A paleothermometer for the northern Andes based on C 3 –C 4 grass phytoliths

Camilla Crifò, Juan Carlos Berrio, Arnoud Boom, Diego A Giraldo-Cañas, Laurent Bremond

A paleothermometer for the Northern Andes based on C_3-C_4 grass phytoliths

Camilla Crifò[a], Juan Carlos Berrio[b], Arnoud Boom[b], Diego A. Giraldo-Cañas[c], Laurent Bremond[a]

[a]Institut des Sciences de l'Évolution de Montpellier (ISEM), EPHE, PSL Research University, Université de Montpellier, CNRS, IRD, Place Eugène Bataillon, CC 065, 34095 Montpellier, France

[b]School of Geography, Geology and Environment, University of Leicester, University Road, Leicester LE1 7RH, UK

[c]Herbario Nacional Colombiano, Instituto de Ciencias Naturales, Universidad Nacional de Colombia, Avenida Ciudad de Quito # 55-31, Barrio Nicolás de Federmann, Bogotá D.C., Colombia

Abstract.—Grass-dominated ecosystems cover ~40% of Earth's surface with tropical grasses accounting for ~20% of global net primary productivity. C_3 (cool/temperate) and C_4 (tropical and subtropical) grass distribution is driven primarily by temperature. In this work, we used phytolith assemblages collected from vegetation plots along an elevation and temperature gradient in the Northern Andes (Colombia and Ecuador) to develop a paleothermometer for the region. To accomplish this, we created a transfer function based on the relationship between mean annual temperature (MAT) and the phytolith-based Climatic Index (1-*Ic*), which is the proportion of C_4 grass phytoliths (GSSCP) over the sum of GSSCP. To evaluate how accurately the index reflects C_4-C_3 grass abundance in vegetation plots, we compared it with semi-quantitative floristic estimates of C_4-C_3 grass abundance. To further evaluate the 1-*Ic* index as a proxy for C_4-C_3 grass abundance, we compared it with corresponding $\partial^{13}C$ values (an independent proxy for C_4-C_3 vegetation). Results indicate that 1) GSSCP assemblages correctly estimate C_4-C_3 grass abundance in vegetation plots; 2) the *Ic* index outperforms the $\partial^{13}C$ record in estimating C_4-C_3 grass abundance, even in open vegetation types; and 3) our *Ic* index–based model accurately predicts MAT. This new calibrated proxy will help improve paleotemperature reconstructions in the northern Andes since at least the emergence and spread of C_4 grasses in the region during the late Miocene.

Introduction

Grass-dominated ecosystems occupy ~40% of Earth's terrestrial surface (Gibson 2009). Tropical grasses account for ~20% of the global net primary productivity (NPP) (Lloyd and Farquhar 1994). Most of these grasses employ C_4 photosynthesis (Edwards et al. 2010, Still et al. 2003), and comprise over half of all C_4 plant species (Sage et al. 2011). C_4 photosynthesis operates through several morphological and biochemical modifications of the ancestral C_3 photosynthetic pathway resulting in reduced photorespiration and higher photosynthetic efficiency (Ehleringer and Cerling 2002). C_4 grasses tend to be advantaged and dominate under high-light conditions, and hot, seasonally arid climates, such as in tropical and subtropical grasslands and savannas –though they are not restricted to these habitats –and under low pCO_2. C_3 grasses dominate in temperate and boreal grasslands as well as in high elevation ecosystems (Boom et al. 2002, Edwards et al. 2010, Ehleringer et al. 1997, Strömberg 2011). Thus, global patterns of C_3 and C_4 grass distribution are well known (Edwards and Still 2008, Pau et al. 2013, Twiss 1992). A number of studies have shown ecological sorting of C_3 and C_4 grasses along temperature and moisture gradients (e.g., Chazdon 1978, Edwards and Still 2008, Livingstone and Clayton 1980, Rundel 1980, Teeri and Stowe 1976, Tieszen et al. 1979, Young and Young 1983). Large-scale distribution of C_3/C_4 grass seems to be best explained by the temperature crossover model (Collatz et al. 1998, Ehleringer et al. 1997, Still et al. 2003) which predicts that the switch from C_3 to C_4 grass dominance is a function of mean monthly air temperature at a given atmospheric CO_2. The temperature crossover model, which is widely accepted (but see Winslow et al. 2003), assumes that variation in grass photosynthetic pathway is the main driver of grass distribution. However, several authors have pointed out the important role played by grass evolutionary history on C_3/C_4 grass distribution (e.g., Edwards and Still 2008, Pau et al.

2013). As members of the warm-climate PACMAD clade, C_4 grasses are restricted to warmer areas. Thus, given the shared preference for warm temperatures of C_3 and C_4 PACMAD grasses, as opposed to the temperate-climate Pooideae clade, global patterns of C_3/C_4 grass distribution may be really patterns of Pooideae/PACMAD grass distribution (Edwards and Still 2008).

The spread of grasslands during the Neogene is thought to have had important consequences on faunas (Jacobs et al. 1999) and climate, including the global Si cycle (Conley and Carey 2015). Geochemical, ecological, paleontological and systematics studies investigating the drivers of C_4 grass expansion point to two main factors: water availability (aridification and/or changes in growing season precipitations) and pCO_2 levels (e.g., Cotton et al. 2016, Fox et al. 2018, Sage et al. 2018, e.g., Zhou et al. 2018). However, there is no consensus on the specific environmental dynamics that have led to the ecological success of C_4 grasses (Cotton et al. 2016). Understanding the timing and drivers of the ecological expansion of grassland ecosystems may help us better understand the factors that control both their present and future distribution. Hence, the evolution and spread of grass-dominated ecosystems during the Cenozoic has been widely investigated by paleoecologists and evolutionary biologists. Yet the grass fossil record is scant and provides limited tools to study the timing and evolution of grasses and their response to climate change (Strömberg 2011). Fossil impressions and compressions of vegetative grass structures are rare, and provide little taxonomic information. Fossilized grass inflorescences are even rarer. Lastly, grass pollen can be hardly identified beyond family level (Palmer 1976). The $\partial^{13}C$ signature of C_3 and C_4 plants is among the most common indirect proxies used to reconstruct grass and grassland evolution. In fact, differences between C_3 and C_4 photosynthesis result in differences in carbon stable isotope fractionation (Farquhar 1983), with $\partial^{13}C$ values of C_3 plants ranging from ca. -24‰ to -32‰, and $\partial^{13}C$ values of C_4 plants ranging from ca. -10‰

to -14‰ (Cerling et al. 1997b, Smith and Epstein 1971). The vegetation $\partial^{13}C$ signature is preserved in pedogenic carbonates (Quade et al. 1994), soil organic matter (Boutton 1996, Cerling et al. 1989), fossil grass pollen (Urban et al. 2010), herbivore tooth enamel (Cerling et al. 1997a) as well as in lacustrine sediments (Meyers and Ishiwatari 1993) and leaf wax *n*-alkanes (Rieley et al. 1991), and has been used extensively to reconstruct past vegetation in the absence of direct evidence. However, in lake settings, $\partial^{13}C$ interpretations are complicated by the influence of both terrestrial and aquatic plants as well as of microorganisms on the carbon isotope ratio (Meyers and Ishiwatari 1993). The carbon isotope signature of leaf wax *n*-alkanes, which are generally abundant and well preserved in sediments, especially in lacustrine settings (Castañeda and Schouten 2011), provides finer C_3-C_4 vegetation reconstructions (e.g., Boom et al. 2002) because *n*-alkanes from terrestrial plants can be differentiated from those of aquatic plants (but see Diefendorf et al. 2010). Although the $\partial^{13}C$ signature can be used to distinguish between C_3 and C_4 plants, it cannot separate C_3 grasses and trees, limiting the inferences that can be made about vegetation type and habitat openness. In fact, forest-grassland dynamics are not necessarily influenced by the same factors that control C_3 vs C_4 plant species abundance, especially in the tropics. Indeed, a number of studies have shown that open and closed-canopy systems can persist under the same climatic conditions (e.g., Aleman and Staver 2018, Bond et al. 2005, Favier et al. 2012).

In the last two decades, phytolith analysis has emerged as one of the most reliable proxies to reconstruct past vegetation, and in particular the evolution and spread of grasses and grass-dominated ecosystems (Chen et al. 2015, Cotton et al. 2012, Dunn et al. 2015, Harris et al. 2017, Loughney et al. 2019, Miller et al. 2012, Strömberg 2005, Strömberg et al. 2007a, Strömberg and McInerney 2016, Thorn 2001). Phytoliths are microscopic bodies of amorphous silica produced

by many vascular plants (Strömberg et al. 2016, Trembath-Reichert et al. 2015) at varying abundances by absorption of monosilicic acid (H_4O_4Si) by plant roots, and subsequent polymerization in plant cells and tissues (Piperno 1988). Because of their relatively inert composition (silica, or opal A), they expand the range of depositional environments that can be studied to include well-oxidized sediments where organic matter such as pollen is typically not preserved. Phytoliths primarily represent local vegetation (Piperno 1988), compared to pollen which is sourced from local and regional vegetation (Prentice 1985) , thus allowing finer landscape-level vegetation reconstructions. For instance, the relative proportions of phytoliths produced by grasses vs woody plants is a proxy for canopy structure (Aleman et al. 2012, Bremond et al. 2005a, Strömberg et al. 2007b). Lastly, phytoliths produced by grasses in grass silica short cells (GSSCP) are highly diagnostic allowing for the discrimination of major grass clades (subfamily level), and functional types (open vs. closed habitat grasses; C_3 vs C_4 photosynthesis), providing direct insights into grass community composition and ecology (e.g., Alexandre et al. 1997, Barboni et al. 1999, McInerney et al. 2016). Specifically, GSSCP produced by the tropical clade PACMAD (i.e., Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, Danthonioideae), which uses predominantly the C_4 photosynthetic pathway, can be distinguished from those produced by C_3 grasses in the BEP clade (Bambusoideae, Ehrhartoideae, and Pooideae), and by C_3 early-diverging grasses (Anomochlooideae, Pharoideae, Puelioideae). Hence, phytolith analysis uses the proportion of C_4/C_3 GSSCP to estimate C_3-C_4 grass abundance from fossil phytolith assemblages. Because C_3-C_4 grass abundance is influenced by climate, phytolith analysis can provide new types of paleoenvironmental reconstruction and improve paleoclimatic inferences. Indeed, the proportion of C_3/C_4 GSSCP has been shown to be strongly

related to temperature and consequently represents a potential paleotemperature proxy (Bremond et al. 2008b).

Here we present a grass phytolith-based paleotemperature proxy from across a Neotropical elevation gradient (~300 to 4300 meters above sea level) in the Northern Andes. The influence of climate and orogeny on the evolution of high elevation, grass-dominated ecosystems in the Andes is intensely debated (e.g., Kirschner and Hoorn 2020, Simpson and Todzia 1990). The use of a phytolith-based paleotemperature proxy in parallel with traditional phytolith analysis and through comparison with other paleovegetation and paleoclimate proxies in Andean geological records could help shed light on the timing of grasses and grassland expansion into Neotropical high elevation ecosystems and their relationship with climate variability. Several modern calibration studies have resulted in the development of phytolith-based climatic and vegetation indices in different ecosystems (Aleman et al. 2012, Alexandre et al. 1997, Bremond et al. 2005a, Bremond et al. 2005b, 2008a, Fredlund and Tieszen 1994, Parmenter and Folger 1974). However, many of these studies suffer from methodological limitations due to the lack of control on taphonomic processes, spanning from vegetation production to sediment deposition. In addition, only a limited number of studies couple phytolith analysis to other C_3-C_4 plant archives for interproxy validation (e.g., Lupien et al. 2021). Furthermore, modern reference collections from plants or soil surface samples along C_3-C_4 gradients are restricted to a few geographic areas and to date, none have been produced in the Neotropics. Several studies have investigated the relationship between grass phytolith assemblages from soils and grass communities in different ecosystems but have led to inconsistent results. For instance, Fredlund and Tieszen (1994) found that the GSSCP assemblage composition of soils from fifteen modern grasslands in the North America Great Plains broadly reflects the regional east-west moisture

147 gradient (with C_4 GSSCP being associated to drier sites) but also noted the extra-local and
148 regional nature of the phytolith signal. Kerns et al. (2001) found only a weak correspondence
149 between GSSCP assemblages and C_4-C_3 grass species abundance in three canopy types within a
150 *Pinus ponderosa* forest in North America. In contrast, Bremond et al. (2008b) found a positive
151 relationship between the proportion of C_3/C_4 grasses and the proportion of C_3/C_4 GSSCP along
152 an altitudinal gradient In Kenya. A similar trend was described by An et al. (2015) in the
153 southern Himalaya. More recently, Biswas et al. (2021) developed a temperature calibration
154 based on the phytolith climatic index (*Ic*)(Bremond et al. 2008b) in the eastern Himalayan
155 sectors.
156 The objectives of this study are 1) to assess the accuracy of grass phytolith assemblages
157 in soils as a tool for reconstructing the C_3/C_4 grass ratio in the tropical Andes; 2) to compare this
158 ratio with that of the $\partial^{13}C$ record; and 3) to develop a transfer function based on modern
159 phytolith assemblages and temperature to estimate paleotemperature from fossil phytolith
160 assemblages. To complete these objectives, floristic data, along with soil surface samples for
161 phytolith and $\partial^{13}C$ analysis were collected and compared along an altitudinal and aridity gradient
162 in Colombia along the northeastern Andean Cordillera. Based on this work, we propose a new
163 phytolith-based transfer function to reconstruct paleotemperatures, and we discuss the prospects
164 and limitations of this new proxy to improve deep-time paleoclimate inferences.
165
166 **Material and Method**
167 *Sample and data collection.*—A total of 35 samples for phytolith and $\partial^{13}C$ analysis were
168 collected along two elevational transects in the eastern Andean Cordillera of Colombia (**Fig. 1**).
169 These transects follow a SW-NE orientation and span from 1621 to 3545 and 331 to 2858 meters

above sea level (m a.s.l.), respectively. This set of samples was complemented by five additional samples spanning a higher elevation range between 3929 and 4303 m a.s.l., and previously collected from a transect oriented SW-SE in the Andean Cordillera of Ecuador (Ledru et al. 2013) (**Fig.1**). Samples from Colombia encompass several vegetation zones from dry tropical forest to lower and upper montane forest (Hooghiemstra and Van der Hammen 2004). Samples from Ecuador corresponds to Paramo vegetation (**Fig. 1**). All samples were collected from the soil subsurface below the litter layer. Each homogenized soil sample consists of about a dozen subsamples (measuring ~1 cm^3 each) collected randomly within a plot measuring about 100 m^2, and then mixed together (Bremond et al. 2005). Plots were installed within open vegetation areas. Forested areas in closed- or mosaic vegetation types were avoided, in order to minimize the influence of woody vegetation (mainly C_3) on the soil $\partial^{13}C$ signal, as well as to maximize the amount of grass silica short cell phytoliths (GSSCP) present in the samples. Furthermore, in order to limit potential biases in the composition of the GSSCP assemblages linked to the presence of exotic grass species, disturbed areas were also avoided. These mainly include pastures, where non-native grass species might have been introduced. A semi-quantitative floristic survey was conducted at each sample site to estimate the relative abundance of C_3 vs. C_4 grasses at each site (see **Supplementary Table 1** and **Supplementary Table 2**). In each plot two 5x5 meter subplots were installed; all grass species present in each subplot were identified and their percent cover was estimated visually. Percent cover values for each species at each site were calculated by averaging the two subplots estimations. The relative abundance of C_3 and C_4 grasses at each site corresponds to the sum of the percent cover off all C_3 and C_4, respectively. The dominant vegetation stratum was also noted for each site. Based on this information, sampling sites were classified into six vegetation types (C_3 grassland, C_4 grassland, C_3 shrubland,

193 C_4 shrubland, C_3 grass forest, and C_4 grass forest). Geographic coordinates and elevation of each
194 sample site were recorded in the field; mean annual precipitation (MAP) and mean annual
195 temperature (MAT) data were gathered from the WorldClim data website (Supplementary Table
196 3). We extracted regional C_3 and C_4 grass distribution, and MAT data from a compilation of
197 grass species richness data from four intertropical countries in South America (Bolivia,
198 Colombia, Ecuador and Peru) and a compilation of climatic variables obtained from the
199 Worldclim 1.4 1-km database by Bremond et al. (2012) (in Supplementary Table 4).
200 *Stable isotope analysis ($\partial^{13}C$).*—Stable carbon isotope analyses were performed on bulk
201 soil from the 35 soil samples collected in Colombia. These analyses were not performed on the
202 six samples collected in Ecuador due to the use of a different sampling protocol for their
203 collection (Ledru et al. 2013). Approximately 5 grams samples were pre-treated with 10%
204 hydrochloric acid (HCl), to remove carbonates and then rinsed in deionized water until neutral,
205 freeze dried, and ball milled. The stable isotope ratio of total organic carbon ($\delta^{13}C_{TOC}$) was
206 measured using a SerCon ANCA GSL elemental analyzer interfaced to a Hydra 20-20
207 continuous flow isotope ratio mass spectrometer. Carbon isotopic values ($\partial^{13}C$) are expressed in
208 per mil with respect to the international V-PDB (Vienna-Pee-Dee-Belemnite) standard (see
209 Supplementary Table 2). The precision/standard deviation is ± 0.16‰.
210 *Phytolith extraction.*—For each soil sample we used 3-5 grams for phytolith extraction.
211 We used the protocol described by Aleman (2013) which includes treatment with HCl for
212 carbonate removal, potassium hydroxide for humic acid removal, hydrogen peroxide for removal
213 of organic material, and sodium citrate and sodium dithionite for removal of iron oxides.
214 We removed soil particles >200 μm by sieving, and deflocculated the remaining soil particles
215 using sodium hexametaphosphate. We used heavy liquid floatation to isolate the biogenic silica

(sponge spicules, diatoms, chrysophyte cysts, and phytoliths) using a heavy liquid solution of zinc bromide, hydrochloric acid and water prepared at a specific gravity of 2.38 g/mL.

Phytolith classification and counting.—Each dry bio-silica sample was mounted for microscopy using immersion oil to allow for rotation during identification and counting. Phytolith morphotypes were classified according to the International Code for Phytolith Nomenclature (ICPN 2.0; Neumann et al. 2019) into 23 morphotypes divided into the following classes: 1) grass silica short cell phytoliths (GSSCP), exclusively produced by and diagnostic of several Poaceae (grass) subfamilies; 2) phytoliths diagnostic of woody dicotyledons; 3) phytoliths distinctive of the monocot family Arecaceae (palms) (SPHEROID ECHINATE), but also possibly diagnostic of the family Bromeliaceae (bromeliads) and of the order Zingiberales (e.g., spiral gingers) (i.e., when the phytolith diameter is ≤5µm; see Benvenuto et al. 2015, Crifò and Strömberg 2021); 4) other phytoliths diagnostic of the Poaceae family (including forms produced mainly but not exclusively by the Poaceae family as well as other morphotypes belonging to the family Poaceae but whose production is influenced by environmental factors such as water availability (e.g., Bremond et al. 2005b, Madella et al. 2009); and 5) other non-diagnostic forms. Morphotype names, corresponding ICPN names and codes, and other information are provided in Table 1. Details regarding GSSCP morphotypes assignement to C_3 or C_4 grasses are given in the Appendix. Morphotype used in our classification are illustrated in Figure 2. Each phytolith count includes all morphotypes present in the assemblage from the four classes mentioned above, and a minimum number (n) of 200 GSSCP diagnostic morphotypes (except for one sample with n =139) for statistical significance (Pearsall 2000, Strömberg 2009). Phytolith counts are provided in Supplementary Table 2.

Statistical Analysis.—All statistical analyses were performed in R version 4.0.3 (R Development Core Team 2019). All R codes are provided in the Appendix.

The C_3/C_4 GSSCP ratio, corresponding to the climate index Ic (Bremond et al. 2008b) was calculated as follow:

$$Ic = \frac{\text{"Stipa type bilobate"} + \text{"trapeziform trilobate"} + \text{"trapeziform polylobate"} + \text{"collapsed saddle"} + \text{"rondel"}}{\text{"short bilobate"} + \text{"long bilobate"} + \text{"short saddle"} + \text{"long saddle"} + \text{"cross"}}$$

To assess robustness of the calculated *Ic* values, a 95% confidence intervals (CI) was constructed using bootstrap analysis of 10,000 replicates (Strömberg 2009). Linear regression was used to test the significance of the relationship between the proportion of C_4 grass phytoliths over the total GSSCP count (i.e., $1 - Ic$) and the proportion of C_4 grasses at our sites. We used a generalized linear model (R function *glm*) with a logit link function to build a transfer function for MAT. The logit link function converts the phytolith C_4 proportion Y to the logarithm of the ratio $\frac{Y}{(1-Y)}$ in order to make the data suitable for regression. Thus, we fitted the following equation to our data:

$$log\left(\frac{Y_i}{1 - Y_i}\right) = \beta_0 + \beta_1 T \ + \ w$$

Where, β values correspond to the intercepts and T corresponds to MAT. To account for potential error due to differences in sample size, we assigned a weight (w) to samples according to their confidence intervals so that phytolith assemblages with wider CI have smaller weight in the regression fit than assemblages with narrower CI (see Harris et al. 2017). w was calculated as follow:

$$w = \frac{1}{\left(\left(Y_i^{upper} - Y_i\right) + 0.1\right)}$$

Where Y_i corresponds to $1 - Ic$ and Y_i^{upper} is the upper bound (upper confidence interval) of Y_i. To deal with cases where $Y_i = 0$, which lead to $Y_i^{upper} = Y_i$, and $w = 0$, we added an additional small weight (0.1) into the denominator of the equation.

Differences in GSSCP compositions between sites were evaluated using nonmetric multidimensional scaling (NMDS, R function *metaMDS*) on the GSSCP morphotype relative abundance matrix (see Crifò et al. 2021). We used tests of stress value randomization (R function *nmds.monte*) to determine the minimum number of dimensions required to capture the variation in the data. Further, we used correspondence analysis (CA; R function *cca*) to evaluate the influence of environmental gradients (i.e., MAT, MAP, and elevation) on GSSCP compositional differences between sites. CA was carried out using the GSSCP morphotype relative abundance data, and the log-transformed environmental variables (elevation, MAT, and MAP). For visual comparison of the stable carbon isotope and phytolith (1-*Ic*) proxies, $\partial^{13}C$ values obtained from the samples were plotted together with C_4 proportions against MAT values. We used linear regression to test for a relationship between the proportion of C_4 GSSCP and $\partial^{13}C$ values of the samples, and between the proportion of C_4 GSSCP in the samples and the proportion of C_4 grasses recorded by the floristic survey.

Results

C_4 grass proportion from floristic survey and phytolith counts along elevation and MAT gradients.—The proportion of C_4 grasses shows a positive relationship with MAT and a negative relationship with elevation ($p << 0.001$; $R^2 = 0.68$ and 0.7 respectively) (**Table 2**). Linear regression analysis also indicates a positive relationship between the C_4 grass proportion from the floristic survey, and the C_4 grass proportion estimated from phytolith assemblages (i.e., $1 -$

Ic) ($p << 0.001$: $R^2 = 0.71$) (**Table 2**). In addition, the C_4/C_3 GSSCP proportion is positively related to MAT and negatively related to elevation ($p << 0.001$, $R^2 = 0.88$ for both regressions) (**Table 3**).

Influence of environmental gradients on the distribution of GSSCP morphotypes across sites.—NMDS ordination of the sites performed on GSSCP assemblage composition adequately represent the data using two dimensions (NMDS, stress value = 0.14; test of stress value randomization, $p < 0.01$). Phytolith morphotype counts are provided in Supplementary Table 2. On NMDS axis 1 (**Fig. 3**), positive loadings are represented by "regular" (RON_1) "truncated", (RON_3), and "keeled" (RON_4) RONDEL morphotypes ($p < 0.001$), "trapeziform polylobate" (CRE_2) CRENATE morphotypes ($p < 0.001$), and by "*Stipa*-type" (BIL_1) BILOBATE morphotypes ($p < 0.05$) while negative loadings are represented by "short" (BIL_2) and "long" (BIL_3) BILOBATE and CROSS (CRO) morphotypes ($p < 0.001$). On NMDS axis 2, positive loadings are represented by the "short" (SAD_2) and "long" (SAD_3) SADDLE morphotypes ($p < 0.001$) while negative loadings are represented by "trapeziform trilobate" (CRE_1) CRENATE morphotypes ($p < 0.001$), "short" (BIL_2) and "long" (BIL_3) BILOBATE morphotypes ($p < 0.001$), as well as by CROSS (CRO) morphotypes ($p < 0.001$). Sites dominated by C_3 grasses tend to group towards the positive end of NMDS axis 1 as opposed to C_4 grass-dominated sites, which are grouped towards the negative end. Among C_4 grass-dominated sites, C_4 grasslands and C_4 shrublands are spread along NMDS axis 2 while C_4 grass forests tend to be grouped towards the negative side of the axis.

CA ordination of the sites performed on the site GSSCP assemblage composition (counts) by environmental variables (MAP, MAT, elevation) shows a similar pattern of site distribution (**Fig. 4**). The first two axes of the CA ordination account for respectively 43% and 15% of the

total variation. CA axis 1 describes a negative MAT gradient ($p < 0.001$) opposite to a positive elevation gradient ($p < 0.01$). C_4-dominated sites are distributed towards the negative end of this axis as opposed to C_3-dominated sites which are distributed towards the positive end. Most C_4-dominated sites tend to group towards the positive end of CA axis 2 with the exception of five sites which appear to form an isolated cluster towards the negative end of CA axis 2. The positive loading of MAP on this axis is not significant ($p > 0.05$).

Predicted temperature as a function of C_4 phytolith and C_4 grass proportion.—Logistic regression shows that the C_4 GSSCP proportion is a good predictor of MAT. C_4 phytolith and grass proportions increase with MAT following a sigmoidal growth pattern ($p << 0.001$) (**Fig. 5**). Accordingly, at lower temperatures (i.e., MAT< 12 ºC) and high elevation the C_4 GSSCP proportion is low, and close to zero in several assemblages (**Fig. 6**). These assemblages are dominated by RONDEL, "*Stipa*-type" BILOBATE, and CRENATE phytolith morphotypes (**Fig. 3**) which are assigned to grass subfamilies in the C_3 BEP clade (Bambusoideae, Ehrhartoideae, and Pooideae). The proportion of C_4 phytoliths sharply increases at higher temperatures. SADDLE, CROSS, and "short" and "long" BILOBATE phytolith morphotypes, assigned to the C_4 clade PACMAD, dominate these assemblages (**Fig. 3**, and **6**). However, unlike C_4 GSSCP, C_3 GSSCP are present across the entire temperature (and elevation) range.

Relationship between C_4 GSSCP proportion and carbon isotope ratio.— The $\partial^{13}C$ values of the samples span -17.04‰ to -26.59 ‰ (**Fig. 5B**), corresponding mainly to intermediate C_3-C_4, and C_3 vegetation signatures (Cerling et al. 1997b). Linear regression indicates no significant relationship between the carbon $\partial^{13}C$ values and the proportion of C_4 GSSCP in the samples ($p >> 0.05$) (**Table 3**). Likewise, sample $\partial^{13}C$ values have no significant relationship with

environmental factors (elevation, MAT, and MAP) ($p >> 0.05$) but are weakly related to the C_4 grass proportion of the sites ($p < 0.01$; R^2=0.21) (**Table 2**).

Discussion

Relationship between the Ic index and C3-C4 grass distribution.—Overall, the grass floristic survey and phytolith counts reflect C_3-C_4 grass distribution along elevation and temperature gradients, as previously documented in the region (Bremond et al. 2012). However, MAT and elevation explain most of the variation in 1- *Ic* phytolith index but explain variations in C_4/C_3 grass proportions from the floristic surveys only to a lesser extent. Further, at least some of the variation in 1- *Ic* seems to not be explained by variation in C_4/ C_3 grass proportions from the floristic surveys. Floristic data indicate higher (lower) C_4/ C_3 grass proportions at high (low) temperatures compared to the phytolith data. These inconsistencies might be explained for several reasons. First, the floristic survey provides instantaneous estimates of C_4/ C_3 grass proportion on a small spatial scale. The C_4/ C_3 grass proportion can vary in relation to the botanical survey season and to local conditions such as microclimate and topography, and in response to short term climate fluctuations. Thus, it might not be representative of the vegetation on a wide spatial and temporal scale; instead, it might reflect super-local vegetation and environmental conditions. On the contrary, C_4/ C_3 phytolith abundance from soil surface samples does not represent a real time snapshot of the vegetation because a phytolith assemblage typically reflect tens to hundreds of years of vegetation occurrence (Song et al. 2016). Further, our floristic survey was conducted in a semi-quantitative way and only provides estimates of relative C_4/ C_3 grass cover. Error associated with these estimates could explain at least in part the mismatch between the floristic survey data and the phytolith signal. Furthermore, although

phytolith production should be similar among C_3 and C_4 grasses growing in the same habitat and under identical environmental conditions (Brightly et al. 2020), the relative proportion of different diagnostic phytolith morphotypes produced for a given plant biomass varies among species and might result in C_3/C_4 proportion under- or overestimation. In addition, coarse classification of RONDEL morphotypes (which are all grouped at the numerator of the *Ic* equation, as diagnostic of C_3 grasses) might result in potential underestimation of C_4 grass abundance (which also produce these morphotypes). However, the error associated to potentially erroneous RONDEL classification is unknown because data on the production of different morphotypes by different species are not available (see also Appendix). Finally, the influence of taphonomic biases on C_3/C_4 GSSCP proportion can be reasonably excluded because the phytolith assemblages extracted from the samples are all finely preserved and do not present evidence of differential preservation or dissolution.

The C_4/C_3 GSSCP proportion as a paleotemperature proxy.—Our model, based on logistic regression, performs better than models based on linear regression (i.e., Biswas et al. 2021), and is in line with the temperature crossover model (Collatz et al. 1998, Ehleringer et al. 1997, Still et al. 2003) and with previous work (Bremond et al. 2012). Indeed, the relationship between C_4/C_3 GSSCP proportion and temperature is not linear, due to the combined effect of the presence of C_3 grasses across the entire elevation and temperature range, and of a crossover temperature (at which the proportion of C_4 grasses and thus of C_4 GSSCP, equals that of C_3 grasses and thus of C_3 GGSCP). This results in a sigmoidal "growth" pattern of the C_4/C_3 GSSCP proportion which is best described by logistic regression. Thus, the transfer function developed herein can be used to obtain paleotemperature estimates in the fossil record of the Andean region. Nevertheless, the accuracy of our model also relies on the fact that phytolith assemblages constitute a local

vegetation signal and on our choice of restricting soil sampling to open areas. In fact, in mountain forests, grass species belonging to the subfamily Panicoideae (which has both C_3 and C_4 representatives), and inhabiting forest understorey, often use C_3 photosynthesis. At intermediate elevation (where mountain forest occurs), their presence can therefore bias the C_4/C_3 proportion towards higher values. While our sampling strategy enabled us to limit such bias associated to the presence of C_3 grasses of the subfamily Panicoideae (by sampling preferentially in open-vegetation areas), it might complicate the interpretation and subsequent use of the transfer function to estimate MAT (which might be potentially overestimated) from fossil phytolith assemblages, where the vegetation type is unknown. As a consequence, for correct application of this method for paleotemperature reconstructions, fossil phytolith assemblages should be first studied in detail (including forest indicator phytoliths) in order to select preferentially those representing open-canopy vegetation. Phytolith-based, vegetation indices commonly used to asses canopy cover, such as the *D/P* index (the sum of SPHEROID ORNATE phytoliths, produced by many tropical ligneous dicotyledonous, over the sum of Poaceae phytoliths; Alexandre et al. 1997), and the *FI-t* ratio (the sum of forest indicator phytoliths morphotypes over the sum of all morphotypes; Strömberg et al. 2007b) should be used.

Relationship between the C_4/C_3 *GSSCP proportion and the carbon isotope ratio* ($\partial^{13}C$).—Soil carbon isotopic ratios ($\partial^{13}C$), are not consistent with the *Ic* index and the C_3-C_4 grass composition from floristic surveys along elevation and temperature gradients. In fact, most $\partial^{13}C$ values indicate C_3 or mixed C_3-C_4 vegetation, while none of the samples is characterized by $\partial^{13}C$ values indicative of C_4-dominated vegetation. This discrepancy can be explained by several factors. First, the soil $\partial^{13}C$ signature represents a mixed vegetation signal (which reflects the presence of grasses as well as forbs, shrubs and trees) rather than a pure grass signal. Because the

relationship between the prevalence of C_3 vs C_4 photosynthesis and temperature differs in grasses and trees, a certain degree of inconsistency between C_3-C_4 grass abundance from floristic surveys and soil $\partial^{13}C$ values, as well as between temperature and $\partial^{13}C$ values is therefore expected. However, as mentioned above, in order to obtain a $\partial^{13}C$ signal representative mostly of the grass component of the vegetation, sample collection was deliberately conducted in open-canopy areas only. Thus, other factors must explain the poor match between $\partial^{13}C$, C_4/C_3 GSSCP proportion, and grass community composition. C_3 and C_4 plants display $\partial^{13}C$ spectra that are quite wide and result in a bimodal distribution of $\partial^{13}C$ values (Cerling et al. 1997b) which renders interpretations of C_4 grass abundance based on $\partial^{13}C$ values particularly complex. Additionally, time-averaging of the soil might result in $\partial^{13}C$ values representing a mix of past and present vegetation, and less negative $\partial^{13}C$ values might record past vegetation shifts between closed and open-canopy in sites that today are dominated by C_4 grasses. However, biases related to time-averaging would have also affected the phytolith record. We found no evidence for such bias in the phytolith assemblages. Nonetheless, it is possible that the shorter turnover time of phytoliths compared to the $\partial^{13}C$ counteracts the effect of time-averaging on phytolith assemblages but not on the carbon isotopic ratio of the samples (Aleman et al. 2012). Finally, differences between the C_3-C_4 grass signal of $\partial^{13}C$ and phytoliths may also be due to differential degradation of organic matter. Higher decomposition rates have been observed for C_4 organic material compared to C_3 organic material in mixed C_3-C_4 soils (Wynn and Bird 2007) and could explain less negative $\partial^{13}C$ ratios at least in mixed C_3-C_4 vegetation areas (Cotton et al. 2012).

Applications to the deep-time fossil record.—The application of the new proxy proposed herein is of particular interest in tropical montane, high elevation ecosystems where available paleotemperature proxies tend to be scarce –at lower elevation, the *Ic* index is less accurate

419 due to the abundance of several C_3 tropical grasses in closed-canopy tropical ecosystems.
420 The use of the *Ic* index to reconstruct paleotemperature is independent on whether the
421 paleoelevation of a fossil site is known. However, paleoelevation data might improve
422 paleoclimatic interpretations based on the *Ic* index. For instance, a change in temperature
423 in the absence of a shift in elevation might be the result of global climate change as opposed
424 to local temperature decrease which might be driven by mountain building. In the Andean
425 region, the application of the *Ic* index to the deep-time phytolith record has the potential to
426 advance our understanding of the expansion of grasses and grass-dominated ecosystems across
427 the region. However, there are some caveats to consider. First, the application of our model to
428 the deep-time record assumes that the relationship between C_3-C_4 grass distribution and MAT
429 observed today also occurred in the past. Indeed, other factors, such as aridity and atmospheric
430 pCO_2, are known to influence C_3-C_4 grass distribution. While increasing evidence points to
431 relatively stable pCO_2 level during at least the last 7 Ma, and possibly since 23 Ma (Cui et al.
432 2020), the uplift of the Andes is thought to have driven a progressive shift towards aridification,
433 the timing and magnitude of which are still debated (Martínez et al. 2020). Caution is therefore
434 required, and phytolith-based inferences should be coupled with other proxies. Second, the
435 temporal range to which the proposed proxy can be applied is dependent on the timing of the
436 expansion of C_4 grasses and C_4-dominated habitats. Although the use of the C_4 photosynthetic
437 pathway likely developed during the Oligocene (Edwards et al. 2010) the expansion of C_4
438 grasses has not been documented until the late Miocene, starting at around 8 Ma (Cerling et al.
439 1997b, Strömberg 2011). In South America the grass fossil record is scarce, and most of our
440 knowledge on the expansion of grasslands comes from Patagonia (Strömberg 2011) where C_3
441 grasses remained dominant at least until the early Miocene (ca. 20 Ma) whereas grassland

442 expansion did not take place until the early-late Miocene (ca. 10 Ma) (Barreda and Palazzesi
443 2007, Dunn et al. 2015, Strömberg et al. 2013). In northern South America, it has been
444 hypothesized that the retreat of the Pebas system and the establishment of the Amazon drainage
445 basin around 9 Ma (Hoorn et al. 2010) favored the expansion of grasses in the newly established
446 ecosystems of the Andean slopes and floodplains. The expansion of grasses in Andean
447 ecosystems would have been further favored by Andean orogenesis and subsequent climate
448 change (Kirschner and Hoorn 2020) creating new ecological niches. By the Plio-Pleistocene the
449 Andes had reached near-modern elevation (Garzione et al. 2017, Gregory-Wodzicki 2000) and
450 the first Paramo and Puna-like ecosystems are reported around 5 Ma (Bermúdez et al. 2017,
451 Martínez et al. 2020). No data recording the occurrence of C_4 grasses in these ecosystems are
452 currently available although work is in progress at least in the Bolivian and Peruvian Altiplano
453 (e.g., Crifò C. 2021 personal communication; Saylor B. 2021 personal communication). Thus,
454 we argue that phytolith based paleotemperature reconstructions in the region might be reasonably
455 made onwards from at least the late-Miocene-early Pliocene. Finally, local calibrations of the *Ic*
456 index, may allow its application to montane regions across a wider geographic range
457

458 **Conclusion**

459 The relationship between C_3-C_4 grass distribution and climate has long been studied.
460 Likewise, in the last two decades the notion that phytoliths provide accurate reconstruction of
461 grass community composition has been widely recognized. However, only a handful of studies
462 have focused on grass phytoliths as a tool for reconstructing paleoclimate quantitatively. Our
463 study is the first to use the proportion of C_4/C_3 GSSCP in soil samples collected along an
464 elevation gradient to develop a paleothermometer in the Andean region. Our data also support

the notion that grass phytoliths represent grass community composition, and in particular the proportion of C_3 versus C_4 grasses, more accurately than indirect proxies such as the carbon isotope ratio of soil organic matter. Moreover, the use of phytoliths presents several advantages over other direct and indirect grass vegetation proxies, which are either rarely available in the fossil record (e.g., grass macrofossils, pollen grains, or leaf cuticles), or subject to many biases (e.g., leaf wax n-alkanes, $\partial^{13}C$ of soil carbonates, and herbivore tooth enamel).

Applications of this new tool include paleoclimatic and paleoenvironmental reconstructions in the Andean region from at least the late Miocene-early Pliocene on. Coupling this new method with the analysis of whole phytolith assemblages and with other proxies will improve paleoclimatic and paleoenvironmental reconstructions in the Andes, potentially contributing a better understanding on the complex dynamics involving tectonics, climate, and vegetation that shaped most modern Andean ecosystems.

Based on the strong, inverse relationship between temperature and elevation, future work should focus on the development of the *Ic* index as a tool for paleoaltimetry. However, this requires accurate calibrations. The link between C_3/C_4 grass distribution and elevation is only indirect and relies on the relationship between temperature and elevation. Thus, changes in regional and temporal factors which might influence temperature must be taken into account. For instance, calibrations of the IC index as a paleoaltimetry proxy, should include modeling C_3/C_4 grass distribution along temperature and elevation gradients under varying pCO_2

Acknowledgements

We thank Marie-Pierre Ledru for providing Paramo samples from Ecuador. This work has been initiated during an Intra-European Marie Curie Fellowships of LB (MEIF-CT-2005-024625).

References

Aleman, J. C., B. Leys, R. Apema, I. Bentaleb, M. A. Dubois, B. Lamba, J. Lebamba, C. Martin, A. Ngomanda, L. Truc, J.-M. Yangakola, C. Favier, L. Bremond, and K. Woods. 2012. Reconstructing savanna tree cover from pollen, phytoliths and stable carbon isotopes. Journal of Vegetation Science 23(1):187-197.

Aleman, J. C., A. Saint-Jean, B. Leys, C. Carcaillet, C. Favier, and L. Bremond. 2013. Estimating phytolith influx in lake sediments. Quaternary Research 80(2):341-347.

Aleman, J. C., and A. C. Staver. 2018. Spatial patterns in the global distributions of savanna and forest. Global Ecology and Biogeography 27(7):792-803.

Alexandre, A., J.-D. Meunier, A.-M. Lezine, A. Vincens, and D. Schwartz. 1997. Phytoliths: indicators of grassland dynamics during the late Holocene in intertropical Africa. Palaeogeography, Palaeoclimatology, Palaeoecology 136(1-4):213-229.

An, X., H. Lu, and G. Chu. 2015. Surface soil phytoliths as vegetation and altitude indicators: a study from the southern Himalaya. Scientific reports 5:15523.

Barboni, D., R. Bonnefille, A. Alexandre, and J.-D. Meunier. 1999. Phytoliths as paleoenvironmental indicators, West sid Middle Awash Valley, Ethiopia. Palaeogeography Palaeoclimatology Palaeoecology 152:87-100.

Barreda, V., and L. Palazzesi. 2007. Patagonian vegetation turnovers during the Paleogene-early Neogene: origin of arid-adapted floras. The Botanical Review 73(1):31-50.

510 Benvenuto, M. L., M. FernÁNdez Honaine, M. L. Osterrieth, and E. Morel. 2015. Differentiation
511 of globular phytoliths in Arecaceae and other monocotyledons: morphological
512 description for paleobotanical application. Turkish Journal of Botany 39:341-353.
513 Bermúdez, M. A., C. Hoorn, M. Bernet, E. Carrillo, P. A. van der Beek, J. I. Garver, J. L. Mora,
514 and K. Mehrkian. 2017. The detrital record of late-Miocene to Pliocene surface uplift and
515 exhumation of the Venezuelan Andes in the Maracaibo and Barinas foreland basins.
516 Basin Research 29(S1):370-395.
517 Biswas, O., R. Ghosh, S. Agrawal, P. Morthekai, D. K. Paruya, B. Mukherjee, M. Bera, and S.
518 Bera. 2021. A comprehensive calibrated phytolith based climatic index from the
519 Himalaya and its application in palaeotemperature reconstruction. Science of The Total
520 Environment 750:142280.
521 Bond, W. J., F. I. Woodward, and G. F. Midgley. 2005. The global distribution of ecosystems in
522 a world without fire. New Phytologist 165(2):525-538.
523 Boom, A., R. Marchant, H. Hooghiemstra, and J. S. Sinninghe Damsté. 2002. CO2- and
524 temperature-controlled altitudinal shifts of C4- and C3-dominated grasslands allow
525 reconstruction of palaeoatmospheric pCO2. Palaeogeography, Palaeoclimatology,
526 Palaeoecology 177(1):151-168.
527 Boutton, T. W. 1996. Stable carbon isotope ratios of soil organic matter and their use as
528 indicators of vegetation and climate change. Mass spectrometry of soils.:47-82.
529 Bremond, L., A. Alexandre, C. Hély, and J. Guiot. 2005a. A phytolith index as a proxy of tree
530 cover density in tropical areas: calibration with Leaf Area Index along a forest–savanna
531 transect in southeastern Cameroon. Global and Planetary Change 45(4):277-293.

532 Bremond, L., A. Alexandre, O. Peyron, and J. Guiot. 2005b. Grass water stress estimated from
533 phytoliths in West Africa. Journal of Biogeography 32(2):311-327.
534 Bremond, L., A. Alexandre, O. Peyron, and J. Guiot. 2008a. Definition of grassland biomes from
535 phytoliths in West Africa. Journal of Biogeography 35(11):2039-2048.
536 Bremond, L., A. Alexandre, M. J. Wooller, C. Hély, D. Williamson, P. A. Schäfer, A. Majule,
537 and J. Guiot. 2008b. Phytolith indices as proxies of grass subfamilies on East African
538 tropical mountains. Global and Planetary Change 61(3-4):209-224.
539 Bremond, L., A. Boom, and C. Favier. 2012. Neotropical C3/C4 grass distributions - present,
540 past and future. Global Change Biology 18(7):2324-2334.
541 Brightly, W. H., S. E. Hartley, C. P. Osborne, K. J. Simpson, and C. A. E. Strömberg. 2020.
542 High silicon concentrations in grasses are linked to environmental conditions and not
543 associated with C4 photosynthesis. Global Change Biology 26(12):7128-7143.
544 Castañeda, I. S., and S. Schouten. 2011. A review of molecular organic proxies for examining
545 modern and ancient lacustrine environments. Quaternary Science Reviews 30(21):2851-
546 2891.
547 Cerling, T. E., J. M. Harris, S. H. Ambrose, M. G. Leakey, and N. Solounias. 1997a. Dietary and
548 environmental reconstruction with stable isotope analyses of herbivore tooth enamel from
549 the Miocene locality of Fort Ternan, Kenya. Journal of Human Evolution 33(6):635-650.
550 Cerling, T. E., J. M. Harris, B. J. MacFadden, M. G. Leakey, J. Quade, V. Eisenmann, and J. R.
551 Ehleringer. 1997b. Global vegetation change through the Miocene/Pliocene boundary.
552 Nature 389(6647):153-158.
553 Cerling, T. E., J. Quade, Y. Wang, and J. R. Bowman. 1989. Carbon isotopes in soils and
554 palaeosols as ecology and palaeoecology indicators. Nature 341(6238):138-139.

555 Chazdon, R. L. 1978. Ecological aspects of the distribution of C4 grasses in selected habitats of
556 Costa Rica. Biotropica:265-269.
557 Chen, S. T., S. Y. Smith, N. D. Sheldon, and C. A. E. Strömberg. 2015. Regional-scale
558 variability in the spread of grasslands in the late Miocene. Palaeogeography,
559 Palaeoclimatology, Palaeoecology 437:42-52.
560 Collatz, G. J., J. A. Berry, and J. S. Clark. 1998. Effects of climate and atmospheric CO2 partial
561 pressure on the global distribution of C4 grasses: present, past, and future. Oecologia
562 114(4):441-454.
563 Conley, D. J., and J. C. Carey. 2015. Silica cycling over geologic time. Nature Geoscience
564 8(6):431-432.
565 Cotton, J. M., T. E. Cerling, K. A. Hoppe, T. M. Mosier, and C. J. Still. 2016. Climate, CO2, and
566 the history of North American grasses since the Last Glacial Maximum. Science
567 Advances 2(3):e1501346.
568 Cotton, J. M., N. D. Sheldon, and C. A. E. Strömberg. 2012. High-resolution isotopic record of
569 C4 photosynthesis in a Miocene grassland. Palaeogeography, Palaeoclimatology,
570 Palaeoecology 337-338:88-98.
571 Crifò, C., and C. A. E. Strömberg. 2021. Spatial patterns of soil phytoliths in a wet vs. dry
572 Neotropical forest: implications for paleoecology. Palaeogeography Palaeoclimatology
573 Palaeoecology 526:110100.
574 Cui, Y., B. A. Schubert, and A. H. Jahren. 2020. A 23 m.y. record of low atmospheric CO2.
575 Geology 48(9):888-892.

576 Diefendorf, A. F., K. E. Mueller, S. L. Wing, P. L. Koch, and K. H. Freeman. 2010. Global
577 patterns in leaf 13C discrimination and implications for studies of past and future climate.
578 Proceedings of the National Academy of Sciences 107(13):5738.
579 Dunn, R. E., C. A. E. Strömberg, R. H. Madden, M. J. Kohn, and A. A. Carlini. 2015. Linked
580 canopy, climate, and faunal change in the Cenozoic of Patagonia. Science
581 347(6219):258-61.
582 Edwards, E. J., C. P. Osborne, C. A. E. Strömberg, S. A. Smith, C. G. Consortium, W. J. Bond,
583 P. A. Christin, A. B. Cousins, M. R. Duvall, D. L. Fox, R. P. Freckleton, O. Ghannoum,
584 J. Hartwell, Y. Huang, C. M. Janis, J. E. Keeley, E. A. Kellogg, A. K. Knapp, A. D.
585 Leakey, D. M. Nelson, J. M. Saarela, R. F. Sage, O. E. Sala, N. Salamin, C. J. Still, and
586 B. Tipple. 2010. The origins of C4 grasslands: integrating evolutionary and ecosystem
587 science. Science 328(5978):587-91.
588 Edwards, E. J., and C. J. Still. 2008. Climate, phylogeny and the ecological distribution of C4
589 grasses. Ecology Letters 11(3):266-276.
590 Ehleringer, J. R., and T. E. Cerling. 2002. C3 and C4 photosynthesis. Encyclopedia of global
591 environmental change 2(4).
592 Ehleringer, J. R., T. E. Cerling, and B. R. Helliker. 1997. C4 photosynthesis, atmospheric CO2,
593 and climate. Oecologia 112(3):285-299.
594 Farquhar, G. D. 1983. On the nature of carbon isotope discrimination in C4 species. Functional
595 Plant Biology 10(2):205-226.
596 Favier, C., J. Aleman, L. Bremond, M. A. Dubois, V. Freycon, and J.-M. Yangakola. 2012.
597 Abrupt shifts in African savanna tree cover along a climatic gradient. Global Ecology and
598 Biogeography 21(8):787-797.

599 Fox, D. L., S. Pau, L. Taylor, C. A. Strömberg, C. P. Osborne, C. Bradshaw, S. Conn, D. J.
600 Beerling, and C. J. Still. 2018. Climatic controls on C4 grassland distributions during the
601 Neogene: A model-data comparison. Frontiers in Ecology and Evolution:147.
602 Fredlund, G. G., and L. T. Tieszen. 1994. Modern Phytolith Assemblages from the North
603 American Great Plains. Journal of Biogeography 21(3):321-335.
604 Garzione, C. N., N. McQuarrie, N. D. Perez, T. A. Ehlers, S. L. Beck, N. Kar, N. Eichelberger,
605 A. D. Chapman, K. M. Ward, M. N. Ducea, R. O. Lease, C. J. Poulsen, L. S. Wagner, J.
606 E. Saylor, G. Zandt, and B. K. Horton. 2017. Tectonic Evolution of the Central Andean
607 Plateau and Implications for the Growth of Plateaus. Annual Review of Earth and
608 Planetary Sciences 45(1):529-559.
609 Gibson, D. J. 2009. Grasses and grassland ecology. Oxford University Press.
610 Gregory-Wodzicki, K. M. 2000. Uplift history of the Central and Northern Andes: A review.
611 GSA Bulletin 112(7):1091-1105.
612 Harris, E. B., C. A. E. S. Strömberg, N.D., S. Y. Smith, and D. A. Vilhena. 2017. Vegetation
613 response during the lead-up to the middle Miocene warming event in the Northern Rocky
614 Mountains, USA. Palaeogeography, Palaeoclimatology, Palaeoecology 485:401-415.
615 Hooghiemstra, H., and T. Van der Hammen. 2004. Quaternary Ice-Age dynamics in the
616 Colombian Andes: developing an understanding of our legacy. Philosophical
617 Transactions of the Royal Society of London. Series B: Biological Sciences
618 359(1442):173-181.
619 Hoorn, C., F. P. Wesselingh, H. ter Steege, M. A. Bermudez, A. Mora, J. Sevink, I. Sanmartin,
620 A. Sanchez-Meseguer, C. L. Anderson, J. P. Figueiredo, C. Jaramillo, D. Riff, F. R.
621 Negri, H. Hooghiemstra, J. Lundberg, T. Stadler, T. Sarkinen, and A. Antonelli. 2010.

622 Amazonia through time: Andean uplift, climate change, landscape evolution, and
623 biodiversity. Science 330(6006):927-31.
624 Jacobs, B. F., J. D. Kingston, and L. L. Jacobs. 1999. The origin of grass-dominated ecosystems.
625 Annals of the Missouri Botanical Garden:590-643.
626 Kerns, B. K., M. M. Moore, and S. C. Hart. 2001. Estimating forest-grassland dynamics using
627 soil phytolith assemblages and d13C of soil organic matter. Écoscience 8(4):478-488.
628 Kirschner, J. A., and C. Hoorn. 2020. The onset of grasses in the Amazon drainage basin,
629 evidence from the fossil record. Frontiers of Biogeography 12(2).
630 Ledru, M. P., V. Jomelli, P. Samaniego, M. Vuille, S. Hidalgo, M. Herrera, and C. Ceron. 2013.
631 The Medieval Climate Anomaly and the Little Ice Age in the eastern Ecuadorian Andes.
632 Clim. Past 9(1):307-321.
633 Livingstone, D., and W. Clayton. 1980. An altitudinal cline in tropical African grass floras and
634 its paleoecological significance. Quaternary Research 13(3):392-402.
635 Lloyd, J., and G. D. Farquhar. 1994. 13C discrimination during CO2 assimilation by the
636 terrestrial biosphere. Oecologia 99(3-4):201-215.
637 Loughney, K. M., M. T. Hren, S. Y. Smith, and J. L. Pappas. 2019. Vegetation and habitat
638 change in southern California through the Middle Miocene Climatic Optimum:
639 Paleoenvironmental records from the Barstow Formation, Mojave Desert, USA. GSA
640 Bulletin 132(1-2):113-129.
641 Lupien, R. L., J. M. Russell, C. L. Yost, J. D. Kingston, A. L. Deino, J. Logan, A. Schuh, and A.
642 S. Cohen. 2021. Vegetation change in the Baringo Basin, East Africa across the onset of
643 Northern Hemisphere glaciation 3.3–2.6 Ma. Palaeogeography, Palaeoclimatology,
644 Palaeoecology 570:109426.

645 Madella, M., M. Jones, P. Echlin, A. Powers-Jones, and M. Moore. 2009. Plant water availability
646 and analytical microscopy of phytoliths: implications for ancient irrigation in arid zones.
647 Quaternary International 193(1-2):32-40.
648 Martínez, C., C. Jaramillo, A. Correa-Metrío, W. Crepet, J. E. Moreno, A. Aliaga, F. Moreno, M.
649 Ibañez-Mejia, and M. B. Bush. 2020. Neogene precipitation, vegetation, and elevation
650 history of the Central Andean Plateau. Science Advances 6(35):eaaz4724.
651 McInerney, F. A., C. A. E. Strömberg, and J. W. C. White. 2016. The Neogene transition from
652 C3 to C4 grasslands in North America: stable carbon isotope ratios of fossil phytoliths.
653 Paleobiology 37(1):23-49.
654 Meyers, P. A., and R. Ishiwatari. 1993. Lacustrine organic geochemistry—an overview of
655 indicators of organic matter sources and diagenesis in lake sediments. Organic
656 geochemistry 20(7):867-900.
657 Miller, L. A., S. Y. Smith, N. D. Sheldon, and C. A. E. Stromberg. 2012. Eocene vegetation and
658 ecosystem fluctuations inferred from a high-resolution phytolith record. Geological
659 Society of America Bulletin 124(9-10):1577-1589.
660 Neumann, K., C. A. E. Strömberg, T. Ball, R. M. Albert, L. Vrydaghs, and L. S. Cummings.
661 2019. International code for phytolith nomenclature (ICPN) 2.0. Annals of Botany
662 124(2):189-199.
663 Palmer, P. G. 1976. Grass cuticles: a new paleoecological tool for East African lake sediments.
664 Canadian Journal of Botany 54(15):1725-1734.
665 Parmenter, C., and D. W. Folger. 1974. Eolian Biogenic Detritus in Deep Sea Sediments: A
666 Possible Index of Equatorial Ice Age Aridity. Science 185(4152):695-698.

667 Pau, S., E. J. Edwards, and C. J. Still. 2013. Improving our understanding of environmental
668 controls on the distribution of C3 and C4 grasses. Global Change Biology 19(1):184-196.
669 Pearsall, D. M. 2000. Paleoethnobotany: a handbook of procedures. Academic Press.
670 Piperno, D. R. 1988. Phytolyth analysis: an archaeological and geological perspective. Academic
671 Press, San Diego.
672 Piperno, D. R., and D. M. Pearsall. 1998. The silica bodies of tropical American grasses:
673 morphology, taxonomy, and implications for grass systematics and fossil phytolith
674 identification. Smithsonian contributions to botany.
675 Prentice, I. C. 1985. Pollen Representation, Source Area, and Basin Size: Toward a Unified
676 Theory of Pollen Analysis. Quaternary Research 23(1):76-86.
677 R Development Core Team. 2019. R: A language and environment for statistical computing. R
678 Foundation for Statistical Computing, Vienna, Austria. Vienna, Austria.
679 Rieley, G., R. J. Collier, D. M. Jones, G. Eglinton, P. A. Eakin, and A. E. Fallick. 1991. Sources
680 of sedimentary lipids deduced from stable carbon-isotope analyses of individual
681 compounds. Nature 352(6334):425-427.
682 Rundel, P. W. 1980. The ecological distribution of C4 and C3 grasses in the Hawaiian Islands.
683 Oecologia 45(3):354-359.
684 Sage, R. F., P. A. Christin, and E. J. Edwards. 2011. The C4 plant lineages of planet Earth. J Exp
685 Bot 62(9):3155-69.
686 Sage, R. F., R. K. Monson, J. R. Ehleringer, S. Adachi, and R. W. Pearcy. 2018. Some like it hot:
687 the physiological ecology of C4 plant evolution. Oecologia 187(4):941-966.
688 Simpson, B. B., and C. A. Todzia. 1990. Patterns and processes in the development of the high
689 Andean flora. American Journal of Botany 77(11):1419-1432.

690 Smith, B. N., and S. Epstein. 1971. Two categories of 13C/12C ratios for higher plants. Plant
691 physiology 47(3):380-384.
692 Song, Z., K. McGrouther, and H. Wang. 2016. Occurrence, turnover and carbon sequestration
693 potential of phytoliths in terrestrial ecosystems. Earth-Science Reviews 158:19-30.
694 Still, C. J., J. A. Berry, G. J. Collatz, and R. S. DeFries. 2003. Global distribution of C3 and C4
695 vegetation: carbon cycle implications. Global Biogeochemical Cycles 17(1):6-1-6-14.
696 Strömberg, C. A. E. 2003. The origin and spread of grass-dominated ecosystems during the
697 Tertiary of North America and how it relates to the evolution of hypsodonty in equids.
698 University of California, Berkeley.
699 Strömberg, C. A. E. 2005. Decoupled taxonomic radiation and ecological expansion of open-
700 habitat grasses in the Cenozoic of North America. Proc Natl Acad Sci U S A
701 102(34):11980-4.
702 Strömberg, C. A. E. 2009. Methodological concerns for analysis of phytolith assemblages: Does
703 count size matter? Quaternary International 193(1):124-140.
704 Strömberg, C. A. E. 2011. Evolution of Grasses and Grassland Ecosystems. Annual Review of
705 Earth and Planetary Sciences 39:517-544.
706 Strömberg, C. A. E., V. S. Di Stilio, and Z. Song. 2016. Functions of phytoliths in vascular
707 plants: an evolutionary perspective. Functional Ecology 30(8):1286-1297.
708 Strömberg, C. A. E., R. E. Dunn, R. H. Madden, M. J. Kohn, and A. A. Carlini. 2013.
709 Decoupling the spread of grasslands from the evolution of grazer-type herbivores in
710 South America. Nat Commun 4:1478.

711 Strömberg, C. A. E., E. M. Friis, M.-M. Liang, L. Werdelin, and Y.-L. Zhang. 2007a.
712 Palaeoecology of an Early-Middle Miocene lake in China: preliminary interpretations
713 based on phytoliths from the Shanwang Basin. Vertebrata PalAsiatica 45(2):145-160.
714 Strömberg, C. A. E., and F. A. McInerney. 2016. The Neogene transition from C3 to C4
715 grasslands in North America: assemblage analysis of fossil phytoliths. Paleobiology
716 37(1):50-71.
717 Strömberg, C. A. E., L. Werdelin, E. M. Friis, and G. Saraç. 2007b. The spread of grass-
718 dominated habitats in Turkey and surrounding areas during the Cenozoic: Phytolith
719 evidence. Palaeogeography, Palaeoclimatology, Palaeoecology 250(1-4):18-49.
720 Teeri, J., and L. Stowe. 1976. Climatic patterns and the distribution of C4 grasses in North
721 America. Oecologia 23(1):1-12.
722 Thorn, V. 2001. Oligocene and early Miocene phytolits from CRP-2/2A and CRP-3, Victoria
723 Land Basin, Antarctica. Terra Antartica 8(4):407-422.
724 Tieszen, L. L., M. M. Senyimba, S. K. Imbamba, and J. H. Troughton. 1979. The distribution of
725 C3 and C4 grasses and carbon isotope discrimination along an altitudinal and moisture
726 gradient in Kenya. Oecologia 37(3):337-350.
727 Trembath-Reichert, E., J. P. Wilson, S. E. McGlynn, and W. W. Fischer. 2015. Four hundred
728 million years of silica biomineralization in land plants. Proc Natl Acad Sci U S A
729 112(17):5449-54.
730 Twiss, P. C. 1992. Predicted world distribution of C3 and C4 grass phytoliths. Pp. 113-128.
731 Phytolith systematics. Springer.

732 Urban, M. A., D. M. Nelson, G. Jiménez-Moreno, J.-J. Châteauneuf, A. Pearson, and F. S. Hu.
733 2010. Isotopic evidence of C4 grasses in southwestern Europe during the Early
734 Oligocene–Middle Miocene. Geology 38(12):1091-1094.
735 Winslow, J. C., E. R. Hunt Jr, and S. C. Piper. 2003. The influence of seasonal water availability
736 on global C3 versus C4 grassland biomass and its implications for climate change
737 research. Ecological Modelling 163(1-2):153-173.
738 Wynn, J. G., and M. I. Bird. 2007. C4-derived soil organic carbon decomposes faster than its C3
739 counterpart in mixed C3/C4 soils. Global Change Biology 13(10):2206-2217.
740 Young, H. J., and T. P. Young. 1983. Local distribution of C3 and C4 grasses in sites of overlap
741 on Mount Kenya. Oecologia 58(3):373-377.
742 Zhou, H., B. R. Helliker, M. Huber, A. Dicks, and E. Akcay. 2018. C4 photosynthesis and
743 climate through the lens of optimality. Proc Natl Acad Sci U S A 115(47):12057-12062.
744

Table and Figure captions

Table 1. List of phytolith types included in this study. Full names and acronyms according to the ICPN 2.0 given in columns 2 and 3 (numbers in parentheses in column 3 refers to morphotype subtypes). The dominant photosynthetic pathway of producing taxa (GSSCP only) is given in column 4. Phytolith classes (1-5) are given in column 5.

Table 2. Statistics of the linear regression between C_4/C_3 grass proportion of the samples, and MAT, elevation of the sites, and the $\delta^{13}C$ of the samples.

Table 3. Statistics of the linear regression between C_4/C_3 GSSCP proportion of the samples, and MAT, elevation of the sites, and the $\delta^{13}C$ of the samples.

Figure 1. Map of the study site. Red dots represent sample sites. Site names are given in red and correspond to: PES01-06; SOC01, BAR01, CHI01-06, IGU01-07, GUA01-02, PPAL01, 06, NEI01-9, 11-2, Anti2-5. Map was created using the ETOPO2 Global Relief Model map (NOAA) in R 4.0.3.

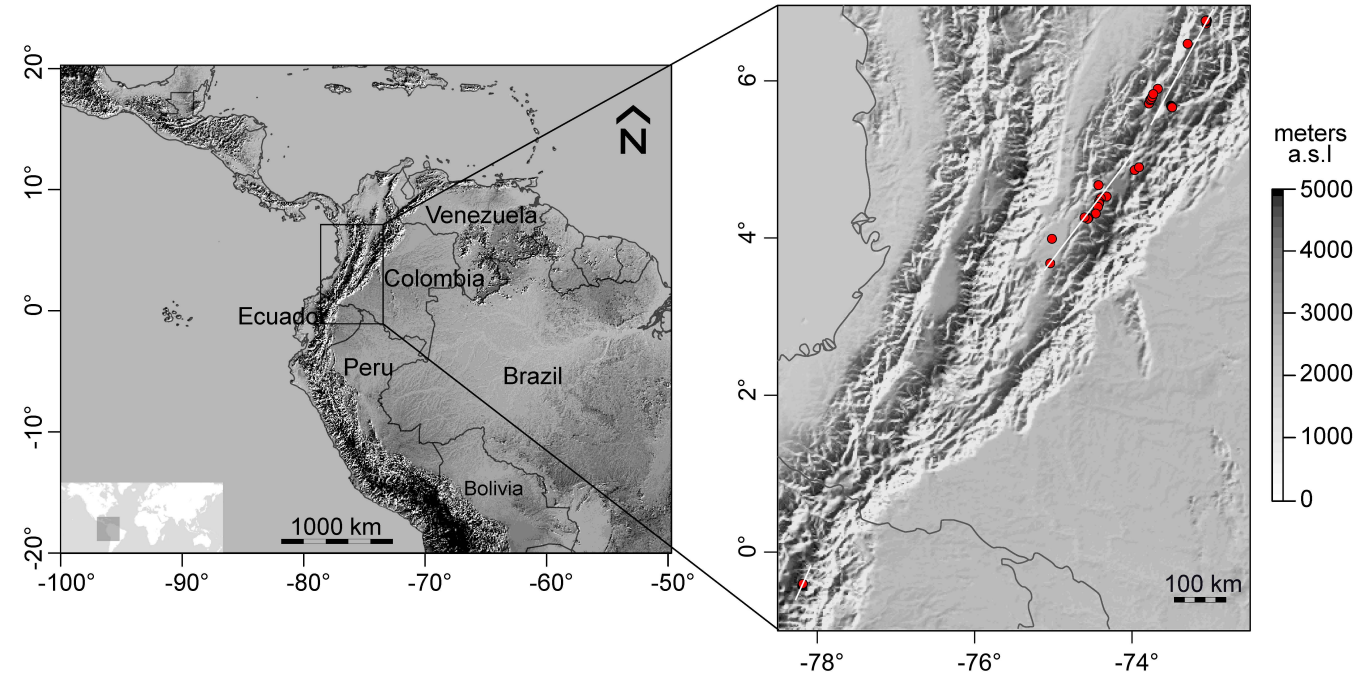
Figure 2. Examples of phytolith morphotypes from soil assemblages diagnostic of: Poaceae (A-Z), woody dicotyledons (A_1-D_1); Palms/bromeliads/ Zingiberales (E_1-F_1), Other Poaceae (G_1-N_1), and non-diagnostic morphotypes (M_1-V_1). Morphotype names in small capitals (e.g., BILOBATE) and acronyms (e.g., BIL) signify morphotypes as formally defined in the

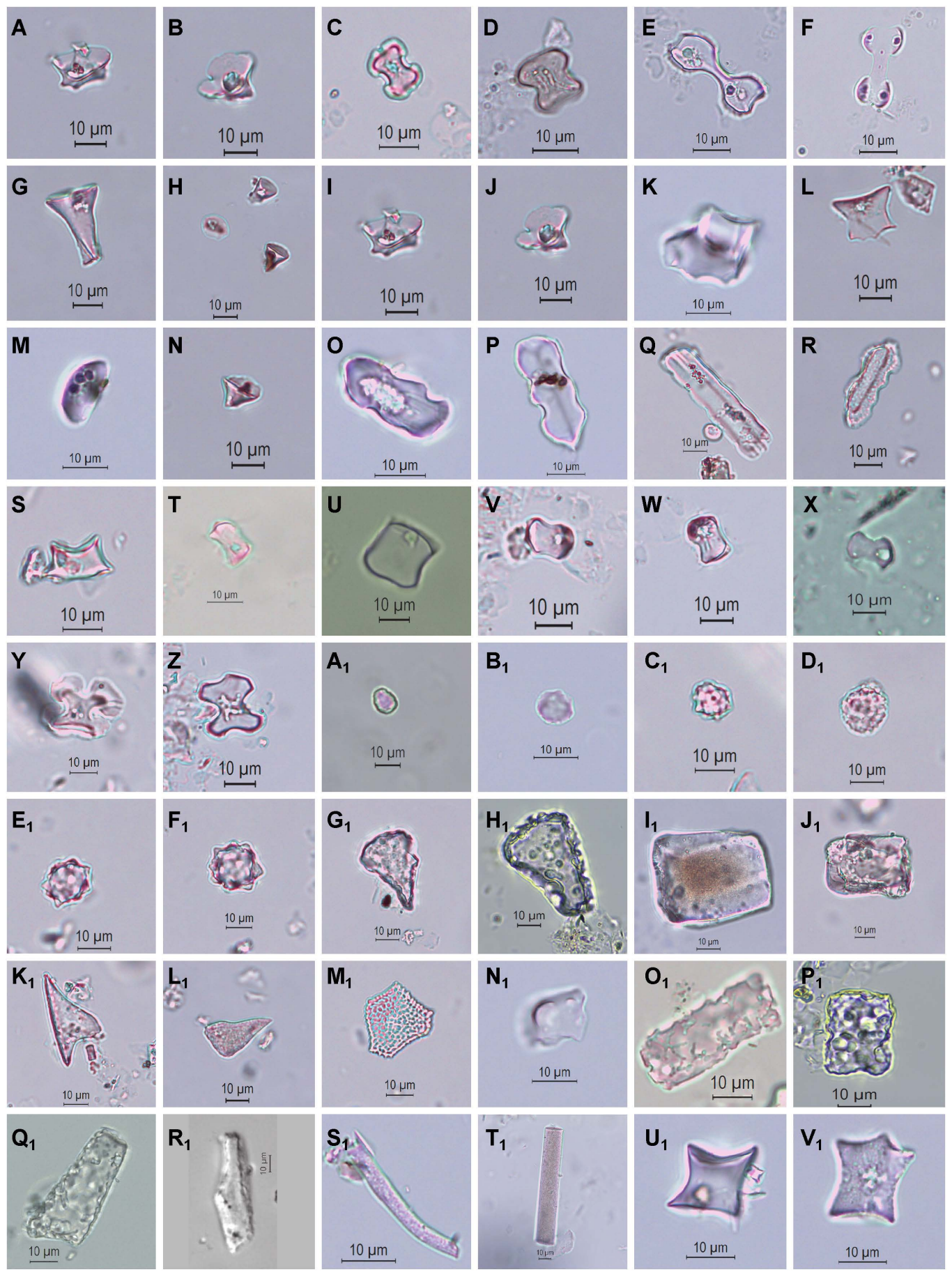
767 International Code for Phytolith Nomenclature 2.0 (Neumann et al. 2019); morphotype names in
768 square brackets correspond to the classification scheme by Bremond (2005a, and thereafter).

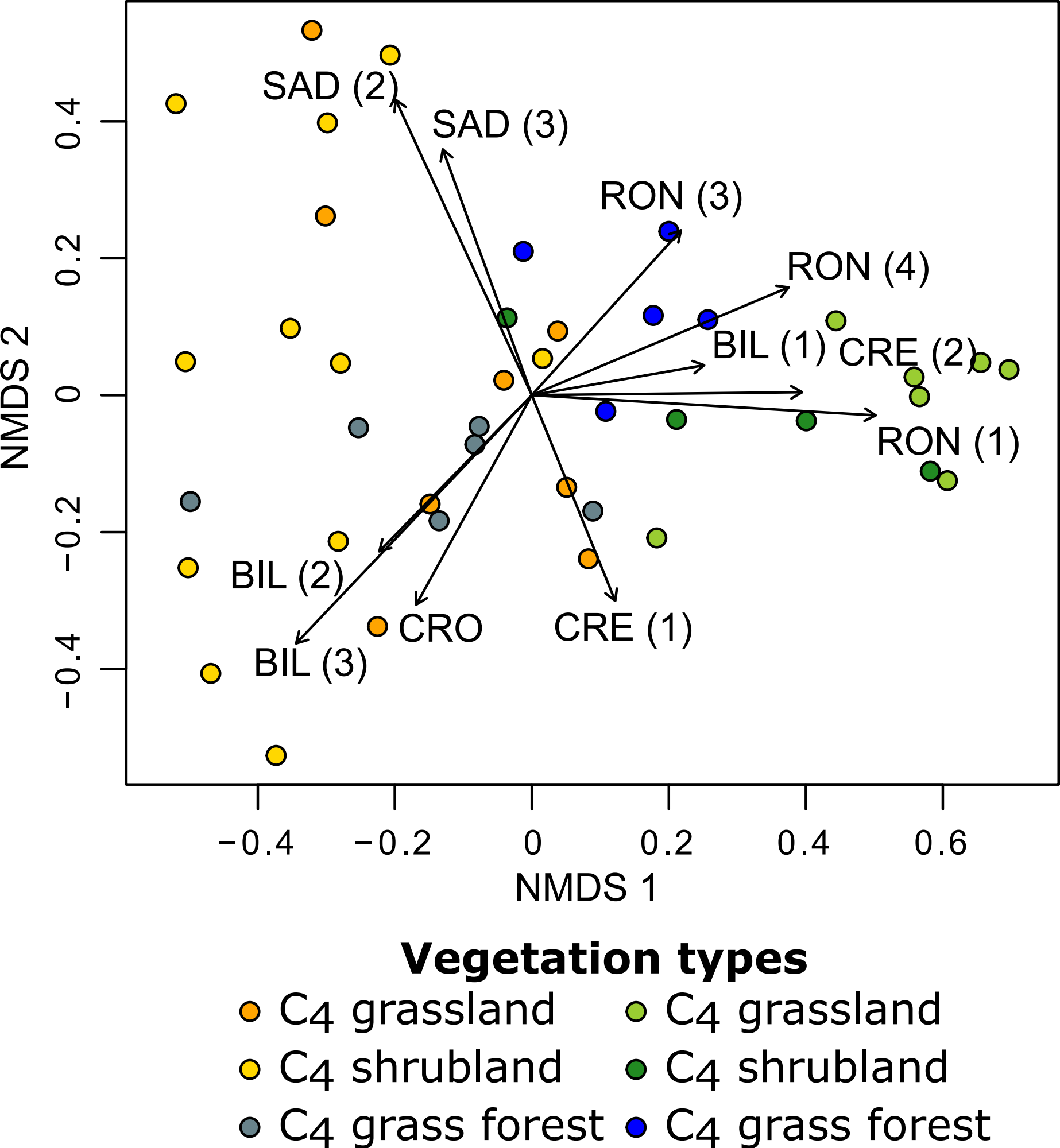
769 **A-F.** BILOBATE (BIL) morphotypes, including **A-B.** BIL (1) [*Stipa* type bilobate], **C-D.** BIL (2)
770 [short bilobate], and **E-F.** BIL (3) [long bilobate]. **G-J.** RONDEL (RON) morphotypes; **K-N-.**
771 CRENATE (CRE) morphotypes including **K-L.** CRE (1), [trapeziform trilobate], and **M-N.** CRE
772 (2) [trapeziform polylobate]; S-X. SADDLE (SAD) morphotypes including **O-P.** SAD (1),
773 [collapsed saddles], **S-U.** SAD (2), [short saddle], and **V-W.** SAD (3), [long saddle]; **Y-Z.** Cross
774 morphotypes; **A1.** SPHEROID PSILATE (SPH_PSI) morphotypes, [globular smooth]. **C_1-D_1.**
775 SPHEROID ORNATE (SPH_ORN) morphotypes, [globular granulate]; **E_1-F_1.** SPHEROID ECHINATE
776 (SPH_ECH) morphotypes, [globular echinate]; **G_1-H_1.** BULLIFORM FLABELLATE (BUL_FLA)
777 morphotypes, [cuneiform bulliform cell]; **I_1.** BLOCKY (BLO) morphotypes, [parallelepipedal
778 bulliform cell]; **K_1-L_1**. ACUTE BULBOSUS (ACU_BUL) morphotypes, [acicular air cell]; **M_1-N_1**.
779 PAPILLATE (PAP) morphotypes, [papillae]; O_1-T_1. ELONGATE ENTIRE (ELO_ENT) morphotypes
780 including **O_1-P_1.** ELO_ENT (1), [parallelepipedal long cell], **Q_1-R_1.** ELO_ENT (2), [thick
781 trapezoidal rectangle], and **S_1-T_1.** ELO_ENT (3); [elongates]; **U_1-V_1.** Undetermined
782 morphotypes (n.a.), [unclassified].
783
784 **Figure 3.** Nonmetric multidimensional scaling (NMDS) ordination of the phytolith assemblages
785 across sites (colored dots), and phytolith classes (black vectors). Only phytolith classes with
786 significant ($p < 0.05$) loadings on the ordination axes are shown here. The direction of vectors
787 indicates maximum correlation of the phytolith classes with between sample scores in the
788 ordination space. Sample sites are colored according to vegetation type (see figure legend).
789

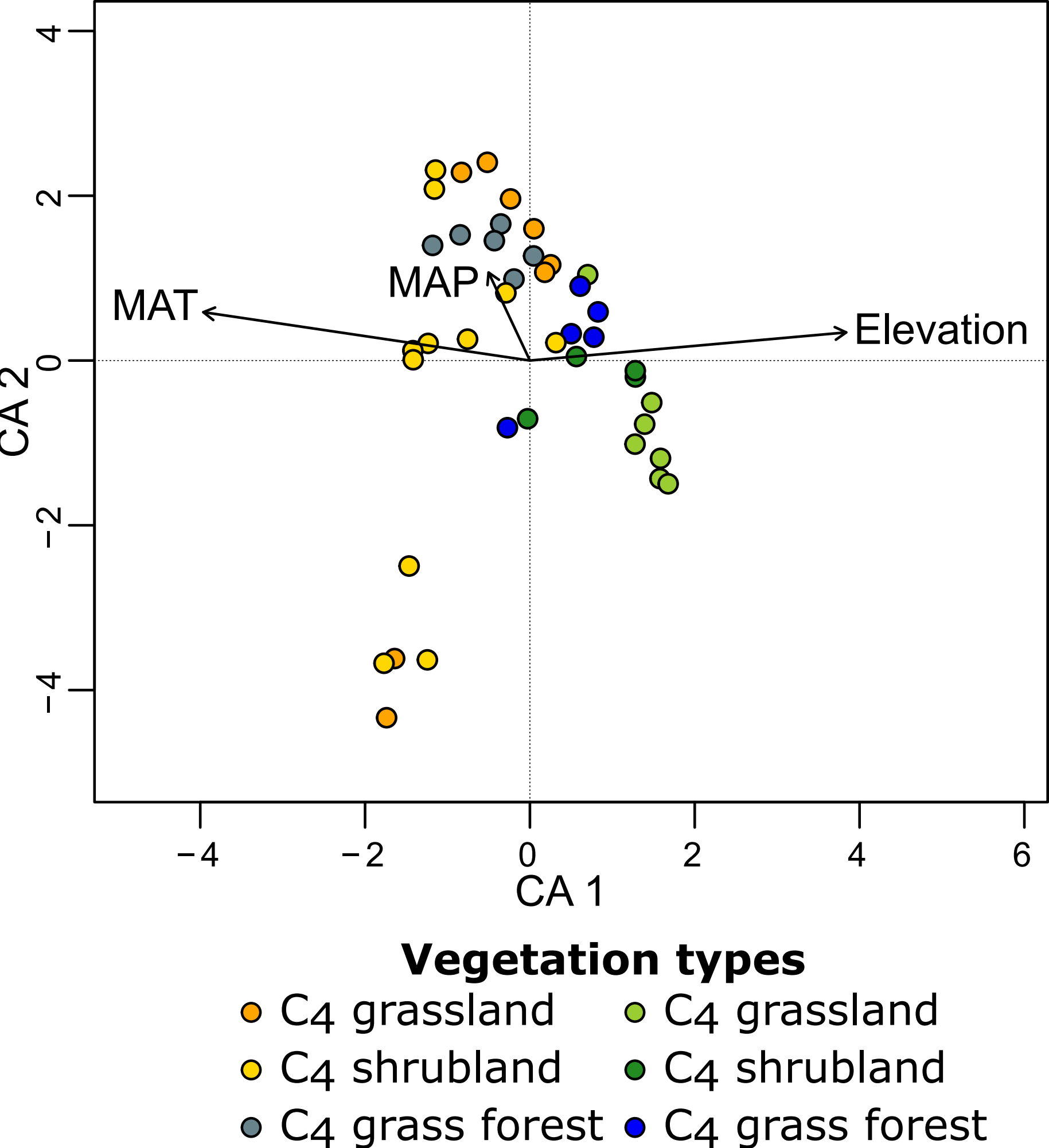
790 **Figure 4.** Correspondence analysis ordination of the phytolith assemblages across sites (colored
791 dots), and environmental variable (black vectors). Sample sites are colored according to
792 vegetation type (see figure legend).
793
794 **Figure 5.** Proportion of C_4 grass phytoliths (GSSCP) and carbon isotope ratios ($\partial^{13}C$) of
795 samples against mean annual temperature (MAT; lower x axis) and mean annual precipitation
796 (MAP; upper x axis). Blue dots represent estimates of C_4 GSSCP proportion from phytolith
797 counts; vertical blue bars correspond to 95% confidence intervals calculated by bootstrap
798 analysis of each sample. Black crosses represent carbon isotope ratios ($\partial^{13}C$) of the samples.
799 Solid and dotted blue lines represent respectively the logistic regression curve performed
800 between the proportion of C_4 GSSCP in phytolith assemblages and MAT ($p << 0.01$), and the
801 upper and lower confidence interval bounds. The solid gray line represents the linear regression
802 ($p << 0.001$, $R^2 = 0.87$) performed between the proportion of C_4 GSSCP in phytolith assemblages
803 and MAT. Orange dots represent the C_4 proportion in the study sites (calculated from the floristic
804 survey). The dot-dashed green line represents the logistic regression based on a compilation of
805 floristic data for Colombia, Ecuador, Peru and Bolivia in Bremond et al. (2012) ($p << 0.01$).
806
807 **Figure 6.** Phytolith assemblages organized according to increasing elevation. Bars represent
808 percentage of phytolith classes. Palm phytoliths include SPHEROID ECHINATE MORPHOTYPES;
809 Woody dicots phytoliths include SPHEROID PSILATE and SPHEROID ORNATE morphotypes; C_3
810 GSSCP include "*Stipa*-type" BILOBATE [BIL (1)], RONDEL (RON), CRENATE (CRE) and
811 "collapsed" SADDLE [SAD (1)] morphotypes; C_4 GSSCP include "short" and "long" BILOBATE
812 [BIL (2) and BIL (3)], short" and "long" SADDLE [SAD (2), and SAD (3)], and CROSS (CRO)

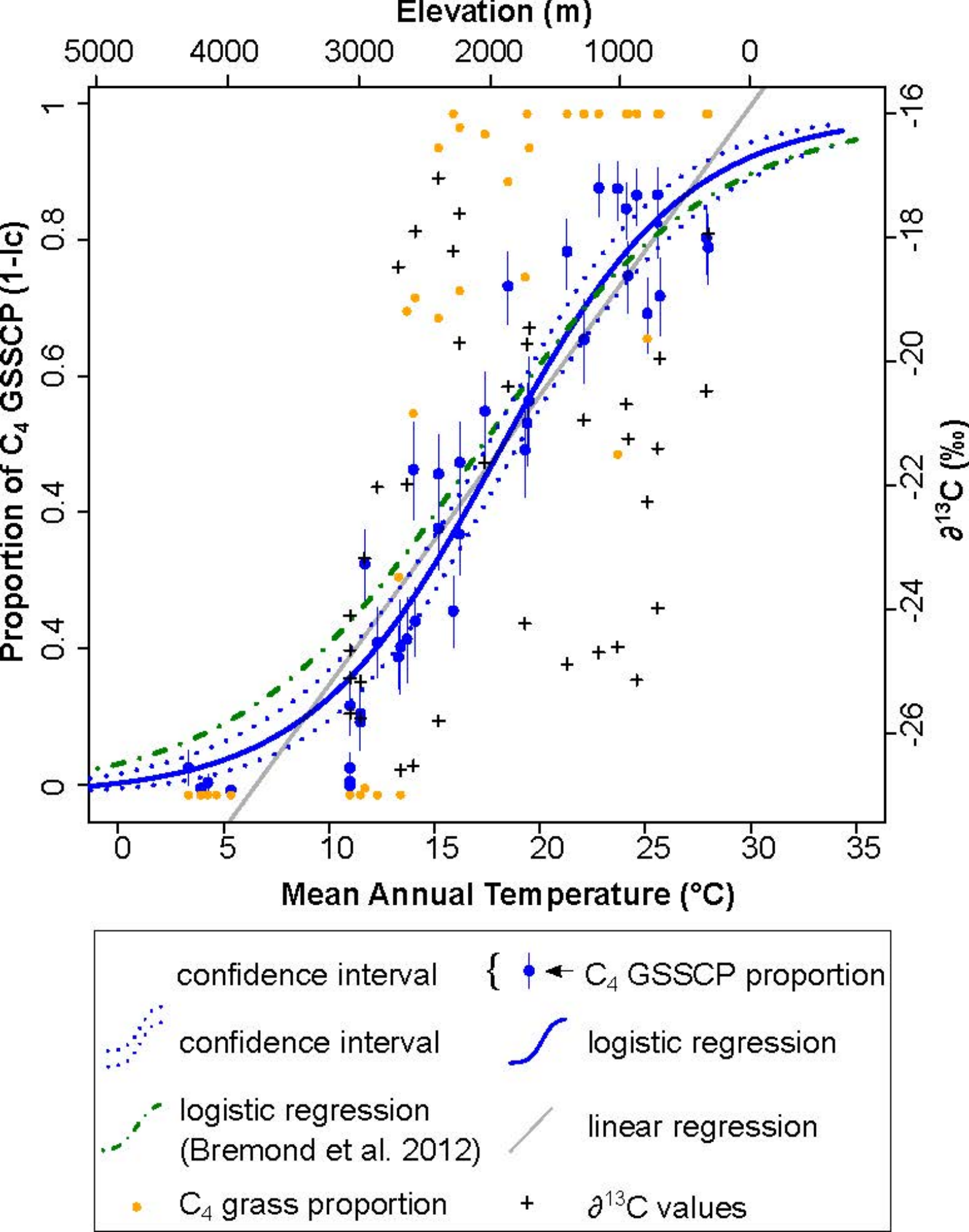
morphotypes; Other Poaceae morphotype include BULLIFORM FLABELLATE (BUL_FLA), BLOCKY (BLO), ACUTE BULBOSUS (ACU_BUL), and PAPILLATE (PAP) morphotypes.

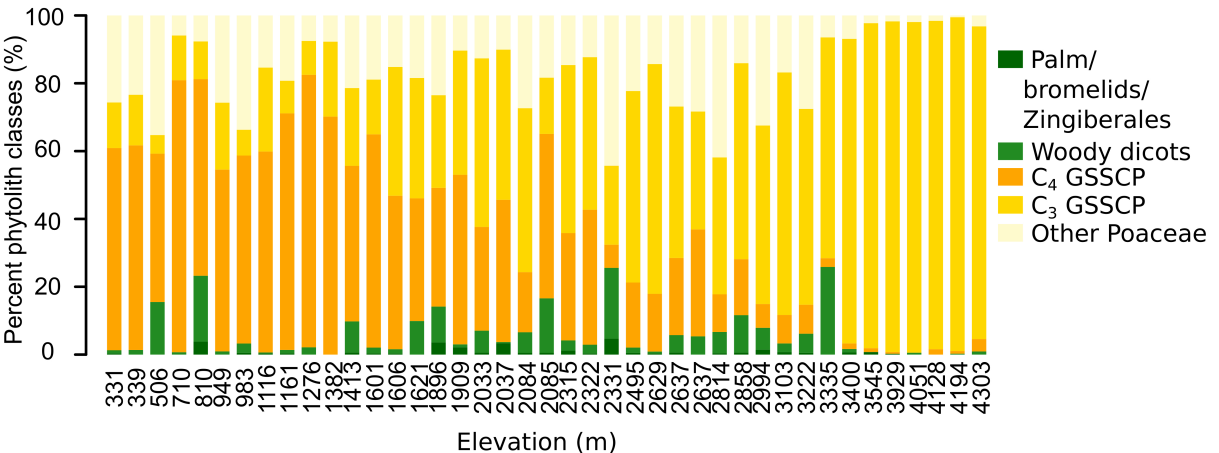

Venezuela
Colombia
Ecuador
Peru
Brazil
Bolivia
1000 km
N
20°
10°
0°
-10°
-20°
-100°
-90°
-80°
-70°
-60°
-50°
6°
4°
2°
0°
-78°
-76°
-74°
100 km
meters a.s.l
5000
4000
3000
2000
1000
0


A
10 µm
B
10 µm
C
10 µm
D
10 µm
E
10 µm
F
10 µm
G
10 µm
H
10 µm
I
10 µm
J
10 µm
K
10 µm
L
10 µm
M
10 µm
N
10 µm
O
10 µm
P
10 µm
Q
10 µm
R
10 µm
S
10 µm
T
10 µm
U
10 µm
V
10 µm
W
10 µm
X
10 µm
Y
10 µm
Z
10 µm
A1
10 µm
B1
10 µm
C1
10 µm
D1
10 µm
E1
10 µm
F1
10 µm
G1
10 µm
H1
10 µm
I1
10 µm
J1
10 µm
K1
10 µm
L1
10 µm
M1
10 µm
N1
10 µm
O1
10 µm
P1
10 µm
Q1
10 µm
R1
10 µm
S1
10 µm
T1
10 µm
U1
10 µm
V1
10 µm


SAD (2)
SAD (3)
RON (3)
RON (4)
BIL (1)
CRE (2)
RON (1)
BIL (2)
CRO
CRE (1)
BIL (3)
NMDS 2
NMDS 1
0.4
0.2
0
−0.2
−0.4
−0.4
−0.2
0
0.2
0.4
0.6
Vegetation types
C4 grassland
C4 shrubland
C4 grass forest
C4 grassland
C4 shrubland
C4 grass forest


MAT
MAP
Elevation
CA 1
CA 2
−4
−2
0
2
4
6
Vegetation types
C4 grassland
C4 shrubland
C4 grass forest
C4 grassland
C4 shrubland
C4 grass forest


Elevation (m)
5000
4000
3000
2000
1000
0
Proportion of C_4 GSSCP (1-Ic)
0
0.4
0.6
0.8
1
$\partial^{13}C$ (‰)
-16
-18
-20
-22
-24
-26
0
5
10
15
20
25
30
35
Mean Annual Temperature (°C)
confidence interval
confidence interval
logistic regression
(Bremond et al. 2012)
C_4 grass proportion
C_4 GSSCP proportion
logistic regression
linear regression
$\partial^{13}C$ values


Percent phytolith classes (%)
100
80
60
40
20
0
331
339
506
710
810
949
983
1116
1161
1276
1382
1413
1601
1606
1621
1896
1909
2033
2037
2084
2085
2315
2322
2331
2495
2629
2637
2637
2814
2858
2994
3103
3222
3335
3400
3545
3929
4051
4128
4194
4303
Elevation (m)
Palm/
bromelids/
Zingiberales
Woody dicots
C_4 GSSCP
C_3 GSSCP
Other Poaceae