1 **Towards a more objective inference of local and regional vegetation types using pollen assemblages**

2 **from the East Africa Rift (Ethiopia)**

3

4 Benjamin Bourel*[,1,2], Jean-Charles Mazur[2] , Doris Barboni[2,3]

5 [1] IMBE, Aix Marseille Université, Avignon Université, CNRS, IRD, Technopole Arbois, Aix-en-Provence,

6 France

7 [2] CEREGE, Aix Marseille Université, CNRS, IRD, INRAE, Coll. France, Technopole Arbois, Aix-en-

8 Provence, France

9 [3] Institut Français de Pondichéry, UMIFRE 21, UAR 3330 « Savoirs et Mondes Indiens », Pondicherry,

10 India

11 * Corresponding author: benjaminbourel1@gmail.com

12

15

16 **Abstract:** Pollen grains are particularly important for the study of Hominini sites in the Plio-Pleistocene

17 of East Africa Rift, which have long been known to be associated with riparian environments. These

18 riparian environments allow the development of a wide variety of local vegetation that differs greatly

19 from the surrounding regional vegetation. However, it is difficult to differentiate the local pollen signal

20 from the regional pollen signal. In addition, in East Africa, an important part paleo-environmental

21 reconstructions from pollen are based at least in part on empirical observations. These two elements

22 make it difficult to appreciate the relevance of paleo-environmental interpretations, mainly their

23 relevance regarding local vegetation. To help solve these problems, this study presents a proof of

24 concept based on the calculation of pollen scores to objectively discriminate local vegetation (in

25 particular riparian vegetation) from a modern data set of 283 surface pollen samples . The aim is to

26 see and discuss the effectiveness and limitations of our concept in the modern data before potentially

27 applying it to fossil data. Our approach allows the identification of 24 types and sub-types of

28 vegetation, nine of which are local riparian and azonal vegetation. It differs from previous pollen

29 scoring approaches in the spatial precision of the reconstruction of vegetation (lower than the regional

30 scale), the integration of riparian vegetation, and further treatment of aquatic plant pollen. We

31 highlight that relative abundance is not the best way to deal with pollen data and the importance of

32 treating pollen data differently depending on the conditions of deposit: relative abundance scale for

33 non-dynamic deposition environments (e.g., soils or lake edges) and presence/absence for dynamic

34 deposition environments (e.g., river beds). The interest of this approach is not limited to the

35 determination of vegetation but also to the inference of climatic parameters, the percentage of tree

36 cover, and river morphology.

37

38 **1. Introduction**

39   Many paleontological areas in Ethiopia have preserved abundant fauna, including hominin

40 remains of Late Miocene, Pliocene, and Pleistocene age (e.g., Campisano et al., 2017; Haile-Selassie et

41 al., 2015; Johanson, 2017; Saylor et al., 2019; Villmoore et al., 2015). Ethiopia is an important place to

42 study the history of human evolution and to understand how the environment and climate may have

43 influenced it (Levin, 2015). For some paleontological areas in Afar (Ethiopia), such as the

44 *Australopithecus afarensis*-bearing Hadar Formation (Bonnefille et al., 1987), the *Ardipithecus*

45 *ramidus*-bearing Aramis member (WoldeGabriel et al., 2009), more recently the *Australopithecus*

46 *anamensis*-bearing deposits at Woranso-Mille (MRD locality) (Saylor et al., 2019), and other

47 *Australopithecus* sites such as Meshellu/Dikika (Bonnefille, 2010), pollen were extensively studied and

48 provided direct botanical evidences of the plant taxa associated with early hominids. By comparison

49 with modern pollen data obtained from surface soil samples, fossil pollen relative abundances allowed

50 to appraise the composition of past plant communities (Bonnefille et al., 1987). This comparative

51 approach is possible because the floristic composition of plant ecosystems in East Africa has been

52 relatively stable over the last 11 Ma. (Bonnefille, 2010).

53   In the 1990s, based on the work of Prentice et al. (1992) on the characterization of biomes in

54 Europe by a combination of PFTs, Jolly et al. (1998) proposed a method of reconstructing vegetation

55 and climate from pollen for Africa and the Arabian Peninsula: the Plant Functional Type (PFT) method.

56 This method is based on the allocation of one or more PFTs to the different pollen types and was

57 calibrated with 966 modern pollen samples from Africa, sub-Saharan Africa and the Arabian Peninsula.

58 It excluded pollen from aquatic plants and allowed the reconstruction of biomes *in equilibrium* with

59 climate (e.g., desert, steppe, savannah, warm mixed forest, tropical rain forest). The PFT method

60 produced consistent reconstructions of past (mid-Holocene and Pleistocene) and present climate

61 parameters (Bonnefille et al., 2004; Jolly et al., 1998; Peyron et al., 2000). However, the interpretation

62 by the PFT method of past and present biome is unfortunately not efficient in East African (Bonnefille

63 et al., 2004; Peyron et al., 2000). Vincens et al. (2006) highlight the difficulties of reconstructing, with

64 the PFT method, the mosaic of open and closed vegetation, which is the most frequently defined

65 habitat type for early hominids (Reynolds et al., 2015; Stewart, 2014).). This method has shown its

66 limitations. It is necessary to find another way to extract a more precise signal from the vegetation

67 than the biome if we want to apply it to the reconstruction of hominid paleohabitats

68   Today, for the Afar and more generally for the East Africa Rift, the PFT method has been

69 abandoned for paleo-pallynological studies and we returned to a more empirical approach (R.

70 Bonnefille, personal communication). This makes it common to see oppositions between different

71 authors on how to interpret the presence of a pollen type. This is the case of *Hyphaene* pollen which

72 are considered as a signal of edaphic forest by Barboni et al. (2019) and a signal of sub-desert steppe

73 by Bonnefille et al. (1987). In addition, the same author may, without clear justification, interpret the

74 same pollen spectra differently. For example, Pliocene pollen samples from sub-member SH-2 at Hadar

75 (sample KMa to KMi) have been interpreted as swamp "close to evergreen bushland and forest" by

76 Bonnefille et al. (1987) and swamp "close to wet/dry grassland" by Bonnefille et al. (2004). Moreover,

77 the difficulty in separating the local pollen signal from the regional pollen signal in the East Africa Rift

78 (Bonnefille et al., 1987) makes the environmental reconstructions based on pollen often unclear and

79 subject to interpretation. This is reflected in the use of terms such as "dry shrubland with varying

80 proportions of grassland, wetland and riparian forest" (Saylor et al., 2019) or "woodland-grassland

81 landscape" (Bonnefille, 2010). In addition to this problem, the comparisons between modern and fossil

82 pollen data are more often made in terms of relative abundance of taxa than in terms of

83 presence/absence as it is generally the case with paleontological (faunal) data. Working with relative

84 abundance pollen data can be a serious issue when dealing with fossil assemblages that definitely

85 underwent significant taphonomical processes. Unfortunately, this is the case for the sites presented

86 above, and more globally for the East Africa Rift sites of Late Miocene, Pliocene, and Pleistocene age

87 (Bonnefille, 1984; El-Moslimany, 1983). On the other hand, ignoring the difference of pollen relative

88 abundance resulting from a real difference in the vegetation between two sites is also not satisfactory.

89 The use of a relative abundance scale is a compromise that still needs to be tested.

90   Finally, reconstructing vegetation (biomes) *in equilibrium* with climate may provide plausible

91 paleo-climate estimates, but is a too coarse approach to capture the habitat diversity that likely made

92 the attractive places for a wide variety of animals including several hominin species as recorded by the

93 abundance and diversity of faunal taxa in this area (Faith et al., 2019; Reed et al., 2013). Although the

94 presence of regional vegetation types and sub-types with different climatic conditions is clearly

95 established by botanists in the East African Rift (Friis et al., 2010), the potential of pollen to characterize

them objectively has not been developed. In arid and semi-arid eastern Africa, surface water such as rivers, lakes and groundwater discharge areas (springs) promotes the development of micro-habitats, and strongly influences vegetation and wildlife distribution (e.g., Friis et al., 2010; Geheb and Abebe, 2003). These azonal wooded riparian vegetation associated with permanent water were particularly attractive for Hominini in the past (Barboni et al., 2019). Yet, few vegetation proxies allow reconstructing paleo-vegetation at the spatial scale of hominid sites. Phytoliths and organic compounds were used for this purpose at a few sites in Ethiopia (WoldeGabriel et al., 2009) and Tanzania (e.g., Arráiz et al., 2017; Magill et al., 2016), but their taxonomic resolution is too low to characterize paleo-vegetation beyond its physiognomy (i.e., the paleo-tree cover), or beyond indicating the presence of palms, sedges, ferns, or details within the grass paleo-communities. To our knowledge, there is no more accurate proxy than pollen to describe the composition of past vegetation. Yet, little interest has been given so far to the local vegetation (riverine and fed by springs) signals contained in pollen assemblages despite the fact that "riverine forests are represented by a long range of floristic sub-types from nearly all parts of Ethiopia, and differ in composition according to topographic and climatic factors, and the surrounding vegetation with which they more or less share similar species" (Friis et al., 2010). The potential of pollen proxy to characterize different riparian vegetation types, which typically occur as narrow and strip-like habitats along perennial and non-perennial rivers below 1800 m above mean sea level (a.s.l.) and associated with lakes and springs, has yet to be formally tested.

To solve these problems and improve our understanding of Hominini habitats, it is necessary to develop a new approach to reconstructing vegetation from pollen in East Africa. This new method must respect several elements. (1) It must be based on pollen scores without excluding aquatic plants in order to be objective and reproducible. (2) It must be able to separating the local pollen signal from the regional pollen signal in the East Africa Rift, and thus to distinguish as clearly as possible between regional (climatically driven) and local (edaphically-driven) vegetation. (3) It must minimize the use of relative abundance data and maximize the use of relative abundance scales or presence/absence data, in order to minimize taphonomic biases in future comparisons between modern and fossil pollen data to better assess past vegetation.

With this in mind, we developed a new method based on a pollen score: the vegetation score of sample β for the vegetation group α ($V_{\alpha,\beta}$ score). To see if the $V_{\alpha,\beta}$ effectively separates different regional and local vegetations , we used modern pollen assemblages. This modern pollen assemblages includes new lowlands samples from the Awash Valley and the southern extremity of the East Highlands that complement existing data. Here, we will see (1) which types of data (relative

129 abundance, relative abundance scale, and presence/absence) for the $V_{\alpha,\beta}$ score are most effective in
130 discriminating these 24 regional and local vegetation types and subtypes. (2) Whether the vegetation
131 discriminations performed with the $V_{\alpha,\beta}$ score are good enough to consider its use to assist in the
132 reconstruction of paleo-vegetation from pollen data. (3) The environmental information that can be
133 obtained with the $V_{\alpha,\beta}$ score in addition to vegetation.

134 **2. Study area**

135   This study focuses on the present-day vegetation and climate of the Horn of Africa, specifically
136 in Ethiopia, for which pollen assemblages from surface soil/mud samples are available (**Figure 1**). In
137 Ethiopia, topography, which is with the seasonal movement of the Intertropical Convergence Zone
138 (ITCZ) one of the main factors influencing precipitation and precipitation patterns, is characterized by
139 highlands generally between 1600 and 2600 m a.s.l. (up to 4500 m a.s.l.) of elevation on the western
140 and eastern sides of the Rift valley. In Central Ethiopia, the floor of the Rift valley, where large lakes
141 are located, lies at ca. 1650 m a.s.l., but falls down to -125 m below the sea level in the northeast, in
142 the Awash Valley (Danakil depression). The Western Ethiopian highlands receive maximum rainfall
143 (heavy rains coming from the west Atlantic front mainly from June to September, or almost all year
144 round in the far west of the country), while the Eastern Ethiopian highlands receive monsoon
145 precipitations coming from the Indian Ocean in two rainy seasons due to seasonal variation of the ITCZ
146 (Diro et al., 2011). Rainfall is unimodal in the west of the Rift valley and bimodal everywhere else (Diro
147 et al., 2009, 2008). Dry season, expressed here as the number of months with <30 mm rainfall lasts
148 just three months in the western regions of Amhara and Benishangul-Gumuz Regional States, but is as
149 long as 6 months in the Afar Region and Somali Regional State (Fick and Hijmans, 2017; Friis et al.,
150 2010). In Ethiopia, temperature decreases with increasing elevation according to a thermal gradient of
151 about -0.4°C/100 m. In the rift valley, mean annual temperature is about 30°-34°C in the northernmost
152 area of the rift valley (Danakil Depression), and about 20°-23°C in the Great Rift Lakes Region. Mean
153 annual temperature in Ethiopia is as low as 6-7°C above 4100 m a.s.l. where freezing is common.

154   Vegetation composition and physiognomy are largely influenced by precipitation and
155 temperature. Friis et al. (2010) describe regional vegetation types and vegetation sub-types for
156 Ethiopia based on dominant species, vegetation physiognomy, and environmental parameters (e.g.,
157 soil, climate, elevation). A georeferenced map of vegetation types and sub-types over Ethiopia and
158 East Africa (with homogenized names and abbreviations) was produced under the Vegetation and
159 Climate change in East Africa (VECEA) project (van Breugel et al., 2015), and has been used here.
160 Vegetation distribution can be summarized as follows. Afroalpine vegetation (A) including Ericaceous
161 zones (E) occur at high elevation (>3000 m a.s.l.) where mean annual temperature ranges between ca.

162   15° and 6°C (**Figure 2**). A complex of Afromontane forest with wooded grasslands and evergreen or

163   semi-evergreen bushland and thicket (Fb/Wd) occurs below the Afroalpine A and E zones and covers

164   most of the Western and Eastern Ethiopian highlands where mean annual temperature ranges is ca

165   22.5-12.5°C between ca 1800-3000 m a.s.l. (but upper and lower limits can vary). Afromontane

166   rainforest (Fa) are restricted to the south-western Highlands (Illubator and Kefa regions), where mean

167   annual temperature ranges between ca. 12.5° to 20°C, and mean annual precipitations are of ca 1800-

168   2000 mm/yr. At low and mid-elevation (ca. 400-1800 m a.s.l.), *Combretum-Terminalia* woodland and

169   wooded grassland (Wcd) (with small to moderate sized trees with large deciduous leaves) occur in the

170   west where mean annual precipitation is ca. 600-2000 mm/yr, rainfall is unimodal with a dry season

171   of 3 to 6 months, and mean annual temperature ranges from ca. 18°C to 28°C. At similar altitude (ca.

172   400-1800 m a.s.l.) and comparable mean annual temperature range (ca. 18°-30°C) but with lower

173   annual precipitations (<1000 mm/yr) and a bimodal rainfall regime, the eastern and southern regions

174   are covered with *Acacia-Commiphora* woodlands and bushlands (Bd). In the central part of the Rift

175   Valley below 1800 m a.s.l., where mean annual temperature is ca. 15°-24°C, rainfall is also bimodal but

176   higher (600-1000 mm/yr), the vegetation is less xerophytic and is described as Upland Acacia wooded

177   grassland (We). Desert (D) and semi-desert grassland and shrubland (S) occur in northernmost area of

178   Ethiopia rift valley and eastern Ethiopia below ca 400 m a.s.l., where mean annual temperature is ca.

179   25°-34°C. The mean annual precipitations is < 200 mm/yr for D and generally is 200-400 mm/yr for S

180   but can but can be on a few S sites up to 1150 mm/yr. In this most arid land, small trees, shrubs and

181   herbs may be succulent, geophytic or annual. We note that boundaries of main vegetation types match

182   the bioclimatic boundaries except in south-western Ethiopia, where the *Acacia-Commiphora*

183   woodland and bushland (Bd) (in the east) meets with *Combretum-Terminalia* woodland and wooded

184   grassland (Wcd) (in the west) (Friis et al. 2010) (**Figure 2**). Edaphic vegetation associated with rivers,

185   freshwater lakes and salty lakes was also studied in detail and mapped by Friis et al. (2010), who

186   distinguished up to seven types and sub-types. They noted that riverine vegetation is quite similar to

187   vegetation in the major type above 1800 m a.s.l., but shows very different floristic composition from

188   the surrounding vegetation below 1800 m a.s.l. (Friis et al., 2010, p. 42). Riverine habitats below 1800

189   m a.s.l. may contain species found in forests under higher rainfall regimes at the same elevations.

190   Floristic composition of riverine vegetation was estimated at 242 species. 26% of these (including 60

191   woody species) are exclusively or predominantly recorded in this habitat (Friis et al., 2010, p. 127).

3. Material and method

3.1. Modern pollen samples

Pollen names were homogenized according to Vincens et al. (2007) (**Table S1**). Pollen types difficult to differentiate in the modern samples and rarely differentiated in the fossil register were grouped together (e.g., *Nuxia/Dobera*). Taxonomic precision was limited at the subfamily level for Asteraceae. Spores were excluded from our analyses because it is not clear if they were consistently identified and counted by all analysts. *Prosopis*-type and *Prosopis*-type *africana*, which exhibit little morphological differences were also removed from the dataset after the calculation of relative abundance, because *Prosopis*-type is produced by the recently introduced and very invasive species *Prosopis chilensis.*

This study includes 283 modern pollen assemblages (**Table S2**) obtained from 207 soil samples and 76 mud samples from Ethiopia (251 assemblages), the Arabian Peninsula (17 assemblages from Yemen, Saudi Arabia, and Qatar) and Tanzania (15 assemblages) (**Figure 1** and **Table S3**). Out of these 283 assemblages, 182 assemblages were extracted from the APD (African Pollen Database, 2019), 93 assemblages were extracted from the literature, and 62 assemblages (22% of the dataset) are new. This dataset samples most of the climate domain of Ethiopia, except where annual temperature is above 30°C (cf., the Danakil Depression) (**Figure 2** and **Table S4**). The 15 samples from Tanzania were added to improve representation of spring-associated wetlands and forests (Barboni et al., 2019). The 17 samples from the Arabian Peninsula (Bonnefille and Riollet, 1988; El-Moslimany, 1983; Lézine et al., 1998) were included to represent very arid conditions (annual precipitation < 125 mm/yr) that are not currently found in the modern Ethiopian Rift Valley.

The 62 new assemblages were collected in 2009 by D. Barboni and R. Bonnefille (Ethiopian samples DB09) and in 2017 by D. Barboni, B. Bourel and J.-C. Mazur (Ethiopian samples BB17). Our new samples fill several gaps. First, they document the driest range of the *Acacia-Commiphora* deciduous bushland and thicket (Bd), which occurs under 400-600 mm/yr precipitation within the 20°-30°C mean annual temperature zone (**Figure 2**). Second, they document the hottest and driest range of occurrence of the Complex of Afromontane forest with wooded grasslands and evergreen or semi-evergreen bushland and thicket (Fb/Wd) where mean annual temperature is 18-21°C and mean annual precipitation 700-800 mm/yr. And third, they document the lowland azonal vegetation fed by springs or groundwater. Despite these additions, sampling of the climate spaces could still be improved for Afroalpine vegetation and Ericaceous belt (A.E), Dry *Combretum* wooded grassland (Wcd), Complex of Afromontane forest with wooded grasslands and evergreen or semi-evergreen bushland and thicket (Fb/Wd), and Semi-desert grassland and shrubland (S) and Desert (D).

3.2. Geopositioning and assignment of environmental data

To make local vegetation reconstructions possible, geolocation of some pollen samples required to be reworked. This was especially true for samples that were collected at a time when GPS was not available or not very accurate (Bonnefille, 2018). To do so, we used all material available: authors' notes, descriptions of vegetation at sampling sites, mission preparation documents, road maps, elevation data obtained with altimeters, and mission reports, in addition to information provided in the original publications. The data we used are from (Barboni et al., 2019; Bonnefille et al., 1987, 2004, 1993; Bonnefille, 1972; Bonnefille and Buchet, 1986; Bonnefille and Riollet, 1988; Buchet, 1982; El-Moslimany, 1983; Garnier et al., 2013; Lézine, 1981; Lézine et al., 1998; Mohammed, 1992; Mohammed et al., 2007; Mohammed and Bonnefille, 2002) (column 6 and 9 in **Table S5**). For example, in the Hadar area (north of the Awash valley), samples E010 to E026 are archived under the same geographical coordinates in the APD, although author's notes and a map in the original (Bonnefille et al., 1987) show that they were collected at different locations along the Awash River, and not some 10 to 16 km away in the xerophytic bushland . For the purpose of our study, their geographical coordinates were therefore corrected. In total, we corrected the geographical coordinates for 68% of the "old" pollen samples (150/221 pollen samples) with an average correction of 9.7 km (column 9 in **Table S3**).

Using a geographic information system software (GIS QGis, v3.2.2), we assigned to all 283 sites (with corrected coordinates) 12 soil parameters (SoilGrids, 2000-2015, 0.25×0.25 km, Hengl et al., 2017), 32 climatic parameters (WordClim2, 1970-2000, 0.93×0.93 km, Fick and Hijmans, 2017), five aridity parameters (Global High-Resolution Soil-Water Balance, 1950-2000, 0.93×0.93 km, Trabucco and Zomer, 2019), Net Primary Productivity (NPP) (MODIS MOD17A3, 2000-2014, 0.93×0.93 km, Zhao et al., 2005), and percent of tree cover (Global Maps - Vegetation, year 2008, 0.46×0.46 km, Kobayashi et al., 2017) (**Table S4Erreur ! Source du renvoi introuvable.**). Satellite percent tree cover values for the 2000-2014 period were sometimes very different from the authors' notes at the time of sampling (some 40 years ago in some cases). Coherence between modern satellite tree cover values and authors' notes was therefore checked for every site, and when deforestation was observable on Google Earth images ("time travel" function) (column 10 in **Table S5**), then tree cover values of neighboring, unaffected or less affected areas within 1-kilometer radius of the sampling point were considered. Similarly, we corrected tree cover values when a slight discrepancy (<250 m) was observed between Google Earth images and the tree cover grid. In total, tree cover values were corrected for 116 sites (over 283). GIS was also used to attribute to each surface sample (with corrected geographical coordinates) a potential vegetation type according to the VECEA project (van Breugel et al., 2015) for African samples and according to vegetation map of Frey and Kürschner (1989) for Arabian samples

258 (column 7 in **Table S5**). We gave priority to the collector's notes and vegetation observed with Google

259 Earth to correct the assigned potential vegetation accordingly (column 6 and 8 in **Table S5**).

260 ### 3.3. Pre-processing of data

261 To define a threshold pollen percentage optimized to our study area, data transformations and

262 analyses were carried out using R (R Development Core Team, 2018). Pollen type relative abundance

263 (without spores and undetermined pollen) were calculated for each of the 283 sites and compiled in

264 double entry matrix of n pollen taxa x m sites (data in **Table S2**). The double entry matrix was then

265 transformed into a column vector $\vec{V}$. Classified in ascending order, values of $\vec{V}$ show a break at 0.2% in

266 the evolution of pollen percentages. We interpret this break as the signal of allochthonous taxa, which

267 occurrence in pollen assemblages is erratic. Before further processing of the data, all $\vec{V}$ entries with

268 values <0.2% were therefore excluded from the column vector to reduce the noise associated with the

269 presence of allochthonous pollen (**Figure 3**). In other words, we chosen a threshold pollen percentage

270 of 0.2% for this study.

271 ### 3.4. Determination of the relative abundance scale

272 To define the relative abundance scale optimized to our study area, after removal of pollen

273 percentage < 0.2% in each pollen sample, we used the Vegan function cascadeKM() (Vegan v2.5-2:

274 Oksanen et al., 2010) that applies the k-means partition comparison method and the simple structure

275 index (ssi) criterion to identify the minimum number of classes of relative abundances that keep a

276 maximum of information (Dolnicar et al., 1999). The larger the ssi, the more significant is the

277 partitioning (Borcard et al., 2018). Based on this index, 2 to 6 classes were optimal (**Figure S1**). We

278 retained six classes. To define the classes we used the Fisher-Jenks method of the Cartography function

279 getBreaks() (Cartography v2.2.0: Giraud and Lambert, 2017): relative abundance class 1 ∈]0.2% ; 3%],

280 relative abundance class 2 ∈]3% ; 11%], relative abundance class 3 ∈]11% ; 23%], relative abundance

281 class 4 ∈]23% ; 38%], relative abundance class 5 ∈]38% ; 60%] and relative abundance class 6 ∈]60%

282 ; 100%]. A 7th class was added for the threshold pollen percentage of 0.2%: relative abundance classes

283 0 ∈]0% ; 0.2%].

284 ### 3.5. Inference of vegetation groups

285 The pollen assemblages were grouped into vegetation groups from 6 observations: (1) the

286 partition of the pollen assemblages according to the pollen data using a k-means partitioning

287 constrained by a hierarchical classification with the function vegdist(), function cascadeKM() and

288 function hclust() of Vegan v2.5-2 and Stats v3.5.1 R package (Oksanen et al., 2010; R Core Team, 2010),

289 (2) the sample type (mud or soil), (3) the potential natural vegetation (Frey and Kürschner, 1989; Friis

290 et al., 2010; van Breugel et al., 2015), (4) the vegetation described at the sampling sites in the original

291 publications (Barboni et al., 2019; Bonnefille et al., 1987, 1993; Bonnefille, 1972; Bonnefille et al., 2004;

292 Bonnefille and Buchet, 1986; Bonnefille and Riollet, 1988; El-Moslimany, 1983; Garnier et al., 2013;

293 Lézine, 1981; Lézine et al., 1998; Mohammed, 1992; Mohammed et al., 2007; Mohammed and

294 Bonnefille, 2002), (5) the vegetation described at the sampling sites in mission reports, (6) and some

295 hydrological studies providing information on the edaphic context at the sampling sites (e.g., spring)

296 (Ayenew et al., 2008; Cuthbert et al., 2017) (**Table S5**). Of the 283 pollen assemblages, 21 were

297 removed: 7% of the dataset. Within these 7%, 12 samples are from anthropic sites (e.g., DB09-04,

298 Karasoditi coffee forest) and 9 samples are from very poorly sampled vegetation (e.g., BB17-41,

299 pioneer vegetation growing in small patches on 200 years-old lava flow) (**Table S5**). In the end, the

300 pollen assemblages were distributed in 24 vegetations types or subtypes, including nine local riparian

301 and azonal vegetation. (**Table 2**). These 24 groups are divided into two sets: 18 groups composed of

302 pollen samples from non-dynamic depositional environments (soils, muds of lake edges and soils/muds

303 of swamps) and 6 groups composed of pollen samples from dynamic depositional environments (muds

304 of river beds or edges).

305 ### 3.6. Environmental parameters for inferred vegetation groups

306 To see which environmental, climatic, and/or edaphic variables would best relate to the 24

307 vegetation groups (**Table 2**), we ran a Random Forest model of 10000 iterations with the function

308 Random Forest () of the Random Forest v4.6 R package. We used 10 variables measuring temperature,

309 10 for precipitation, 5 for aridity, 11 for soil, tree cover percentage, and net primary production (NPP)

310 (columns 8-26 ; 39-44 ; 46 ; 48-59, **Table S4**). We then applied the function varImpPlot () of the same

311 R package with our model to see which variables best explain the distribution of vegetation with the

312 function.

313 ### 3.7. Vegetation score ($V_{\alpha,\beta}$ score)

314 Pollen taxa affinities for a vegetation group, were estimated with the Value Test (VT) (Lebart,

315 2000). VT compares for one variable (pollen type) the values of descriptive statistic indicators (means,

316 variances, and group size) computed on all groups (the full dataset: all vegetation groups tested) and

317 on a subsample related to one group (one vegetation group). The VT of a pollen taxon for a vegetation

318 group is calculated as follows:

319
$$VT = \frac{\mu_g - \mu}{\sqrt{\dfrac{n - n_g}{n - 1} \times \dfrac{\sigma^2}{n_g}}}$$

320   where μ_g is the mean relative abundance of the pollen taxon within the vegetation group, μ the mean

321   of the pollen type (in relative abundance, relative abundance scale, or presence/absence) in the whole

322   dataset, n_g the size of the group (number of samples in the vegetation group), n the size of the whole

323   dataset (total number of samples), and σ^2 the variance associated with mean of the pollen type (in

324   relative abundance, relative abundance scale, or presence/absence) in all vegetation groups tested

325   (the whole dataset) (Lebart et al., 1995). The VT of a pollen type for a vegetation group is a simple

326   measure of similarity between the pollen type and the vegetation group. Similarity is indicated by

327   positive VT, and dissimilarity by negative VT (**Table S6**).

328   Using VT values of the pollen taxa for the vegetation groups, we then developed the calculation

329   of a score to measure the affinity of the pollen samples for vegetation groups. The affinity of a pollen

330   assemblage to a vegetation group is calculated as follows:

331
$$A_{\alpha,\beta} = \frac{\sum_{i=0}^{9}\left(\sum P_{VT\%\in\left]\frac{10i\times VT_{max}}{100}\,;\,\frac{10(i+1)\times VT_{max}}{100}\right]} \times i\right)}{\sum_{i=0}^{9}\left(N_{VT\%\in\left]\frac{10i\times VT_{max}}{100}\,;\,\frac{10(i+1)\times VT_{max}}{100}\right]} \times i\right)} - \frac{\sum_{i=0}^{9}\left(\sum P_{VT\%\in\left[\frac{10(i+1)\times VT_{min}}{100}\,;\,\frac{10i\times VT_{min}}{100}\right[} \times i\right)}{\sum_{i=0}^{9}\left(N_{VT\%\in\left[\frac{10(i+1)\times VT_{min}}{100}\,;\,\frac{10i\times VT_{min}}{100}\right[} \times i\right)}$$

332   where $A_{\alpha,\beta}$ is affinity of sample β for the vegetation group α, $P_{VT\%\in]x;y]}$ is the relative abundance scale

333   of a pollen type in sample β with VT percentage between $]x; y]$ for the vegetation group α, $N_{VT\%\in]x;y]}$ is

334   the number of pollen types in the sample β with VT percentage between $]x; y]$ for the km-group α. The

335   VT percentage of a pollen type for the vegetation group α is [VT$_{pollen\ type}$ x 100]/VT$_{max}$, VT$_{max}$ is the most

336   positive VT of pollen type for the vegetation group α, and VT$_{min}$ is more negative VT of pollen type for

337   the vegetation group α. In this formula, the first member of the formula (left) is the weighted mean of

338   the relative abundance scale of all pollen types in sample β with a positive VT for the vegetation group

339   α. The weight of each relative abundance scale is based on the VT percentage of its pollen type such

340   as: VT%$\in$]10n ; 10(i+1)] for weight=i+1. The second member of the formula (right) does the same thing

341   but for negative VT. $A_{\alpha,\beta}$ therefore represents the weighted mean of pollen of sample β with a similarity

342   for the vegetation group α minus the weighted mean of pollen of sample β with a dissimilarity for the

343   km-group α.

344   In order to have a Vegetation score (V$_{\alpha,\beta}$ score), the standard score of $A_{\alpha,\beta}$ for $A_{all,\beta}$ (all A scores

345   of sample β for all vegetation groups tested) is expressed as follows:

346
$$V_{\alpha,\beta} = \frac{A_{\alpha,\beta} \times \mu_{all,\beta}}{\sigma_{all,\beta}}$$

347 where $V_{\alpha,\beta}$ is the vegetation score of sample β for the vegetation group α, $A_{\alpha,\beta}$ is affinity of pollen

348 sample β for the vegetation group α, $\mu_{all,\beta}$ is the mean of all A scores of sample β for all vegetation

349 groups tested, and $\sigma_{all,\beta}$ is the standard deviation of $A_{all,\beta}$ (all A scores of sample β for all vegetation

350 groups tested). This last step was added in anticipation of using this method on fossil data. Indeed, for

351 similar vegetation, fossil pollen samples may have a lower taxonomic diversity than their modern

352 analogues due to taphonomic bias. In this particular case, the A-scores of fossils will be lower than the

353 A-scores of their modern analogues. The numeric centering and scaling used in V-scores minimize this

354 effect.

355 **3.8. Significant and non-significant score ranges for $V_{\alpha,\beta}$ score**

356 For a selected vegetation group in all vegetation groups tested, we calculate the $V_{\alpha,\beta}$ score values

357 of this selected vegetation group (α) for all pollen assemblages of all vegetation groups tested: $V_{\alpha,\beta \in all}$

358 $_{tested\ groups}$. The $V_{\alpha,\beta}$ score values of the pollen assemblages classified in this vegetation group ($V_{\alpha,\beta \in \alpha}$)

359 define the range of significant values for $V_{\alpha,\beta}$ score of this selected vegetation group. The $V_{\alpha,\beta}$ score

360 values of the pollen assemblages that are not classified in this selected vegetation group ($V_{\alpha,\beta \notin \alpha}$) define

361 the range of non-significant values for $V_{\alpha,\beta}$ score of this selected vegetation group.

362 **4. Results**

363 **4.1. $V_{\alpha,\beta}$ score applied to the vegetation groups**

364 The $V_{\alpha,\beta}$ score was applied to pollen assemblages from the 24 vegetation groups but by grouping

365 the six groups linked to dynamic depositional environments into two sets: the S-, Bd- and, We-river

366 grouped as We/Bd/S-river and the Fb/Wd-river, Wcd-river 1, and Wcd-river 2 grouped as Wcd/Fb/Wd-

367 river (**Figure 4**). We have chosen this grouping because the pollen assemblages of these 6 groups linked

368 to dynamic depositional environments include a high proportion of exogenous pollen (pollen from

369 vegetation groups not linked to dynamic depositional environments) which biases the determination

370 of pollen taxa affinities with the VT score.

371 For pollen data in relative abundance, the $V_{\alpha,\beta}$ score of 18 vegetation groups (all groups except

372 the Fb/Wd-low/middle and Fa-w) have overlapping significant and non-significant values (**Figure 4**).

373 $V_{\alpha,\beta}$ score of 10 vegetation show large overlap of significant values by non-significant values ($\geq$ 50%).

374 $V_{\alpha,\beta}$ score of S-xs ($V_{S-xs,\beta}$) has non-significant values that completely overlap with significant values.

375 Depending on the vegetation (α) of the $V_{\alpha,\beta}$ score, the pollen assemblages that cause overlaps with

376 significant values come from one to 10 different vegetation groups (on average 4). The taxonomic and

377 geographical proximity of the vegetation groups explain some overlaps. This is the case for the close

378 $V_{Bd-wb\,2,\beta}$ score values between pollen assemblages from vegetation subtypes Bd-wb-1, -2, and, -3 (e.g.,

379 overlap of $V_{Bd-wb\,2,\beta\in Bd-wb\,2}$ linked to $V_{Bd-wb\,2,\beta\in(Bd-wb\,1\,\cup\,Bd-wb\,3)}$). But 25% of these overlaps, such as the

380 overlap of V score values of the pollen assemblages from S-xs and Fa-upper (e.g., overlap of $V_{S-xs,\beta\in S-xs}$

381 related $V_{S-xs,\beta\in Fa-upper}$), cannot be explained. The use of the $V_{\alpha,\beta}$ score with pollen data in relative

382 abundance seems to discriminate 10 vegetation in an acceptable way but for the other 10 vegetation

383 the discrimination is not good (overlap of significant values by non-significant values ≥ 50%).

384 For pollen data in relative abundance scale, the $V_{\alpha,\beta}$ score of only four vegetation groups

385 (We/Bd/S-river, Wcd/Fd/Wd-river, Bd-wb 2, and Bd-wb 3) have overlapping significant and non-

386 significant values (**Figure 4**). These are slight overlaps of significant scores by non-significant scores (<

387 50%). Depending on the vegetation (α) of the $V_{\alpha,\beta}$ score, pollen assemblages that cause overlaps with

388 significant values come from one to two different vegetation groups (on average 1). The taxonomic

389 and geographical proximity of the vegetation groups explain all overlaps. This is the case for the close

390 $V_{Bd-wb\,2,\beta}$ score values between pollen assemblages from vegetation subtypes Bd-wb-1 and-3 (e.g.,

391 overlap of $V_{Bd-wb\,3,\beta\in Bd-wb\,3}$ linked to $V_{Bd-wb\,3,\beta\in Bd-wb\,1}$). In contrast to the relative abundance pollen data,,

392 significant values of $V_{Wcd/Fb/Wd-river,\beta}$ score overlap only with pollen assemblages from We/Bd/S-river

393 group and the significant values of $V_{We/Bd/S-river,\,all}$ score overlap only with the pollen assemblages from

394 Wcd/Fd/Wd-river group. Despite these overlaps, all pollen assemblages with a significant value for

395 $V_{We/Bd/S-river,\,\beta}$ or/and $V_{Wcd/Fb/Wd-river,\,\beta}$ scores can only be from We/Bd/S-river or Wcd/Fd/Wd-river groups,

396 and by extension they are necessarily from the set of dynamic depositional environments (the S-, Bd-,

397 We-, Fb/Wd-, and the two Wcd-river vegetation group). The use of the $V_{\alpha,\beta}$ score with pollen data in

398 relative abundance scale is therefore appropriate to discriminated vegetation liked with non-dynamic

399 depositional environments and for isolating the pollen assemblages from dynamic depositional

400 environments.

401 For pollen data in presence/absence, $V_{\alpha,\beta}$ scores of four vegetation (We/Bd/S-river, Bd-spring1,

402 Bd-wb 1, and Fa-upper) show large overlap of significant values by non-significant values (≥ 50%)

403 (**Figure 4**). $V_{\alpha,\beta}$ score of 6 vegetation (Wcd/Fb/Wd-river, Bd-wb 2, Bd-wb 3, Bd-rtf, Fb/Wd-upper, and

404 Fa-sc 1) are slight overlaps of significant scores by non-significant scores (< 50%). Depending on the

405 vegetation (α) of the $V_{\alpha,\beta}$ score, the pollen assemblages that cause overlaps with significant values

406 come from one to 11 different vegetation groups (on average 3). Except for $V_{We/Bd/S-river,\beta}$, the taxonomic

407 and geographical proximity of the vegetation groups explain the overlaps. As a result, to discriminate

408 vegetation, $V_{\alpha,\beta}$ score with pollen data in presence/absence is more efficient than $V_{\alpha,\beta}$ score with pollen

409 data in relative abundance scale but less efficient than the $V_{\alpha,\beta}$ score with pollen data in relative

410 abundance. Moreover, as $V_{\alpha,\beta}$ score with pollen data in relative abundance and unlike $V_{\alpha,\beta}$ score with

411   pollen data in relative abundance, $V_{\alpha,\beta}$ score with pollen data in presence/absence does not efficiently

412   isolate the pollen assemblages from dynamic depositional environments.

4.2. V $_{\alpha,\beta}$ score applied only to the vegetation groups linked to dynamic depositional environments

415   As seen in the previous section, in some cases, it is possible to isolate with $V_{\alpha,\beta}$ score the set of

416   vegetation groups related to dynamic depositional environments (the S-, Bd-, We-, Fb/Wd-, and the

417   two Wcd-river vegetation group). The question, then, is whether it is possible to further improve the

418   discrimination of vegetation groups with the $V_{\alpha,\beta}$ score once this set is isolated. We present here the

419   results of the $V_{\alpha,\beta}$ score applied only to the six vegetation groups linked to dynamic depositional

420   environments: S-, Bd-, We-, Fb/Wd-, and both Wcd- vegetation groups.

421   For pollen data in relative abundance, the $V_{\alpha,\beta}$ score of all vegetation groups have overlapping

422   significant and non-significant values (**Figure 5**). Depending on the vegetation (α) of the $V_{\alpha,\beta}$ score, the

423   pollen assemblages that cause overlaps with significant values come from one to five different

424   vegetation groups (on average 3). $V_{\alpha,\beta}$ score of four vegetation groups show large overlap of significant

425   values by non-significant values ($\geq$ 50%). Using the $V_{\alpha,\beta}$ score with pollen data in relative abundance

426   discriminates acceptably only two of the six vegetation groups.

427   For pollen data in relative abundance scale, the $V_{\alpha,\beta}$ score of only two vegetation groups have

428   overlapping significant and non-significant values (**Figure 5**). Depending on the vegetation (α) of the

429    $V_{\alpha,\beta}$ score, the pollen assemblages that cause overlaps with significant values come from one to three

430   different vegetation groups (on average 2). $V_{\alpha,\beta}$ score of only one vegetation group show large overlap

431   of significant values by non-significant values ($\geq$ 50%). Using the $V_{\alpha,\beta}$ score with pollen data in relative

432   abundance scale discriminates acceptably five of the six vegetation groups. This is much better than

433   with the data in relative abundance.

434   For pollen data in presence/absence, there is no overlap between the significant and non-

435   significant values of $V_{\alpha,\beta}$ score (**Figure 5**). The presence/absence data allow us to better discriminate

436   between vegetation groups than the relative abundance data.

4.3. Relation between vegetation groups and environmental parameters

438   The Random Forest model was able to recreate the 24 vegetation groups (**Table 2**) with a success

439   rate of 80%. This indicates that 80% of our dataset has a pollen signal that provides information about

440   one or several environmental parameters. Three parameters emerge as most correlated with

441   vegetation groups inferred from their pollen signature: the tree cover percentage, mean temperature

442   of warmest quarter, and mean precipitation of driest quarter (**Figure S2**). Globally, the climatic

443   parameters and the NPP are more correlated with the vegetation groups than the edaphic parameters.

444   For the two climatic parameters that best explain our vegetation groups (mean temperature of

445   warmest quarter and mean precipitation of driest quarter) (**Figure 6**), it appears that the 24 groups

446   have well-defined and very diverse climatic spaces. The climatic spaces of these vegetation groups

447   range from less than 10°C to more than 35°C for the mean temperature of warmest and from almost

448   0 mm to more than 150 mm for the mean precipitation of driest quarter. These 24 vegetation groups

449   are distinguished by the two main climatic parameters (e.g., Bd-wb1 vs. Bd-wb3) or by edaphic

450   parameters (e.g., Bd-wb 1 vs S-river). Sometimes it is secondary climatic parameters that explain the

451   differences in vegetation groups. This is the case for the mean annual precipitation for Bd-wb 3 and

452   Wcd, respectively from 446 mm to 919 mm and from 1013 mm to 1336 mm (column 26 in **Table S4**).

453   Our vegetation groups correspond to distinct physiognomic types of vegetation with different

454   tree cover (**Figure 7**). With a 95% confidence interval, the estimated intervals of tree cover percentage

455   are different and complementary for the 24 vegetation groups (e.g., Fa-sc 1 and Fa-sc 2). These

456   estimated tree covers range from 0 to 82% depending on the vegetation group.

457   **4.4. Morphologies and vegetation strip of rivers**

458   This study differentiates several types of azonal and local vegetation, including seven river-

459   related vegetation (S-river, Bd-river, Bd-rtf, We-river, Fb/Wd-river, Wcd-river 1, and Wcd-river2).

460   These seven vegetations can each be associated with the morphology of the river (meandering or

461   straight) and with a width of vegetation strip along the river banks (**Table 2**). For example, it is possible

462   to say that a Bd-river pollen-inferred vegetation implies the presence of a meandriform and perennial

463   river with a 50-100 m wide strip of open forest along its banks. The combination of these two

464   parameters is specific to each group. Details are given for the seven vegetation groups in **Table 2**.

465   **4.5. Most characteristic pollen types**

466   For the sets of non-dynamic depositional environments (18 vegetation groups: **Table 2**), if we

467   look at the most characteristic pollen types of each of these vegetation (**Table 3**) on the base of T-

468   values with data in relative abundance scale (**Table S6**) , we note that arboreal pollen types are more

469   often used than non-arboreal pollen types for vegetation characterization. This is true for all vegetation

470   groups except those related to the most open and xerophilic vegetation (S-river, S-xs, S-xds, D) or to

471   Afromontane vegetation (E.A-e). For some vegetation types, such as Fa-w, the pollen taxa markers

472   correspond to botanical markers identified by botanists (e.g., *Macaranga* or *Pouteria*) (Friis et al.,

473   2010). Similarly, characteristic pollen types for riparian and azonal vegetation types were easily

474   identified as *Tamarix* for Bd-rft and *Hyphaene* for Bd-spring 1 et 2 in agreement with the relative

475   abundance of these trees in riparian *Tamarix* forests and groundwater-fed woodlands. Overall, 10% to

476   40% of the pollen taxa identified as the most characteristic correspond to botanical markers identified

477   by botanists.

478   For the sets of non-dynamic depositional environments (6 vegetation groups: **Table 2**), if we

479   look at the most characteristic pollen types of each vegetation (**Table 3**) on the base of T-values with

480   data in presence/absence (**Table S6**), correlations with the surrounding vegetation are not obvious,

481   which is not surprising because the pollen spectra integrate pollen grains likely coming from vegetation

482   much further upstream. Yet, six different river vegetation types were identified by our approach, which

483   shows that riparian formations carry distinct pollen signals according to the vegetation zones crossed

484   by the river .

485   **5.   Discussion**

486   **5.1. The distinction of vegetation**

487   The $V_{\alpha,\beta}$ score presented in this paper effectively distinguishes 24 vegetation groups distributed

488   in eight vegetation types (**Table 2**). Of these 24 vegetation types, 22 are subtypes of the regional

489   vegetation, and of these 22, nine are local riparian and azonal vegetation. To be able to do this, an

490   initial analysis must be carried out with the pollen data in relative abundance scale. This is to

491   discriminated vegetation liked with non-dynamic depositional environments and for isolating the

492   pollen assemblages from dynamic depositional environments (pollen assemblage from S-, Bd-, We-,

493   Fb/Wd-, and the two Wcd-river vegetation group). A second analysis of pollen assemblages identified

494   as pollen assemblages from dynamic depositional environments to specify their vegetation: S-river,

495   Bd-river, We-river, Fb/Wd-river, Wcd-river 1, or Wcd-river 2. This second operation can be carried out

496   either with presence/absence data (preferably) or in relative abundance scale.

497   This is the first time that pollen score has been successfully developed to discriminate

498   vegetation with a spatial resolution lower than the regional scale. In East Africa, vegetation

499   discriminations using pollen score calculations on modern pollen assemblages (and by extension on

500   fossil pollen assemblages) were limited to the supra-regional or regional scale for Ethiopia, Kenya,

501   Tanzania and Uganda. (Bonnefille et al., 2004; Vincens et al., 2006b). This is also the case for Central

502   and West Africa (Hély et al., 2006; Izumi and Lézine, 2016; Lebamba et al., 2009). The difficulty of

503   separating the local and regional pollen signal has been known for a long time (Bonnefille, 1984). Until

504   now, no solution to this problem had been found, either by traditional palynological approaches or by

505   calculation methods.

506   We also show that it is possible to distinguish nine azonal and riparian vegetation very well via
507   pollen assemblages from dynamic or non-dynamic environments. These are also new results because
508   these vegetation are generally neglected in the palynological studies of our study area (Bonnefille et
509   al., 2004; Hély et al., 2006; Izumi and Lézine, 2016; Lebamba et al., 2009; Vincens et al., 2006b), and
510   aquatic taxa are usually removed from the analyses as originally proposed by Jolly et al. (1998). Our
511   results are a step forward for future application in the study of Hominini sites in East Africa, because
512   they are very frequently associated with riparian or spring contexts (Barboni et al., 2019;
513   Behrensmeyer and Reed, 2013; Campisano and Feibel, 2008; Levin, 2015). The reconstructions of these
514   local and azonal environments where the Hominini fossils are located, are not clearly defined yet by
515   palynological or faunal studies, which normally define them as "swamp", "wetland" or "floodplain"
516   (Bonnefille et al., 1987, 2004; Bonnefille, 2010; Reed, 2008).

517   We have not been able to separate the Afroalpine vegetation (A) and Ericaceous belt (E) pollen
518   signal. The samples of these two vegetation are in the A.E-e vegetation sub-type. An explanation is
519   that all samples of A are in the Sanetti plateau (Ethiopia) and correspond to a particular area with
520   isolated *Erica* shrubs (Mohammed et al., 2007). Miehe and Miehe (1994) suggested that this area
521   represents the potential extent of *Erica*-dominated vegetation in the absence of natural or anthropic
522   fires, hence our difficulty in the separation of A and E samples. Another difficulty is at the lower limit
523   of the Fa where Friis et al. (2010) observe a "Moist evergreen Afromontane forest, bushland, woodland
524   and wooded grassland". Although several Fa-sc 1, Fa-sc 2, and Fa-w samples are available in this area,
525   our analyses did not reveal any specific pollen composition for this "Moist evergreen Afromontane
526   forest, bushland, woodland and wooded grassland". The potential samples related to this vegetation
527   are mainly in 3 sample sets: E201 to E209, E210-E235 and E268 to E282. The first set consists of
528   unpublished data in APD and for the third set the authors do not mention this sub-type of vegetation
529   (Mohammed, 1992). For the second set, the palynologists refer to this vegetation as "Intermediate
530   forest" (cf., next paragraph) (Bonnefille et al., 1993). With their 25 pollen samples from South West
531   Ethiopia, they show that the "Moist evergreen Afromontane forest, bushland, woodland and wooded
532   grassland" is well distinguished from the other samples of Fa. Our results suggest that this distinction
533   may not be valid on a larger scale.For Fa-w, Bonnefille et al. (1993) indicate that the Fa-w samples
534   collected near Gore (western Ethiopia) come from the "Intermediate forest" (corresponds to the
535   Transitional rain forest type of Friis et al. (2010)). However, Friis et al. (2010) describe the vegetation
536   at the location of the sampling sites of Fa-w samples (near Gore) as "Primary or mature secondary
537   moist evergreen Afromontane forest", a sub-type of Fa type. Normally, the elevation of Fa-w samples
538   (lowest sample: 1020 m a.s.l.) is too low for Fa (low limit of Fa: 1550-1450 m a.s.l.). But in this part of
539   Western Ethiopia, the "Primary or mature secondary moist evergreen Afromontane forest" has the

540 particularity of occurring below the calculated lower elevation limit of the Fa, and grows at the same
541 altitude as the Transitional rain forest (Friis et al., 2010). In addition, the present plants described by
542 Bonnefille et al. (1993) corresponds to vegetation of "Primary or mature secondary moist evergreen
543 Afromontane forest" (Friis et al., 2010). This is why we preferred to define Fa-w as Secondary moist
544 evergreen Afromontane forest rather than as Transitional rainforest

545 Finally, our results show that the pollen type *Hyphaene*-type is a main marker of spring as
546 suggested by Barboni et al. (2019). This is an important step because it is supposed that springs have
547 an important role in the hominin evolution in East Africa (Ashley et al., 2009; Cuthbert et al., 2017;
548 Cuthbert and Ashley, 2014). It is interesting to note that there is clear distinction between Ethiopian
549 (Bd-spring 1) and Tanzanian (Bd-spring 2) springs related to the regional differences in vegetation,
550 groundwater (alkaline water in Ethiopia and fresh water in Tanzania: personal observation) and mean
551 annual temperature.

552 **5.2. The extrapolated information from the inferred vegetation**

553 In the East African Rift, the variations in river vegetation are not only linked to the variations in
554 climatic conditions and surrounding vegetation (Friis et al., 2010), but also to the variations in river
555 morphology (Carr, 1998). This relationship allowed us to associate these seven types of river vegetation
556 with river morphologies (**Table 2**). One of the most interesting cases in this study concerns the pollen
557 assemblages from Wcd where the subtypes Wcd-river 1 and Wcd-river 2 are associated respectively
558 with straight rivers and meandering rivers. In the same way, it is possible to associate the seven river-
559 related vegetation groups with the width of the vegetative strip along the rivers (**Table 2**). This agrees
560 with the observations of Carr (1998). Showing that it is possible to do this type of discrimination with
561 pollen scores is interesting because the structure of the river and the width of the vegetation strip are
562 important elements as they influence the type of fauna present (Geheb and Abebe, 2003).

563 Like previous studies in East Africa (Bonnefille et al., 2004; Peyron et al., 2000) or in other parts
564 of the world (Barboni and Bonnefille, 2001), our vegetation groups can be linked to several parameters,
565 mainly climatic. It is the mean temperature of warmest quarter and the mean precipitation of driest
566 quarter are the climate parameters that best explain our vegetation groups. This is not surprising
567 because in our study area, the limiting parameters for vegetation are high temperatures and low
568 precipitation(Friis et al., 2010). The new contributions of this study are the ability to strongly associate
569 our 24 vegetation groups, and by extension our pollen assemblages, with ranges of percent tree cover.
570 We show that pollen can be a strong proxy for tree cover and that they can help to quantify it. Pollen-

571   inferred tree cover will be interesting to compare with paleo-tree cover estimates based on $\delta^{13}C$ at

572   hominin sites (e.g., Cerling et al., 2011).

573   **5.3. Importance of data type**

574   With $V_{\alpha,\beta}$ score, the relative abundance scale data allowed a very good discrimination of

575   vegetation related to non-dynamic depositional environments and effectively isolated the set of

576   vegetation related to dynamic depositional environments (**Figure 5**). This was not possible with the

577   relative abundance data and was partially possible with the presence/absence data. The reason for

578   this result is that for these depositional environments, the abundance scale minimizes the magnitude

579   of variations in the pollen type relative abundance within the vegetation group while avoiding the loss

580   of information of variation between vegetation groups (Borcard et al., 2018). The relative abundance

581   scale data minimizes also the problems of over-representation (e.g., Cyperaceae, Poaceae) or under-

582   representation (e.g., arboreal pollen) of pollen types and has the same role as the square root (Prentice

583   et al., 1996; Vincens et al., 2006b), but is more relevant because our relative abundance scale are

584   specifically calculated for the pollen counts of East Africa.

585   For all 6 vegetation groups related to dynamic depositional environments, discrimination is poor

586   with relative abundance data, good with relative abundance scale data and very good with

587   presence/absence data. (**Figure 6**). We assumed that this was due to the depositional environments,

588   which created very large the magnitude of variations in the pollen type abundance (relative or scaled)

589   within the vegetation group, caused by taphonomic bias (transport and weathering) and by large influx

590   of pollen from upstream vegetation. The variation in abundance of pollen types within each vegetation

591   group then becomes so large that these variations between vegetation groups loses all informative

592   character. This is why the use of presence/absence is most effective.

593   We emphasize here that, contrary to what one might think, discrimination of vegetation from

594   pollen is more effective when the data are in presence/absence or abundance scale than when they

595   are in relative abundance. It is also important to match the type of data to the deposition environment.

596   These are important results because currently, in our study area, pollen studies tend to focus on

597   relative abundance data and analyze pollen assemblages from different depositional environments in

598   the same way (Bonnefille et al., 2004; Peyron et al., 2000; Saylor et al., 2019). Finally, for future use on

599   fossil data, the use of the abundance scale and the presence/absence makes more sense in our study

600   areas where taphonomic biases are assumed to be important (Bonnefille, 1984; El-Moslimany, 1983).

5.4. A more objective approach

Some authors such as Izumi and Lézine (2016) have used thresholds for Poaceae that can be compared to our relative abundance classes, but these thresholds are defined subjectively for each pollen type. We have seen that it is possible to overcome this by using relative abundance scale defined by statistical analysis, or simple observations such as presence/absence. Similarly, the choice to attribute a pollen taxon to a vegetation type is often arbitrary, especially for herbaceous species (cf., Poaceae in Izumi and Lézine, 2016). This choice is based on botanical observations (Bonnefille, 1984), however, the statistical analyses of this study show that the most characteristic pollen types of a vegetation are not always those associated with the plants used by botanists to describe this vegetation (**Table 3**). We have shown here that it is possible to solve these problems of vegetation attributions types by using statistical analyses to define the weight of each pollen type for each vegetation.

To reduce the noise due to occasional long-transport pollen grains, the current value used in East Africa is 0.5% and it is based on Jolly et al. (1998) and Prentice and Webb III (1998). But, these two studies transpose the Prentice et al. (1996) value used for Europe without justifying this transposition to Africa. Based on our graphical analyses (**Figure 3**), we believe that a threshold of 0.2% is more appropriate for Africa. Izumi and Lézine (2016) also suggest this with empirical observations. This changes a lot for pollen score calculations in Africa. For our study, for example, using a threshold of 0.2% removes only 9.5% of total pollen occurrences compared to 43.5% for a threshold of 0.5% (**Figure 3**).

6. Conclusions

Using the $V_{\alpha,\beta}$ score presented in this paper and based on a dataset of 283 modern pollen assemblages, we show that it is possible to discriminate 24 vegetation groups distributed in eight regional vegetation types and including 22 subtypes. This is a solution to the difficulties of separating the local and regional pollen signal for the East East African Rift. This highlights the possibility of using pollen to inform vegetation at a scale smaller than the region. It was shown that riparian vegetation in north part of East African Rift carry distinct pollen and environmental signals, which was largely ignored before. Our new approach allows identifying nine riparian vegetation (two of springs and seven of rivers) which not only relate to regional vegetation composition and climate, but also to the type of wet edaphic conditions.

Except for the removal of non-native taxa from the analysis, it was possible to develop a functional and standardized approach, without preconditions on the pollen types to be considered or

633   the importance to be attributed to the different taxa. As in previous studies, it was possible to link the

634   pollen signal of modern assemblages to climatic parameters. The ability to quantitatively define tree

635   cover via the pollen signal was also highlighted, which is new. Also in the novelty, the pollen signal has

636   been associated with the river morphology (meandering or straight) and type of vegetation strip along

637   the river banks (notably the width).

638   Finally, this study highlighted several elements for optimizing palynological analyses in the East

639   African Rift. It was shown that for this study area, the use of the 0.5% threshold was not optimal and

640   that a 0.2% threshold was preferable. Variations in relative abundance of pollen taxa within the

641   vegetation group was found to be a limiting parameter for vegetation discrimination below the

642   regional scale using pollen data in relative abundance. In order to discriminate between vegetations

643   at this geographic scale, it is necessary to use relative abundance scale or presence/absence pollen

644   data. We highlight the importance of adapting the type of data used according to the variations in

645   relative abundance of pollen taxa within the vegetation group, which is itself related to the

646   depositional environment of pollen assemblages: relative abundance scale data for non-dynamic

647   depositional environments and presence/absence for dynamic depositional environments.

648   These results provide food for thought for the processing and analysis of pollen data in East

649   Africa. It is also a conclusive proof of concept of our pollen score approach for modern pollen

650   assemblages that legitimizes its test on the fossil record. This is both to interpret new fossil data and

651   to reflect on previous pollen analyses. The results of these future works will help to confirm or not the

652   ability of the $V_{\alpha,\beta}$ score to be an alternative to the PFT method.

653

660

661

662   **References**

663   African Pollen Database, 2019 [Online dataset]. URL http://apd.sedoo.fr/accueil.htm (accessed
664   5.6.19).

665   Arráiz, H., Barboni, D., Uribelarrea, D., Mabulla, A., Baquedano, E., Domínguez-Rodrigo, M.,
666   2017. Paleovegetation changes accompanying the evolution of a riverine system at the
667   BK paleoanthropological site (Upper Bed II, Olduvai Gorge, Tanzania). Palaeogeogr.
668   Palaeoclimatol. Palaeoecol., Paleoecological reconstructions of the Bed I and Bed II
669   lacustrine basins of Olduvai Gorge (Tanzania) and insights into early human behavior
670   488, 84–92. https://doi.org/10.1016/j.palaeo.2017.05.010

671   Ashley, G.M., Tactikos, J.C., Owen, R.B., 2009. Hominin use of springs and wetlands:
672   Paleoclimate and archaeological records from Olduvai Gorge (~1.79–1.74 Ma).
673   Palaeogeogr.   Palaeoclimatol.   Palaeoecol.   272,   1–16.
674   https://doi.org/10.1016/j.palaeo.2008.10.016

675   Ayenew, T., Demlie, M., Wohnlich, S., 2008. Hydrogeological framework and occurrence of
676   groundwater in the Ethiopian aquifers. J. Afr. Earth Sci. 52, 97–113.
677   https://doi.org/10.1016/j.jafrearsci.2008.06.006

678   Barboni, D., Ashley, G.M., Bourel, B., Arráiz, H., Mazur, J.-C., 2019. Springs, palm groves, and
679   the record of early hominins in Africa. Rev. Palaeobot. Palynol. 266, 23–41.
680   https://doi.org/10.1016/j.revpalbo.2019.03.004

681   Barboni, D., Bonnefille, R., 2001. Precipitation signal in pollen rain from tropical forests, South
682   India. Rev. Palaeobot. Palynol. 114, 239–258. https://doi.org/10.1016/S0034-
683   6667(01)00057-4

684   Behrensmeyer, A.K., Reed, K.E., 2013. Reconstructing the Habitats of Australopithecus:
685   Paleoenvironments, Site Taphonomy, and Faunas, in: Reed, K.E., Fleagle, J.G., Leakey,
686   R.E. (Eds.), The Paleobiology of Australopithecus, Vertebrate Paleobiology and
687   Paleoanthropology.   Springer   Netherlands,   Dordrecht,   pp.   41–60.
688   https://doi.org/10.1007/978-94-007-5919-0_4

689   Bonnefille, R., 2018. Sur les pas de Lucy - Expédition en Ethiopie, Odile Jacob. ed. Paris, France.

690   Bonnefille, R., 2010. Cenozoic vegetation, climate changes and hominid evolution in tropical
691   Africa.   Glob.   Planet.   Change   72,   390–411.
692   https://doi.org/10.1016/j.gloplacha.2010.01.015

693 Bonnefille, R., 1984. Palynological research at Olduvai Gorge. Natl. Geogr. Soc. Res. Rep. 17,
694   227–243.

695 Bonnefille, R., 1972. Associations polliniques actuelles et quaternaires en Ethiopie (vallées de
696   l'Awash et de l'Omo) (Associations polliniques actuelles et quaternaires en Ethiopie).
697   Paris VI, Paris, France.

698 Bonnefille, R., Buchet, G., 1986. Forest history in Ethiopia, a pollen diagram from Wenchi. SINET
699   Ethiop J Sci 9, 169–178.

700 Bonnefille, R., Buchet, G., Friis, I.B., Kelbessa, E., Mohammed, M.U., 1993. Modern pollen rain
701   on an altitudinal range of forests and woodlands in South West Ethiopia. Opera Bot.
702   121, 71–84.

703 Bonnefille, R., Riollet, G., 1988. Palynologie des sediments holocènes de sites archeologiques
704   du Qatar. Rech. Sur Civilis. Préhistoire À Qatar Paris 137–146.

705 Bonnefille, R., Riollet, G., Chalié, F., Jolly, D., Peyron, O., 2004. High-resolution vegetation and
706   climate change associated with Pliocene Australopithecus afarensis. Proc. Natl. Acad.
707   Sci. 101, 12125–12129.

708 Bonnefille, Vincens, A., Buchet, 1987. Palynology, stratigraphy and palaeoenvironment of a
709   pliocene hominid site (2.9-3.3 M.Y.) at Hadar, Ethiopia. Palaeogeogr. Palaeoclimatol.
710   Palaeoecol. 60, 249–281. https://doi.org/10.1016/0031-0182(87)90035-6

711 Borcard, D., Gillet, F., Legendre, P., 2018. Numerical Ecology with R. Springer.

712 Buchet, G., 1982 [PhD thesis]. Transport des pollens dans les fleuves Omo et Awash (Ethiopie):
713   étude des vases actuelles. École pratique des hautes études, Montpellier, France.

714 Campisano, C.J., Cohen, A.S., Arrowsmith, J.R., Asrat, A., Behrensmeyer, A.K., Brown, E.T.,
715   Deino, A.L., Deocampo, D.M., Feibel, C.S., Kingston, J.D., Lamb, H.F., Lowenstein, T.,
716   Noren, A., Olago, D., Owen, R.B., Pelletier, J.D., Potts, R., Reed, K.E., Renaut, R.W.,
717   Russell, J.M., Russell, J.L., Schäbitz, F., Trauth, M.H., Wynn, J.G., 2017. The Hominin Sites
718   and Paleolakes Drilling Project: High-resolution paleoclimate records from the East
719   African Rift System and their implications for understanding the environmental context
720   of hominin evolution. PaleoAnthropology.

721 Campisano, C.J., Feibel, C.S., 2008. Depositional environments and stratigraphic summary of
722   the Pliocene Hadar Formation at Hadar, Afar Depression, Ethiopia, in: The Geology of
723   Early Humans in the Horn of Africa. Geological Society of America, pp. 179–201.

724 Carr, C.J., 1998. Patterns of vegetation along the Omo River in southwest Ethiopia. Plant Ecol.
725   135, 135–163.

726 Cerling, T.E., Wynn, J.G., Andanje, S.A., Bird, M.I., Korir, D.K., Levin, N.E., Mace, W., Macharia,
727   A.N., Quade, J., Remien, C.H., 2011. Woody cover and hominin environments in the past
728   6 million years. Nature 476, 51–56.

729 Cuthbert, M.O., Ashley, G.M., 2014. A Spring Forward for Hominin Evolution in East Africa. PLoS
730   ONE 9, e107358.

731 Cuthbert, M.O., Gleeson, T., Reynolds, S.C., Bennett, M.R., Newton, A.C., McCormack, C.J.,
732   Ashley, G.M., 2017. Modelling the role of groundwater hydro-refugia in East African
733   hominin evolution and dispersal. Nat. Commun. 8, 15696.

734 Diro, G.T., Black, E., Grimes, D.I.F., 2008. Seasonal forecasting of Ethiopian spring rains.
735   Meteorol. Appl. 15, 73–83. h

736 Diro, G.T., Grimes, D.I.F., Black, E., O'Neill, A., Pardo-Iguzquiza, E., 2009. Evaluation of reanalysis
737   rainfall estimates over Ethiopia. Int. J. Climatol. J. R. Meteorol. Soc. 29, 67–78.

738 Diro, G.T., Toniazzo, T., Shaffrey, L., 2011. Ethiopian Rainfall in Climate Models, in: Williams,
739   C.J.R., Kniveton, D.R. (Eds.), African Climate and Climate Change: Physical, Social and
740   Political Perspectives, Advances in Global Change Research. Springer Netherlands,
741   Dordrecht, pp. 51–69.

742 Dolnicar, S., Grabler, K., Mazanec, J.A., 1999. A tale of three cities: perceptual charting for
743   analyzing destination images, in: Woodside, A., Mazanec, J., Crouch, G.I., Oppermann,
744   M., Sakai, M.Y. (Eds.), Consumer Psychology of Tourism, Hospitality and Leisure. CABI
745   Publishing, New York (USA), pp. 39–62.

746 El-Moslimany, A.P., 1983 [PhD thesis]. History of climate and vegetation in the Eastern
747   Mediterranean and the Middle East from the Pleniglacial to the mid-Holocene.
748   University of Washington, Washington, USA.

749 Faith, J.T., Rowan, J., Du, A., 2019. Early hominins evolved within non-analog ecosystems. Proc.
750   Natl. Acad. Sci. 116, 21478–21483.

751 Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for
752   global land areas. Int. J. Climatol. 37, 4302–4315.

753 Frey, W., Kürschner, H., 1989. A VI 1 Vorderer Orient. Vegetation.

754   Friis, I., Demissew, S., Breugel, P. van, 2010. Atlas of the potential vegetation of Ethiopia., The
755   Royal Danish Academy of Sciences and Letters. ed. Det Kongelige Danske
756   Videnskabernes Selskab, Copenhagen.

757   Garnier, A., Neumann, K., Eichhorn, B., Lespez, L., 2013. Phytolith taphonomy in the middle- to
758   late-Holocene fluvial sediments of Ounjougou (Mali, West Africa). The Holocene 23,
759   416–431.

760   Geheb, K., Abebe, Y.D., 2003. Wetlands of Ethiopia: Proceedings of a Seminar on the Resources
761   and Status of Ethiopia's Wetlands. IUCN.

762   Giraud, T., Lambert, N., 2017. Reproducible Cartography, in: Peterson, M.P. (Ed.), Advances in
763   Cartography and GIScience, Lecture Notes in Geoinformation and Cartography. Springer
764   International Publishing, Cham, pp. 173–183.

765   Haile-Selassie, Y., Gibert, L., Melillo, S.M., Ryan, T.M., Alene, M., Deino, A., Levin, N.E., Scott, G.,
766   Saylor, B.Z., 2015. New species from Ethiopia further expands Middle Pliocene hominin
767   diversity. Nature 521, 483.

768   Hély, C., Bremond, L., Alleaume, S., Smith, B., Sykes, M.T., Guiot, J., 2006. Sensitivity of African
769   biomes to changes in the precipitation regime. Glob. Ecol. Biogeogr. 15, 258–270.

770   Hengl, T., Jesus, J.M. de, Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotić, A.,
771   Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A.,
772   Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I.,
773   Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on
774   machine learning. PLOS ONE 12, e0169748.

775   Izumi, K., Lézine, A.-M., 2016. Pollen-based biome reconstructions over the past 18,000 years
776   and atmospheric CO2 impacts on vegetation in equatorial mountains of Africa. Quat.
777   Sci. Rev. 152, 93–103.

778   Johanson, D., 2017. The paleoanthropology of Hadar, Ethiopia. Comptes Rendus Palevol 16,
779   140–154.

780   Jolly, D., Prentice, I.C., Bonnefille, R., Ballouche, A., Bengo, M., Brenac, P., Buchet, G., Burney,
781   D., Cazet, J.-P., Cheddadi, R., 1998. Biome reconstruction from pollen and plant
782   macrofossil data for Africa and the Arabian peninsula at 0 and 6000 years. J. Biogeogr.
783   25, 1007–1027.

784   Kobayashi, T., Tateishi, R., Alsaaideh, B., Sharma, R., Wakaizumi, T., Miyamoto, D., Bai, X., Long,
785   B., Gegentana, G., Maitiniyazi, A., Cahyana, D., Haireti, A., Morifuji, Y., Abake, G.,

786   Pratama, R., Zhang, N., Alifu, Z., Shirahata, T., Mi, L., Iizuka, K., Yusupujiang, A., Rinawan,
787   F., Bhattarai, R., Phong, D., 2017. Production of Global Land Cover Data – GLCNMO2013.
788   J. Geogr. Geol. 9, p1.

789   Lebamba, J., Ngomanda, A., Vincens, A., Jolly, D., Favier, C., Elenga, H., Bentaleb, I., 2009.
790   Central African biomes and forest succession stages derived from modern pollen data
791   and plant functional types. Clim. Past 5.

792   Lebart, L., 2000. Contiguity Analysis and Classification, in: Gaul, W., Opitz, O., Schader, M. (Eds.),
793   Data Analysis: Scientific Modeling and Practical Application, Studies in Classification,
794   Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg, pp. 233–243.

795   Lebart, L., Morineau, A., Piron, M., 1995. Statistique exploratoire multidimensionnelle. Dunod,
796   Paris.

797   Levin, N.E., 2015. Environment and climate of early human evolution. Annu. Rev. Earth Planet.
798   Sci. 43, 405–429.

799   Lézine, A.-M., 1981 [PhD thesis]. Le Lac Abiyata (Ethiopie) - Palynologie et paléoclimatologie du
800   Quaternaire récent. Université de Bordeaux, Bordeaux, France.

801   Lézine, A.-M., Saliège, J.-F., Robert, C., Wertz, F., Inizan, M.-L., 1998. Holocene Lakes from
802   Ramlat as-Sab'atayn (Yemen) Illustrate the Impact of Monsoon Activity in Southern
803   Arabia. Quat. Res. 50, 290–299.

804   Magill, C.R., Ashley, G.M., Domínguez-Rodrigo, M., Freeman, K.H., 2016. Dietary options and
805   behavior suggested by plant biomarker evidence in an early human habitat. Proc. Natl.
806   Acad. Sci. 113, 2874–2879.

807   Miehe, S., Miehe, G., 1994. Ericaceous forests and heathlands in the Bale Mountains of South
808   Ethiopia: ecology and man's impact, Stiftung Walderhaltung in Afrika and
809   Bundesforschungsanstalt für Forst und Holzwirtschaft. ed. T. Warnke Verlag, Hamburg.

810   Mohammed, M.U., 1992 [PhD thesis]. Paléoenvironnement et paléoclimatologie des derniers
811   millénaires en Ethiopie. Contribution palynologique. University of Aix-Marseille III,
812   Marseille, France.

813   Mohammed, M.U., Bonnefille, R., 2002. Late Holocene Climatic Fluctuations and Historical
814   Records of Famine in Ethiopia, in: Droughts, Food and Culture. Springer, pp. 83–94.

815   Mohammed, M.U., Lamb, H.F., Bonnefille, R., Lézine, A.-M., Tiercelin, J.-J., Gibert, E., Cazet, J.-
816   P., Watrin, J., 2007. Late pleistocene and holocene vegetation history of the bale
817   mountains, Ethiopia. Quat. Sci. Rev. 26, 2229–2246.

818 Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., O'hara, R.B., Simpson, G.L., Solymos, P.,
819   Stevens, M.H.H., Wagner, H., 2010. Vegan: community ecology package. R package
820   version 1.17-4. URL HttpCRAN R-Proj. Orgpackage Vegan.

821 Peyron, O., Jolly, D., Bonnefille, R., Vincens, A., Guiot, J., 2000. Climate of East Africa 6000 14 C
822   yr BP as inferred from pollen data. Quat. Res. 54, 90–101.

823 Prentice, C., Guiot, J., Huntley, B., Jolly, D., Cheddadi, R., 1996. Reconstructing biomes from
824   palaeoecological data: a general method and its application to European pollen data at
825   0 and 6 ka. Clim. Dyn. 12, 185–194.

826 Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A., Solomon, A.M., 1992.
827   Special paper: a global biome model based on plant physiology and dominance, soil
828   properties and climate. J. Biogeogr. 117–134.

829 Prentice, I.C., Webb III, T., 1998. BIOME 6000: reconstructing global mid-Holocene vegetation
830   patterns from palaeoecological records. J. Biogeogr. 25, 997–1005.

831 R Core Team, 2010. R: A language and environment for statistical computing., R Foundation for
832   Statistical Computing. ed. Vienna (Austria).

833 Reed, K.E., 2008. Paleoecological patterns at the Hadar hominin site, Afar Regional State,
834   Ethiopia. J. Hum. Evol. 54, 743–768.

835 Reed, K.E., Spencer, L.M., Rector, A.L., 2013. Faunal approaches in early hominin paleo-ecology,
836   in: Early Hominin Paleoecology. University Press of Colorado.

837 Reynolds, S.C., Wilkinson, D.M., Marston, C.G., O'Regan, H.J., 2015. The 'mosaic
838   habitat'concept in human evolution: past and present. Trans. R. Soc. South Afr. 70, 57–
839   69.

840 Saylor, B.Z., Gibert, L., Deino, A., Alene, M., Levin, N.E., Melillo, S.M., Peaple, M.D., Feakins, S.J.,
841   Bourel, B., Barboni, D., 2019. Age and context of mid-Pliocene hominin cranium from
842   Woranso-Mille, Ethiopia. Nature 573, 220–224.

843 Stewart, K.M., 2014. Environmental change and hominin exploitation of C4-based resources in
844   wetland/savanna mosaics. J. Hum. Evol. 77, 1–16.

845 Trabucco, A., Zomer, R.J., 2019 [Online dataset]. Global High-Resolution Soil-Water Balance.
846   URL https://doi.org/10.6084/m9.figshare.7707605.v3 (accessed 12.6.19).

847 van Breugel, P., Breugel, P. van, Kindt, R., Lillesø, J.-P.B., Bingham, M., Demissew, S., Dudley, C.,
848   Friis, I., Gachathi, F., Kalema, J., Mbago, F.M., Moshi, H.N., Mulumba, J., Namaganda,
849   M., Ndangalasi, H.J., Ruffo, C.K., Védaste, M., Jamnadass, R., Graudal, L., 2015. Potential

850   Natural Vegetation Map of Eastern Africa (Burundi, Ethiopia, Kenya, Malawi, Rwanda,

851   Tanzania, Uganda and Zambia). Version 2.0.

852   Villmoare, B., Kimbel, W.H., Seyoum, C., Campisano, C.J., DiMaggio, E.N., Rowan, J., Braun, D.R.,

853   Arrowsmith, J.R., Reed, K.E., 2015. Early Homo at 2.8 Ma from Ledi-Geraru, Afar,

854   Ethiopia. Science 347, 1352–1355.

855   Vincens, A., Bremond, L., Brewer, S., Buchet, G., Dussouillez, P., 2006a. Modern pollen-based

856   biome reconstructions in East Africa expanded to southern Tanzania. Rev. Palaeobot.

857   Palynol. 140, 187–212.

858   Vincens, A., Bremond, L., Brewer, S., Buchet, G., Dussouillez, P., 2006b. Modern pollen-based

859   biome reconstructions in East Africa expanded to southern Tanzania. Rev. Palaeobot.

860   Palynol. 140, 187–212.

861   Vincens, A., Lézine, A.-M., Buchet, G., Lewden, D., Le Thomas, A., 2007. African pollen database

862   inventory of tree and shrub pollen types. Rev. Palaeobot. Palynol. 145, 135–141.

863   WoldeGabriel, G., Ambrose, S.H., Barboni, D., Bonnefille, R., Bremond, L., Currie, B., DeGusta,

864   D., Hart, W.K., Murray, A.M., Renne, P.R., 2009. The geological, isotopic, botanical,

865   invertebrate, and lower vertebrate surroundings of Ardipithecus ramidus. Science 326,

866   65-65e5.

867   Zhao, M., Heinsch, F.A., Nemani, R.R., Running, S.W., 2005. Improvements of the MODIS

868   terrestrial gross and net primary production global data set. Remote Sens. Environ. 95,

869   164–176.

870

871   **Table captions:**

872   **Table 1:** Comparison of nomenclatures for the main regional vegetation types in Ethiopia.

873   **Table 2:** Regional vegetation types identified by botanists, number of pollen samples, and types of

874   depositional environments of samples for each vegetation group inferred. 0: van Breugel et al. (2015).

875   Non-dynamic depositional environments: soils, muds of lake edges, and soils/muds of swamps;

876   dynamic depositional environments: muds of river beds or edges. Notes: a: observation using Google

877   Earth images (1984 to 2020) ; b: in situ observations ; 1:Barboni et al. (2019) ; 2: Bonnefille and Riollet

878   (1988) ; 3:Bonnefille et al. (1987) ; 4: Buchet (1982) ; 5: Carr (1998) ; 6: Frey and Kürschner (1989) ; 7:

879   Friis et al. (2010) ; 8: Kobayashi et al. (2017) ; 9: Mohammed et al. (2007) ; 10: van Breugel et al. (2015)

880   ; 11: El-Moslimany (1983).

881   **Table 3:** The 10 most characteristic pollen types of each vegetation group inferred based on the
882   calculated T-values). Taxa are listed from the most to the least important. For abbreviations of
883   vegetation groups, please refer to Table 2.

884

885   **Figure captions:**

886   **Figure 1:** Distribution of regional vegetation types and pollen assemblages from this study. Red labels
887   represent samples collected prior to 2009 and available in the APD or the literature, and white labels
888   represent new (unpublished) samples collected in 2009 and 2017. Vegetation distribution is based on
889   (van Breugel et al., 2015) for Africa and on Global Maps - Land Cover (year 2013, 0.46×0.46 km:
890   (Kobayashi et al., 2017) for Arabian Peninsula. For abbreviations of vegetation types, please refer to
891   Table 1. NA: not available.

892   **Figure 2:** Present-day distribution of the vegetation types and pollen samples in the climate domain of
893   Ethiopia (grey shading) as defined by mean annual precipitation (mm/yr) and mean annual
894   temperature (WordClim2, 1970-2000, Fick and Hijmans, 2017). The grey shading is made of 53381
895   dots, each representing 4.63×4.63 km. A-G, Climatic domain occupied by the main vegetation types of
896   Ethiopia obtained using the georeferenced map of vegetation types and sub-types over Eastern Africa
897   produce under the Vegetation and Climate change in Eastern Africa (VECEA) project (van Breugel et
898   al., 2015). H, Position of our 283 surface samples within the present-day climatic domain of Ethiopia
899   and according to to the vegetation identified at the time of sampling(Table S5). For abbreviations of
900   vegetation types, refer to Table 1.

901   **Figure 3:** Occurrence of each pollen type for each pollen assemblage accorded to its relative
902   abundance.

903   **Figure 4:** Comparison according to the type of pollen data (relative abundance, relative abundance
904   scale and presence/absence) of $V_{\alpha,\beta}$ scores (the vegetation score of pollen assemblage β for the
905   vegetation group α) applied to the 262 pollen assemblages for each vegetation group (α). (1) The range
906   of scores of pollen assemblages that belong to vegetation α ($V_{\alpha,\,\beta\in\alpha}$) represented by the black bars (this
907   is the range of significant values of each vegetation α) and the range of scores of pollen assemblages
908   that do not belong to vegetation α ($V_{\alpha,\,\beta\notin\alpha}$) represented by the grey bars (this is the range of non-
909   significant values of each vegetation α). (2) For each vegetation α, the vegetation groups from which
910   pollen assemblages β originated with $V_{\alpha,\,\beta}$ scores intersecting the range of significant values ($V_{\alpha,\,\beta} \in V_{\alpha,}$
911    $_{\beta\in\alpha}$). For abbreviations of vegetation groups, refer to Table 2.

Figure 5: Comparison according to the type of pollen data (relative abundance, relative abundance scale and presence/absence) of $V_{\alpha,\beta}$ score (the vegetation score of pollen assemblage β for the vegetation group α) applied to pollen assemblages from dynamic depositional environments for the each vegetation group α linked with dynamic depositional environments. (1) The range of scores of pollen assemblages that belong to vegetation α ($V_{\alpha, \beta \in \alpha}$) represented by the black bars (this is the range of significant values of each vegetation α) and the range of scores of pollen assemblages that do not belong to vegetation α ($V_{\alpha, \beta \notin \alpha}$) represented by the grey bars (this is the range of non-significant values of each vegetation α). (2) For each vegetation α, the vegetation groups from which pollen assemblages β originated with $V_{\alpha, \beta}$ scores intersecting the range of significant values ($V_{\alpha, \beta} \in V_{\alpha, \beta \in \alpha}$). For abbreviations of vegetation groups, please refer to Table 2.

Figure 6: Modern range of mean temperature of warmest quarter by mean precipitation of driest quarter for pollen assemblages grouped by vegetation group (WordClim2, 1970-2000, 0.93x0.93 km, Fick and Hijmans, 2017), with 95% confidence intervals. For abbreviations of vegetation groups, please refer to Table 2.

Figure 7: Percentage of tree cover associated with pollen assemblages for each vegetation group (Global Maps, year 2008, 0.46×0.46 km, Kobayashi et al., 2017), with a confidence interval of 95%. For abbreviations of vegetation groups, please refer to Table 2.

Supplementary Online Material

Table S1: List of accepted pollen types and their synonyms.

Table S2: Pollen count with harmonized pollen types according to the table S1.

Table S3: Metadata of pollen samples.

Table S4: Climatic data, edaphic data, net primary production value, and % of tree cover for each pollen sample.

Table S5: vegetation information for each pollen sample.

Table S6: list of Value Test for pollen data in relative abundance (sheet n°1), relative abundance scale (sheet n°2), and presence/absence (sheet n°3).

Figure S1: Comparison of K-means partitions for the number of abundance classes with the simple structure index (ssi) criterion.

941   **Figure S2:** Mean Decrease Accuracy for the 38 parameters used for reconstruct our 20 k-mean groups

942   with Random Forest.