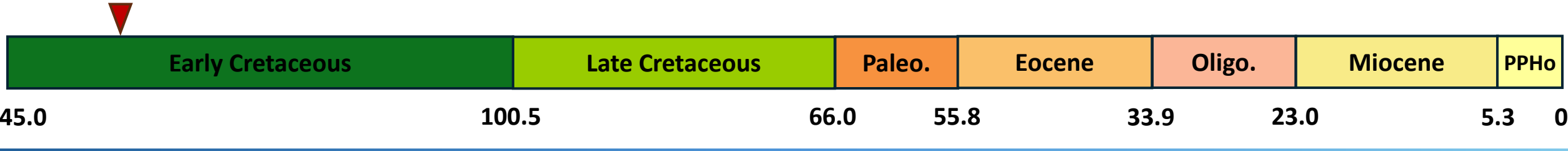
# Biogeographic methods

Early Cretaceous
135.0 My



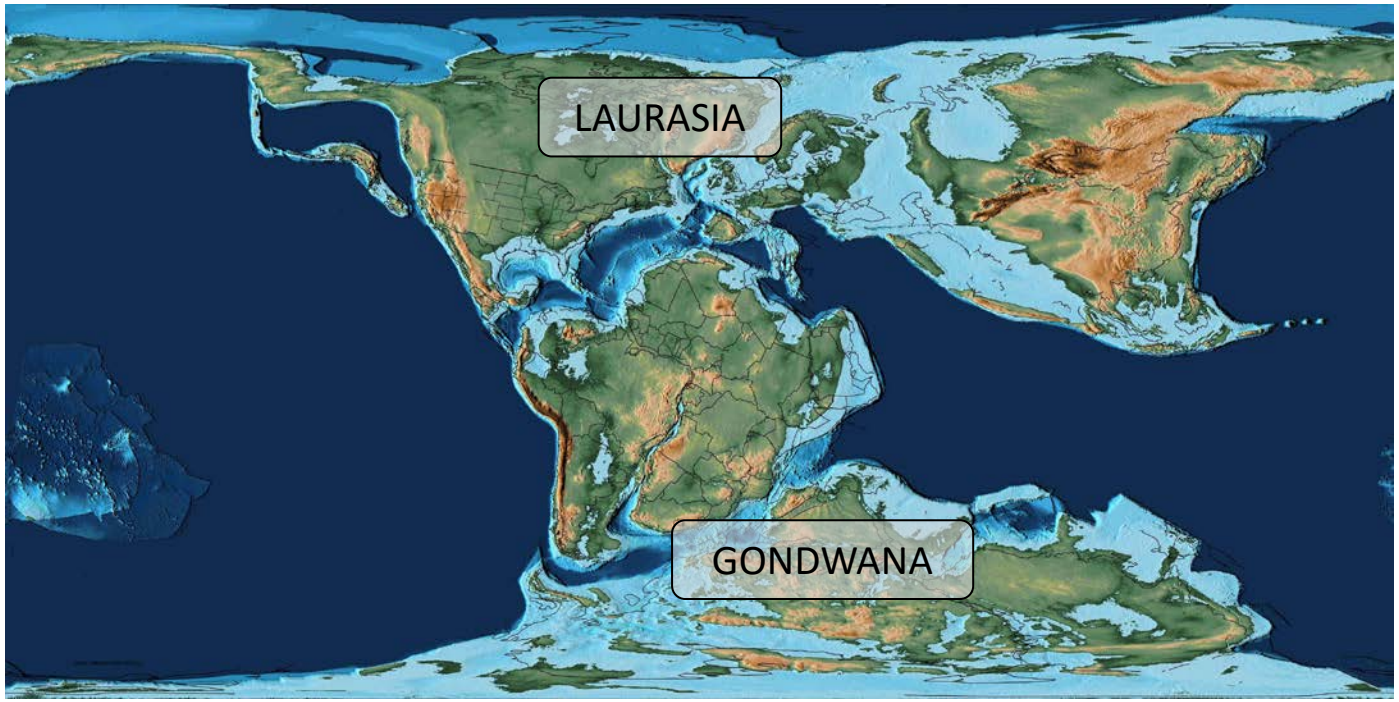




Source: PALEOMAP Project by C. Scotese

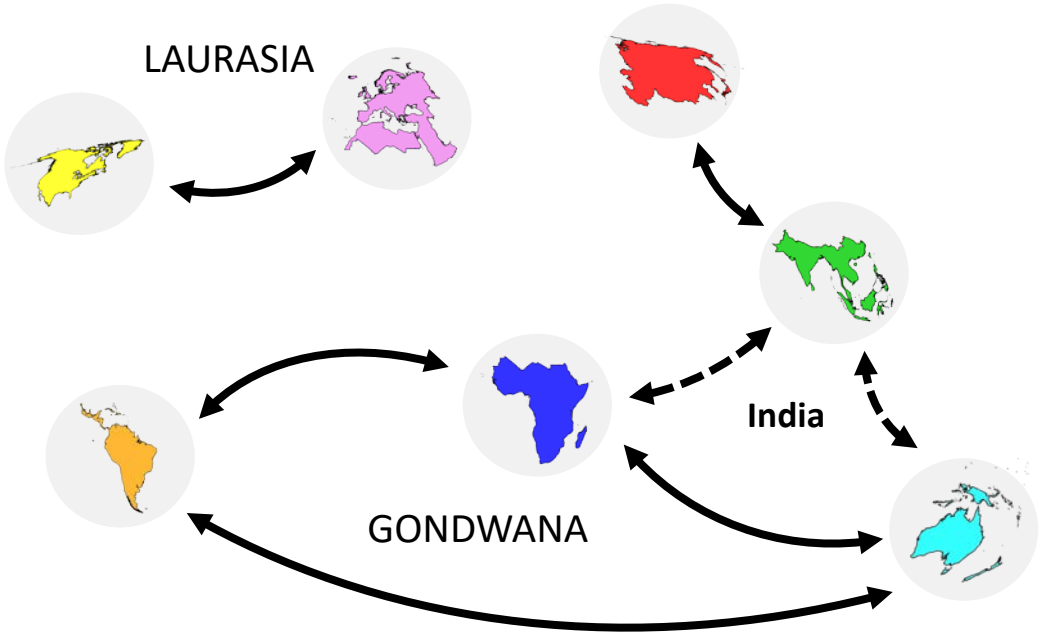


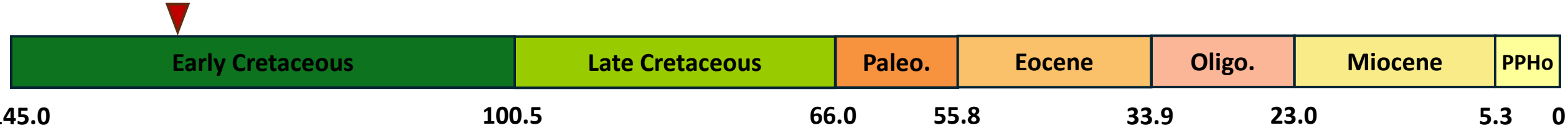
# Biogeographic methods

Early Cretaceous
130.0 My

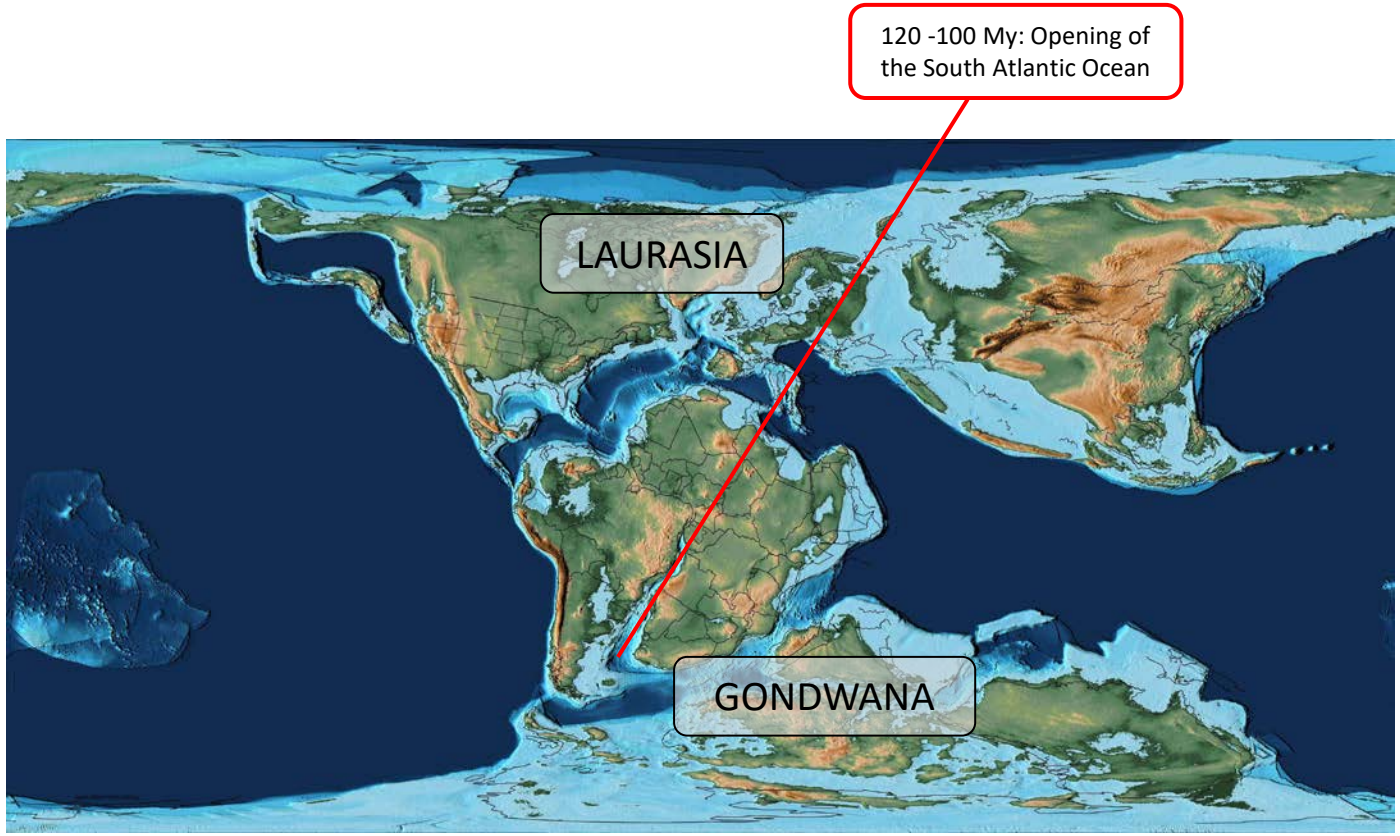



Source: PALEOMAP Project by C. Scotese

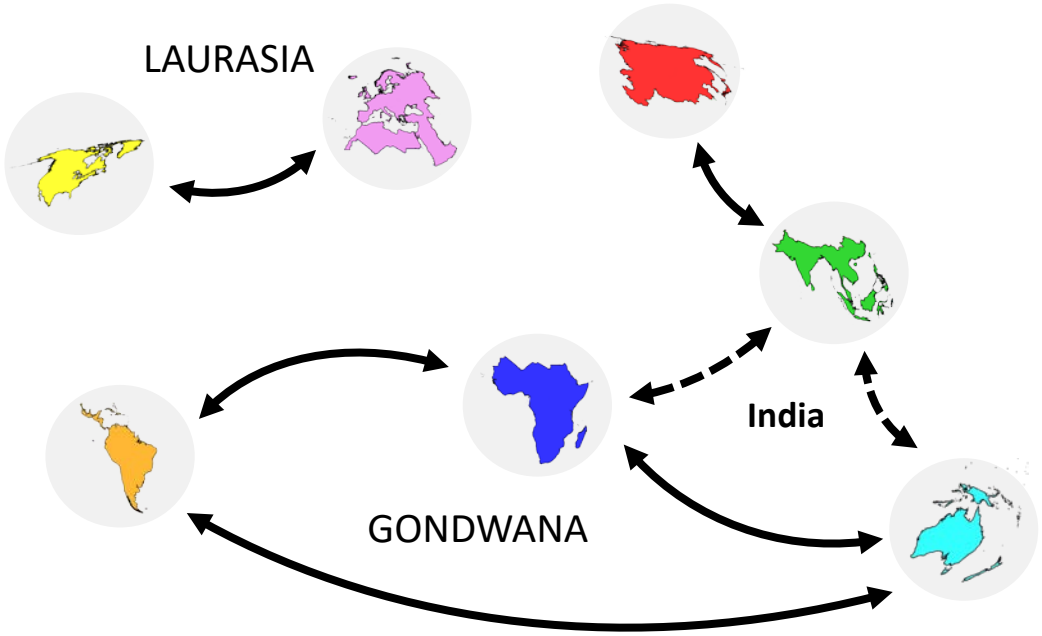


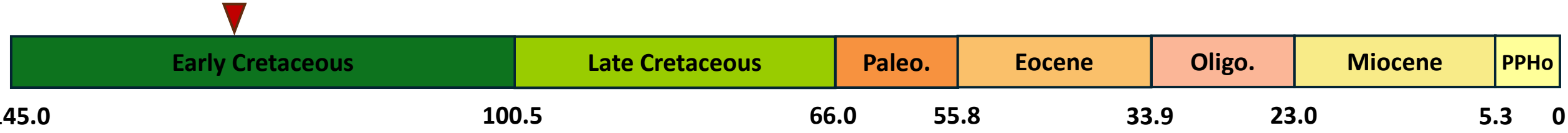



# Biogeographic methods

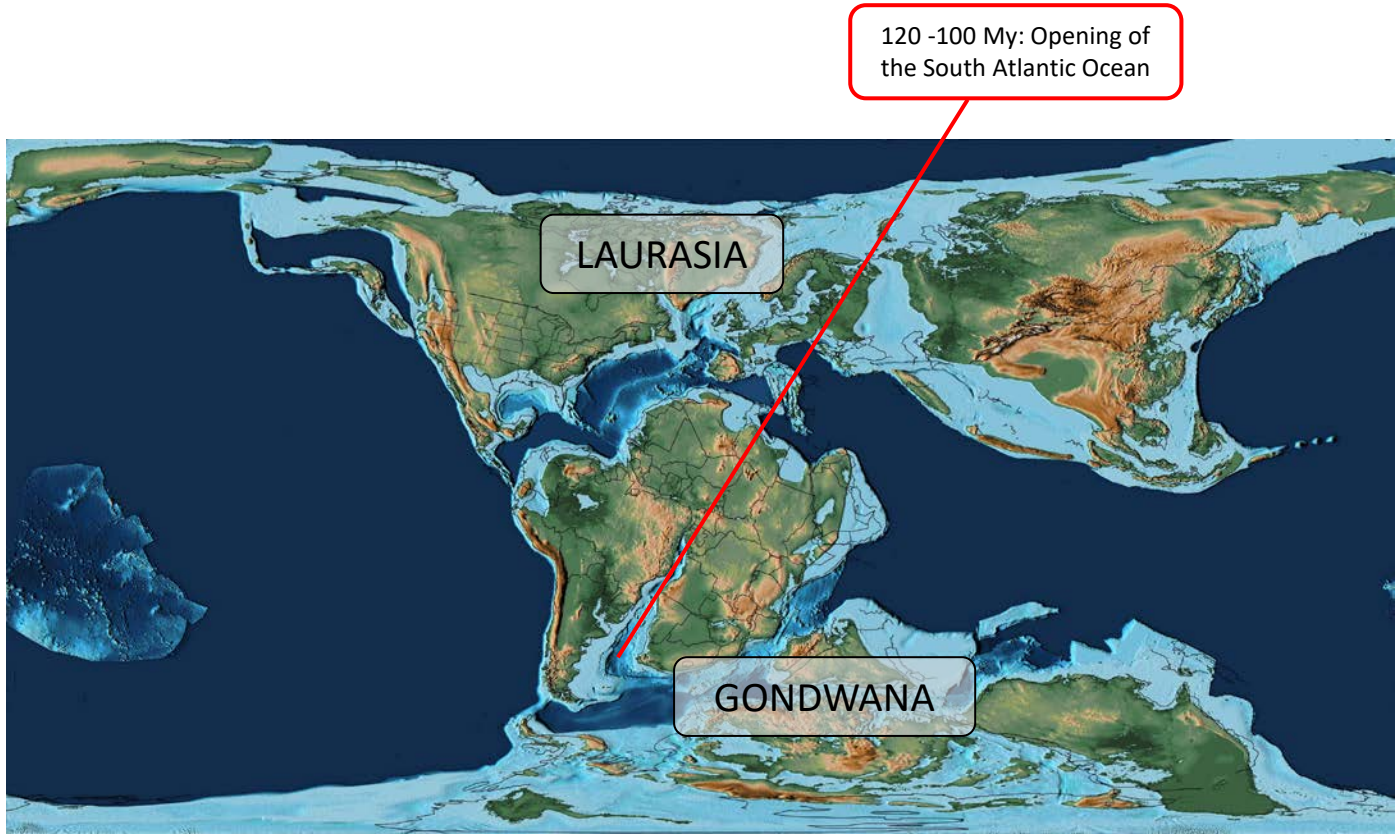



Source: PALEOMAP Project by C. Scotese

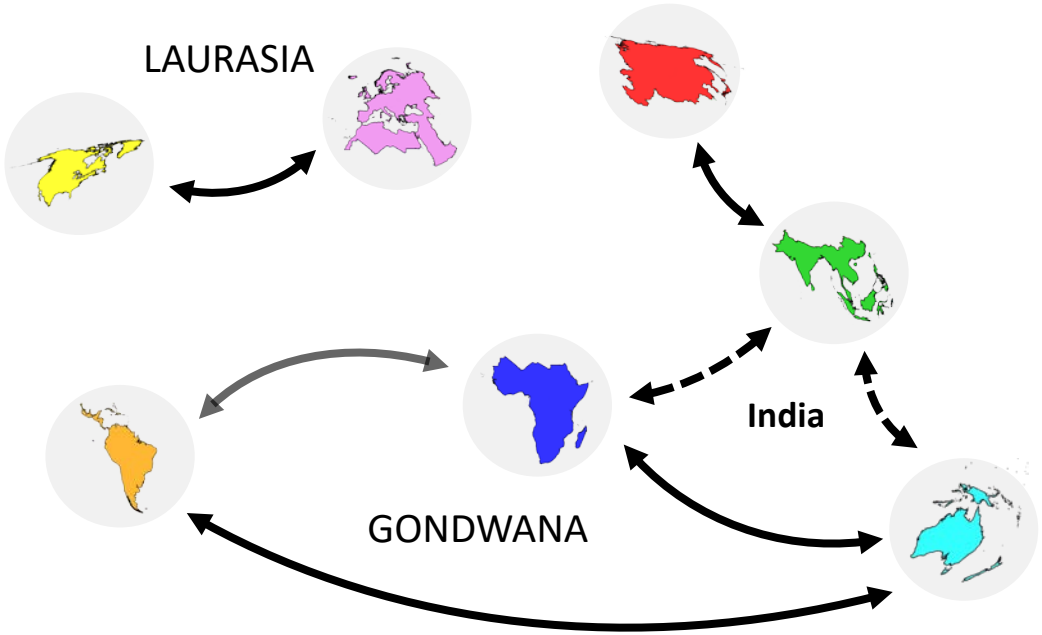


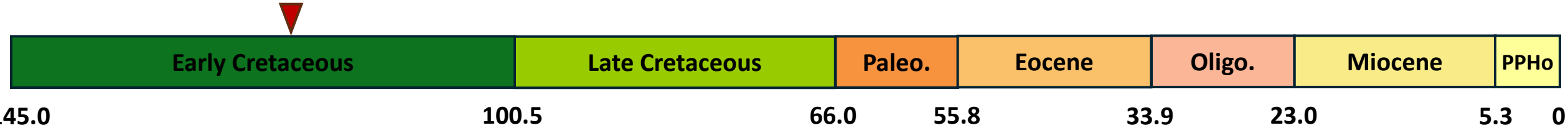



# Biogeographic methods

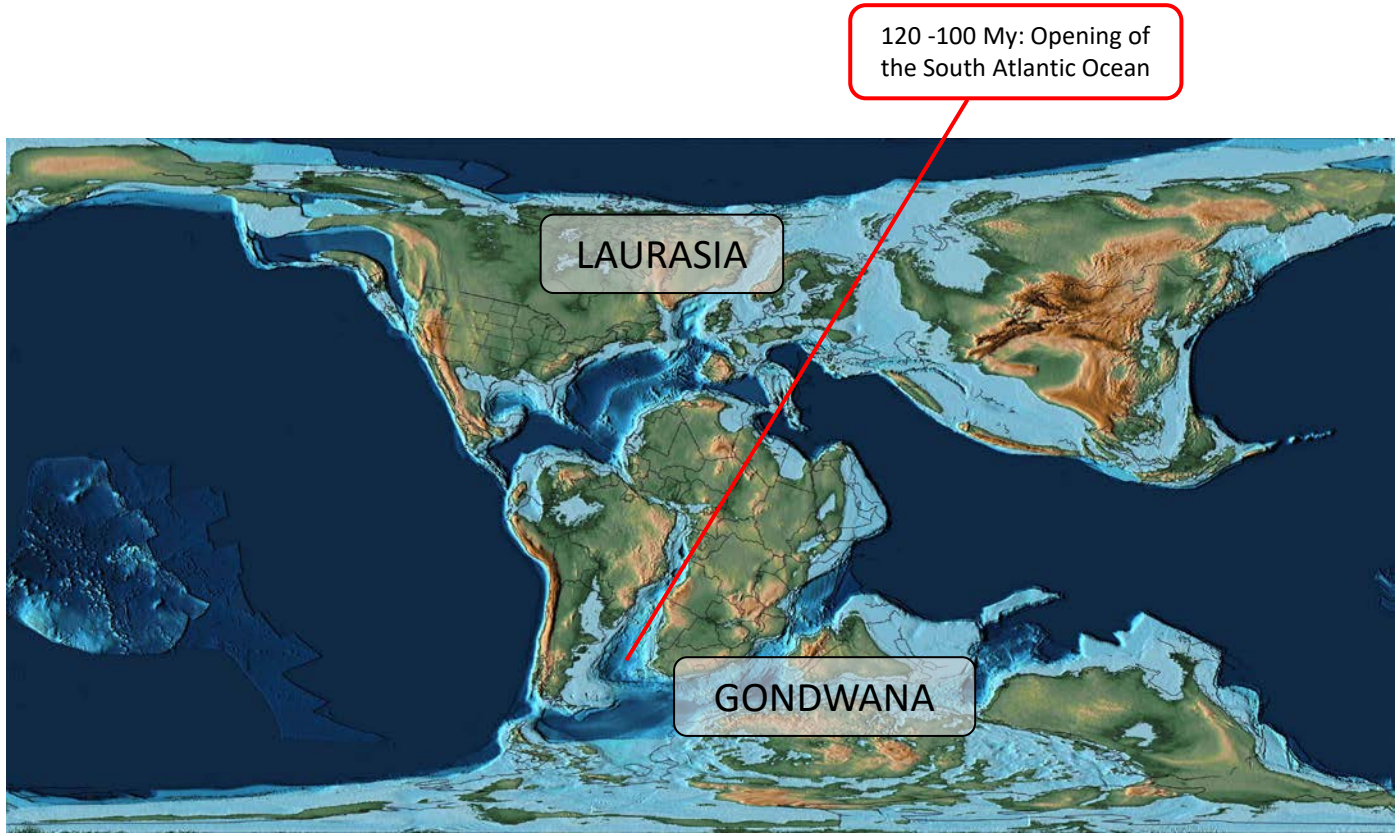



Source: PALEOMAP Project by C. Scotese

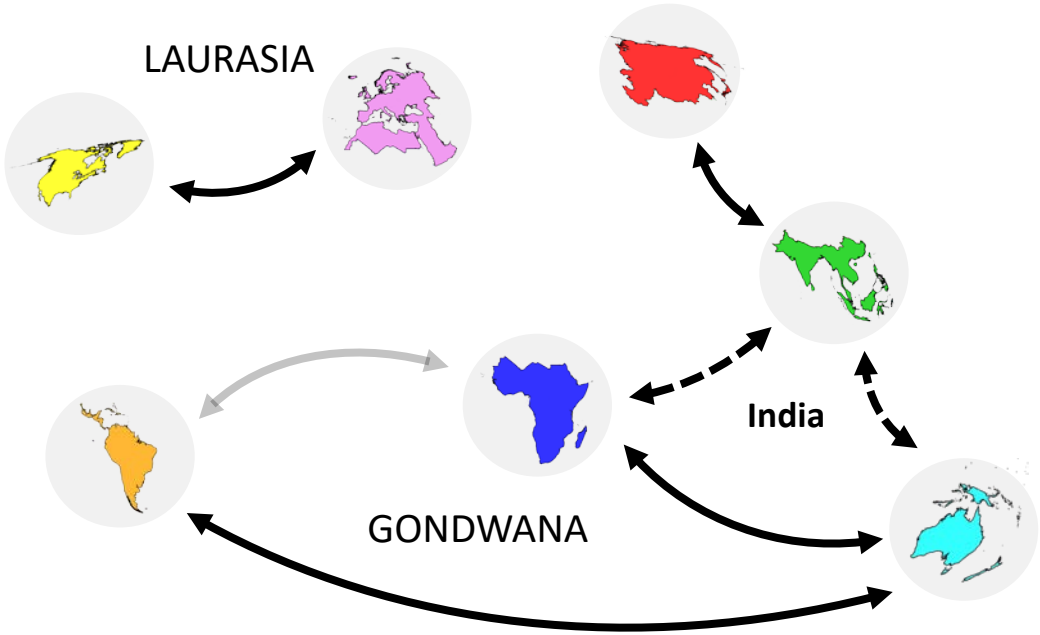


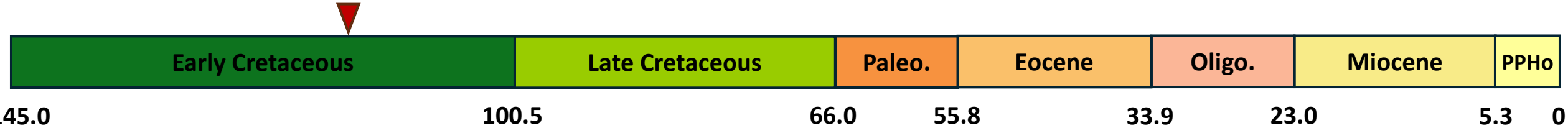



# Biogeographic methods

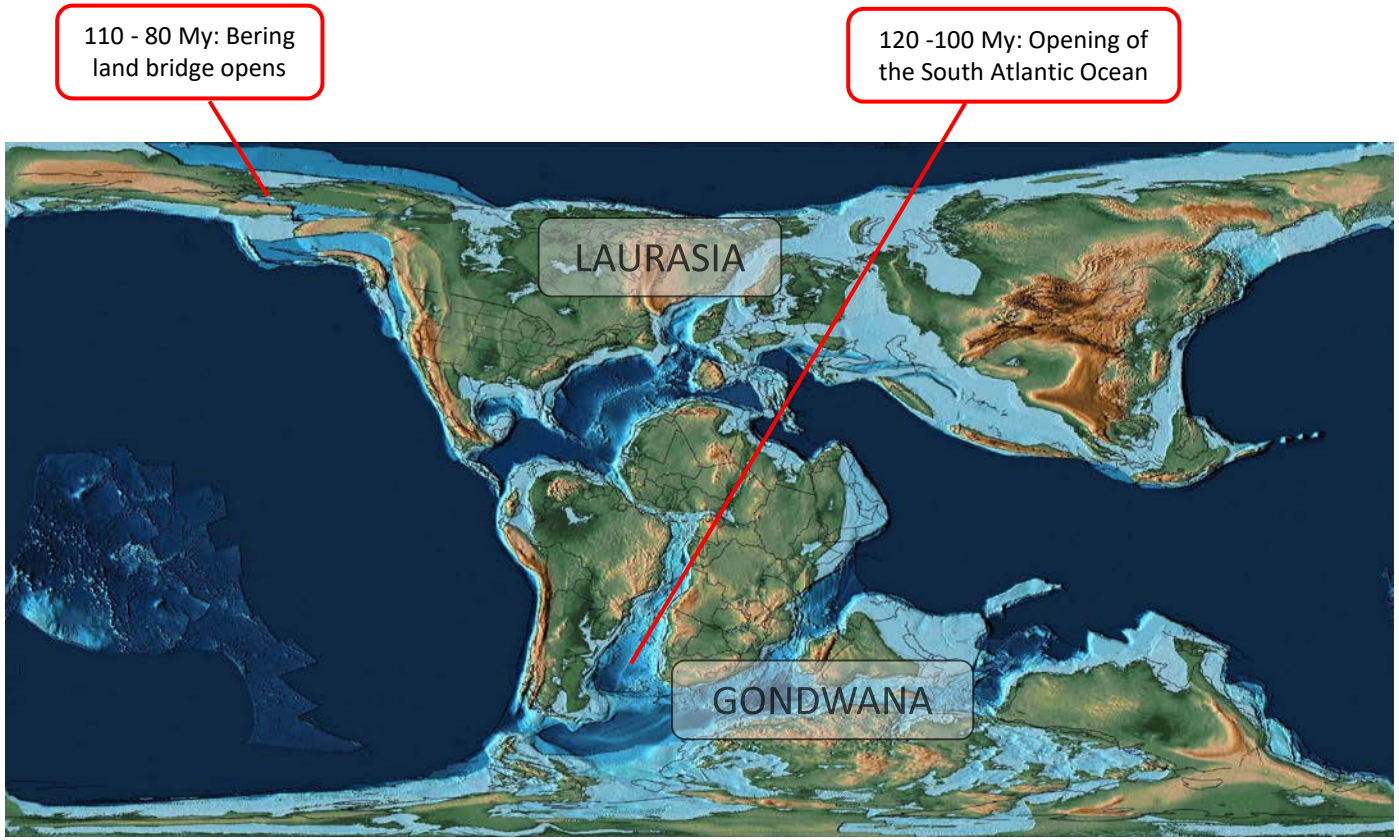



Source: PALEOMAP Project by C. Scotese

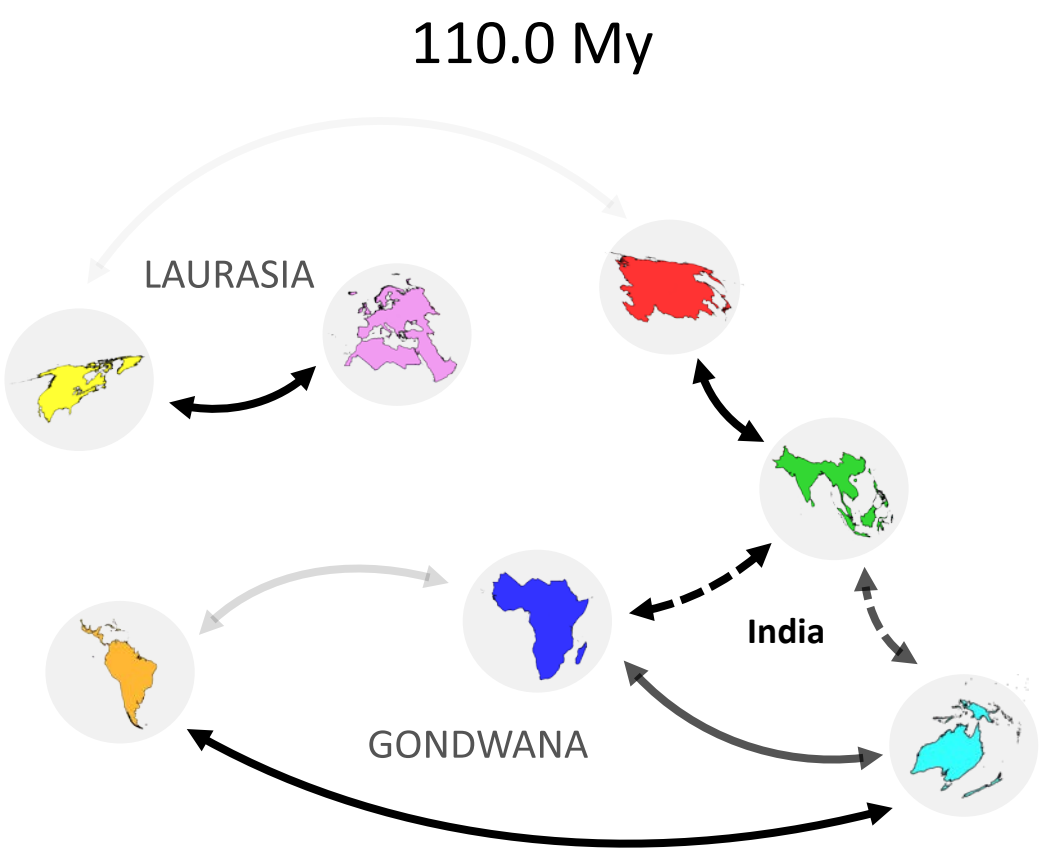


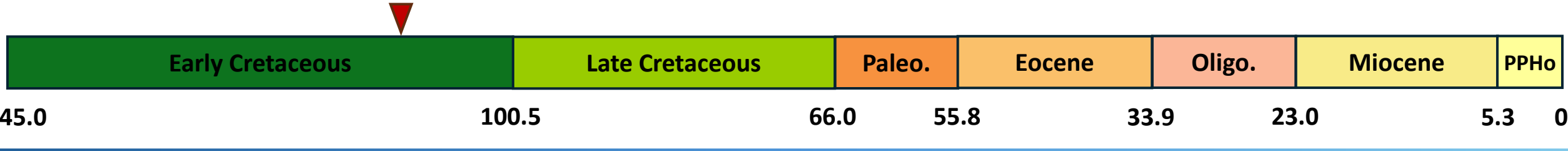



# Biogeographic methods

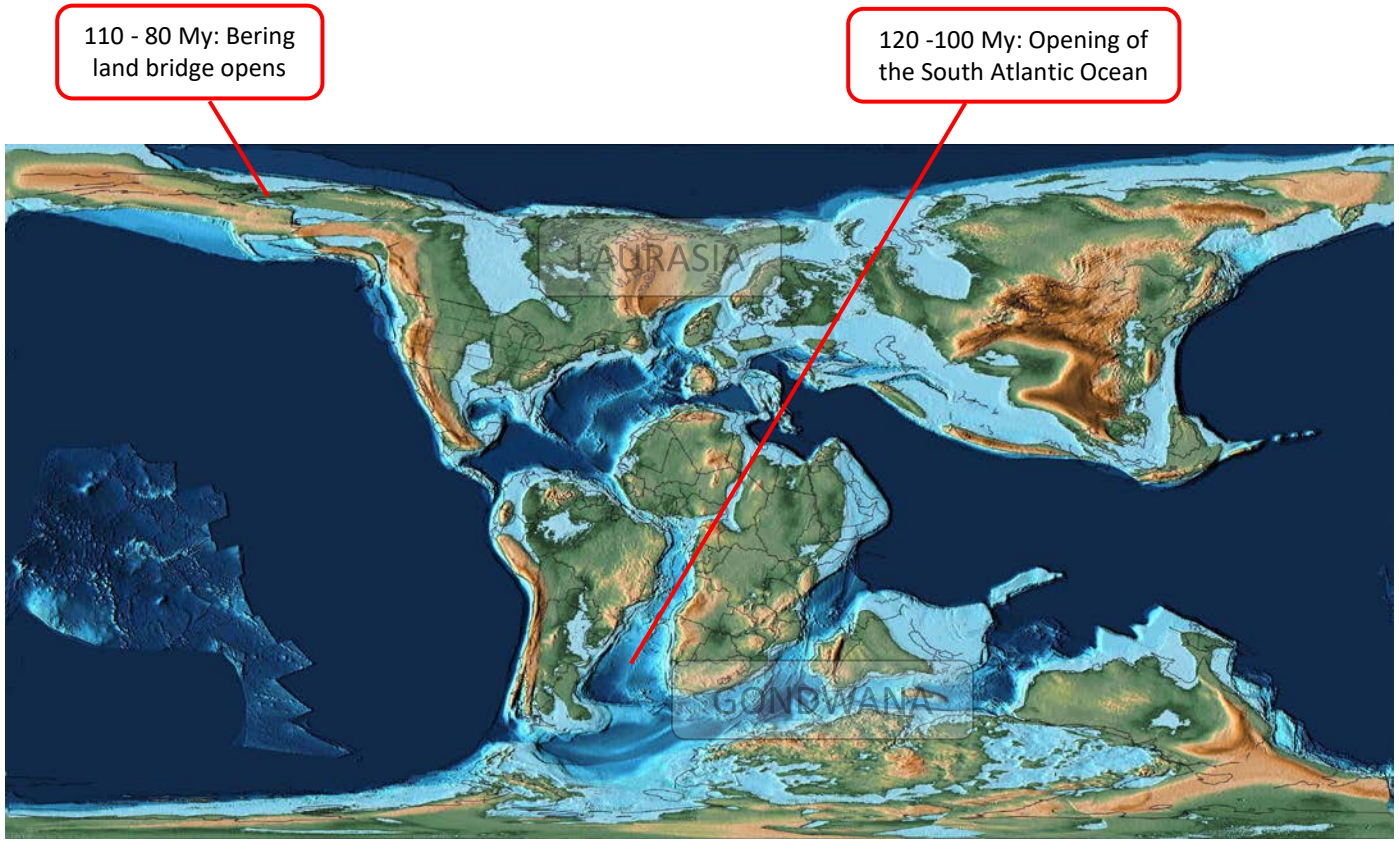



Source: PALEOMAP Project by C. Scotese

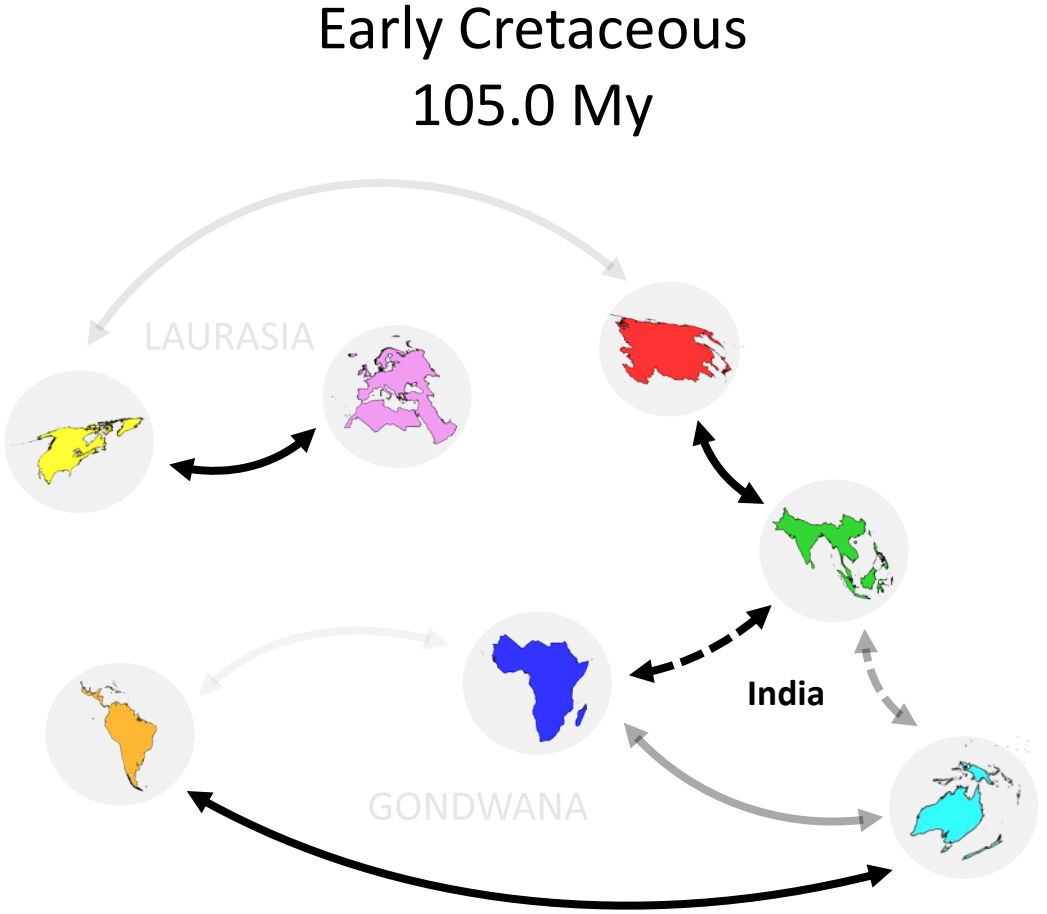


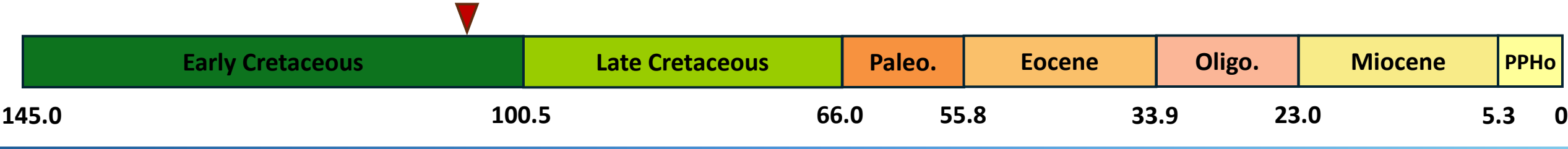



# Biogeographic methods

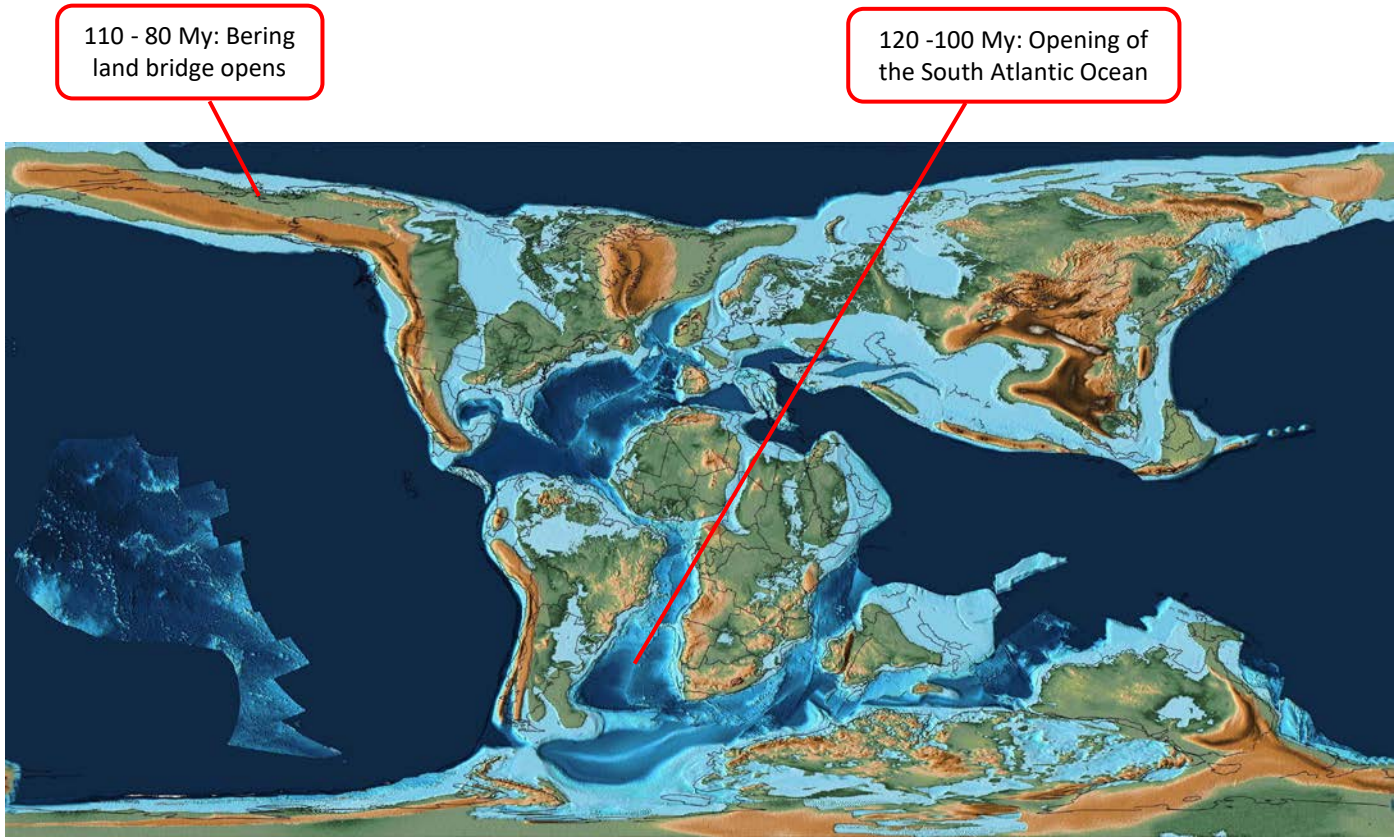



Source: PALEOMAP Project by C. Scotese

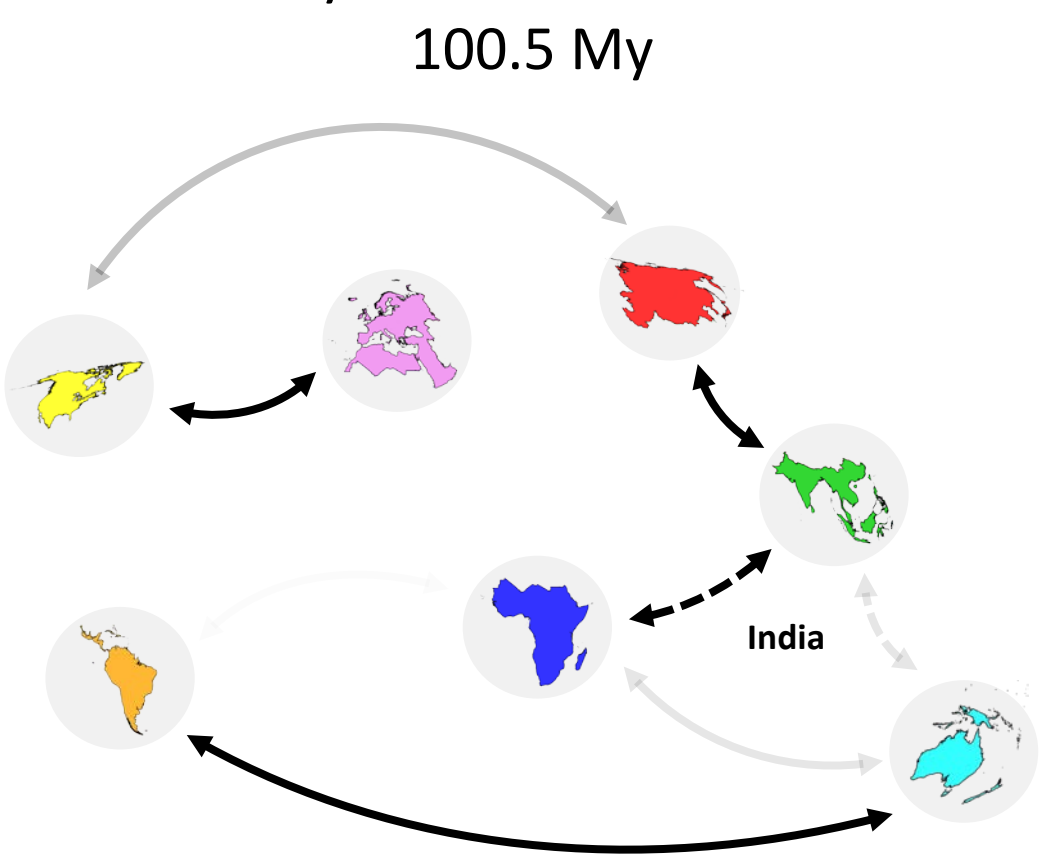


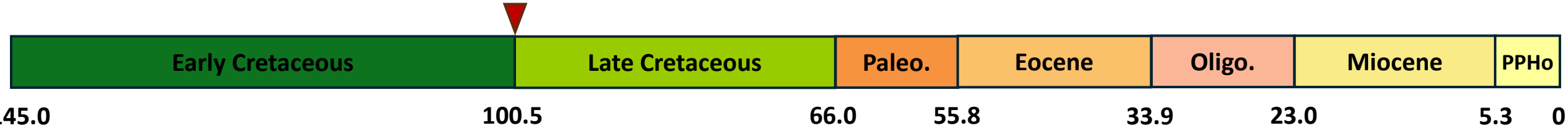



# Biogeographic methods



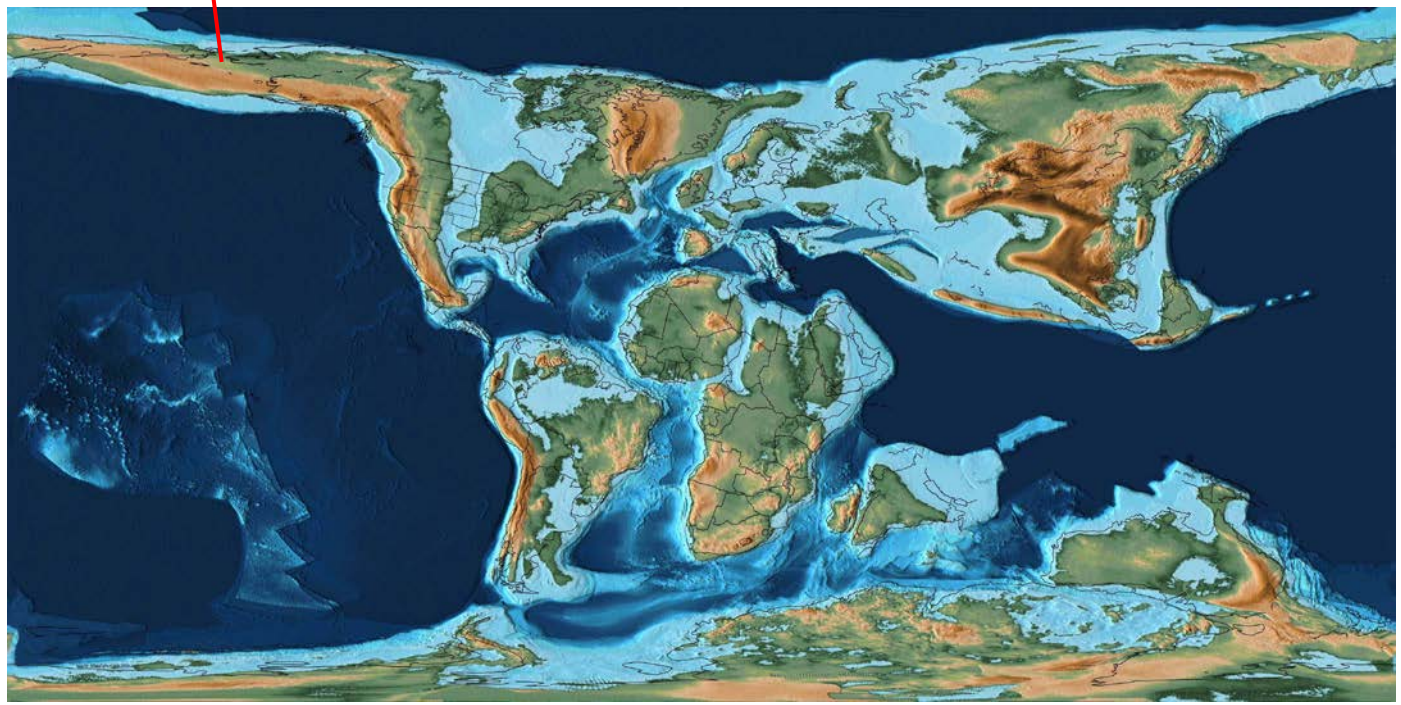

Source: PALEOMAP Project by C. Scotese



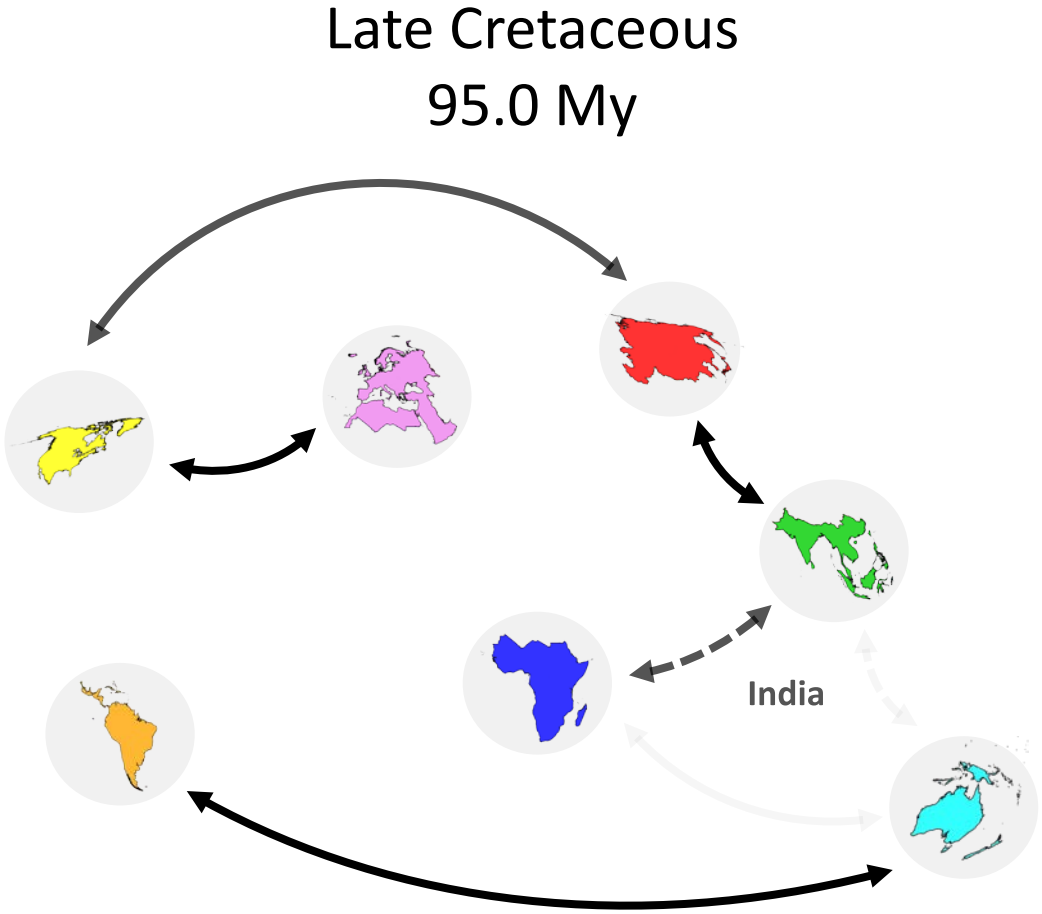



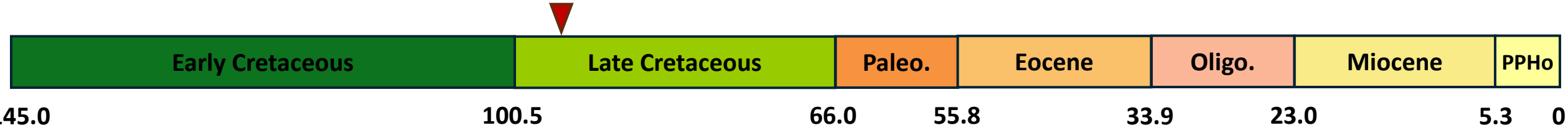
# Biogeographic methods

110 - 80 My: Bering land bridge opens

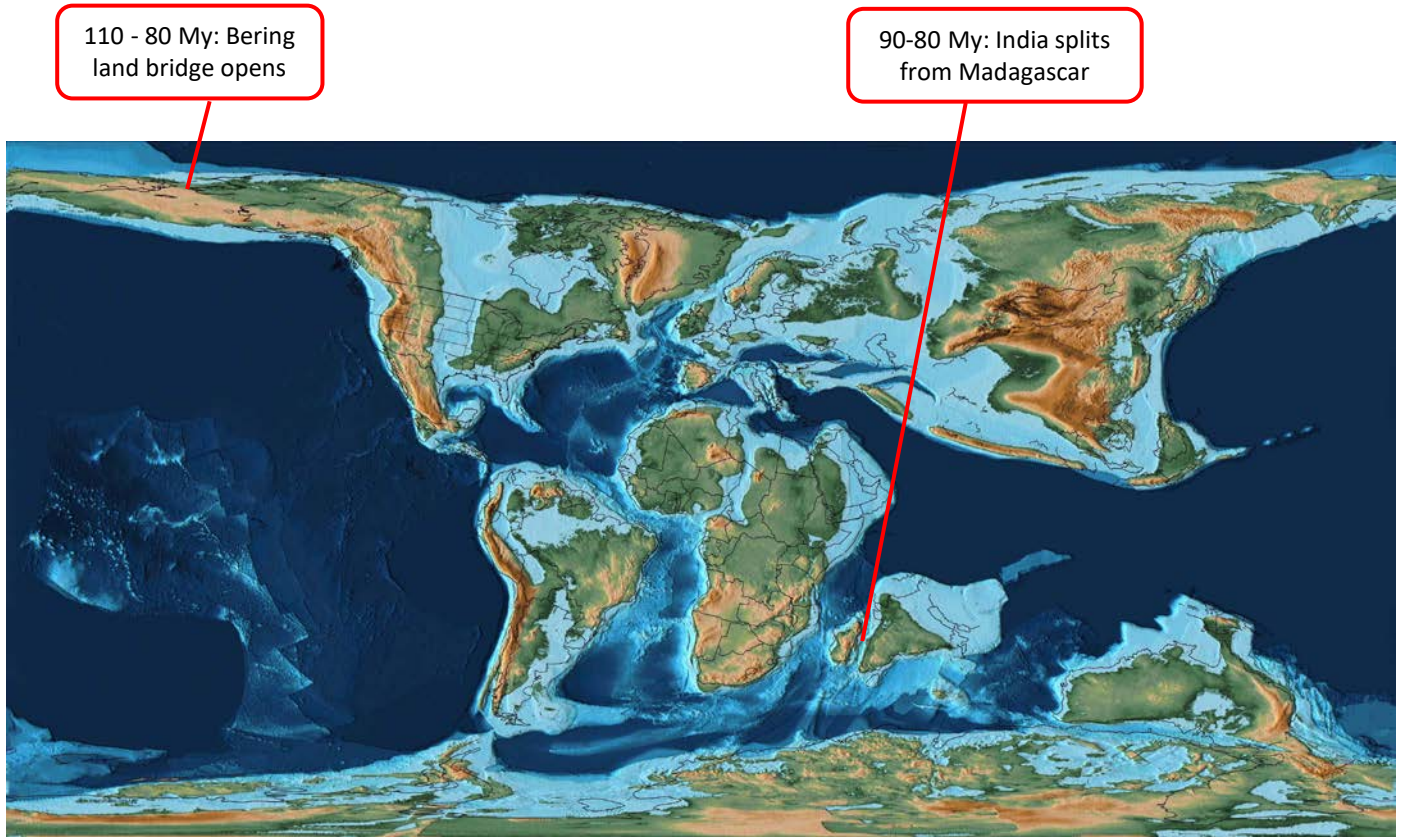


Source: PALEOMAP Project by C. Scotese

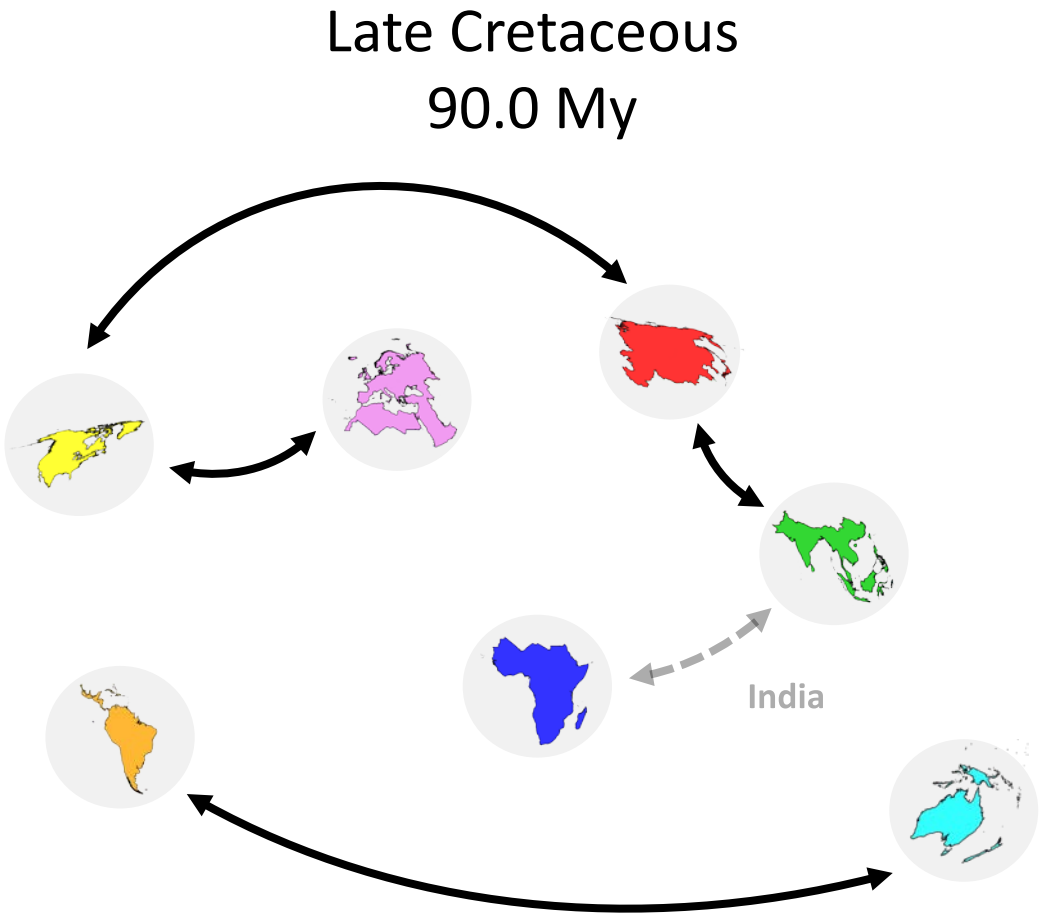


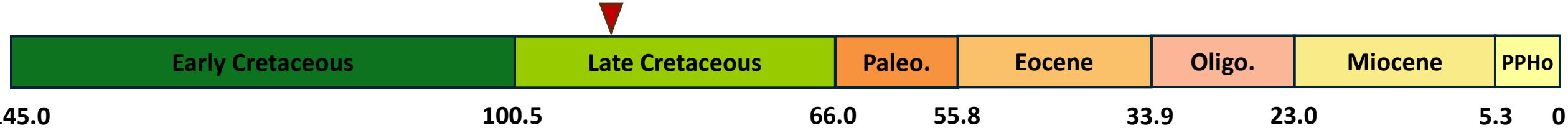



# Biogeographic methods




Source: PALEOMAP Project by C. Scotese

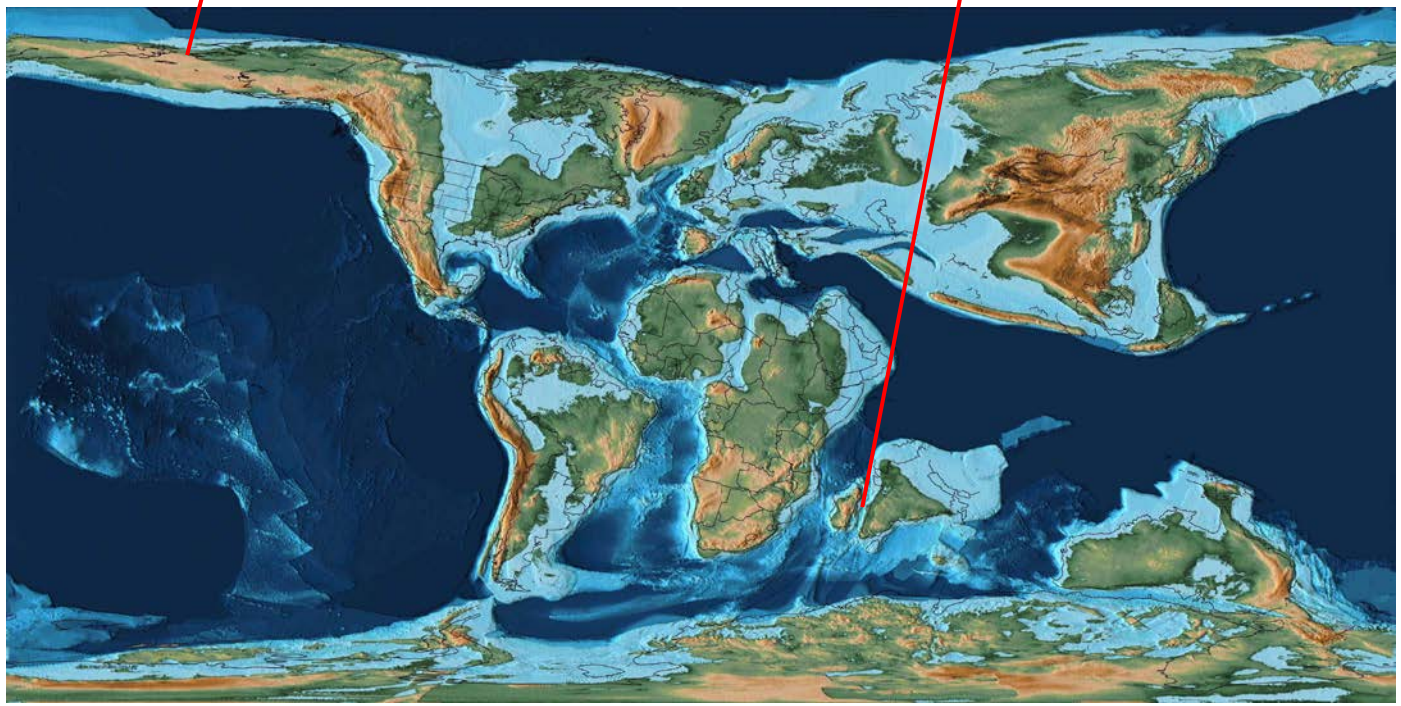


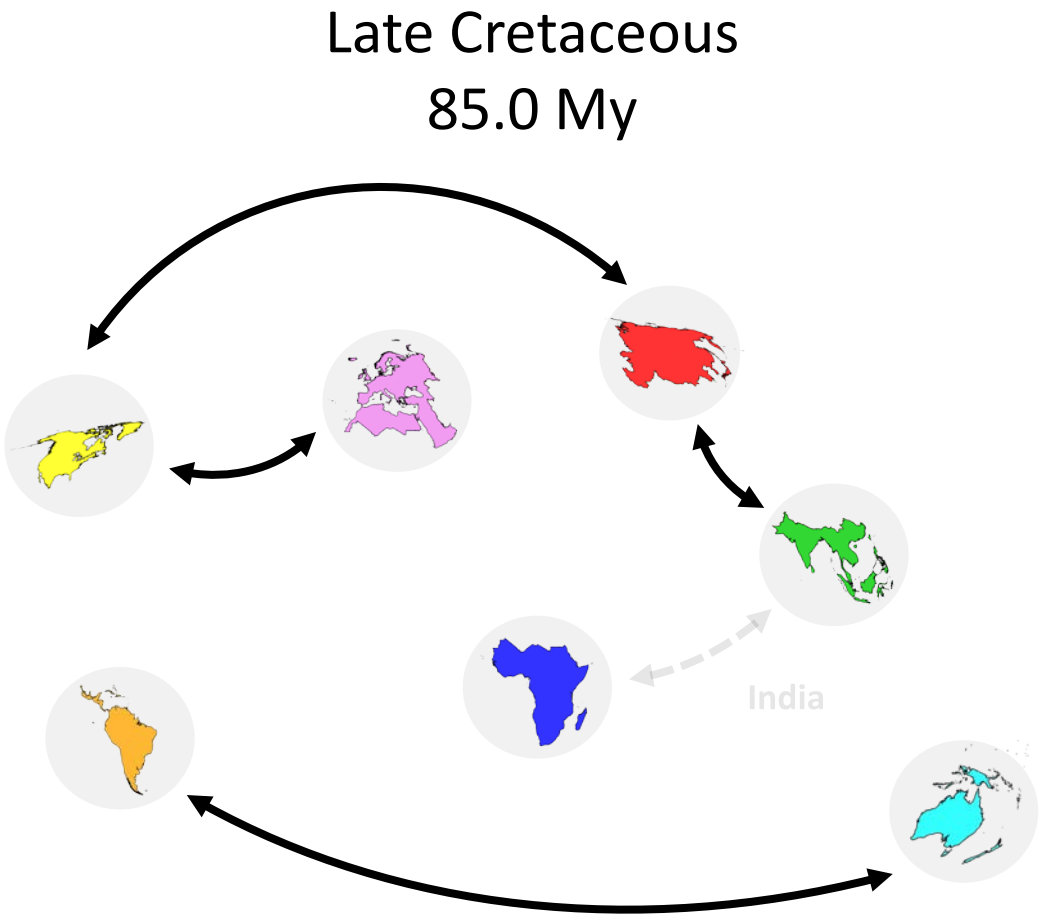
# Biogeographic methods

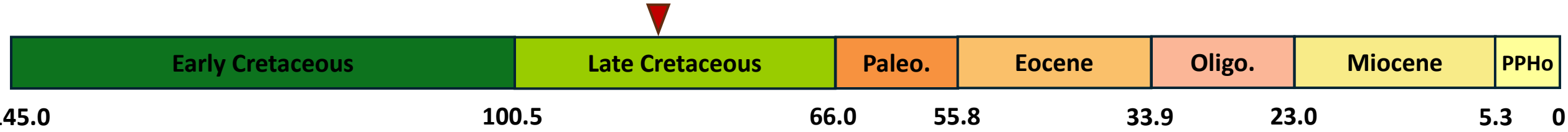
110 - 80 My: Bering land bridge opens

90-80 My: India splits from Madagascar



Source: PALEOMAP Project by C. Scotese

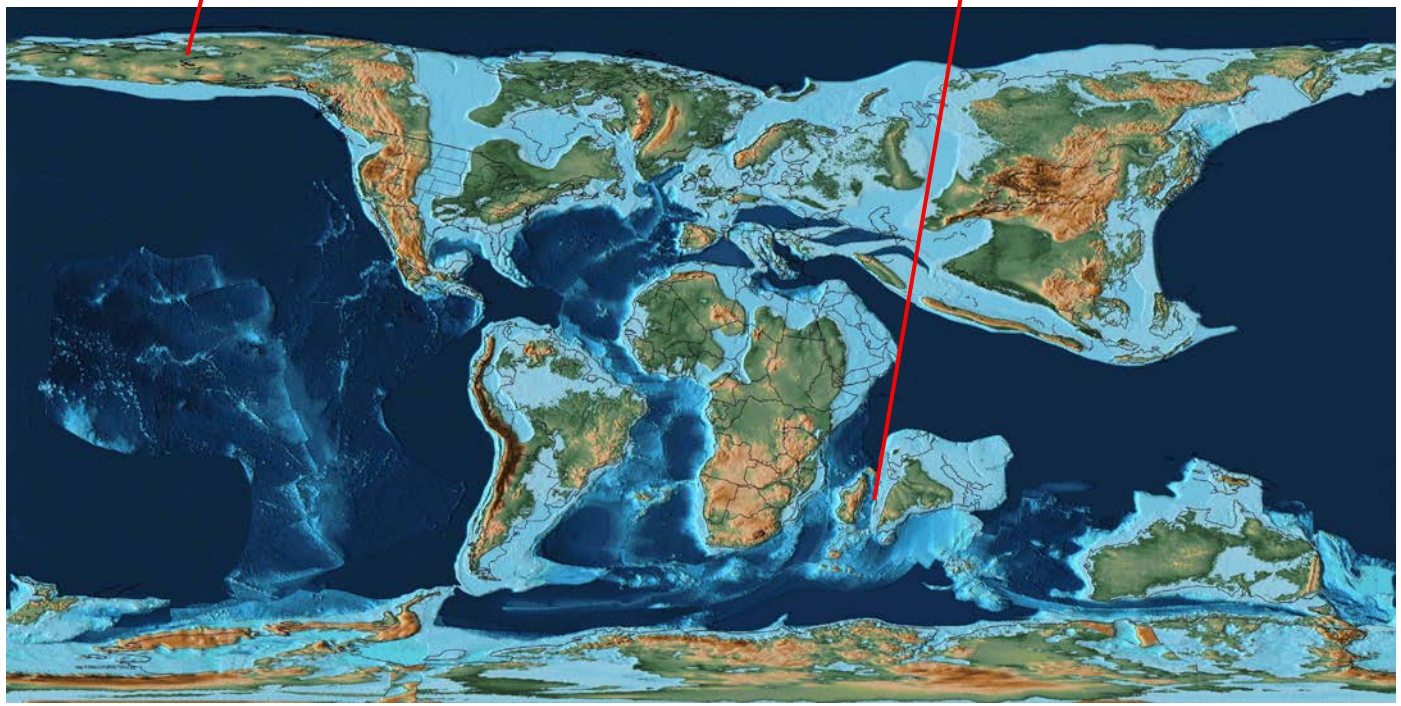


# Biogeographic methods

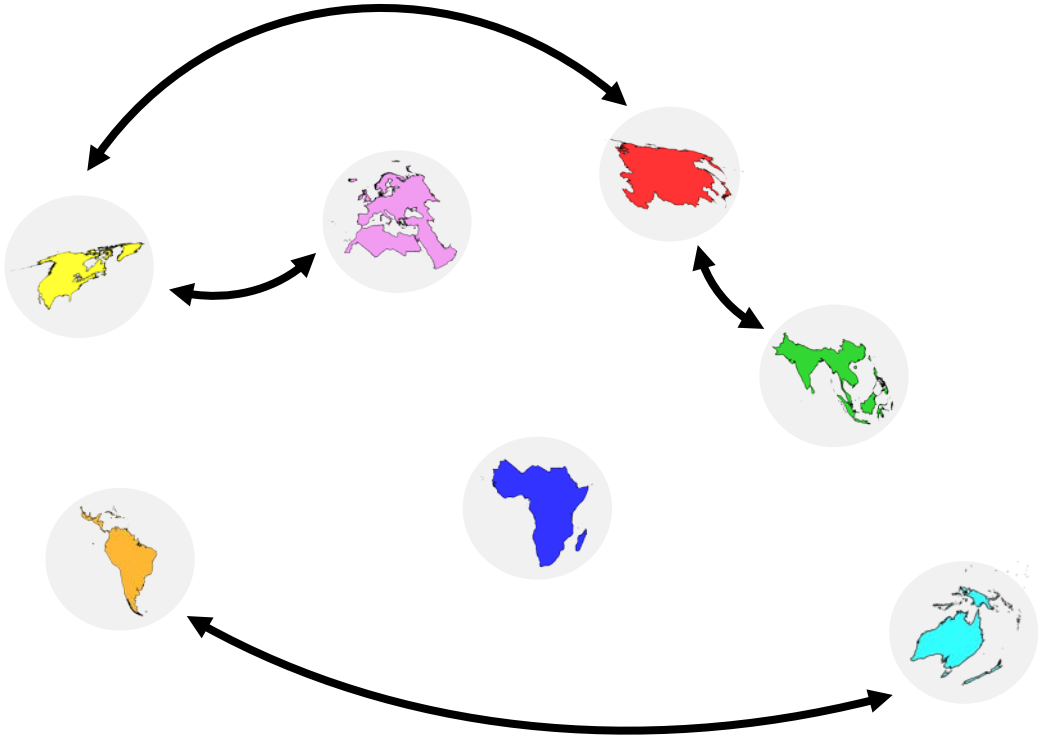
110 - 80 My: Bering land bridge opens

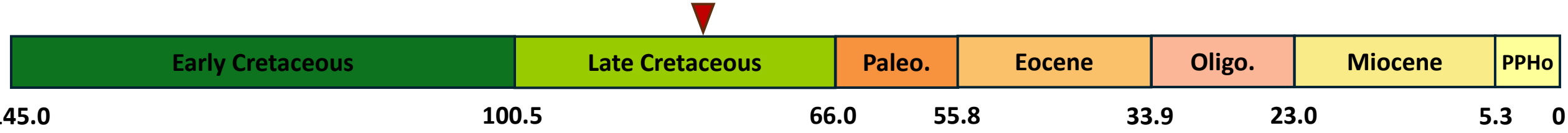
90-80 My: India splits from Madagascar



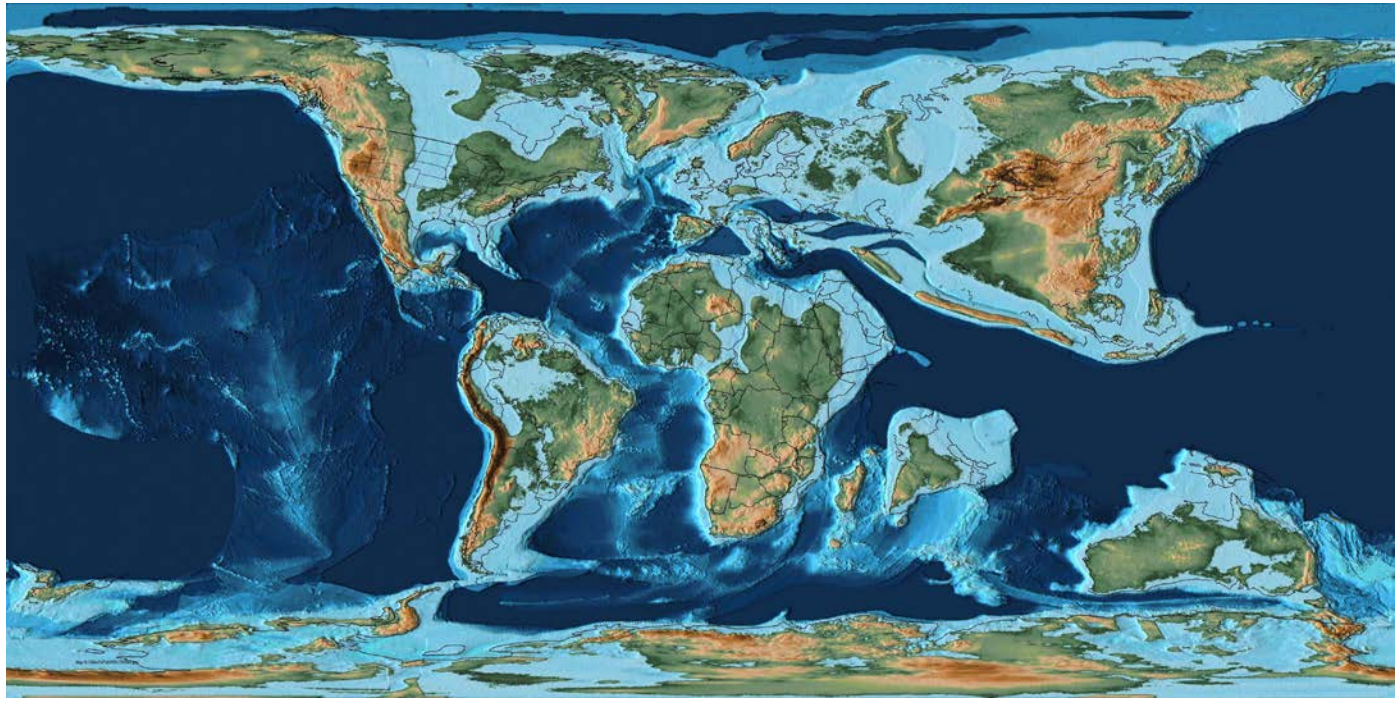
Source: PALEOMAP Project by C. Scotese

Late Cretaceous
80.0 My

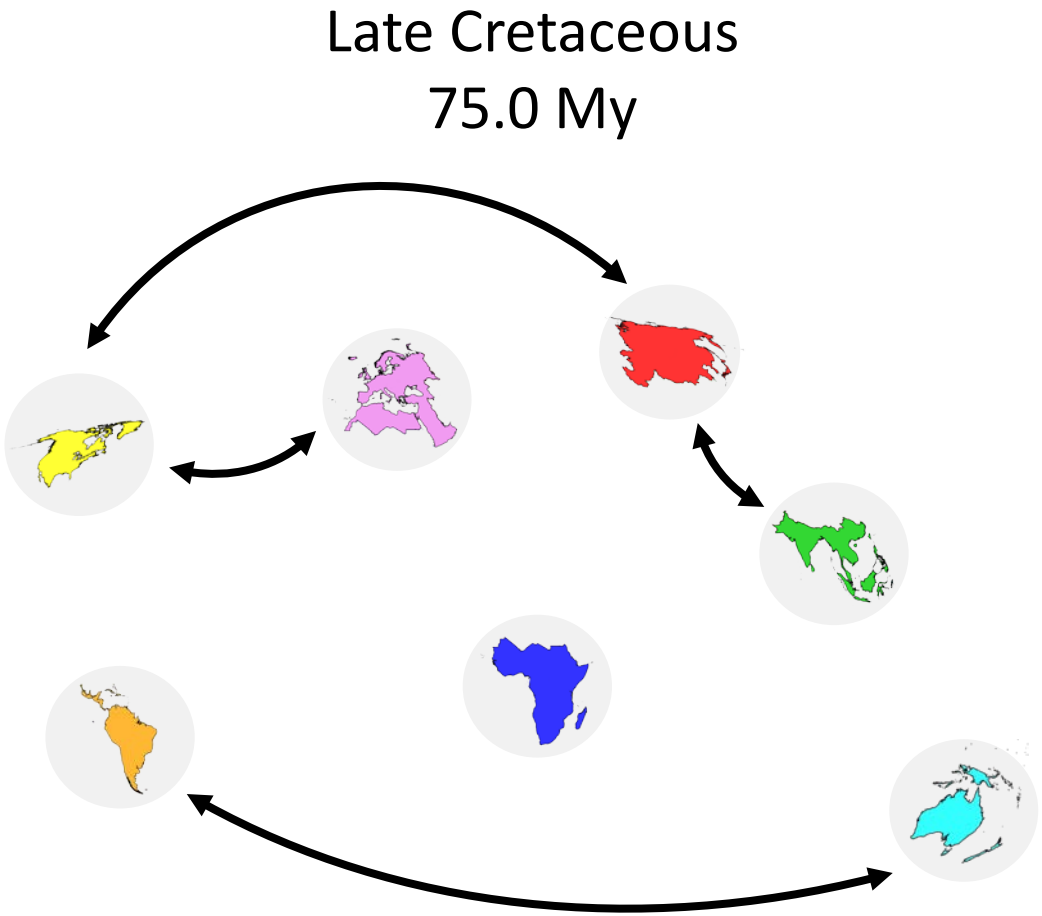


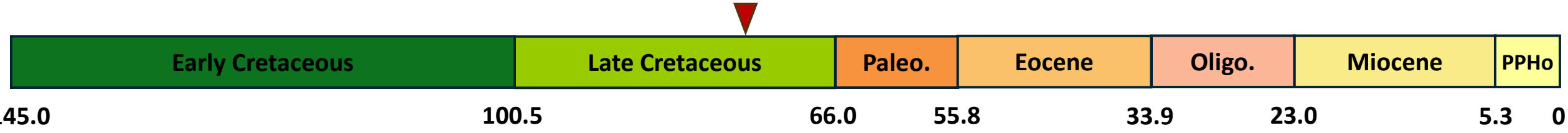


# Biogeographic methods

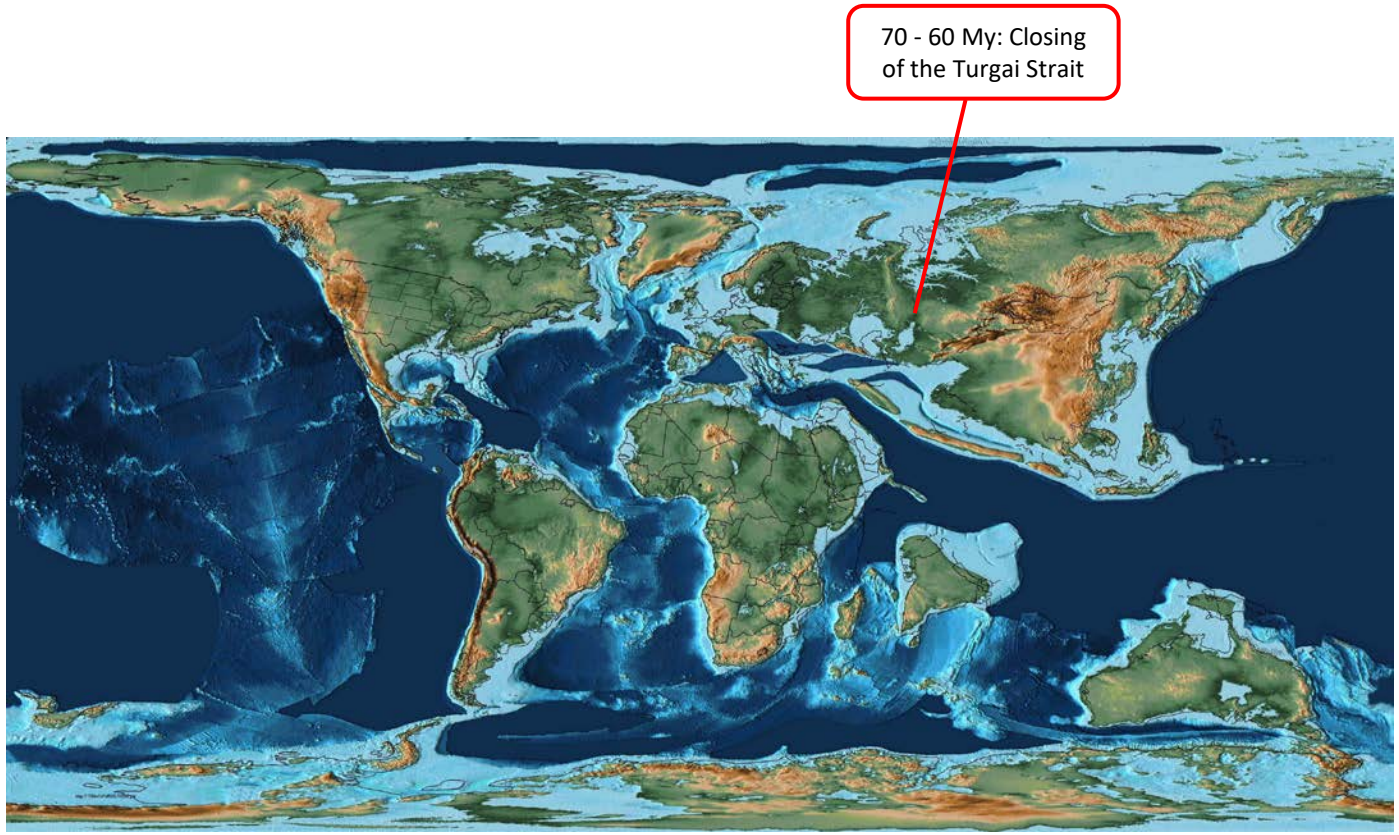


Source: PALEOMAP Project by C. Scotese

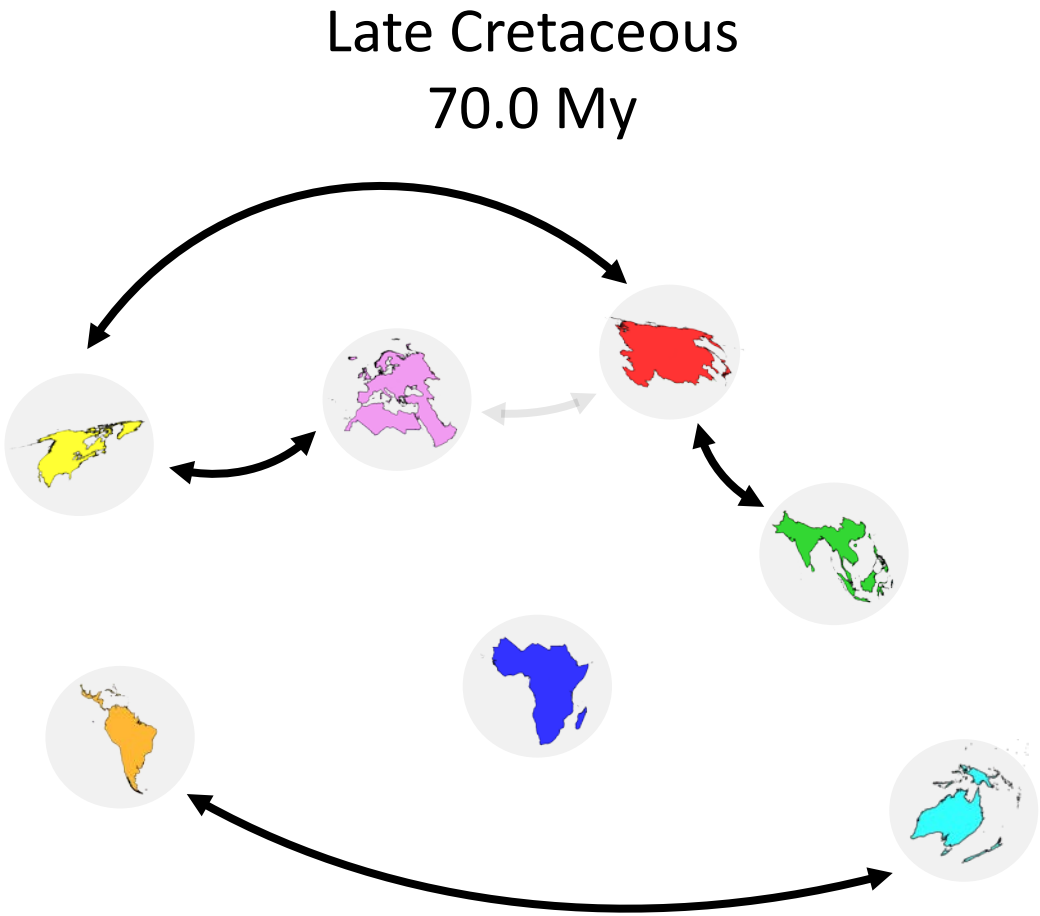


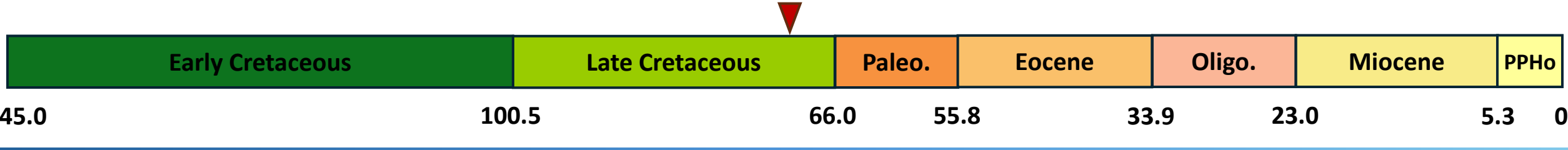



# Biogeographic methods

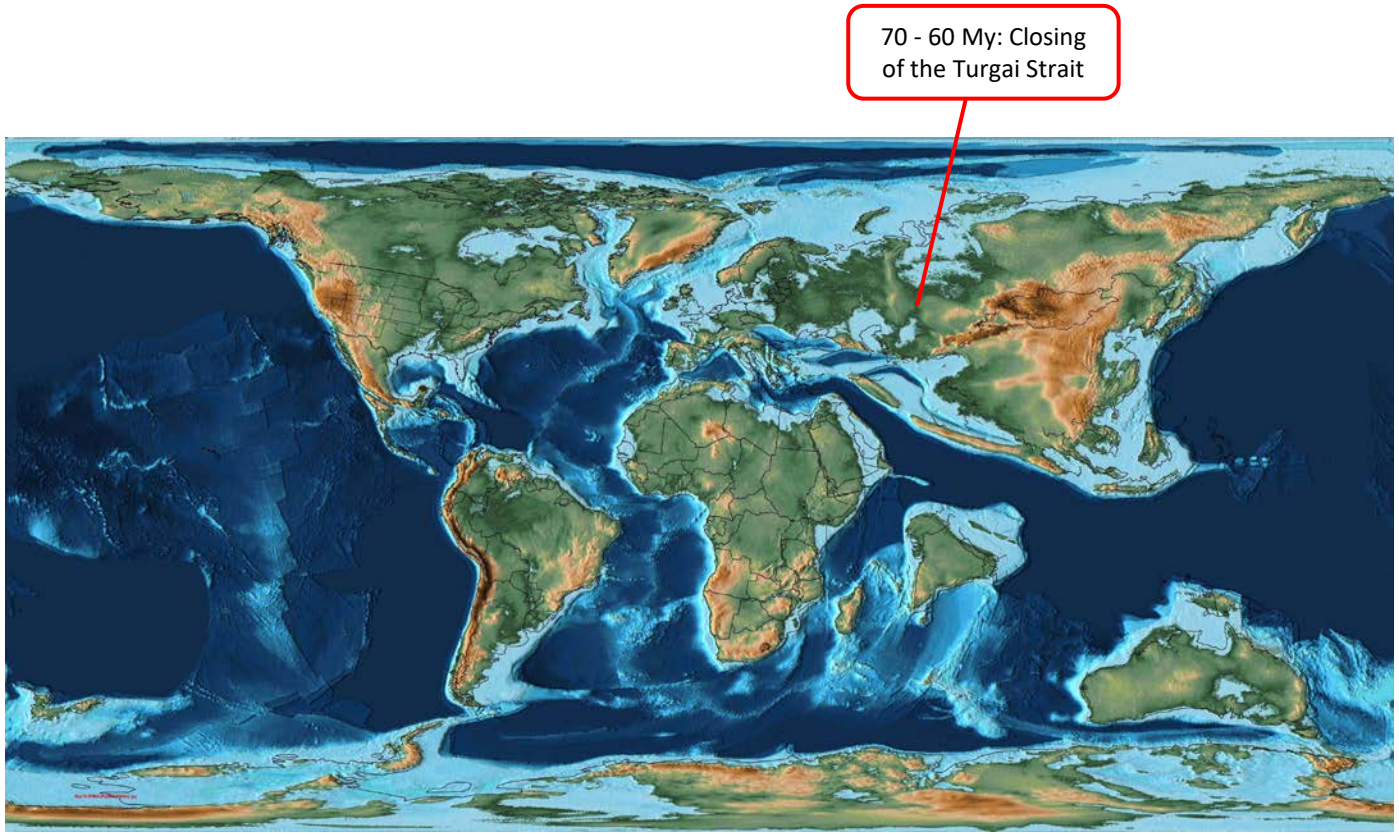



Source: PALEOMAP Project by C. Scotese



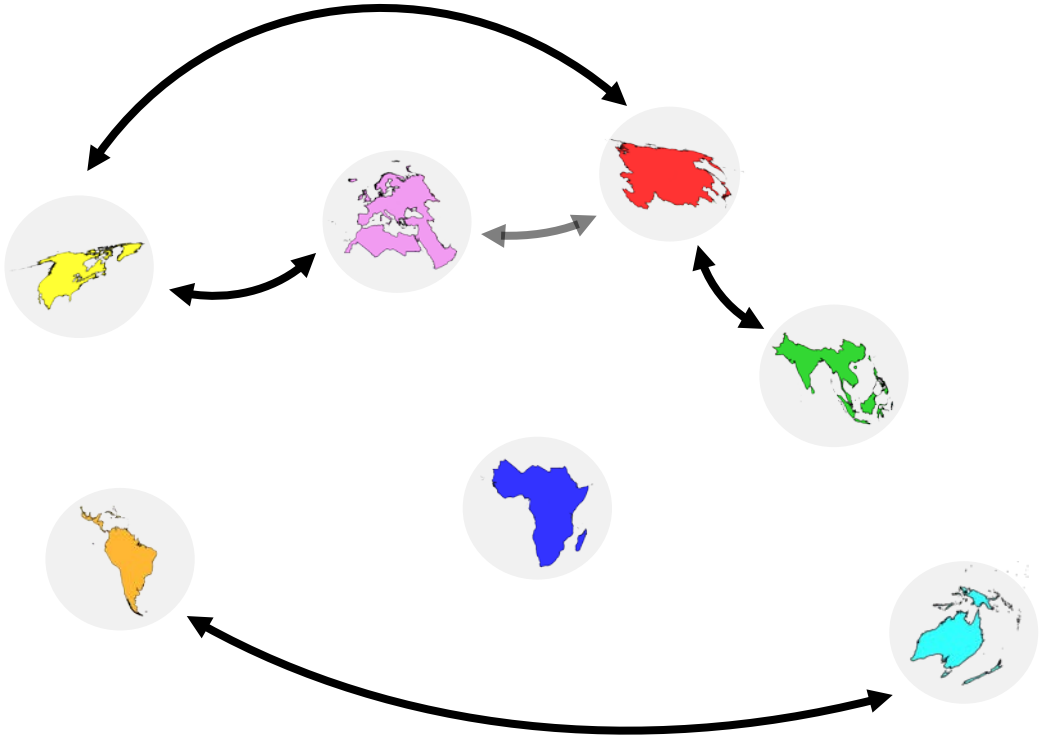



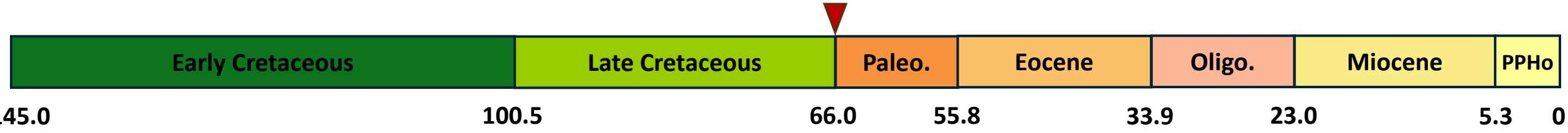
# Biogeographic methods



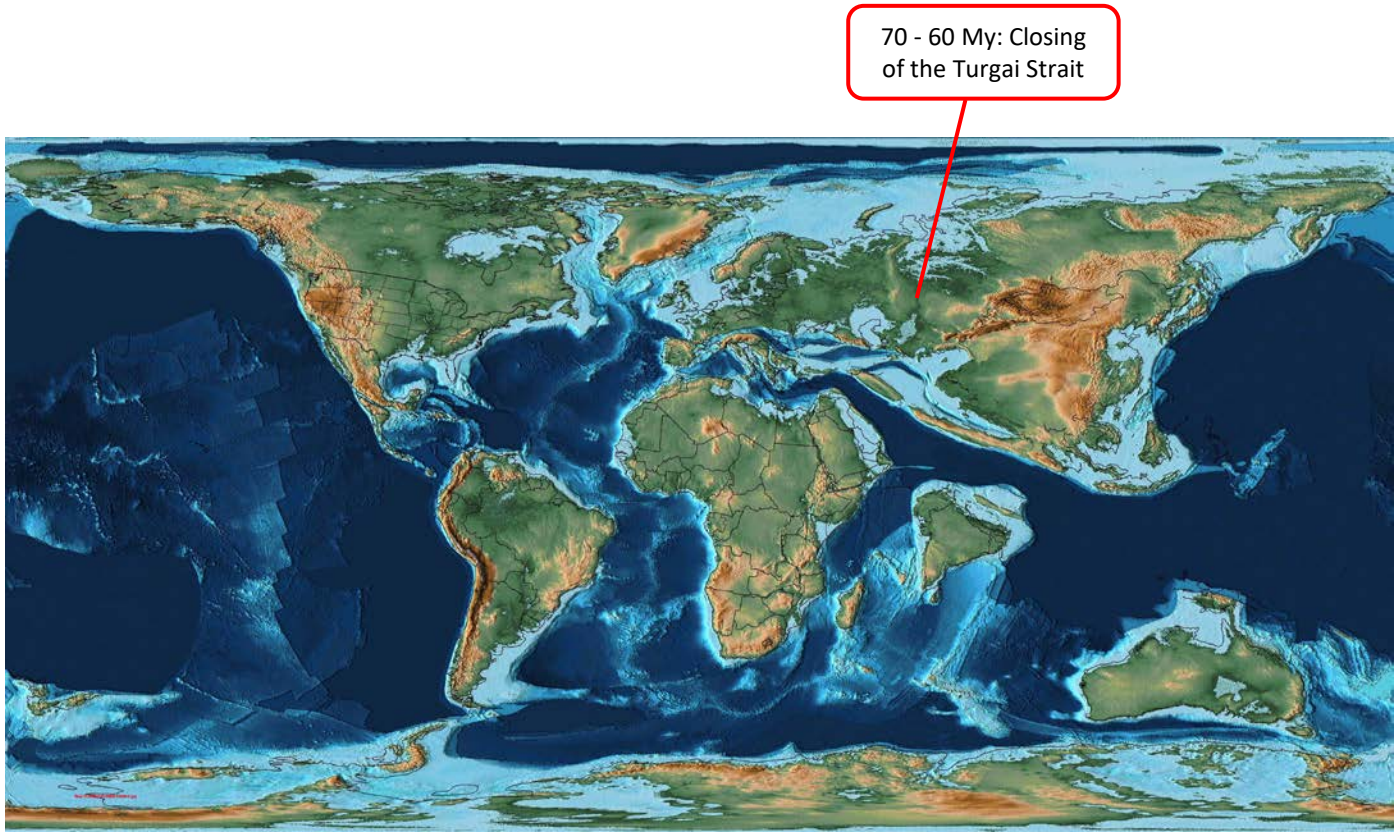

Source: PALEOMAP Project by C. Scotese

Late Cretaceous to Paleocene
66.0 My

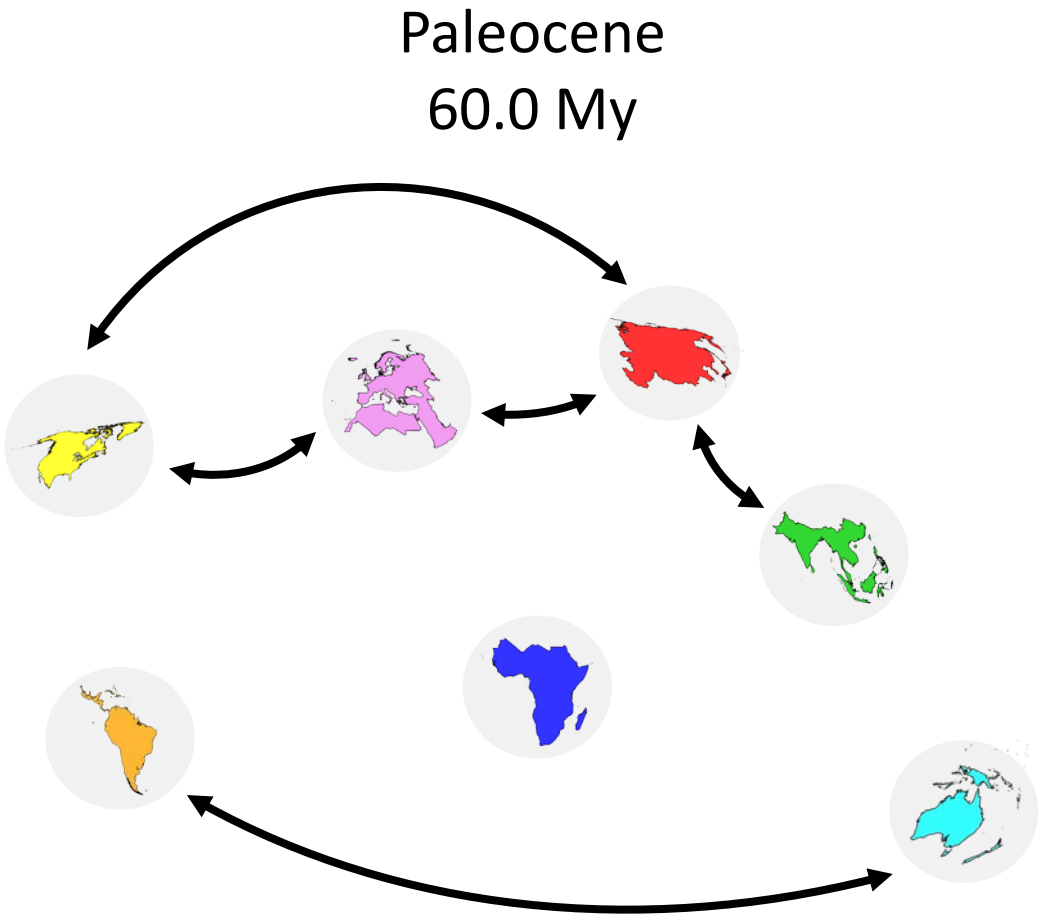


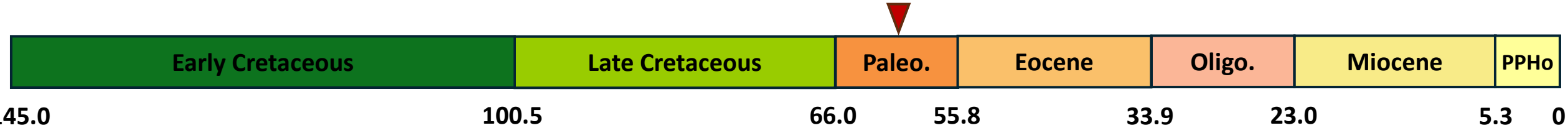


# Biogeographic methods

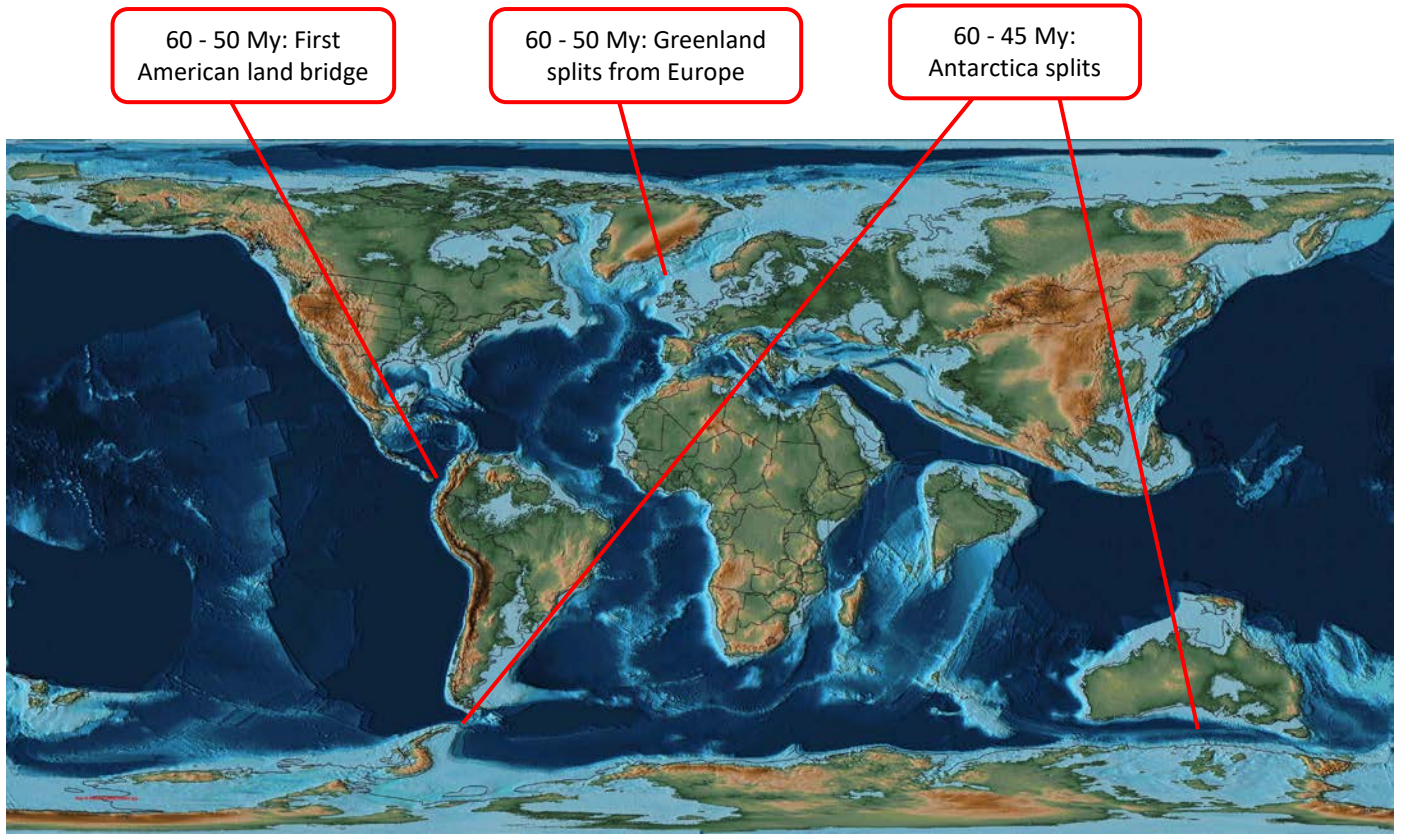



Source: PALEOMAP Project by C. Scotese



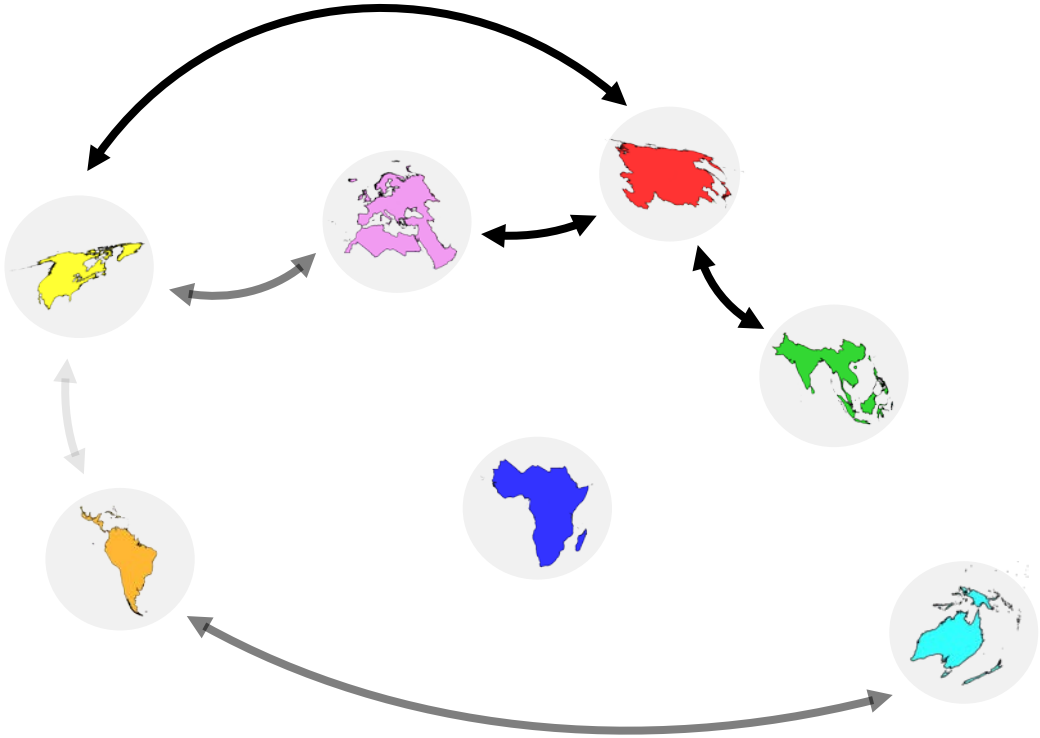



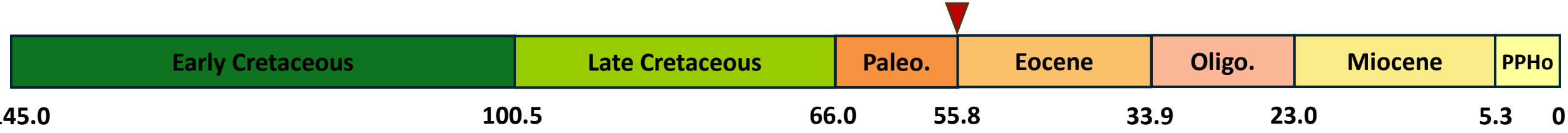
# Biogeographic methods



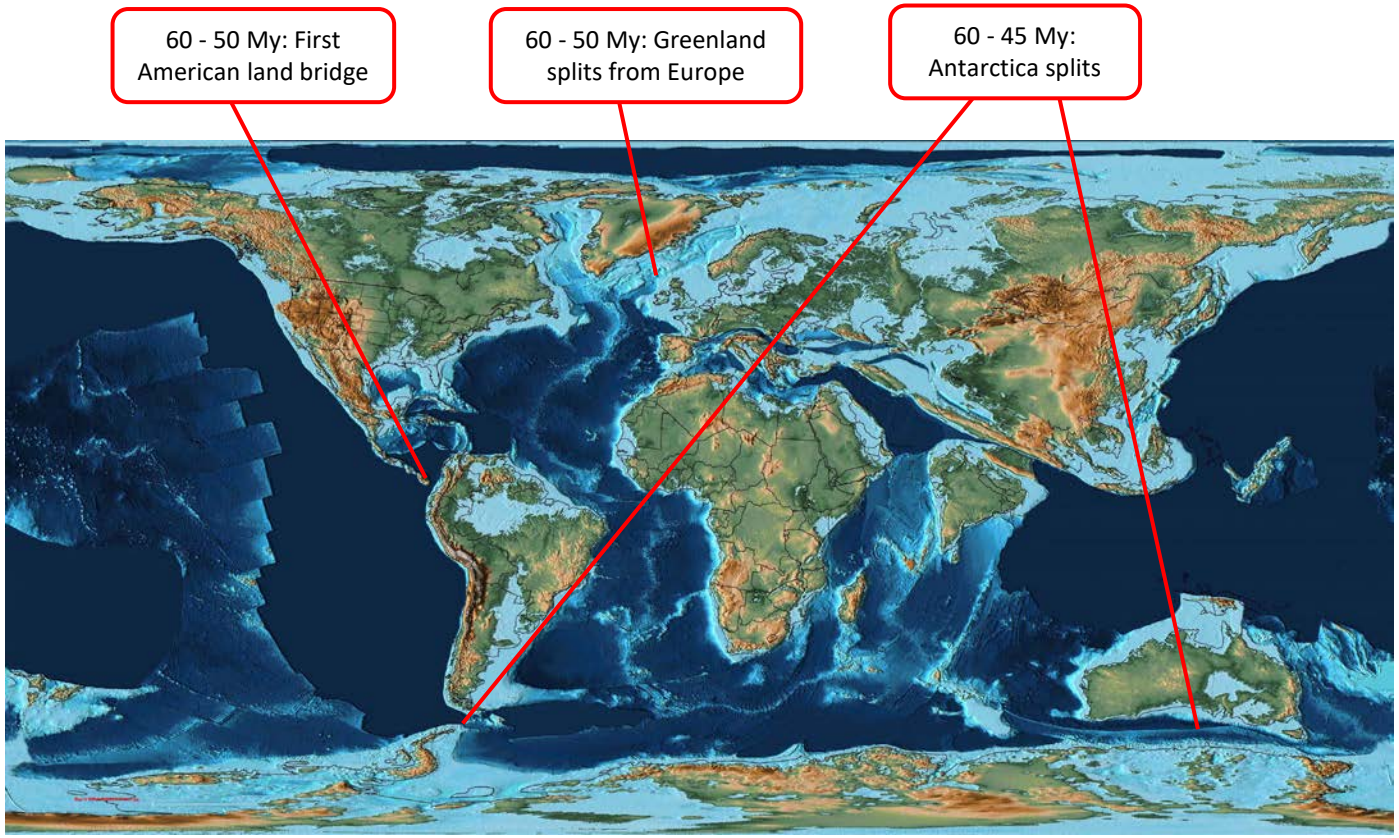

Source: PALEOMAP Project by C. Scotese

Paleocene to Eocene
55.8 My

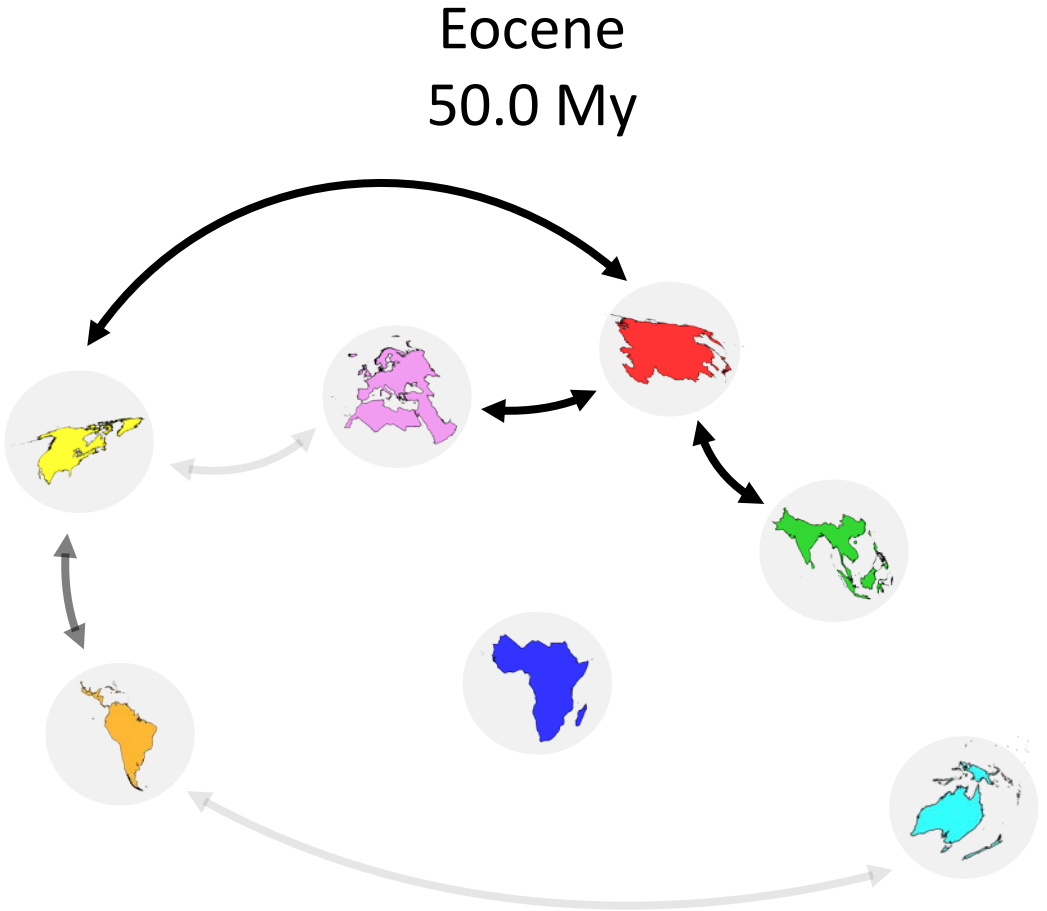


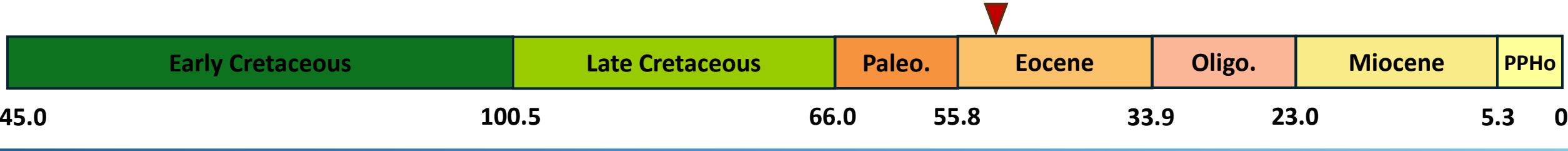


# Biogeographic methods

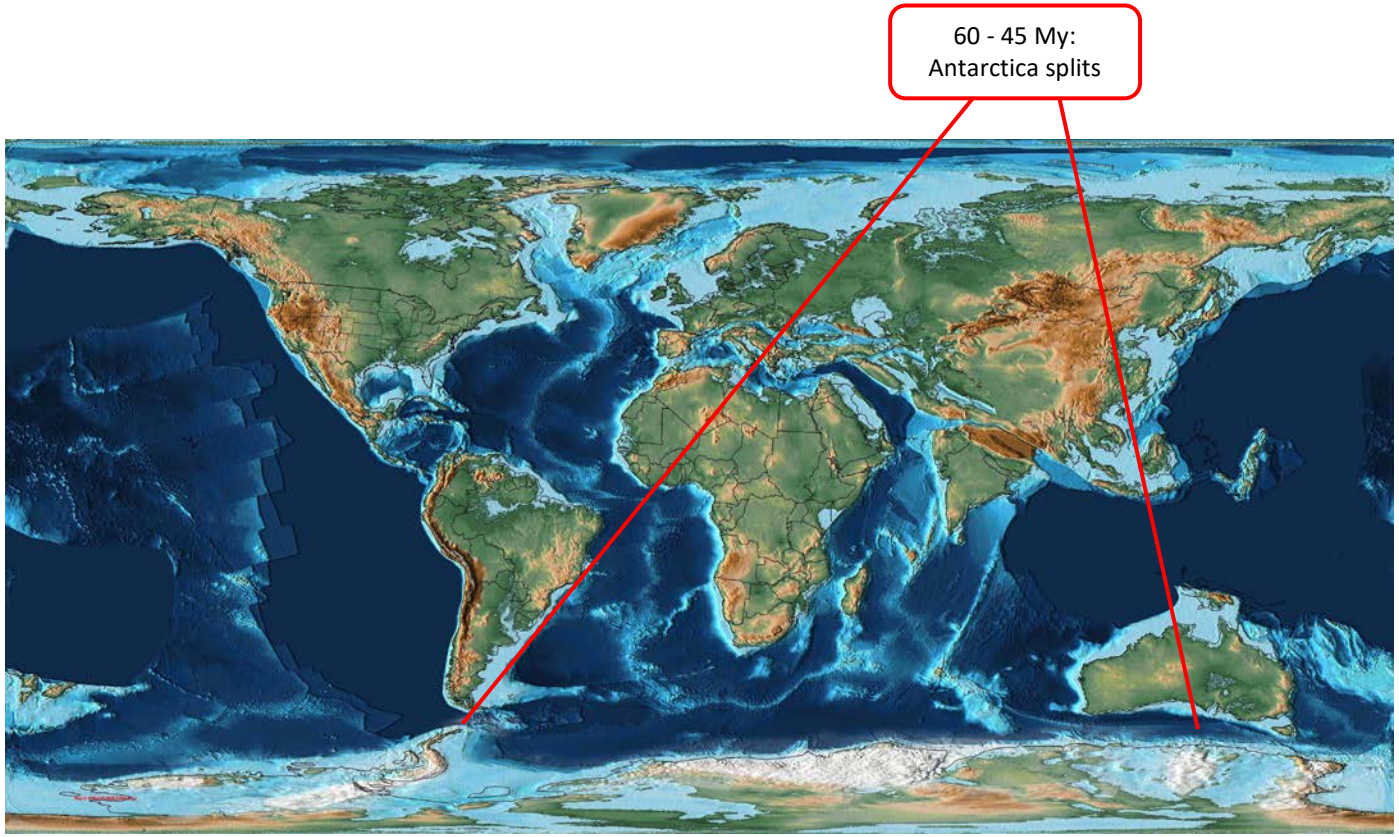



Source: PALEOMAP Project by C. Scotese

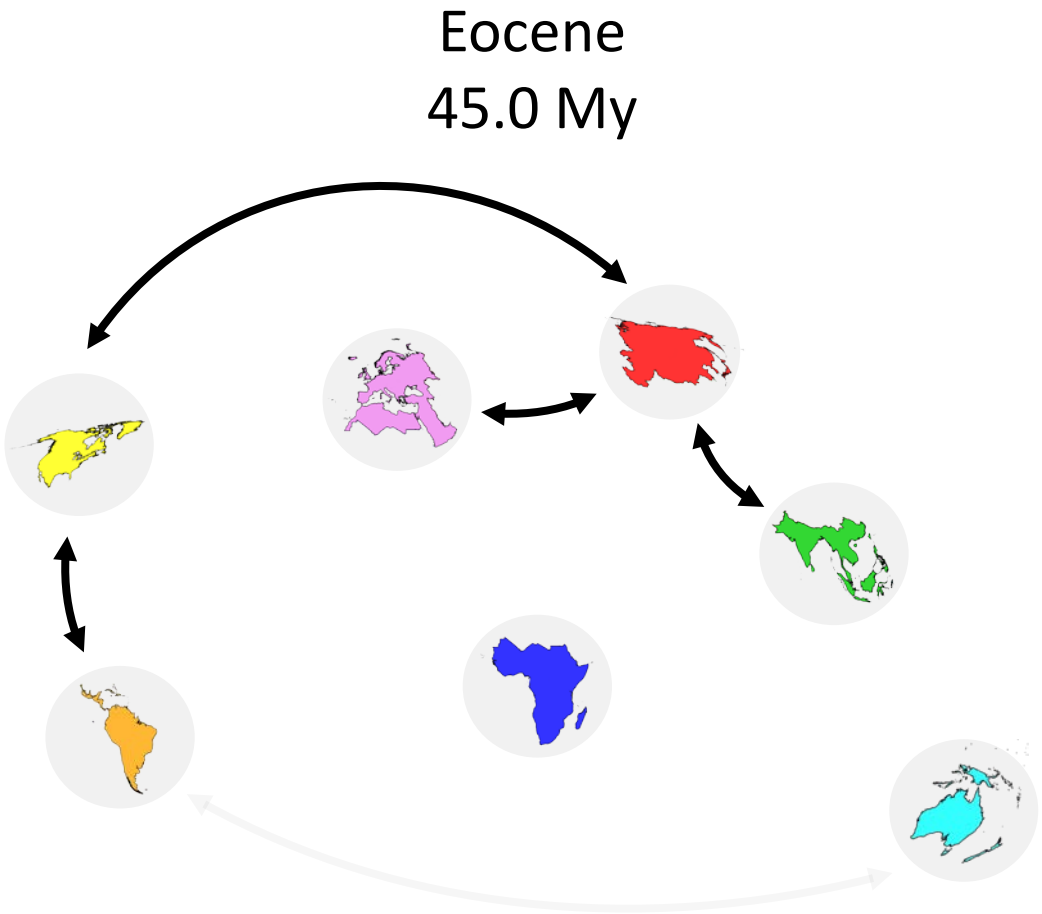


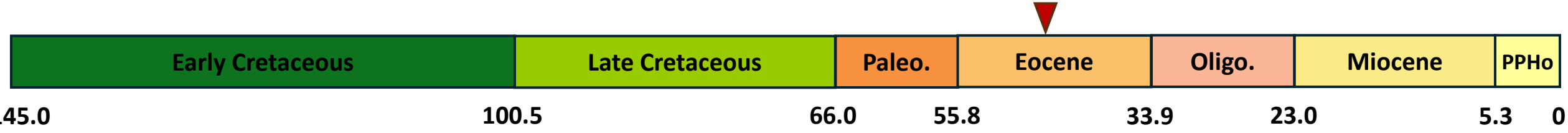



# Biogeographic methods

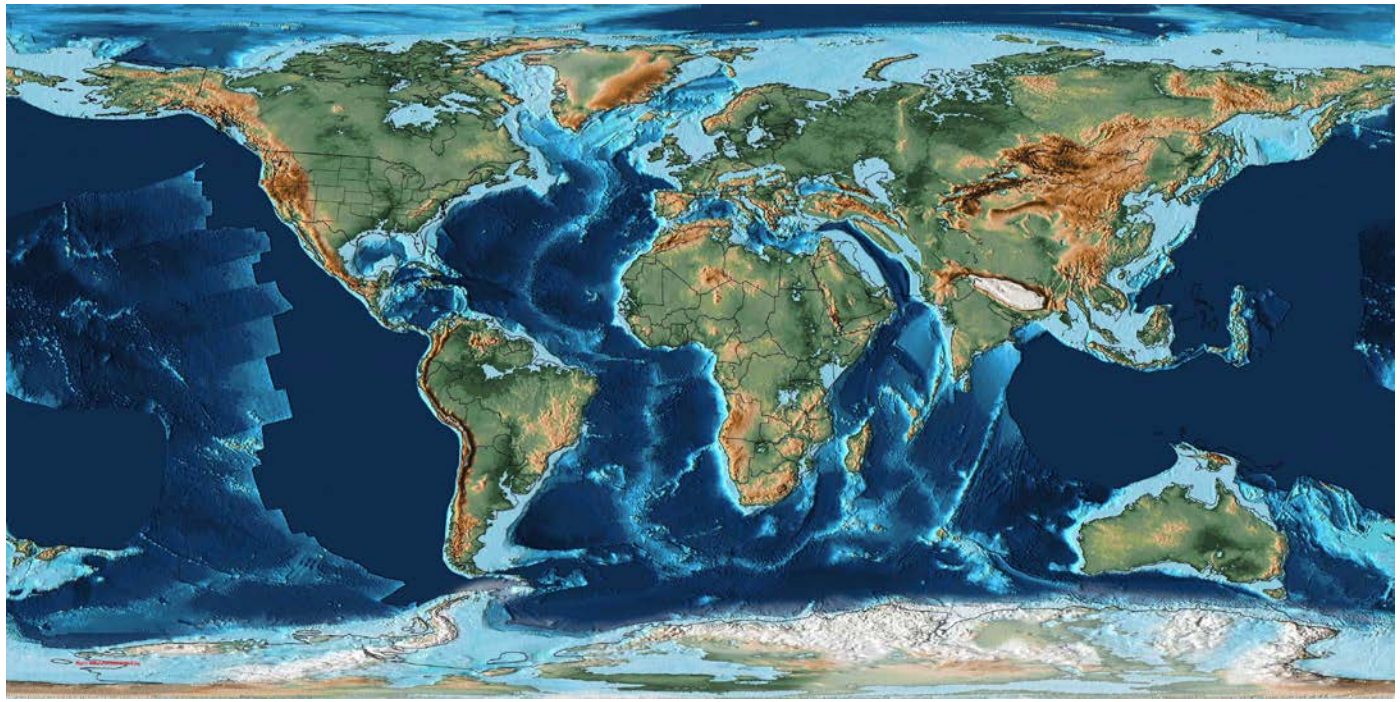



Source: PALEOMAP Project by C. Scotese

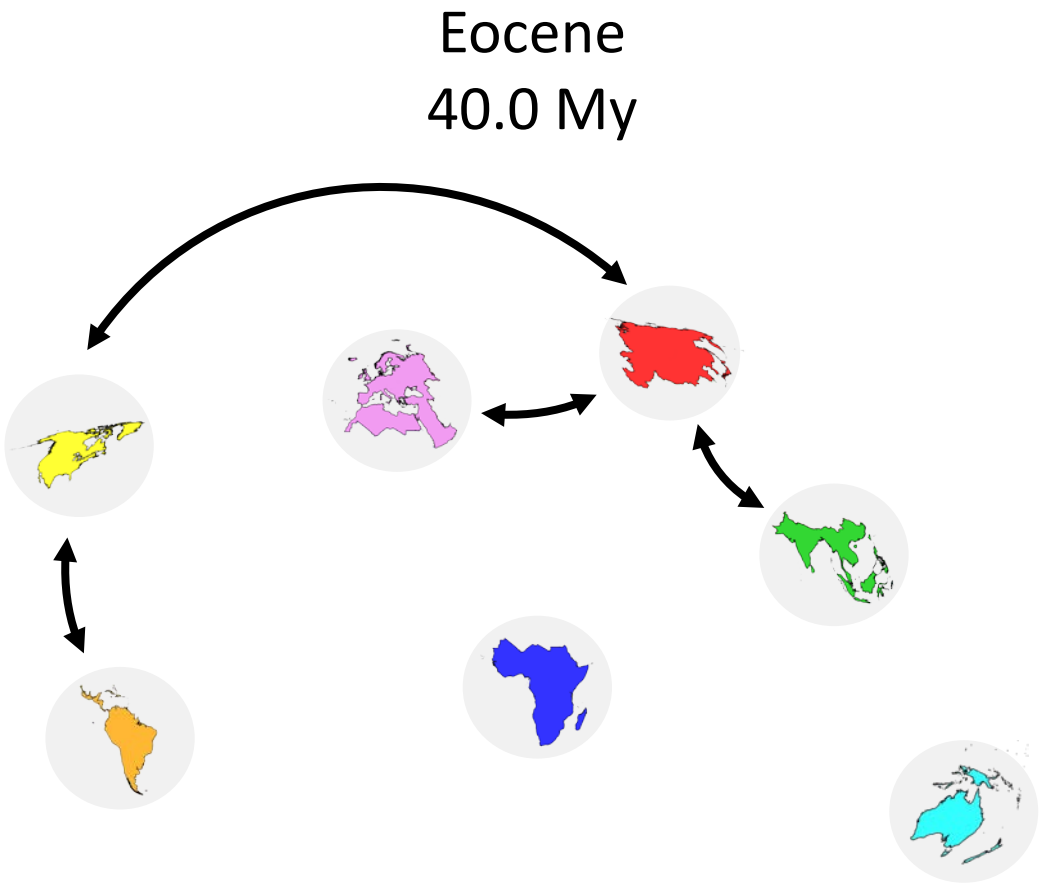


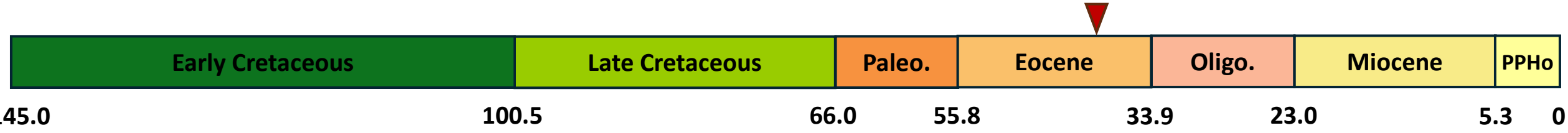



# Biogeographic methods



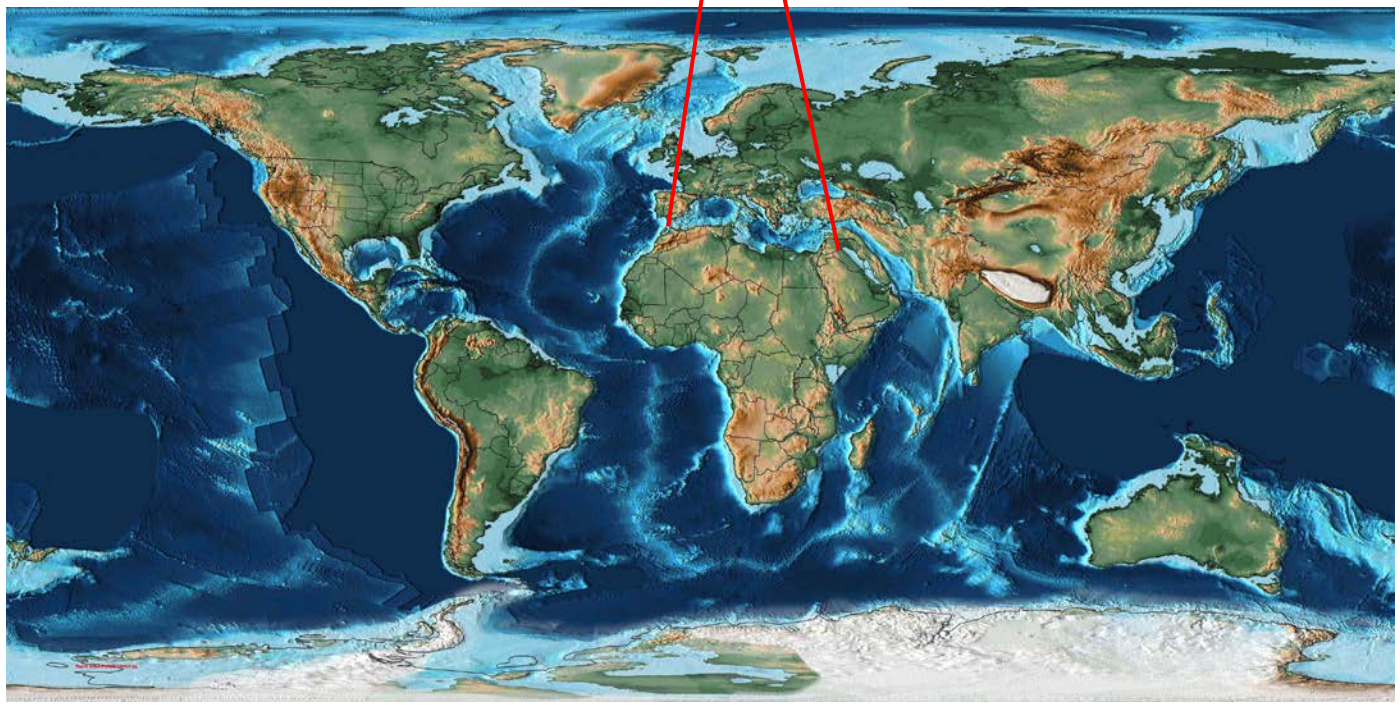
Source: PALEOMAP Project by C. Scotese



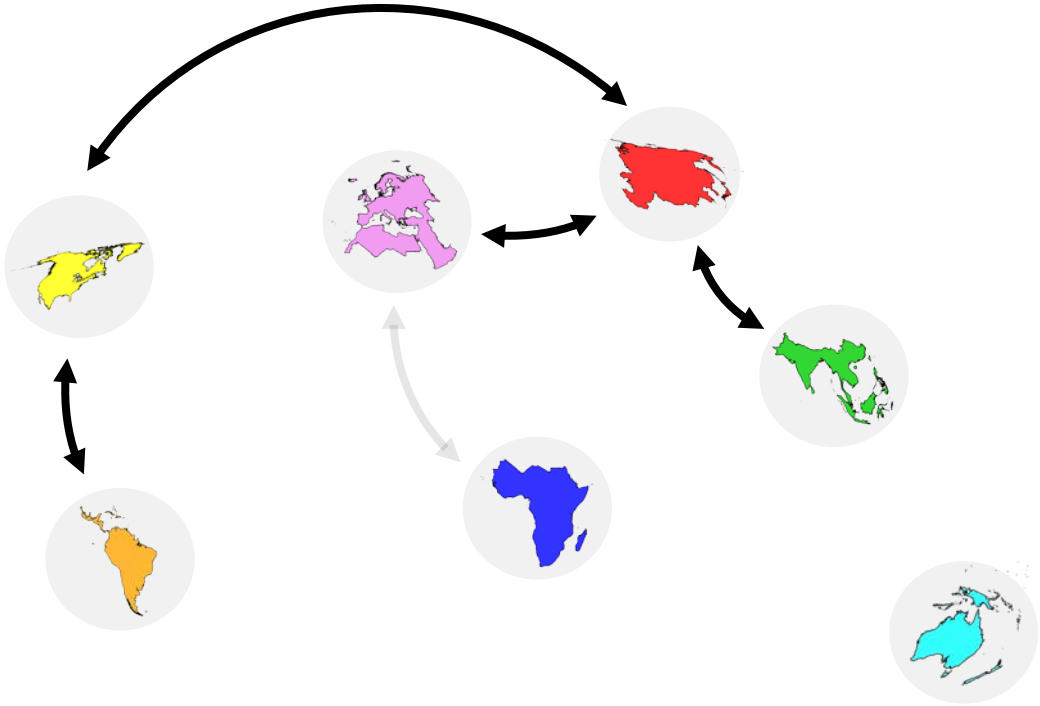
# Biogeographic methods

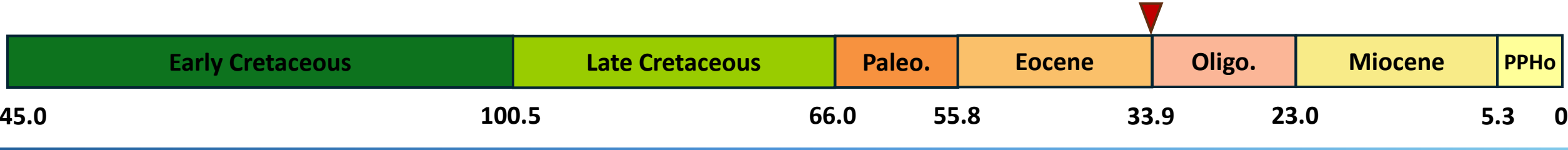
35 – 30 My: Africa collides with Palearctic



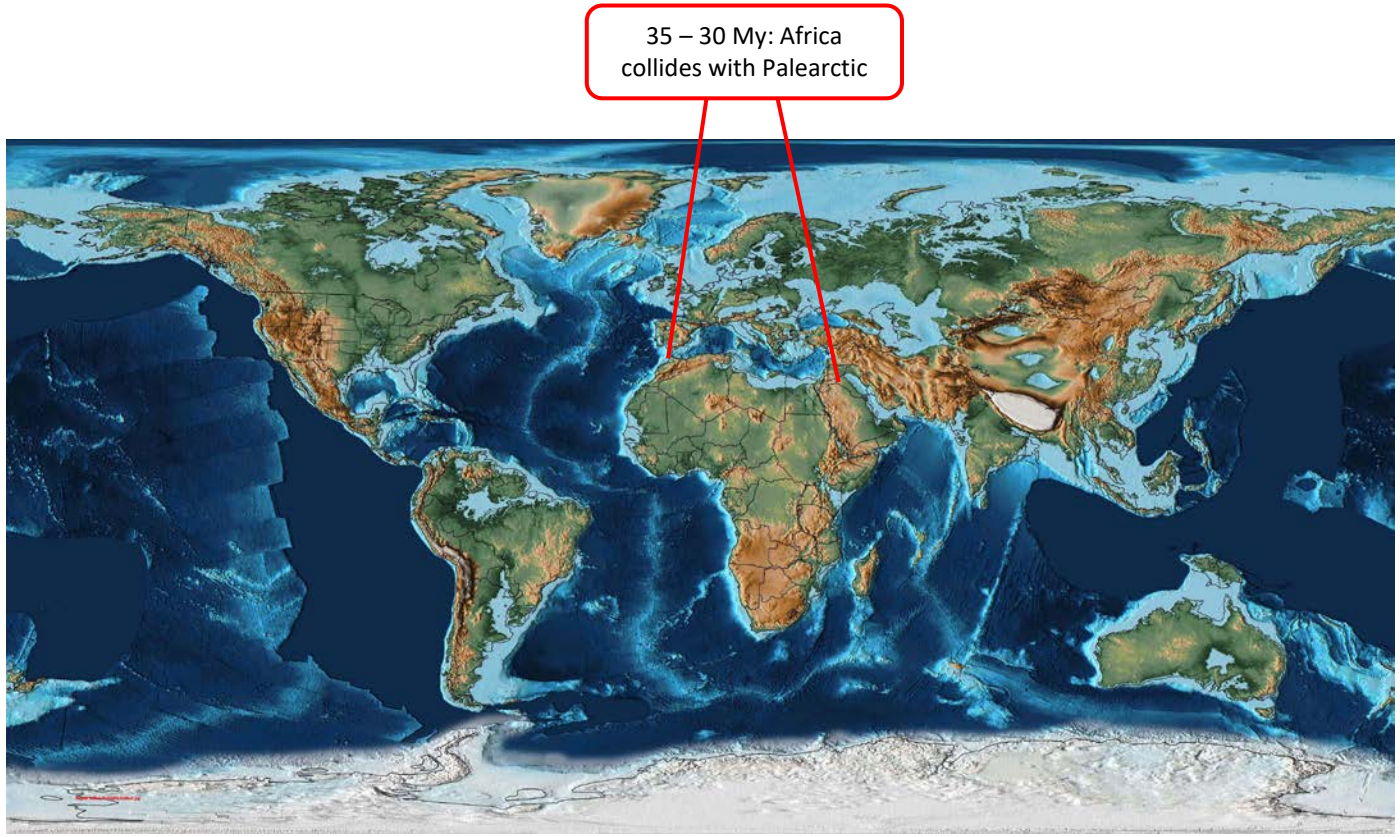
Source: PALEOMAP Project by C. Scotese

Eocene to Oligocene
33.9 My

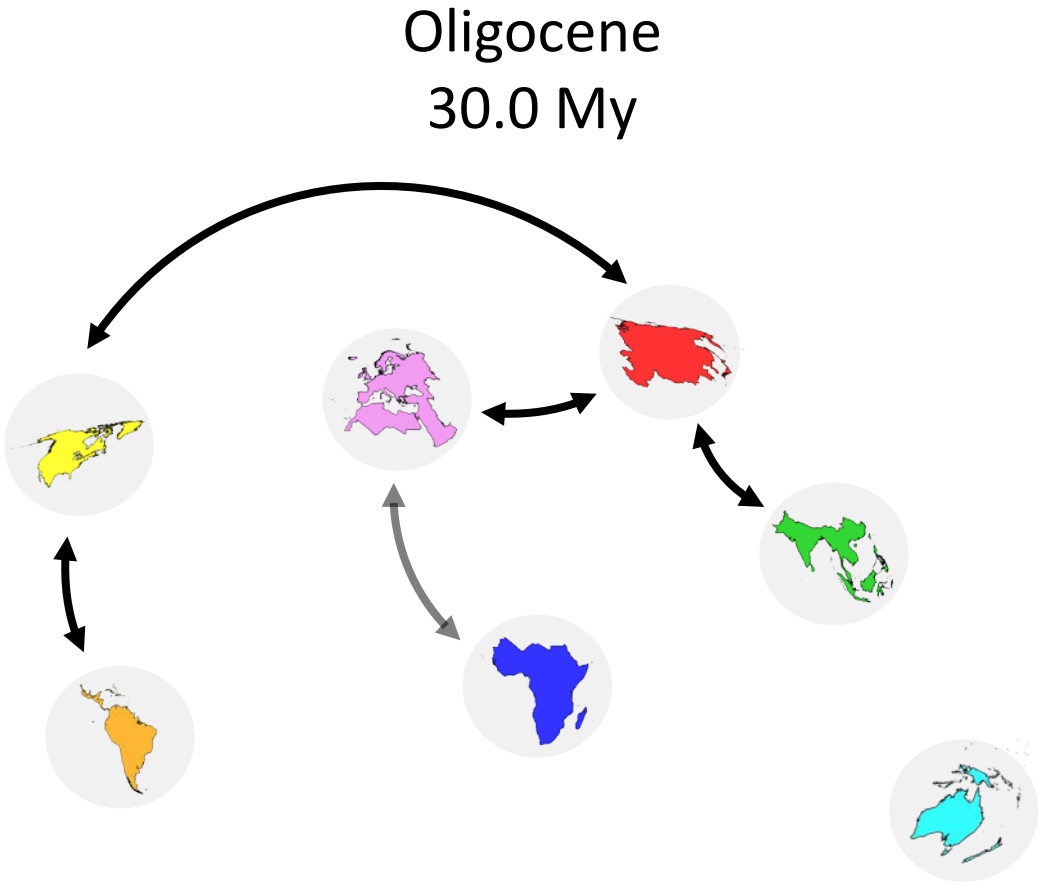


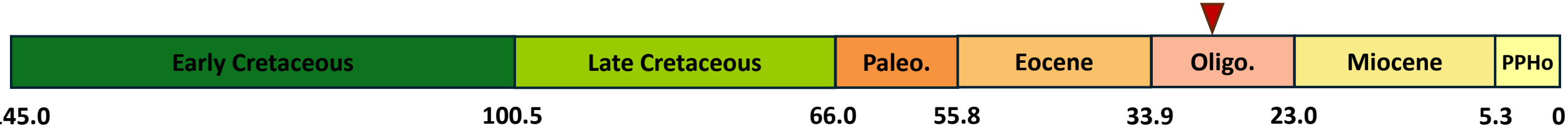


# Biogeographic methods

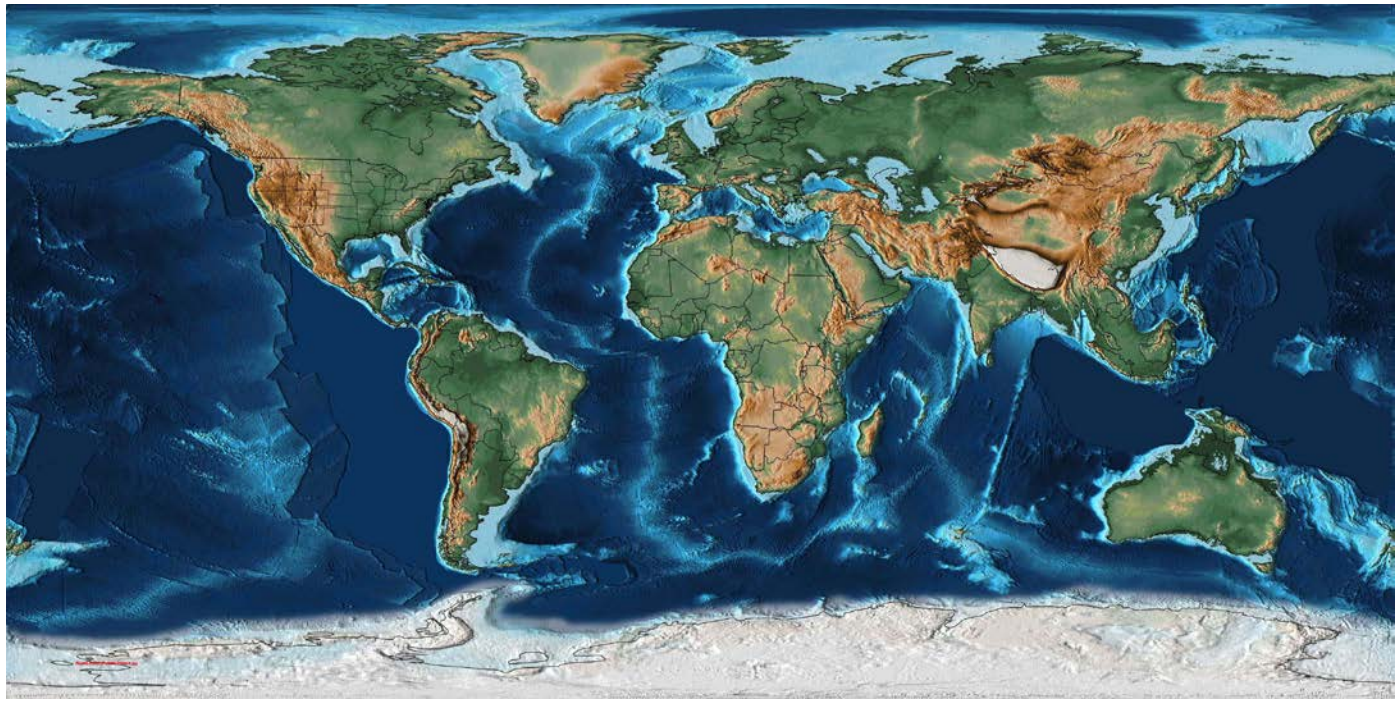



Source: PALEOMAP Project by C. Scotese

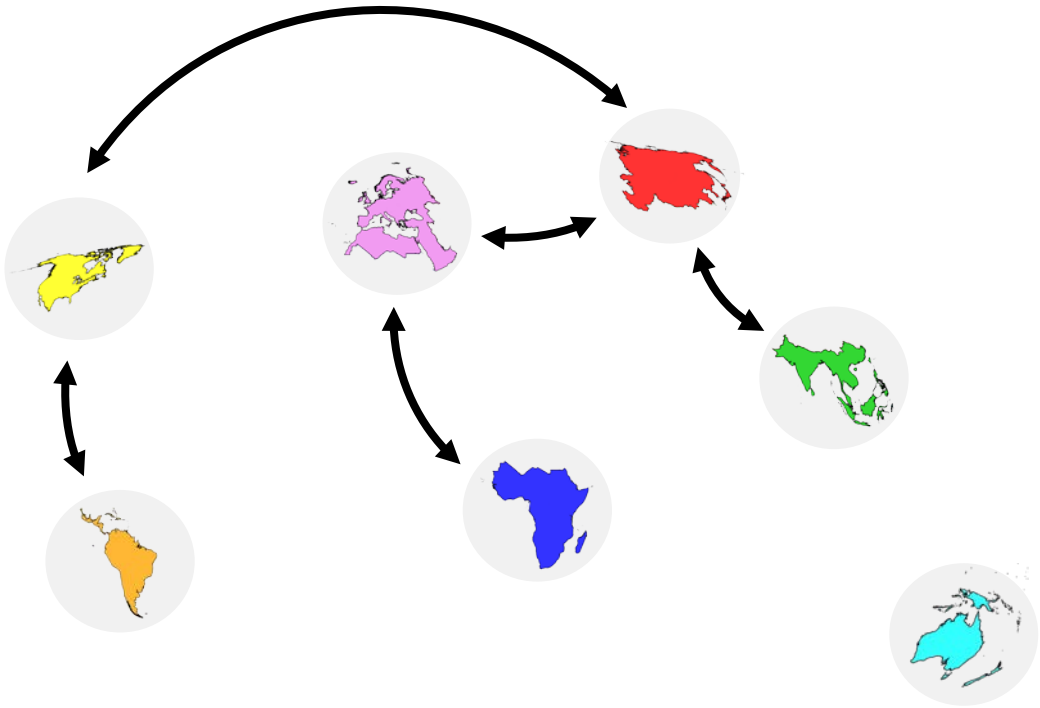


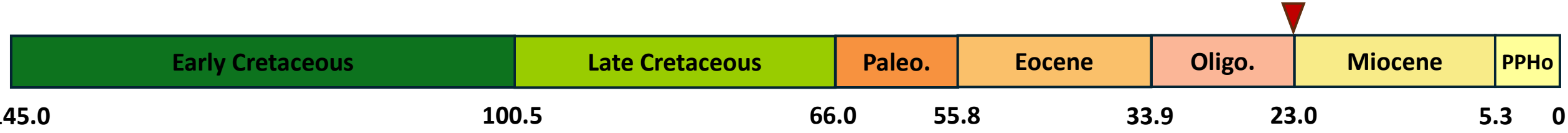



# Biogeographic methods

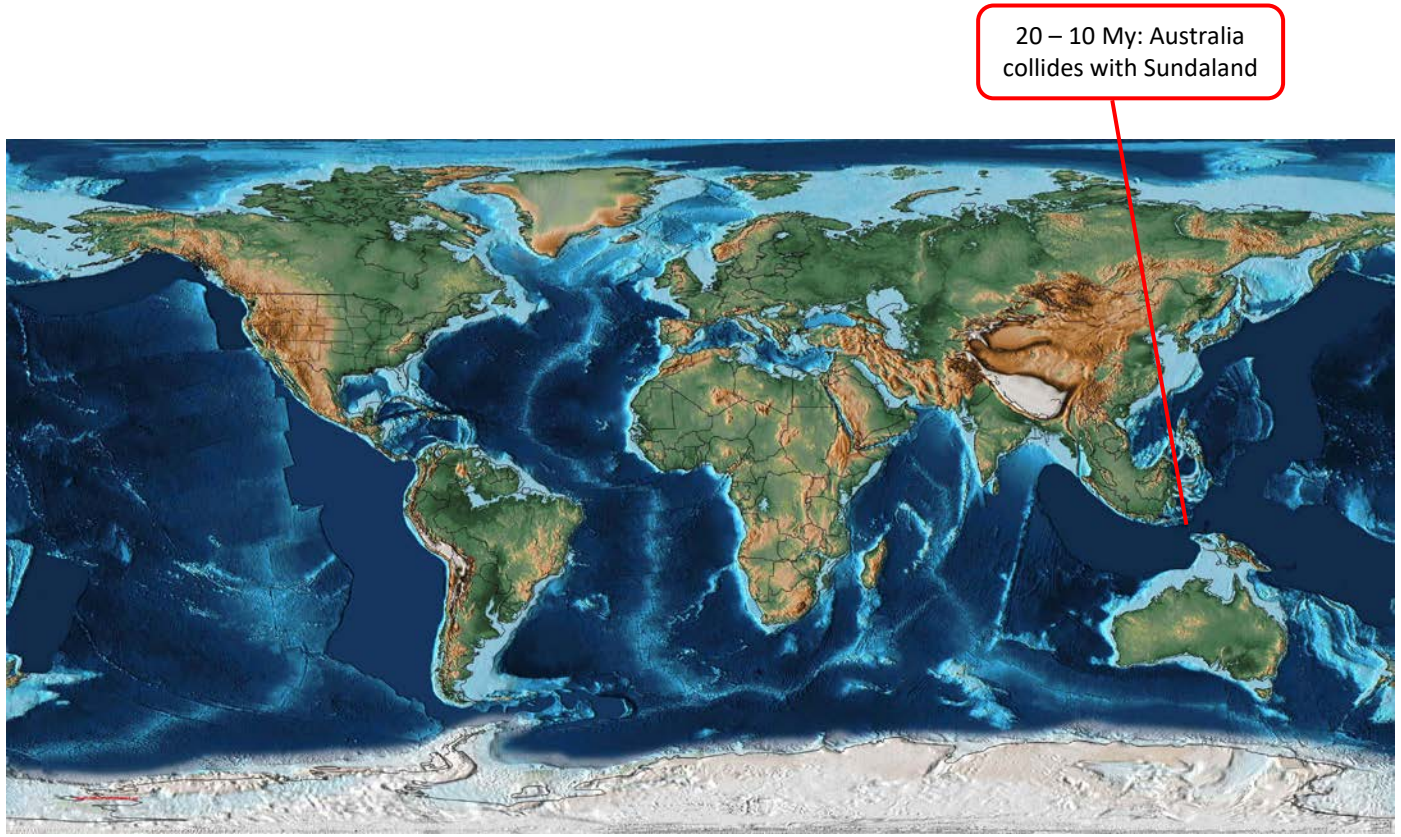


Source: PALEOMAP Project by C. Scotese

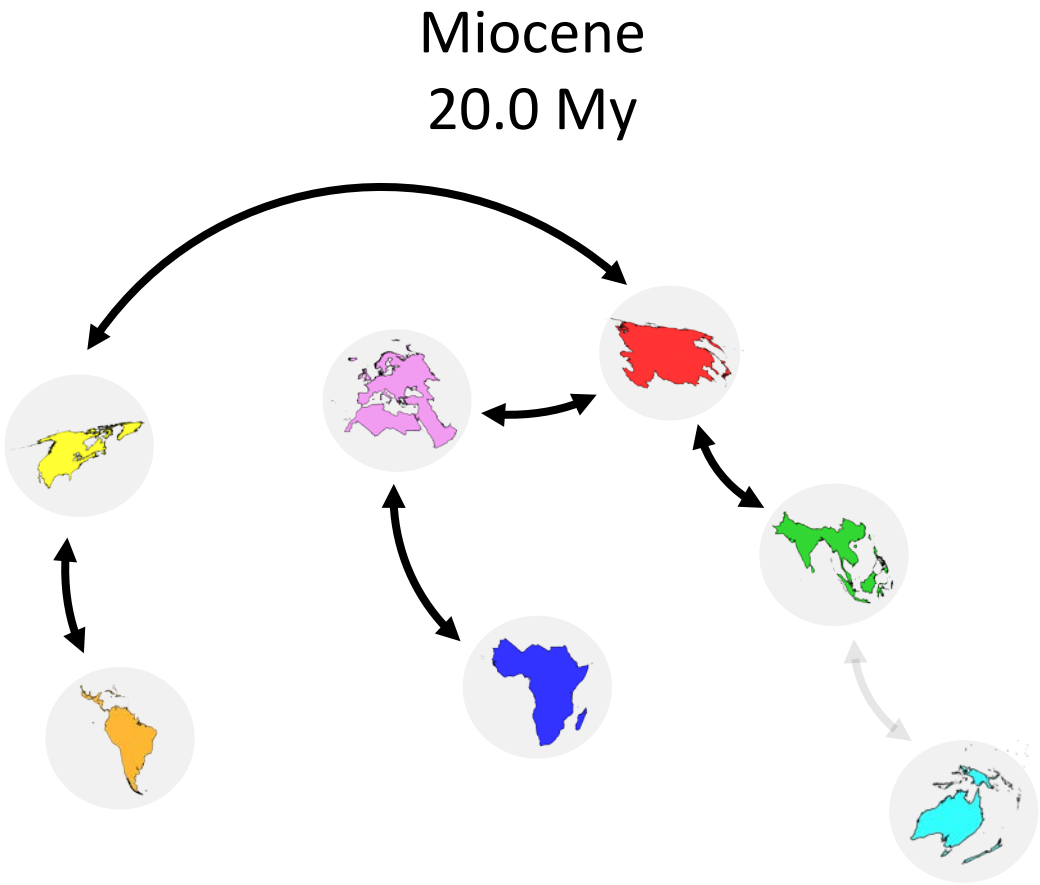


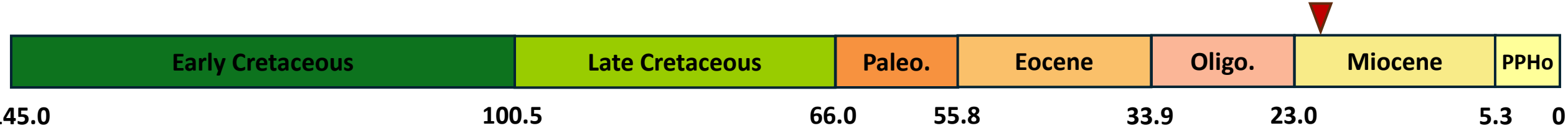



# Biogeographic methods

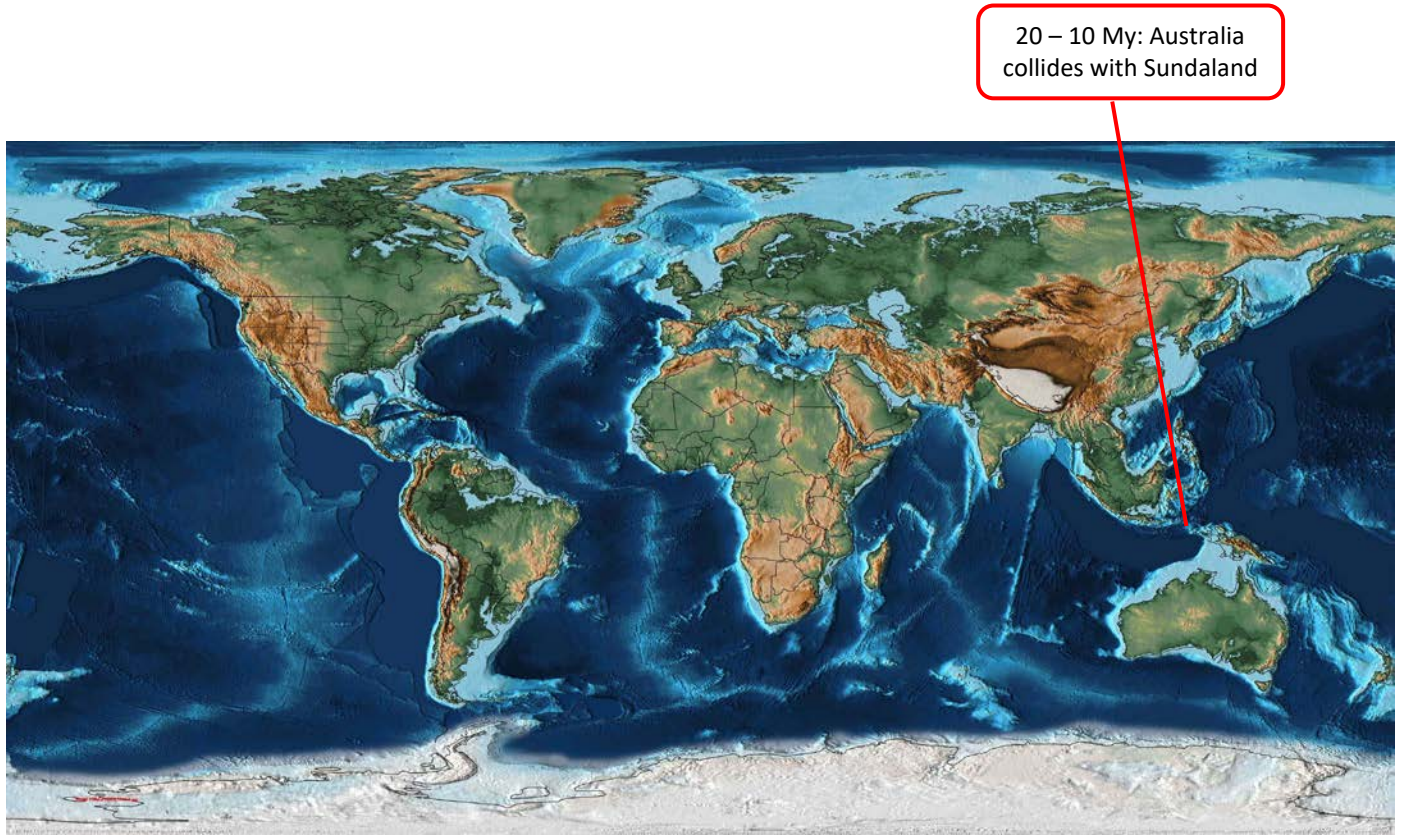



Source: PALEOMAP Project by C. Scotese

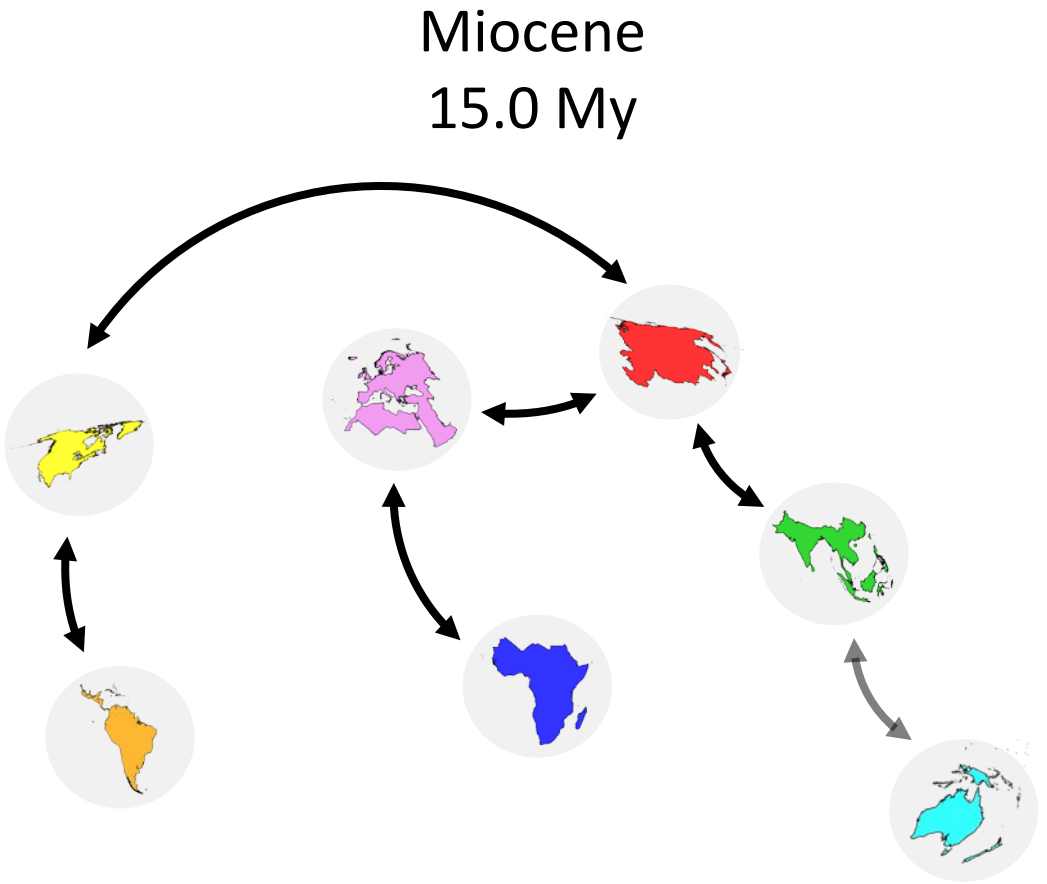


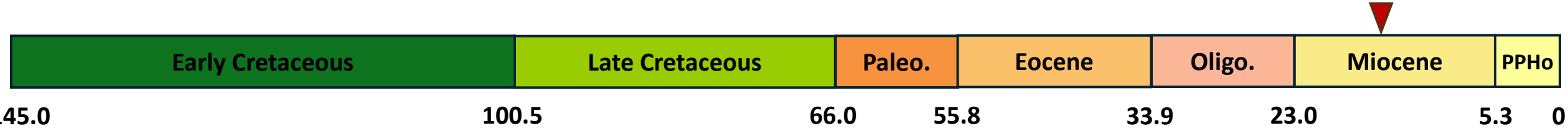



# Biogeographic methods

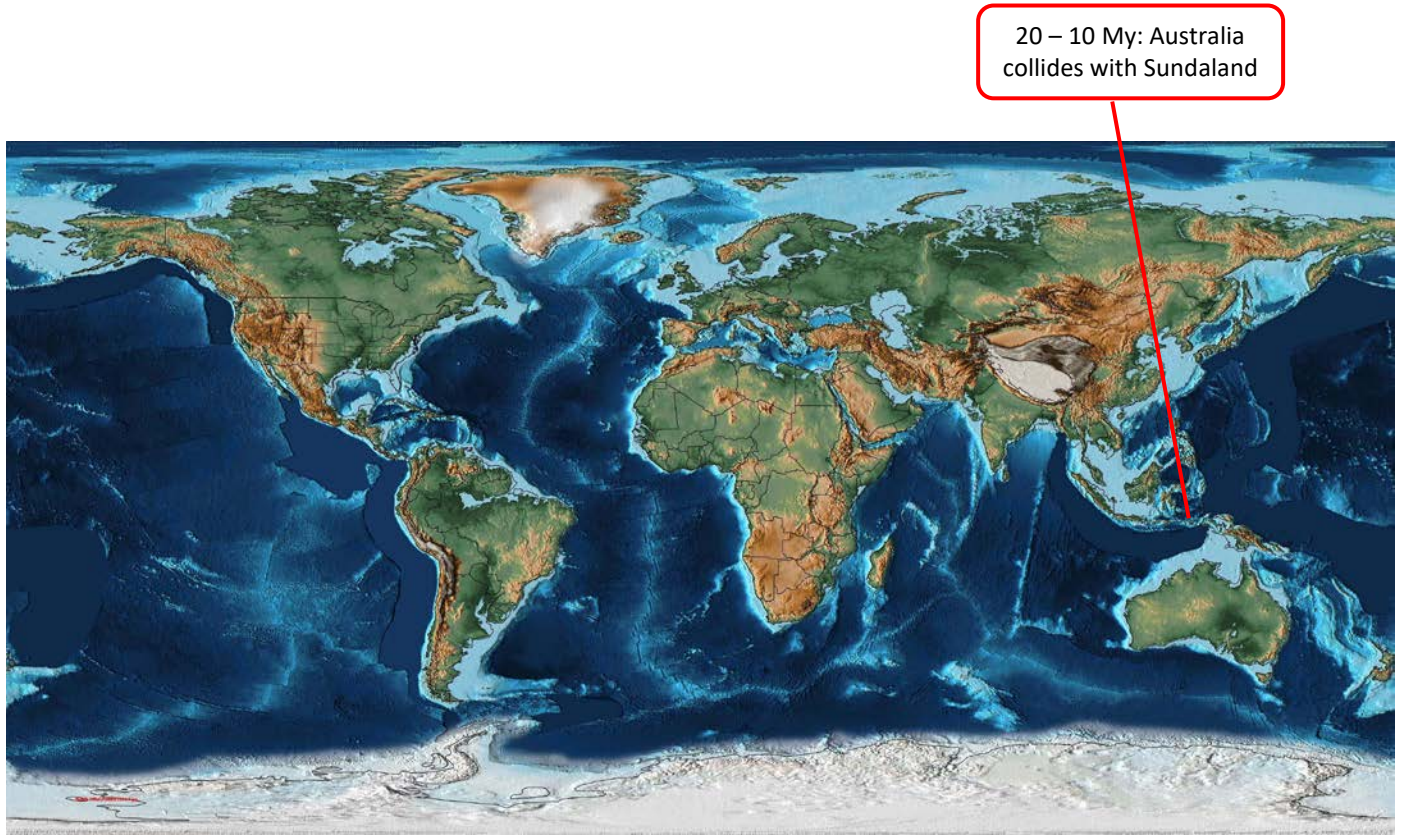



Source: PALEOMAP Project by C. Scotese

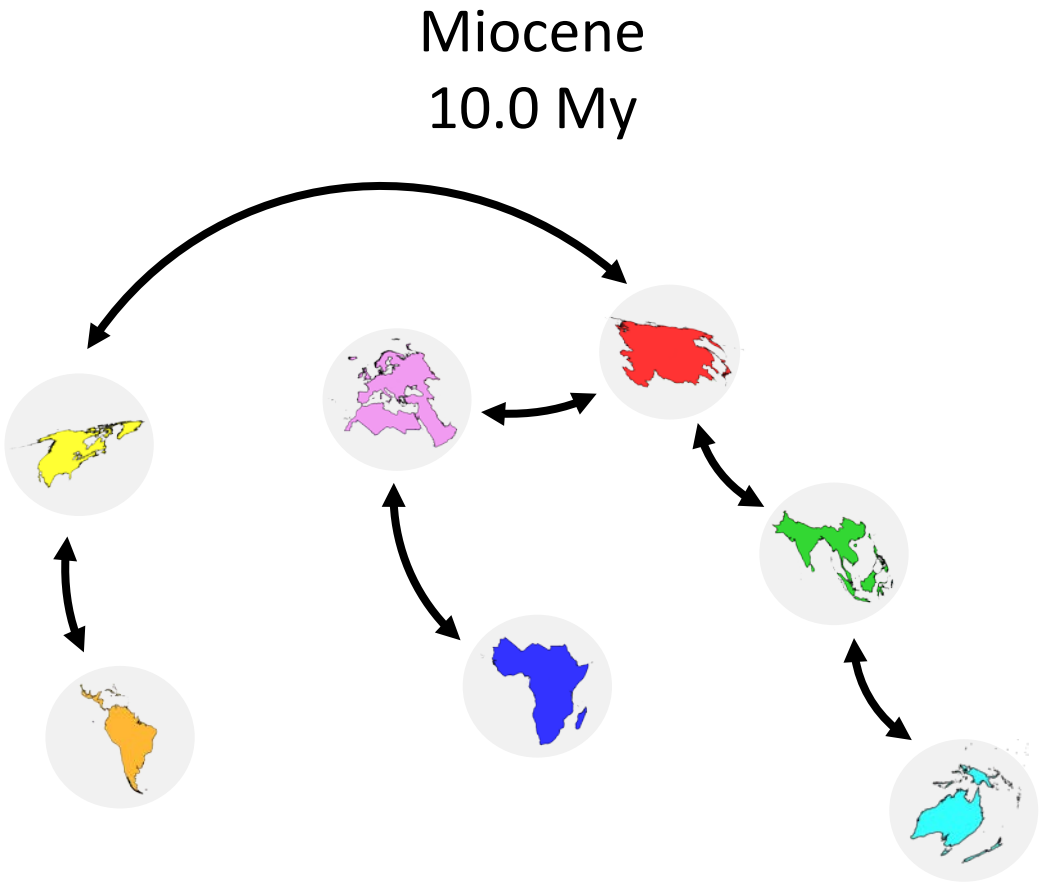


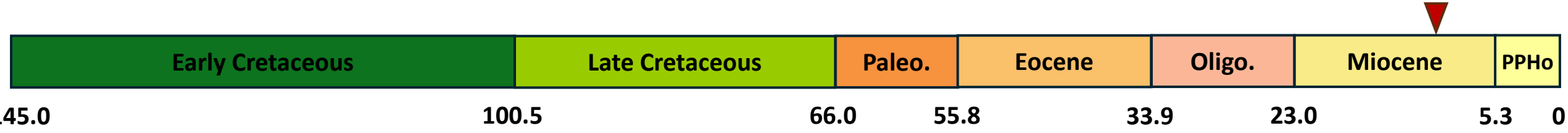



# Biogeographic methods

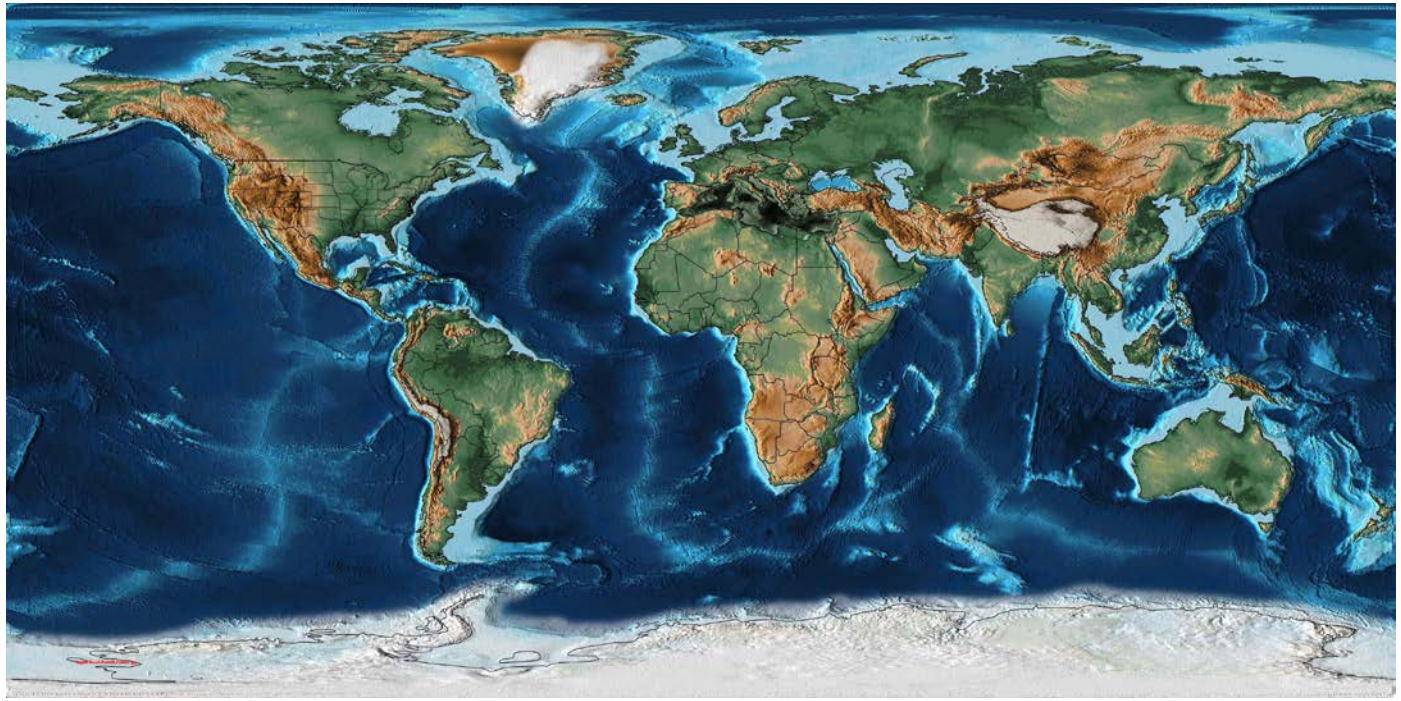



Source: PALEOMAP Project by C. Scotese



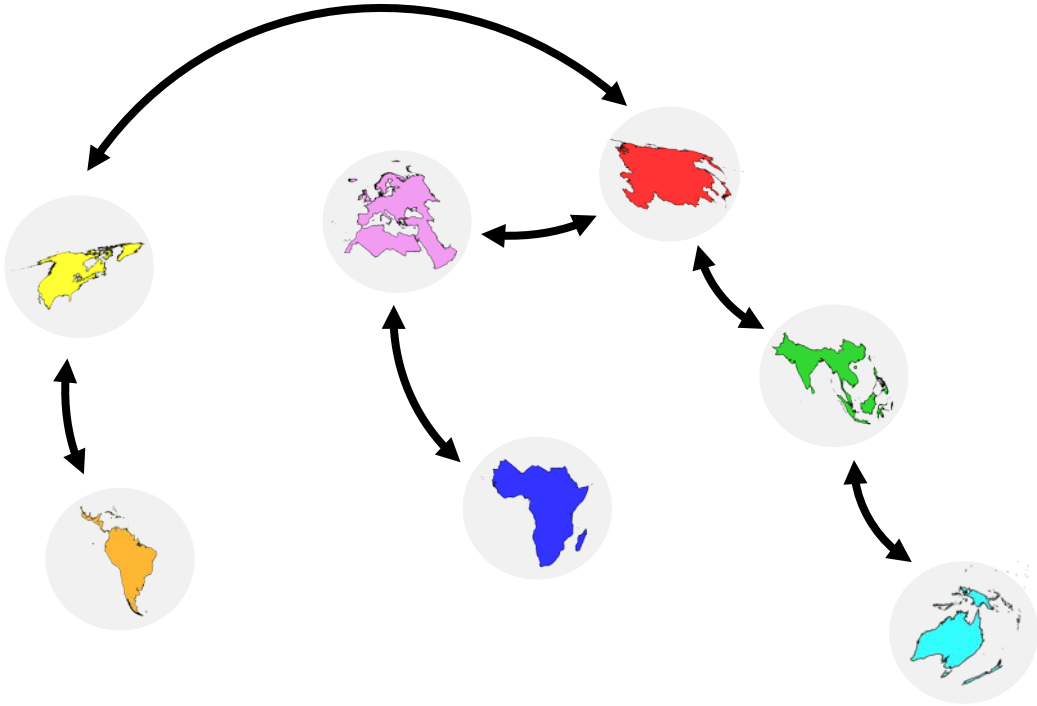



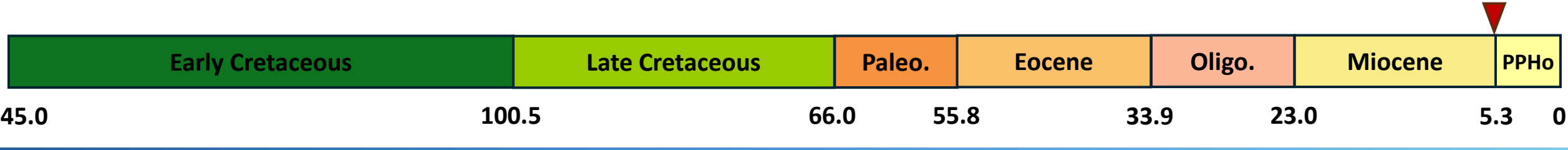
# Biogeographic methods



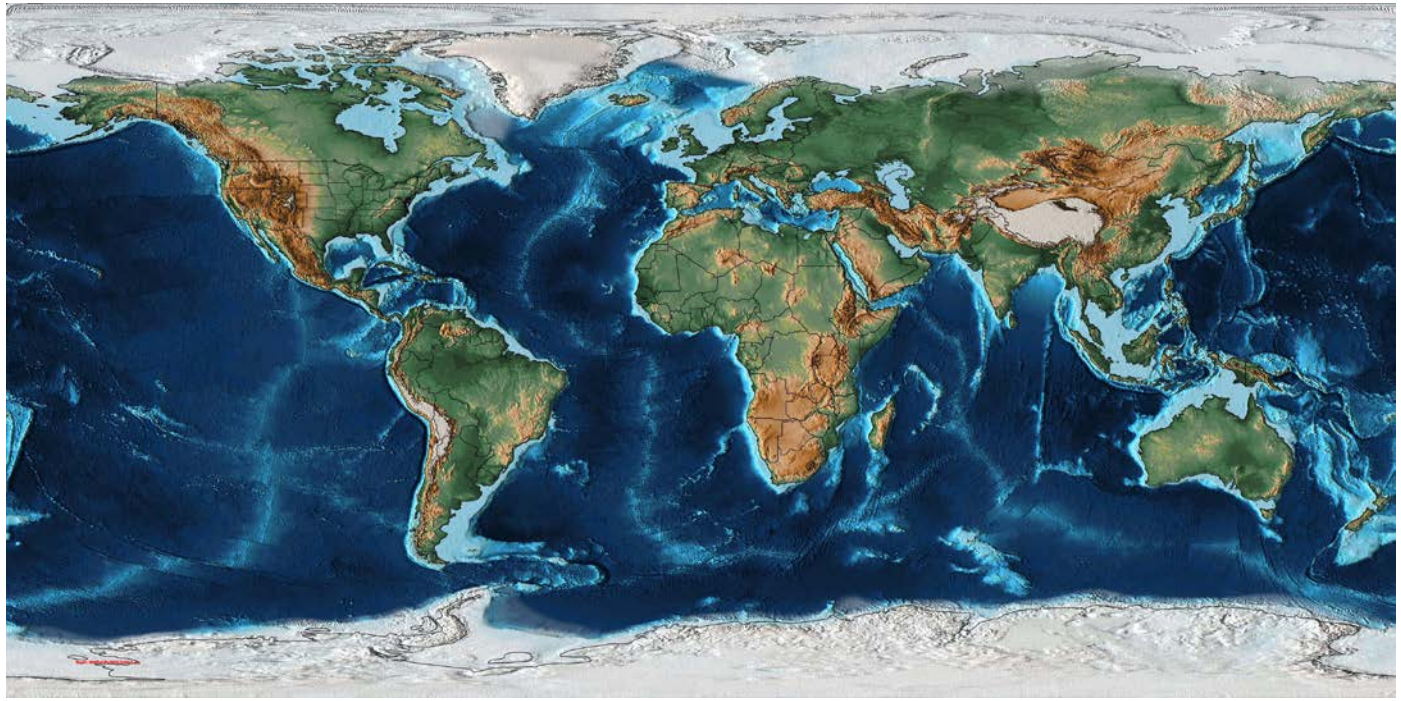
Source: PALEOMAP Project by C. Scotese

Miocene to Plio-Pleisto-Holocene
5.3 My

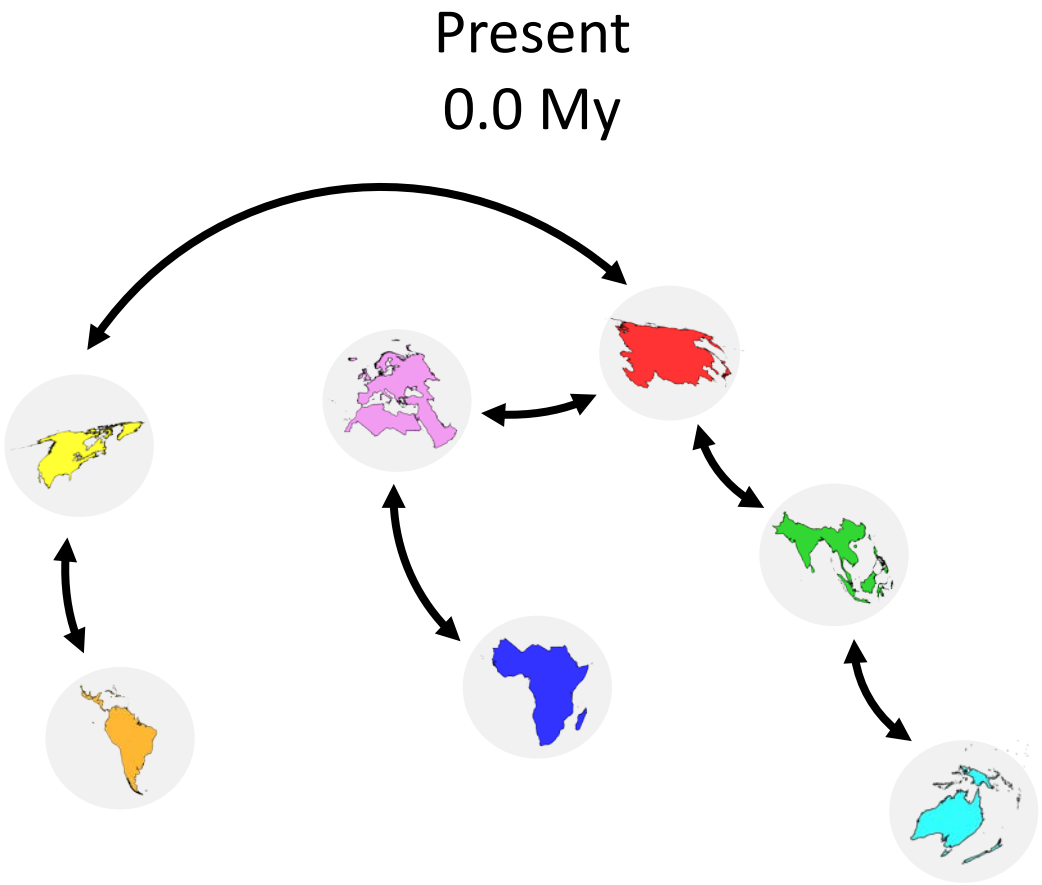


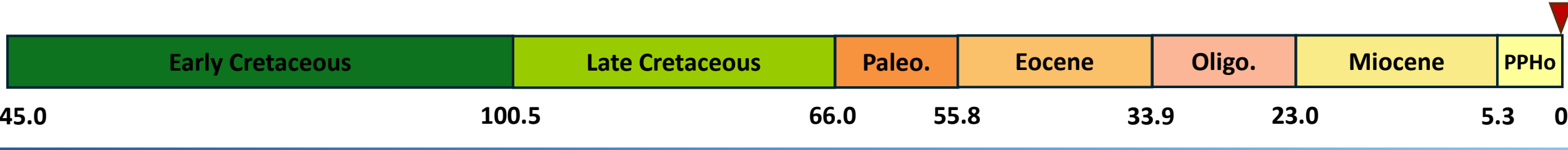


# Biogeographic methods

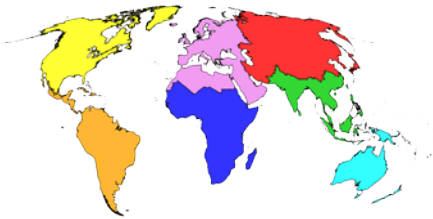


Source: PALEOMAP Project by C. Scotese



# Biogeographic methods



**BioGeoBEARS** framework:

- DIVALIKE, DEC, DIVALIKE + J, DEC + J
- Model selection based on AICc
- Best model = **Time-stratified DEC + J**

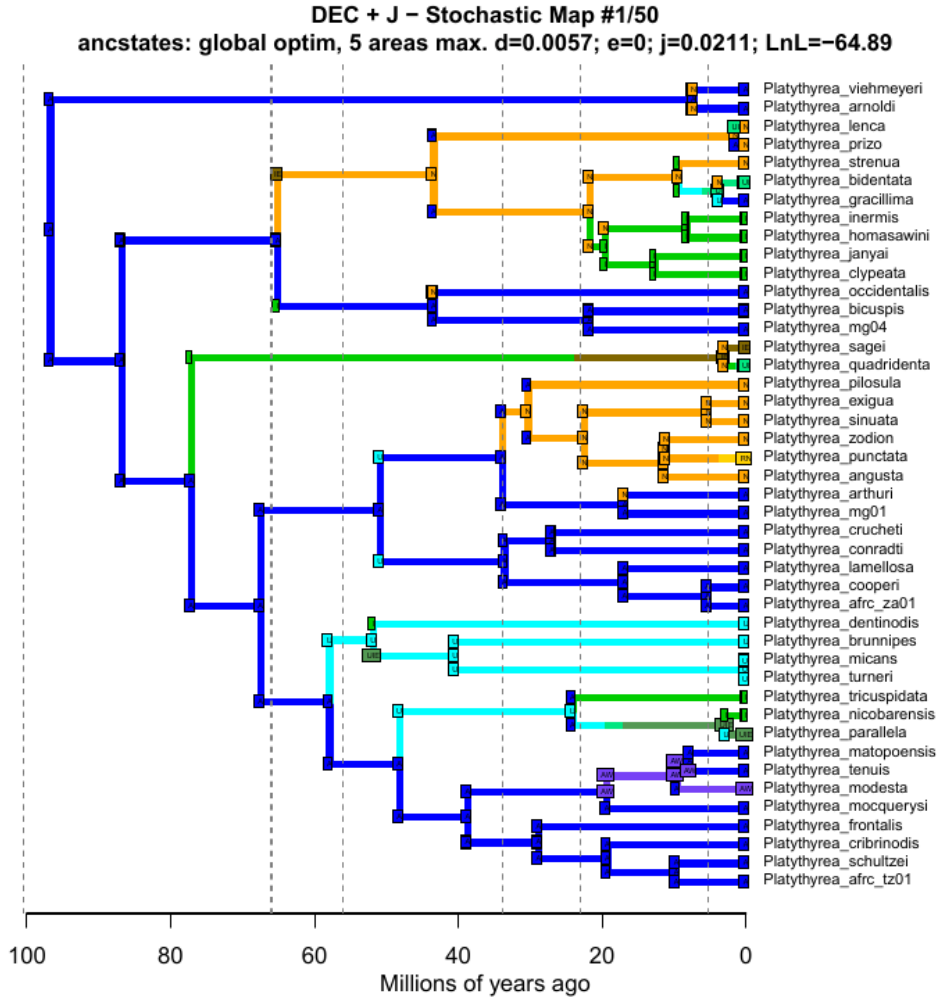
Model of **continental connectivity**

**Ancestral Range Estimates (ARE)**:

- Marginal likelihoods of node ranges

**Biogeographic Stochastic Mapping (BSM):**

- Simulation of **biogeographic histories** compatible with ancestral ranges estimates
- Summary statistics across 1000 BSM



# Origins of clades

Overall picture:

- Some important **regional radiations**
- Many **independent colonizations**

Origins of S&S clades:

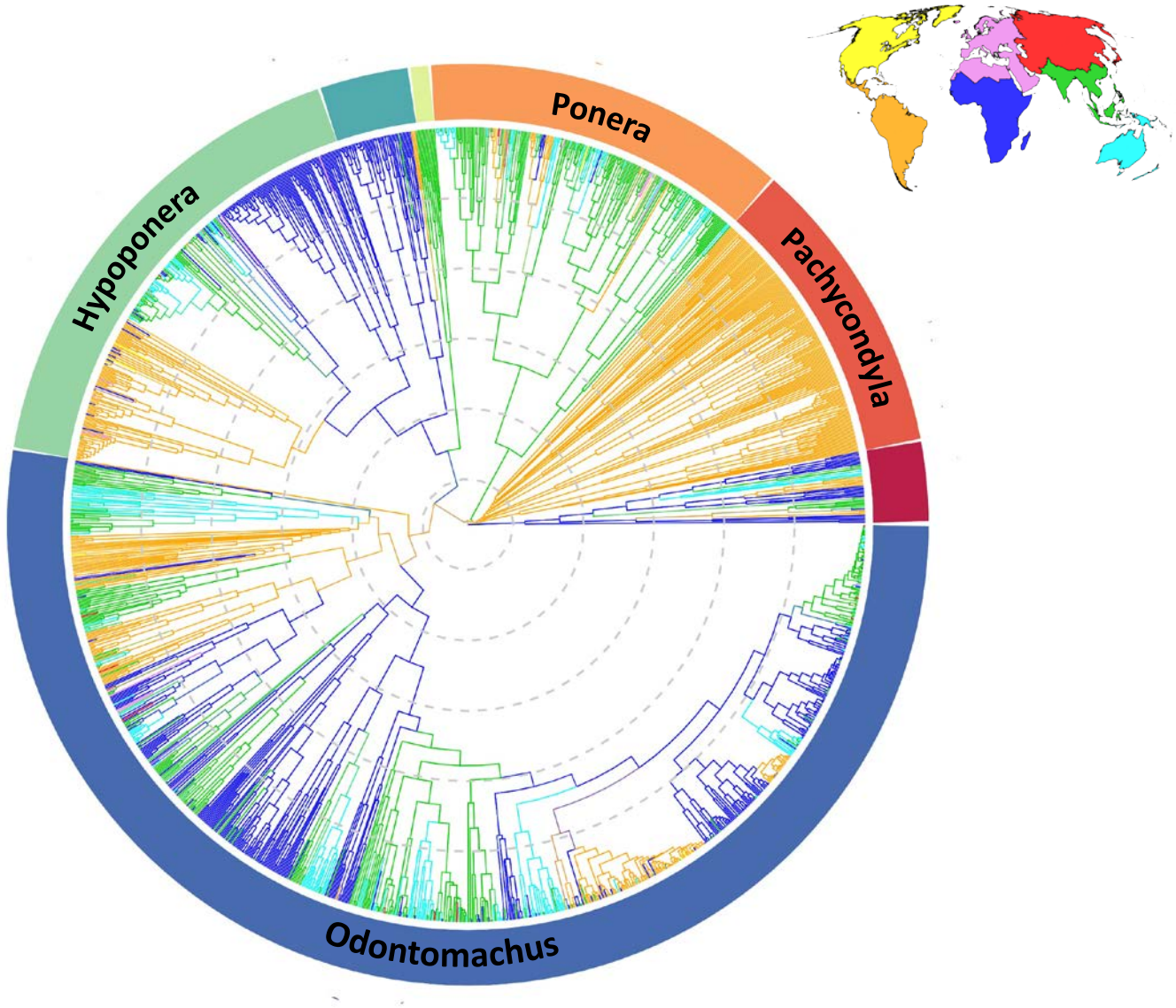
- **Afrotropics**: Platytherea, Plectroctena, Hypoponera
- **Neotropics**: Pachycondyla, Odontomachus
- **IndoMalaya**: Ponera, Harpegnathos

Origin of Ponerinae:

**Afrotropics** + **Neotropics** [ + **Indo(Malaya)** ]

=

**Gondwana**




**Ancestral Range Estimates**
**All Ponerinae species**

Doré et al., *in prep.*

# Origins of clades

Overall picture:

- Some important **regional radiations**
- Many **independent colonizations**

Origins of S&S clades:

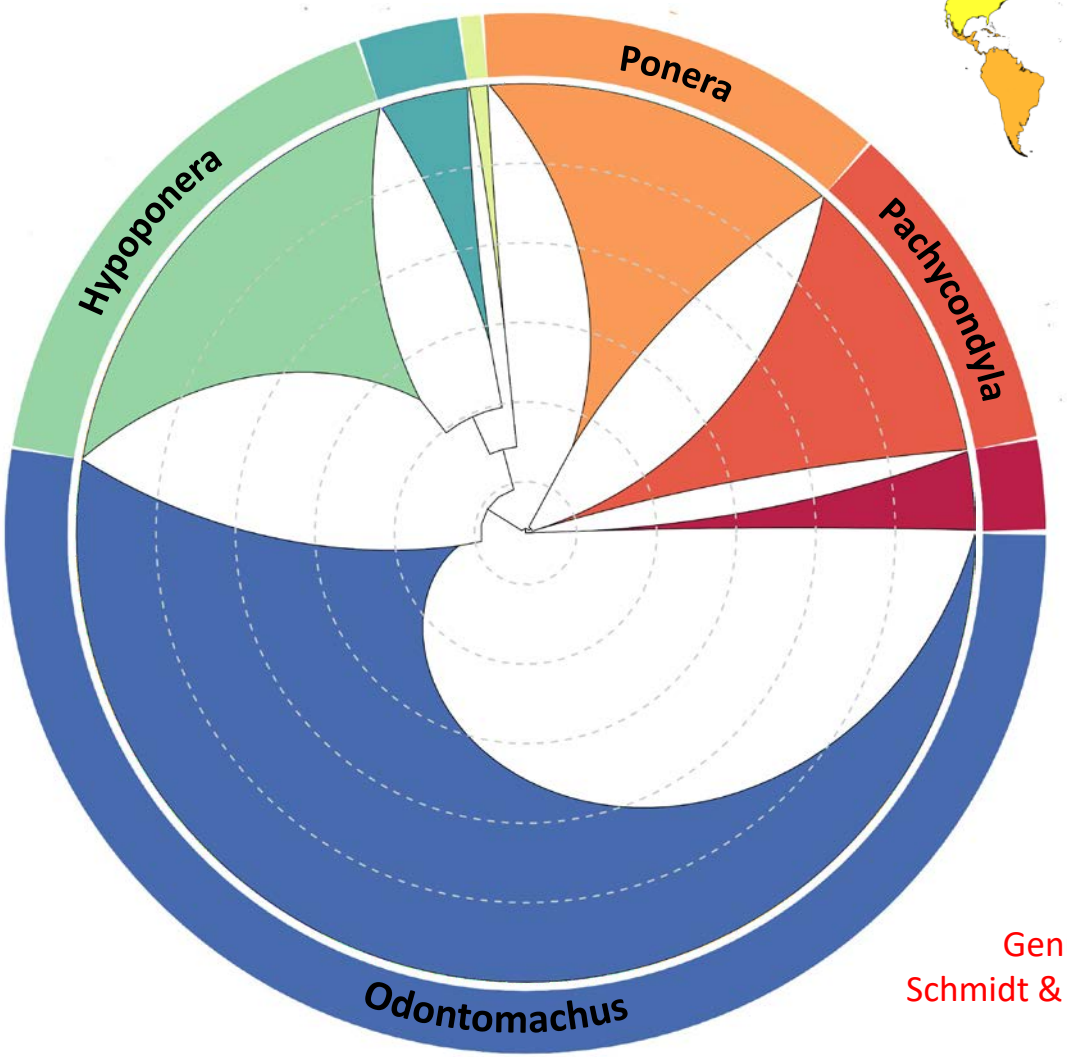
- **Afrotropics**: Platytherea, Plectroctena, Hypoponera
- **Neotropics**: Pachycondyla, Odontomachus
- **IndoMalaya**: Ponera, Harpegnathos

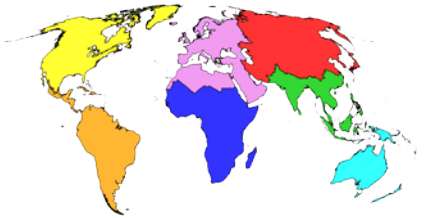
Origin of Ponerinae:

**Afrotropics** + **Neotropics** [ + **Indo(Malaya)** ]

=

**Gondwana**



Genus-groups from Schmidt & Shattuck, 2014

**Ancestral Range Estimates**
**S&S Genus-groups**

Doré et al., *in prep.*

# Origins of clades

Overall picture:

- Some important **regional radiations**
- Many **independent colonizations**

Origins of S&S clades:

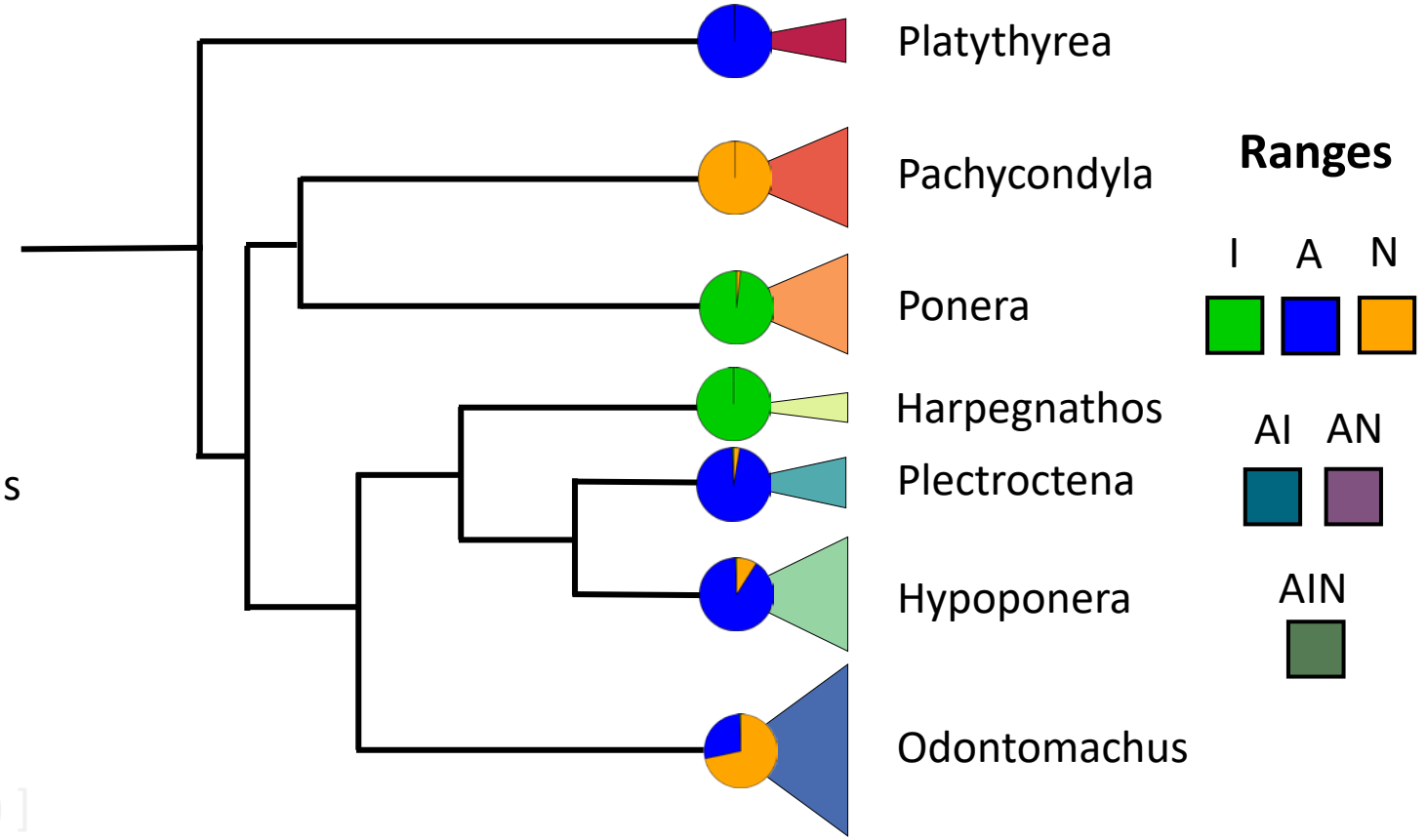
- **Afrotropics**: Platytherea, Plectroctena, Hypoponera
- **Neotropics**: Pachycondyla, Odontomachus
- **IndoMalaya**: Ponera, Harpegnathos

Origin of Ponerinae:

Afrotropics + Neotropics [ + Indo(Malaya) ]

=

Gondwana




**Ancestral Range Estimates**
**S&S Genus-groups**

Doré et al., *in prep.*

# Origins of clades

Overall picture:

- Some important **regional radiations**
- Many **independent colonizations**

Origins of S&S clades:

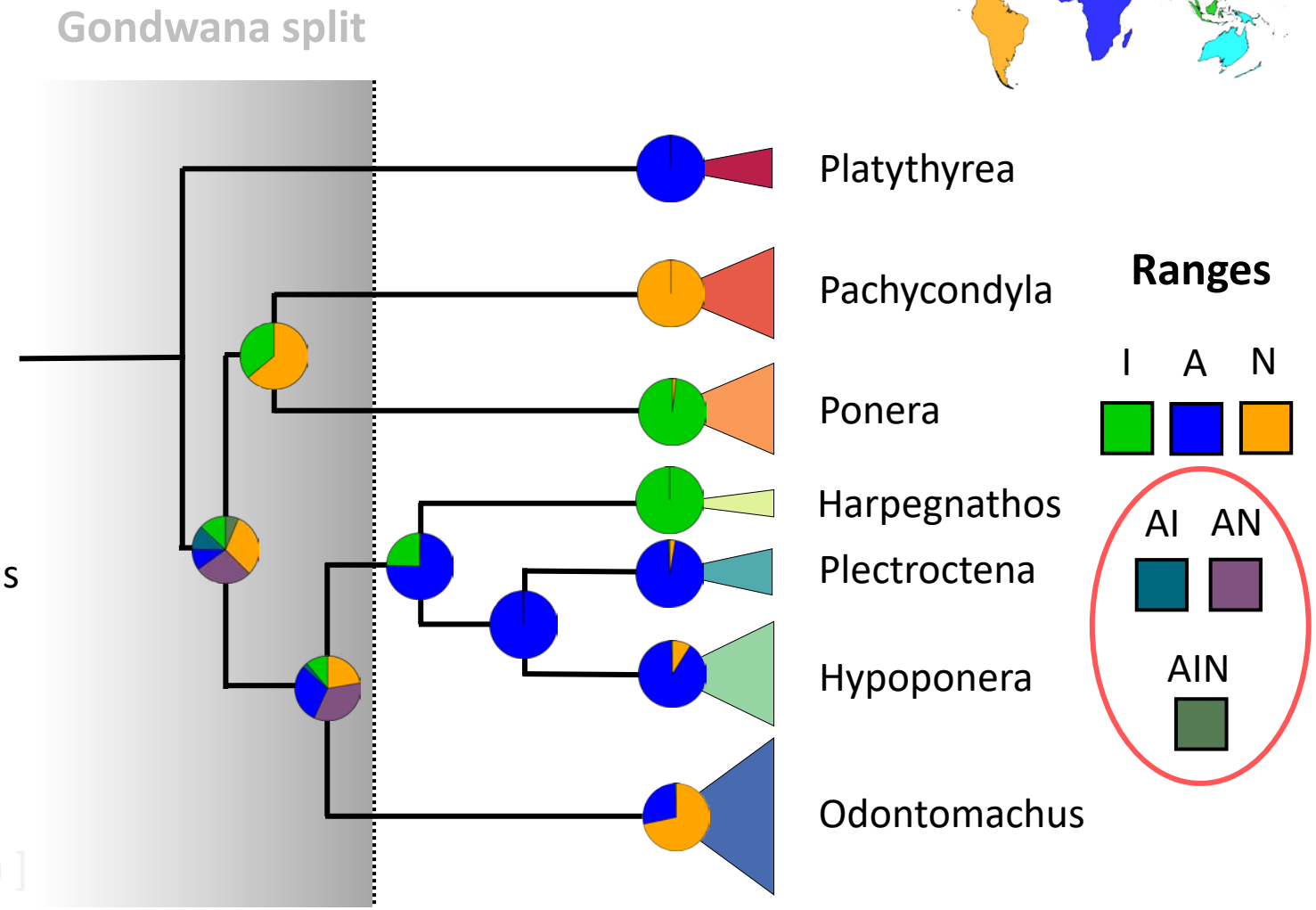
- **Afrotropics**: Platytherea, Plectroctena, Hypoponera
- **Neotropics**: Pachycondyla, Odontomachus
- **IndoMalaya**: Ponera, Harpegnathos

Origin of Ponerinae:

Afrotropics + Neotropics [ + Indo(Malaya) ]

=

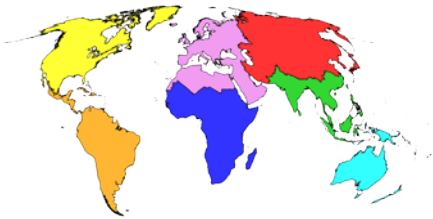
Gondwana




**Ancestral Range Estimates**
**S&S Genus-groups**

Doré et al., *in prep.*

# Origins of clades



Overall picture:

- Some important **regional radiations**
- Many **independent colonizations**

Origins of S&S clades:

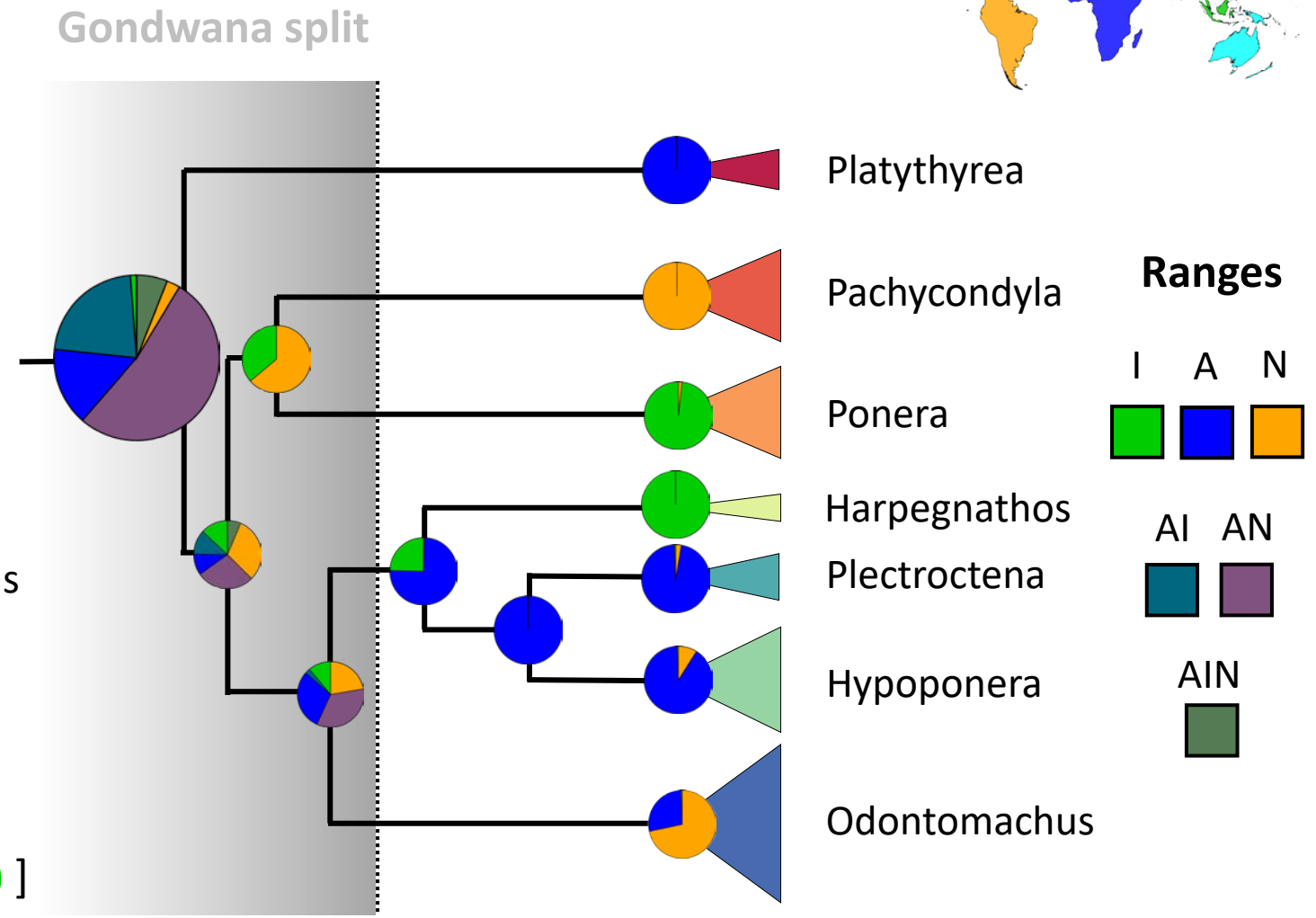
- **Afrotropics**: Platytherea, Plectroctena, Hypoponera
- **Neotropics**: Pachycondyla, Odontomachus
- **IndoMalaya**: Ponera, Harpegnathos

Origin of Ponerinae:

**Afrotropics** + **Neotropics** [ + **Indo(Malaya)** ]

=

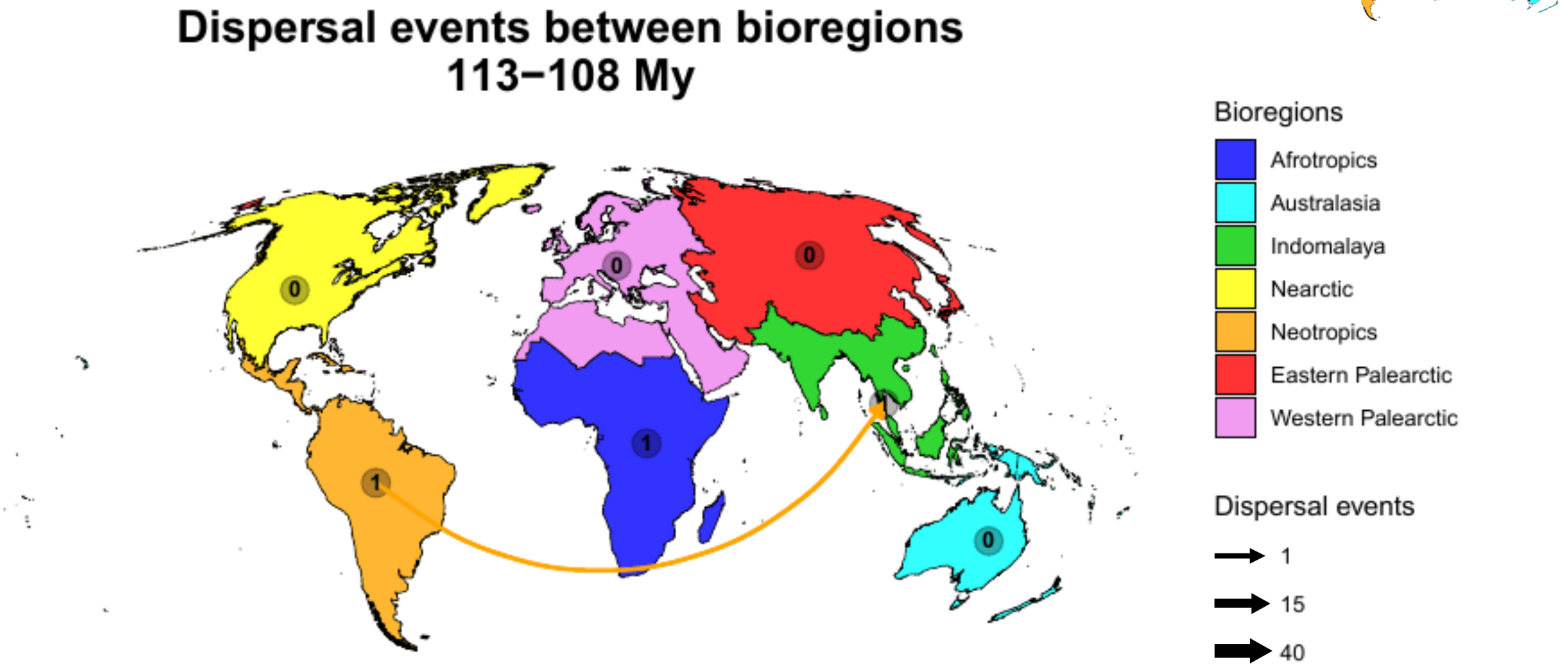
**Gondwana**




**Ancestral Range Estimates**
**S&S Genus-groups**

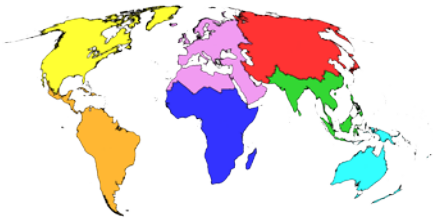
Doré et al., *in prep.*

# Dispersal events in time

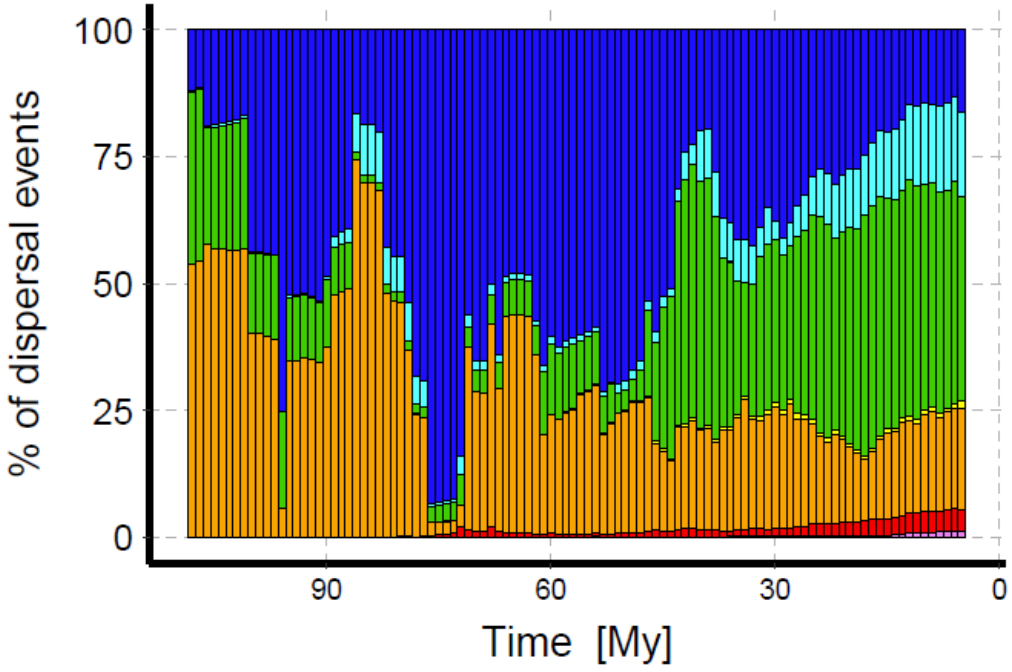



Doré et al., *in prep.*

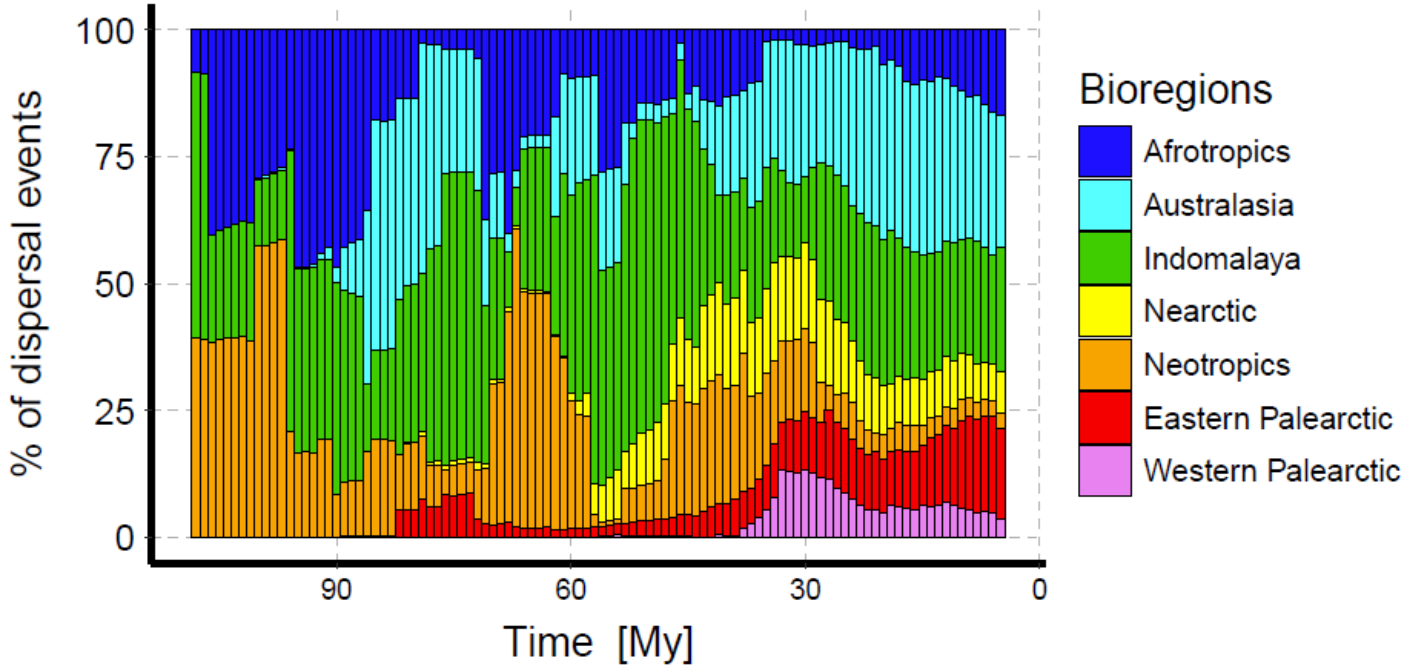
# Dispersal events in time



Dispersal events
per Source bioregions

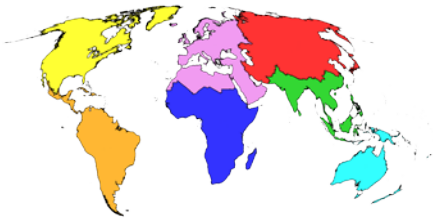



Dispersal events
per Destination bioregions

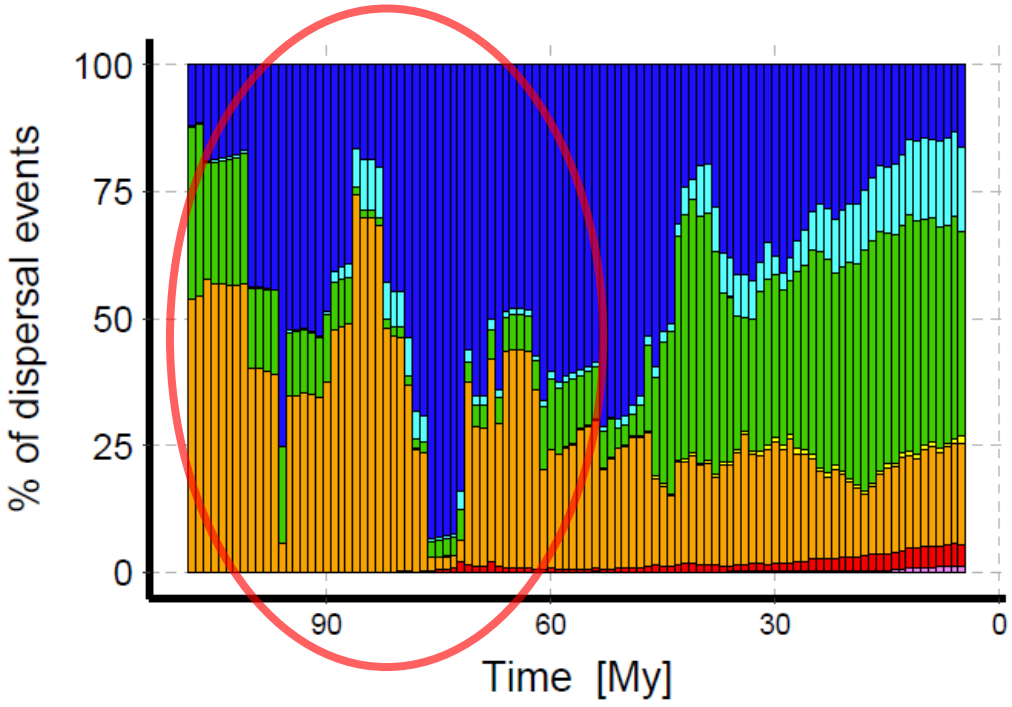



Doré et al., *in prep.*

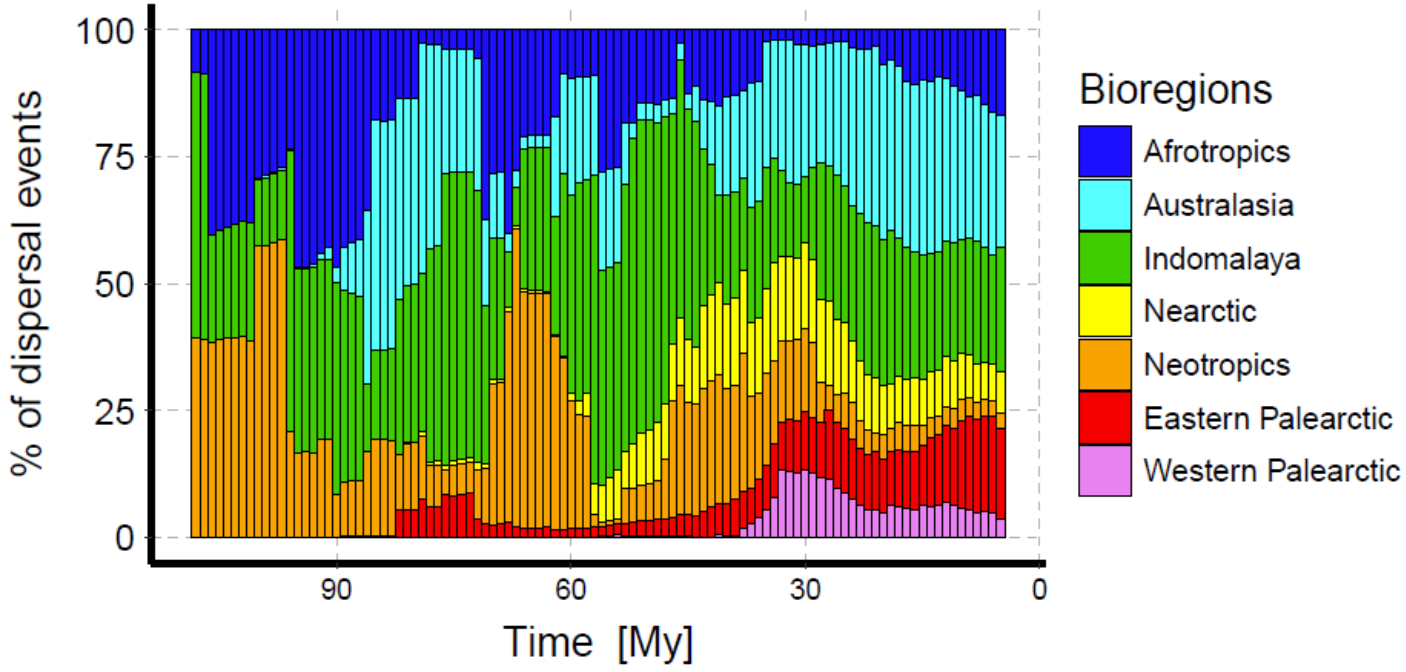
# Dispersal events in time



Dispersal events
per Source bioregions

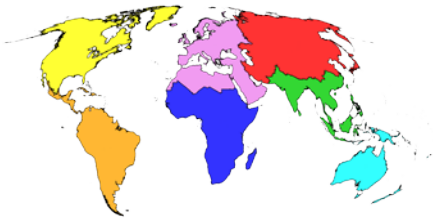



Dispersal events
per Destination bioregions

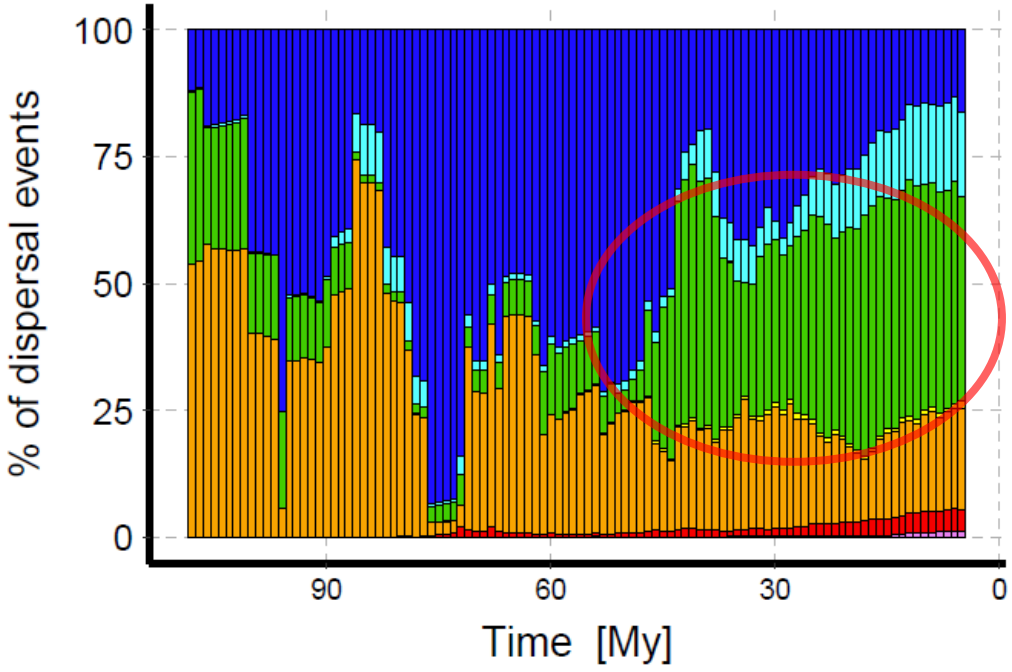



Doré et al., *in prep.*

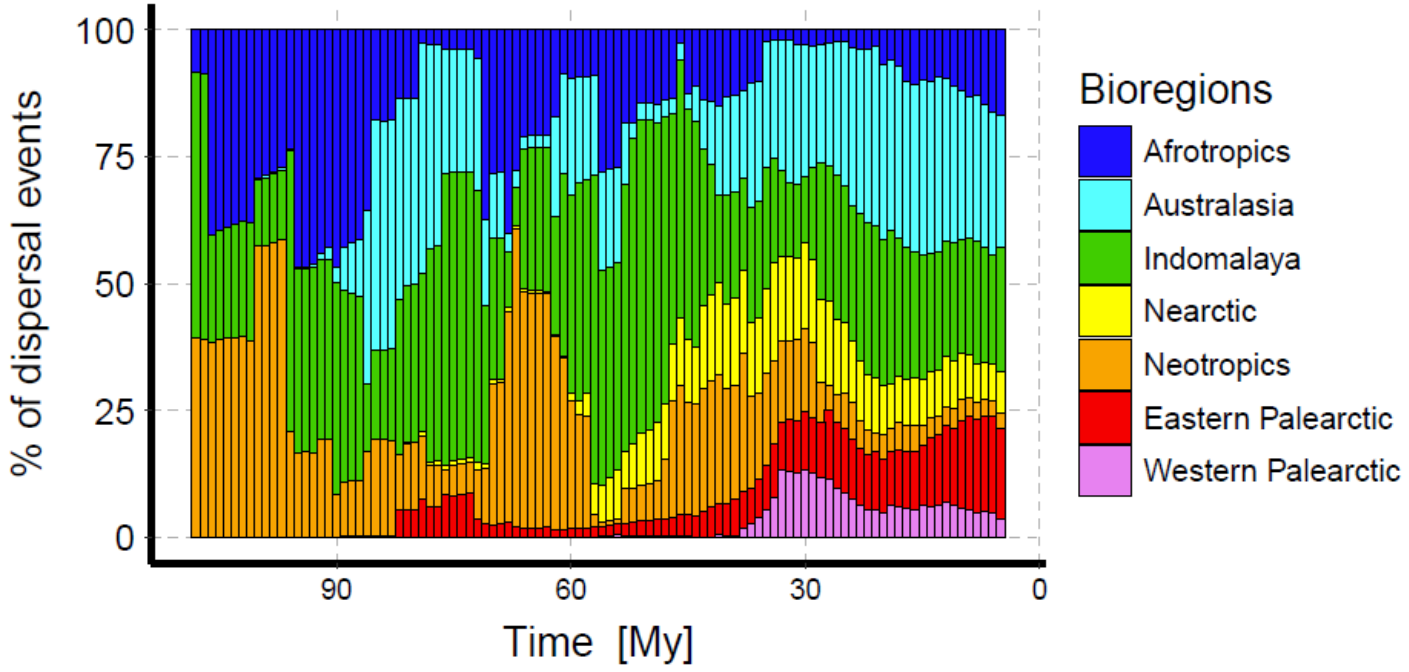
# Dispersal events in time



Dispersal events
per Source bioregions

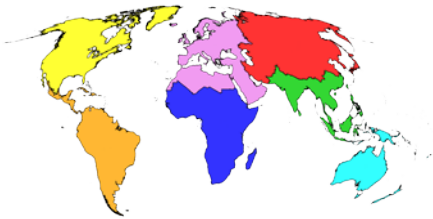



Dispersal events
per Destination bioregions

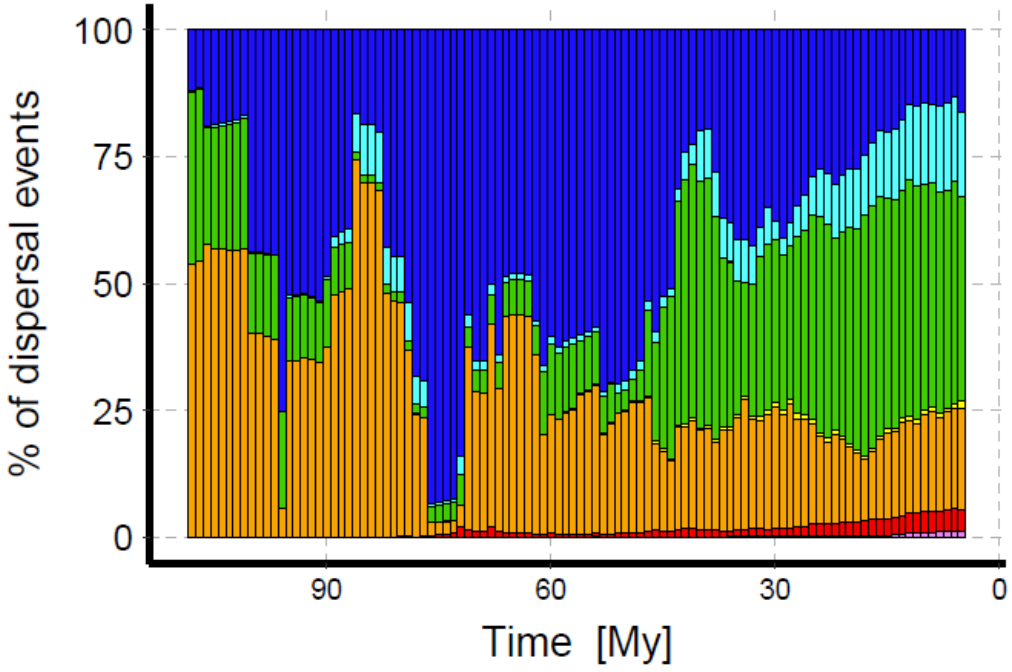



Doré et al., *in prep.*

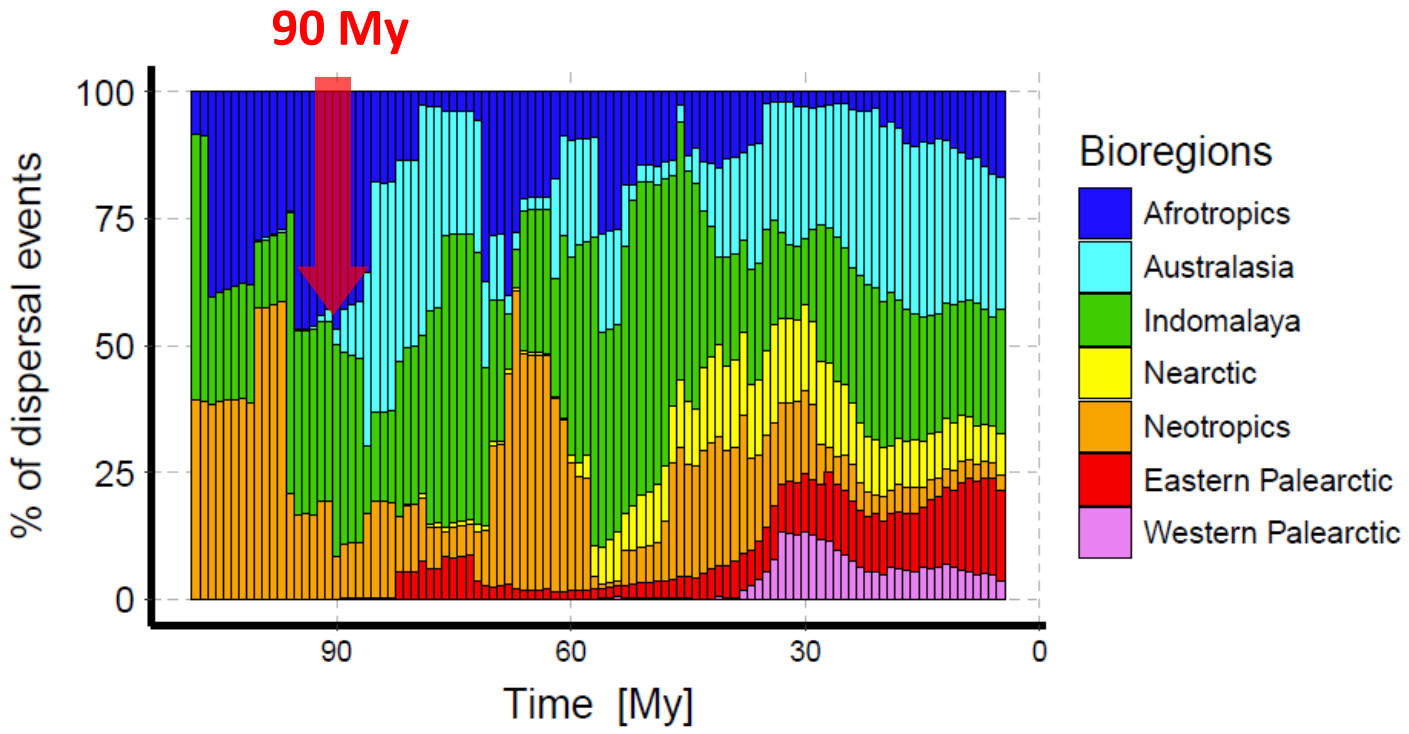
# Dispersal events in time



Dispersal events
per Source bioregions

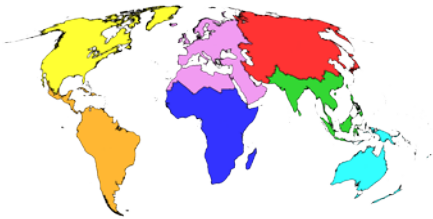



Dispersal events
per Destination bioregions

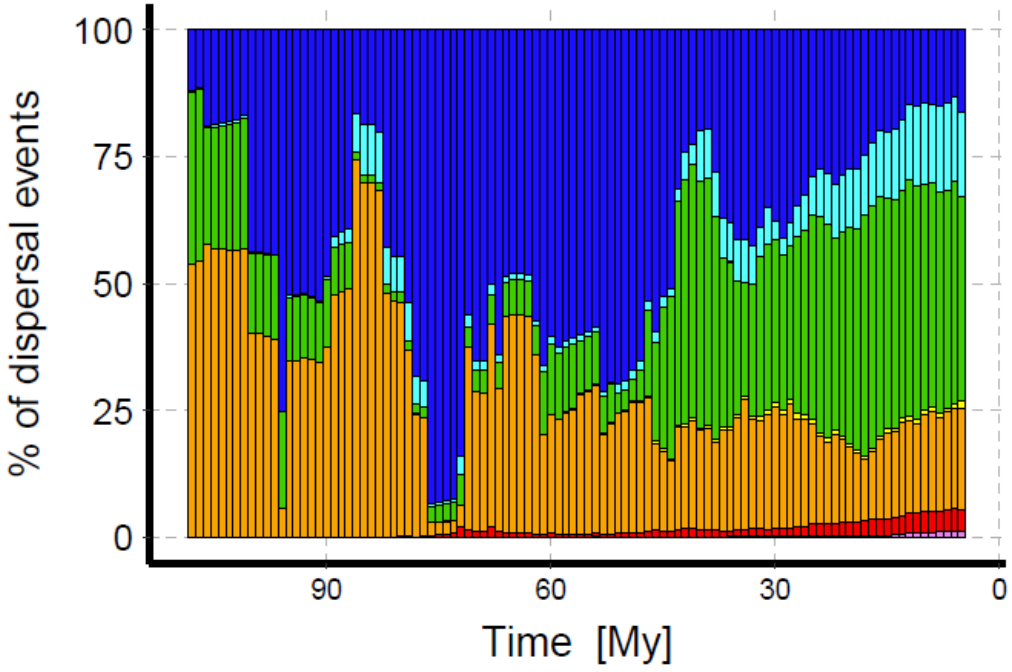



Doré et al., *in prep.*

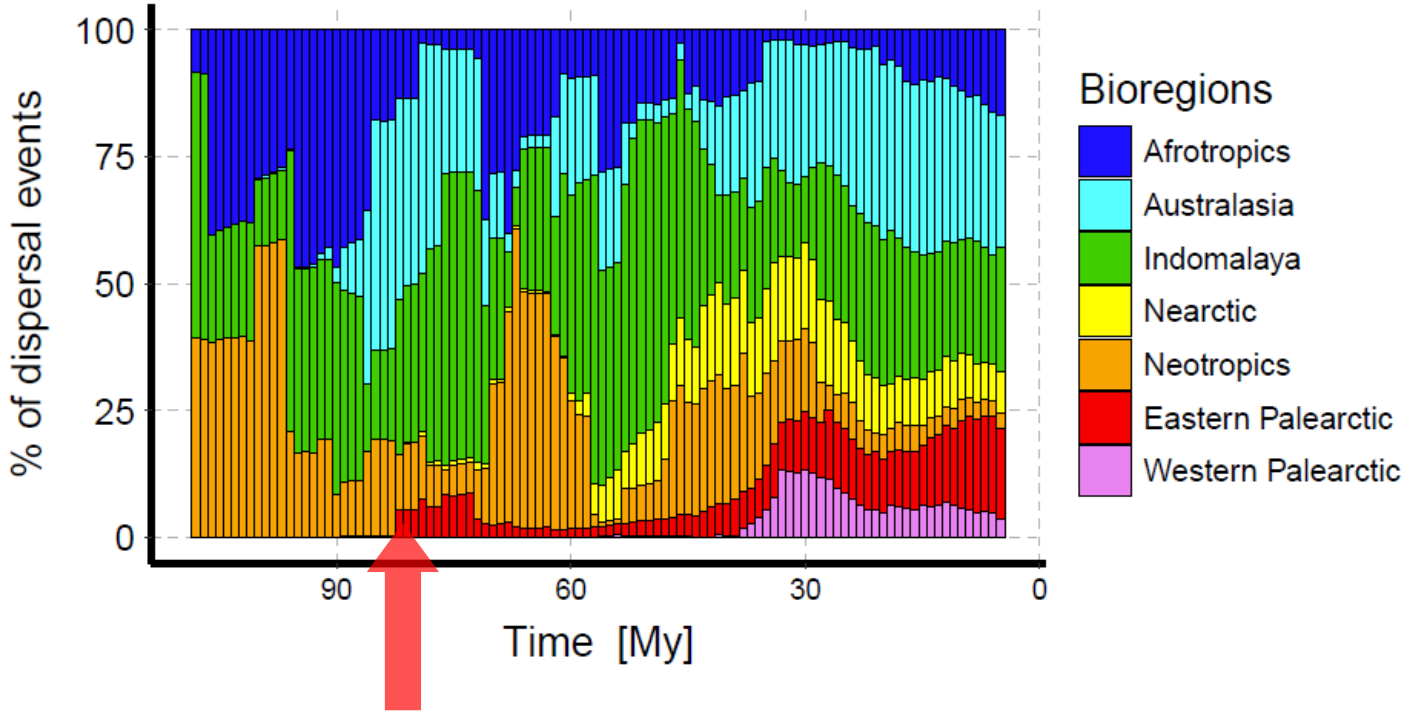
# Dispersal events in time



Dispersal events
per Source bioregions



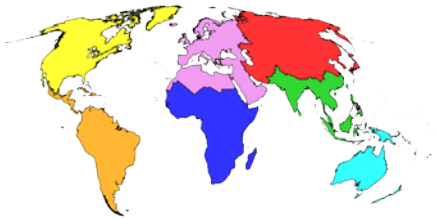

Dispersal events
per Destination bioregions



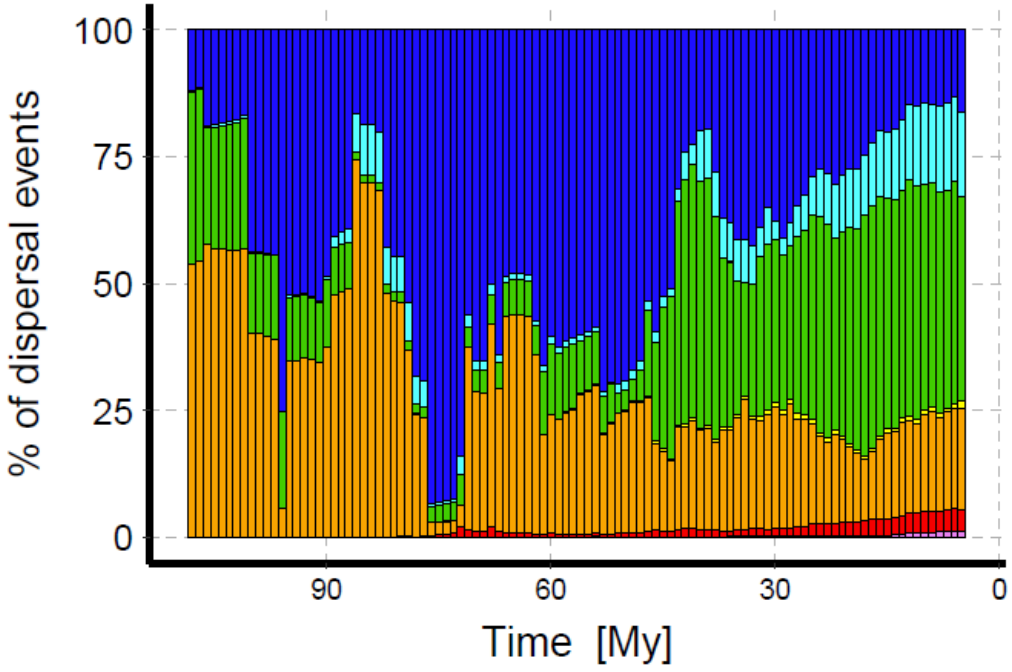

**80 My**

Doré et al., *in prep.*

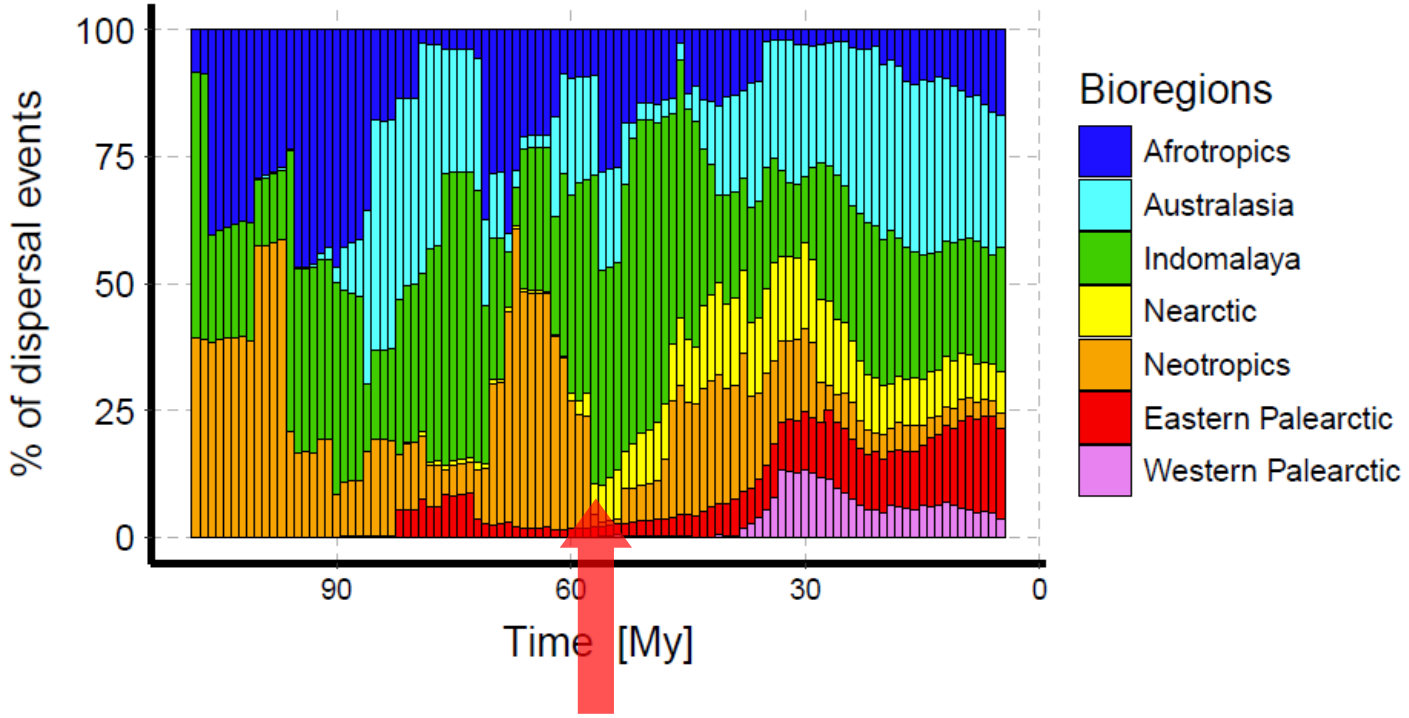
# Dispersal events in time



Dispersal events
per Source bioregions

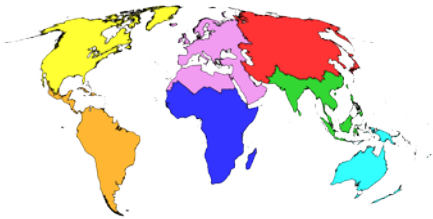



Dispersal events
per Destination bioregions

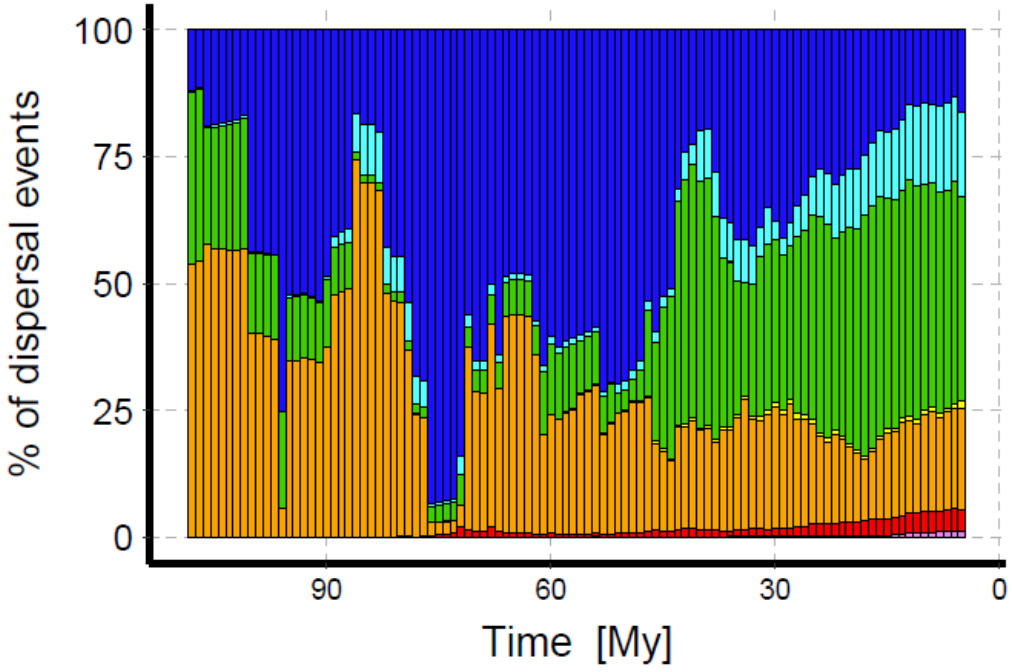



Doré et al., *in prep.*

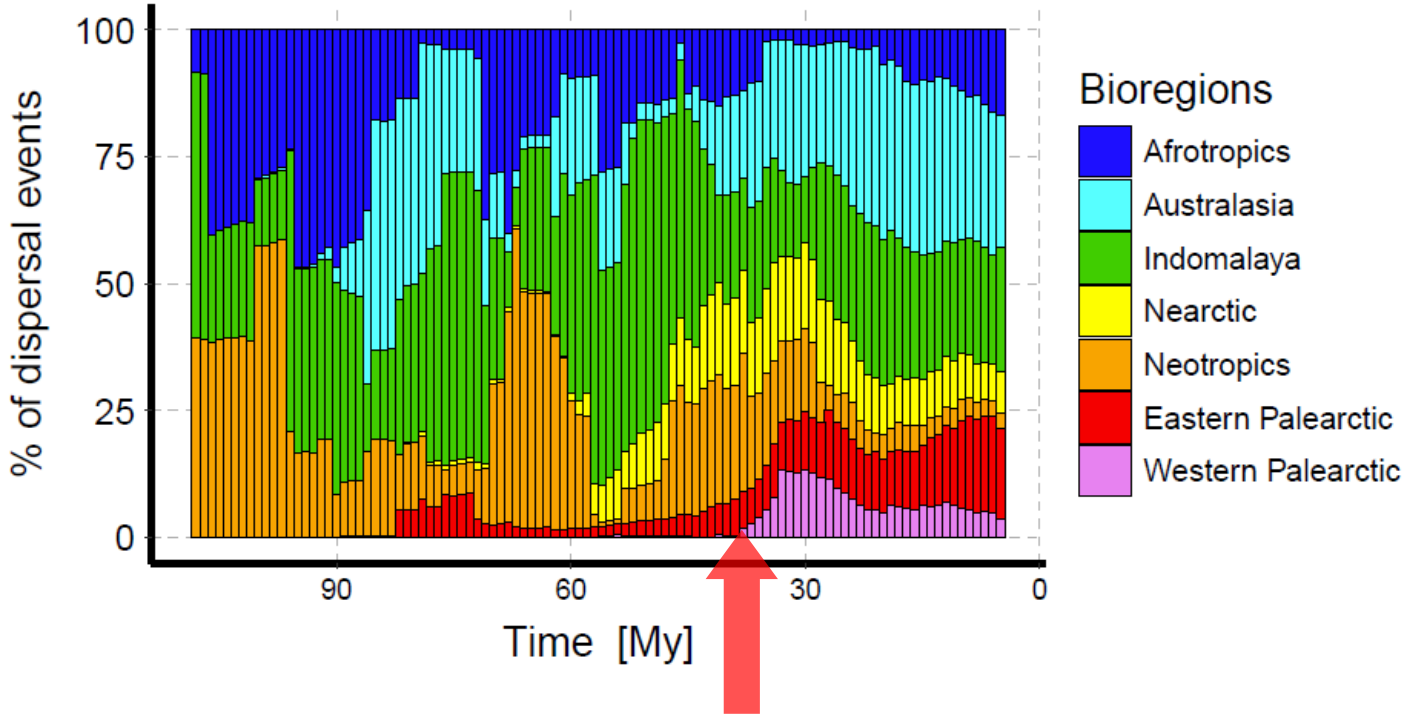
# Dispersal events in time



Dispersal events
per Source bioregions



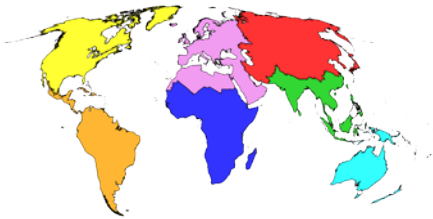

Dispersal events
per Destination bioregions



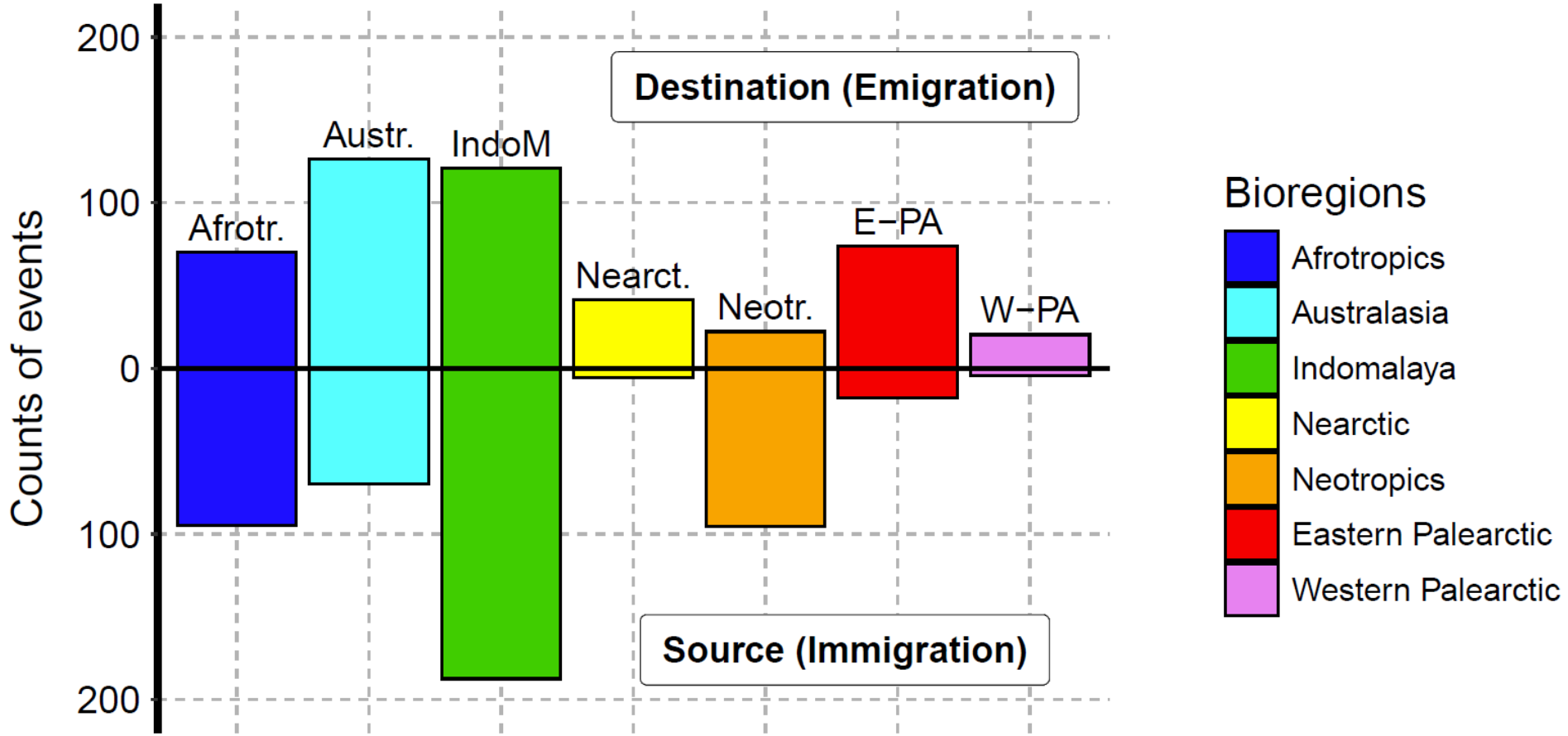

**35 My**

Doré et al., *in prep.*

# Macroevolutionary sources & sinks

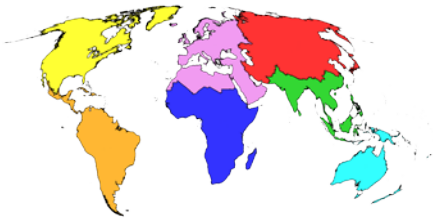


Dispersal events per Bioregions

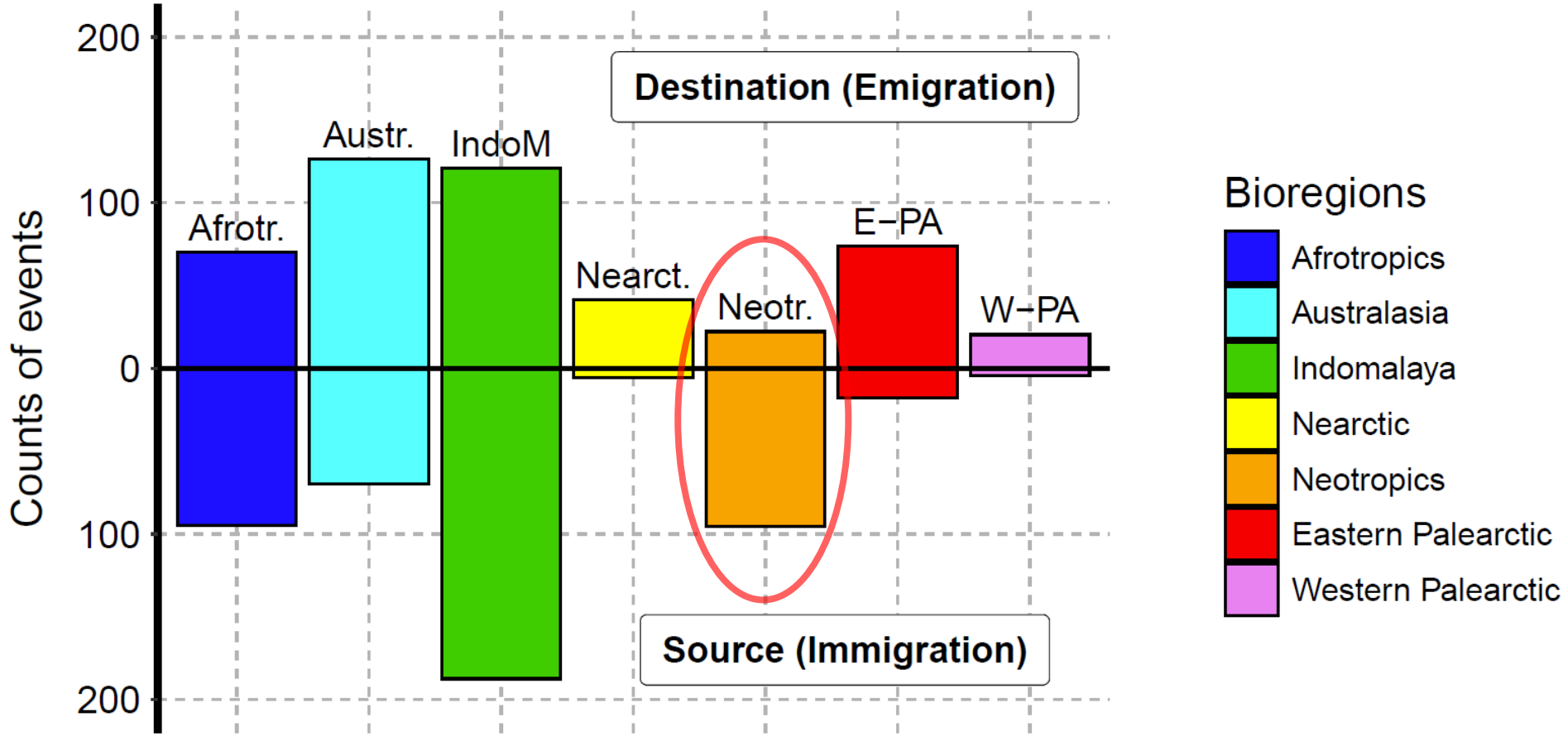



Doré et al., *in prep.*

# Macroevolutionary sources & sinks

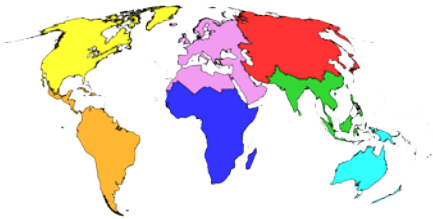


Dispersal events per Bioregions

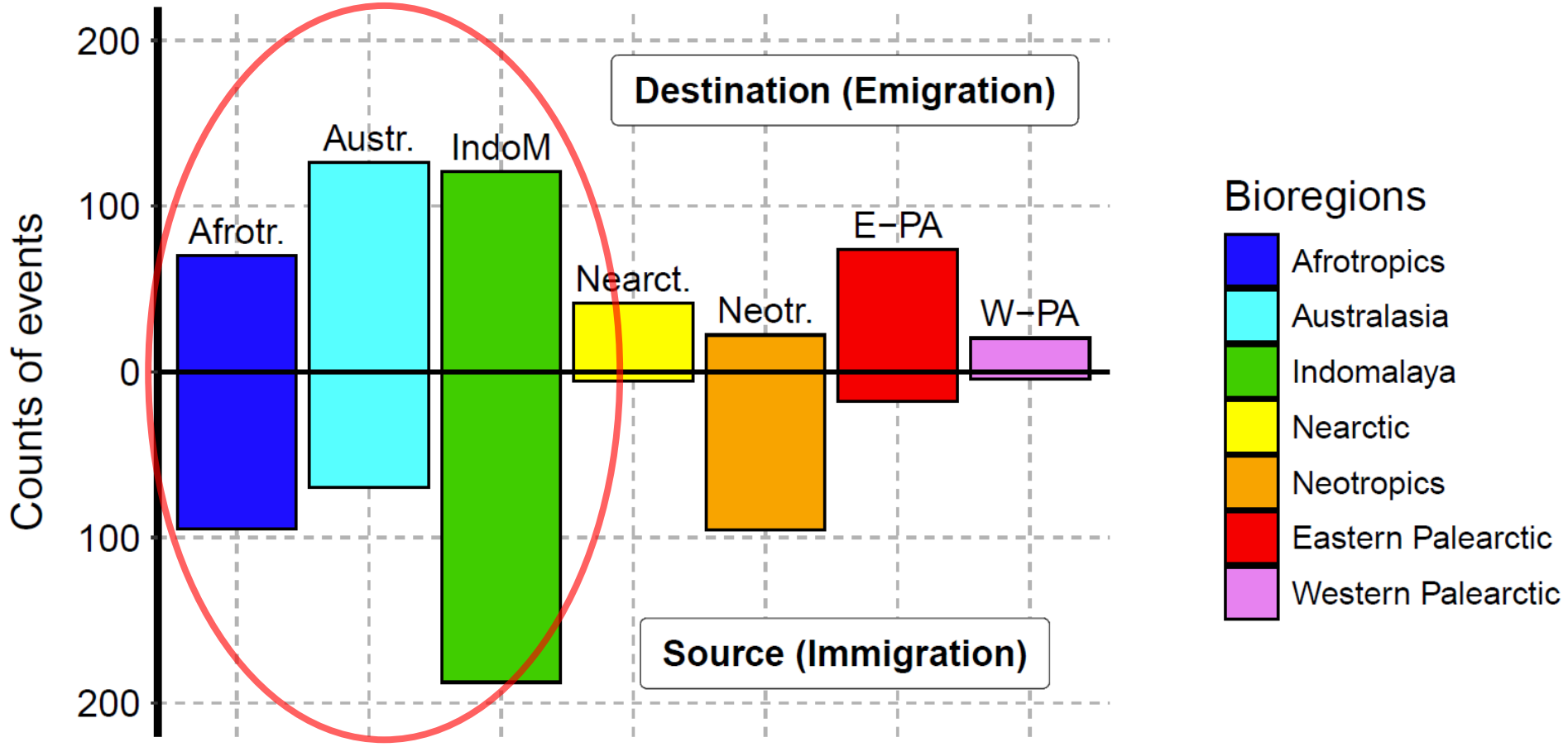



Doré et al., *in prep.*

# Macroevolutionary sources & sinks

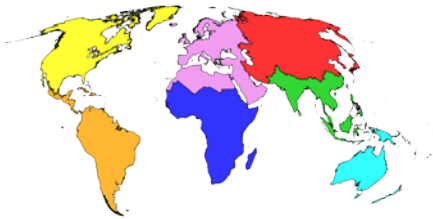


Dispersal events per Bioregions

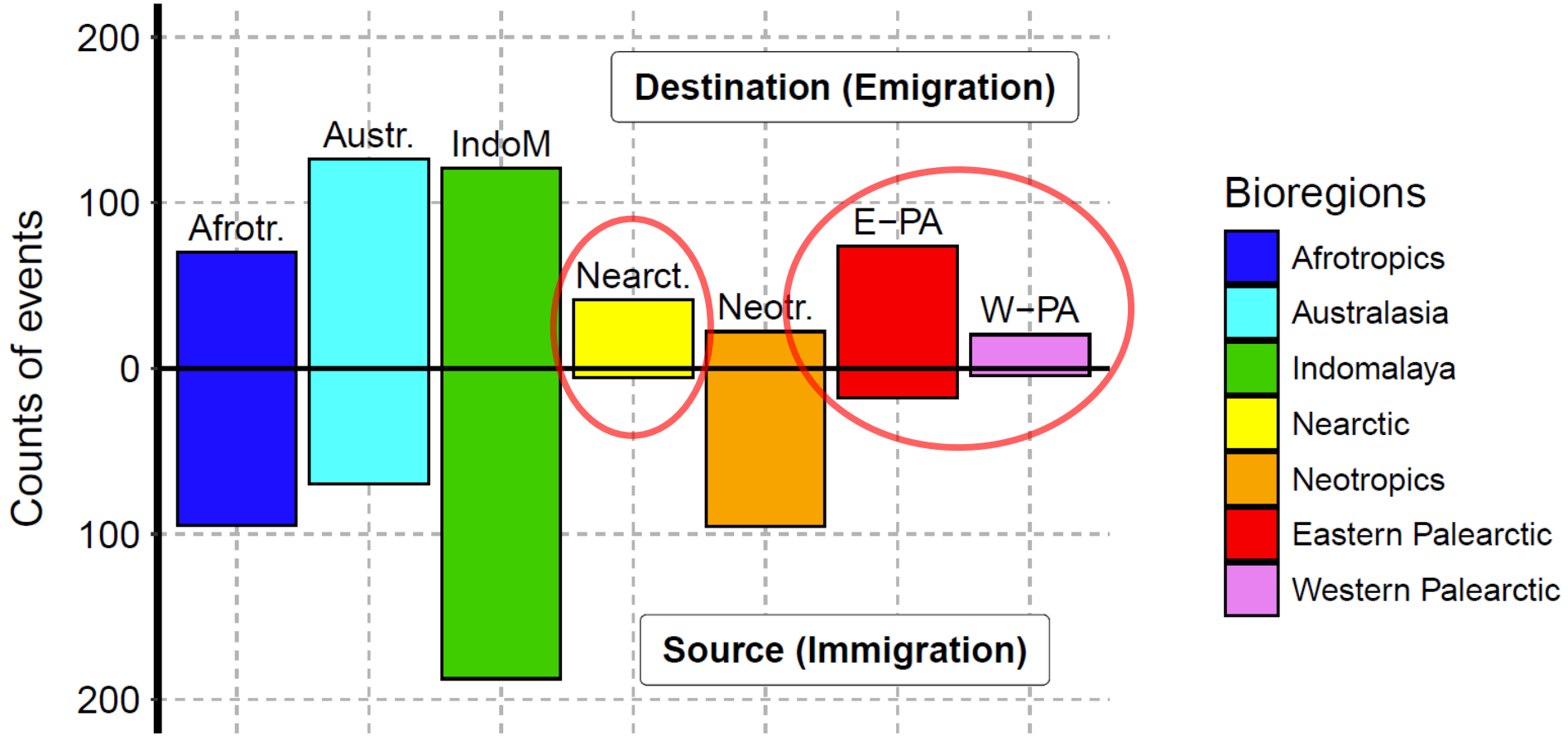



Doré et al., *in prep.*

# Macroevolutionary sources & sinks

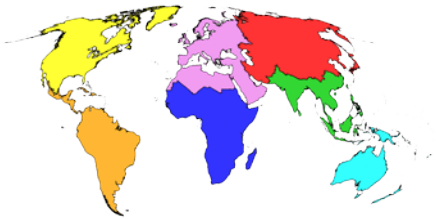


Dispersal events per Bioregions

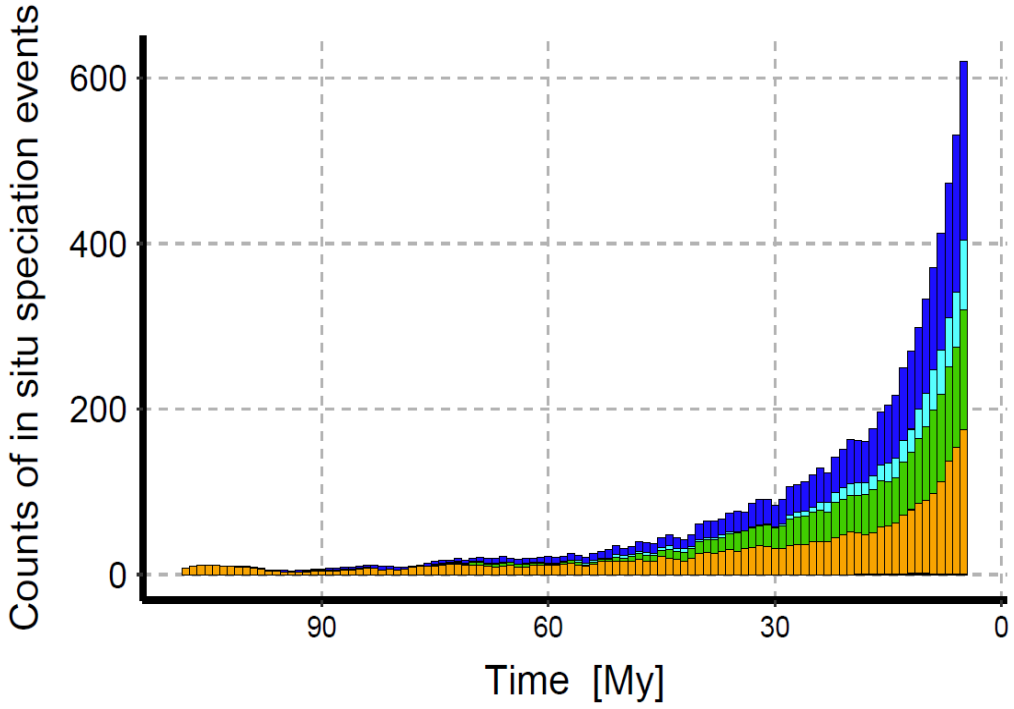



Doré et al., *in prep.*

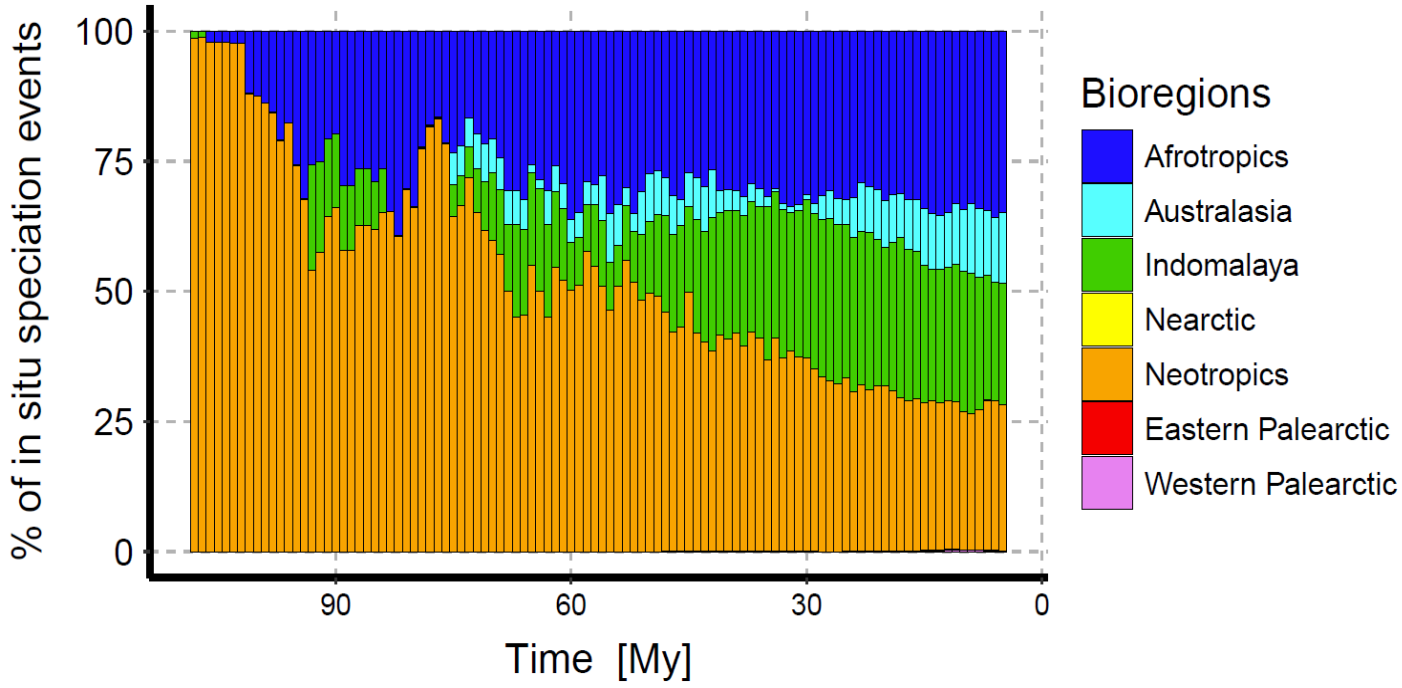
# *In situ* diversification in time



Counts
*in situ* speciation events
per Bioregions

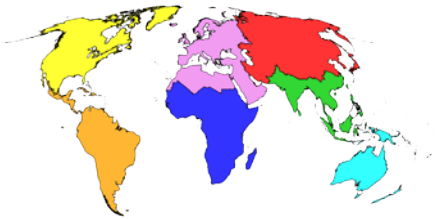



Percentages
*in situ* speciation events
per Bioregions

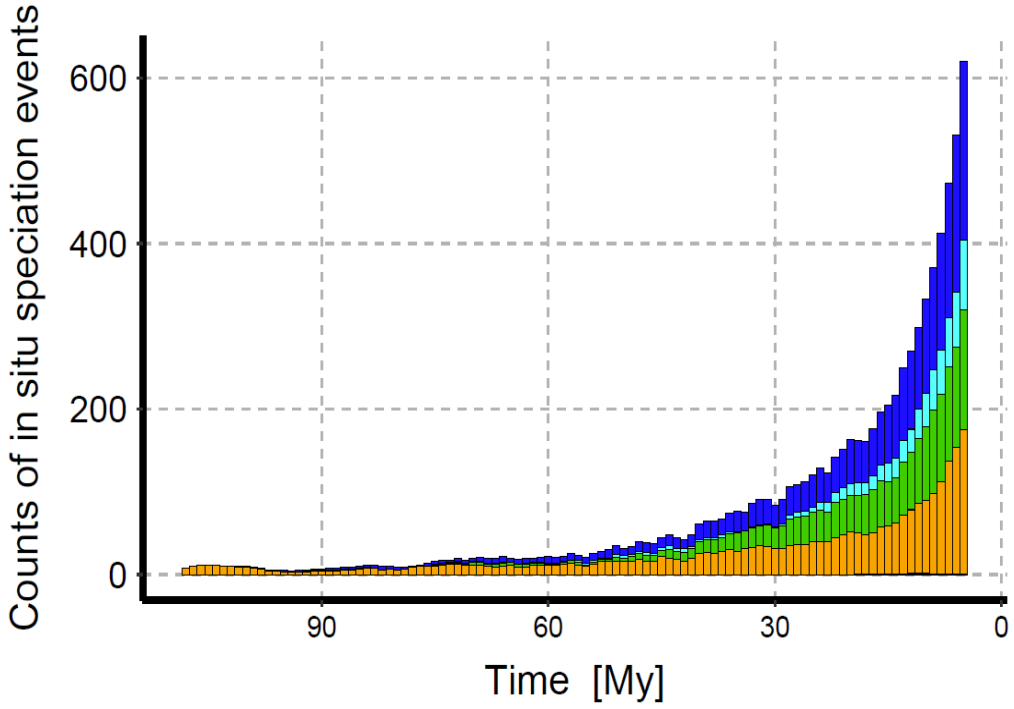



Doré et al., *in prep.*

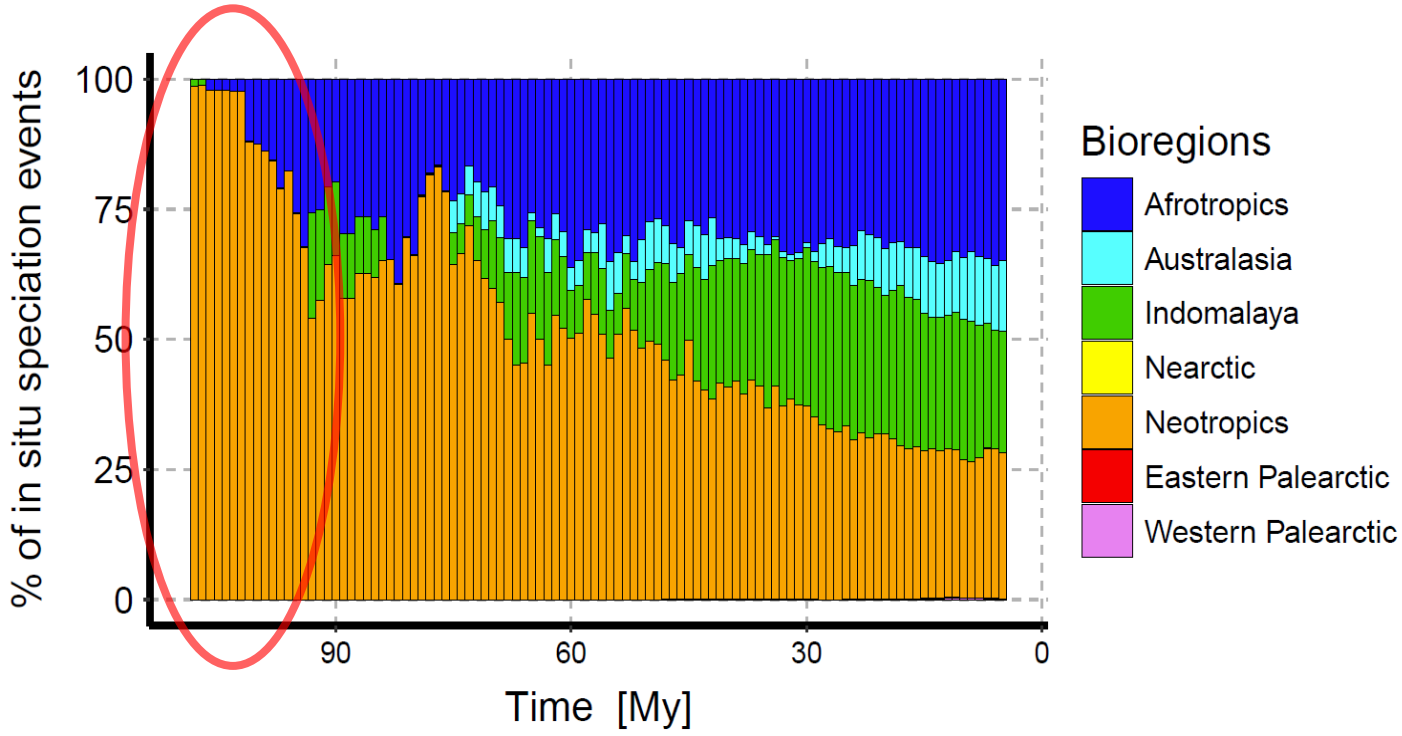
# *In situ* diversification in time



Counts
*in situ* speciation events
per Bioregions

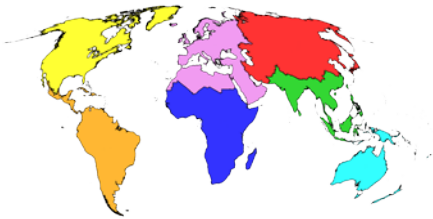



Percentages
*in situ* speciation events
per Bioregions

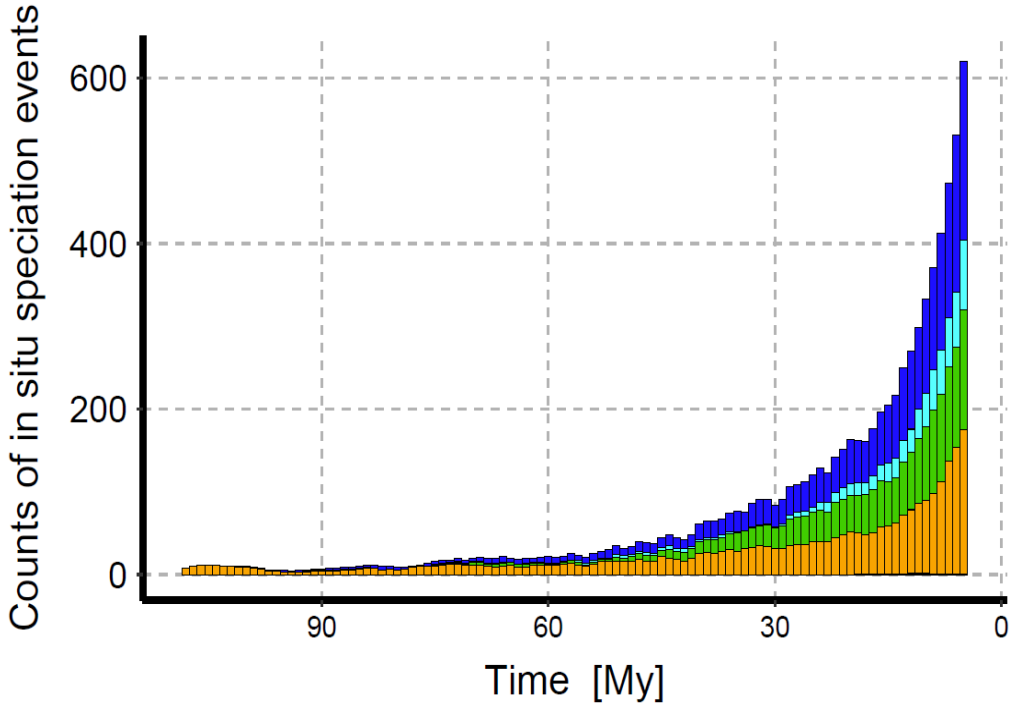



Doré et al., *in prep.*

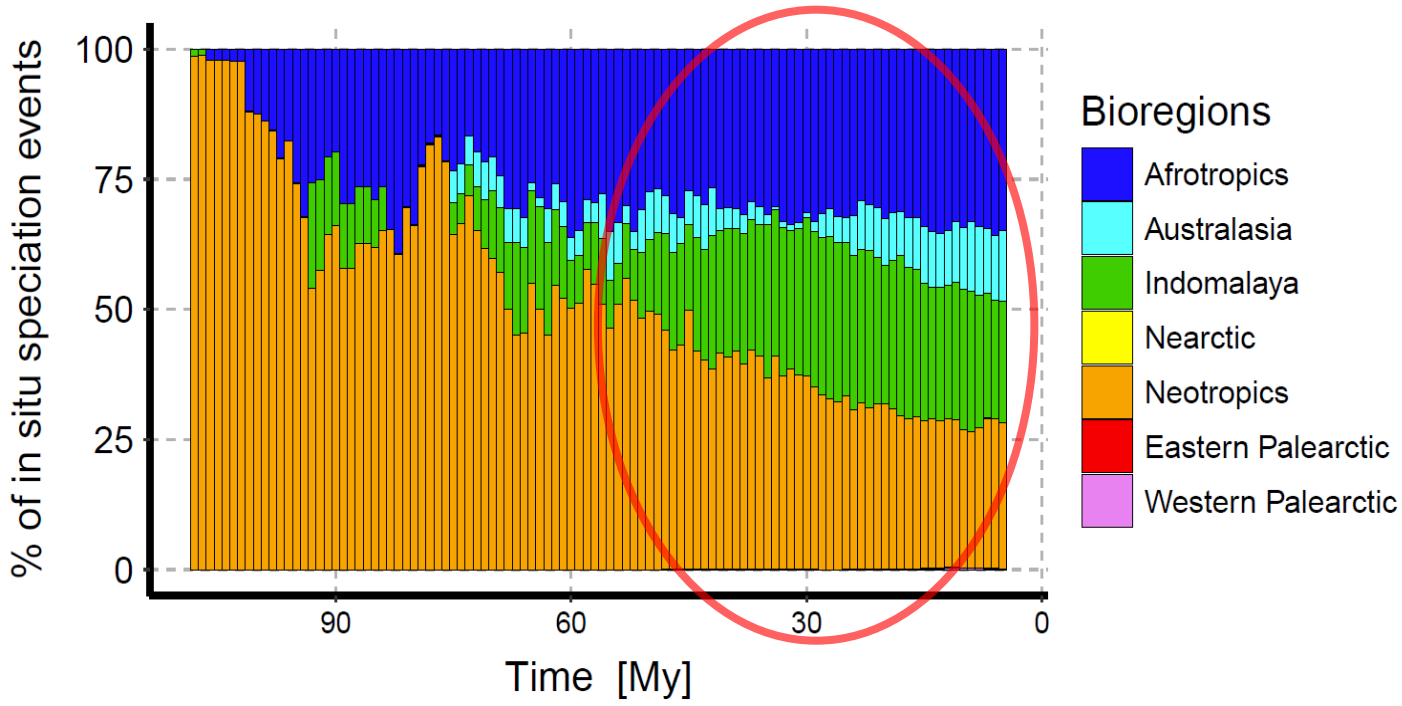
# *In situ* diversification in time



Counts
*in situ* speciation events
per Bioregions

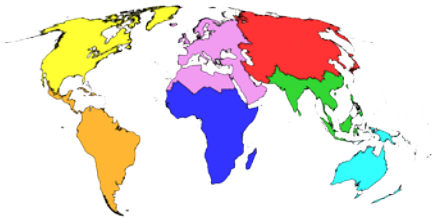


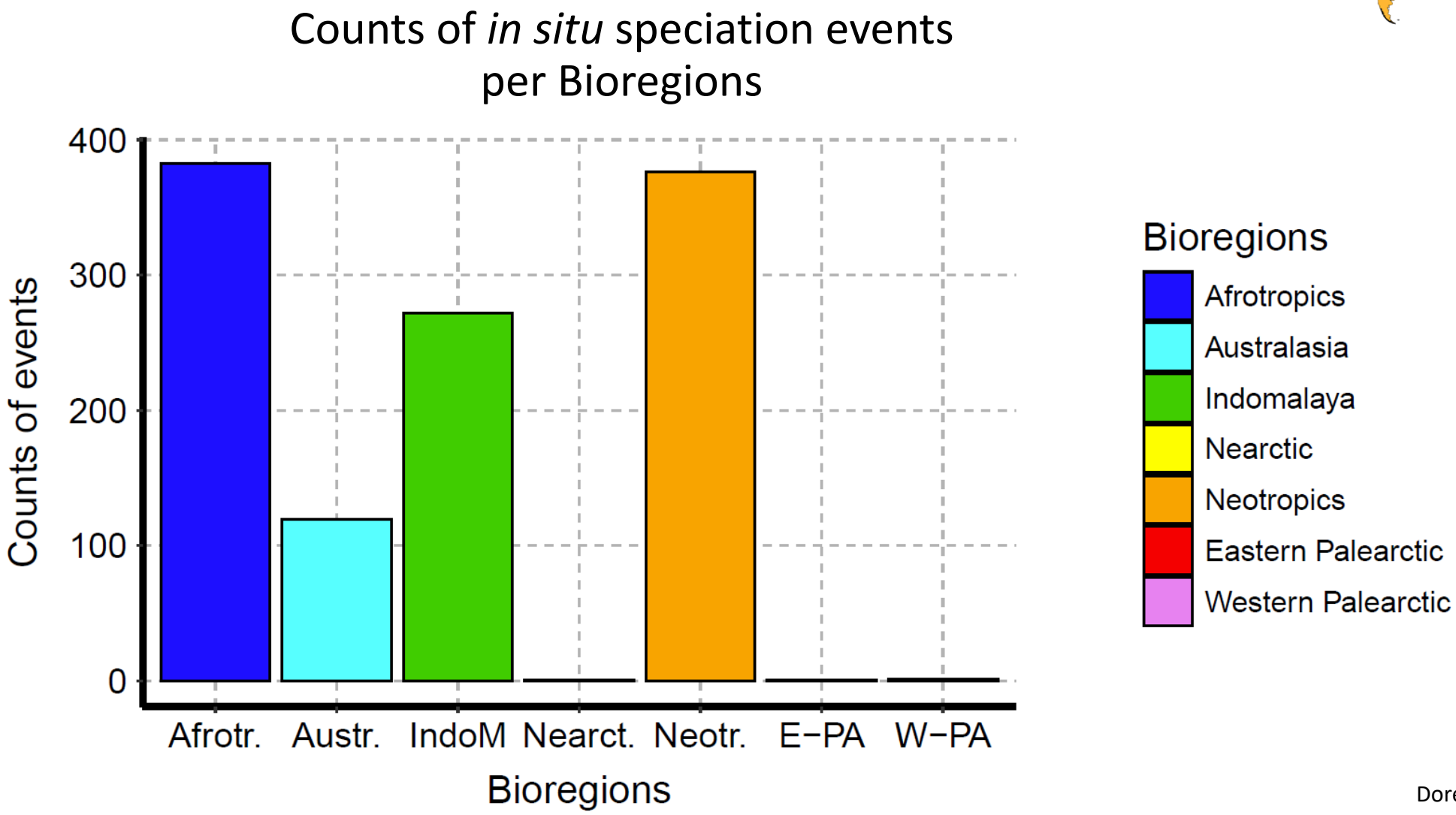

Percentages
*in situ* speciation events
per Bioregions




Doré et al., *in prep.*

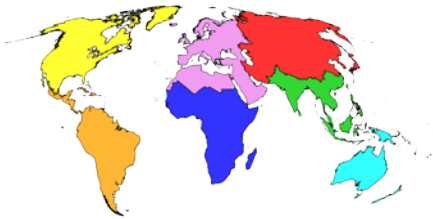
# Overall *in situ* diversification

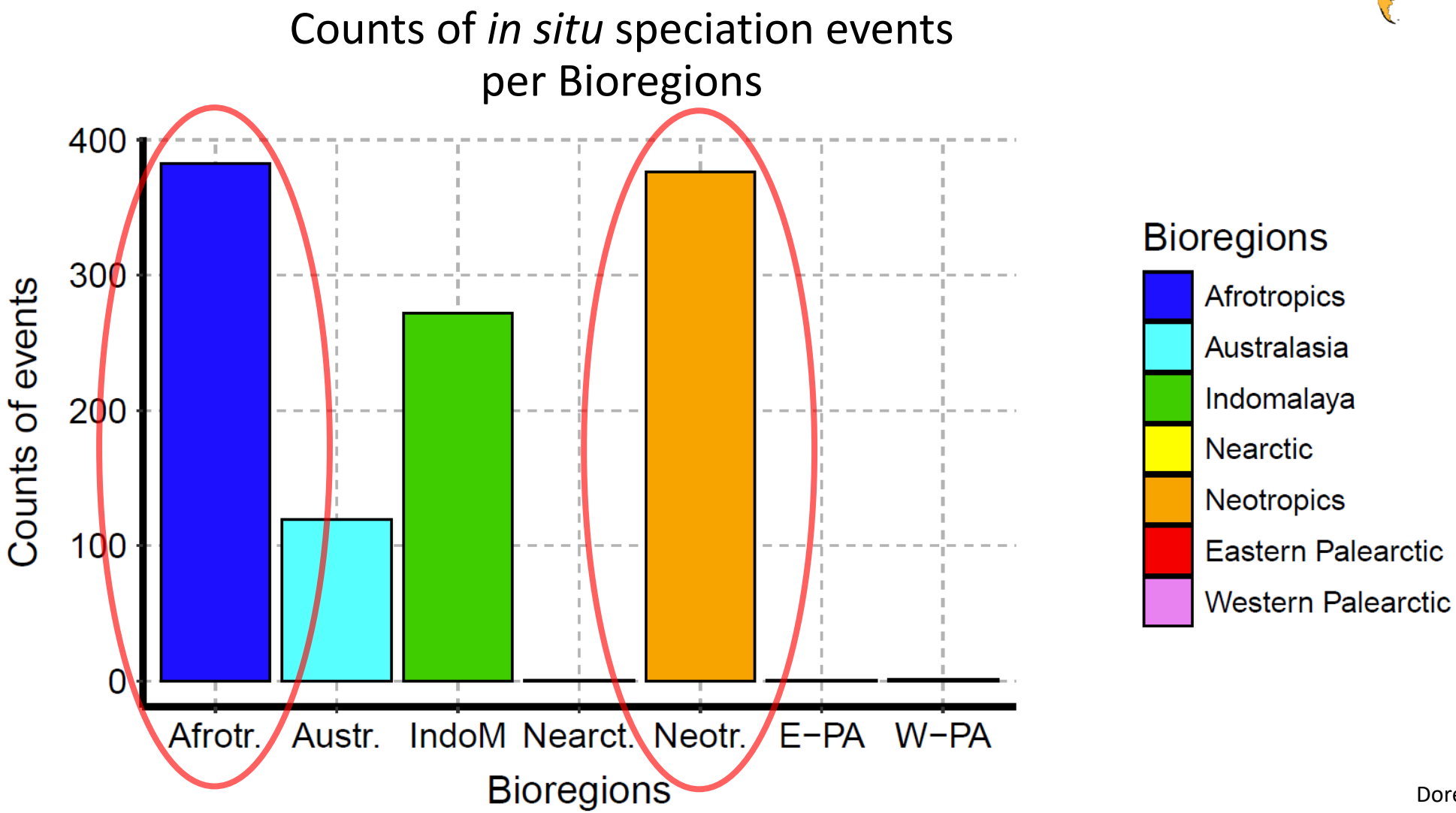





Doré et al., *in prep.*

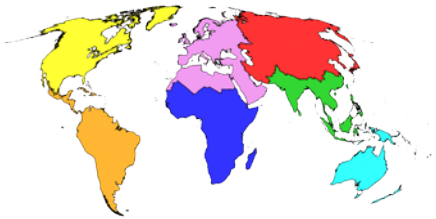
# Overall *in situ* diversification

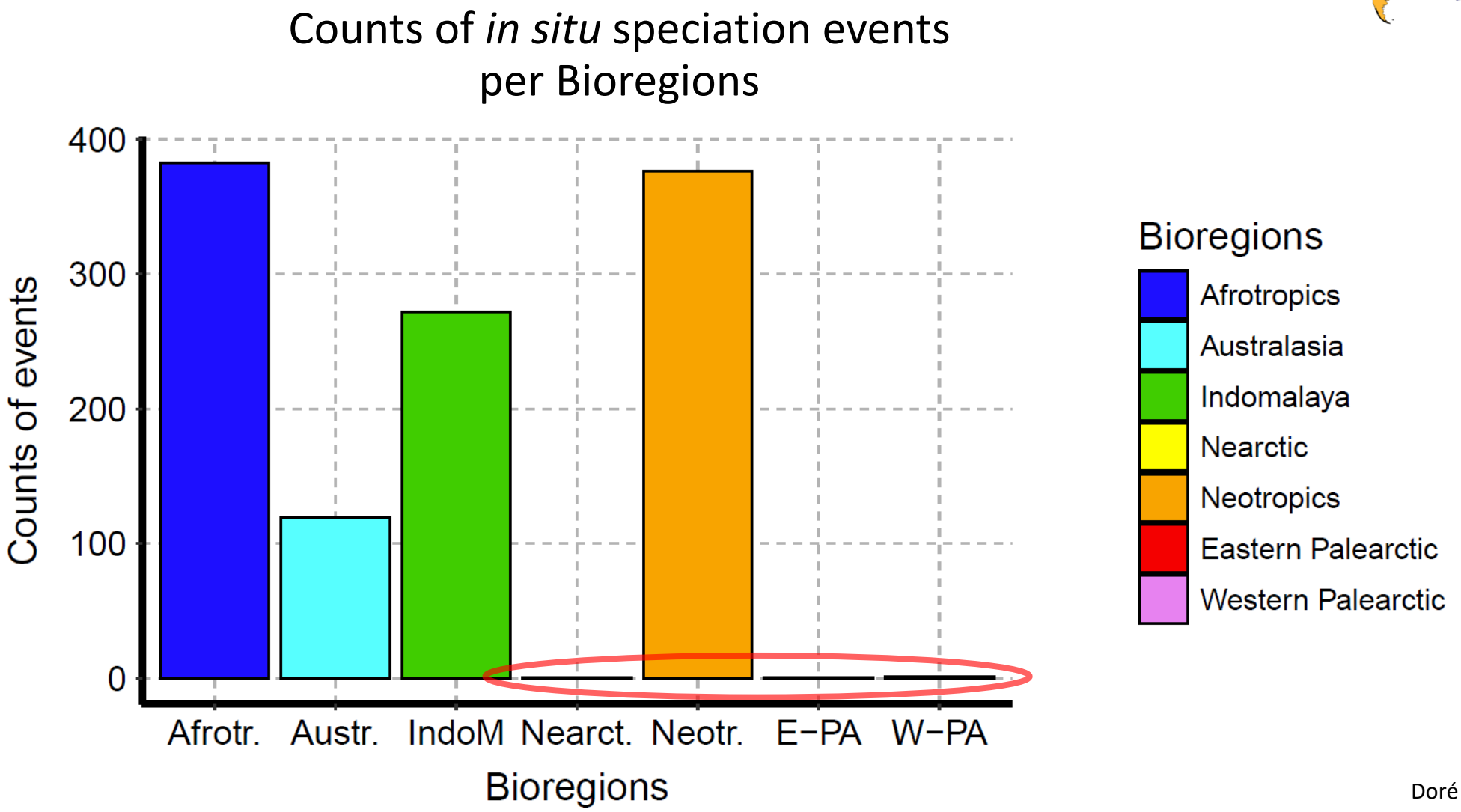





Doré et al., *in prep.*

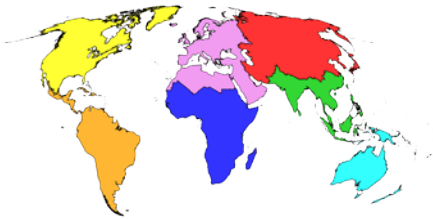
# Overall *in situ* diversification






Doré et al., *in prep.*

# Conclusions



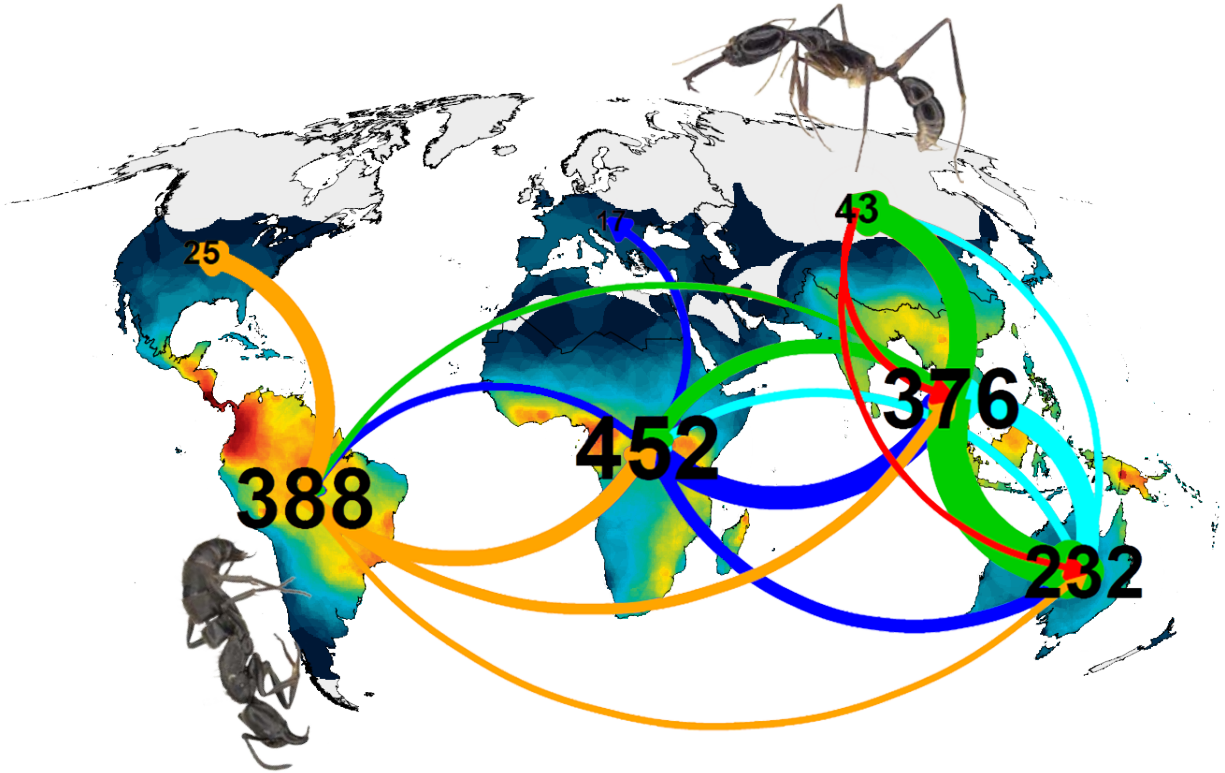
Origin in **Gondwana** ≈ 110-120 My

Macroevolutionary sources/sinks:

- **Source** = **Neotropics**
- **Sinks** = Holoarctic = **Palearctic** + **Nearctic**
- **Centers** = **Afrotropics/IndoMalaya/Australasia**

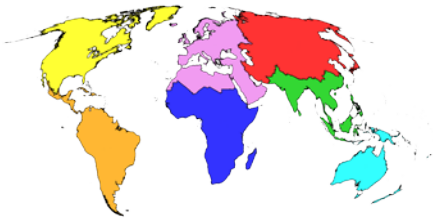
Mechanisms?

- Tropical **niche conservatism**
- Early diversification and **niche occupation** in **Neotropics** limits further immigration




Doré et al., *in prep.*

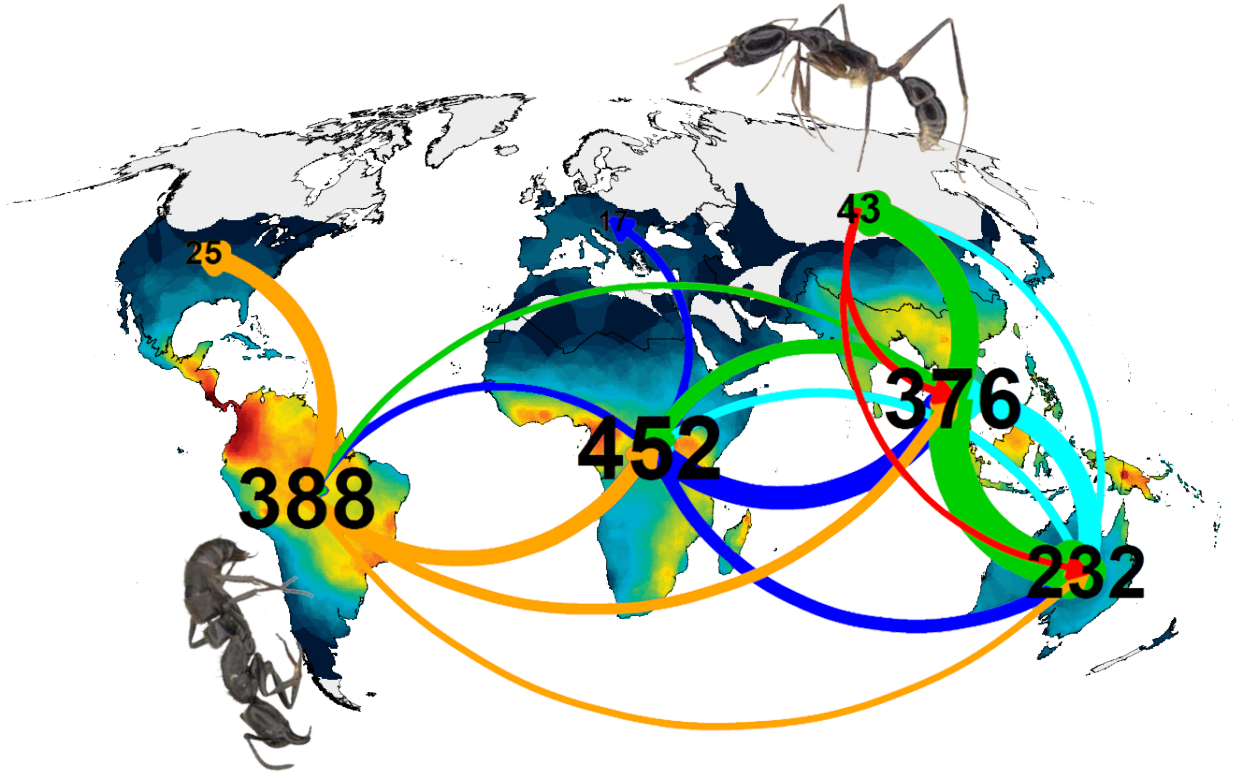
# Conclusions



Origin in **Gondwana** ≈ 110-120 My

Macroevolutionary sources/sinks:

- **Source** = Neotropics
- **Sinks** = Holoarctic = Palearctic + Nearctic
- **Centers** = Afrotropics/IndoMalaya/Australasia

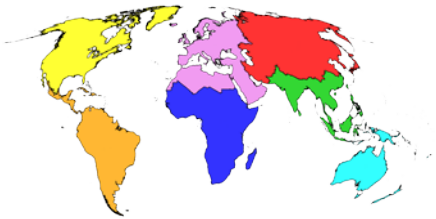



Mechanisms?

- Tropical **niche conservatism**
- Early diversification and **niche occupation** in **Neotropics** limits further immigration

Doré et al., *in prep.*

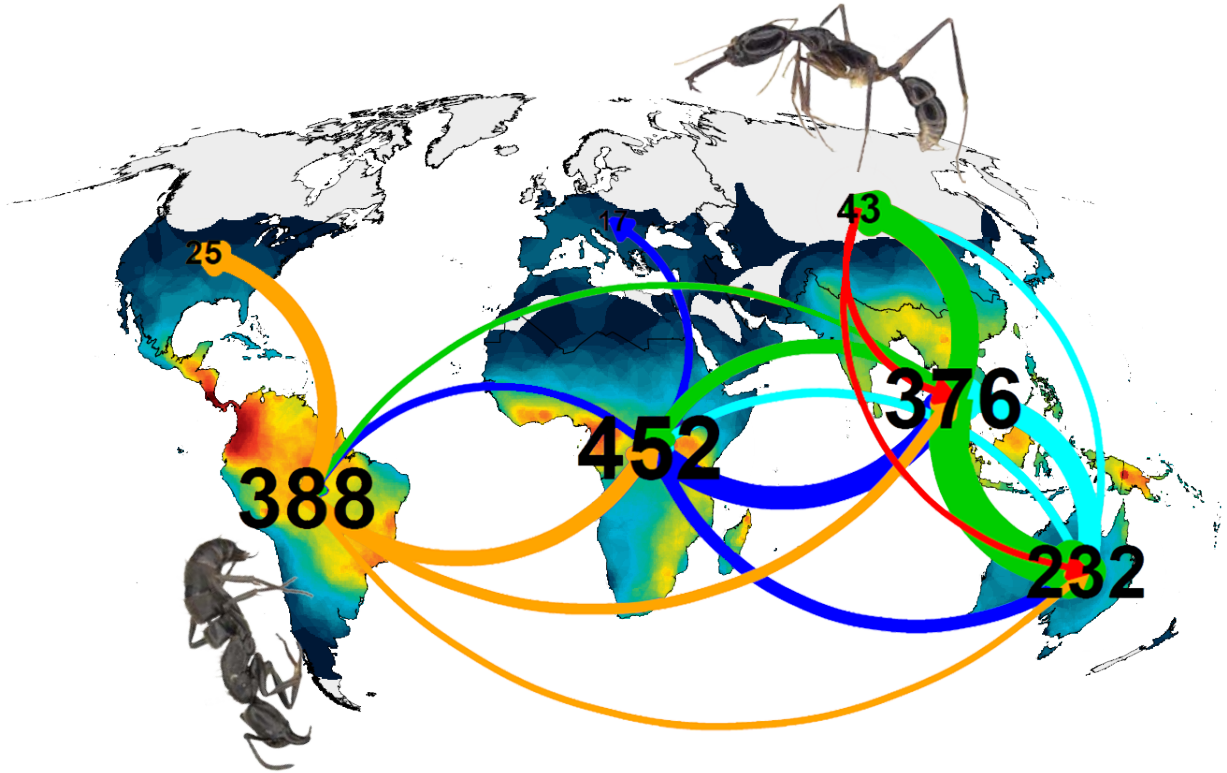
# Conclusions



Origin in **Gondwana** ≈ 110-120 My

Macroevolutionary sources/sinks:

- **Source** = Neotropics
- **Sinks** = Holoarctic = Palearctic + Nearctic
- **Centers** = Afrotropics/IndoMalaya/Australasia

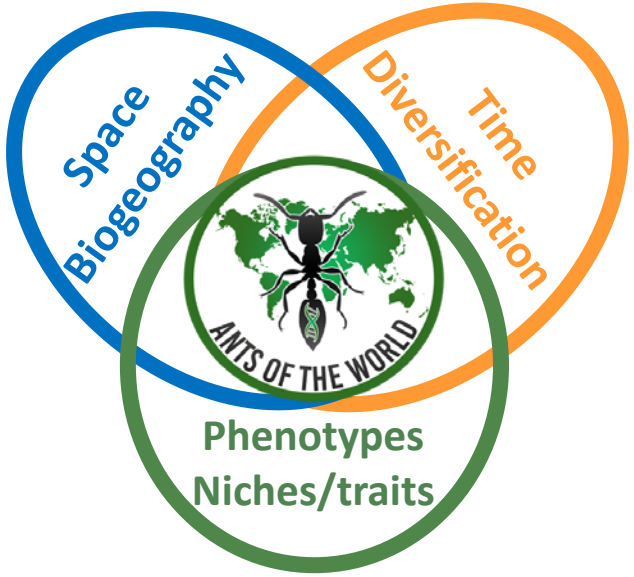



Mechanisms?

- Tropical **niche conservatism**
- Early diversification and **niche occupation** in Neotropics limits further immigration

Doré et al., *in prep.*

# Perspectives




**Climatic niche evolution**:

- Is evolution repeating itself? Compare the **filling of the niche** across bioregions

**Morphology** & **Diversification**:

- Do **key morphological innovations** trigger diversification uplifts?


**ANTS OF THE WORLD**

DEB-1932467, DEB-1932062, DEB-2019431


**Bonnie Blaimer**

MUSEUM FÜR NATURKUNDE BERLIN

CIBD Center for Integrative Biodiversity Discovery

**Marek Borowiec**

CSU

**Michael Branstetter**

USDA ARS

STATE U

**Gabi Camacho**

USP Universidade de São Paulo

**Brian Fisher**

CALIFORNIA ACADEMY OF SCIENCES

**Jack Longino**

THE UNIVERSITY OF UTAH

**Phil Ward**

UCDAVIS UNIVERSITY OF CALIFORNIA

SCAN ME

# Questions?

Photo credits: F. Brassard

# References


Runfola, D. et al. (2020) geoBoundaries: A global database of political administrative boundaries. PLoS ONE 15(4): e0231866. https://doi.org/10.1371/journal.pone.0231866

Schmidt, C. A.; Shattuck, S. O. 2014. The higher classification of the ant subfamily Ponerinae (Hymenoptera: Formicidae), with a review of ponerine ecology and behavior. Zootaxa 3817 (1):1-242. 10.11646/zootaxa.3817.1.1

Scotese, C. (2004). Maps from the PALEOMAP Project. Retrieved from http://www.scotese.com/ on May 2024.

Zizka A, Silvestro D, Andermann T, Azevedo J, Duarte Ritter C, Edler D, Farooq H, Herdean A, Ariza M, Scharn R, Svanteson S, Wengstrom N, Zizka V, Antonelli A (2019). "CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases." *Methods in Ecology and Evolution*, -7. doi:10.1111/2041-210X.13152, R package version 3.0.1


.

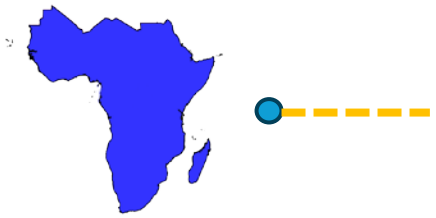
# Grafting missing taxa

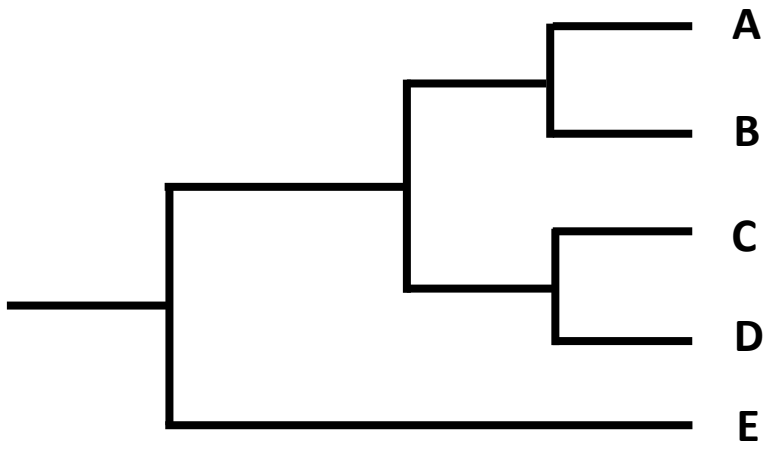
**Backbone UCE phylogeny:** 789 species (51.4%)

**Aim:** Impute all **745 missing valid species** with no UCE data using available **taxonomic** and **biogeographic** information

Step 1: Define grafting **location** Step 2: Define **divergence time** Step 3: **Randomize** and **repeat**

Missing taxa
from Afrotropics





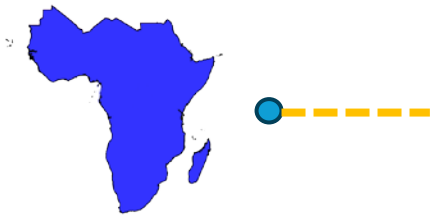
# Grafting missing taxa

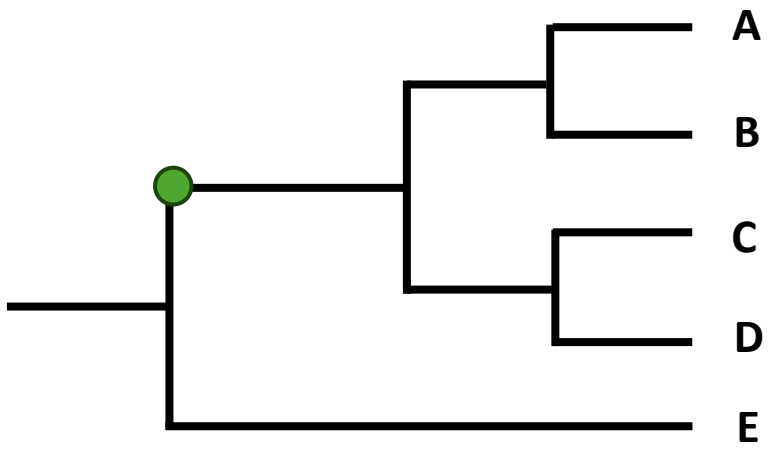
**Backbone UCE phylogeny:** 789 species (51.4%)

**Aim:** Impute all **745 missing valid species** with no UCE data using available **taxonomic** and **biogeographic** information

Step 1: Define grafting **location** Step 2: Define **divergence time** Step 3: **Randomize** and **repeat**

Missing taxa
from Afrotropics





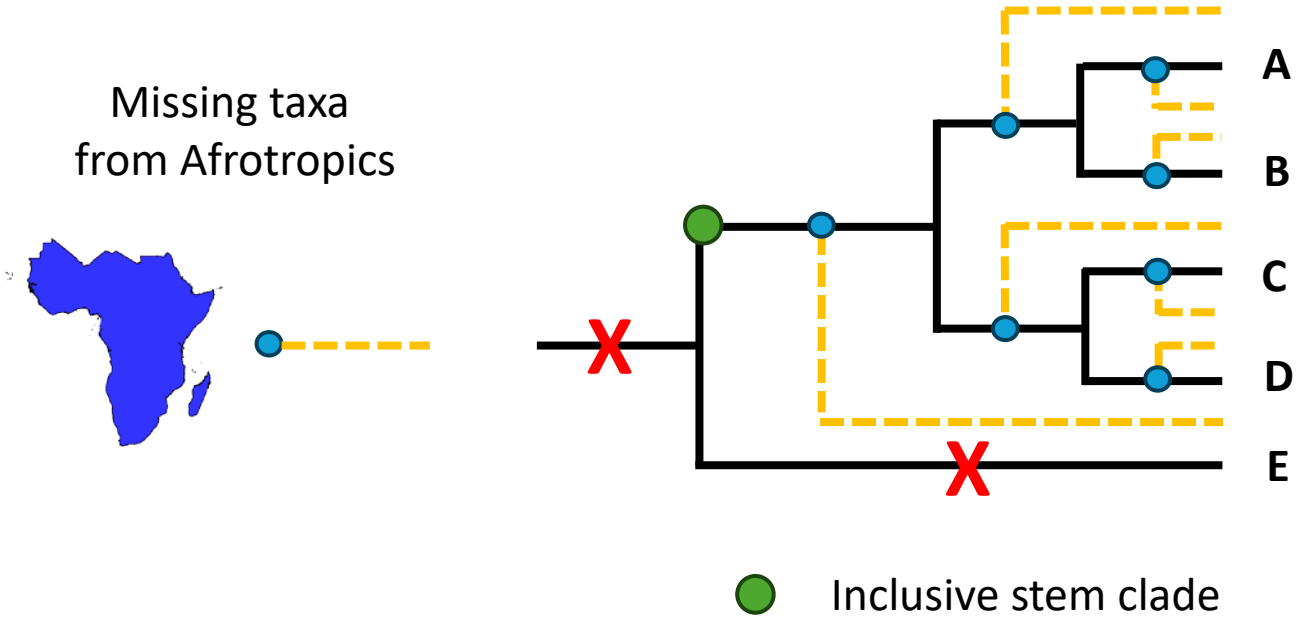

Inclusive stem clade

# Grafting missing taxa

**Backbone UCE phylogeny:** 789 species (51.4%)

**Aim:** Impute all **745 missing valid species** with no UCE data using available **taxonomic** and **biogeographic** information

Step 1: Define grafting **location** Step 2: Define **divergence time** Step 3: **Randomize** and **repeat**

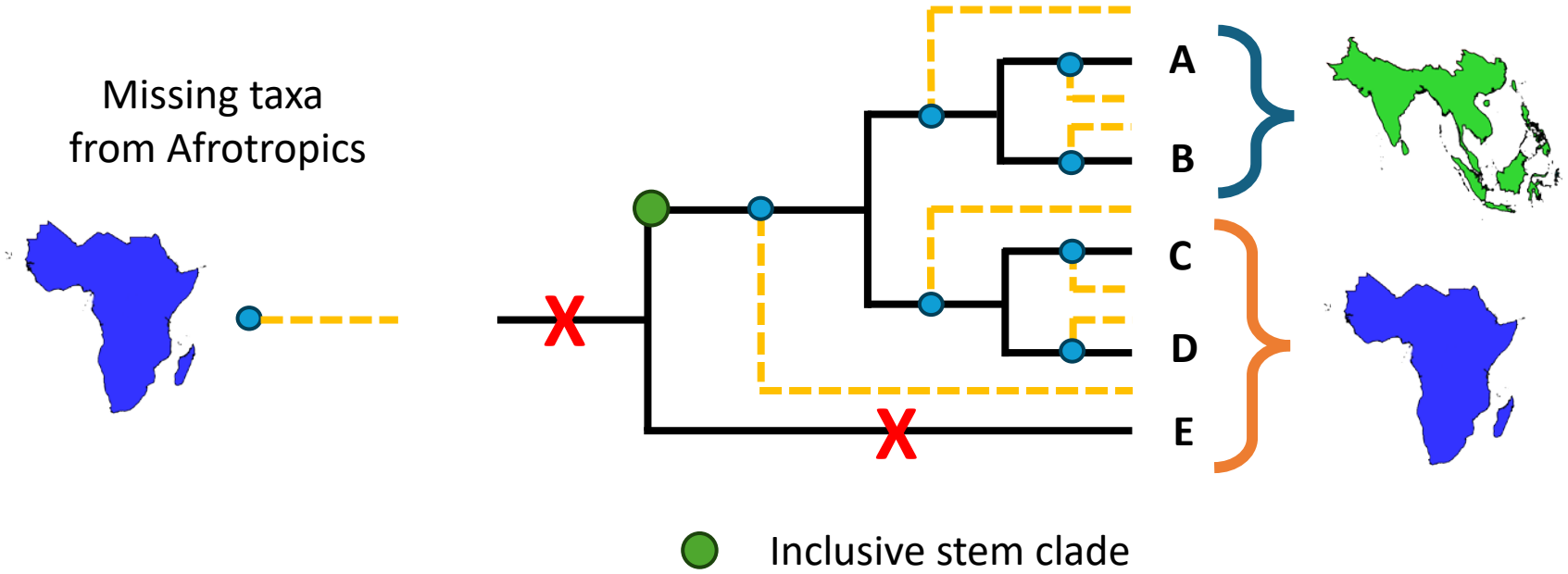


# Grafting missing taxa

**Backbone UCE phylogeny:** 789 species (51.4%)

**Aim:** Impute all **745 missing valid species** with no UCE data using available **taxonomic** and **biogeographic** information

Step 1: Define grafting **location** Step 2: Define **divergence time** Step 3: **Randomize** and **repeat**



# Grafting missing taxa

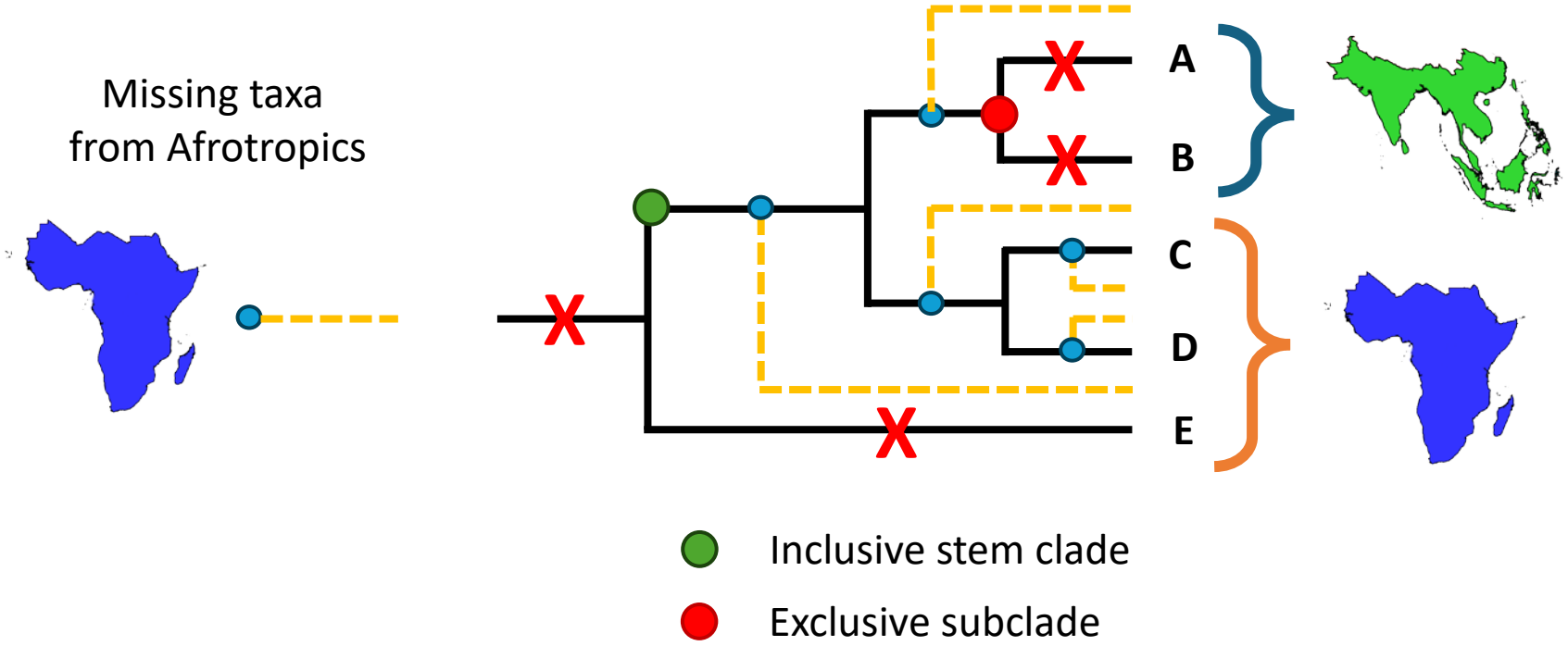
**Backbone UCE phylogeny:** 789 species (51.4%)

**Aim:** Impute all **745 missing valid species** with no UCE data using available **taxonomic** and **biogeographic** information

Step 1: Define grafting **location**

Step 2: Define **divergence time**

Step 3: **Randomize** and **repeat**



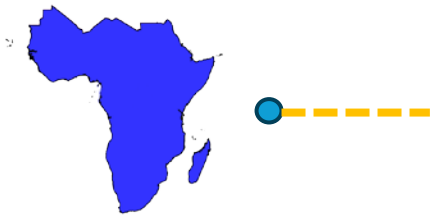
# Grafting missing taxa

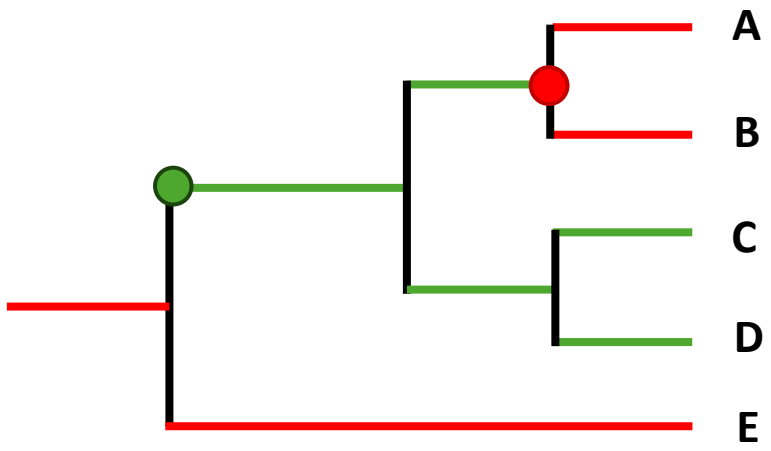
**Backbone UCE phylogeny:** 789 species (51.4%)

**Aim:** Impute all **745 missing valid species** with no UCE data using available **taxonomic** and **biogeographic** information

Step 1: Define grafting **location** Step 2: Define **divergence time** Step 3: **Randomize** and **repeat**

Missing taxa
from Afrotropics






Inclusive stem clade

Exclusive subclade

# Grafting missing taxa

**Backbone UCE phylogeny:** 789 species (51.4%)

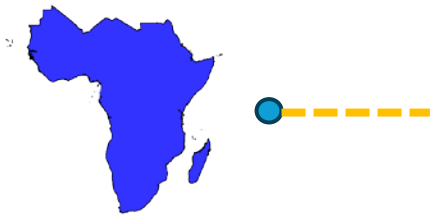
**Aim:** Impute all **745 missing valid species** with no UCE data using available **taxonomic** and **biogeographic** information

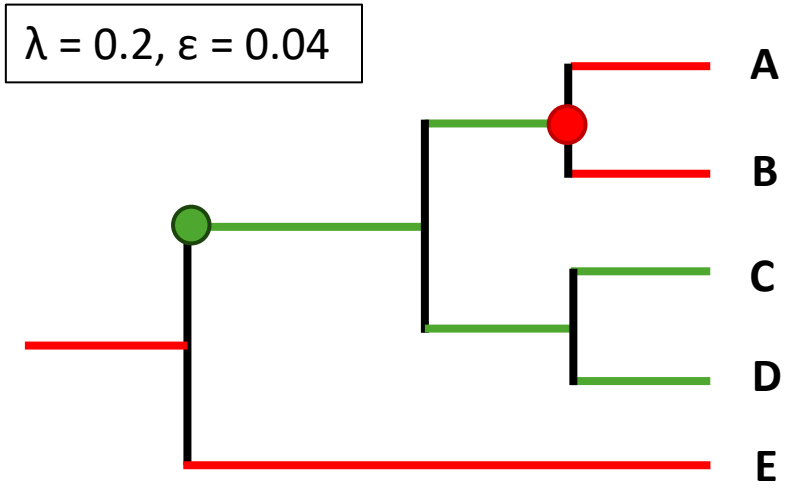
Step 1: Define grafting **location**

Step 2: Define **divergence time**

Step 3: **Randomize** and repeat

Missing taxa
from Afrotropics






1/ Fit BD model within
with *phytools::fit.bd*

2/ Simulate age of missing taxa
with *TreeSim::corsim*

3/ Select randomly location
among **valid branches**

Inclusive stem clade

Exclusive subclade

# Grafting missing taxa

**Backbone UCE phylogeny:** 789 species (51.4%)

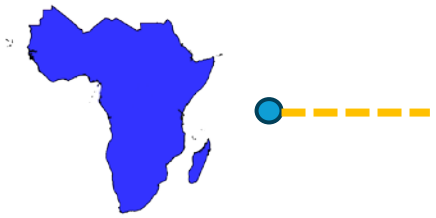
**Aim:** Impute all **745 missing valid species** with no UCE data using available **taxonomic** and **biogeographic** information

Step 1: Define grafting **location**

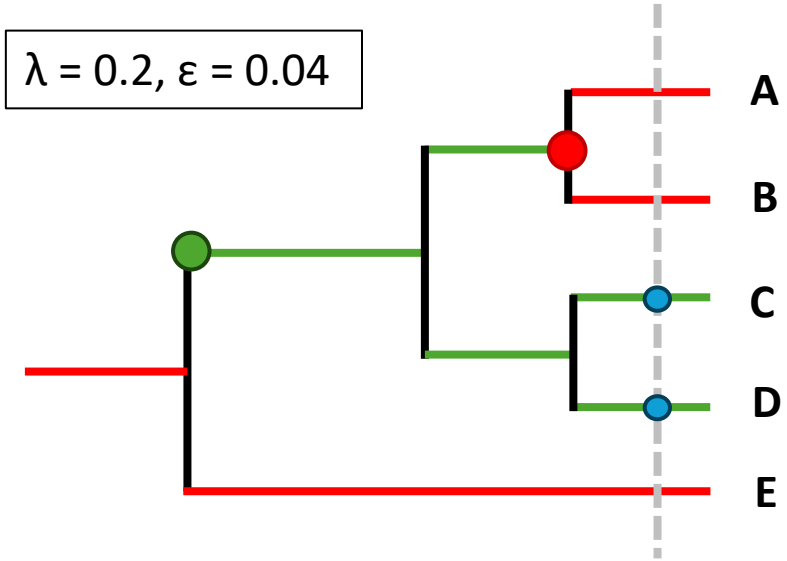
Step 2: Define **divergence time**

Step 3: **Randomize** and repeat

Missing taxa
from Afrotropics



$\lambda = 0.2, \varepsilon = 0.04$




1/ Fit BD model within
with *phytools::fit.bd*

2/ Simulate age of missing taxa
with *TreeSim::corsim*

3/ Select randomly location
among **valid branches**

Inclusive stem clade

Exclusive subclade

# Grafting missing taxa

**Backbone UCE phylogeny:** 789 species (51.4%)

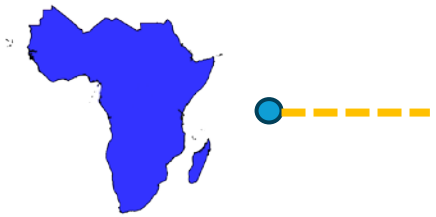
**Aim:** Impute all **745 missing valid species** with no UCE data using available **taxonomic** and **biogeographic** information

Step 1: Define grafting **location**

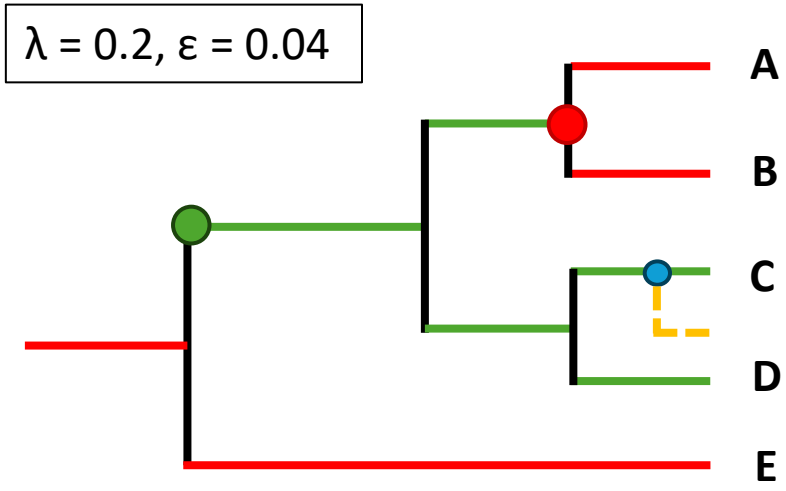
Step 2: Define **divergence time**

Step 3: **Randomize** and **repeat**

Missing taxa
from Afrotropics



$\lambda = 0.2, \varepsilon = 0.04$




1/ Fit BD model within
with *phytools::fit.bd*

2/ Simulate age of missing taxa
with *TreeSim::corsim*

3/ Select randomly location
among **valid branches**

Inclusive stem clade

Exclusive subclade

# Grafting missing taxa

**Backbone UCE phylogeny:** 789 species (51.4%)

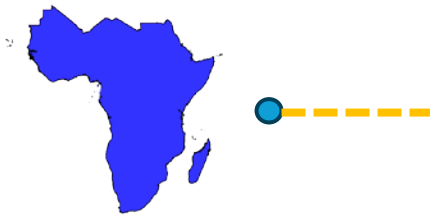
**Aim:** Impute all **745 missing valid species** with no UCE data using available **taxonomic** and **biogeographic** information

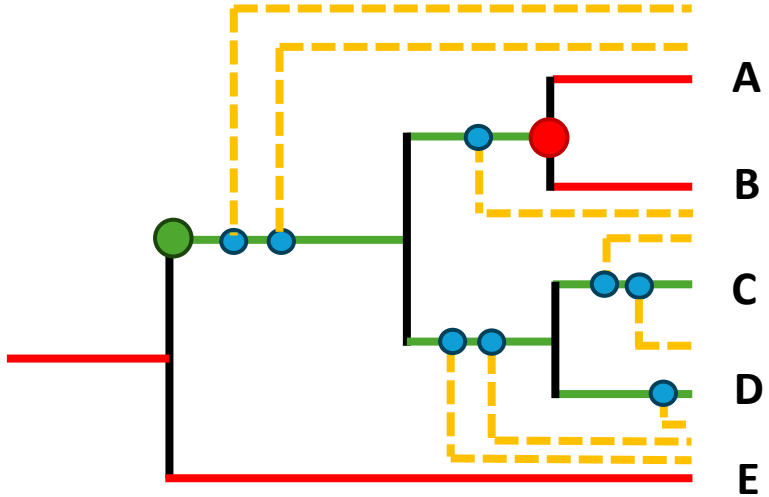
Step 1: Define grafting **location**

Step 2: Define **divergence time**

Step 3: **Randomize** and **repeat**

Missing taxa
from Afrotropics






1/ Fit BD model within 
with *phytools::fit.bd*

2/ Simulate age of missing taxa
with *TreeSim::corsim*

3/ Select randomly location
among **valid branches**

**x 1000**

Inclusive stem clade

Exclusive subclade

# Biogeographic curation

**Data sources**: GABI v1.0 / Antmaps.org (93,568 records) and AntWeb (62,292 records)

**Goal:** Curate occurrence coordinates to use them for **biogeographic** inferences, **climatic niche** evolution, **diversity** maps, **latitudinal** and **elevation** gradient analyses, …

- 1/ **Automatic detection** of potential errors
- 2/ **Manual checking** of outliers:
- 3/ **Retrieve occurrences** from administrative locations

# Biogeographic curation

**Data sources**: GABI v1.0 / Antmaps.org (93,568 records) and AntWeb (62,292 records)

**Goal:** Curate occurrence coordinates to use them for **biogeographic** inferences, **climatic niche** evolution, **diversity** maps, **latitudinal** and **elevation** gradient analyses, …

- 1/ **Automatic detection** of potential errors
- 2/ **Manual checking** of outliers:
- 3/ **Retrieve occurrences** from administrative locations

# Automatic detection

Detect ‘classic’ errors: *CoordinateCleaner* package (Zizka et al., 2019)

- Fall in **water**
- Fall in **wrong country**
- Coordinates as country **centroid**
- Coordinates as country **capital**
- In **institutions**’ collections

Detect **outliers**: need expert evaluation

- 1000 km min **distance**
- With less than 25% of coordinates in a single **bioregion**
- **Interactive maps** can be made available for **tricky cases** to evaluate
- **Conservative approach** = remove all outliers

# Automatic detection

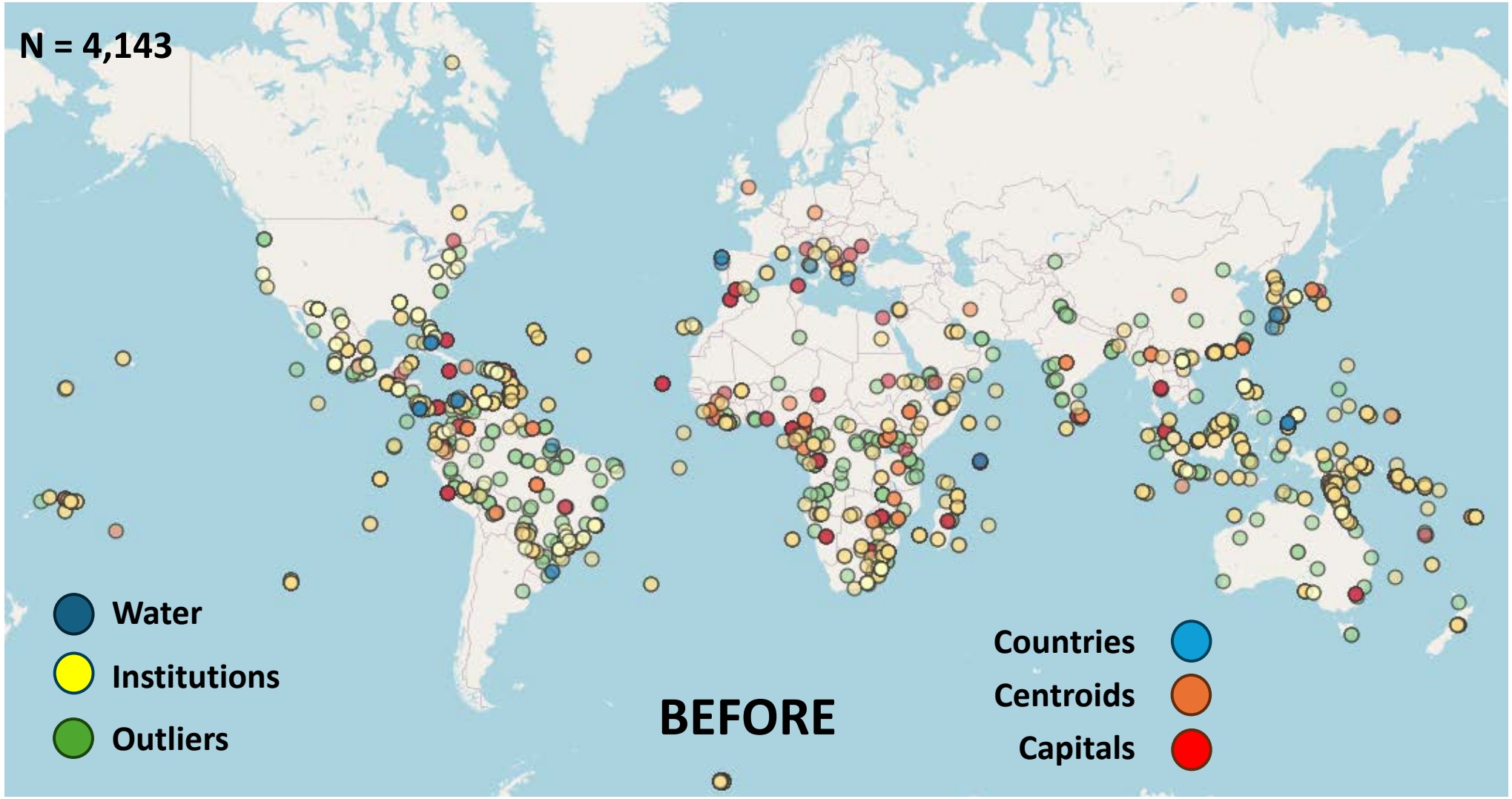
Detect ‘classic’ errors: *CoordinateCleaner* package (Zizka et al., 2019)

- Fall in **water**
- Fall in **wrong country**
- Coordinates as country **centroid**
- Coordinates as country **capital**
- In **institutions**’ collections

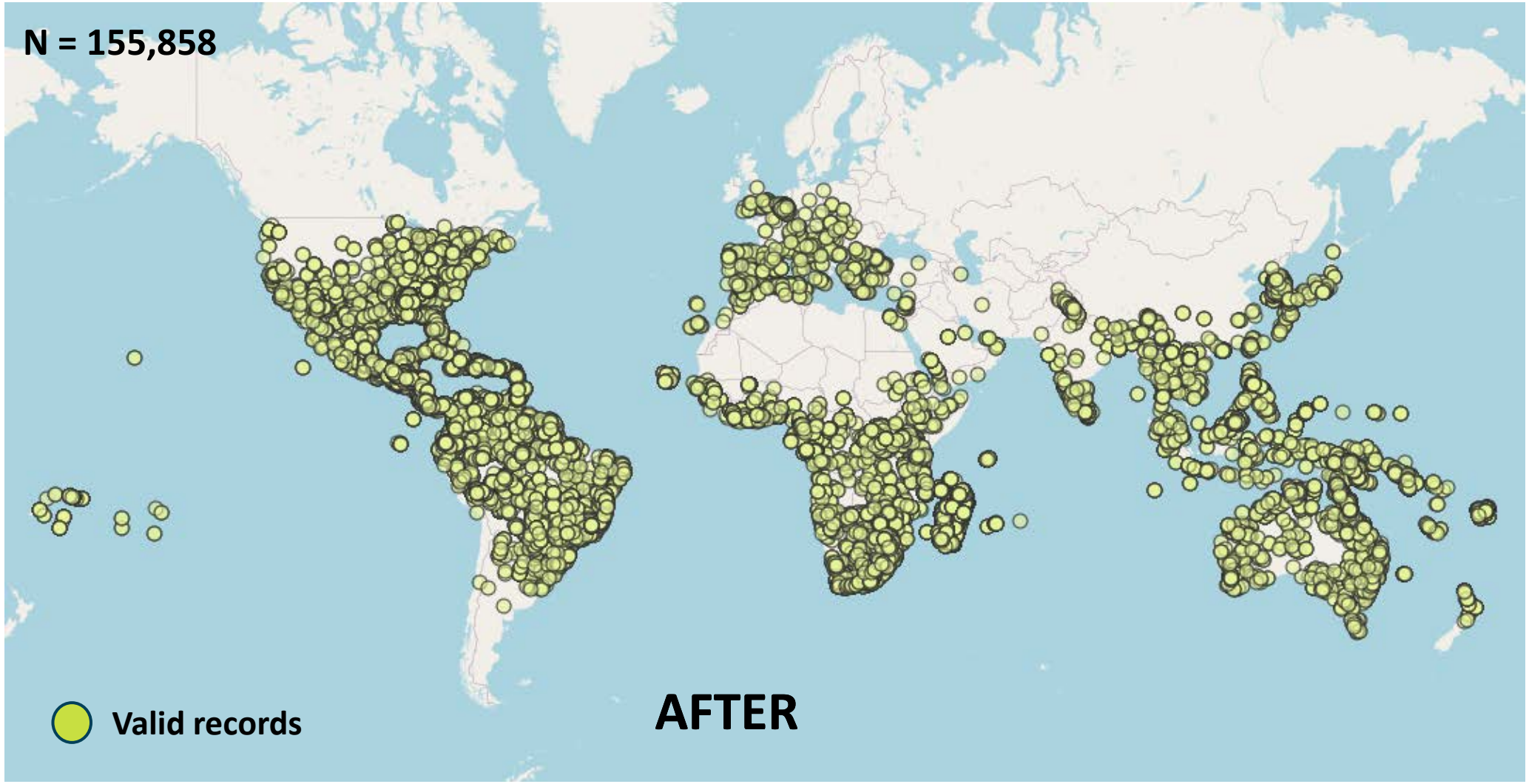
Detect **outliers**: need expert evaluation

- 1000 km min **distance**
- With less than 25% of coordinates in a single **bioregion**
- **Interactive maps** can be made available for **tricky cases** to evaluate
- **Conservative approach** = remove all outliers

# Automatic detection



# Automatic detection



# Biogeographic curation

**Data sources**: GABI v1.0 / Antmaps.org (93,568 records) and AntWeb (62,292 records)

**Goal:** Curate occurrence coordinates to use them for **biogeographic** inferences, **climatic niche** evolution, **diversity** maps, **latitudinal** and **elevation** gradient analyses, …

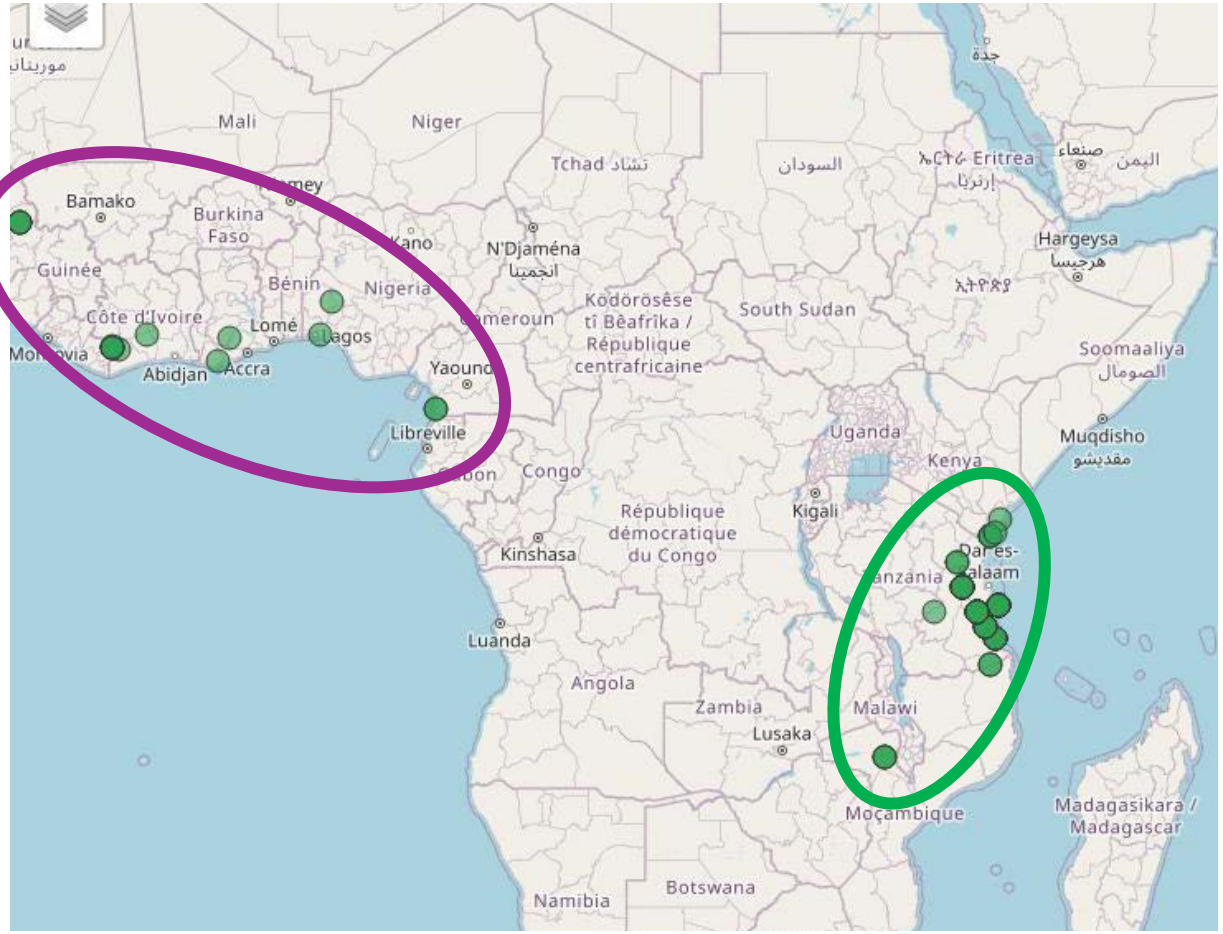
- 1/ **Automatic detection** of potential errors
- 2/ **Manual checking** of outliers:
- 3/ **Retrieve occurrences** from administrative locations

# Manual checking of outliers

Outlier


Example of distance outlier detection

# Manual checking of outliers




Example of limits of outlier detection
Case of recently split taxa

# Biogeographic curation

**Data sources**: GABI v1.0 / Antmaps.org (93,568 records) and AntWeb (62,292 records)

**Goal:** Curate occurrence coordinates to use them for **biogeographic** inferences, **climatic niche** evolution, **diversity** maps, **latitudinal** and **elevation** gradient analyses, …

- 1/ **Automatic detection** of potential errors
- 2/ **Manual checking** of outliers:
- 3/ **Retrieve occurrences** from administrative locations

# Retrieve occurrences

**Data sources**:

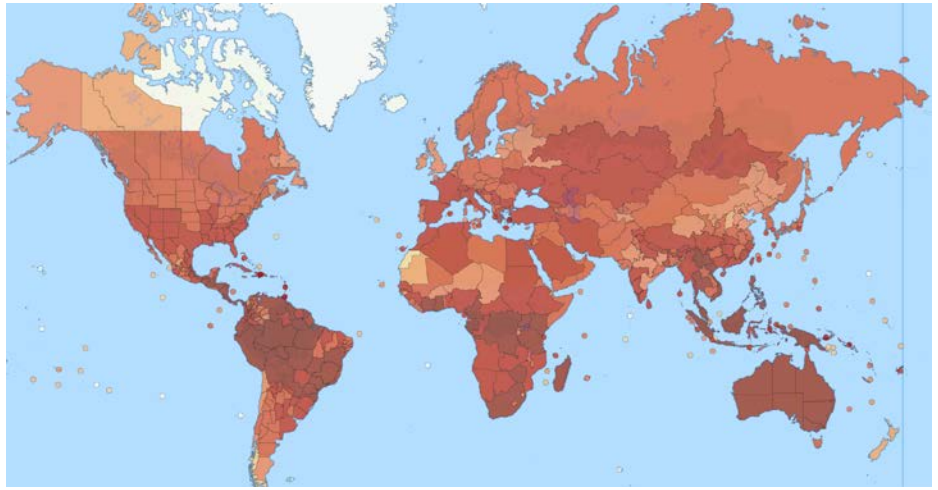
- AntWeb: 2,372 occurrences **without coordinates** (3.8%)
- GABI: 47,972 occurrences **without coordinates** (51.2% !)

Solution: Draw them at random in **geounit** when known

- Quantify **uncertainty** to allow filtering = area of the geounit

Challenge: **Homogenize** geographic unit systems

- Convert to a new curated system: **GeoBoundaries** (Runfola et al., 2020)



**GABI/AntMaps** = bentity2 system

**AntWeb** = Geographic Names Server (GNS)

Point data, but no polygons available?!

# Retrieve occurrences

**Data sources**:

- AntWeb: 2,372 occurrences **without coordinates** (3.8%)
- GABI: 47,972 occurrences **without coordinates** (51.2% !)

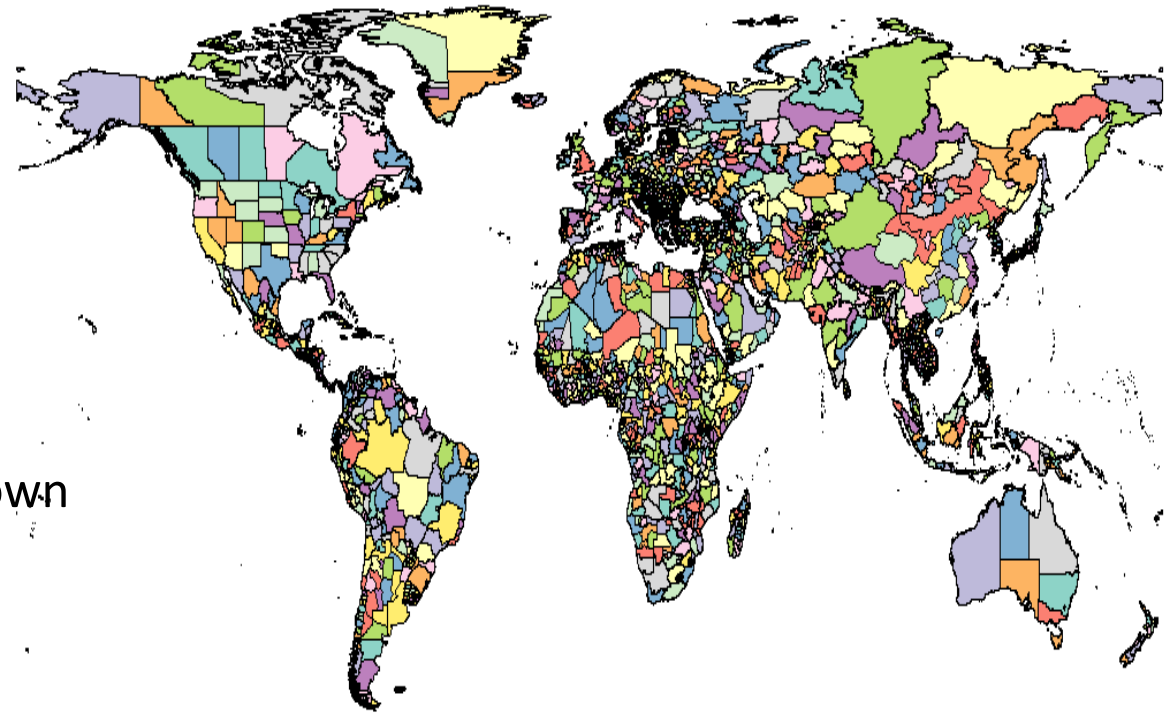
Solution: Draw them at random in **geounit** when known

- Quantify **uncertainty** to allow filtering = area of the geounit

Challenge: **Homogenize** geographic unit systems

- Convert to a new curated system: **GeoBoundaries**

(Runfola et al., 2020)



**GeoBoundaries**

GeoLab, William & Mary College
3,223 adm1 polygons

# Retrieve occurrences

**Data sources**:

- AntWeb: 2,372 occurrences **without coordinates** (3.8%)
- GABI: 47,972 occurrences **without coordinates** (51.2% !)

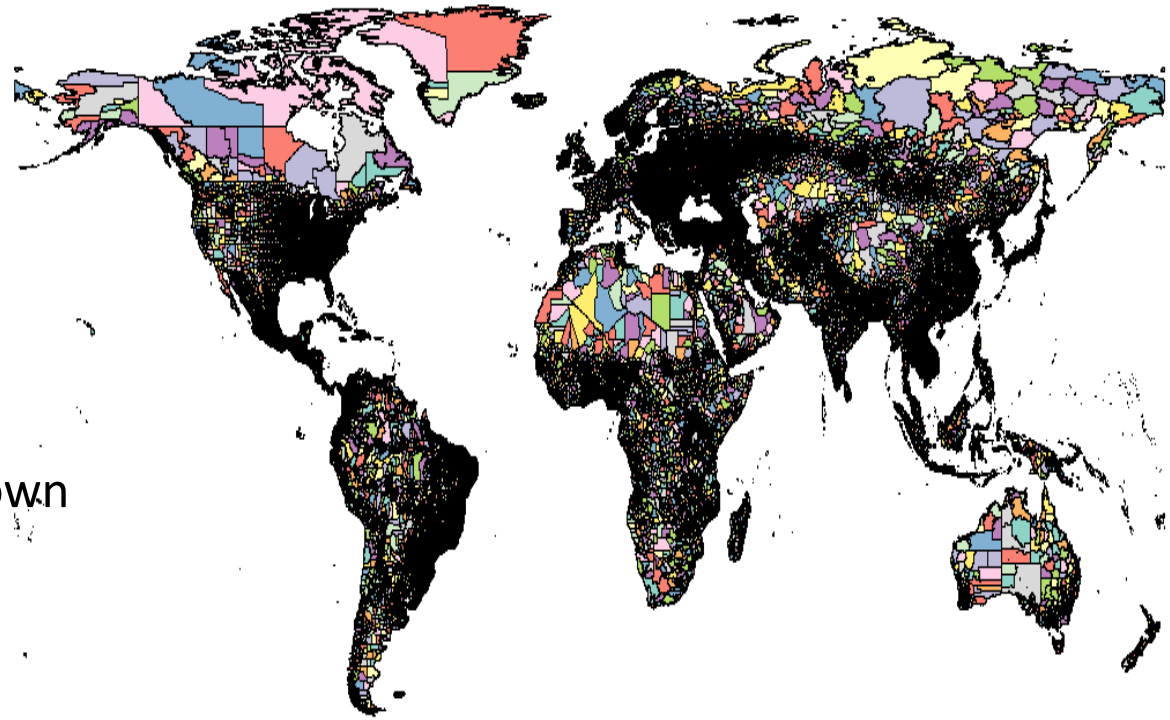
Solution: Draw them at random in **geounit** when known

- Quantify **uncertainty** to allow filtering = area of the geounit

Challenge: **Homogenize** geographic unit systems

- Convert to a new curated system: **GeoBoundaries**

(Runfola et al., 2020)



**GeoBoundaries**

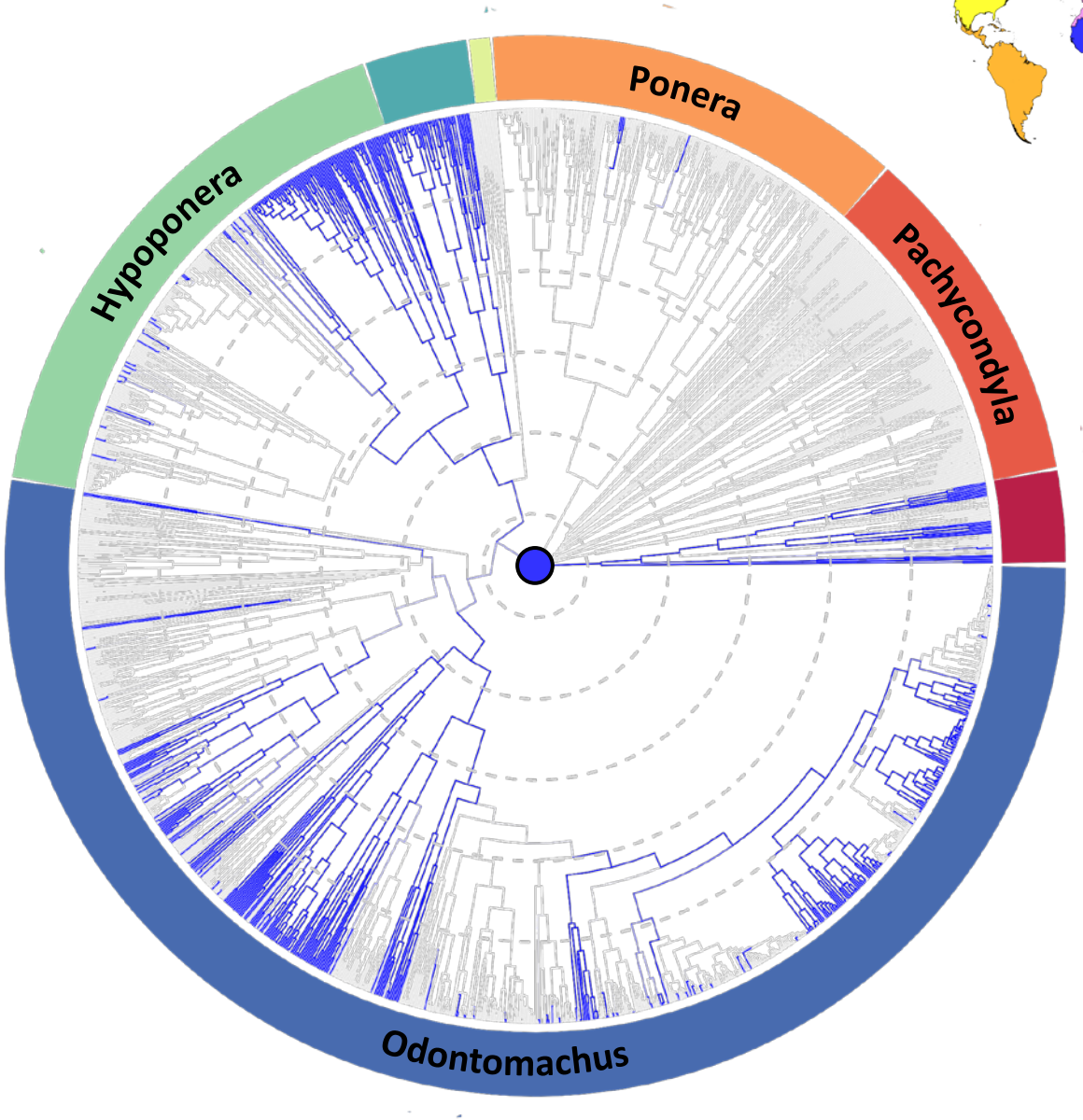
GeoLab, William & Mary College
3,223 adm1 polygons
49,476 adm2 polygons

# Origins in bioregions

## Afrotropics

Current richness = **452 sp.**

First colonization = **113 My?**



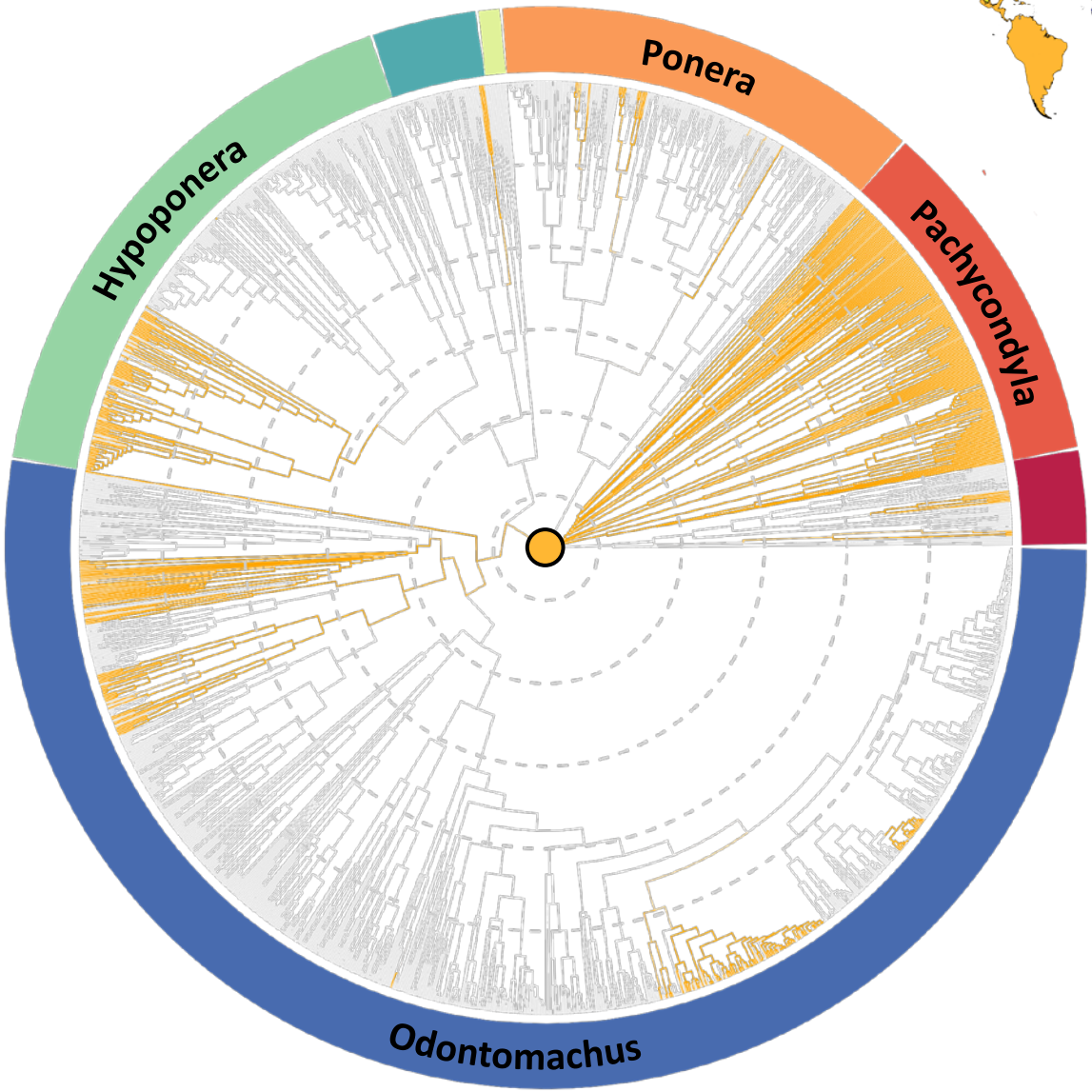

Doré et al., *in prep.*

# Origins in bioregions

## Neotropics

Current richness = **388 sp.**

First colonization = **113 My?**



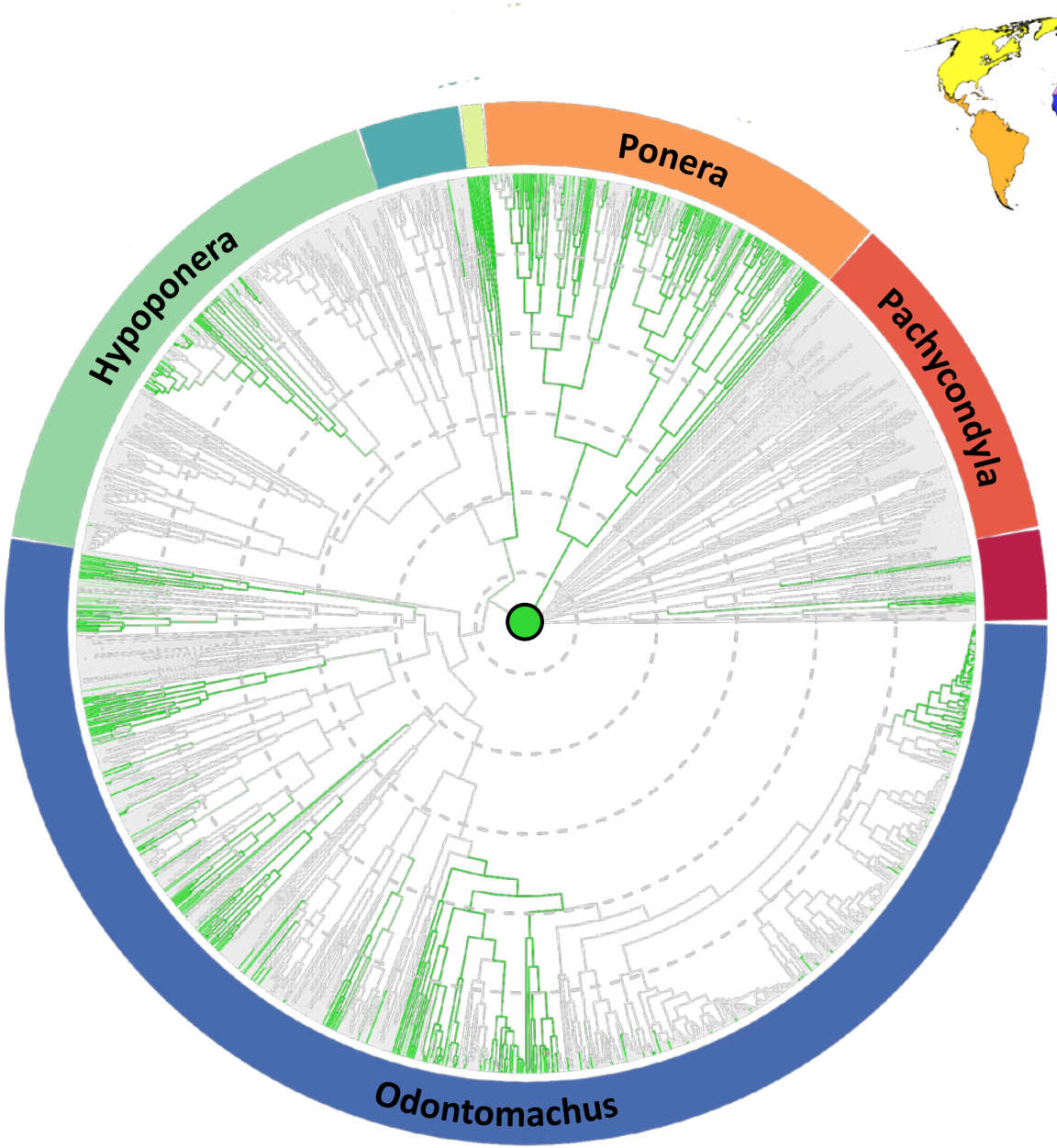

Doré et al., *in prep.*

# Origins in bioregions

## IndoMalaya

Current richness = **376 sp.**

First colonization = **113 My?**



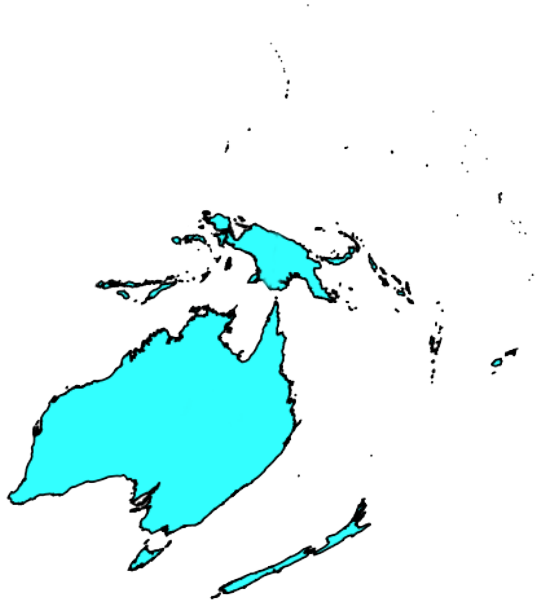

Doré et al., *in prep.*

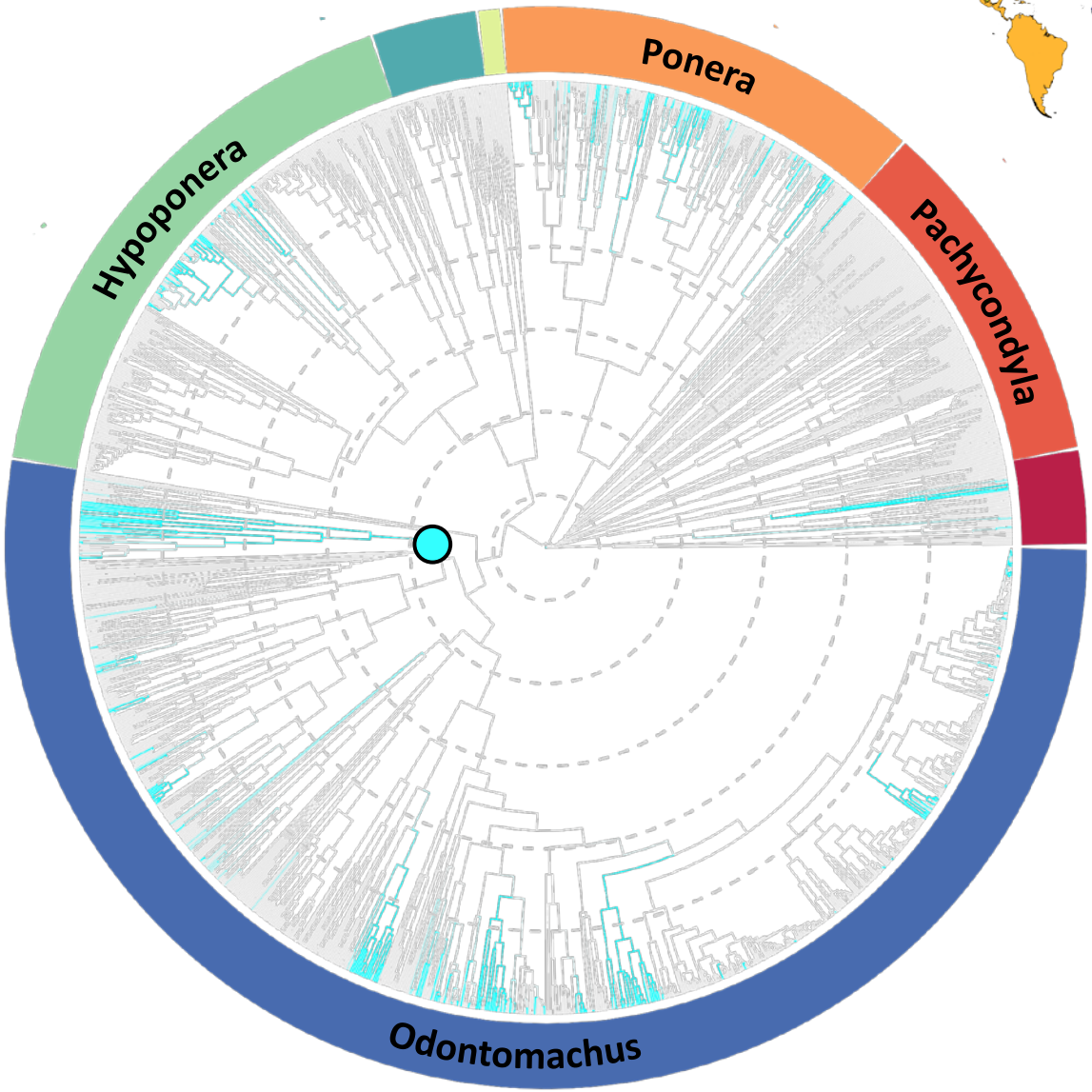
# Origins in bioregions

## Australasia

Current richness = **232 sp.**

First colonization = **90 My?**





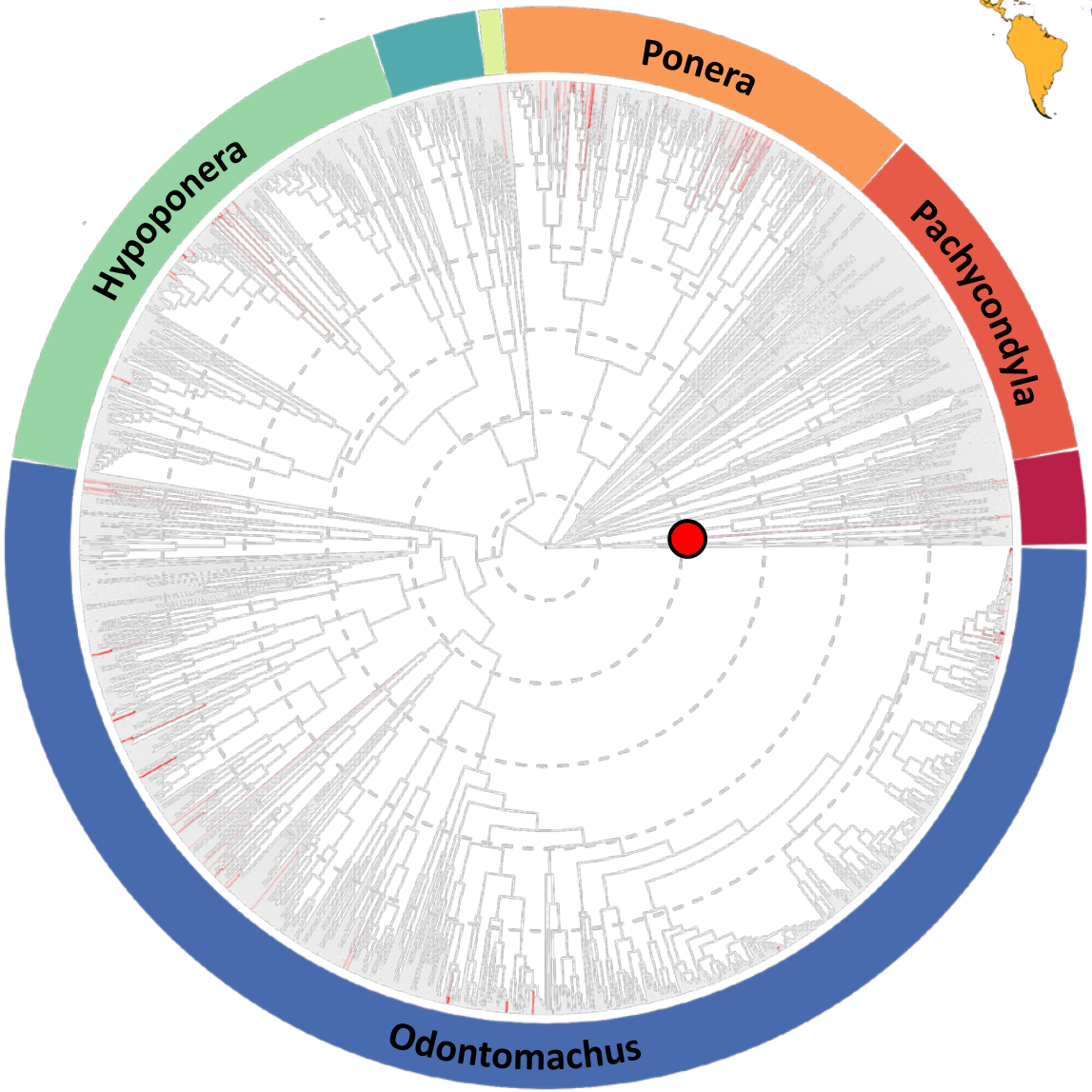

Doré et al., *in prep.*

# Origins in bioregions

## Eastern Palearctic

Current richness = **43 sp.**

First colonization = **80 My?**



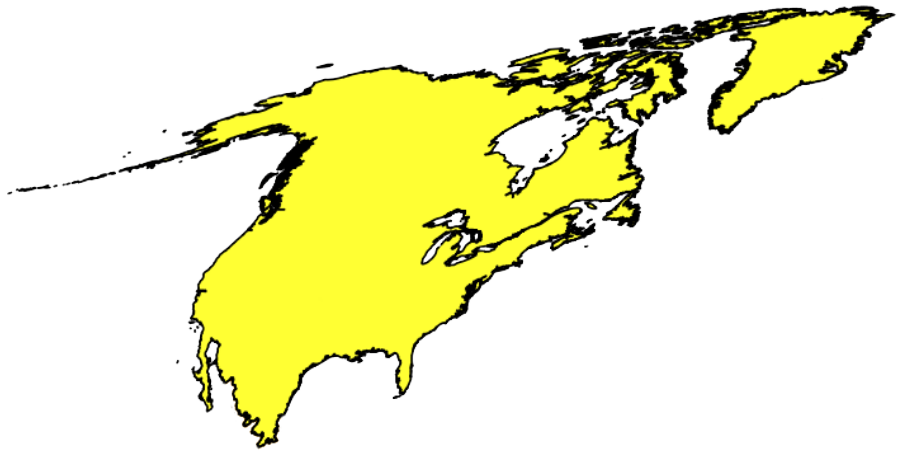

Doré et al., *in prep.*

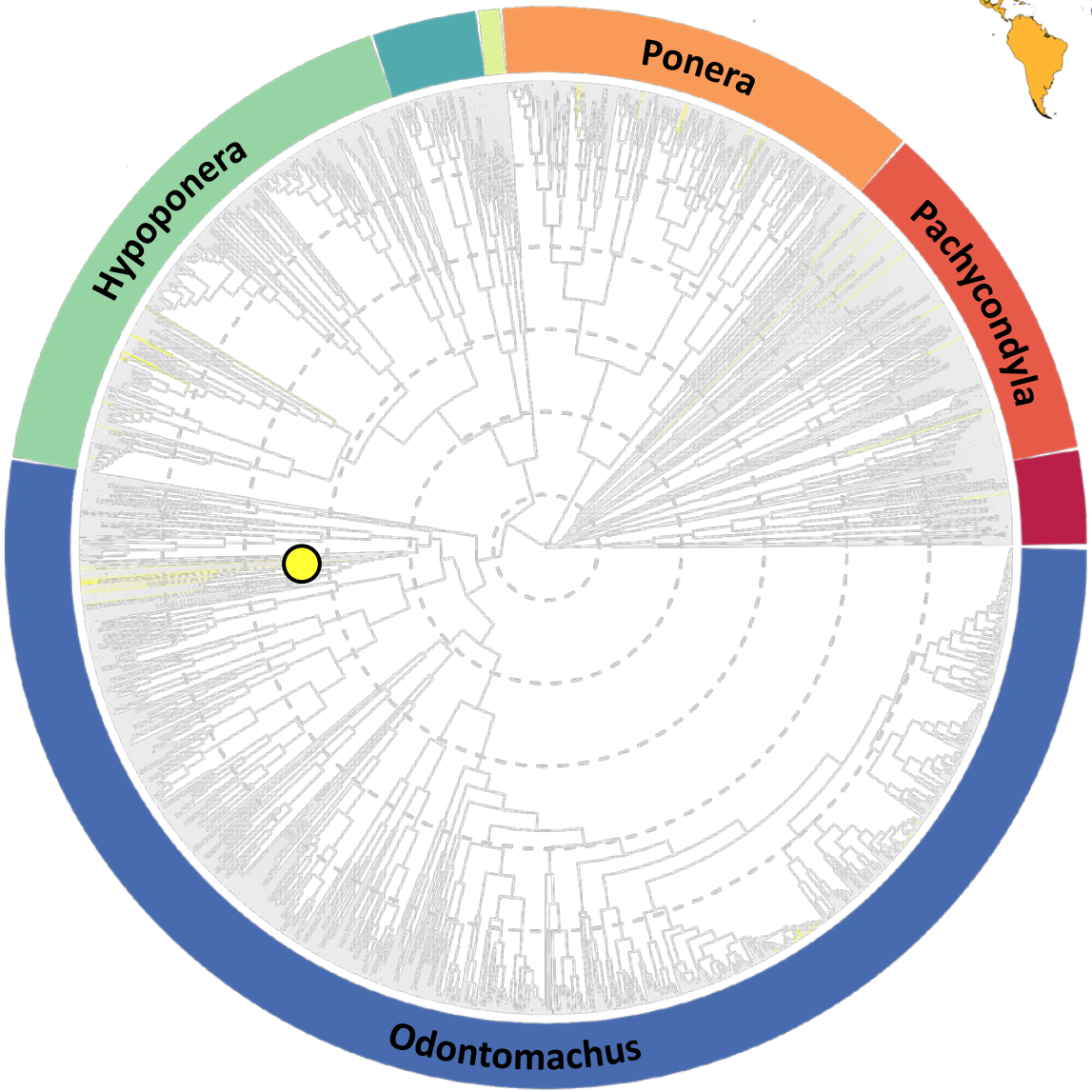
# Origins in bioregions

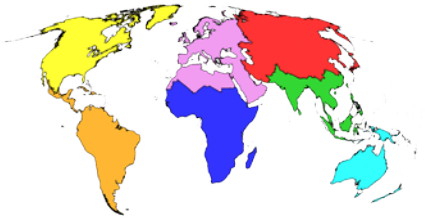
## Nearctic

Current richness = **25 sp.**

First colonization = **55 My?**





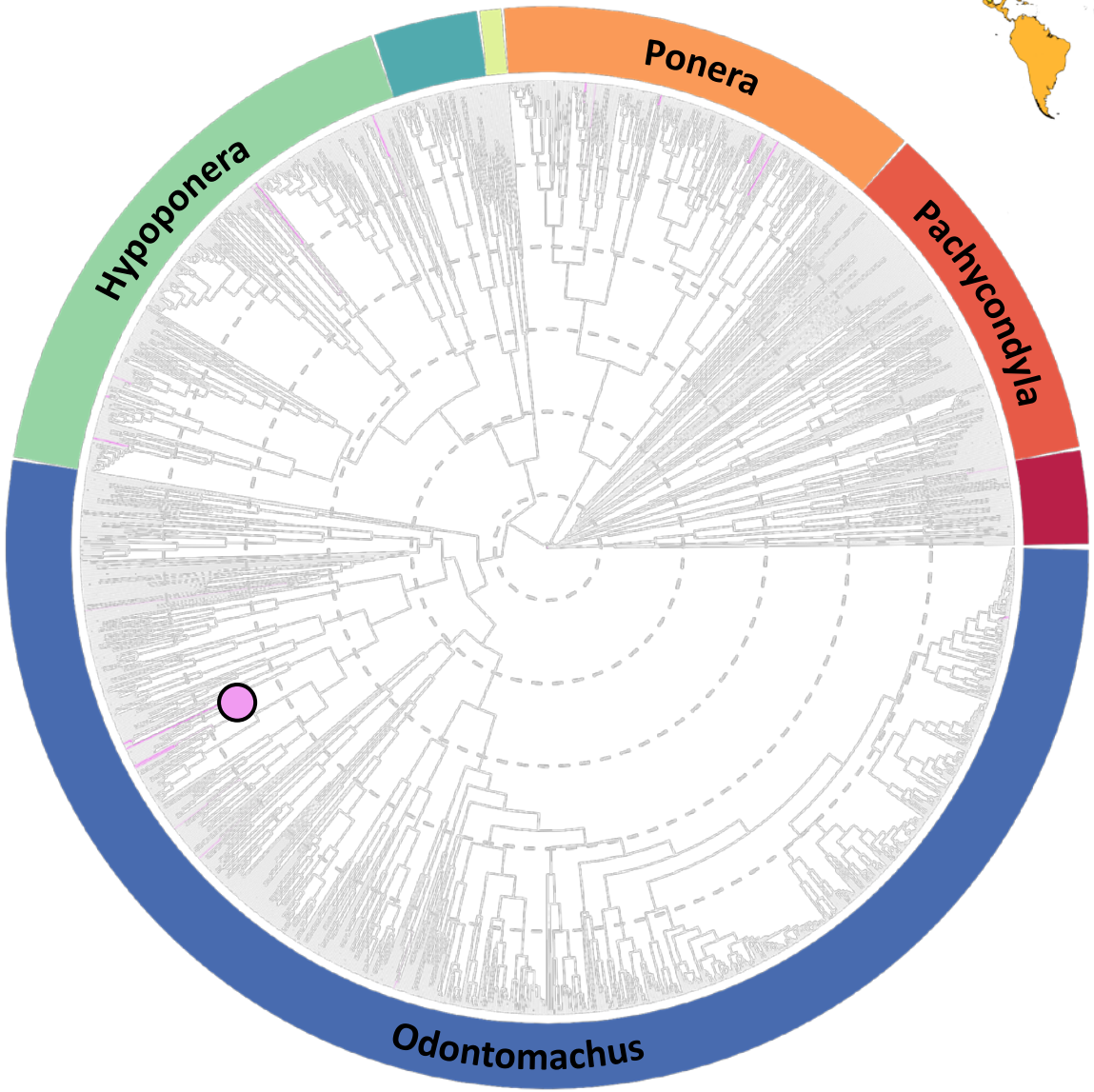
Doré et al., *in prep.*

# Origins in bioregions

## Western Palearctic

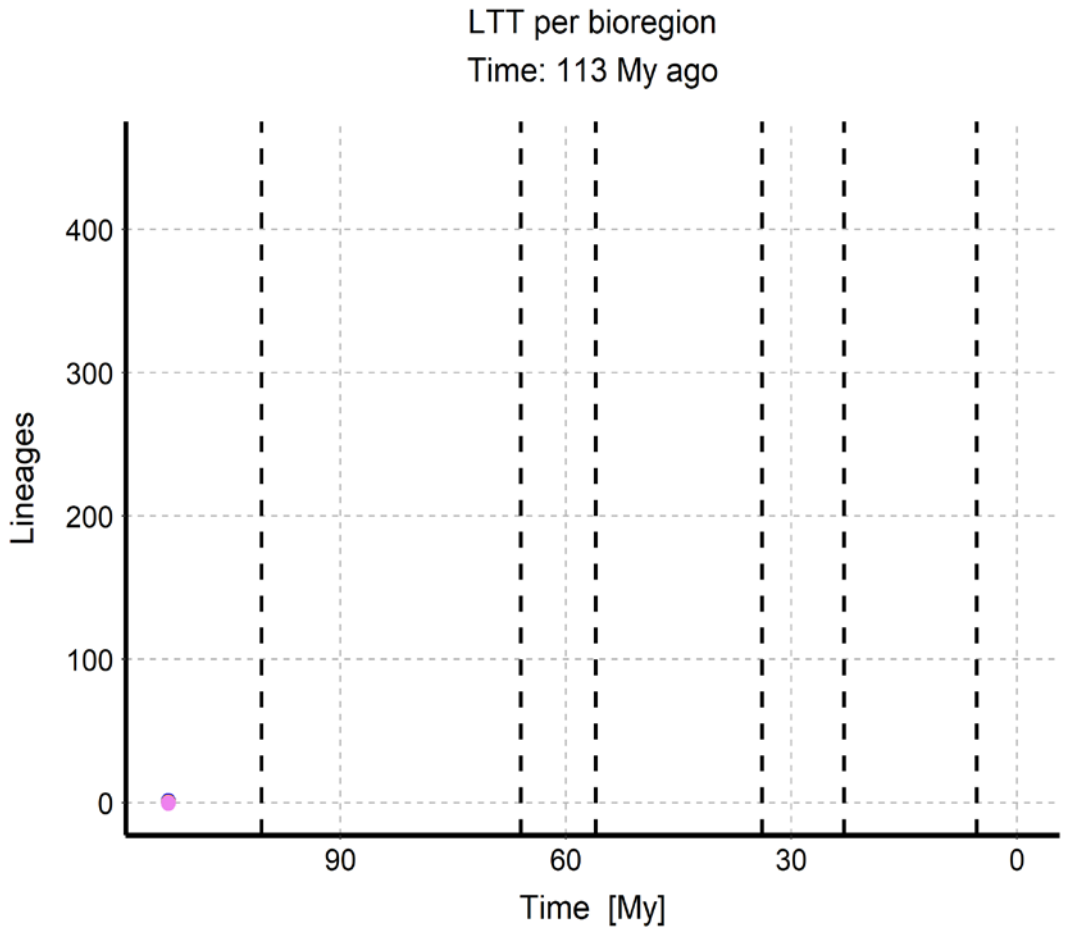
Current richness = **17 sp.**

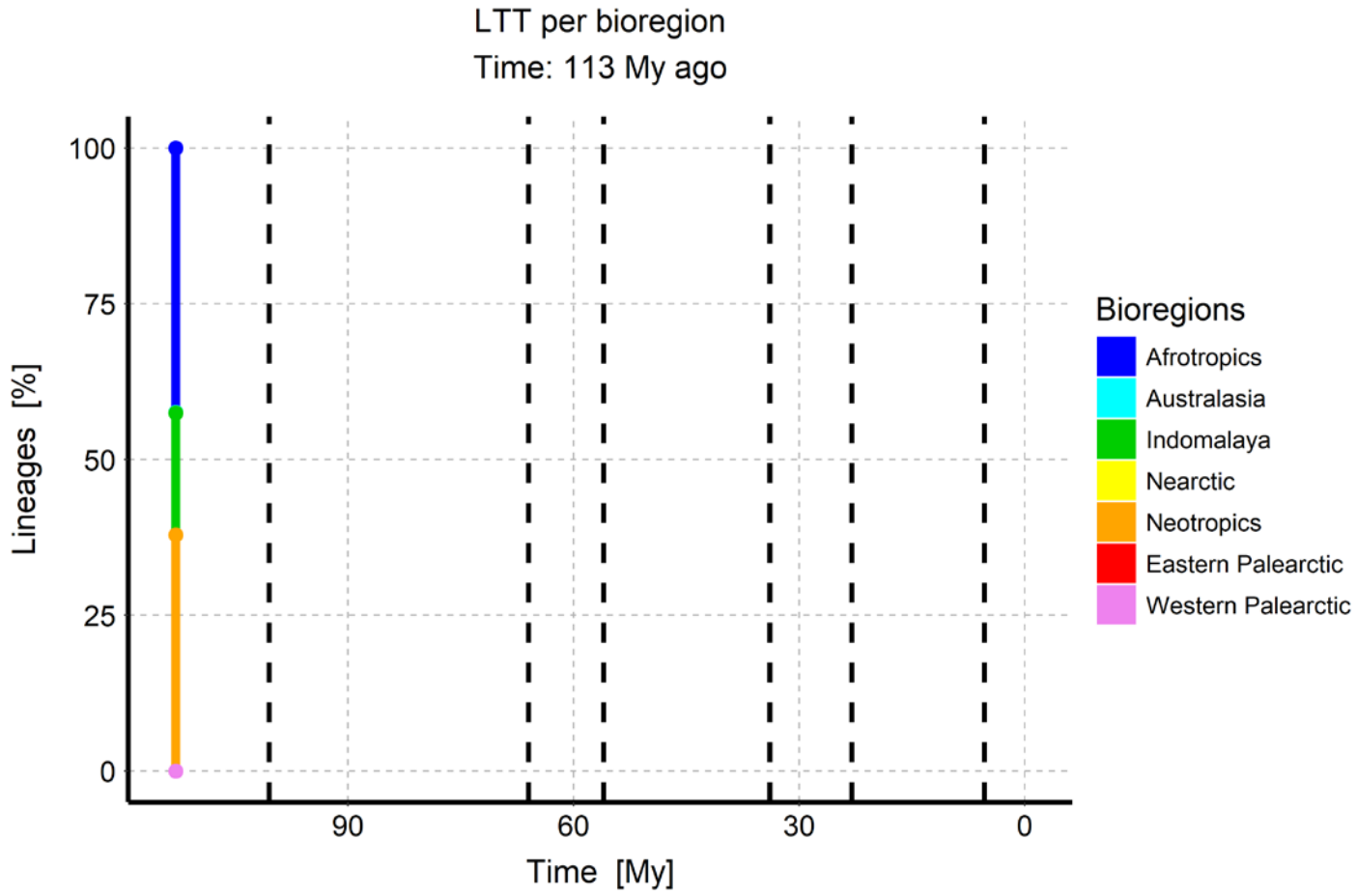
First colonization = **35 My?**




Doré et al., *in prep.*

# Lineages through time

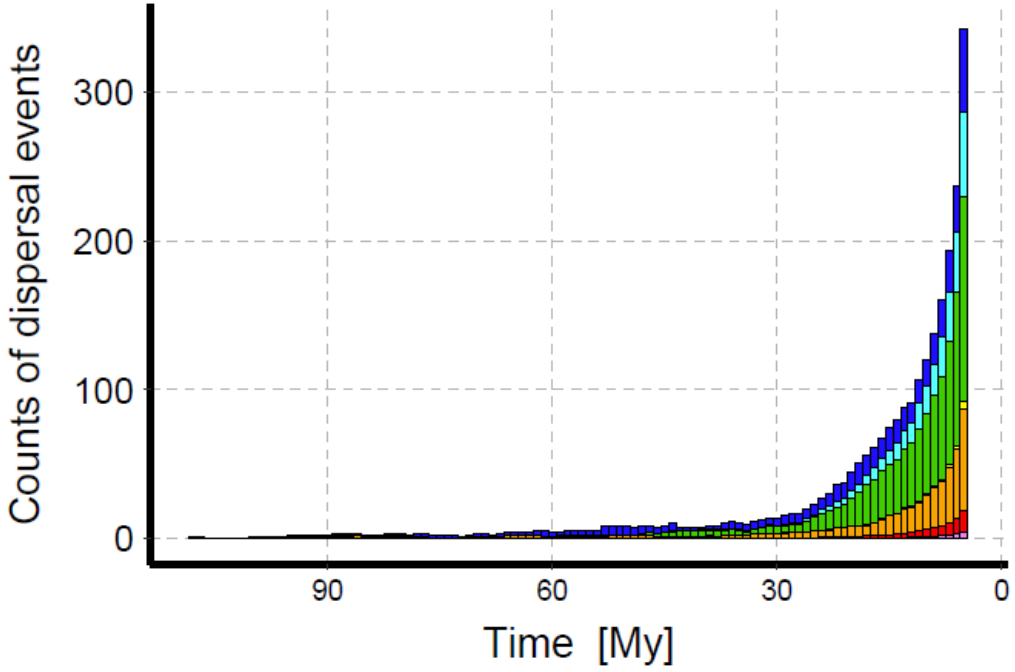


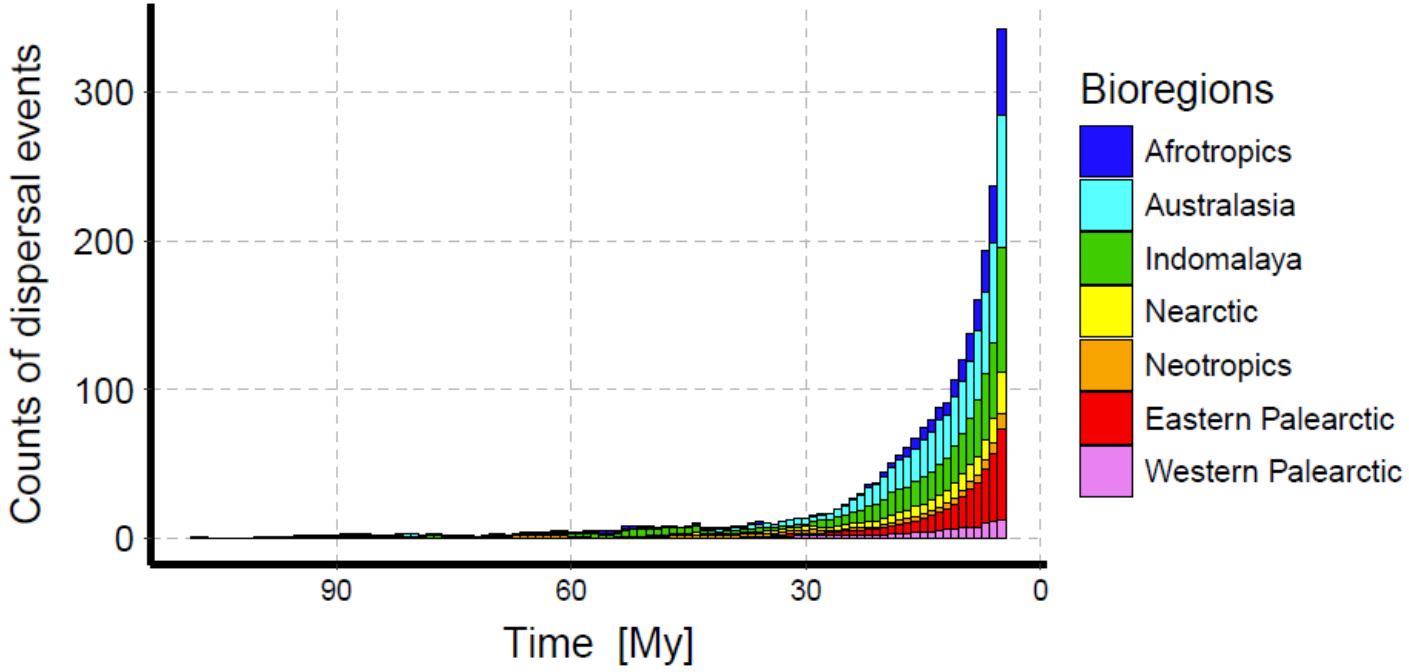

Doré et al., *in prep.*

# Dispersal events in time

Dispersal events
per Source bioregions

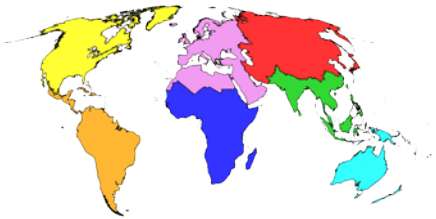



Dispersal events
per Destination bioregions

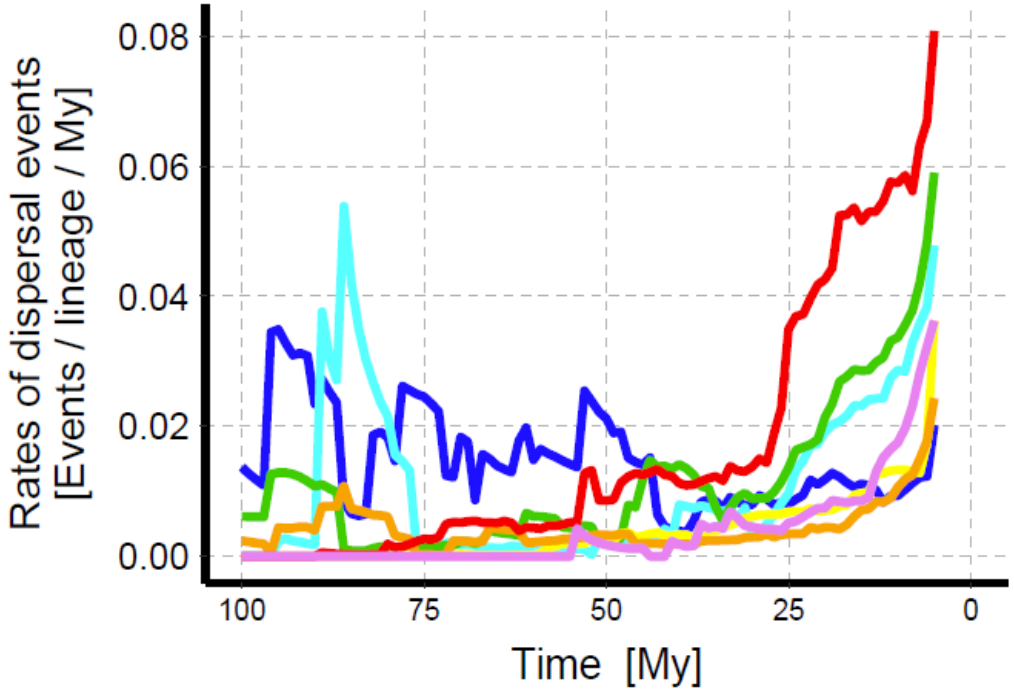



Doré et al., *in prep.*

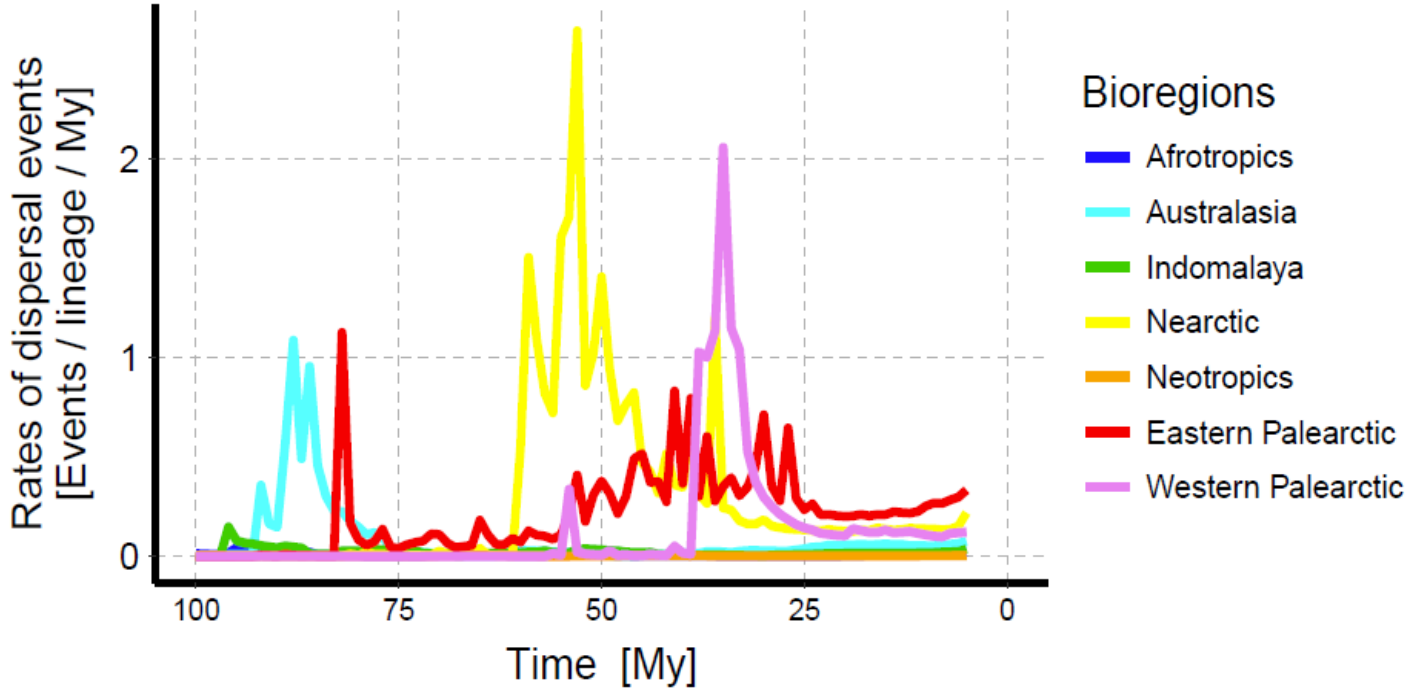
# Dispersal events in time



Rates of dispersal events
per Source bioregions



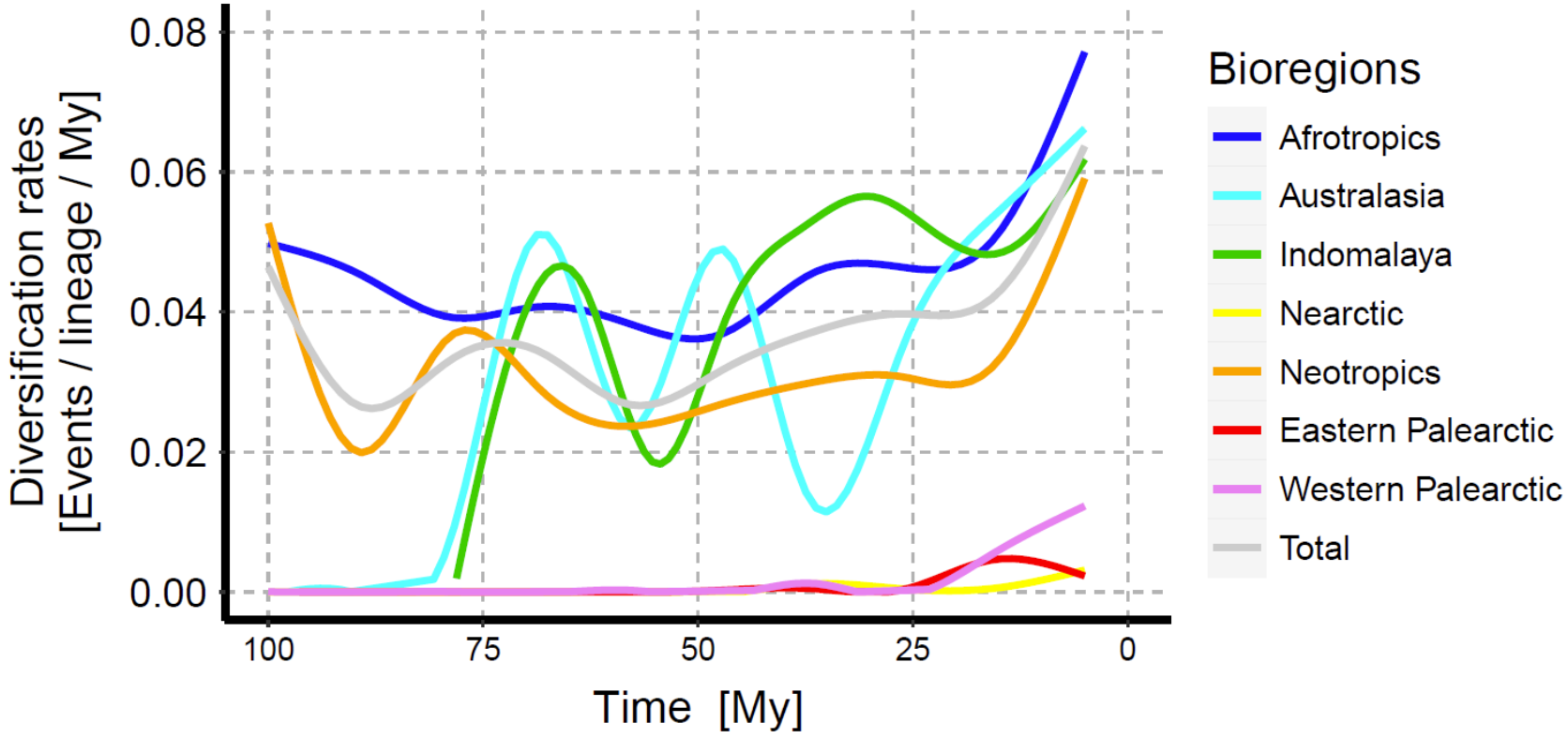

Rates of dispersal events
per Destination bioregions




Doré et al., *in prep.*

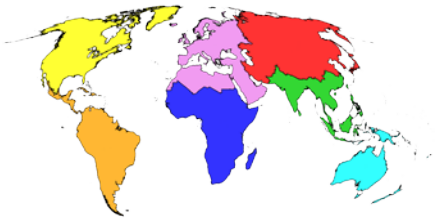
# *In situ* diversification in time

Rates of *in situ* speciation events
per Bioregions

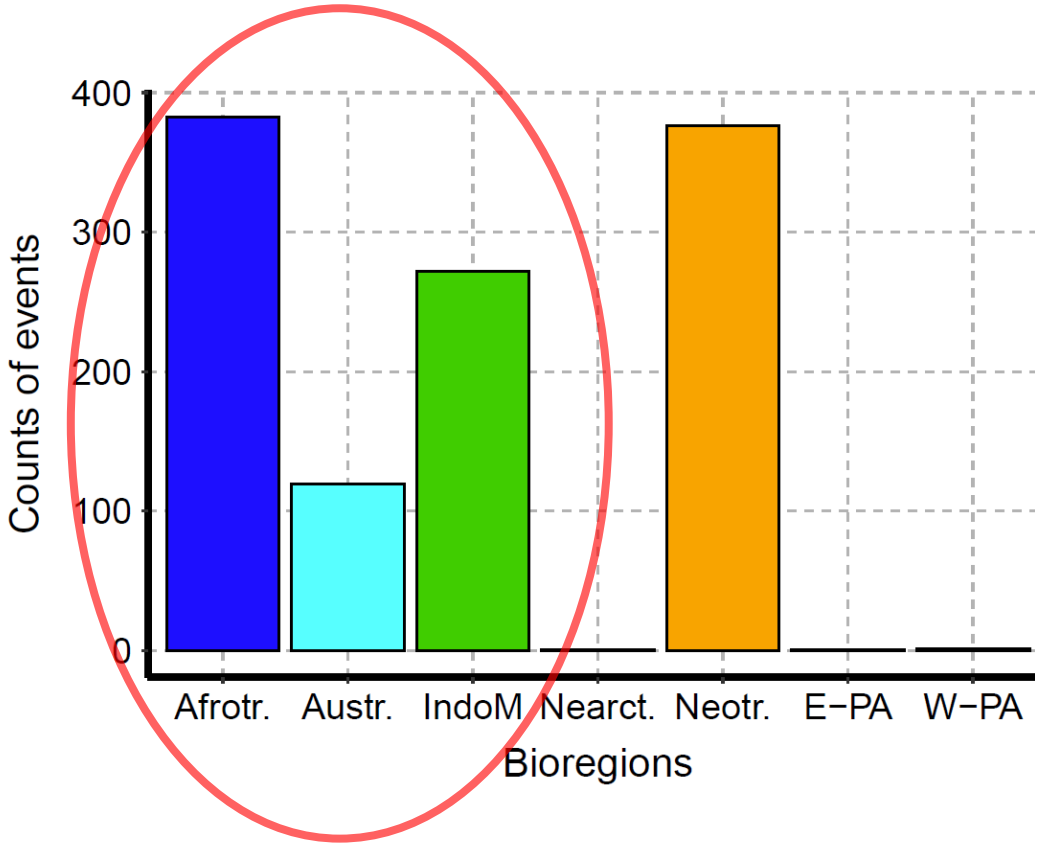



Doré et al., *in prep.*

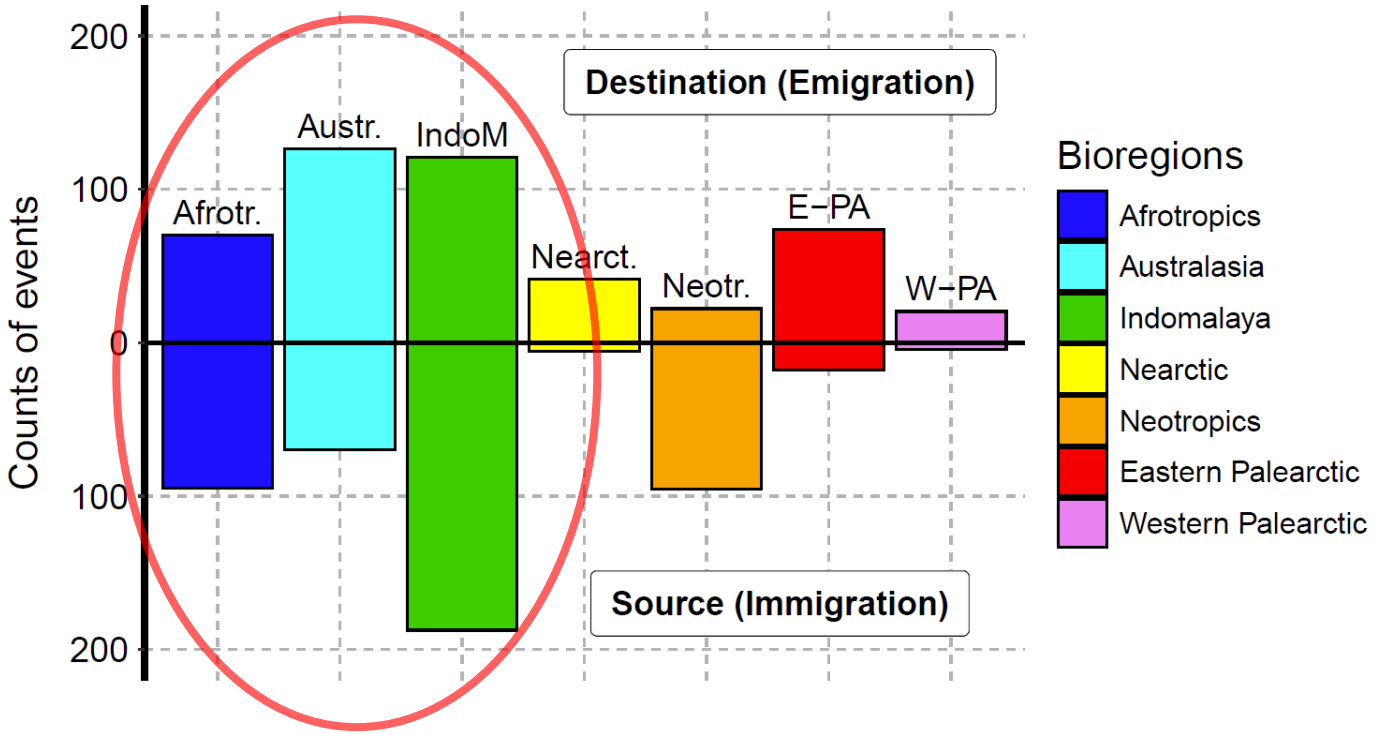
# Origin of diversity



*In situ* diversification per Bioregions

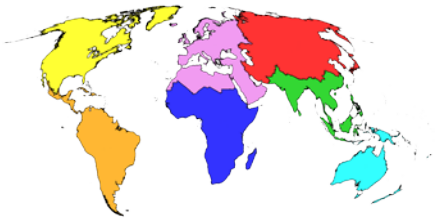



Dispersal events per Bioregions

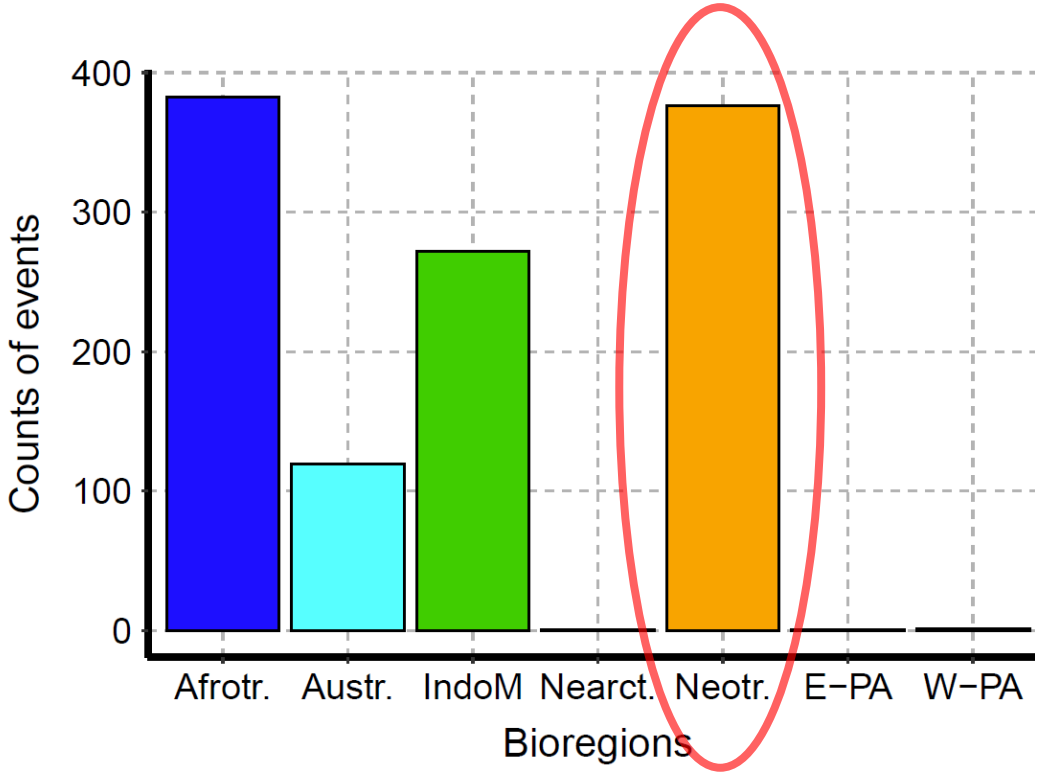



Doré et al., *in prep.*

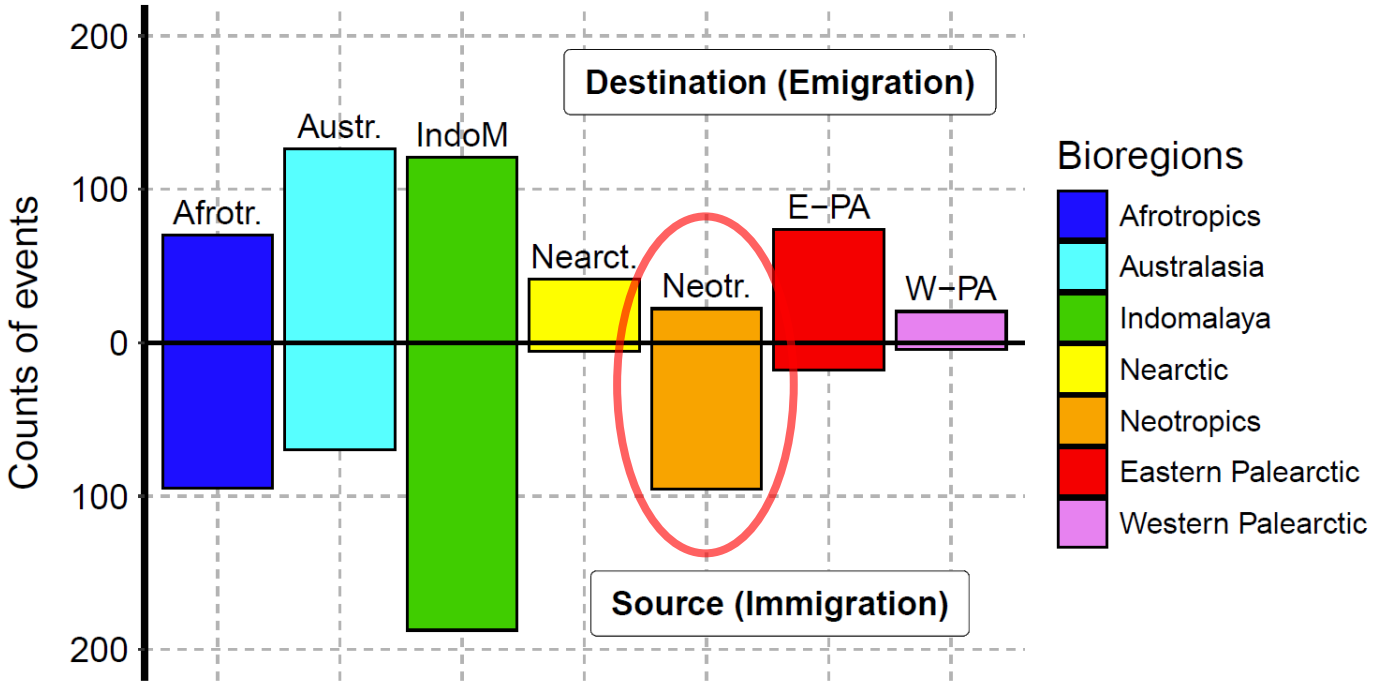
# Origin of diversity



*In situ* diversification per Bioregions

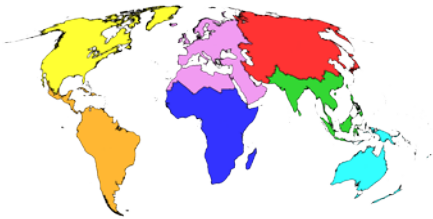



Dispersal events per Bioregions

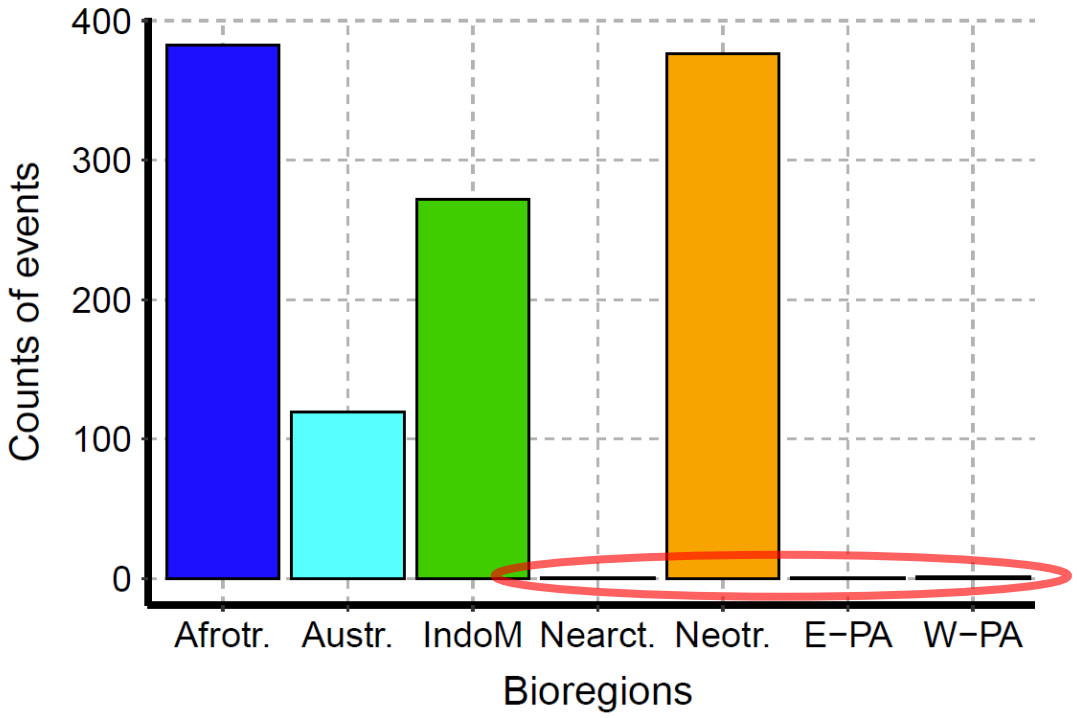



Doré et al., *in prep.*

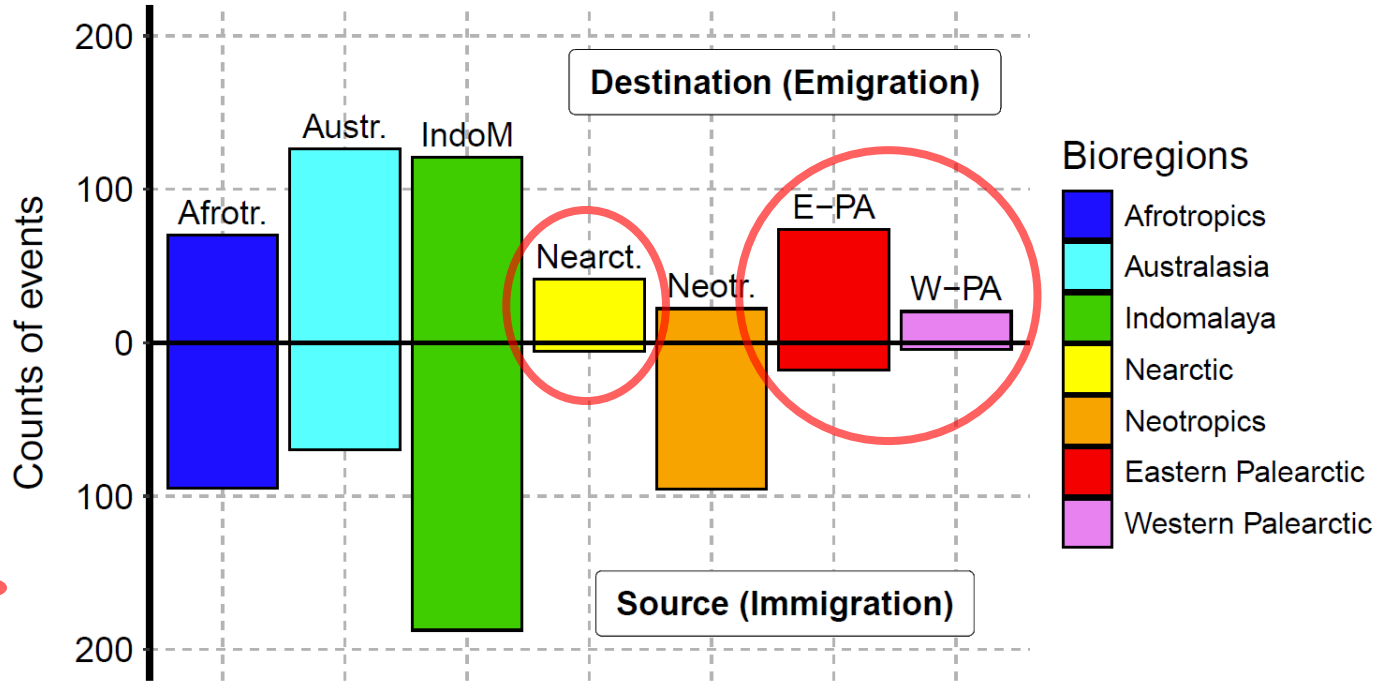
# Origin of diversity



*In situ* diversification per Bioregions

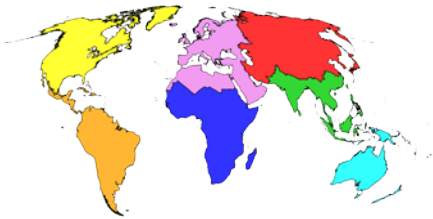



Dispersal events per Bioregions



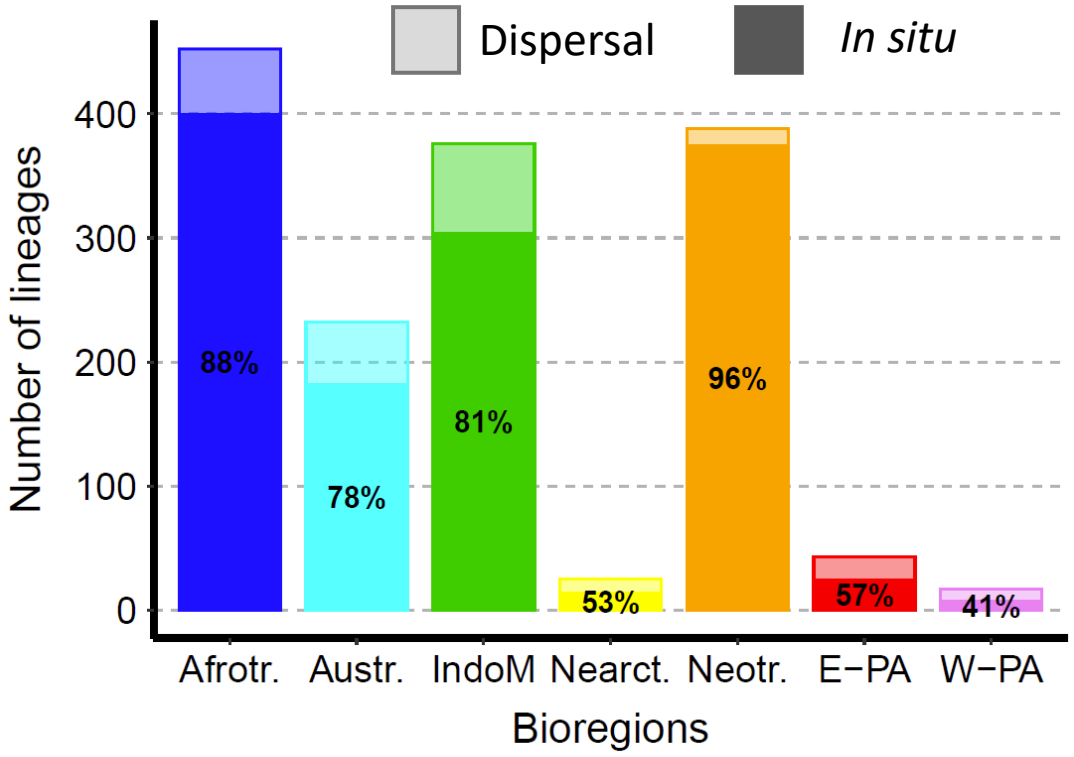

Doré et al., *in prep.*

# Origin of diversity

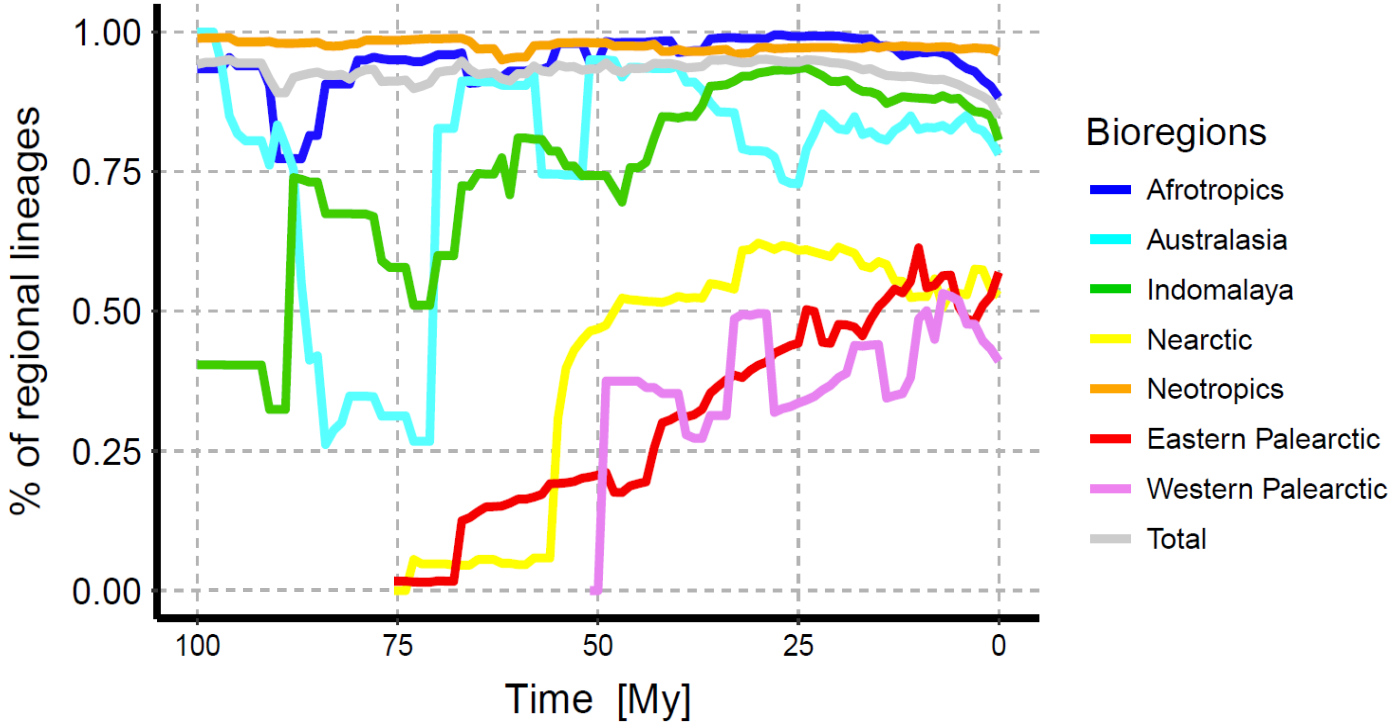


Lineage origins per Bioregions

## Now

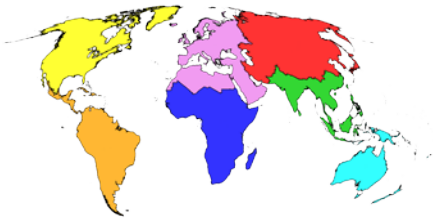



## Across time



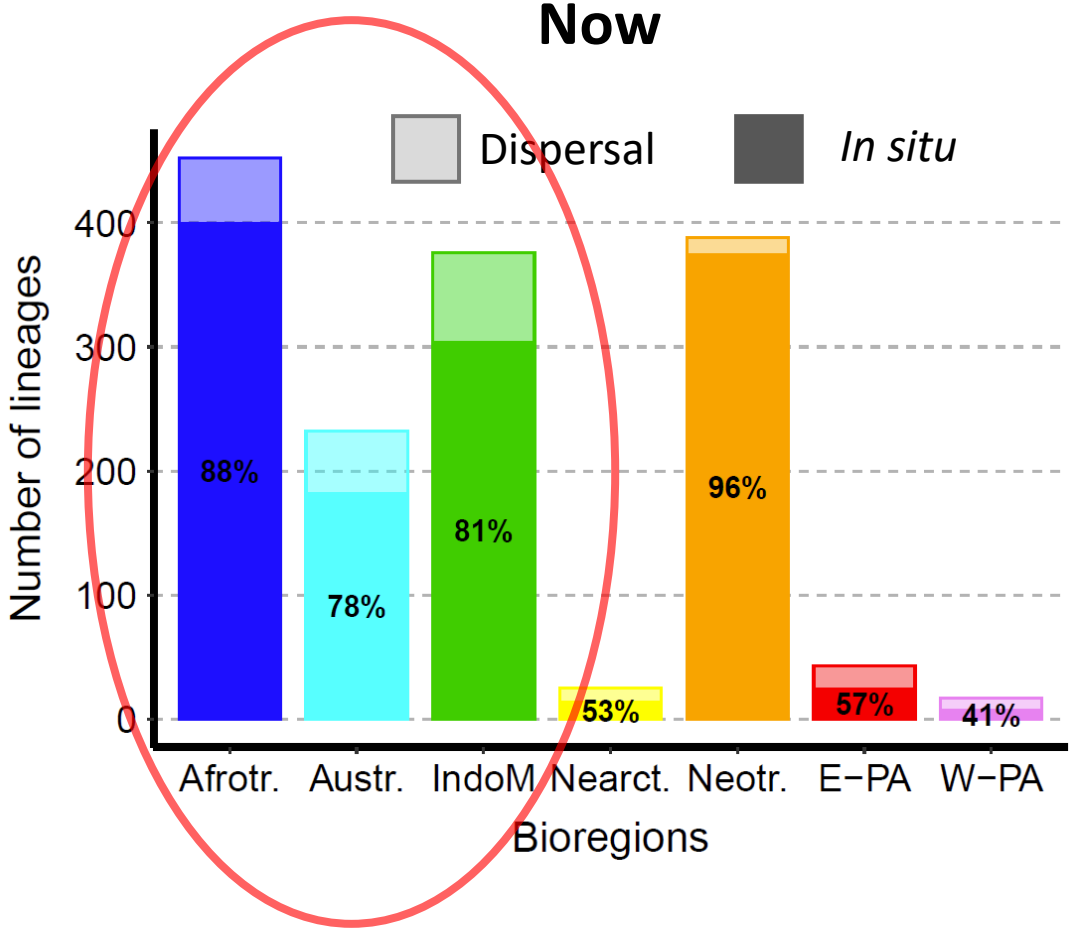

Doré et al., *in prep.*

# Origin of diversity

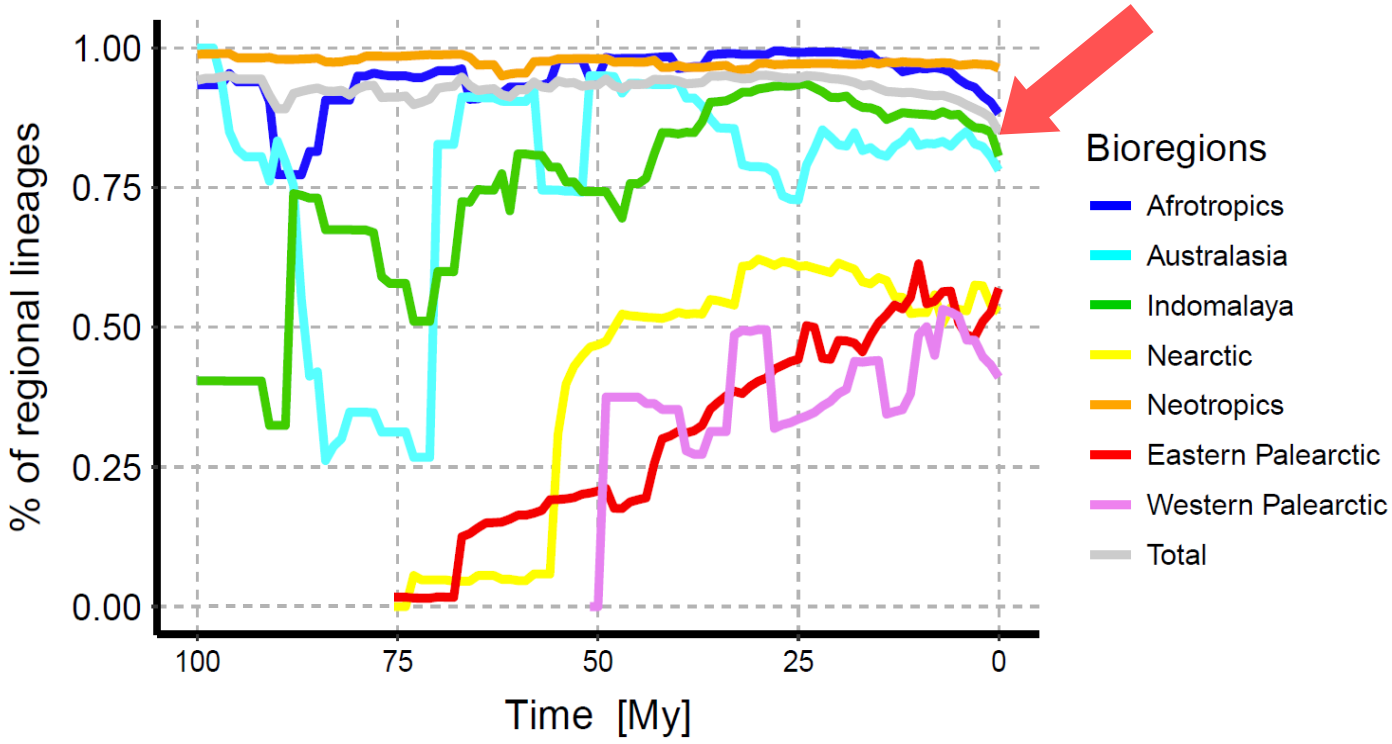


Lineage origins per Bioregions

## Now

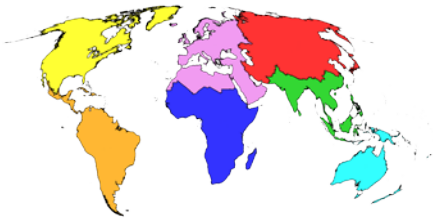



## Across time



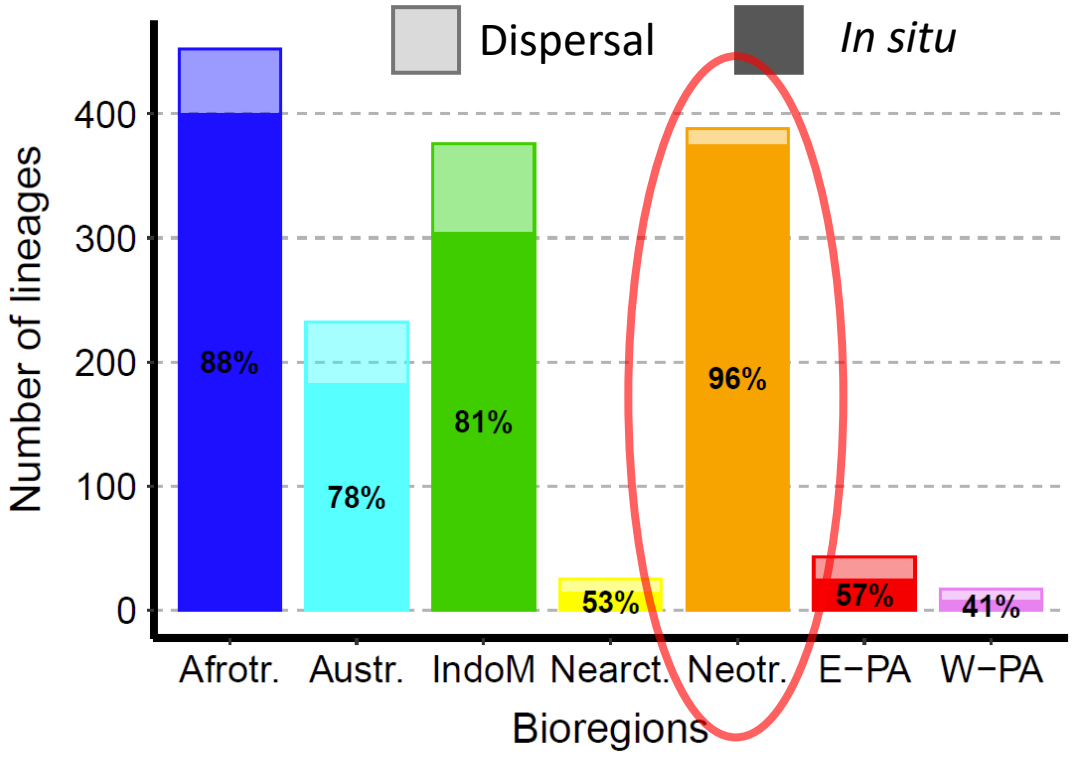

Doré et al., *in prep.*

# Origin of diversity

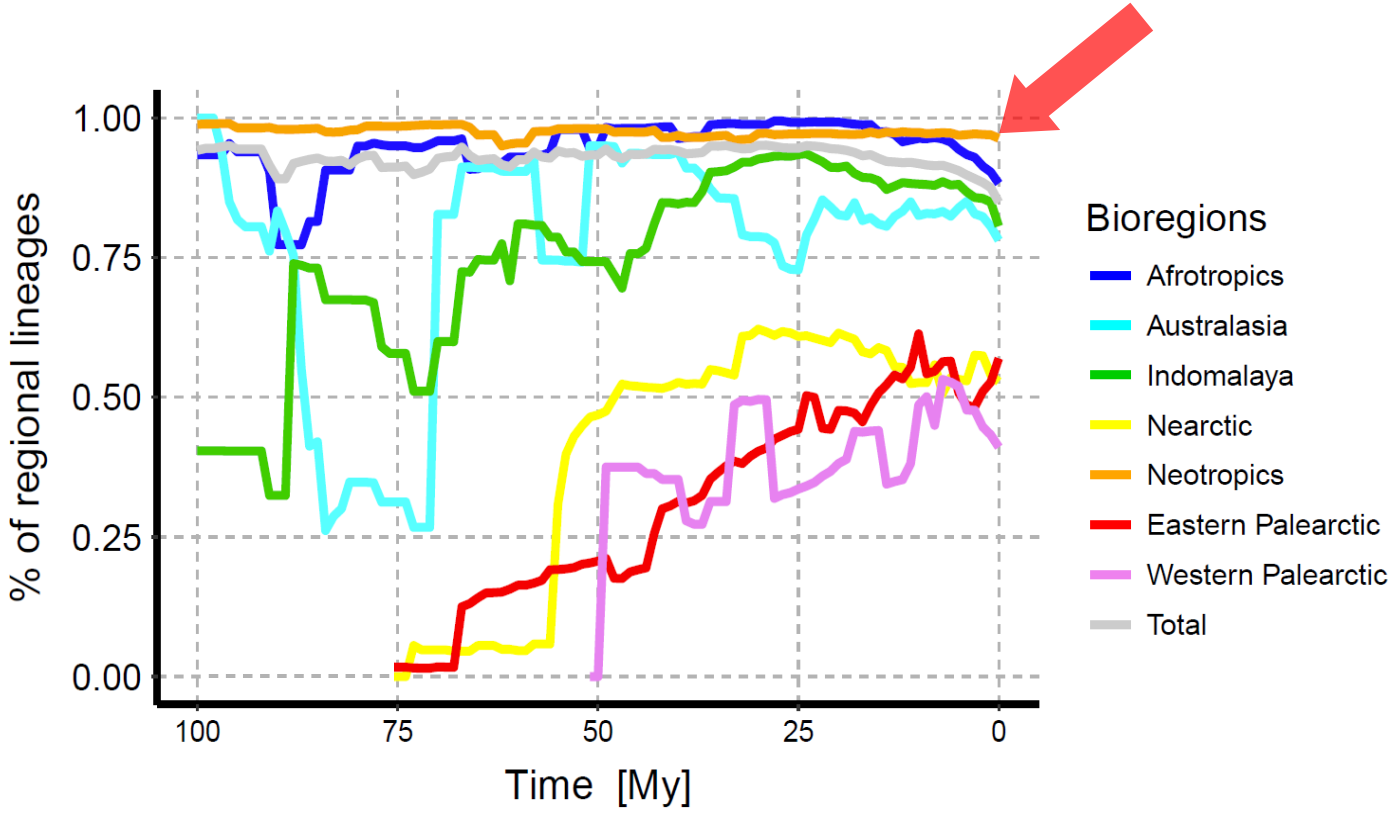


Lineage origins per Bioregions

## Now

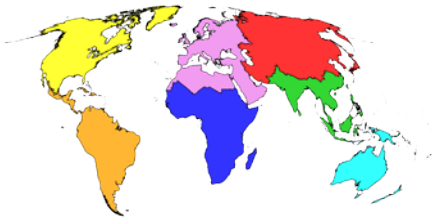



## Across time



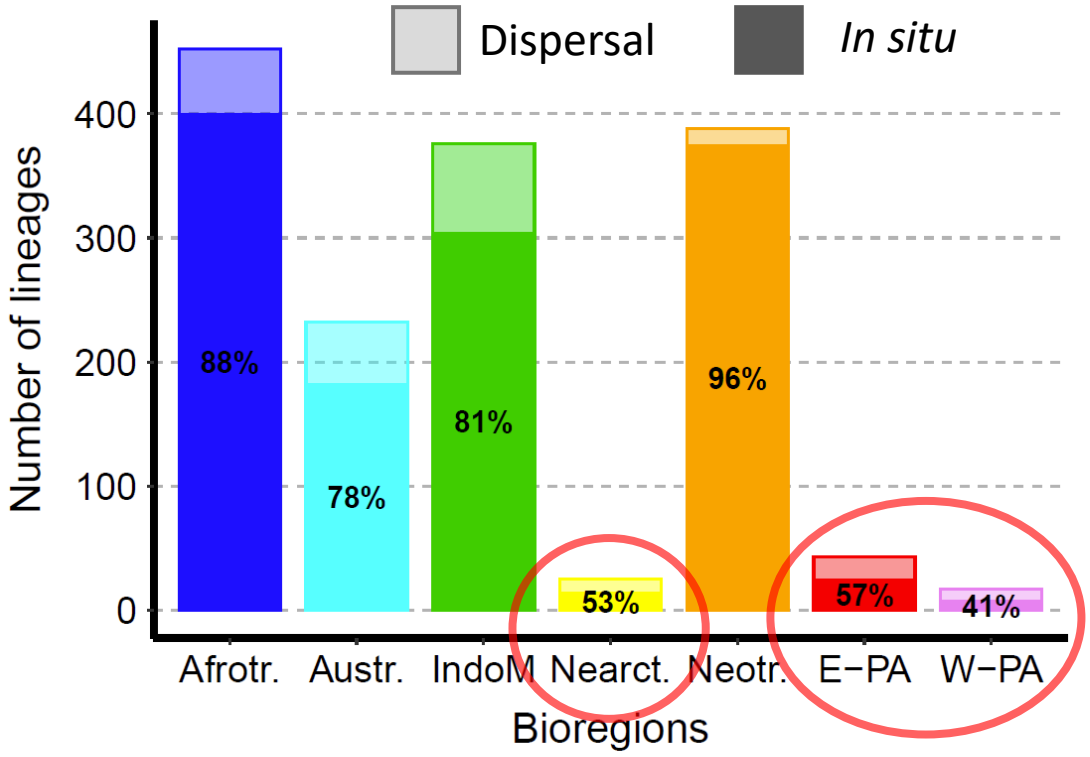

Doré et al., *in prep.*

# Origin of diversity

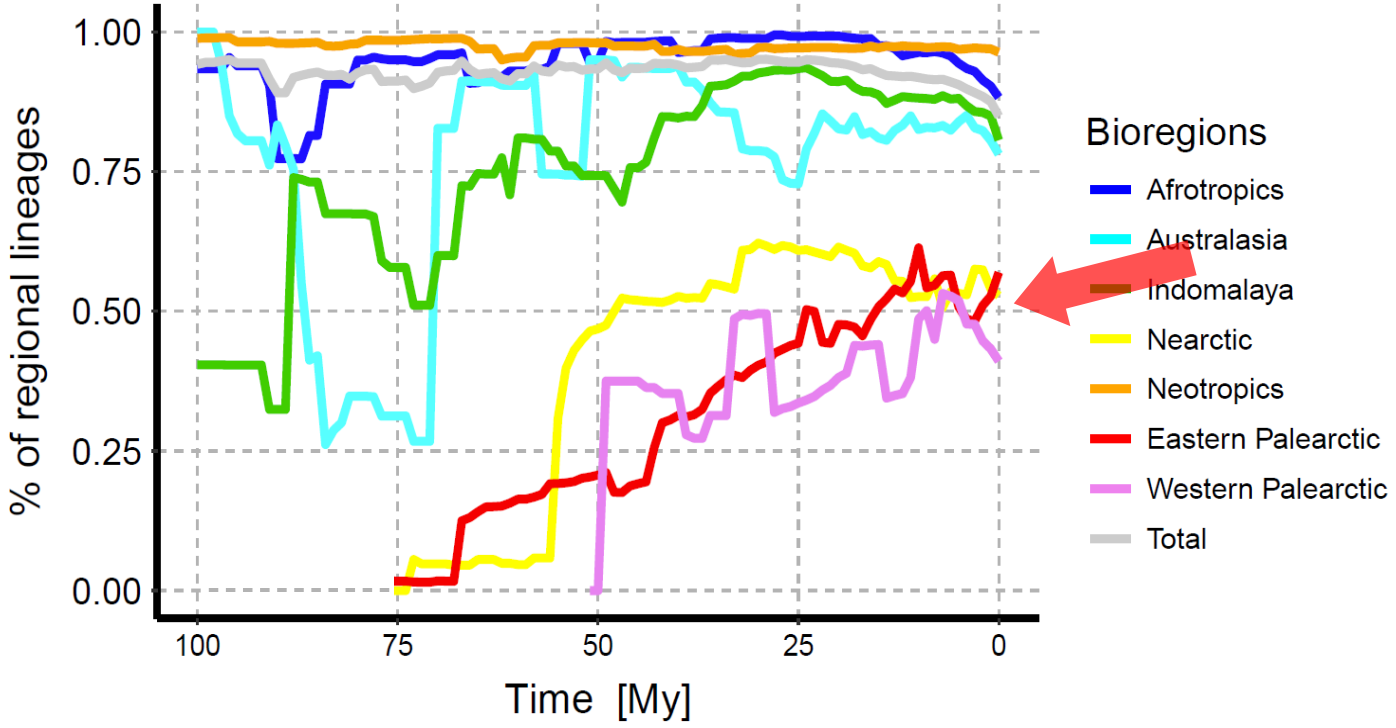


Lineage origins per Bioregions

## Now




## Across time



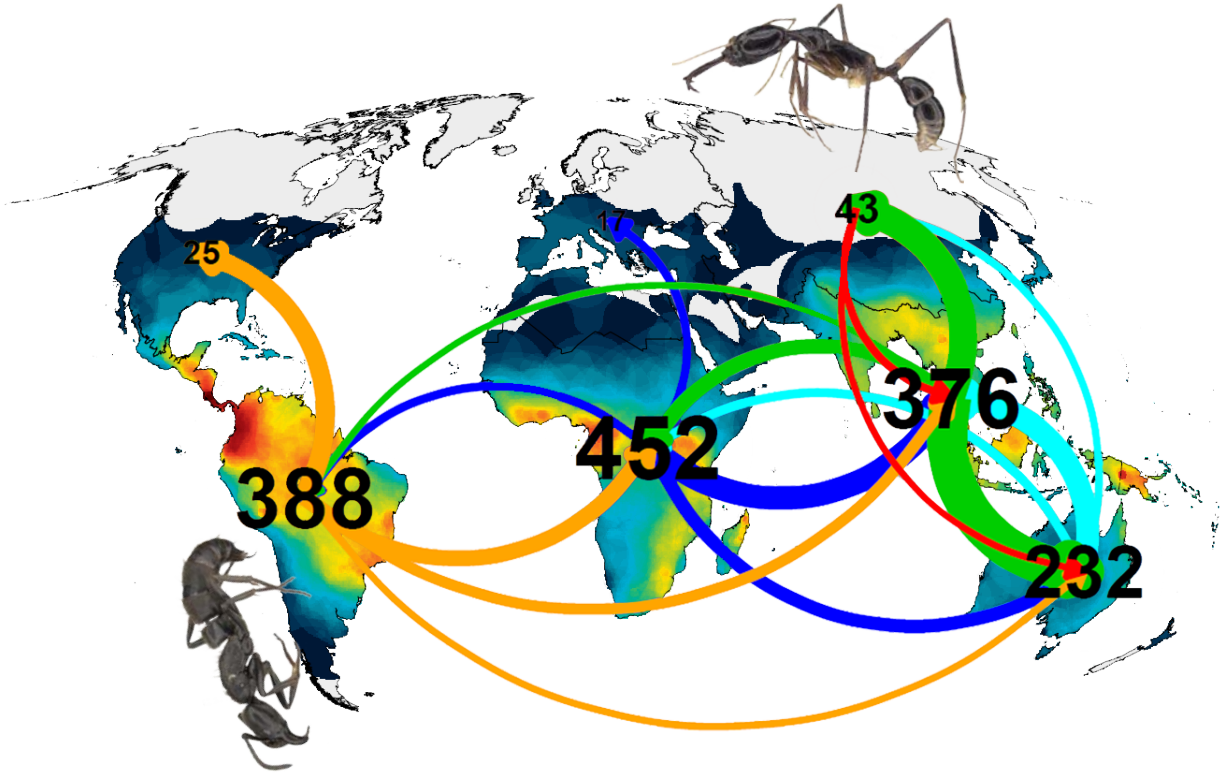

Doré et al., *in prep.*

# Conclusions

Origin in **Gondwana** ≈ 110-120 My

Timing of **first colonizations**:

- **Eastern Palearctic** from (Indo)Malaya = **80 My**
- **Nearctic** from Neotropics = **55 My**
- **Western Palearctic** from Afrotropics = **35 My**




**Dispersal** dynamics:

- Early sources: **Neotropics** + **Afrotropics**
- Overall/Recent source: **IndoMalaya**

**Diversification** dynamics:

- Early source: **Neotropics**
- Overall/Recent sources: **Afrotropics** + **IndoMalaya**

Doré et al., *in prep.*