Journal of Biogeography

SUPPLEMENTARY INFORMATION

Holocene vegetation dynamics of the Eastern Mediterranean region: old controversies addressed by a new analysis

Esmeralda Cruz-Silva[1], Sandy P. Harrison[1], I. Colin Prentice[2], Elena Marinova[3]

1: School of Archaeology, Geography & Environmental Science, Reading University, Whiteknights, Reading, RG6 6AH, UK
2: Georgina Mace Centre for the Living Planet, Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot SL5 7PY, UK
3: Laboratory for Archaeobotany, Baden-Württemberg Cultural Heritage State Office, Geienhofen-Hemmenhofen, Germany

This Supplementary contains:

Supplementary appendix 1: Summary of the reconstruction method

Supplementary Table 1. Quantitative comparison of predicted and observed dominant and (sub-dominant) biomes in modern (<150 years) samples from long pollen records of the eastern Mediterranean-Black Sea Caspian corridor region. The observed dominant and sub-dominant biomes are taken from the Hengl et al. (2018) potential natural vegetation map. The biome codes are: CENF, cold evergreen needleleaf forest; CMIX, cool mixed evergreen needleleaf and deciduous broadleaf forest; DESE, desert; ENWD, evergreen needleleaf woodland; GRAM, graminoids with forbs; TEDE, temperate deciduous malacophyll broadleaf forest; TUND, tundra; WTFS, warm-temperate evergreen needleleaf and sclerophyll broadleaf forest; XSHB, xeric shrubland. The total number of predicted and observed records for each biome are also shown (Σ). The grey diagonal shows the number of correctly predicted samples, while off-diagonal elements are those incorrectly predicted.

Supplementary Figure 1. Plot showing the resolution and length of the records from the Eastern Mediterranean-Black Sea Caspian corridor (EMBSeCBIO) region.

Supplementary Figure 2: Plots showing reconstructed biomes at 0.5 ky intervals from 12 ky to the present. each plot shows the dominant biome within the 300-year window around each time point. The biome codes are: CMIX, cool mixed evergreen needleleaf and deciduous broadleaf forest; DESE, desert; ENWD, evergreen needleleaf woodland; GRAM, graminoids with forbs; TEDE, temperate deciduous malacophyll broadleaf forest; WTFS, warm-temperate evergreen needleleaf and sclerophyll broadleaf forest; XSHB, xeric shrubland. Note that cold evergreen needleleaf forest (CENF) and tundra (TUND) were never reconstructed. Sites which had samples that were considered to have no modern analogue are indicated as NONA.

Supplementary Figure 3: Proportion of records in different elevation bands (> 1500m, between 500 and 1500m, < 500m) in the Eastern Mediterranean-Black Sea Caspian corridor (EMBSeCBIO) region having at least one sample identified as non-analogue in a 300-year time-window over the past 12 ky, where the time windows were constructed with 50% overlap. The red line indicates the 5% threshold to separate false positives (values below the threshold).

Supplementary Figure 4: Taxon abundance in non-analogue samples from the interval between 11 and 8 ky. The box plots represent the mean and interquartile range across all the non-analogue samples.

Supplementary Figure 5: Identification of the starting point of Holocene forest growth in sequences spanning at least the last 12.3 ky with at least 13 samples in the interval. The curves were produced considering the similarity score ratio of biomes defined as moisture-demanding forest types to other vegetation types for each pollen sample. The forest types are temperate deciduous malacophyll broadleaf forest (TEDE), cool mixed evergreen needleleaf and deciduous broadleaf forest (CMIX) and cold evergreen needleleaf forest (CENF). The curves were smoothed using a 300-year window with 50% overlap. A break point analysis was applied to the forest curve to obtain the optimal number (least amount of residuals) of break points that allowed identifying trend changes in the sequence (score ~ age). Linear regressions were performed between the identified break points in each sequence (formula: y ~ x + {x - breaking} * ifelse(x > breaking,1,0)). The earliest point indicating a change from a zero or negative slope (no change or decrease of forest) to a positive slope (an increase of forest) was cidnetified as the starting point of forest growth. The period evaluated was from 7 to 11.6 ky before present. In the plot, the colours represent the proportion of forest vegetation (blue shading) versus other vegetation (orange shading). The red lines indicate the breakpoint identified as the start of forest growth. The white lines represent the linear regressions between adjacent breakpoints.

Supplementary Figure 6: Proportion of records in the eastern Mediterranean-Black Sea Caspian corridor (EMBSeCBIO) region identified as temperate deciduous malacophyll broadleaf forest (TEDE) in successive 300-year windows during the Holocene.

Supplementary Figure 7: Holocene dynamics of forest vegetation during the Holocene. The uppermost panel shows the proportion of records in the region characterised by moisture-demanding forest, as represented temperate deciduous malacophyll broadleaf forest (TEDE) and cool mixed evergreen needleleaf and deciduous broadleaf forest (CMIX), in 300-year windows with 50% overlap. The lowermost panel shows warm-temperate evergreen needleleaf and sclerophyll broadleaf forest (WTSFS), represented in the same way. The central panel represents more open vegetation types, including samples reconstructed as tundra (TUND), desert (DESE), graminoids with forbs (GRAM), xeric shrubland (XSHB), evergreen needleleaved woodland (ENWD).

Supplementary Figure 8. Forest curve at subregional scale (green line) alongside the count of utilized sites within each time-window (represented by the grey bars in the background).

Supplementary appendix 1: The vegetation reconstruction method

Cruz-Silva et al. (2022) used 5765 modern pollen samples from Europe, the Middle East and northern Eurasia derived from the EMSeCBIO and SPECIAL databases (Harrison, 2019; Harrison et al., 2021), which were taxonomically harmonised (Cruz-Silva et al., 2022) and assigned to biomes based on potential natural vegetation reconstructions (Hengl et al., 2018) to characterise the within-biome means and standard deviations of the abundances of individual taxa. These values were then used to calculate first a dissimilarity index and subsequently a similarity index between any pollen sample and every biome and thus assign each pollen sample to the biome which it is most likely to represent:

$$D_{ik} = \sqrt{\Sigma_{jk} \frac{(p_{jk} - \mu_{ji})^2}{(s_{ji} + \varepsilon)^2}}, \quad (1)$$

where D_{ik} is the dissimilarity of pollen sample k from biome i; p_{jk} is the pollen percentage of taxon j in sample k; μ_{ji} is the mean of taxon j in biome i, s_{ji} is the sample standard deviation of taxon j in biome i and ε is a parameter assigned a value of 0.5%. Summation is over all taxa in sample k.

The dissimilarity values were converted to similarities by:

$$S_{ik} = e^{-D_{ik}/100}, \quad (2)$$

where S_{ik} is an approximation of the likelihood of biome i given pollen sample k. A sample was then assigned to the biome with which it had the greatest similarity.

Biome reconstructions were evaluated using confusion matrices based on dominant and subdominant biomes in the sample search window. The balanced accuracy metric was used to assess performance, especially for imbalanced classes. The balanced accuracy metric is given by:

$$\mathrm{BA} = \frac{1}{l} \sum_{i=1}^{l} \frac{k_i}{n_i}, \quad (3)$$

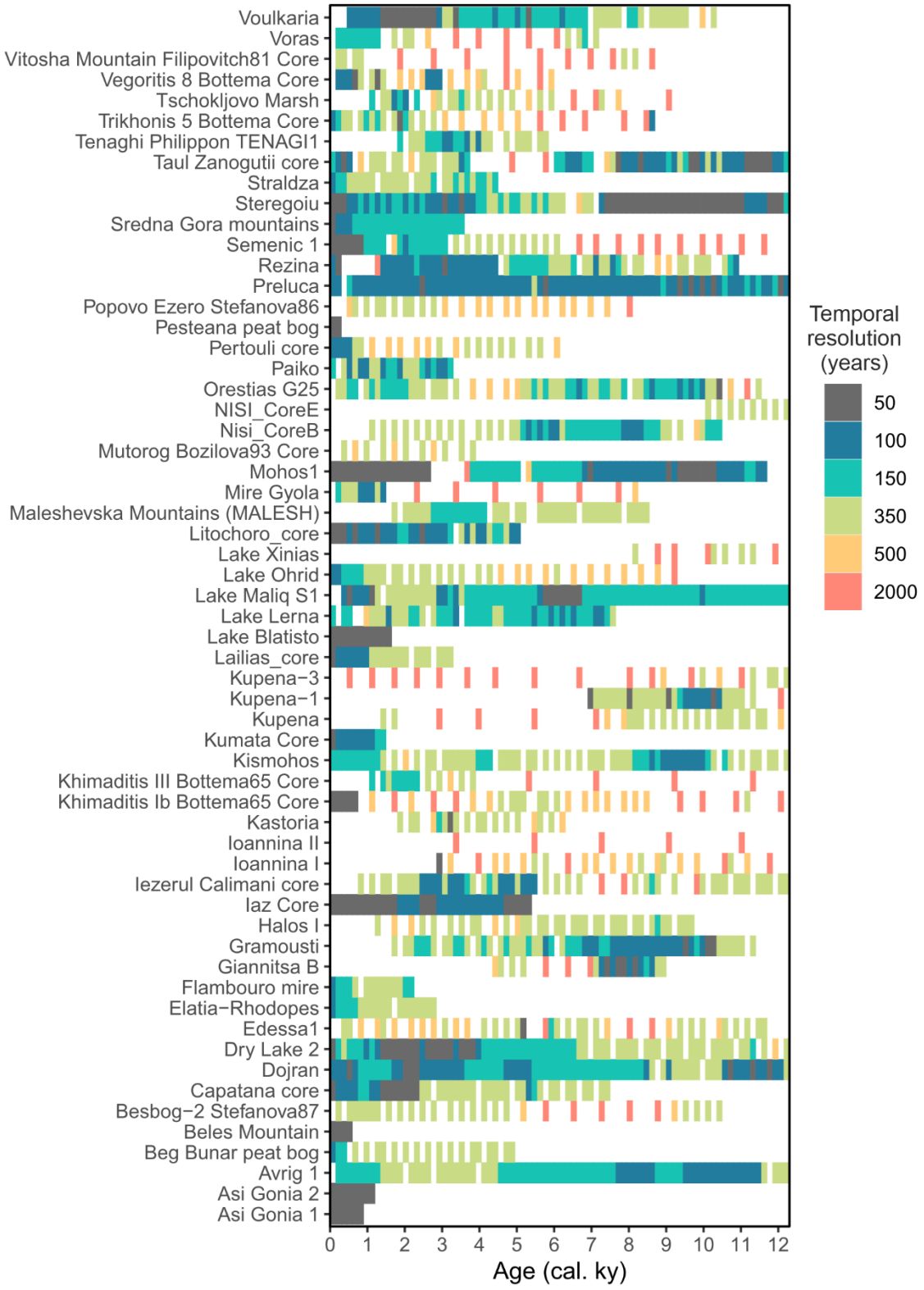
Where where k_i is the number of sites correctly predicted for biome i, l is the number of biomes and n_i is the number of sites in biome i.

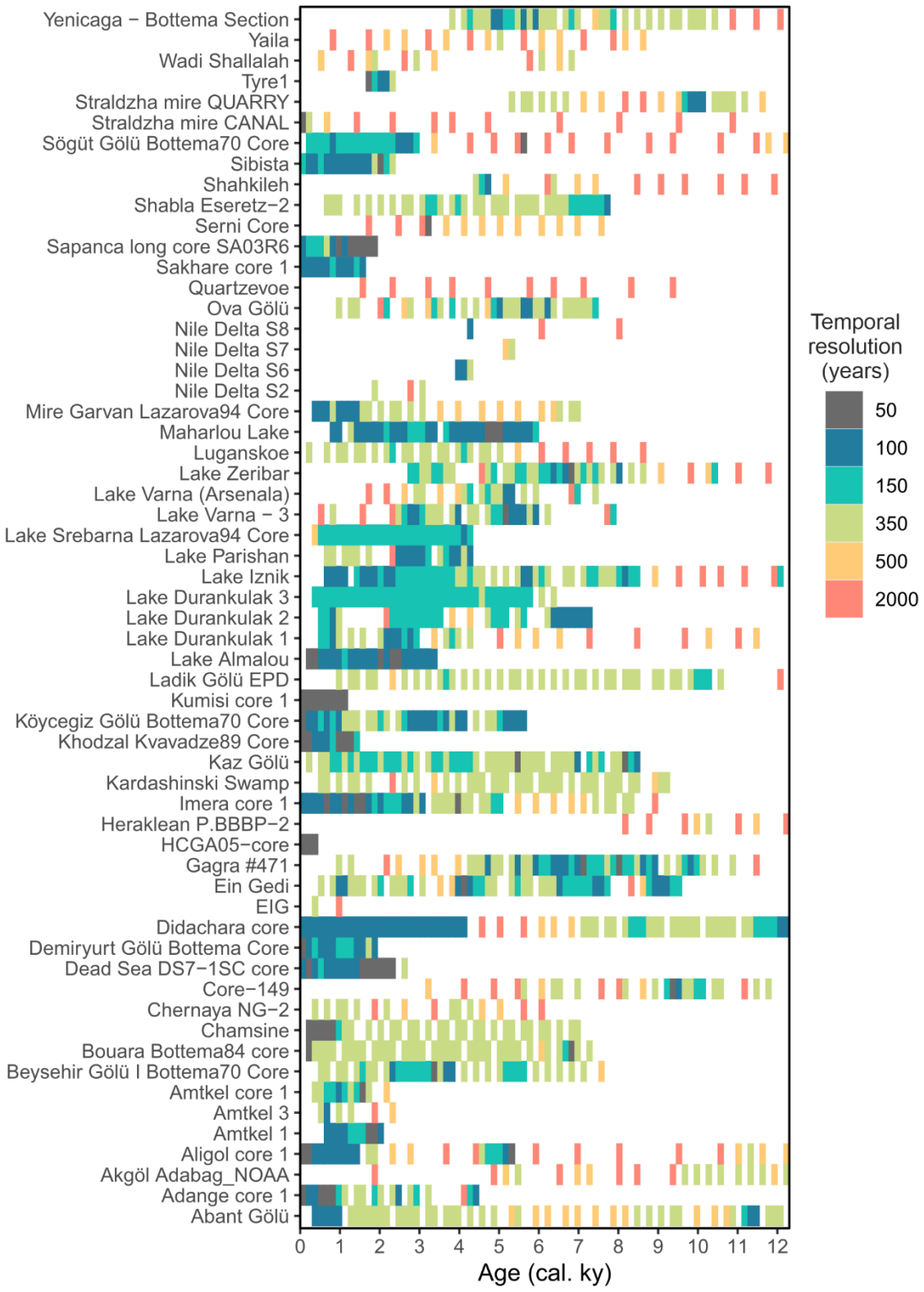
Cruz-Silva et al (2022) showed that reconstructions of the modern vegetation using a modern pollen dataset covering Europe, the Middle East and northern Eurasia used to derive the training data reached a balanced accuracy (i.e. the average accuracy of assignments to all classes: Carrillo et al., 2014) of 66%, which improved to 76% when both dominant and subdominant biomes within a search window of 21 km around the sample were considered. They also showed that the method captured the modern vegetation in the EMBSeCBIO region only with a balanced accuracy of 64%, which improved to 77% when both the dominant and subdominant biomes were considered. The present study includes a further evaluation of the method to determine how accurate it is in predicting the vegetation represented by core top samples (age < 150 cal. yrs. BP) from fossil pollen cores from the EMBSeCBIO region.

Table 1. Quantitative comparison of predicted and observed dominant and (sub-dominant) biomes in modern (<150 years) samples from long pollen records of the eastern Mediterranean-Black Sea Caspian corridor region. The observed dominant and sub-dominant biomes are taken from the Hengl et al. (2018) potential natural vegetation map. The biome codes are: CENF, cold evergreen needleleaf forest; CMIX, cool mixed evergreen needleleaf and deciduous broadleaf forest; DESE, desert; ENWD, evergreen needleleaf woodland; GRAM, graminoids with forbs; TEDE, temperate deciduous malacophyll broadleaf forest; TUND, tundra; WTFS, warm-temperate evergreen needleleaf and sclerophyll broadleaf forest; XSHB, xeric shrubland. The total number of predicted and observed records for each biome are also shown (Σ). The grey diagonal shows the number of correctly predicted samples, while off-diagonal elements are those incorrectly predicted.

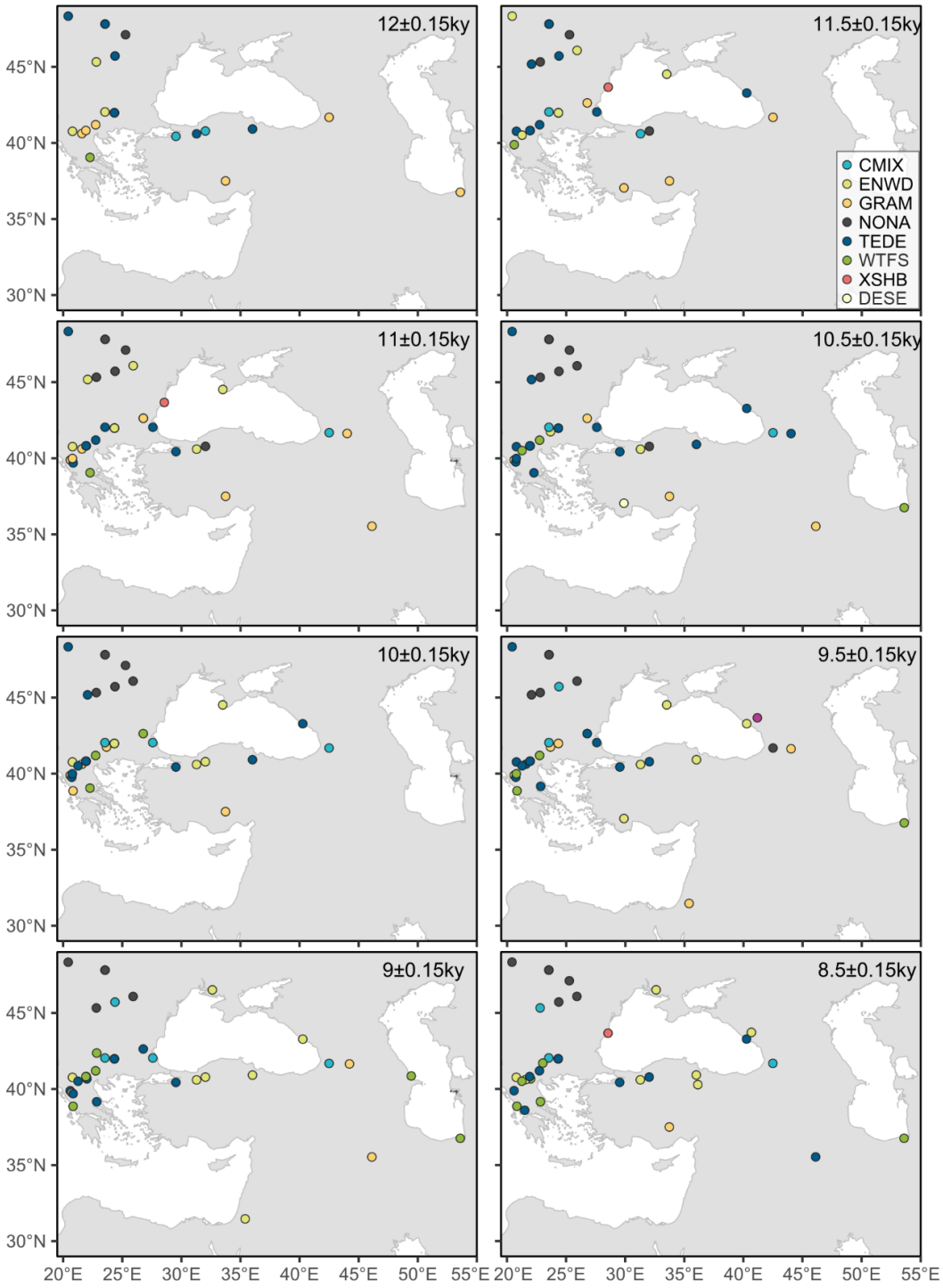
		Predicted									
		DESE	XSHB	WTSFS	GRAM	ENWD	TEDE	CMIX	CENF	TUND	Σ
Observed	DESE	0	0	0	0	0	0	0	0	0	0
	XSHB	0	0	0	0	0	0	0	0	0	0
	WTSFS	0	0	2[1]	1	1	4	0	0	0	9
	GRAM	0	0	0	4[1]	0	1	0	0	0	6
	ENWD	0	0	0	0	7[2]	0	0	0	0	9
	TEDE	0	0	1	0	6	13[3]	1	0	0	24
	CMIX	0	0	0	0	1	4	4[6]	0	0	15
	CENF	0	0	0	0	0	0	0	0	0	0
	TUND	0	0	0	0	0	0	0	0	0	0
	Σ	0	0	4	6	17	25	11	0	0	63

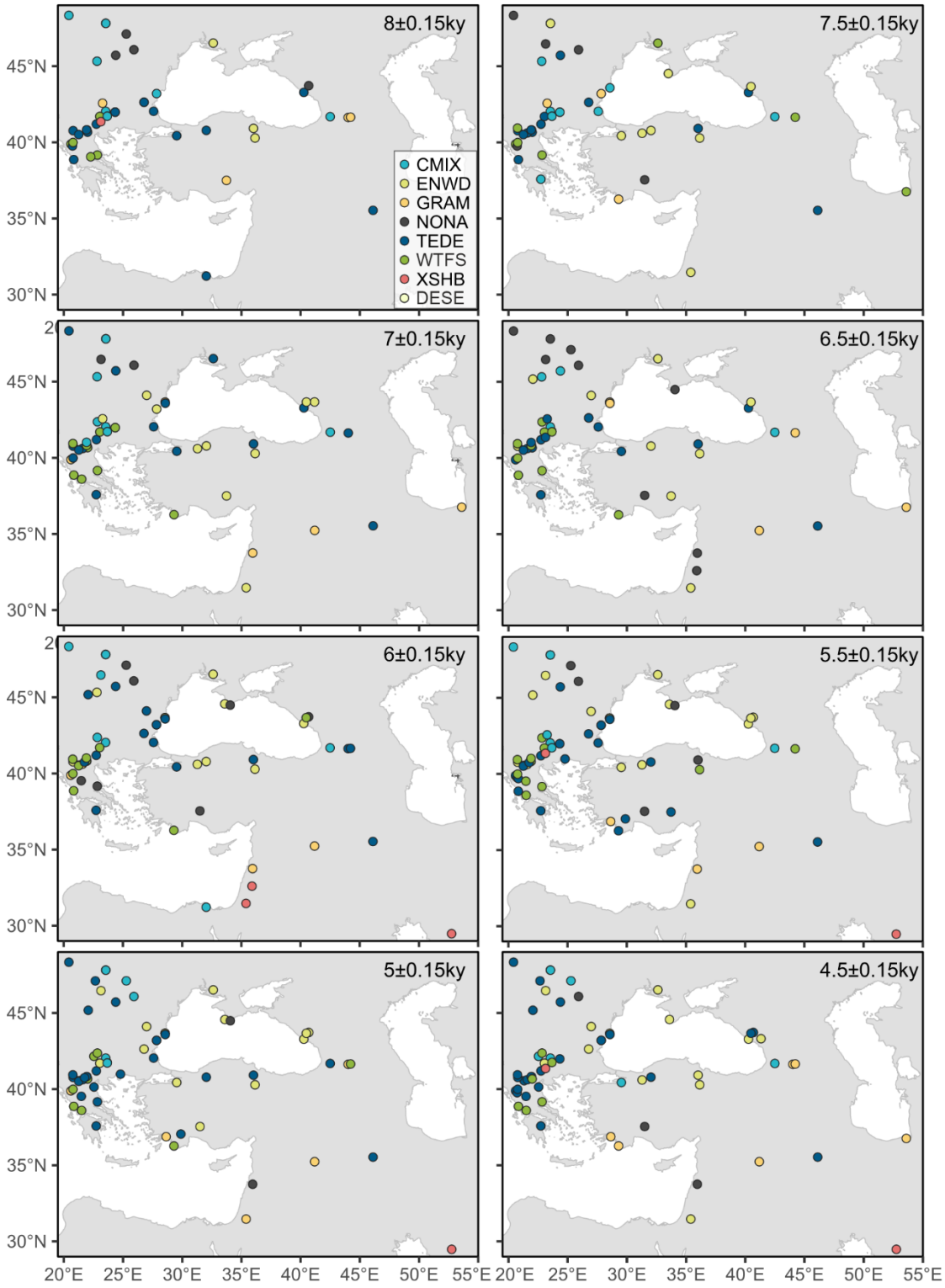
Supplementary Figure 1. Plot showing the resolution and length of the records from the Eastern Mediterranean-Black Sea Caspian corridor (EMBSeCBIO) region.

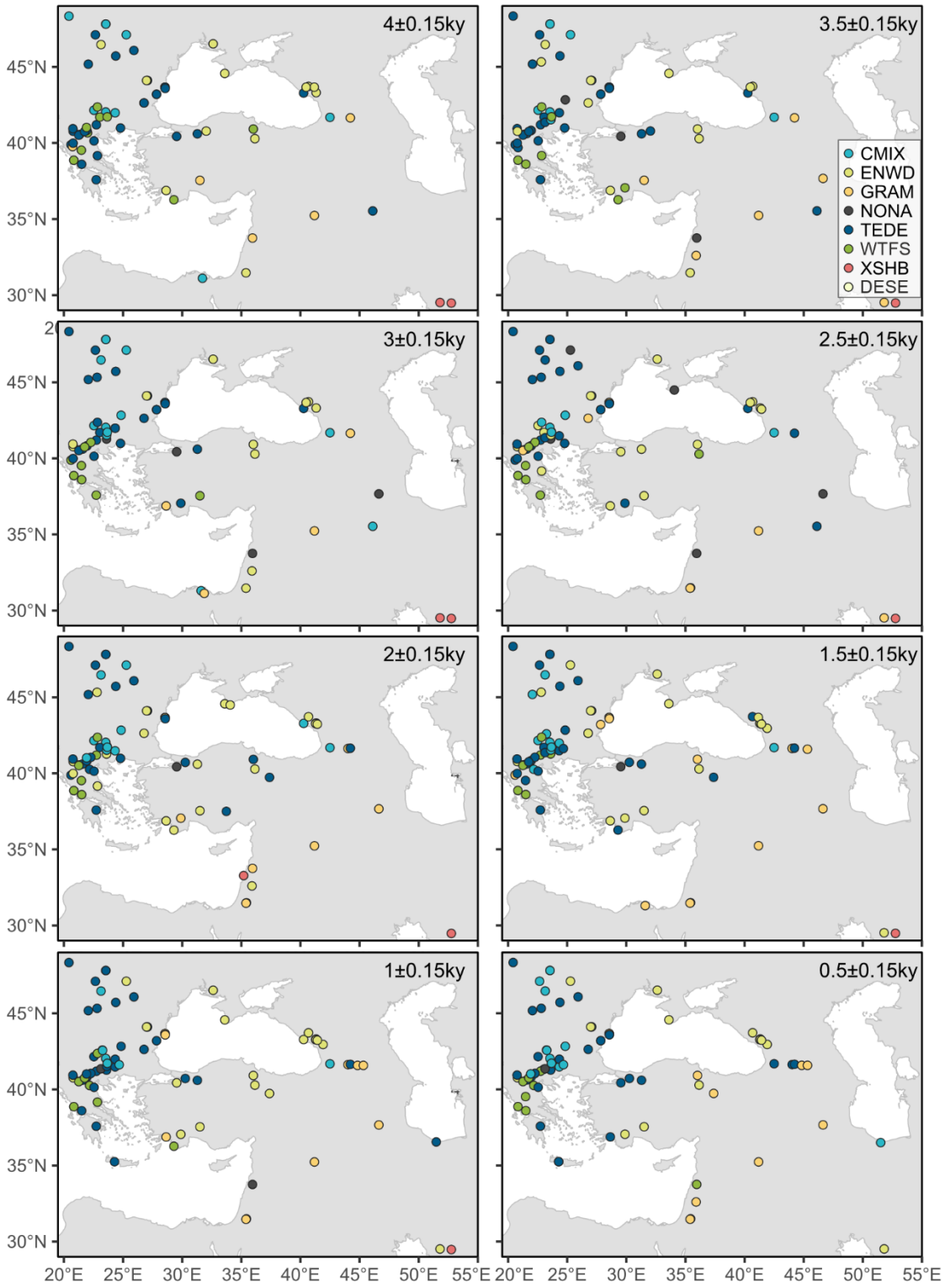



Yenicaga − Bottema Section
Yaila
Wadi Shallalah
Tyre1
Straldzha mire QUARRY
Straldzha mire CANAL
Sögüt Gölü Bottema70 Core
Sibista
Shahkileh
Shabla Eseretz−2
Serni Core
Sapanca long core SA03R6
Sakhare core 1
Quartzevoe
Ova Gölü
Nile Delta S8
Nile Delta S7
Nile Delta S6
Nile Delta S2
Mire Garvan Lazarova94 Core
Maharlou Lake
Luganskoe
Lake Zeribar
Lake Varna (Arsenala)
Lake Varna − 3
Lake Srebarna Lazarova94 Core
Lake Parishan
Lake Iznik
Lake Durankulak 3
Lake Durankulak 2
Lake Durankulak 1
Lake Almalou
Ladik Gölü EPD
Kumisi core 1
Köycegiz Gölü Bottema70 Core
Khodzal Kvavadze89 Core
Kaz Gölü
Kardashinski Swamp
Imera core 1
Heraklean P.BBBP−2
HCGA05−core
Gagra #471
Ein Gedi
EIG
Didachara core
Demiryurt Gölü Bottema Core
Dead Sea DS7−1SC core
Core−149
Chernaya NG−2
Chamsine
Bouara Bottema84 core
Beysehir Gölü I Bottema70 Core
Amtkel core 1
Amtkel 3
Amtkel 1
Aligol core 1
Akgöl Adabag_NOAA
Adange core 1
Abant Gölü
0
1
2
3
4
5
6
7
8
9
10
11
12
Age (cal. ky)
Temporal resolution (years)
50
100
150
350
500
2000

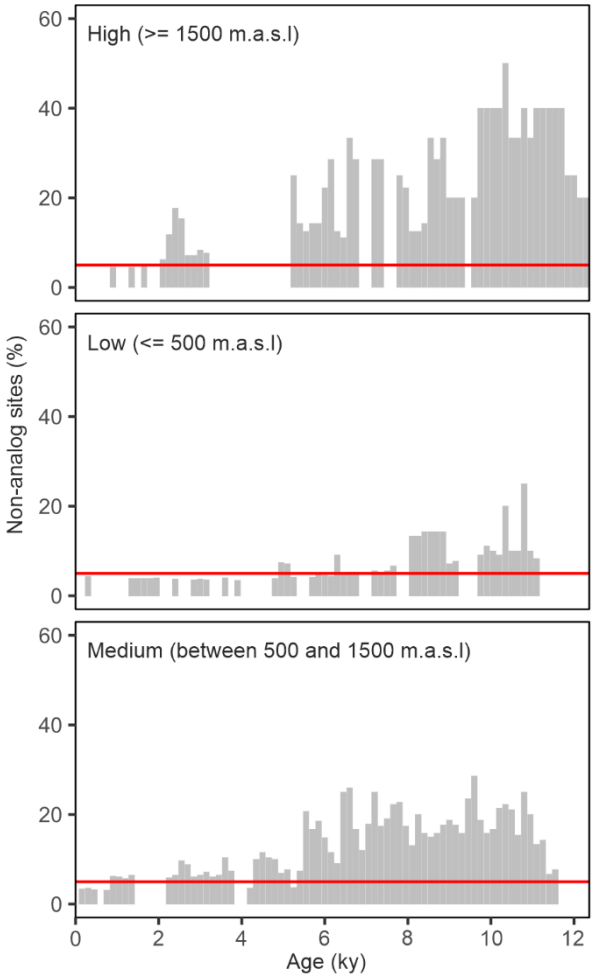
Supplementary Figure 2: Plots showing reconstructed biomes at 0.5 ky intervals from 12 ky to the present. each plot shows the dominant biome within the 300-year window around each time point. The biome codes are: CMIX, cool mixed evergreen needleleaf and deciduous broadleaf forest; DESE, desert; ENWD, evergreen needleleaf woodland; GRAM, graminoids with forbs; TEDE, temperate deciduous malacophyll broadleaf forest; WTFS, warm-temperate evergreen needleleaf and sclerophyll broadleaf forest; XSHB, xeric shrubland. Note that cold evergreen needleleaf forest (CENF) and tundra (TUND) do not occur in these time intervals, although they were reconstructed for other samples. Sites which had samples that were considered to have no modern analogue are indicated as NONA.



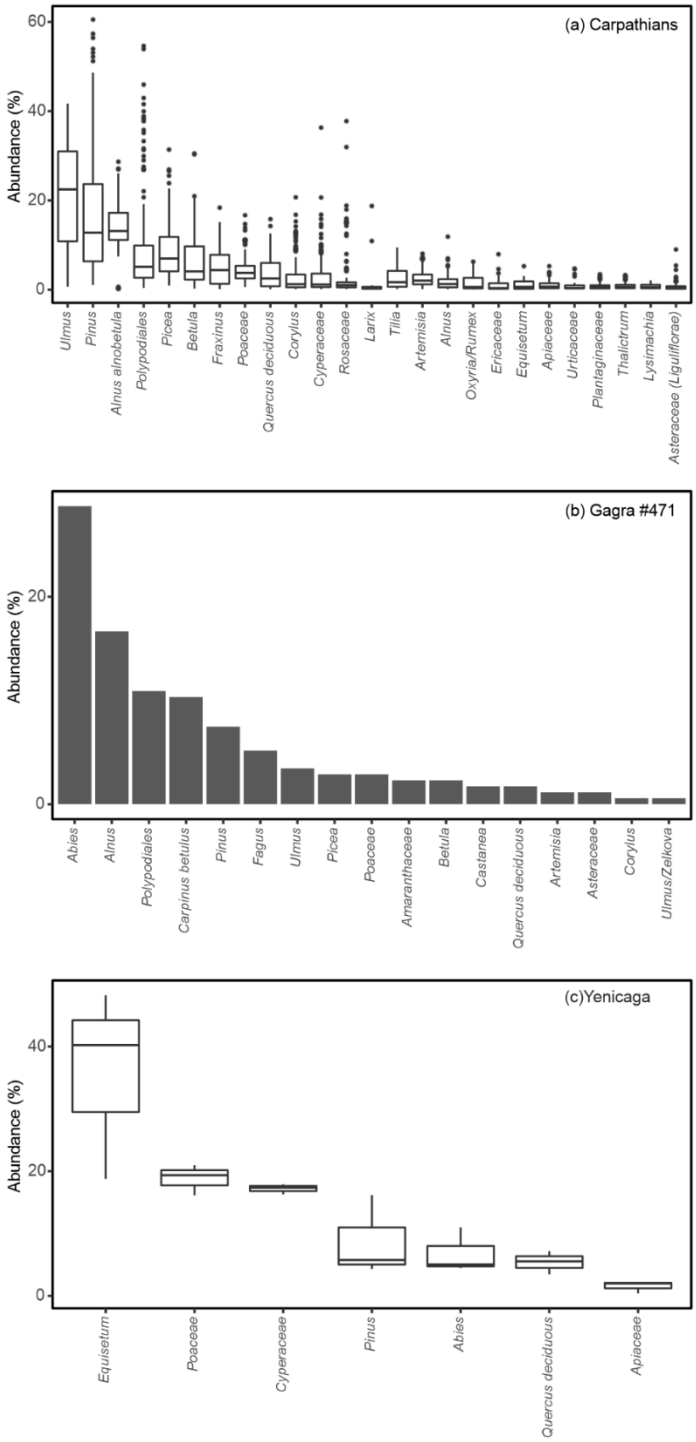

8±0.15ky
7.5±0.15ky
7±0.15ky
6.5±0.15ky
6±0.15ky
5.5±0.15ky
5±0.15ky
4.5±0.15ky
CMIX
ENWD
GRAM
NONA
TEDE
WTFS
XSHB
DESE
45°N
40°N
35°N
30°N
20°E
25°E
30°E
35°E
40°E
45°E
50°E
55°E


4±0.15ky
3.5±0.15ky
3±0.15ky
2.5±0.15ky
2±0.15ky
1.5±0.15ky
1±0.15ky
0.5±0.15ky
CMIX
ENWD
GRAM
NONA
TEDE
WTFS
XSHB
DESE
45°N
40°N
35°N
30°N
20°E
25°E
30°E
35°E
40°E
45°E
50°E
55°E

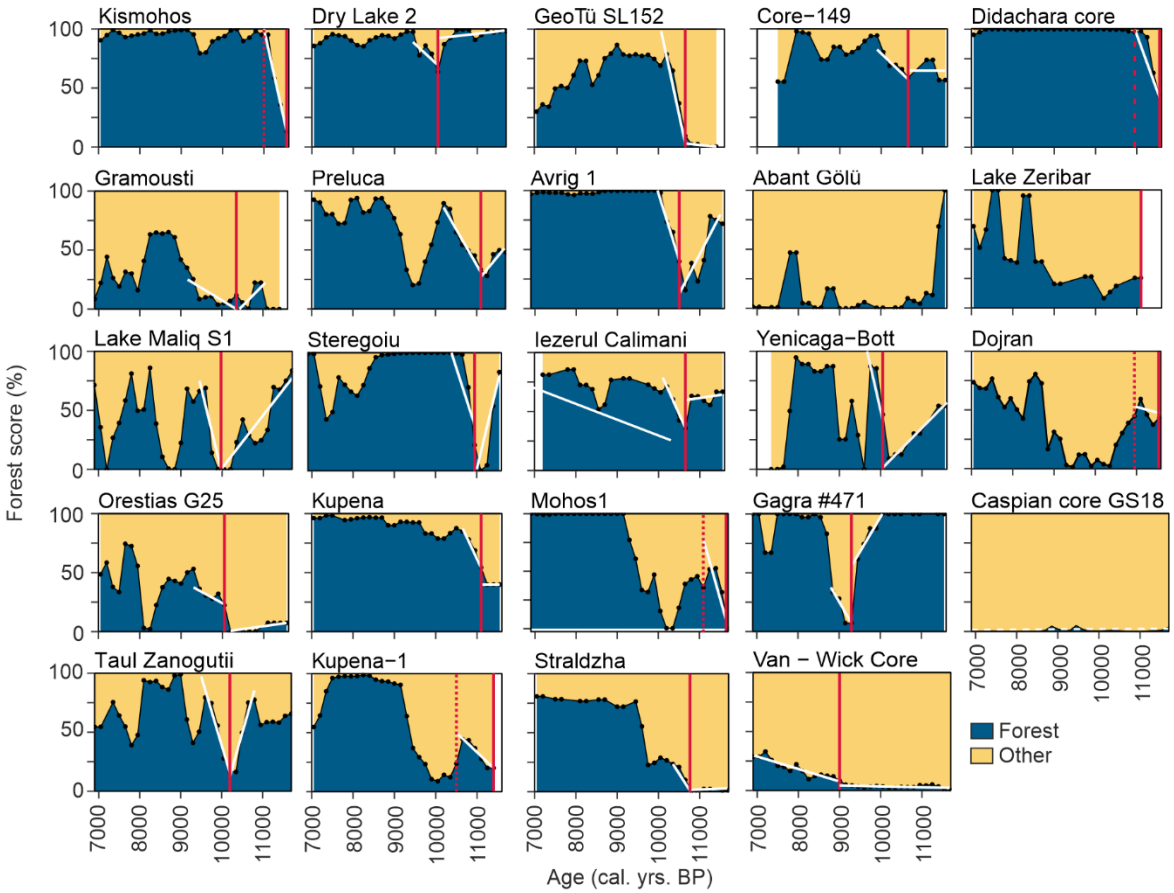
Supplementary Figure 3: Proportion of records in different elevation bands (> 1500m, between 500 and 1500m, < 500m) in the Eastern Mediterranean-Black Sea Caspian corridor (EMBSeCBIO) region having at least one sample identified as non-analogue in a 300-year time-window over the past 12 ky, where the time windows were constructed with 50% overlap. The red line indicates the 5% threshold to separate false positives (values below the threshold).



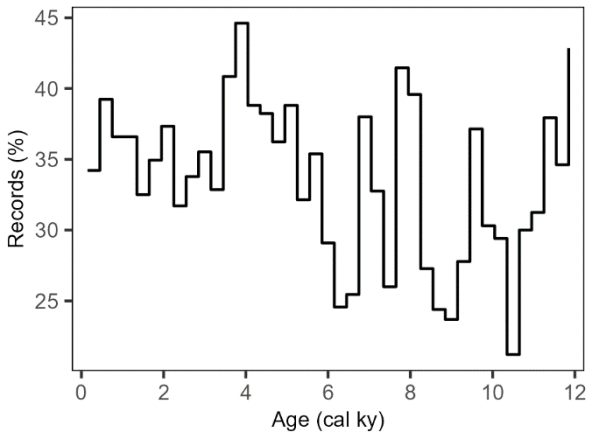
Supplementary Figure 4: Taxon abundance in non-analogue samples from the interval between 11 and 8 ky from (a) the Carpathians, (b) Gagra, and (c) Yenicaga. The box plots represent the mean and interquartile range across all the non-analogue samples.



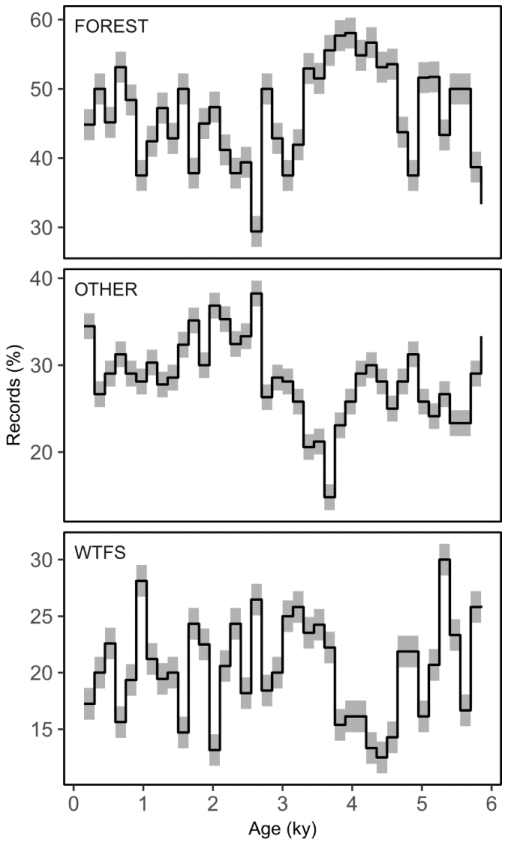
Supplementary Figure 5: Identification of the starting point of Holocene forest growth in sequences spanning at least the last 12.3 ky with at least 13 samples in the interval. The curves were produced considering the similarity score ratio of biomes defined as moisture-demanding forest types to other vegetation types for each pollen sample. The forest types are temperate deciduous malacophyll broadleaf forest (TEDE), cool mixed evergreen needleleaf and deciduous broadleaf forest (CMIX) and cold evergreen needleleaf forest (CENF). The curves were smoothed using a 300-year window with 50% overlap. A break point analysis was applied to the forest curve to obtain the optimal number (least amount of residuals) of break points that allowed identifying trend changes in the sequence (score ~ age). Linear regressions were performed between the identified break points in each sequence (formula: y ~ x + {x - breaking} * ifelse(x > breaking,1,0)). The earliest point indicating a change from a zero or negative slope (no change or decrease of forest) to a positive slope (an increase of forest) was cidnetified as the starting point of forest growth. The period evaluated was from 7 to 11.6 ky before present. In the plot, the colours represent the proportion of forest vegetation (blue shading) versus other vegetation (orange shading). The red lines indicate the breakpoint identified as the start of forest growth. The white lines represent the linear regressions between adjacent breakpoints.



Supplementary Figure 6: Proportion of records the eastern Mediterranean-Black Sea Caspian corridor (EMBSeCBIO) region identified as temperate deciduous malacophyll broadleaf forest (TEDE) in successive 300-year windows during the Holocene.



Supplementary Figure 7: Holocene dynamics of forest vegetation during the Holocene. The uppermost panel shows the proportion of records in the region characterised by moisture-demanding forest, as represented temperate deciduous malacophyll broadleaf forest (TEDE), cool mixed evergreen needleleaf and deciduous broadleaf forest (CMIX) and cold evergreen needleleaf forest (CENF), in 300-year windows with 50% overlap. The lowermost panel shows warm-temperate evergreen needleleaf and sclerophyll broadleaf forest (WTSFS), represented in the same way. The central panel represents more open vegetation types, including samples reconstructed as tundra (TUND), desert (DESE), graminoids with forbs (GRAM), xeric shrubland (XSHB), evergreen needleleaved woodland (ENWD).



Suppelentary Figure 8. Forest curve at subregional scale (green line) alongside the count of utilized sites within each time-window (represented by the grey bars in the background).