Responses to referee #2

This manuscript provides a wealth of data on the pollen morphology of *Artemisia* and its relation to the habitat preferences of their mother plants. Such data will serve in the future for better reconstructions of paleoenvironmental setting, which in turn will help to better understand the earth's climate past.

The manuscript is well written, well structured and with a very informative introduction. The methods used are all well established and robust. I could envision more sophisticated statistical analyses to be applied, but this may be done in the future with the database grown further.

Thanks for these constructive suggestions. We have revised the manuscript according to your comments. Please see the following point-by-point responses.

C1. One critique I may raise here, is the assessment of the distribution of taxa in different biomes (here wrongly called ecoregions). First of all, the results shown in Figure 18 do not support the claims made in the text with respect to habitat preferences of the species or morpho-groups. Second, biomes are too large a vegetation unit to be able to reflect habitat preferences. Here, probably the actual habitat of the sampled specimen needs to be considered, or a different approach to be developed.

R1. Following your suggestions, we have replaced "ecoregions" with "biomes", and removed the assessment of the distribution of species specimens from GBIF in different biomes from Fig. 18. Therefore, the problem that Fig. 18b does not support habitat preference well in the text no longer exists. Section 4.2 The Ecological Implications of *Artemisia* pollen Types are modified as follows.

Lines 368-399: "Plotting the distribution data of 33 species from 9 main branches of *Artemisia* constrained by the phylogenetic framework (Fig. 1) onto the global terrestrial biomes (Fig. 20), we noticed that the genus is widely distributed from forest to grassland, desert, and saline habitats (Figs. 16, 17a, 20). Furthermore, different species of *Artemisia* with SWS pollen type (Fig. 20a) and LNS type (Fig. 20b) have a rather wide distribution with severely overlapping ranges while those with SG type (Fig. 20c) have narrow and isolated distributions.
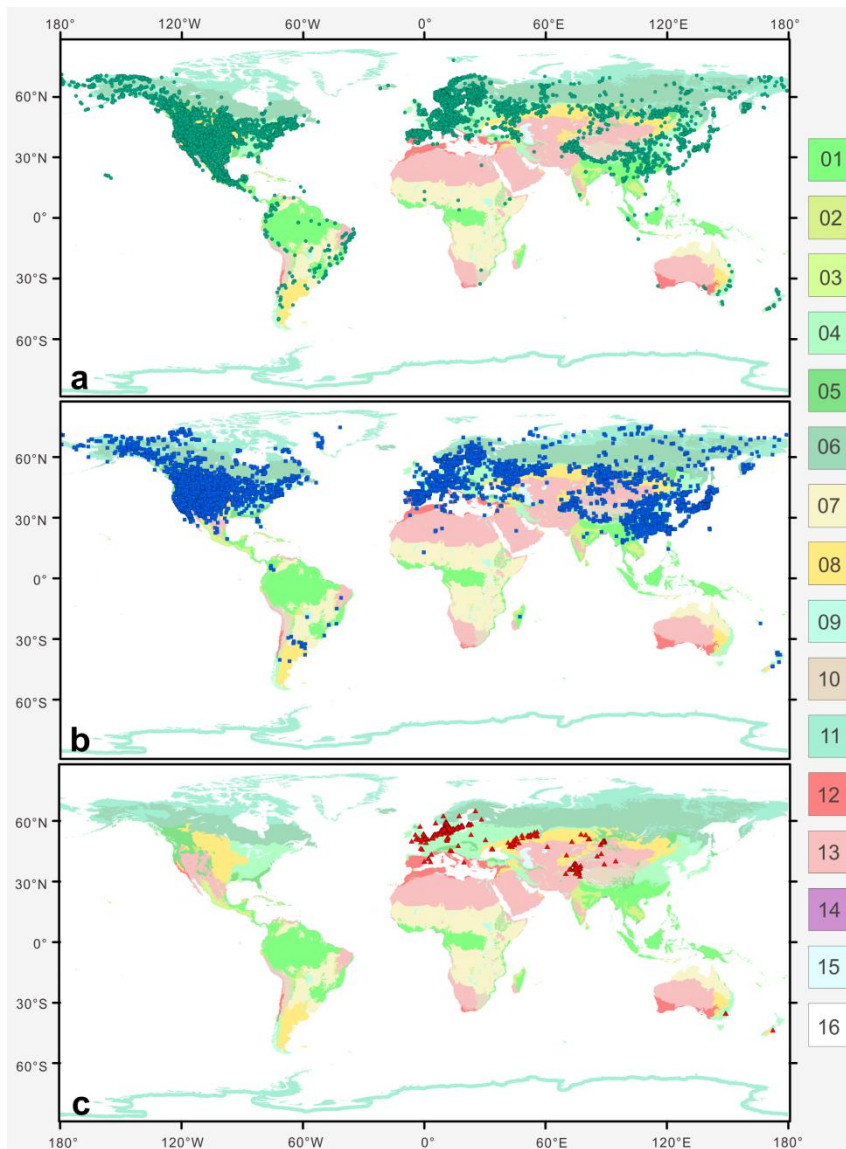
Figure 20. The global distribution pattern of 3 *Artemisia* pollen types in terrestrial biomes (modified from Olson et al., 2001). a. SG type; b. LNS type; c. SWS type.

14 terrestrial biomes: 01. Tropical and Subtropical Moist Broadleaf Forests; 02. Tropical and Subtropical Dry Broadleaf Forests; 03. Tropical and Subtropical Coniferous Forests; 04. Temperate Broadleaf and Mixed Forests; 05. Temperate Coniferous Forests; 06. Boreal Forests/Taiga; 07. Flooded Grasslands and Savannas; 08. Montane Grasslands and Shrublands; 09. Tundra; 10: Mediterranean Forests, Woodlands, and Shrub; 11. Tropical and Subtropical Grasslands, Savannas, and Shrublands; 12. Temperate Grasslands, Savannas, and Shrublands; 13. Deserts and Xeric Shrublands; 14. Mangroves; 15. Lakes; 16. Rock and Ice.

The ecological implications of *Artemisia* pollen types mentioned above fall into four categories. (i) *Artemisia* with the SG pollen type all belong to the subg. *Seriphidium*, which generally grows in dry habitats ranging from grassland desert to desert and coastal saline-alkaline environments, with their distribution largely limited to Eurasia and growing at low altitude (Figs. 17b, 20c, 21). (ii) The habitats of *Artemisia* with LNS pollen type have a global distribution and occur in forest, grassland and desert, and even coastal areas (Figs. 17b, 20b, 21), with the highest mean annual temperature (MAT). Hence, the LNS pollen type is a

generalist. (iii) *Artemisia* with SWS pollen type include Sect. *Artemisia* and its habitats range from forest to desert, although most of the taxa are confined to humid environments from forest to grassland with a global distribution and the highest mean annual precipitation (MAP, Figs. 17b, 20c, 21). (iv) If the SWS pollen type and the SG pollen type appear together, the range of vegetation types could be reduced to grassland desert and desert through niche coexistence (Fig. 17b).
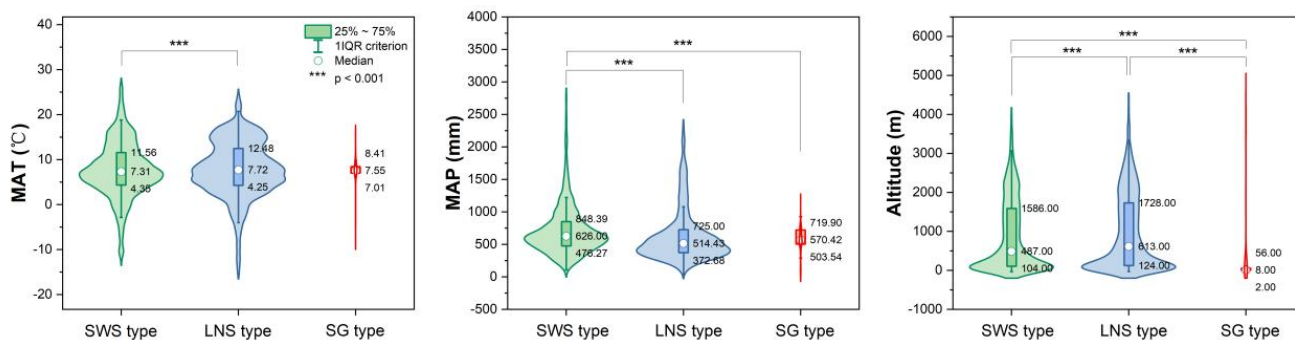

Figure 21. Violin diagrams of three pollen types from *Artemisia*, showing the variations (25%-75%) in MAT, MAP, and altitude. Asterisks indicate statistically significant differences (p < 0.001).

In addition, we noticed that *Kaschgaria brachanthemoides* as an outgroup of *Artemisia* lives in dry mountain valleys or dry riverbeds of Northwest China (Toksun) and Kazakhstan, with highly characteristic pollen (Fig. 14a), narrow habitats (Fig. 17b), and regional distribution (Fig. 16-34) and has the potential to indicate some specific habitats."

Following your constructive suggestion, we randomly sampled the coordinates of the distribution of 33 species of *Artemisia* using the R language package "sampling" (Tillé and Matei, 2021). If the number of species distribution coordinates was more than ten, ten coordinates are randomly selected, while any number below 10 was used completely. The refined biomes of these sampling coordinates were obtained using Arcgis 10.2 software according to 867 terrestrial biomes distinct units by Olson et al. (2001).

The proportion of the three pollen types of *Artemisia* in the four vegetation types (forest, grassland, desert, and saline) was determined by a specimen sampling survey, with the SWS pollen type having the highest proportion in the forest, the LNS pollen type having a more balanced distribution, and the SG pollen type having the highest proportion in the desert and saline (Figure R1). These were generally consistent with the results of our literature review (Fig. 16b).
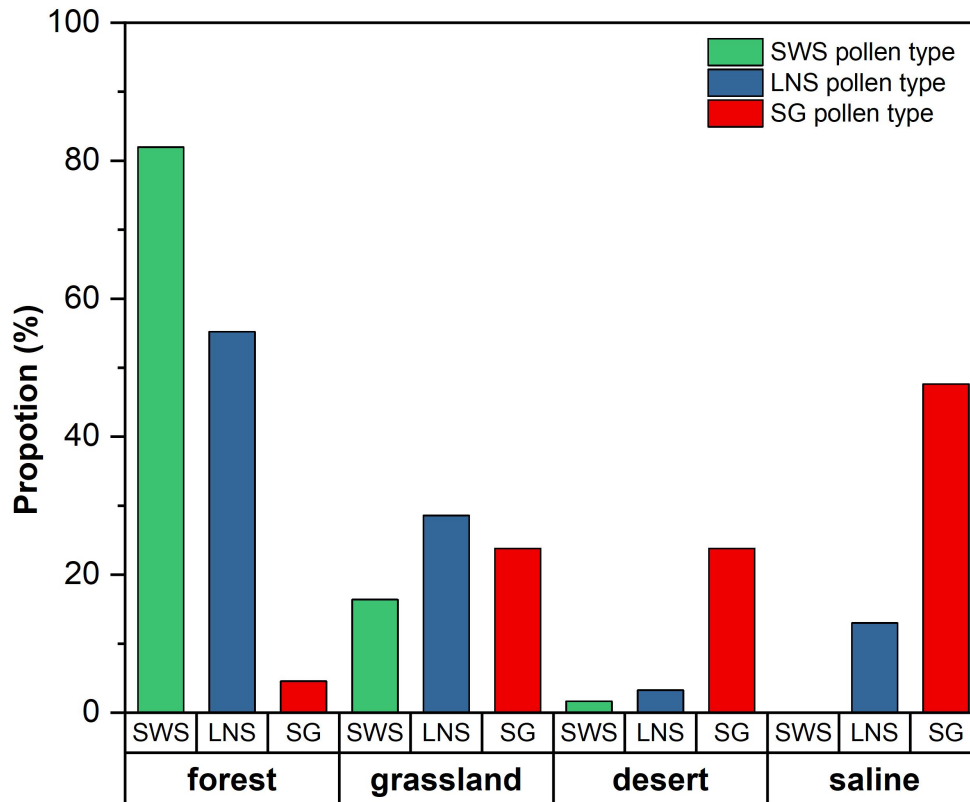
Figure R1. Distribution of randomly sampled specimens showing the proportion of the three pollen types of *Artemisia* in the four vegetation types.

C2. The data are well documented. The rich dataset of detailed high quality pollen fotos is crucial for palynological studies and a treasure in itself. The dataset on measurements of pollen grains is quite small, considering only one specimen per species. It should definitively, and hopefully will, be enlarged in the future, but for sure will serve as a basis or blueprint to other overarching statistical analyses on pollen morphology.

R2. Thank you for your valuable suggestion! We modify the section 6 Summary as follows.

Lines 406-436: "To cover the maximum range of *Artemisia* pollen morphological variation, we provide a pollen dataset of 36 species from 9 clades and 3 outgroups of *Artemisia* constrained by the phylogenetic framework, containing high-quality pollen photographs under LM and SEM, statistical data of pollen morphological traits, together with their source plant distribution, and corresponding environmental factors. Here, we attempt to decipher the underlying causes of the long-standing disagreement in the palynological community on the correlation between *Artemisia* pollen and aridity by recognizing the different ecological implications of *Artemisia* pollen types.

This dataset should work well for identifying and classifying *Artemisia* pollen from Neogene and Quaternary sediments. While *Artemisia* pollen grains are uniform in morphology under LM, different types

can be recognized under SEM. So, the single-grain technique for picking out fossil pollen grains and photographing the same grains under LM and SEM should provide valuable insights in the diversity of fossil *Artemisia* (Ferguson et al., 2007; Grímsson et al., 2011; Grímsson et al., 2012; Halbritter et al., 2018). Furthermore, those *Artemisia* pollen grains could then be compared with the rich photographs from this dataset, and together with the key provided here, possibly attributed to one of the three *Artemisia* pollen types, which in turn may provide a link to the different habitat ranges.

However, the application of this dataset probably may not work well for the Palaeogene, as 1) *Artemisia* might have originated in the Palaeocene, although there is no evidence for a specific location or time interval of its origin (e.g. Ling 1982; Wang 2004; Miao 2011); 2) both the lack of macrofossils of *Artemisia* and the strong pollen similarity between *Artemisia* and its closely related taxa under LM might lead to confusion and more uncertainty in tracing the origin of *Artemisia*. On the other hand, the present dataset provides a potential morphological tool to distinguish *Artemisia* pollen grains from those of its related taxa at the SEM level and may shed light on the origin of this genus in the Palaeogene.

Moreover, these pollen photographs also have potential and the possibility to be used for deep learning research. We are attempting to automatically identify pollen images using pollen assemblages from the eastern Central Asian desert as an example with deep convolutional neural network (DCNN) of artificial intelligence. Pollen images of the many species of *Artemisia* provided here, and the increasing number of intraspecific replications in the future, will all serve for projected image identification research.

Finally and most importantly, the *Artemisia* pollen dataset as designed is open and expandable for new pollen data from *Artemisia* worldwide in order to better serve the global environment assessment and refined reconstruction of vegetation in the geological past as a basis or blueprint for other overarching statistical analyses on pollen morphology."

C3. With respect to the dataset which was downloaded from GBIF and WorldClim, I do not see the necessary to provide them here (except to justify the figures in the manuscript). GBIF is a dynamic database growing constantly and new retrievals from GBIF will give more detailed information with every day. Moreover, GBIF data have to be carefully quality checked before further analyses, but there is no awareness of this mentioned here in the text.

R3. This is a helpful suggestion! The GBIF species distribution data we have used were updated until 09 November 2021 and were included in the dataset to 1) visualize the global distribution of 36 species and three *Artemisia* pollen types in terrestrial biomes, and 2) attempt to determine the preferences of different pollen

types for different climatic indicators by correlating the environmental factors of the distribution coordinates of these specimens.

Although the climatic parameters may not be well-developed at present, the preferences of different pollen types for distinct climatic indicators in these data were consistent with the above literature survey, e.g. *Artemisia* with SWS pollen type is primarily found in humid environments from forest to grassland with the highest MAP (Figs. 17b, 20c, 21), whereas *Artemisia* with the SG pollen generally grows in dry habitats ranging from grassland desert to desert and coastal saline-alkaline environments with the lowest altitude (Figs. 17b, 20c, 21).

We have added some instructions on the quality check of GBIF data in the second paragraph of section 2.2 Data acquisition.

Lines 136-143: "The scientific names of selected taxa were standardized according to Plants of the World Online (https://powo.science.kew.org/). The specimen sampling coordinates of the corresponding taxa were obtained from the Global Biodiversity Information Facility (GBIF, https://www.gbif.org/). Only preserved specimens were filtered for GBIF data given their well-documented geographical information and the availability of specimens as definitive vouchers. The distribution data on observations and cultivated collections provided by GBIF were excluded because they may contain incorrect identification or incorrect geo-referencing (Brummitt et al., 2020). Next, the distribution data was standardized cleaned using R package "CoordinateCleaner" (Zizka et al., 2019); no outliers were found."

C4. In general, I consider this manuscript ready for publications after minor revisions. I have a few remarks annotated in the pdf.

R4. Thank you very much for your kind correction! The manuscript has been revised following your remarks.

The references for the above responses are listed as below:

Brummitt, N., Araujo, A. C., and Harris, T.: Areas of plant diversity-What do we know?, Plants People Planet, 3, 33-44, https://doi.org/10.1002/ppp3.10110, 2021.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'Amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., and Kassem, K. R.: Terrestrial ecoregions of the worlds: A new map of life on Earth, Bioscience, 51, 933-938, https://doi.org/10.1641/0006-3568(2001)051[0933:teotwa]2.0.co;2, 2001.

Tillé, Y. and Matei, A.: sampling: Survey Sampling. R package version 2.9., https://CRAN.R-project.org/package=sampling, 2021.

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V., and Antonelli, A.: CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases, Methods in Ecology

and Evolution, 10, 744-751 %U https://onlinelibrary.wiley.com/doi/abs/710.1111/2041-1210X.13152, 10.1111/2041-210X.13152, 2019.