Soil nematode abundance and functional group composition at a global scale

Johan van den Hoogen[1*], Stefan Geisen[1,2], Devin Routh[1], Howard Ferris[3], Walter Traunspurger[4], David Wardle[5], Ron de Goede[6], Byron Adams[7], Wasim Ahmad[8], Walter S. Andriuzzi[9], Richard D. Bardgett[10], Michael Bonkowski[11,12], Raquel Campos-Herrera[13], Juvenil E. Cares[14], Tancredi Caruso[15], Larissa de Brito Caixeta[14], Xiaoyun Chen[16], Sofia dos Santos da Rocha Costa[17], Rachel Creamer[6], José Mauro da Cunha Castro[18], Marie Dam[19], Djibril Djigal[20], Miguel Escuer[21], Bryan Griffiths[22], Carmen Gutiérrez[21], Karin Hohberg[23], Daria Kalinkina[24], Paul Kardol[25], Alan Kergunteuil[26], Gerard Korthals[2], Valentyna Krashevska[27], Alexey Kudrin[28], Qi Li[29], Wenju Liang[29], Matthew Magilton[15], Mariette Marais[30] José Antonio Rodríguez Martín[31], Elizaveta Matveeva[24], El Hassan Mayad[32], Christian Mulder[33], Peter Mullin[34], Roy Neilson[35], Duong Nguyen[11,36], Uffe N Nielsen[37], Hiroaki Okada[38], Juan Emilio Palomares Rius[39], Kaiwen Pan[40,4], Vlada Peneva[41], Loïc Pellissier[42,43], Julio Carlos Pereira da Silva[44], Camille Pitteloud[42], Thomas O. Powers[34], Kirsten Powers[34], Casper Quist[45,46], Sergio Rasmann[47], Sara Sánchez Moreno[48], Stefan Scheu[27,49], Heikki Setälä[50], Anna Sushchuk[24], Alexei Tiunov[51], Jean Trap[52], Wim van der Putten[2,46], Mette Vestergård[53], Cecile Villenave[54,55], Lieven Waeyenberge[56], Diana H.Wall[9], Rutger Wilschut[2], Daniel Wright[57], Jiue-in Yang[58], Thomas Ward Crowther[1*]

These authors contributed equally: Johan van den Hoogen, Stefan Geisen
[*]Email: johan.vandenhoogen@usys.ethz.ch, tom.crowther@usys.ethz.ch

Affiliations
[1]Crowther Lab, Institute of Integrative Biology, ETH Zürich, 8092 Zürich, Switzerland
[2]Department of Terrestrial Ecology, Netherlands Institute of Ecology, 6708 PB Wageningen, The Netherlands

Further affiliations are listed at the bottom of the document.

Summary

Soil organisms are a crucial part of the terrestrial biosphere. Despite their importance for ecosystem functioning, no quantitative, spatially-explicit models of the active belowground community currently exist. In particular, nematodes are the most abundant animals on Earth, filling all trophic levels in the soil food web. Here, we use 6,579 georeferenced samples to generate a mechanistic understanding of the patterns of global soil nematode abundance and functional group composition. The resulting maps show that $4.4 \pm 0.64 \times 10^{20}$ nematodes (total biomass ~0.3 Gt) inhabit surface soils across the world, with higher abundances in sub-arctic regions (38% of total), than in temperate (24%), or tropical regions (21%). Regional variations in these global trends also provide insights into local patterns of soil fertility and functioning. These high-resolution models provide the first steps towards representing soil ecological processes into global biogeochemical models, to predict elemental cycling under current and future climate scenarios.

As we refine our spatial understanding of the terrestrial biosphere, we improve our capacity to manage natural resources effectively. With ever-growing functional information about the biogeography of aboveground organisms, an outstanding gap in our understanding of the biosphere remains the activity and distribution patterns of soil organisms[1,2]. Soil biota, including bacteria, fungi, protists and animals, play central roles in every aspect of global biogeochemistry, influencing the fertility of soils and the exchange of CO_2 and other gasses with the atmosphere[3]. As such, biogeographic information on the abundance and activity of soil biota is essential for climate modelling and, ultimately, environmental decision making[2,4-6]. Yet, the activity of soil organisms is not explicitly reflected in biogeochemical models due to our limited understanding of their biogeographic patterns at the global scale.

In recent years, pioneering studies in soil biogeography have begun to provide valuable insights into the broad-scale taxonomic diversity patterns of soil bacteria[7-11], fungi[11-13] and nematodes[14-17], and patterns of microbial biomass[11,18,19]. However, until now, we have been unable to generate a high-resolution, quantitative understanding of the abundance or functional composition of active soil organisms because of two major reasons. First, due to the methodological challenges in characterizing soil biota, most previous studies have focused on a relatively limited number of spatially distinct sampling sites (<500), and therefore cannot detect high-resolution regional-scale patterns. Second, most global studies have used molecular sequencing approaches, which provide valuable semi-quantitative information on taxonomic diversity, but not information on absolute abundance or biomass that is essential to link biological communities to ecosystem functioning and global biogeochemistry[20,21]. DNA and RNA-based approaches cannot unambiguously differentiate between living (being either active or dormant) and dead cells, so they cannot be used to quantify the active component of the belowground community[22,23]. To generate a robust, global perspective of belowground biota and their roles in biogeochemical cycling, we need a sampling design that provides a thorough global representation of the belowground community, and direct, quantitative abundance data reflecting the active community. Here, we adopt this approach in order to generate a

quantitative understanding of a critical component of the soil food web, for which direct extraction methods enable quantification of active organisms: nematodes.

Nematodes are a dominant component of the soil community and are by far the most abundant animals on Earth[2]. They account for an estimated four-fifths of all animals on land[24], and feature in all major trophic levels in the soil food web. The functional role of nematodes in soils can be inferred by their trophic position, and hence nematodes are often classified into trophic groups based on feeding guilds (i.e. bacterivores, fungivores, herbivores, omnivores, predators). Given their pivotal roles in processing organic nutrients and control of soil microorganism populations[25-27], they play critical roles in regulating carbon and nutrient dynamics within and across landscapes[26] and are a good indicator of biological activity in soils[28]. Yet, we still lack even a basic understanding of broad-scale biogeographic patterns in nematode abundance and nematode functional group composition. Despite expectations that nematode abundances may peak in warm tropical regions with high plant biomass[14,15], other studies suggest that the opposite pattern might exist, with high nematode abundances in high-latitude regions with larger standing soil carbon stocks[16,17,29-31]. Disentangling the effects of these different environmental drivers of soil nematode communities is critical to generate a mechanistic understanding of the global patterns of soil nematodes, and for quantifying their influence on global biogeochemical cycling.

Here, we use 6,759 spatially distinct soil samples from all terrestrial biomes and continents to examine the environmental drivers of global nematode communities. By making use of 73 global layers of climate, soil, and vegetation characteristics, we then extrapolate these relationships across the globe to generate the first spatially-explicit, quantitative maps of soil nematode density and functional group composition at a global scale.

Results and Discussion

Biome-level patterns of soil nematodes

By compiling soil sampling data from all major biomes and continents we aimed to generate a representative dataset to capture the variation in global nematode densities. Within each sample, we quantified the total abundance of each trophic group using microscopy. In order to standardize sampling protocols, we focus on the top 15 cm of soil, which is the most biologically active zone of soils[6,32]. In line with previous reports[33], nematode abundances are highly variable within and across terrestrial biomes, ranging from dozen to thousands of individuals per 100 g soil (Fig. 1b). This variation highlights the necessity for large datasets in soil biodiversity analyses to reliably predict large-scale patterns, as the accuracy of our mean estimates for any region improves considerably with increasing number of samples (Fig. 2a). Specifically, the confidence in our mean estimates for nematode abundance in any region is relatively low at the individual sample scale, but high only when calculated with larger (i.e. >400) sample size.

Overall, we observed the highest nematode densities in tundra (median = 2,329 nematodes per 100 g dry soil), boreal forests (median = 2,159) and in temperate broadleaf forests (median = 2,136), while the lowest densities are observed in Mediterranean forests (median = 425), Antarctic sites (median = 96) and hot deserts (median = 81) (Fig. 1b, Supplementary Table 2). To examine the mechanisms driving the patterns of soil nematode density and functional group composition across biomes, we integrated the nematode abundance data with 73 global datasets of soil physical and chemical properties, and vegetative, climatic, topographic, anthropogenic, and spectral reflectance information (Supplementary Table 3). Antarctic sampling points were excluded from the modelling dataset due to limited coverage of several covariate layers. To match the spatial resolution of our covariates, all samples were aggregated to the 1-km^2 pixel level to generate 1,876 unique pixel locations across the world. We analysed a suite of machine-learning models (including random forest, L1 and L2 regularised linear regression) to determine the environmental drivers of the variation in nematode abundance and functional group composition across the globe. We iteratively varied the set of covariates and model hyperparameters across 405 models and evaluated model strength using *k*-fold cross validation (with $k = 10$). This approach allowed us to select the best performing model which had high predictive strength (mean cross-validation $R^2 = 0.43$, overall $R^2 = 0.86$), whilst taking

into account issues surrounding multicollinearity, and model overparameterization and overfitting. This final model, an iteration of random forests using all 73 covariates, was then used to create a per-pixel mean and standard deviation values. Mapping the extent of extrapolation highlighted that our dataset covered most environmental conditions, with the least represented pixels and highest proportion of extrapolation in the Sahara and Arabian Desert (Extended Data Figs. 1a, 1b). We acknowledge that our models cannot accurately predict nematode abundances at fine spatial scales, as local environmental heterogeneity can cause considerable variation in nematode abundances, even within individual locations. However, the strength of these predictions increases at the larger scales where our modelled estimates are informed by more data observations (Fig. 2b), ensuring confidence in our estimates. Predicted *vs.* observed plots revealed that, despite the high accuracy in most regions, the models tended to marginally over-represent the observed numbers at low densities and underrepresent at higher nematode densities (Figs. 2c-h). Moreover, our cross-validation accuracy calculations may be optimistically biased, as we cannot entirely account for the potential impacts of overfitting. Our analyses would have ideally included a subset of data removed at the beginning of the analyses for fully independent accuracy assessment. However, as the removal of a subset would mean a loss of geographic representation, we chose instead to maintain the integrity of the entire dataset and generate spatially explicit maps of model confidence that allow for error propagation throughout the final global calculations (Fig. 2i, Extended Data Fig. 1a). These maps provide spatial insight into the prediction uncertainties rather than a single accuracy measure for overall model accuracy.

Our statistical models reveal the dominant drivers of nematode abundance across global soils. As with aboveground animals, climatic variables (i.e., temperature and precipitation) played an important role in shaping the patterns in total soil nematode abundance. However, soil characteristics (e.g. texture, soil organic carbon (SOC) content, pH, cation-exchange capacity (CEC)) were by far the most important factors driving nematode abundance at a global scale, with effects that largely overwhelmed the climate impacts (Supplementary Table 3). Linear models enabled us to assess the directionality of these relationships,

144 revealing that both SOC content and CEC had strong positive correlations, whilst pH had a negative effect
145 on total nematode density (Extended Data Fig. 2). These trends support the suggestion that soil resource
146 availability is a dominant factor structuring belowground communities at broad spatial scales, overriding
147 the impact of climate, in structuring belowground communities at broad spatial scales[2,12,15].
148
149 **Global biogeography of soil nematodes**
150 The high predictive strength of the top model enabled us to extend the relationships across global soils to
151 construct high-resolution (30 arc-seconds, ~ 1 km^2), quantitative maps of total nematode densities. These
152 maps reveal striking latitudinal trends in soil nematode abundance, with the highest densities in sub-arctic
153 regions (Fig. 3), a trend that is consistent across all trophic groups (Extended Data Figs. 3a-e). Specifically,
154 as with the regional averages, the highest abundances of soil nematodes are found in boreal forests across
155 North America, Scandinavia and Russia. Whether nematode abundance is expressed as density per gram of
156 soil or per unit area (thereby controlling for the differences in soil bulk density), the models reveal a striking
157 latitudinal gradient in soil nematode abundance (Fig. 3, Extended Data Figs. 4, 5). Whether soil animals
158 exist at highest abundances in the high or low latitudes has been a contentious issue in the soil ecology
159 literature, with some studies highlighting highest abundances in boreal forests, and others suggesting that
160 tropical forests support the greatest abundance[29,31,14]. Our extensive sample data from every biogeographic
161 region allows us to see beyond these contrasting results to reveal a striking latitudinal pattern of nematode
162 abundance, providing conclusive evidence that soil nematodes are present in considerably higher densities
163 in high-latitude arctic and sub-arctic regions (Fig. 3).
164
165 Along with the latitudinal gradient in nematode abundance, our nematode density map also reveals regional
166 contingencies that stand out against the global trends. Although nematode abundances were relatively low
167 in tropical regions, our sampling data and models reveal high nematode abundance in certain tropical
168 peatlands such as the Peruvian Amazon (Fig. 1a; Fig. 3). These regions are characterized by high SOC
169 stocks, which support high microbial biomasses that serve as the basic resource for most nematode groups.

Similarly, increased SOC stocks at high altitude compared to lowland regions drive higher nematode abundances in mountainous regions and highlands, such as the Rocky Mountains, Himalayan Plateau and the Alps (Fig. 1a; Fig. 3). Although the respective climates of these regions exhibit large differences in mean annual temperature (<0˚C to >10˚C), their soils are all characterized by relatively high SOC stocks (i.e. >50 g kg^{-1}). In contrast, the lowest nematode densities were predicted in hot deserts such as the Sahara, Arabian Desert, Gobi Desert, and Kalahari Desert (Fig. 3), regions characterized by very low SOC stocks. As such, the spatial variability in nematode abundance is highest in equatorial regions, which exhibit the full range of possible abundances from desert to biomes characterized by high SOC stocks. This is reflected by the spatial patterns in our model uncertainty, in which low-latitude arid regions with low sampling density and soil nematode abundances are characterized by larger uncertainty (Fig. 2i, Extended Data Fig 1).

The strong correlation between temperature and SOC content at a global scale[19] makes it challenging to identify the primary driver of the latitudinal gradient in nematode abundances. However, regional deviations from the global biogeographic pattern help to disentangle their relative roles, as they decouple the effects of climate and soil characteristics. For example, low temperatures and high moisture content in high-latitude regions restrict annual decomposition rates, leading to the accumulation of soil organic material[19,30]. But the positive effect of SOC in tropical peatland regions (with high soil carbon but also warm temperatures) suggests that it is organic matter content, rather than climate conditions, that ultimately determines nematode abundance in soil. These models reinforce the dominant role of soil characteristics in driving nematode abundances. These trends suggest that the impacts of climate on nematode density are not direct, but instead act indirectly by modifying soil characteristics.

We next examined how nematode community structure varied across landscapes by exploring the abundance of each trophic group across our dataset. At the global scale, all trophic groups were positively correlated with one another (Extended Data Fig. 6a), suggesting that biogeographic regions with high

196 nematode abundances are generally hospitable for members of all trophic groups. Despite the distinct
197 feeding habits, the global consistency across trophic groups provides some unity in the biogeography of the
198 soil food web. That is, although different nematodes rely on distinct food sources for their energetic
199 demands, the size of the entire food web is ultimately determined by the availability of soil organic matter.
200 Nevertheless, the relative composition of nematode communities did vary across samples. To characterize
201 the main nematode community types, we clustered the observed relative abundances into four types, based
202 on the relative abundance of each trophic group (Extended Data Fig. 6b). Although there were no clear
203 spatial patterns in these community types, vector analysis revealed that the indices of vegetation cover (e.g.,
204 NDVI, EVI) were the best predictors of herbivore-dominated communities, while edaphic factors (sand
205 content, pH) were strong predictors of communities dominated by bacterivores (Extended Data Fig. 6c).
206
207 By summing the nematode density information in each pixel, we can begin to generate a quantitative
208 understanding of soil nematode abundances and biomass at a global scale. We estimate that approximately
209 $4.4 \pm 0.64 \times 10^{20}$ nematodes inhabit the upper layer of soils across the globe (Table 1, Supplementary Table
210 5). Of these, 38.7% exist in boreal forests and tundra, 24.5% in temperate regions and 20.5% in tropical
211 and sub-tropical regions (Supplementary Table 6). By combining our estimates of nematode abundance
212 with mean biomass estimates of each functional group (using a database containing 32,728 nematode
213 samples[34,35]), we can approximate that global nematode biomass in the global topsoil is approximately 0.3
214 Gt (Table 1). This translates to approximately 0.03 Gt of carbon (C) (Table 1, Supplementary Table 7),
215 which is three times greater than a previous estimate of soil nematode biomass[36], and represents 82% of
216 total human biomass on Earth (see Supplementary Methods). Using the same database of nematode
217 metabolic activity[34,35], we estimate that nematodes may be responsible for a monthly C turnover of 0.14 Gt
218 C within the global growing season, of which 0.11 Gt C is respired into the atmosphere (Table 1). For a
219 comparison, the amount of C respired by soil nematodes is equivalent to roughly ~15% of C emissions
220 from fossil fuel use, or ~2.2% of the total annual C emissions from soils (approximately 9 and 60 Gt C per

221 year, respectively[37]). As such, our findings indicate that soil nematodes are a major, and to date poorly
222 recognised, player in global soil C cycling.
223
224 Despite high confidence in our estimates of total nematode abundance and community composition, these
225 approximations of metabolic footprint retain several assumptions that might lead to considerable
226 uncertainty in our estimates. For example, seasonal climatic variation in metabolic activity could influence
227 the values we present here, and total activity levels might be lower than expected based on these growing
228 season estimates. On the other hand, extraction efficiency can be lower than 50% in some samples, which
229 could lead to underestimation of the actual activity levels. Local variation in land use types and bias in our
230 sampling data could cause variation in soil nematode abundances at local scales. Further, even though our
231 sampling locations cover the vast majority of environmental conditions on Earth (Extended Data Figs. 1c,
232 1e), our data underrepresented certain regions such as the Sahara and Arabian Desert, leading to relatively
233 high uncertainties in these regions (Fig. 2i, Extended Data Figs. 1a, 1b, 6). Also, as our sampling approach
234 focusses on the top soil layer, we stress that our analysis will underestimate total nematode abundances, for
235 example in tropical regions where high nematode densities are found in litter layers[38]. Yet, the metabolic
236 footprint that we provide enables us to approximate the magnitude of soil nematode contributions to global
237 carbon cycling and highlights their contribution to the total soil C budget. Further, our findings emphasize
238 the importance of high-latitude regions, characterized by high soil nematode abundances, in our
239 understanding of soil carbon and feedbacks to on-going climate change. These regions compose a major
240 reservoir of soil carbon stocks[6], and may release much more carbon as a result of increased soil animal
241 activity and a prolongation of the plant-growing season due to human-induced climate change.
242
243 In conclusion, our maps provide the first spatially-explicit, quantitative information of belowground biota
244 at a global scale. Besides providing baseline information about soil nematodes as a fundamental component
245 of terrestrial ecosystems, it also alters some of our most basic assumptions about the terrestrial biosphere
246 by highlighting that soil animal abundances peak in high latitude zones. The high nematode numbers that

247 are present across all global soils highlights their functional importance in global soil food web dynamics,
248 nutrient cycling terrestrial ecosystem functioning. This quantitative understanding of these belowground
249 animals enables us to begin to comprehend the order of magnitude of their influence on the global carbon
250 cycle, and the spatial patterns in these processes. By providing quantitative information about the variation
251 in biological activity in soils around the world, our models can provide the information necessary to
252 explicitly represent soil biotic activity levels in spatially-explicit biogeochemical models. That is, this
253 information can now be used to parameterize, scale or benchmark spatially-explicit model predictions of
254 organic matter turnover under current or future climate change scenarios. We highlight that this global
255 nematode study can and should be supplemented with similar future efforts to understand the biogeography
256 of other important soil organisms, including fungi, bacteria and protists. Our unique soil nematode
257 abundance and biomass data can serve as a stepping stone to facilitate future modelling efforts that add
258 additional layers of soil biodiversity information to build a thorough understanding of the overwhelming
259 abundance of life belowground and its impact on global ecosystem functioning.
260
261

Table 1 | Total nematode abundance, biomass and carbon budget.

Trophic group	Computed individuals (x 10^{20})	Fresh biomass (Mt)	Biomass (Mt C)	Monthly respiration (Mt C)	Monthly production (Mt C)	Monthly carbon budget (Mt C)
Bacterivores	1.92 ± 0.208	68.57 ± 7.42	7.13 ± 0. 77	34.17 ± 3.69	12.22 ± 1.31	46.39 ± 5.02
Fungivores	0.64 ± 0.065	9.56 ± 0.97	0.99 ± 0.10	6.49 ± 0.66	0.91 ± 0.09	7.40 ± 0.75
Herbivores	1.25 ± 0.114	83.41 ± 7.59	8.67 ± 0.79	26.74 ± 2.43	7.01 ± 0.64	33.75 ± 3.07
Omnivores	0.39 ± 0.046	96.50 ± 11.40	10.25 ± 1.19	27.38 ± 3.17	6.08 ± 0.70	33.46 ± 3.87
Predators	0.20 ± 0.031	42.25 ± 6.59	4.39 ± 0.68	15.06 ± 2.35	3.00 ± 0.46	18.06 ± 2.82
Total	4.40 ± 0.643	302.30 ± 33.99	31.44 ± 3.54	109.82 ± 12.31	29.24 ± 3.23	139.06 ± 15.54

Main text references

1 Cameron, E. K. *et al.* Global gaps in soil biodiversity data. *Nat Ecol Evol*, doi:10.1038/s41559-018-0573-8 (2018).

2 Bardgett, R. D. & van der Putten, W. H. Belowground biodiversity and ecosystem functioning. *Nature* **515**, 505-511, doi:10.1038/nature13855 (2014).

3 Paul, E. A. *Soil microbiology, ecology, and biochemistry*. 4th ed. edn, (Amsterdam : Elsevier, 2015).

4 Wieder, W. R., Bonan, G. B. & Allison, S. D. Global soil carbon projections are improved by modelling microbial processes. *Nature Climate Change* **3**, 909-912, doi:10.1038/nclimate1951 (2013).

5 Bradford, M. A. *et al.* A test of the hierarchical model of litter decomposition. *Nat Ecol Evol* **1**, 1836-1845, doi:10.1038/s41559-017-0367-4 (2017).

6 Crowther, T. W. *et al.* Quantifying global soil carbon losses in response to warming. *Nature* **540**, 104, doi:10.1038/nature20150 (2016).

7 Delgado-Baquerizo, M. *et al.* A global atlas of the dominant bacteria found in soil. *Science* **359**, 320-325, doi:10.1126/science.aap9516 (2018).

8 Thompson, L. R. *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457-463, doi:10.1038/nature24621 (2017).

9 Lozupone, C. A. & Knight, R. Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences* **104**, 11436-11440, doi:10.1073/pnas.0611525104 (2007).

10 Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences* **103**, 626-631, doi:10.1073/pnas.0507535103 (2006).

11 Bahram, M. *et al.* Structure and function of the global topsoil microbiome. *Nature*, doi:10.1038/s41586-018-0386-6 (2018).

12 Tedersoo, L. *et al.* Fungal biogeography. Global diversity and geography of soil fungi. *Science* **346**, 1256688, doi:10.1126/science.1256688 (2014).

13 Davison, J. *et al.* Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science* **349**, 970-973, doi:10.1126/science.aab1161 (2015).

14 Nielsen, U. N. *et al.* Global-scale patterns of assemblage structure of soil nematodes in relation to climate and ecosystem properties. *Global Ecology and Biogeography* **23**, 968-978, doi:10.1111/geb.12177 (2014).

15 Wu, T., Ayres, E., Bardgett, R. D., Wall, D. H. & Garey, J. R. Molecular study of worldwide distribution and diversity of soil animals. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 17720-17725, doi:10.1073/pnas.1103824108 (2011).

16 Boag, B. & Yeates, G. W. Soil nematode biodiversity in terrestrial ecosystems. *Biodiversity and Conservation* **7**, 617-630, doi:10.1023/a:1008852301349 (1998).

17 Song, D. *et al.* Large-scale patterns of distribution and diversity of terrestrial nematodes. *Applied Soil Ecology* **114**, 161-169, doi:10.1016/j.apsoil.2017.02.013 (2017).

18 Xu, X., Thornton, P. E. & Post, W. M. A global analysis of soil microbial biomass carbon, nitrogen and phosphorus in terrestrial ecosystems. *Global Ecology and Biogeography* **22**, 737-749, doi:10.1111/geb.12029 (2013).

19 Serna-Chavez, H. M., Fierer, N. & van Bodegom, P. M. Global drivers and patterns of microbial abundance in soil. *Global Ecology and Biogeography* **22**, 1162-1172, doi:10.1111/geb.12070 (2013).
20 Geisen, S. *et al.* Integrating quantitative morphological and qualitative molecular methods to analyse soil nematode community responses to plant range expansion. *Methods in Ecology and Evolution*, doi:10.1111/2041-210x.12999 (2018).
21 Darby, B. J., Todd, T. C. & Herman, M. A. High-throughput amplicon sequencing of rRNA genes requires a copy number correction to accurately reflect the effects of management practices on soil nematode community structure. *Molecular ecology* **22**, 5456-5471, doi:10.1111/mec.12480 (2013).
22 Carini, P. *et al.* Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nature Microbiology* **2**, 16242, doi:10.1038/nmicrobiol.2016.242 (2016).
23 Blazewicz, S. J., Barnard, R. L., Daly, R. A. & Firestone, M. K. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *ISME J* **7**, 2061-2068, doi:10.1038/ismej.2013.102 (2013).
24 Platt, H. M. in *The phylogenetic systematics of freeliving nematodes* (ed Lorenzen S.) i-ii (The Ray Society, 1994).
25 Ingham, R. E., Trofymow, J. A., Ingham, E. R. & Coleman, D. C. Interactions of Bacteria, Fungi, and their Nematode Grazers: Effects on Nutrient Cycling and Plant Growth. *Ecological Monographs* **55**, 119-140, doi:10.2307/1942528 (1985).
26 Ferris, H. Contribution of nematodes to the structure and function of the soil food web. *J Nematol* **42**, 63-67 (2010).
27 Crowther, T. W., Boddy, L. & Jones, T. H. Species-specific effects of soil fauna on fungal foraging and decomposition. *Oecologia* **167**, 535-545, doi:10.1007/s00442-011-2005-1 (2011).
28 Neher, D. A. Role of nematodes in soil health and their use as indicators. *Journal of Nematology* **33**, 161-168 (2001).
29 Procter, D. L. Global Overview of the Functional Roles of Soil-living Nematodes in Terrestrial Communities and Ecosystems. *J Nematol* **22**, 1-7 (1990).
30 Fierer, N., Strickland, M. S., Liptzin, D., Bradford, M. A. & Cleveland, C. C. Global patterns in belowground communities. *Ecol Lett* **12**, 1238-1249, doi:10.1111/j.1461-0248.2009.01360.x (2009).
31 Sohlenius, B. Abundance, Biomass and Contribution to Energy Flow by Soil Nematodes in Terrestrial Ecosystems. *Oikos* **34**, doi:10.2307/3544181 (1980).
32 Jobbágy, E. G. & Jackson, R. B. The Vertical Distribution of Soil Organic Carbon and Its Relation to Climate and Vegetation. *Ecological Applications* **10**, 423-436, doi:10.1890/1051-0761(2000)010[0423:Tvdoso]2.0.Co;2 (2000).
33 Ettema, C. H. Soil nematode diversity: species coexistence and ecosystem function. *J Nematol* **30**, 159-169 (1998).
34 Ferris, H. *Ecophysiology Parameters*, <http://nemaplex.ucdavis.edu/Ecology/EcophysiologyParms/EcoParameterMenu.html> (2018).

35 Mulder, C. & Vonk, J. A. Nematode traits and environmental constraints in 200 soil systems: scaling within the 60–6000 µm body size range. *Ecology*, doi:10.1890/11-0546.1 (2011).
36 Bar-On, Y. M., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas.1711842115 (2018).
37 Schlesinger, W. H. & Bernhardt, E. S. in *Biogeochemistry* 419-444 (2013).
38 Powers, T. O. *et al.* Tropical nematode diversity: vertical stratification of nematode communities in a Costa Rican humid lowland rainforest. *Molecular ecology* **18**, 985-996, doi:10.1111/j.1365-294X.2008.04075.x (2009).

Acknowledgements

This research was supported by a grant to T.W.C. from DOB Ecology and S.G. by the Netherlands Organisation for Scientific Research (grant 016.Veni.181.078). The authors thank E. Clark and A. Orgiazzi for critical review of the manuscript; R. Bouharroud, Z. Ferji, L. Jackson, and E. Mzough for providing data.

Author contributions

J.vdH., S.G., D.R. and T.W.C. designed and performed the data analyses. J.vdH, D.R., T.W.C. designed and performed geospatial analyses. J.H. S.G., H.F., R.G.M.dG., C.M. designed and performed biomass calculations. S.G., H.F., W.T., D.A.W., R.G.M.G, B.J.A., W.A., W.S.A., R.D.B., M.B., R.C.H., J.E.C., T.C., X.C., S.R.C., R.C., J.M.C.C., M.D., L.B.C., D.D., M.E., B.S.G., C.G., K.H., D.K., P.K., A.K., G.K., V.K., A.A.K., Q.L., W-J.L., M.M., M.M., J.A.R.M., E.M., E.H.M., C.M., P.M., R.N., T.A.D.N., U.N.N., H.O., J.E.P.R., K.P., V.P., L.P., J.C.P.S., C.P., T.O.P., K.P., C.W.Q., S.R., S.M., S.S., H.S., A.S., A.V.T., J.T., W.H.vdP., M.V., C.V., L.W., D.H.W., R.W., D.G.W. and Y-I.Y. contributed data. J.H, S.G. and T.W.C wrote the first draft of the manuscript with input from D.A.W. All authors contributed to editing of the paper.

Author information

Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to thomas.crowther@usys.ethz.ch.

Competing interests

One of the co-authors (WSA) recently became an employee of Nature Communications, a sister journal from the same publisher; he did not have any access to or involvement with the editorial process at Nature. All other authors declare no competing interests.

Author affiliations

[1]Crowther Lab, Institute of Integrative Biology, ETH Zürich, 8092 Zürich, Switzerland
[2]Department of Terrestrial Ecology, Netherlands Institute of Ecology, 6708 PB Wageningen, The Netherlands
[3]Department of Entomology & Nematology, University of California, Davis, CA 95616, USA
[4]Animal Ecology, Bielefeld University, 33615 Bielefeld, Germany
[5]Asian School of the Environment, Nanyang Technological University, 639798 Singapore
[6]Soil Biology Group, Wageningen University & Research, 6700AA Wageningen, The Netherlands
[7]Department of Biology, and Monte L. Bean Museum, Brigham Young University, Provo, UT 84602, USA
[8]Nematode Biodiversity Research Laboratory, Department of Zoology, Aligarh Muslim University, 202002 Aligarh, India
[9]Department of Biology and School of Global Environmental Sustainability, Colorado State University, 80523 - 1036 Fort Collins, USA
[10]School of Earth and Environmental Sciences, The University of Manchester, Manchester, M13 9PT, UK
[11]Department of Biology, Institute of Zoology, University of Cologne, 50674 Köln, Germany
[12] Cluster of Excellence on Plant Sciences CEPLAS), 50674 Köln, Germany
[13]Instituto de Ciencias de la Vid y del Vino, Universidad de La Rioja-Gobierno de La Rioja, Finca La Grajera, 26007 Logroño, Spain
[14]University of Brasília, Institute of Biological Sciences, Department of Phytopathology, 70910-900 Brasília, DF, Brazil
[15]School of Biological Sciences and Institute for Global Food Security, Queen's University of Belfast, BT9 5AH Belfast, Northern Ireland, UK
[16]Soil Ecology Lab, College of Resources and Environmental Sciences, Nanjing Agricultural University, 210095 Nanjing, China
[17]Centre of Molecular and Environmental Biology, University of Minho, 4710-057 Braga, Portugal
[18]Empresa Brasileira de Pesquisa Agropecuária, Embrapa, Centro de Pesquisa Agropecuária do Trópico Semiárido, 56302970 Petrolina, Brazil
[19]Zealand Institute of Business and Technology, 4200 Slagelse, Denmark
[20]Institut Sénégalais de Recherches Agricoles/CDH, BP 3120, Dakar, Senegal
[21]Instituto de Ciencias Agrarias, CSIC, 28006, Madrid, Spain
[22]SRUC, Crop and Soil Systems Research Group, Edinburgh, EH9 3JG, UK
[23]Senckenberg Museum of Natural History Görlitz, 02826 Görlitz, Germany
[24]Institute of Biology of Karelian Research Centre, Russian Academy of Sciences, 185910 Petrozavodsk, Russia
[25]Department of Forest Ecology and Management, Swedish University of Agricultural Sciences, S 901 83 Umeå, Sweden
[26]Laboratory of Functional Ecology, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland
[27]J.F. Blumenbach Institute of Zoology and Anthropology, University of Göttingen, 37073 Göttingen, Germany
[28]Institute of Biology of the Komi Scientific Centre, Ural Branch of the Russian Academy of Sciences, 167982, Syktyvkar, Russia
[29]Erguna Forest-Steppe Ecotone Research Station, Institute of Applied Ecology, Chinese Academy of Sciences, 110016 Shenyang, China
[30]Nematology Unit, Agricultural Research Council, Plant Health and Protection, Pretoria 0001, South Africa
[31]Dept. Environment, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, 28040 Madrid, Spain

[32]Laboratory of Biotechnology and Valorization of Natural Resources, Faculty of Science - Agadir, Ibn Zohr University, B.P 8106, Hay Dakhla, 80000 Agadir, Morocco
[33]Department of Biological, Geological and Environmental Sciences, University of Catania, 95124 Catania, Italy
[34]Department of Plant Pathology, University of Nebraska-Lincoln, Lincoln, NE 68583-0722, USA
[35]Ecological Sciences, The James Hutton Institute, Dundee, DD2 5DA, Scotland, UK
[36]Institute of Ecology and Biological Resources, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Cau Giay, 10000000 Hanoi, Vietnam.
[37]Hawkesbury Institute for the Environment, Western Sydney University, NSW 2751 Penrith, Australia
[38]Nematode Management Group, Division of Applied Entomology and Zoology, Central Region Agricultural Research Center, NARO, 2-1-18, Kan'nondai, Tsukuba, Ibaraki 305-8666, Japan
[39]Institute for Sustainable Agriculture, Spanish National Research Council, 14004 Córdoba, Spain
[40]Ecological Processes and Biodiversity, Center for Ecological Studies, Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610041, China
[41]Institute of Biodiversity and Ecosystem Research, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria
[42]Landscape Ecology, Institute of Terrestrial Ecosystems, Department of Environmental Systems Science, ETH Zürich, 8092 Zürich, Switzerland
[43]Swiss Federal Research Institute WSL, 8903 Birmensdorf, Switzerland
[44]Laboratory of Nematology, Department of Plant Pathology, Universidade Federal de Lavras, 37200000 Lavras, Brazil
[45]Biosystematics Group, Wageningen University, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands
[46]Laboratory of Nematology, Wageningen University, 6700 ES Wageningen, The Netherlands
[47]Institute of Biology, University of Neuchâtel, 2000 Neuchâtel, Switzerland
[48]Plant Protection Products Unit, National Institute of Agricultural and Food Research and Technology, 28040 Madrid, Spain
[49]Centre of Biodiversity and Sustainable Land Use, University of Göttingen, 37075 Göttingen, Germany
[50]Faculty of Biological and Environmental Sciences, Ecosystems and Environment Research Programme, University of Helsinki, FI-15140 Lahti, Finland
[51]A.N. Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, 117079 Moscow, Russia
[52]Eco&Sols, Univ Montpellier, IRD, CIRAD, INRA, Montpellier SupAgro, 34060 Montpellier, France
[53]Department of Agroecology, AU-Flakkebjerg, Aarhus University, Forsøgsvej 1, 4200 Slagelse, Denmark
[54]IRD, UMR ECO&SOLS, 34060 Montpellier, France
[55]ELISOL Environnement, 30111 Congénies, France
[56]Flanders Research Institute for Agriculture, Fisheries and Food, Plant Sciences Unit, B-9820 Merelbeke, Belgium
[57]Centre for Ecology & Hydrology, Lancaster Environment Centre, Lancaster LA1 4AP, UK
[58]Department of Plant Pathology and Microbiology, National Taiwan University, Taipai, 10607, Taiwan

Main figure legends

Figure 1 | Map of sample locations and abundance data. a, Sampling sites. A total of 6,759 samples were collected and aggregated into 1,876 1-km^2 pixels that were used for geospatial modelling and abundance data from 39 1-km^2 pixels from Antarctica. **b,** The median and interquartile range of nematode abundances (n = 1,875) per trophic group (top) and per biome (bottom) from all continents. Axes have been truncated for increased readability. Biomes with observations from more than 20 1-km^2 pixels are shown.

Figure 2 | Model and data validation. The standard error of the observed (**a**) and predicted (**b**) mean values of nematode density decrease with increasing sample size. The operation was repeated with 100 and 1,000 random seeds for the observed and predicted mean values, respectively, and the mean calculated standard errors are shown. **c-h,** Heat plots showing the relationships between predicted versus observed nematode abundance values, for total nematode number and each trophic group. Dashed diagonal lines indicate fitted relationships (R^2 values are indicated in the bottom right), solid diagonal lines indicate a 1:1 relationship between predicted and observed points. **i,** Bootstrapped (100 iterations) coefficient of variation (standard deviation divided by mean predicted value) as a measure of prediction accuracy. Sampling was stratified by biome. Overall, our prediction accuracy is lowest in arid regions and in parts of the Amazon and Malay Archipelago.

Figure 3 | Global map of soil nematode density at the 30 arc-seconds (~1 km^2) pixel scale. Nematodes per 100 g dry soil. Pixel values were binned into seven quantiles to create the colour palette.

Methods

Data acquisition

We collected data on soil nematode abundances that morphologically quantified nematodes and determined taxa to the level of trophic groups or taxonomic groups. Rather than taxonomic diversity, we decided to focus on trophic groups as this gives more functional information. Trophic groups were assigned based on Yeates, et al. [39]. We only collected samples that contained the following metadata: longitude and latitude, season or date sampled, sampling depth, information on land use (agriculture or natural sites) and if samples were collected from soils or litter. We then standardized our efforts by focusing on all samples that were derived from soils and in which samples were representative for nematode functional group composition in the top 15 cm of soils. This resulted in a final subset of 6,759 samples that were used for further analyses. Of these, 32.8% originate from agricultural or managed sites, and 67.2% from natural sites. All data points falling within the same 30 arc-seconds (~1-km^2) pixel were aggregated via an average, resulting in a total of 1,915 unique pixels across the globe as inputs into the models (Extended Data Table 1). 39 pixels located in Antarctica were removed from the dataset as the covariate layers have limited coverage in these regions. This resulted in a total of 1,876 unique pixels that were used for geospatial modelling.

Acquisition of global covariate layers

To create spatial predictive models of nematode abundance, we first sampled our prepared stack of 73 ecologically relevant, global map layers at each of the point locations within the dataset. These layers included climatic, soil nutrient, soil chemical, soil physical, vegetative indices, radiation and topographic variables and one anthropogenic covariate (Extended Data Table 2). All covariate map layers were resampled and reprojected to a unified pixel grid in EPSG:4326 (WGS84) at 30 arc-seconds resolution (≈1km at the equator). Layers with a higher original pixel resolution were downsampled using a mean aggregation method; layers with a lower original resolution were resampled using simple upsampling (i.e., without interpolation) to align with the higher resolution grid.

526 Geospatial modelling

527 Using the ClustOfVar package[40] in R, we reduced the covariates of interest to the most representative and
528 least collinear few. As we did not have a specific number of variables defined *a priori* to use as a parameter
529 for the clustering procedure, we put a range of cluster numbers (i.e., 5, 10, 15, 20) into the ClustOfVar
530 functions in order to compute multiple covariate groups for testing machine learning models. Using these
531 selections of variables, we used a "grid search" procedure to iteratively explore the results of a suite of
532 machine learning models trained on each group of covariates computed from the ClustOfVar function.
533 Moreover, following recent advancements in machine learning for spatial prediction[41], we tested models
534 using all covariates with and without latitude/longitude data as well as a specific selection of covariates
535 representing principal ecosystem components plus satellite-based spectral reflectance. In addition to grid
536 searching through models trained on different groupings of the covariates, we also explored the parameter
537 space of multiple machine learning algorithms (including random forests and regularized linear regression
538 with both L1 and L2 regularization) and optional post-hoc image convolution using kernels of various pixel
539 sizes. During the grid search procedure, we assessed each model using *k*-fold cross validation, to test the
540 performance and overfitting across each of the 405 models. For each fold, a 10% subset of the data was
541 extracted and held back for validation. Then, the model was trained on the remaining data, and tested on
542 the validation data. To test each model on the entire dataset, this process was performed 10 times for each
543 model (i.e., $k = 10$). computing coefficient of determination values for each fold that were then used to
544 compute mean and standard deviation values for the cross validated model. These mean and standard
545 deviation values were the basis for choosing the "best model" of all 405 models explored via the grid search
546 procedure, which was an iteration of random forests using all 73 non-spatial covariates. The grid search
547 procedure was performed using the total nematode abundance data, and this final model was then used to
548 model the sub-functional group abundance. The final R^2 value for the ensembled total nematode abundance
549 model (also assessed using 10-fold cross validation) was 0.43.
550
551 Model uncertainty

552 To create a per-pixel mean and standard deviation we ensembled multiple versions of the "best model"; as
553 the "best model" was an iteration of random forests using all 73 non-spatial covariates, the ensemble
554 procedure was to rerun this model 10 times (each with different random seed values) then averaging the
555 model results. Using these values we calculated the coefficient of variation (standard deviation divided by
556 the mean predicted value) as a measure of the prediction accuracy of our model (Fig 2i).
557
558 To create statistically valid per-pixel confidence intervals, we performed a stratified bootstrapping
559 procedure with the "total number" collection of nematode point data. The stratification category was the
560 sampled biomes of each point feature (to avoid biases), and the number of bootstrap iterations was 100.
561 Each of the bootstrap iterations required the classification of the composite raster data i.e., 209,000,000
562 pixels classified 100 times. Doing so allows us to generate per pixel, statistically robust 95% confidence
563 intervals (Extended Data Fig 1c).
564
565 Next, we tested the extent of extrapolation in our models by examining how many of the Earth's pixels
566 exist outside the range of our sampled data for each of the 73 global covariate layers. To evaluate the
567 sampled range, we extracted the minimum and maximum values of each covariate layer of the pixels in
568 which our sampling sites were located. Then, using the final model, we evaluated the number of variables
569 that fell outside the sampled range, across all terrestrial pixels. Next, we created a per-pixel representation
570 of the relative proportion of interpolation and extrapolation (Extended Data Fig. 1b). This revealed that our
571 samples covered the vast majority of environmental conditions on Earth, with 84% of Earth's pixels values
572 falling within the sampled range of at least 90% percent of all bands (Extended Data Fig. 1e). Across all
573 environmental layers, the percent of pixels with values within the sampled range is generally above 85%
574 (Extended Data Fig. 1f).
575
576 To evaluate how well our data spread throughout the full multivariate environmental covariate space, we
577 performed a Principal Components based approach. After performing a PCA on the sampled data, we used

the centering values, scaling values, and eigenvectors to transform the composite image into the same PCA spaces. Then, we created convex hulls for each of the bivariate combinations from the first 11 principal components (which collectively covered more than 80% of the sample space variation). Using the coordinates of these convex hulls, we classified whether each pixel falls within or outside each of these convex hulls. 62% of the world's pixels fell within the entire set of 55 PCA convex hull spaces computed from our sampled data, with most of the outliers existing in arid regions (Extended Data Fig 1e).

Geospatial analyses and extrapolation were performed in Google Earth Engine[42]. Additional model results can be found in the Extended Data.

Nematode density values

To compute the original nematode density values (which were in "number of nematodes per 100 grams of soil"), we performed the following calculations at a per-pixel level. First, we multiplied the value by 10 in order to compute nematodes per 1 kg of soil; the new units, per-pixel, became "number of nematodes per 1kg of soil". Then, we multiplied this value by the per-pixel bulk density values as produced by SoilGrids[43]; bulk density values were then produced in "kg of soil per 1 cubic meter". Finally, the new units after multiplication are the "number of nematodes per 1 cubic meter of soil". Next, we multiplied this value by 0.15 meters to compute the "number of nematodes per 1 square meter of soil (in the top 15 cm)". For pixels that had a soil layer shallower than 15 cm, the pixel value was multiplied by the depth to bedrock values as produced by SoilGrids[43]. These respective pixel values were then multiplied by the area of each pixel presumed to have soil (i.e., we exclude areas of "permanent snow/ice" and "open water" from the calculations, following the Consensus Land Cover classes found here: https://www.earthenv.org/landcover); the units at this point, per-pixel, are the total number of nematodes (in the first 15cm of soil). Finally, all pixel values were summed to compute the final nematode abundance values across all pixels (i.e., across the entire globe).

Clustering

To delineate main nematode 'community types', i.e. the relative frequency of each trophic group in a given sample, we first defined the number of clusters for the analysis. Based on pairwise distances and Partitioning Around Medoids (*k*-medoids) clustering we chose to select four clusters. Each of the four community types was then plotted (Extended Data Fig. 6b) to reveal their composition. To examine which environmental variables best explained each of the community types, we plotted each of the samples using a non-metric multidimensional scaling (stress = 0.0691) and fitted environmental variables as vectors (Extended Data Fig. 6c).

Biomass estimates

Using publicly available data[34,35], a database with taxon-specific body size values (i.e. length, width) of 32,728 nematode taxa (including 9,497 observations of adult nematodes and 23,231 observations of juveniles) was created to calculate the biomass, and respiration and assimilation rates for each trophic group. A nematode community typically contains numerous juveniles[35], we assume the presence of 70% juveniles and 30% adults. For all calculations described in this section, we calculated per-trophic group means using per-taxon observations. To produce the final values, we multiplied the mean calculated values per trophic group with the predicted number of individuals per trophic group and per biome. The biomass of an assemblage of nematodes can be calculated as the sum of the weights of the number of individuals of each species present. According to Andrassy [44], the fresh weight of individual nematodes is calculated by

$$W_{fresh} = \frac{L \cdot D^2}{1.6 \cdot 10^6}$$

where W_{fresh} is the fresh weight (µg) per individual, L is the nematode length (µm) and D is the greatest body diameter (µm)[44]. Assuming a dry weight of nematodes as 20% of fresh weight and the proportion of carbon in the body as 52% of dry weight[45,46], the dry weight (W_{dry}) of an individual nematode can be calculated as

628 $$W_{dry} = \frac{0.104 \cdot L \cdot D^2}{1.6 \cdot 10^6}$$

629

630 Daily carbon used in production

631 To calculate the total carbon utilized per nematode per day, we assumed that life cycle length in days can
632 be approximated as 12 times the colonizer-persister (cp) scale[47,48] and that the accumulation of fresh weight
633 is linear. Then, the daily increase in fresh weight is

634 $$R_W = \frac{W_t}{12 \cdot cp_t}$$

635 where W_t and cp_t are the adult weight and cp value for a nematode of trophic group *t*, respectively. Then,
636 we calculate the normalized daily carbon used in production (P_C) as

637 $$P_c = \frac{0.104 \cdot W_t}{12 \cdot cp_t}$$

638 where cp_t is the mean cp value of the respective trophic group. For a nematode assemblage, the daily carbon
639 used in production can be calculated as

640 $$P_c = \sum N_t \frac{0.104 \cdot W_t}{12 \cdot cp_t}$$

641 for N_t individuals of each trophic group present in the assemblage.

642

643 Carbon respiration

644 To estimate the carbon respiration rates of an assemblage of nematodes, we assume relationships between
645 respiration rates and body weights for poikilothermic organisms, so that

646 $$R = a \cdot W^b$$

647 where R is the respiration rate, W is the fresh weight (µg) per individual, and *a* and *b* are regression
648 parameters[49,50]. Following literature, we assume that *b* is equal to 0.75[51,52]. The parameter *a* varies with
649 temperature and the time interval on which the rate is based. For example, Klekowski, et al. [53] determined
650 an average *a*-value of approximately 1.40 nl O_2 h^{-1} for 68 nematode species. This converts to an *a*-value of

651 2.43 ng CO_2 h^{-1} at 15 ˚C. To estimate CO_2 respiration in µg per day, we make the assumption of an *a*-value
652 of 2.43 × 24/1000 (= 0.058) for our calculations. Using the relative molecular weights of carbon and oxygen
653 in CO_2 (12/44 = 0.273), we can calculate the total rate of carbon respiration for all nematodes in the system
654 as

655 $$R = \sum N_t \cdot 0.273 \cdot 0.058 \, W_t^{0.75}$$

656 or

657 $$R = \sum N_t \cdot 0.0159 \, W_t^{0.75}$$

658 where N_t is the number of individuals and W_t the median body weight of each of the trophic groups summed
659 over *t* trophic groups.
660
661 Total daily carbon budget
662 The total carbon budget (in µg per day) for each trophic group is the sum amounts that are respired and
663 used for production, that is:

664 $$C_{tot} = \sum \frac{N^t \cdot 0.104 \cdot W_t}{12 \cdot cp_t} + N_t \cdot 0.0159 \cdot (W_t)^{0.75}$$

665
666
667 **Data and code availability**
668 All raw data, source code, sampled covariate layer data, models and maps are available under:
669 https://gitlab.ethz.ch/devinrouth/Crowtherlab_Nematode
670
671
672 **Additional References**

673 34 Ferris, H. *Ecophysiology Parameters*,
674 <http://nemaplex.ucdavis.edu/Ecology/EcophysiologyParms/EcoParameterMenu.html>
675 (2018).

35 Mulder, C. & Vonk, J. A. Nematode traits and environmental constraints in 200 soil systems: scaling within the 60–6000 µm body size range. *Ecology*, doi:10.1890/11-0546.1 (2011).
39 Yeates, G., Bongers, T., De Goede, R., Freckman, D. & Georgieva, S. Feeding habits in soil nematode families and genera—an outline for soil ecologists. *Journal of Nematology* **25**, 315 (1993).
40 Chavent, M., Simonet, V. K., Liquet, B. & Saracco, J. ClustOfVar: An R Package for the Clustering of Variables. *Journal of Statistical Software* **50**, 1-16, doi:arXiv:1112.0295 (2012).
41 Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M. & Gräler, B. Random Forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, doi:10.7287/peerj.preprints.26693v2 (2018).
42 Gorelick, N. *et al.* Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* **202**, 18-27, doi:10.1016/j.rse.2017.06.031 (2017).
43 Hengl, T. *et al.* SoilGrids250m: Global gridded soil information based on machine learning. *PloS one* **12**, e0169748, doi:10.1371/journal.pone.0169748 (2017).
44 Andrassy, I. Die rauminhalts-und gewichtsbestimmung der fadenwürmer (Nematoden). *Acta Zoologica Hungarica* **2**, 1-5 (1956).
45 Mulder, C., Cohen, J. E., Setälä, H., Bloem, J. & Breure, A. M. Bacterial traits, organism mass, and numerical abundance in the detrital soil food web of Dutch agricultural grasslands. *Ecol. Lett.* **8**, 80-90, doi:10.1111/j.1461-0248.2004.00704.x (2005).
46 Persson, T. in *Proc. VIII Int. Colloq. Soil Zool* (eds P Lebrun *et al.*) 117-126 (1983).
47 Bongers, T. The maturity index: an ecological measure of environmental disturbance based on nematode species composition. *Oecologia* **83**, 14-19, doi:10.1007/BF00324627 (1990).
48 Bongers, T. The Maturity Index, the evolution of nematode life history traits, adaptive radiation and cp-scaling. *Plant Soil* **212**, 13-22, doi:10.1023/a:1004571900425 (1999).
49 Kleiber, M. Body size and metabolism. *Hilgardia* **6**, 315-353, doi:10.3733/hilg.v06n11p315 (1932).
50 West, G. B., H., B. J. & Enquist, J. B. A General Model for the Origin of Allometric Scaling Laws in Biology. *Science* **276**, 122-126, doi:10.1126/science.276.5309.122 (1997).
51 Atkinson, H. J. in *Nematodes as Biological Models* Vol. 2 (ed B. M. Zuckerman) 122-126 (Academic Press, 1980).
52 Klekowski, R. Z., Wasilewska, L. & Paplinska, E. Oxygen Consumption in the Developmental Stages of Panagrolaimus Rigid Us. *Nematologica* **20**, 61-68, doi:10.1163/187529274x00591 (1974).
53 Klekowski, R. Z., Paplinska, E. & Wasilewska, L. Oxygen Consumption By Soil-Inhabiting Nematodes. *Nematologica* **18**, 391-403, doi:10.1163/187529272x00665 (1972).

Extended Data Legends

Extended Data Fig. 1 | Model accuracy assessment and extent of interpolation and extrapolation across all terrestrial pixels in 73 global covariate layers. a, coefficient of variation (standard deviation as a fraction of the mean predicted value) as a measure of the prediction accuracy of our model. **b,** proportional extent of interpolation (purple) *vs.* extrapolation (red) in univariate space. **c,** Percentage of pixels that fall within the convex hulls of the first 11 principal component spaces (collectively covering >80% of the sample space variation). **d,** percentage of pixels interpolated as a function of the percent of global environmental conditions covered by the sample set. On the global scale, 86% of the Earth's pixels have at least 90% of the covariate bands falling within the sampled range of environmental conditions. **e,** percentage of pixels falling within the 55 convex hull spaces of the first 11 Principal Components (collectively explaining >80% of the variation. On the global scale, 62% of the Earth's pixels fell within 100% of 55 PCA convex hull spaces. **f,** percent of terrestrial pixels falling within the sampled range, per covariate band.

Extended Data Fig. 2 | Linear regression models of the most important variables from the final random forest model and annual mean temperature. Soil organic carbon and cation-exchange capacity have a positive correlation with total nematode abundance, pH is negatively correlated. These linear regression models (n = 1,809) were not used to create global perspectives of nematode distribution patterns. The grey area represents the 95% confidence interval for the mean.

Extended Data Fig. 3 | Global maps of nematode trophic group abundance. a, bacterivores. **b,** fungivores. **c,** herbivores. **d,** omnivores. **e,** predators. Scales differ per map. Most trophic groups show similar patterns, but predators (**e**) are predicted to be present in particularly high abundances in some arid soils e.g. in the Sahara and Arabian Desert. Pixel values were binned into seven quantiles to create the colour palette.

Extended Data Fig. 4 | Global map of total nematode abundance per unit area (m^2). Correcting for the lower bulk density in soils high in organic matter, this map shows the same global patterns of nematode abundance as in Fig. 3. Hence, it is not low soil bulk density in boreal regions resulting in the observed patterns, but rather the high nematode abundances. Pixel values were binned into seven quantiles to create the colour palette.

Extended Data Fig. 5 | Global maps of nematode trophic group abundance per unit area (m^2). a, bacterivores. **b,** fungivores. **c,** herbivores. **d,** omnivores. **e,** predators. Scales differ per map. Correcting for the lower bulk density in soils high in organic matter, these maps show the same global patterns of nematode trophic group abundance as in Extended Data Figs. 3a-e. Pixel values were binned into seven quantiles to create the colour palette.

Extended Data Fig. 6 | Community types and driving variables of community type composition. a, Correlations between trophic groups. Overall, correlations of predators with other trophic groups are the least positive. **b,** based on the relative abundance of each trophic group, soil nematode communities can be classified in four distinct types. We find that these soil nematode communities are dominated by either herbivores (1), herbivores and bacterivores (2), bacterivores (3), or have a mixed composition (4). **c,** non-metric multidimensional scaling to highlight environmental conditions that drive the composition of each of the four main community types. Vegetation-type indices, such as NDVI and Fpar, drive the dominance of herbivores in nematode communities (type 1), while edaphic characteristics are correlated with communities dominated by microbivores (types 3 and 4). The names of the environmental variables are listed in Supplementary Table 3.

Supplementary Table 1 | Nematode abundance data and corresponding metadata values. Abundance data for each trophic group and associated metadata from 1,876 1-km^2 pixels that were used for geospatial modelling and abundance data from 39 1-km^2 pixels from Antarctica. (.csv file)

Supplementary Table 2 | Summary of mean, median and sample size values per biome. The number of sites corresponds to the number of 1-km^2 pixels into which the samples were aggregated. (.csv file)

Supplementary Table 3 | Global covariate layers used for geospatial modelling. A total of 73 global covariate layers was used in our modelling approach. The 7 Nadir Reflectance Band layers (i.e., MCD43A4.005 BRDF-Adjusted Reflectance 16-Day Global 500m) are summarised as one entry in the table. (.xlsx file)

Supplementary Table 4 | Variable importance metrics. Edaphic characteristics emerged as the most important variables. As the full dataset includes collinear variables leading to a false representation of the variable importance metrics, analysis was performed on a selection of main variables. (.xlsx file)

Supplementary Table 5 | Number of soil nematodes per trophic group, per biome. Summing the predicted number of nematodes per 1 km^2 pixel across biomes we estimate a total of 4.4×10^{20} nematodes are present in the top 15 cm of soil across the globe. (.csv file)

Supplementary Table 6 | Relative abundance of soil nematodes per trophic group, per biome. (.csv file)

Supplementary Table 7 | Nematode biomass per trophic group, per biome. Note that values are presented in megatons (10^6 tons) carbon. (.csv file)

Supplementary Table 8 | Relative nematode biomass per trophic, per biome. (.csv file)