**Biogeographic Analyses**

To evaluate the biogeographic history of Caesalpinia + Cassiae group, and in more details for *Chamaecrista*, we employed maximum-likelihood inference of geographic range using the Dispersal, Extinction, and Cladogenesis model (DEC) (Ree et al. 2005; Ree and Smith 2008) implemented in R using the "BiogeoBears" package (Matzke, XXXX)

To study geographic range evolution through time, we evaluated alternative biogeographical hypotheses by modeling area connectivity in three ways (Figure 2). *Model A*: We defined a simple model with no dispersal constraint (with no information based on paleogeography). According to this model, the rate of dispersal had a value of 1 (the highest dispersal rate) between all areas along the whole diversification period of *Chamaecrista*. Additionally, we defined two alternative hypotheses (*Models B* and *C*) with spatio-temporal dispersal constraints reflecting the likely paleogeographic history of tropical America from the Eocene onward (Gottesberger & Gottesberger 2006, Pennington et al. 2006, Antonelli et al. 2009) adopting the time-stratified parameters. *Model B*: First, we defined three time frames: time frame 1 (53.9 – 28 mya, Eocene and early Oligocene), time frame 2 (28 – 10 mya, Oligocene and late Miocene), and time frame 3 (10 mya – until to present, from Oligocene and all Quaternary) (Figure 2). For each time frame we applied different rates of dispersal between areas. For the time frame 1 we considered a low dispersal rate from Amazon Region (area A) to the other biogeographical areas (rate=0.1), based mainly on the putative lack of savannas (areas B and C), restingas (area E) and of a connection between South America and the rest of Americas (Panama isthmus not yet established, area G). On the other hand, dispersal between Amazon (area A) and Paranaense Subregion (area D) was probably facilitated by the existence of larger extensions of forests during that period, so we accept the value 1 only between these two areas. In the time frame 2, we considered the arising of Neotropical savannas defined as a higher dispersal rate between Amazon and savannas (rate=1). At the same time dispersal from the Amazon to the Paranaense Subregion would be more difficult in a putative scenario of retraction of rainforests and expansion of savannas (rate=0.1). For time frame 3 we considered a high dispersal rate between almost all areas due to the effective connection between the Americas and the establishment of the biogeographical regions within South America known today. In this case the dispersal to Central and North America from the South could be easier and the dispersal among Cerrado (area B), Caatinga (area C), Campos rupestres (area F), and Restingas (area E) could also be facilitated (rate =1). *Model C*: In this model we used the same time frame from *Model B*. The differences between *Model C* and *B* are in the time frame 1 and 2. In the time frame 1 we hypothesized that ancestral species could disperse from the Amazon region to Paranaense subregion (similar *Model B*) and also to Campos rupestres (rate=1). This hypothesis was implemented for considering that the geological history of Espinhaço Range (the main area presenting campo rupestre vegetation) is very old, probably arising in the proterozoic (Marshak et al. 2006). And in the time frame 2 we considered just dispersion from campos rupestres to savanna areas (rate=1) using a lower dispersal rate (rate=0.1) between Amazon subregion and savannas areas. The best biogeographic model was chosen using the likelihood ratio test (LRT) comparing the pairwise models.Detailed dispersal rates for each model and the scripts are available in DRYAD?. Parâmetro J.