Supplement of Clim. Past, 18, 2181–2210, 2022
https://doi.org/10.5194/cp-18-2181-2022-supplement

Climate of the Past
Open Access
EGU

Supplement of

Integrating plant wax abundance and isotopes for paleo-vegetation and paleoclimate reconstructions: a multi-source mixing model using a Bayesian framework

Deming Yang and Gabriel J. Bowen

Correspondence to: Deming Yang (deming.yang@utah.edu)

Supplementary Material

This file includes:

Supplementary text

Supplementary references

Supplementary tables S1 – S7

Supplementary figures S1 – S8

Supplementary Text

Compilation of data and user-defined source groups

In the first two case studies presented in this article, the source groups are user-defined, based on information from published plant *n*-alkane records. In case study 1, the three source groups (terrestrial plants, submerged aquatic macrophytes, and algae) were assigned to the published data records in Supplementary Data EA-2 (Yang, 2022) based on either taxonomic identification or sample description from the original publications. Details on the taxonomic identification, sample description, lipid extraction methods, additional environmental information, and referenced publications are listed in Supplementary Data EA-2 (Yang, 2022). In case study 2, the three source groups (rainforest C_3 plants, savanna C_3 plants, C_4 plants) were assigned to the published data records in in Supplementary Data EA-3 and EA-4 (Yang, 2022) based on either taxonomic identification or description of sample location from the original publications. Details on the taxonomic identification, sample description, lipid extraction methods, and referenced publications are listed in Supplementary Data EA-3 and EA-4 (Yang, 2022).

In case study 3, the same generic biome groups (rainforest C_3 plants, savanna C_3 plants, C_4 plants) were adopted from case study 2, to account for the vegetation effect on *n*-alkane δ^2H. Ideally, *n*-alkane δ^2H of the same African plants summarized in Supplementary Data EA-3 (Yang, 2022) would make the best estimates of group specific ^2H fractionation between *n*-alkane and mean annual precipitation (ε_a). Unfortunately, there are relatively few published records of δ^2H of *n*-alkanes in African plants. As a substitute, the global compilation by Sachse et al. (2012) was adopted as described in Sect. 2.2.3.5, then the three source groups were assigned to the record based on either taxonomic identification or description of sample location from the original publications. Data from more recent publications (Kahmen et al., 2013; Feakins et al., 2016; Griepentrog et al., 2019) were also used to supplement the compilation of Sachse et al. (2012). For data with multiple samples of a single species, species means were calculated to avoid overrepresentation of one species/sample location. The source groups were assigned as described in the previous paragraph. Details on the taxonomic identification, sample location, *n*-alkane δ^2H, referenced OIPC values, calculated ε_a, and referenced publications are listed in Supplementary Data, EA-5 (Yang, 2022).

The potential influence of different lipid extraction methods

We listed the details regarding the lipid extraction methods used in each publication in the Supplementary Data (Yang, 2022), due to the potential influence of the different methods as well as the organic solvents on leaf wax lipid extraction yield and composition. In this study, the influence of extraction methods on the lipid yield and composition is assumed to be minimal. While extraction yield has been found to be different with different methods and solvents (Ardenghi et al., 2017; Weber and Schwark, 2020), it most likely affects the chain-specific *n*-alkane concentrations of the samples, which are known to be associated with large uncertainties due to their log-normal distributions. The potential influence of extraction methods on sample specific *n*-alkane concentrations should have little effects on the prior distributions when a large number of samples are included in the calculation of prior parameters (Supplementary Data EA-2 to EA-4, Yang, 2022). Moreover, the results of sensitivity tests in Sect. 3.3 illustrate that the model is more sensitive to the $\delta^{13}C$ values of the prior distribution and the likelihood evaluation than relative abundance. Since $\delta^{13}C$ values are independent to the *n*-alkane concentrations, extraction methods should have little influence on the central tendency of the posterior distribution of the mixing ratios or chain-specific mixing ratios reported in Sects. 3.1 and 3.2. More systematic investigations will help to evaluate the influence of extraction methods on chain-specific *n*-alkane concentrations, and subsequently, to constrain model uncertainty.

Comparing prior and posterior distributions

Because the model exhaustively explores the multivariate parameter space, the posterior distributions of some parameters may differ from their priors, depending on the data values. Such deviations can be informative on how constrained the parameters are by the data, as well as any potential biases in the prior parameter estimations. Here, we report the comparisons between prior and posterior distributions using examples from the first two case studies.

The first comparison is based on the model output of the QHS-5S sample. For per sample *n*-alkane concentrations, the posterior distributions are quite similar to the prior distributions, with some small deviations observed in the terrestrial source in the *n*-C_{29} and the *n*-C_{31} chains (Fig. S4). For per sample *n*-alkane $\delta^{13}C$, the posterior distributions are identical to the prior distributions (Fig. S5).

The second comparison is based on the model output of the Asso sample. For per sample *n*-alkane concentrations, the posterior distributions are quite similar to the prior distributions, with some small deviations observed in the C_4 plants in all three chains, in the savanna C_3 plants in the *n*-C_{33} chain, as well as in the rainforest C_3 plants in the *n*-C_{31} and the *n*-C_{33} chains (Fig. S6). For per sample *n*-alkane $\delta^{13}C$, the posterior distributions are identical to the prior distributions (Fig. S7).

Because the proxy system model used here is simple, the posterior distributions of the two case studies are generally quite similar to the priors (Figs. S4 to S7). When deviations from the prior distribution do occur (Figs. S4 and S6), they reflect the fact that the *n*-alkane chain length distribution of the sedimentary samples differ substantially from the central tendency of any source. For instance, the QHS-5S sample displays a *n*-C_{31} dominance (Fig. 5), which is only possible if the terrestrial source has a greater *n*-C_{31} dominance than reflected in the prior (Fig. S4). Similarly, the strong *n*-C_{29} dominance in the Asso sample

(Fig. 8) is best explained by somewhat higher than expected *n*-C_{29} abundance in the rainforest C_3 source (Fig. S6). Such information can be used to refine our prior assumptions associated with the model, and to provide information on potential biases or alternative interpretations.

Comparing vegetation reconstruction in case study 3 to published pollen record in another marine core off the Zambezi River Mouth

Terrestrial pollen records from marine sediments have been widely used to reconstruct terrestrial vegetation change through time. The advantage of using the pollen record is that it can reveal taxon- or biome-specific vegetation change, which is much more detailed than interpretations based on e.g., stable carbon isotope record of *n*-alkanes. Given that the two methods are inherently associated with their own limitations and biases, such as those related to pollen/*n*-alkane production, transportation, and preservation, we carry out a qualitative comparison between the FLMC estimations based on an *n*-alkane record from the GIK 16160-3 marine core (Wang et al., 2013), and pollen fractions from the GeoB9311-1 marine core (Dupont and Kuhlmann, 2017), taken ~3 degrees south of GIK 16160-3. The latter has a temporal range of *ca*. 0.9 – 250 ka BP (Dupont and Kuhlmann, 2017). We use the portion of the record that overlaps with the GIK 16160-3 core (*ca*. 0.9 – 33.5 ka BP) in our comparison. Due to a low pollen count, the data point dated at 19.1 ka BP of the GeoB9311-1 core is excluded from the comparison.

One of the challenges with interpreting the pollen record is determining how the taxa should be grouped to represent the three source groups presented in case studies 2 and 3. We have opted to reference the biome-consistent grouping criteria of Dupont and Kuhlmann (2017), and identified 6 groups as follows: Grasses and sedges, Forest, Woodland, Afromontane, Swamp, and ferns (Table S6, Supplementary Data EA-6, Yang, 2022). To represent the three source groups used in case studies 2 and 3, the Grasses and sedges group are considered quasi-comparable to the C_4 plants source; Woodland taxa are quasi-comparable to the Savanna C_3 plants; and Forest, Afromontane, and Swamp (excluding sedges) taxa are quasi-comparable to the Rainforest C_3 plants (Table S6). We grouped the Afromontane taxa with forest taxa primarily because they both represent habitats with high relative humidity, which influences the distributions of their *n*-alkane $\delta^{13}C$ and ε_a. It is worth mentioning that Afromontane taxa are not well studied for their *n*-alkane concentrations and isotopes, but that altitude effects on *n*-alkane concentrations have been found in the Andes (Feakins et al., 2016). Fern spores are excluded in the comparison primarily because ferns are not represented in the datasets used to specify the prior distributions of *n*-alkane concentrations and isotopes.

The MAP estimates of the three *n*-alkane sources (Sect. 3.3, Fig. 11) are quite different from the calculated pollen fractions (Fig. S8). The pollen fractions of grasses and sedges are in general about 0.2 higher than the C_4 plant FLMCs. The pollen fractions of woodland/savanna are consistently low (< 0.2), while FLMCs of savanna C_3 plants are consistently the dominant biome through time (FLMC > 0.5). The pollen fractions of forest, Afromontane and swamp taxa are consistently high (> 0.3), while FLMCs of rainforest C_3 plants are consistently low (< 0.2), except for the peak value during HS1. It is worth noting that the FLMC estimates for the Holocene (Sect. 3.3) are more consistent with the present-day distribution of biomes within the catchment area of the Zambezi River than are the pollen fractions (Wang et al., 2013; Dupont and Kuhlmann, 2017).

Despite the differences, both the *n*-alkane and the pollen records show a similar trend of decrease in grasses and sedges/ C_4 plants, and a similar trend of increase in woodland/savanna C_3 plants from the Last Glacial Maximum to the Holocene (Fig. S8). None of the rank-based correlation tests shows significance (Table S7), although the results are based on a comparison of 8 data points only. The correlation coefficients of the grasses and sedges/ C_4 plants and the woodland/savanna C_3 plants are positive, while the correlation coefficients of the forest, swamp and Afromontane/rainforest group are negative (Table S7). One clear discrepancy is in the estimated forest fraction during HS1, a peak is reconstructed using the *n*-alkane method, but the pollen fraction is low. The negative correlation may be due to the influence from the Afromontane biome, which is not well characterized in the prior distributions of *n*-alkanes (Supplementary Data EA-3, Yang, 2022). If the Afromontane biome also shows a *n*-C_{29} dominance, this could explain the *n*-C_{29} dominance in the record from the LGM to the end of HS1. Future studies that aim to document the distribution of *n*-alkane concentration and stable isotopes of the Afromontane taxa will help to account for this biome in *n*-alkane-based analyses and better understand the discrepancies between the two methods. On the other hand, the positive correlations suggest that both methods can detect similar patterns of vegetation change or change in the source area of the river catchment. This further suggests that the two methods can be used together to better understand the influence of vegetation on the interpretation of other climate proxies in sedimentary archives.

Supplementary References

Ardenghi, N., Mulch, A., Pross, J., and Maria Niedermeyer, E.: Leaf wax n-alkane extraction: An optimised procedure, Org Geochem, 113, 283-292, https://doi.org/10.1016/j.orggeochem.2017.08.012, 2017.

Dupont, L. M. and Kuhlmann, H.: Glacial-interglacial vegetation change in the Zambezi catchment, Quaternary Sci Rev, 155, 127-135, https://doi.org/10.1016/j.quascirev.2016.11.019, 2017.

Feakins, S. J., Peters, T., Wu, M. S., Shenkin, A., Salinas, N., Girardin, C. A. J., Bentley, L. P., Blonder, B., Enquist, B. J., Martin, R. E., Asner, G. P., and Malhi, Y.: Production of leaf wax *n*-alkanes across a tropical forest elevation transect, Org Geochem, 100, 89-100, https://doi.org/10.1016/j.orggeochem.2016.07.004, 2016.

Griepentrog, M., De Wispelaere, L., Bauters, M., Bodé, S., Hemp, A., Verschuren, D., and Boeckx, P.: Influence of plant growth form, habitat and season on leaf-wax *n*-alkane hydrogen-isotopic signatures in equatorial East Africa, Geochim Cosmochim Ac, 263, 122-139, https://doi.org/10.1016/j.gca.2019.08.004, 2019.

Kahmen, A., Hoffmann, B., Schefuß, E., Arndt, S. K., Cernusak, L. A., West, J. B., and Sachse, D.: Leaf water deuterium enrichment shapes leaf wax n-alkane δD values of angiosperm plants II: Observational evidence and global implications, Geochim Cosmochim Ac, 111, 50-63, https://doi.org/10.1016/j.gca.2012.09.004, 2013.

Sachse, D., Billault, I., Bowen, G. J., Chikaraishi, Y., Dawson, T. E., Feakins, S. J., Freeman, K. H., Magill, C. R., McInerney, F. A., van der Meer, M. T. J., Polissar, P., Robins, R. J., Sachs, J. P., Schmidt, H.-L., Sessions, A. L., White, J. W. C., West, J. B., and Kahmen, A.: Molecular paleohydrology: Interpreting the hydrogen-isotopic composition of lipid biomarkers from photosynthesizing organisms, Annu Rev Earth Pl Sc, 40, 221-249, 10.1146/annurev-earth-042711-105535, 2012.

Wang, Y. V., Larsen, T., Leduc, G., Andersen, N., Blanz, T., and Schneider, R. R.: What does leaf wax δD from a mixed C3/C4 vegetation region tell us?, Geochim Cosmochim Ac, 111, 128-139, https://doi.org/10.1016/j.gca.2012.10.016, 2013.

Weber, J. and Schwark, L.: Epicuticular wax lipid composition of endemic European Betula species in a simulated ontogenetic/diagenetic continuum and its application to chemotaxonomy and paleobotany, Sci Total Environ, 730, 138324, https://doi.org/10.1016/j.scitotenv.2020.138324, 2020.

Yang, D.: SPATIAL-Lab/LipidMM: Bok choy (v. 1.0.5), Zenodo [code], https://doi.org/10.5281/zenodo.7025765, 2022.

Supplementary Tables

Table S1. Means and variance-covariance matrices of *n*-alkane $\delta^{13}C$ of terrestrial, aquatic macrophyte and algae n-alkane sources around Lake Qinghai.

$\delta^{13}C$ Means (‰, VPDB)	*n*-C_{27}	*n*-C_{29}	*n*-C_{31}
Terrestrial	−32.8	−33.5	−33.5
Aquatic macrophyte	−22.6	−23.3	−23.8
Algae	−31.2	−31.0	−32.1
Terrestrial $\delta^{13}C$	*n*-C_{27}	*n*-C_{29}	*n*-C_{31}
V-covariance matrix			
n-C_{27}	1.9684	−0.3273	0.8149
n-C_{29}	-	1.4068	−0.0304
n-C_{31}	-	-	2.0529
Aquatic macrophyte $\delta^{13}C$	*n*-C_{27}	*n*-C_{29}	*n*-C_{31}
V-covariance matrix			
n-C_{27}	18.1873	16.5000	16.1755
n-C_{29}	-	15.6888	15.4095
n-C_{31}	-	-	18.9184
Algae $\delta^{13}C$	*n*-C_{27}	*n*-C_{29}	*n*-C_{31}
V-covariance matrix			
n-C_{27}	6.3562	−4.2521	−1.9812
n-C_{29}	-	9.9662	4.9155
n-C_{31}	-	-	2.8562

Table S2. Means and variance-covariance matrices of ln-transformed n-alkane concentration of terrestrial, aquatic macrophyte and algae n-alkane sources around Lake Qinghai.

ln(concentration) ln(μg/g)	*n*-C_{27}	*n*-C_{29}	*n*-C_{31}
Terrestrial	3.6781	4.3077	3.5228
Aquatic macrophyte	2.7477	2.2110	−0.0476
Algae	−0.8770	−1.0454	−1.6011
Terrestrial ln(concentration) V-covariance matrix	*n*-C_{27}	*n*-C_{29}	*n*-C_{31}
n-C_{27}	1.1286	0.9208	0.5662
n-C_{29}	-	1.8794	1.3238
n-C_{31}	-	-	2.7186
Aquatic macrophyte ln(concentration) V-covariance matrix	*n*-C_{27}	*n*-C_{29}	*n*-C_{31}
n-C_{27}	1.1519	0.9627	0.4091
n-C_{29}	-	0.9453	0.5098
n-C_{31}	-	-	1.1875
Algae ln(concentration) V-covariance matrix	*n*-C_{27}	*n*-C_{29}	*n*-C_{31}
n-C_{27}	0.7614	0.6063	0.4345
n-C_{29}		0.7170	0.5284
n-C_{31}			0.6045

Table S3. Means and variance-covariance matrices of n-alkane $\delta^{13}C$ of C_4 grasses, savanna C_3 plants and rainforest C_3 plants of sub-Saharan Africa.

$\delta^{13}C$ Means (‰, VPDB)	*n*-C_{29}	*n*-C_{31}	*n*-C_{33}
C_4 grasses	−21.7	−21.6	−22.0
Savanna C_3	−34.0	−34.5	−34.9
Rainforest C_3	−38.1	−38.4	−38.9
C_4 grasses $\delta^{13}C$	*n*-C_{29}	*n*-C_{31}	*n*-C_{33}
V-covariance matrix			
n-C_{29}	3.2710	2.4707	2.2354
n-C_{31}	-	3.1060	2.6191
n-C_{33}	-	-	3.5775
Savanna C_3 $\delta^{13}C$	*n*-C_{29}	*n*-C_{31}	*n*-C_{33}
V-covariance matrix			
n-C_{29}	6.6406	5.5126	3.8458
n-C_{31}	-	6.7642	5.4598
n-C_{33}	-	-	6.8613
Rainforest C_3 $\delta^{13}C$	*n*-C_{29}	*n*-C_{31}	*n*-C_{33}
V-covariance matrix			
n-C_{29}	6.8823	7.4840	7.4830
n-C_{31}	-	9.0163	9.4427
n-C_{33}	-	-	11.1913

145

Table S4. Means and variance-covariance matrices of ln-transformed n-alkane concentration of C_4 grasses, savanna C_3 plants and rainforest C_3 plants of sub-Saharan Africa.

ln(concentration) ln(µg/g)	*n*-C_{29}	*n*-C_{31}	*n*-C_{33}
C_4 grasses	3.4433	4.4287	3.7499
Savanna C_3	3.4862	3.6974	2.4946
Rainforest C_3	3.9780	3.8278	2.0584
C_4 grasses ln(concentration) V-covariance matrix	*n*-C_{29}	*n*-C_{31}	*n*-C_{33}
n-C_{29}	1.3710	0.6492	0.4509
n-C_{31}	-	1.5669	0.6833
n-C_{33}	-	-	1.6065
Savanna C_3 ln(concentration) V-covariance matrix	*n*-C_{29}	*n*-C_{31}	*n*-C_{33}
n-C_{29}	1.9701	1.3619	1.3795
n-C_{31}	-	2.5040	2.3458
n-C_{33}	-	-	4.0725
Rainforest C_3 ln(concentration) V-covariance matrix	*n*-C_{29}	*n*-C_{31}	*n*-C_{33}
n-C_{29}	2.7555	1.1830	0.6894
n-C_{31}	-	1.9080	1.4103
n-C_{33}	-	-	2.2133

Table S5. Means and variance-covariance matrices of ^2H fractionation between *n*-alkane and mean annual precipitation (ε_a) of C_4 grasses, savanna C_3 plants and rainforest C_3 plants based on the global data compilation in Supplementary Data EA-5 (Yang, 2022).

ε_a Means	*n*-C_{27}	*n*-C_{29}	*n*-C_{31}	*n*-C_{33}
C_4 grasses	−136.0	−139.8	−143.5	−132.0
Savanna C_3	−110.7	−112.5	−105.6	−107.1
Rainforest C_3	−117.8	−118.9	−118.1	−126.9
ε_a C_4 grasses	*n*-C_{27}	*n*-C_{29}	*n*-C_{31}	*n*-C_{33}
V-covariance matrix				
n-C_{27}	392.0	397.2	298.9	180.7
n-C_{29}	-	575.3	445.5	340.6
n-C_{31}	-	-	414.6	309.6
n-C_{33}	-	-	-	262.5
ε_a Savanna C_3	*n*-C_{27}	*n*-C_{29}	*n*-C_{31}	*n*-C_{33}
V-covariance matrix				
n-C_{27}	597.3	487.1	547.1	275.4
n-C_{29}	-	452.1	449.6	231.1
n-C_{31}	-	-	735.6	264.3
n-C_{33}	-	-	-	187.6
ε_a Rainforest C_3	*n*-C_{27}	*n*-C_{29}	*n*-C_{31}	*n*-C_{33}
V-covariance matrix				
n-C_{27}	1182.7	1217.3	1215.4	1238.7
n-C_{29}	-	1323.5	1357.8	1393.7
n-C_{31}	-	-	1455.6	1497.7
n-C_{33}	-	-	-	1581.3

Table S6. Biome designations used in the GeoB9311-1 pollen record (Dupont and Kuhlmann, 2017) and their abbreviations used in Supplementary Data, EA-6 (Yang, 2022).

Biome	Abbreviation used in Supplementary Data, EA-6	Designated source group in case study 3
Forest	RF	Rainforest C_3
Woodland	SV	Savanna C_3
Grasses and sedges	C4	C_4 plants
Swamp (except for sedges)	SW	Rainforest C_3
Afromontane	MT	Rainforest C_3

Table S7. Results of non-parametric correlation tests between the maximum a posteriori estimate (MAPE) of estimated FLMC based on simultaneous evaluation of 4 n-alkane chains in the GIK 16160-3 core, and pollen fractions from the GeoB9311-1 core (n = 8); abbreviations of biomes are listed in Table S6.

Source group FLMC	Pollen fraction	Spearman's ρ	Kendal's τ
MAPE Rainforest C_3	RF + SW + MT	$\rho = -0.3810$ $S = 116$, P-value = 0.3599	$\tau = -0.2857$ $T = 10$, P-value = 0.3988
MAPE Savanna C_3	SV	$\rho = 0.6190$ $S = 32$, P-value = 0.1150	$\tau = 0.4286$ $T = 20$, P-value = 0.1789
MAPE C_4 plants	C4	$\rho = 0.7143$ $S = 24$, P-value = 0.0576	$\tau = 0.5714$ $T = 22$, P-value = 0.0601

Supplementary Figures

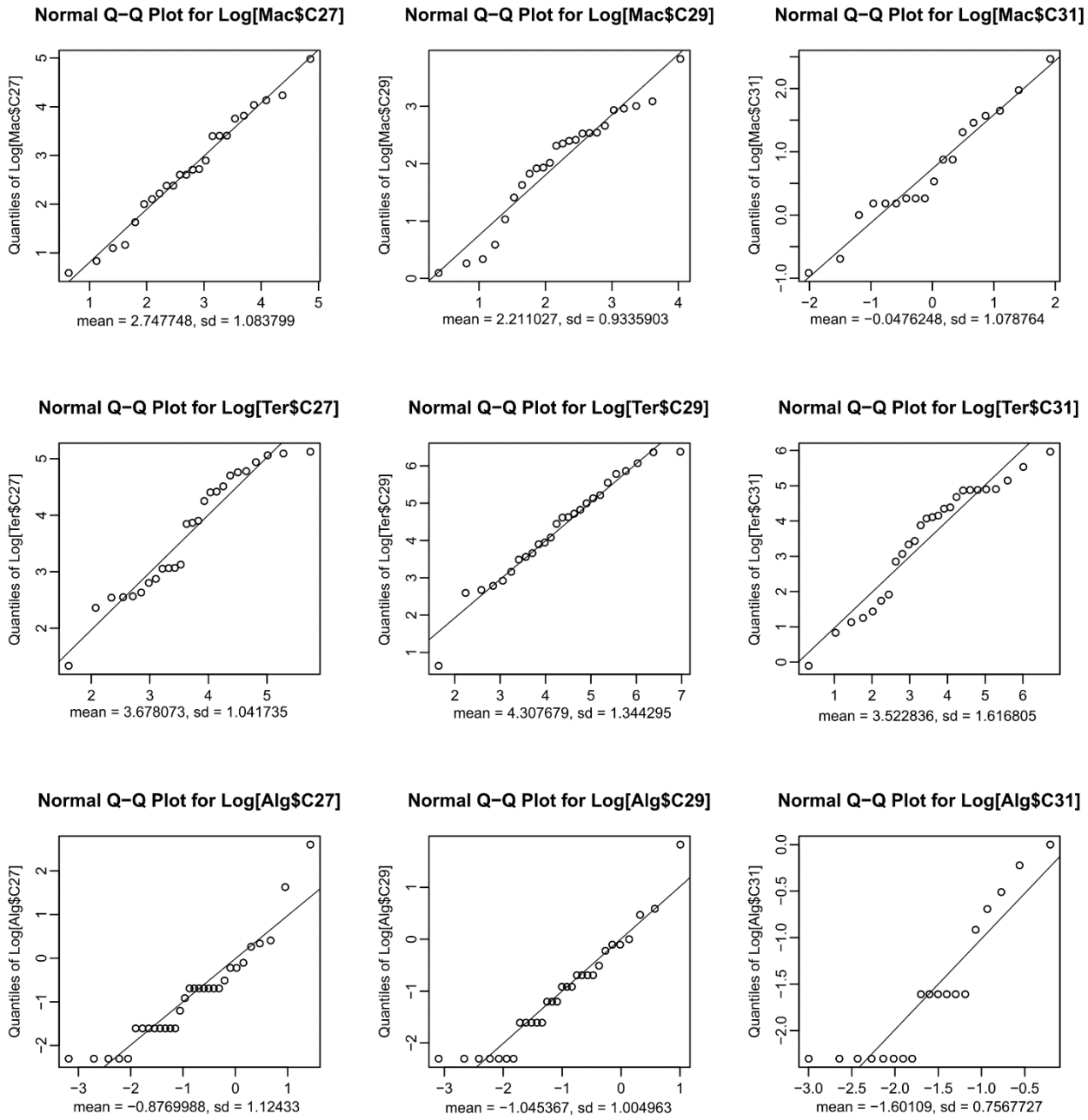

Figure S1. Quantile-quantile plots between parameterized log-normal distribution and empirical data used in case study 1, demonstrating goodness of fit between the data and the model; columns are the three *n*-alkane chains used in the case study, rows are the end members used in the case study; diagonal lines indicate the 1:1 relationship.

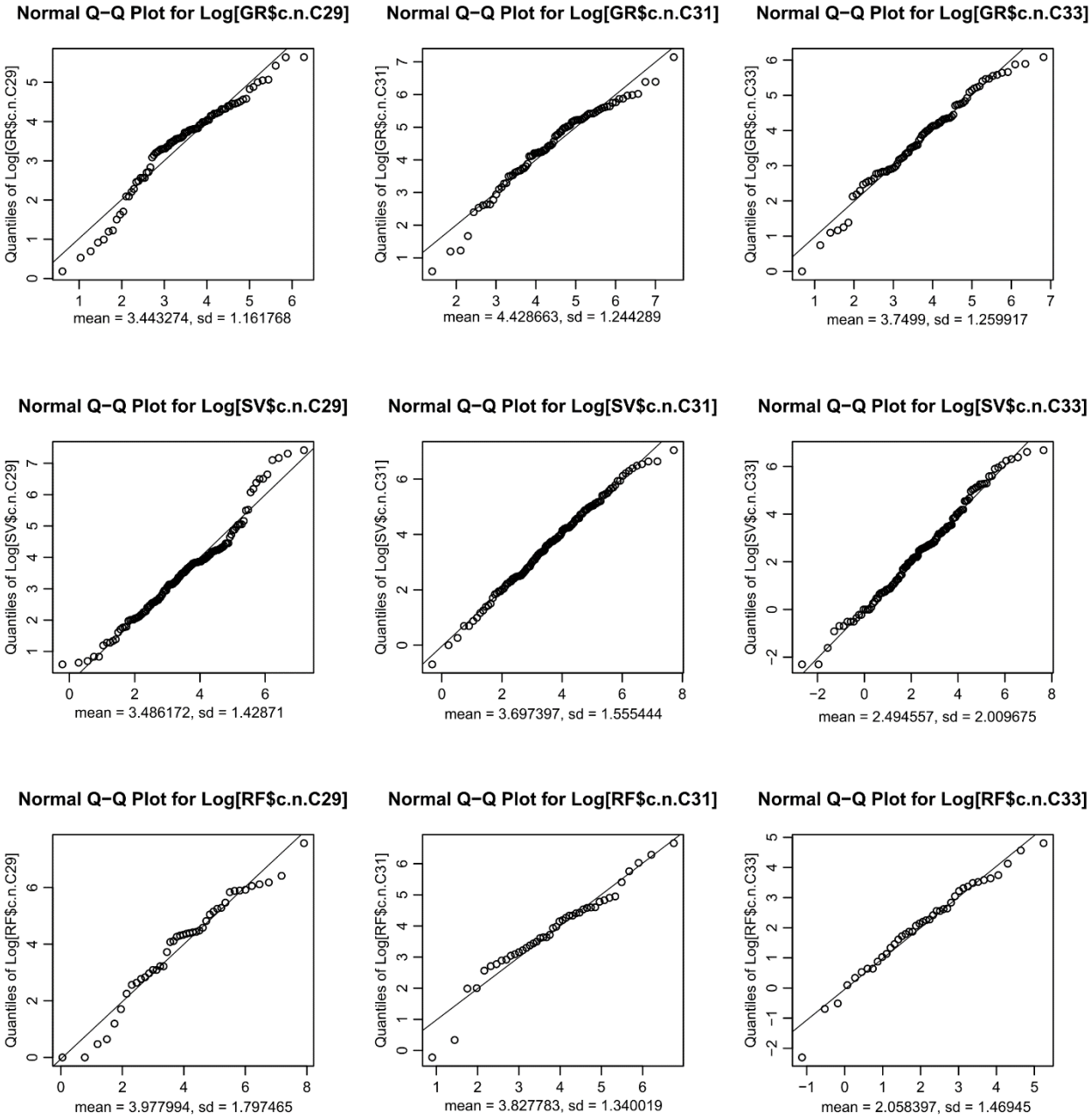

Figure S2. Quantile-quantile plots between parameterized log-normal distribution and empirical data used in case study 2, demonstrating goodness of fit between the data and the model; columns are the three *n*-alkane chains used in the case study, rows are the end members used in the case study; diagonal lines indicate the 1:1 relationship.

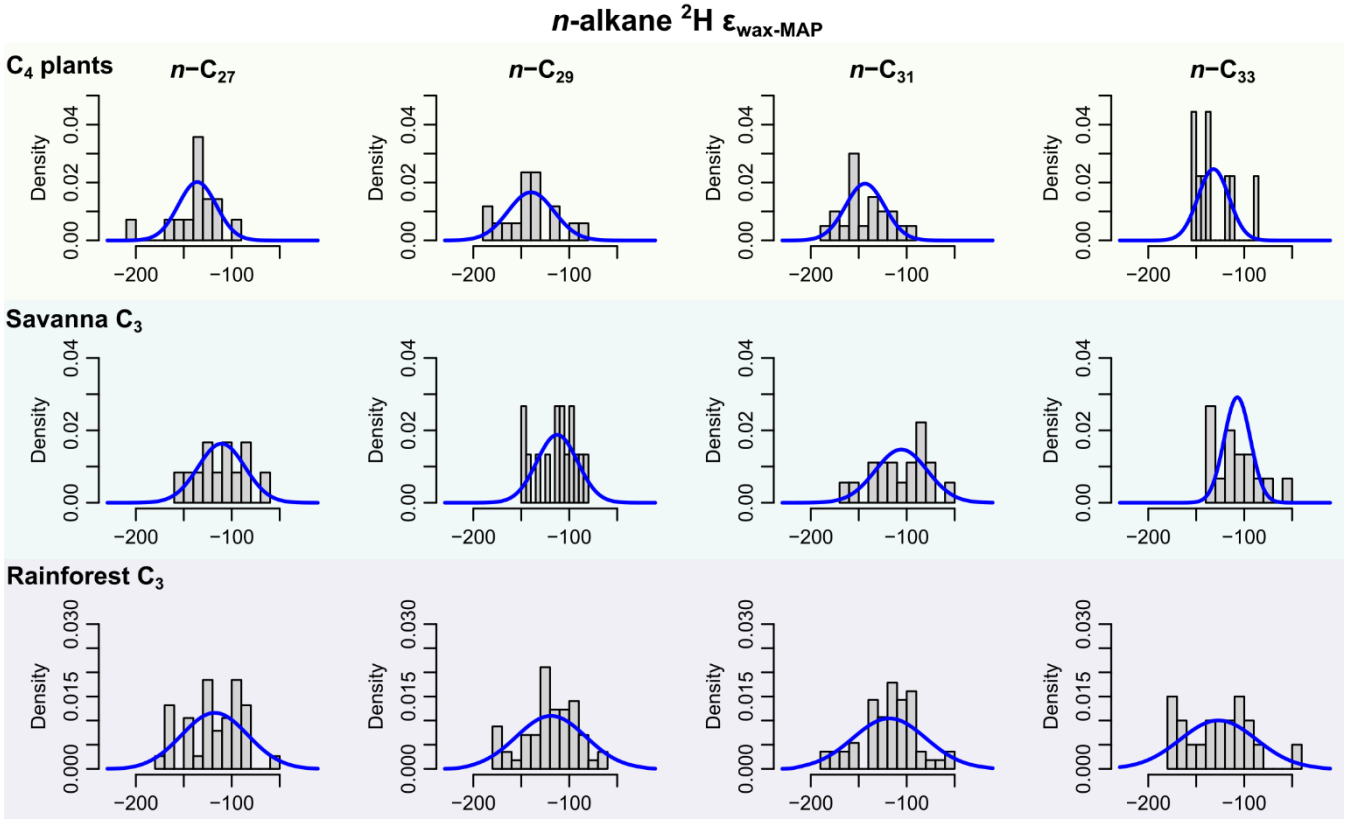

Figure S3. Prior distributions of calculated ^{2}H fractionation (ε) between plant wax and MAP of the three source groups; raw data of the global compilation are presented in Supplementary Data, EA-5; estimated prior parameters of each source, including the means and variance-covariance matrices, are reported in Table S5.

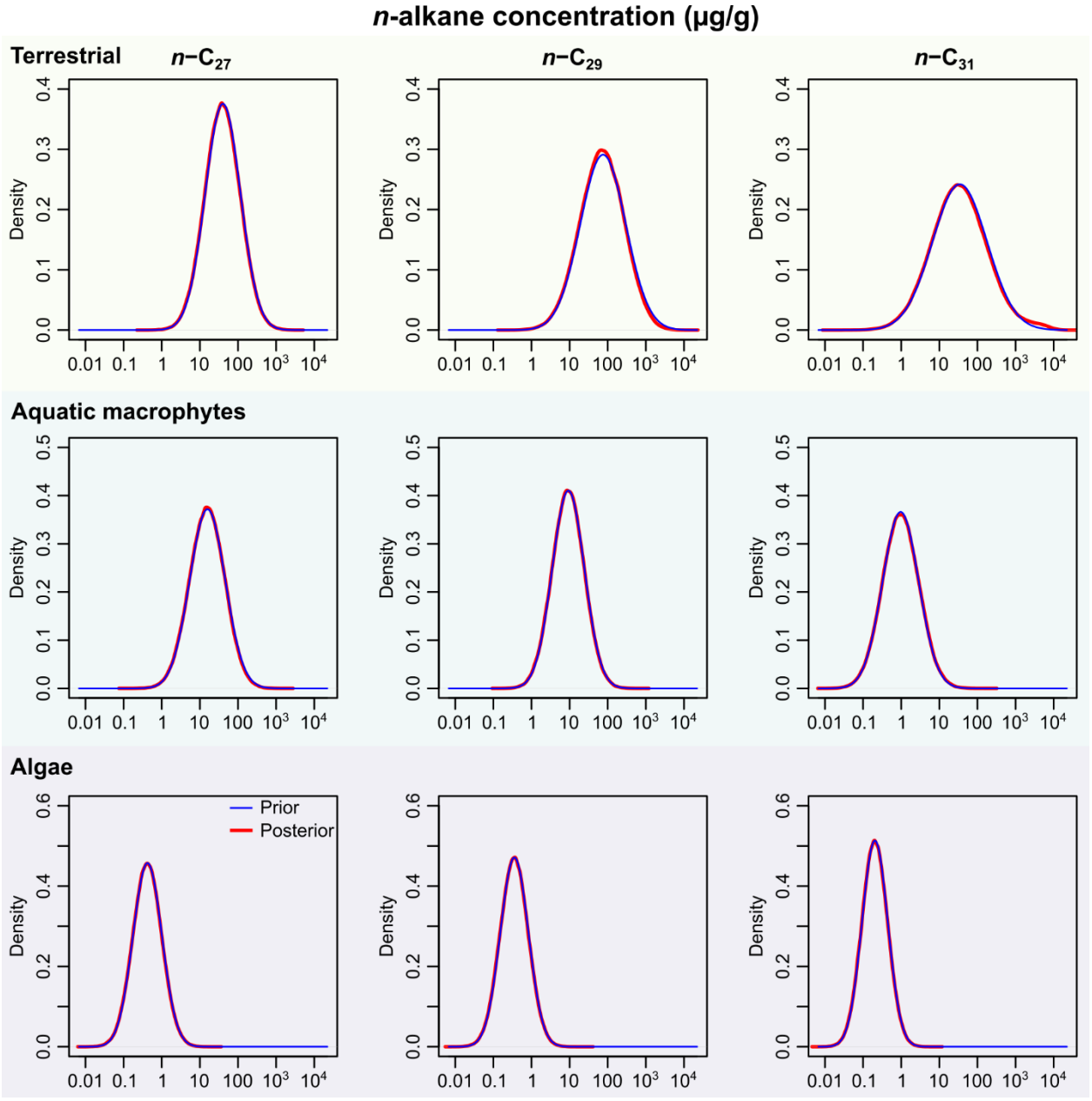

Figure S4. Comparison between the *n*-alkane concentration prior distribution of case study 1 and the posterior distribution based on the QHS-5S sample.

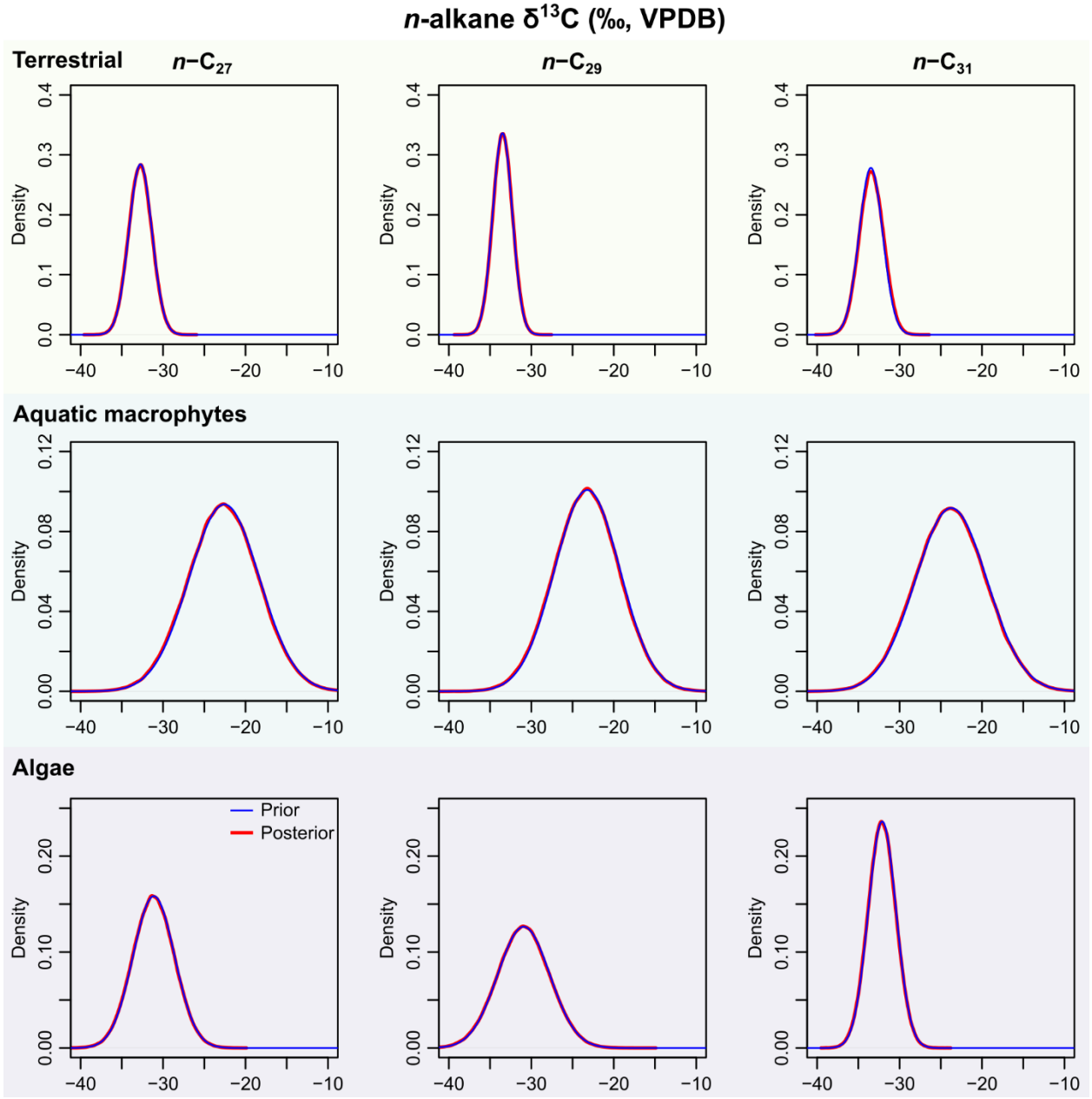

Figure S5. Comparison between the *n*-alkane $\delta^{13}C$ prior distribution of case study 1 and the posterior distribution based on the QHS-5S sample.

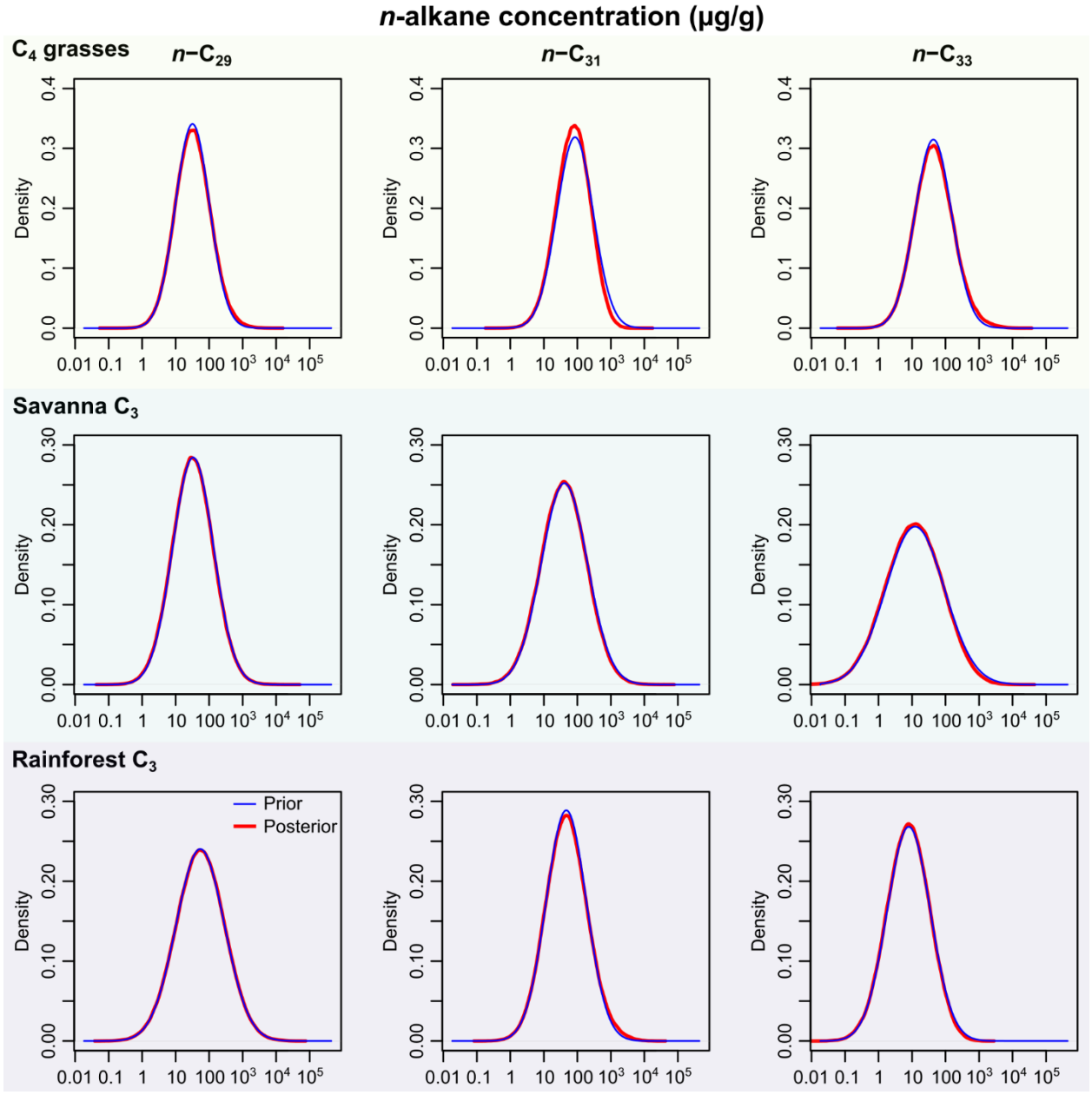

Figure S6. Comparison between the *n*-alkane concentration prior distribution of case study 2 and the posterior distribution based on the Asso sample.

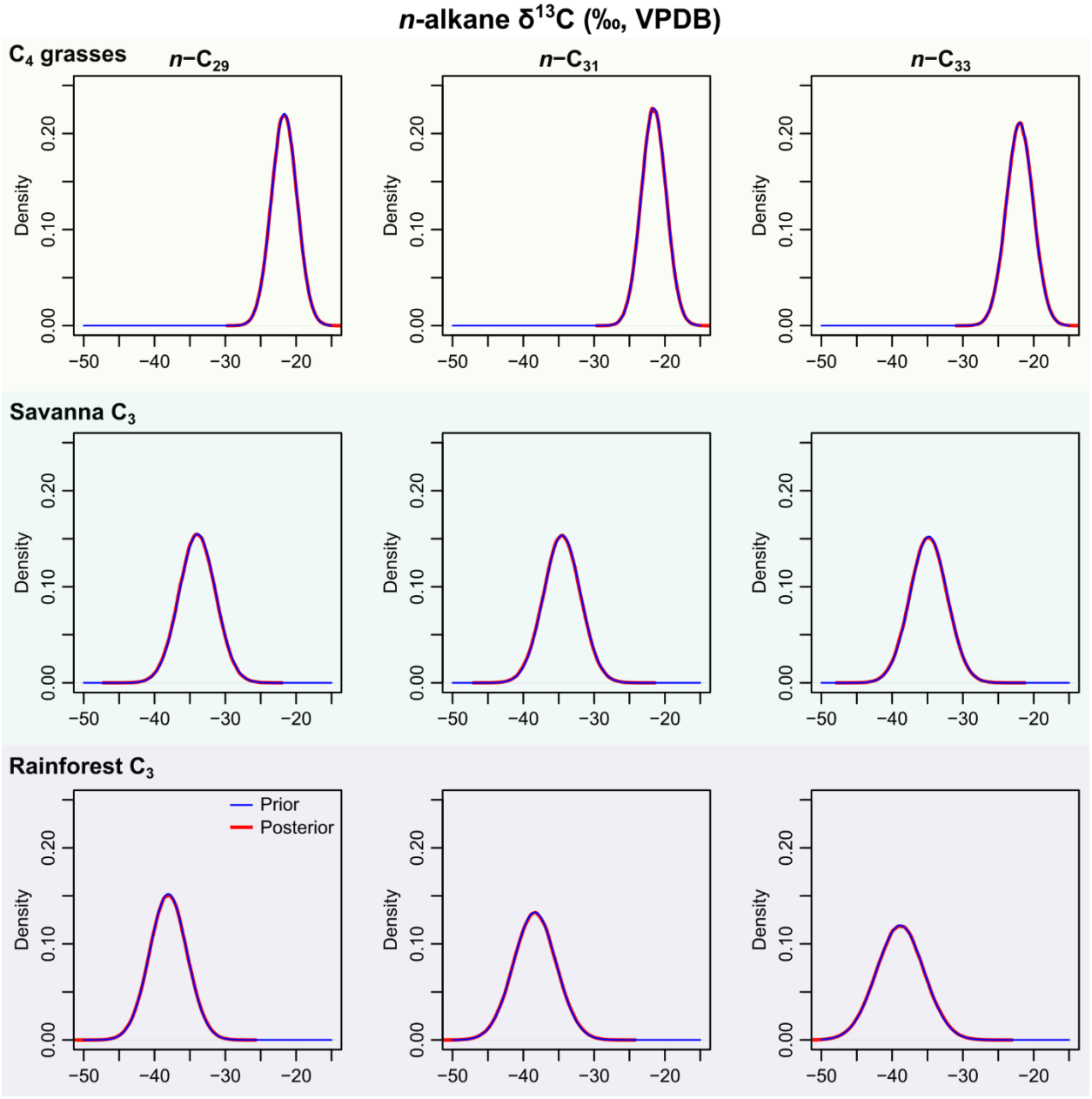

Figure S7. Comparison between the *n*-alkane $\delta^{13}C$ prior distribution of case study 2 and the posterior distribution based on the Asso sample.

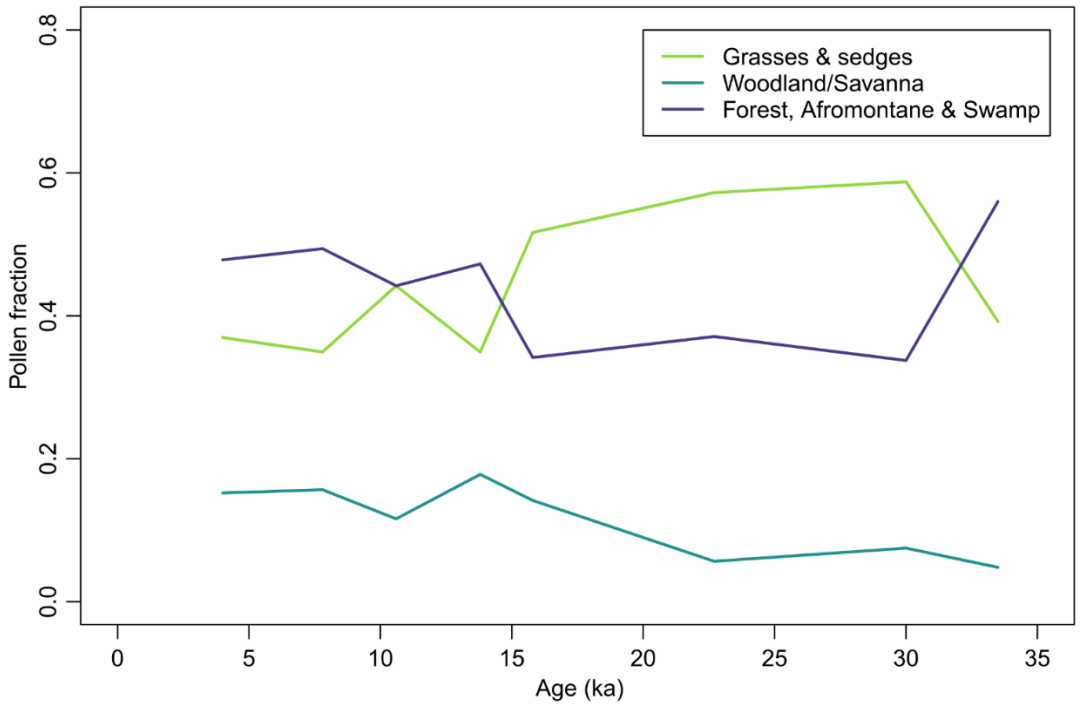

Figure S8. Pollen fractions of three main groups of biomes from the GeoB9311-1 marine sediment core (Dupont and Kuhlmann, 2017), which is ~ 3 degrees south of the GIK16160-3 core (Wang et al., 2013); the rationale for the grouping is explained in supplementary text, Sect. 4; results of correlation tests between the *n*-alkane based vegetation reconstruction and pollen based vegetation reconstruction are reported in Table S6.