Supplementary information

Figure S1–S7

Table S1–S2

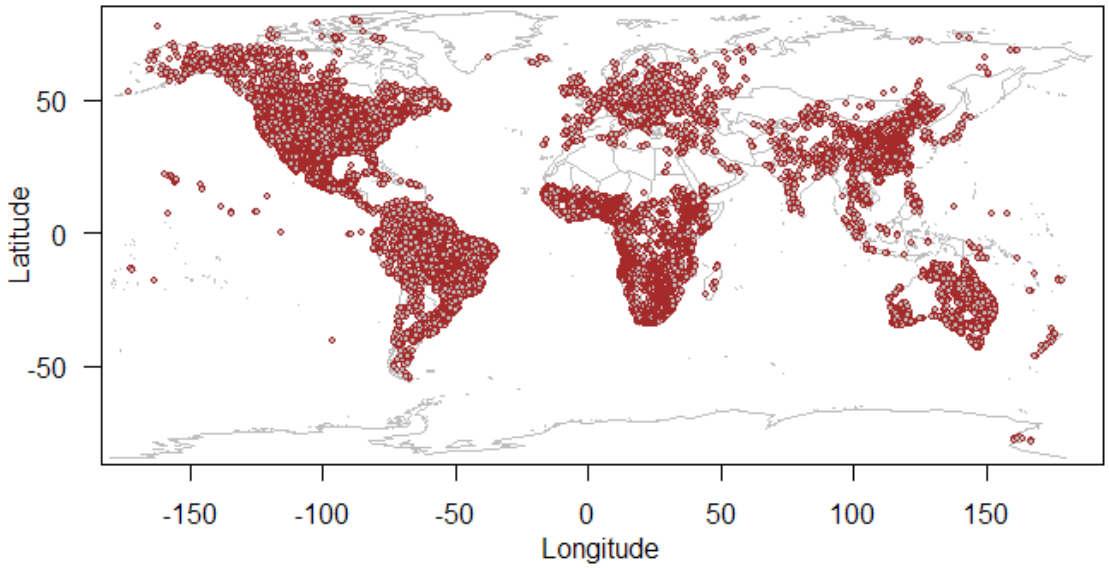

Figure S1. The location of 141,584 soil profiles used in this study. The soil profile database is publicly accessible at http://www.isric.org/explore/wosis/accessing-wosis-derived-datasets.

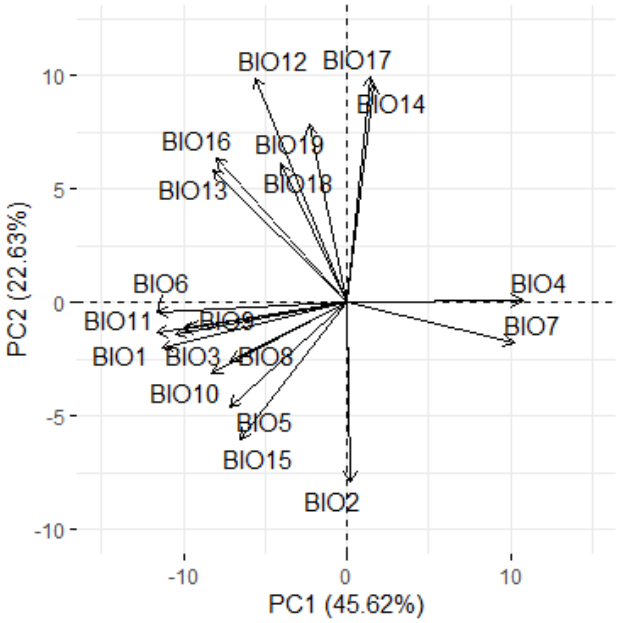

Figure S2. Loadings of 19 bioclimatic variables to the most important two principal components.

components.

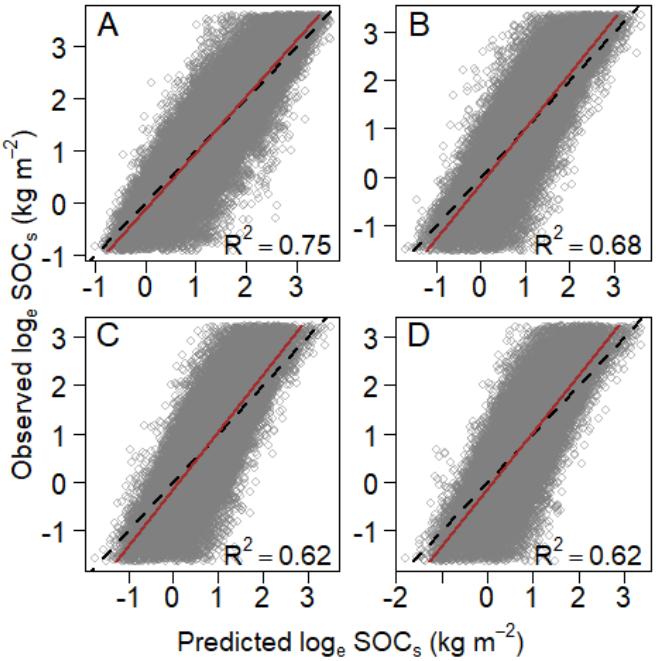

Figure S3. The performance of boosted regression trees in explaining soil organic carbon stock in four standard soil depths across the globe. (A) 0–20 cm, (B) 20–50 cm, (C) 50–100 cm, and (D) 100–200 cm.

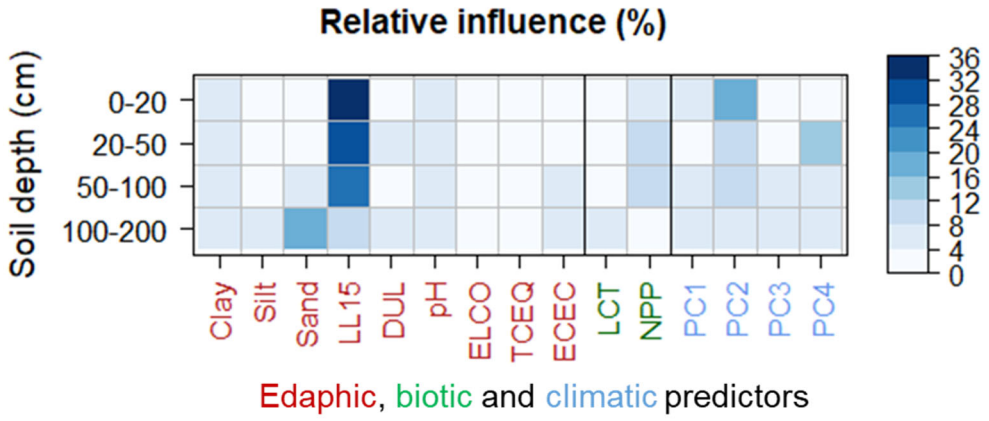

Figure S4. The relative influence of individual biotic, climatic and edaphic variables influencing global soil organic carbon stocks.

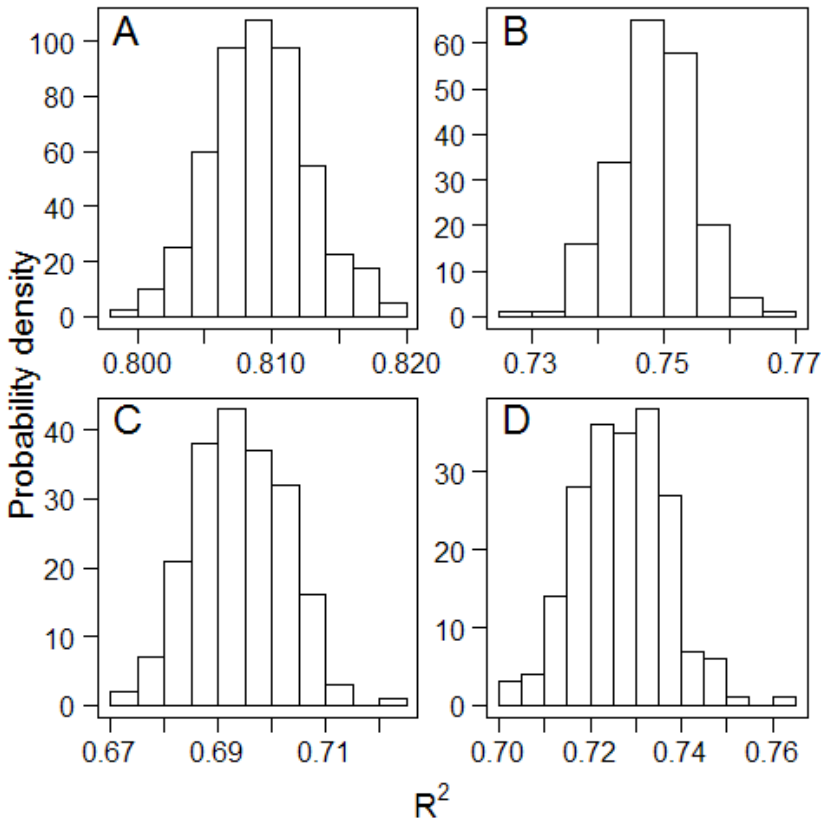

Figure S5. The probability density of model performance in explaining global soil organic carbon stocks at four soil depths. (A) 0–20 cm, (B) 20–50 cm, (C) 50–100 cm, and (D) 100–200 cm. This result is generated by running 200 Monte Carlo simulations taking into account uncertainties in soil bulk density and gravel estimations for calculating soil organic carbon stock.

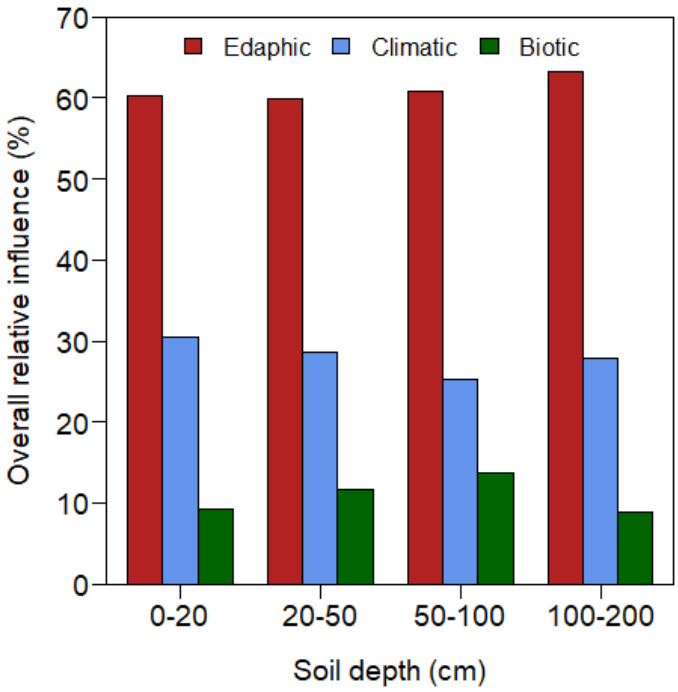

Figure S6. The overall relative influence of edaphic, climatic and biotic variables on soil organic carbon stock in four soil depths across the globe.

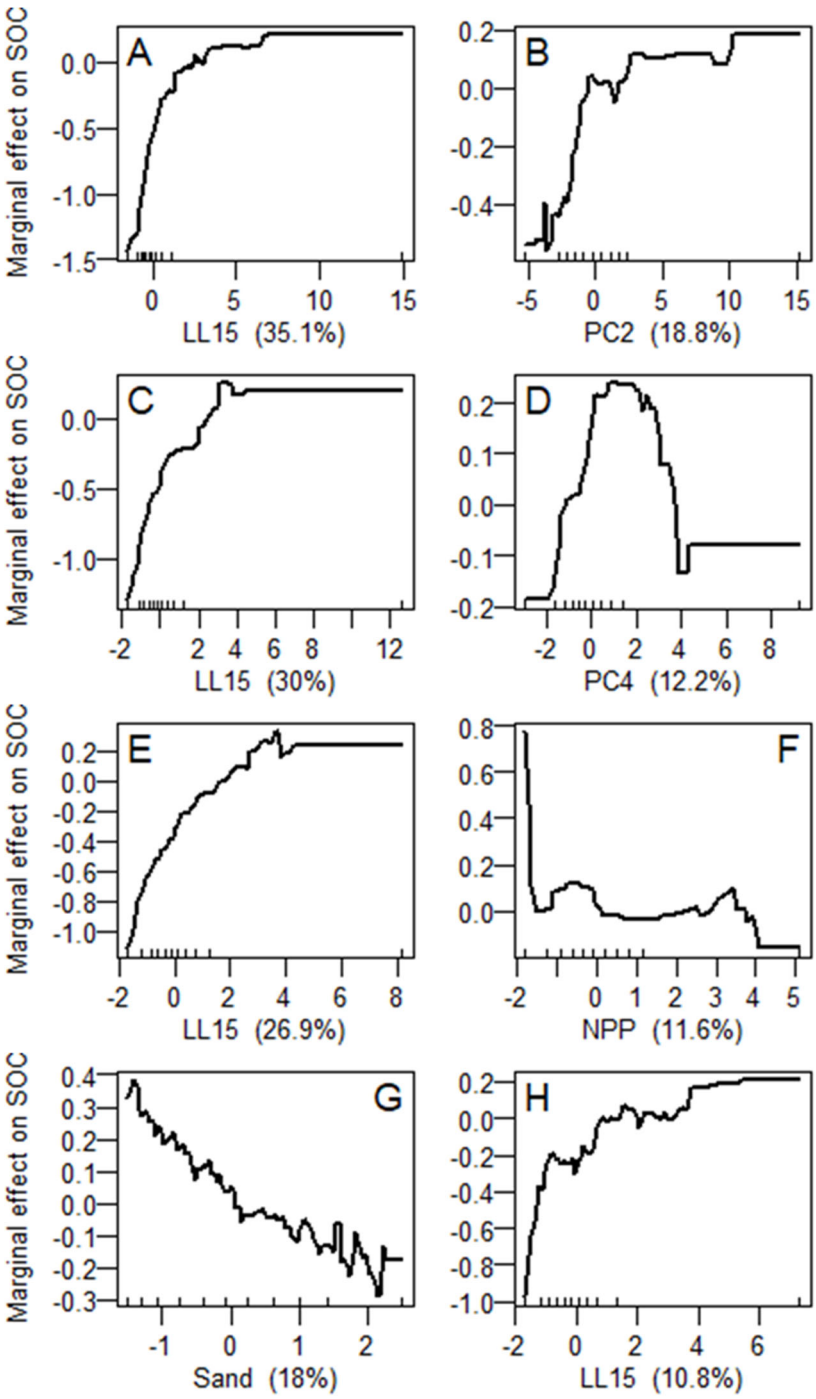

Figure S7. Partial dependence of soil organic carbon (SOC) stock on the two most important controls. Panels from top to bottom show the results for 0–20, 20–50, 50–100, and 100–200 cm depths. Y-axes are centered over the distribution of natural logarithm-transformed SOC stock. Marks on the inside x-axis indicate the distribution of the variable in deciles. All x-axis variables are standardized. Numbers in parenthesis show the relative influence of the variable.

Table S1. Edaphic, climatic and biotic variables used in the modelling of global soil carbon stocks.

Covariates	Code	Description	Unit
Edaphic variables	TCEQ	Calcium carbonate content	g kg^{-1}
	ECEC	Effective cation exchange capacity	cmol kg^{-1}
	ELCO	Electrical conductivity	dS m^{-1}
	Clay	Clay content	%
	Sand	Sand content	%
	Silt	Silt content	%
	pH	pH measured in H_2O	-
	LL15	Lower limit obtained at a matric potential of 1,500 kPa	%
	DUL	Drained upper limit obtained at a matric potential of 33 kPa	%
Climatic variables	T1	Annual mean temperature	°C
	T2	Mean diurnal range	°C
	T3	Isothermality (T2/T7×100)	%
	T4	Temperature seasonality (standard deviation×100)	°C
	T5	Max temperature of warmest month	°C
	T6	Min temperature of coldest month	°C
	T7	Temperature annual range (T5–T6)	°C
	T8	Mean temperature of wettest quarter	°C
	T9	Mean temperature of direst quarter	°C
	T10	Mean temperature of warmest quarter	°C
	T11	Mean temperature of coldest quarter	°C
	P1	Annual precipitation	mm
	P2	Precipitation of wettest month	mm
	P3	Precipitation of driest month	mm
	P4	Precipitation seasonality (coefficient of variation)	%
	P5	Precipitation of wettest quarter	mm
	P6	Precipitation of driest quarter	mm
	P7	Precipitation of warmest quarter	mm
	P8	Precipitation of coldest quarter	mm
Biotic variables	NPP	Net primary productivity	kg C yr^{-1}
	LCT	Land cover type	-

Table S2. Cross-validation statistics for imputation of missing values of soil bulk density (BD) and gravel content (G). RMSE, root mean square error; R^2, coefficient of determination; MAE, mean absolute error.

Soil depth	Statistics	BD	G
0-20 cm	RMSE	0.06	1.47
	R^2	0.93	0.74
	MAE	0.03	0.68
20-50 cm	RMSE	0.06	2.08
	R^2	0.93	0.76
	MAE	0.03	0.95
50-100 cm	RMSE	0.06	3.20
	R^2	0.92	0.72
	MAE	0.03	1.46
100-200 cm	RMSE	0.06	3.44
	R^2	0.92	0.72
	MAE	0.03	1.54