Globally invariant metabolism but density-diversity mismatch in springtails

Anton M. Potapov[1,2]*, Carlos A. Guerra[3,4], Johan van den Hoogen[5], Anatoly Babenko[2], Bruno C. Bellini[6], Matty P. Berg[7,8], Steven L. Chown[9], Louis Deharveng[10], Ľubomír Kováč[11], Natalia A. Kuznetsova[12], Jean-François Ponge[13], Mikhail B. Potapov[12], David J. Russell[14], Douglas Alexandre[15], Juha M. Alatalo[16], Javier I. Arbea[17], Ipsa Bandyopadhyay[18], Verónica Bernava[19], Stef Bokhorst[7], Thomas Bolger[20,21], Gabriela Castaño-Meneses[22], Matthieu Chauvat[23], Ting-Wen Chen[24,1], Mathilde Chomel[25], Aimee T. Classen[26], Jerome Cortet[27], Peter Čuchta[24], Ana Manuela de la Pedrosa[25] Susana S. D. Ferreira[7], Cristina Fiera[27], Juliane Filser[28], Oscar Franken[7,8,85], Saori Fujii[29], Essivi Gagnon Koudji[30], Meixiang Gao[31], Benoit Gendreau-Berthiaume[32], Diego F. Gomez-Pamies[33], Michelle Greve[34], I. Tanya Handa[30], Charlène Heiniger[35], Martin Holmstrup[36], Pablo Homet[37], Mari Ivask[38], Charlene Janion-Scheepers[39,4], Malte Jochum[3,4], Sophie Joimel[40], Bruna Claudia S. Jorge[41], Edite Jucevica[42], Luís Carlos Iuñes de Oliveira Filho[43], Osmar Klauberg-Filho[43], Dilmar Baretta[44], Eveline J. Krab[45,46], Annely Kuu[47], Estevam C. A. de Lima[48], Dunmei Lin[49], Amy Liu[9], Jing-Zhong Lu[1], María José Luciañez[25], Michael T. Marx[50], Matthew M. McCary[51], Maria A. Minor[52], Taizo Nakamori[53], Ilaria Negri[54], Raúl Ochoa-Hueso[55,56], José G. Palacios-Vargas[57], Melanie M. Pollierer[1], Pascal Querner[58,59], Natália Raschmanová[11], Muhammad Imtiaz Rashid[60], Laura J. Raymond-Léonard[30], Laurent Rousseau[30], Ruslan A. Saifutdinov[2], Sandrine Salmon[61], Emma J. Sayer[62,63], Nicole Scheunemann[1,64], Cornelia Scholz[59], Julia Seeber[65,66], Yulia B. Shveenkova[67], Sophya K. Stebaeva[2], Maria Sterzynska[68], Xin Sun[69], Winda I Susanti[1], Anastasia A. Taskaeva[70], Madhav P. Thakur[71], Maria A. Tsiafouli[72], Matthew S. Turnbull[73], Mthokozisi N. Twala[34], Alexei V. Uvarov[2], Lisa A. Venier[74], Lina A. Widenfalk[75,76], Bruna R. Winck[41], Daniel Winkler[77], Donghui Wu[78,79,80], Zhijing Xie[78], Rui Yin[81,82], Douglas Zeppelini[83], Thomas W. Crowther[5], Nico Eisenhauer[3,81], Stefan Scheu[1,84]

[1]Johann Friedrich Blumenbach Institute of Zoology and Anthropology, University of Göttingen, Göttingen, Germany. [2]A.N. Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Moscow, Russia. [3]Experimental Interaction Ecology Group, German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany. [4]Institute of Biology, Martin Luther University Halle Wittenberg, Halle (Saale), Germany. [5]Department of Environmental Systems Science, Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland. [6]Department of Botany and Zoology, Federal University of Rio Grande do Norte, RN, Brazil. [7]Department of Ecological Science, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands.

[8]Community and Conservation Ecology Group, Groningen Institute of Evolutionary Life Science, University of Groningen, Amsterdam, the Netherlands. [9]Securing Antarctica's Environmental Future, School of Biological Sciences, Monash University, Australia. [10]ISYEB, Muséum National d'Histoire Naturelle, Paris, France. [11]Department of Zoology, Institute of Biology and Ecology, Faculty of Science, Pavol Jozef Šafárik University in Košice, košice, Slovakia. [12]Institute of Biology and Chemistry, Moscow Pedagogical State University, Moscow, Russia. [13]Département Adaptations du Vivant, Muséum National d'Histoire Naturelle, Brunoy, France. [14]Department of Soil Zoology, Senckenberg Society for Nature Research, Görlitz, Germany. [15]Department of Soil Science, Centre for Agriculture and Veterinary Science, Santa Catarina State University University (UDUESC- Lages), Lages, SC, Brazil. [16]Environmental Science Center, Qatar University, Doha, Qatar. [17]Department of Sciences, CEPA Camargo, Astillero, Spain. [18]Visva Bharati University, Bengal, India. [19]Administración de Parques Nacionales, Argentina. [20]School of Biology and Environmental Science, University College Dublin, Dublin, Ireland. [21]Earth Institute, University College Dublin, Dublin, Ireland. [22]Unidad Multidisciplinaria de Docencia e Investigación, Facultad de Ciencias, Campus Juriquilla, Universidad Nacional Autónoma de México, Querétaro, México. [23]Normandie University - UNIROUEN, INRAE, ECODIV, Rouen, France. [24]Biology Centre of the Czech Academy of Sciences, Institute of Soil Biology, České Budějovice, Czech Republic. [25]Departmento de Biología (Zoología, Universidad Autónoma de Madrid, Spain. [26]Department of Ecology & Evolutionary Biology, University of Michigan, MI 48109. [27]Institute of Biology Bucharest, Romanian Academy, Romania. [28]FB 02, UFT, General and Theoretical Ecology, University of Bremen, Bremen, Germany. [29]Department of Forest Entomology, Forestry and Forest Products Research Institute, Tsukuba, Japan. [30]Département des Sciences Biologiques, Université du Québec à Montréal, Québec, Canada. [31]Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo, China. [32]Département des Sciences Naturelles, Université du Québec en Outaouais, Québec, Canada. [33]Instituto de Biología Subtropical, Consejo Nacional de Investigaciones Científicas y Técnicas-Universidad Nacional de Misiones, Puerto Iguazú, Argentina. [34]Department of Plant and Soil Sciences, University of Pretoria, Pretoria, South Africa. [35]HES-SO University of Applied Sciences and Arts Western Switzerland. [36]Section of Terrestrial Ecology, Department of Ecoscience, Aarhus University, Aarhus, Denmark. [37]Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS), Consejo Superior de Investigaciones Científicas (CSIC), Sevilla, Spain. [38]Tartu College, Tallinn University of Technology, [39]Department of Biological Sciences, University of Cape Town, Rondebosch, South Africa. [40]Université Paris-Saclay, INRAE, AgroParisTech, UMR EcoSys, 78850 Thiverval-Grignon. [41]Quantitative Ecology Lab, Department of Ecology, Universidade Federal do Rio Grande

do Sul, Porto Alegre, Brazil. [42]Institute of Biology, University of Latvia, Riga, Latvia. [43]Department of Soil Science, Centre for Agriculture and Veterinary Science, Santa Catarina State University (UDESC-Lages), Lages, SC, Brazil. [44]Department of Animal Science, Santa Catarina State University (UDESC Oeste), Chapecó, SC, Brazil. [45]Department of Soil and Environment, Swedish University of Agricultural Sciences, Uppsala, Sweden. [46]Climate Impacts Research Centre, Department of Ecology and Environmental Science, Umeå University, Abisko, Sweden. [47]Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences,Tartu, Estonia. [48]Departamento de Entomologia, Museu Nacional, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil. [49]Key Laboratory of the Three Gorges Reservoir Region's Eco-Environment, Ministry of Education, Chongqing University, Chongqing, China. [50]Institute of Zoology, Johannes Gutenberg University Mainz, Mainz, Germany. [51]Department of BioSciences, Rice University, Houston, US. [52]Wildlife & Ecology Group, School of Agriculture and Environment, Massey University, New Zealand. [53]Graduate School of Environment and Information Sciences, Yokohama National University, Yokohama, Japan. [54]Department of Sustainable Crop Production (DI.PRO.VE.S.), Università Cattolica del Sacro Cuore, Piacenza, Italy. [55]Department of Biology, IVAGRO, University of Cádiz, Puerto Real, Spain. [56]Department of Terrestrial Ecology, Netherlands Institute of Ecology (NIOO KNAW), Wageningen, NL-6700 AB the Netherlands. [57]Lab. Ecología y Sistemática de Microartrópodos, Depto. Ecología y Recursos Naturales, Facultad de Ciencias, Universidad Nacional Autónoma de México, México, México. [58]Natural History Museum Vienna, 1. Zoology, Vienna, Austria. [59]Department of Integrated Biology and Biodiversity Research, University of Natural Resources and Life Sciences, Vienna, Austria. [60]Center of Excellence in Environmental Studies, King Abdulaziz University, Saudi Arabia. [61]UMR 7179 MECADEV - AVIV department, Muséum National d'Histoire Naturelle, Brunoy, France. [62]Lancaster Environment Centre, Lancaster University, Lancaster, UK. [63]Smithsonian Tropical Research Institute, Balboa, Ancon, Panama, Rep. Panama. [64]Department of Soil Zoology, Senckenberg Museum of Natural History Görlitz, Görlitz, Germany. [65]Institute for Alpine Environment, Eurac Research, Bozen, Italy. [66]Department of Ecology, University of Innsbruck, Innsbruck, Austria. [67]State Nature Reserve "Privolzhskaya Lesostep", Penza, Russia. [68]Department of Systematics, Zoogeography and Ecology of Invertebrates, Museum and Institute of Zoology Polish Academy of Science, Warsaw, Poland. [69]Key Laboratory of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, China. [70]Institute of Biology, Komi Science Centre, Ural Branch of Russian Academy of Sciences, Syktyvkar, Russia. [71]Institute of Ecology & Evolution, University of Bern, Bern, Switzerland. [72]Department of Ecology, School of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece. [73]Unaffiliated, Edmonton, Canada. [74]Canadian

Forest Service, Natural Resources Canada, Sault Ste. Marie, Canada. [75]Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden. [76]Greensway AB, Uppsala, Sweden. [77]Institute of Wildlife Management and Wildlife Biology, University of Sopron, Sopron, Hungary. [78]Key laboratory of Wetland Ecology and Environment, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China. [79]Key Laboratory of Vegetation Ecology, Ministry of Education, Northeast Normal University, Changchun 130024, China. [80]Jilin Provincial Key Laboratory of Animal Resource Conservation and Utilization, Northeast Normal University, Changchun 130117, China. [81]Institute of Biology, Leipzig University, Leipzig, Germany. [82]Community Department, Helmholtz Center for Environmental Research, Halle, Germany. [83]Department of Biology, Paraiba State University, Brazil. [84]Centre of Biodiversity and Sustainable Land Use, University of Göttingen, Göttingen, Germany. [85]Department of Coastal Systems, Royal Netherlands Institute for Sea Research, Landsdiep 4, 't Horntje (Texel), 1797 SZ the Netherlands. *e-mail: potapov.msu@gmail.com

Soil life supports the functioning and biodiversity of terrestrial ecosystems[1,2]. Springtails (Collembola) are among the most abundant soil animals regulating soil fertility and flow of energy through above- and belowground food webs[3–5]. However, the global distribution of springtail diversity and density, and how these relate to energy fluxes remains unknown. Here, using a global dataset collected from 2,470 sites, we estimate total soil springtail biomass at 29 Mt carbon (threefold higher than wild terrestrial vertebrates[6]) and record peak densities up to 2 million individuals per m² in the Arctic. Despite a 20-fold biomass difference between tundra and the tropics, springtail energy use (community metabolism) remains similar across the latitudinal gradient, owing to the increase in temperature. Neither springtail density nor community metabolism were predicted by local species richness, which was highest in the tropics, but comparably high in some temperate forests and even tundra. Changes in springtail activity may emerge from latitudinal gradients in temperature, predation[7,8], and resource limitation[7,9,10] in soil communities. Contrasting temperature

responses of biomass, diversity and activity of springtail communities suggest that climate warming will alter fundamental soil biodiversity metrics in different directions, potentially restructuring terrestrial food webs and affecting major soil functions.

Soil biodiversity is an essential component of every terrestrial habitat that affects nutrient cycling, soil fertility and plant-soil feedbacks, among other ecosystem functions and services[1,2,11]. Soil functioning is jointly driven by multiple components of soil biota that are closely interconnected, including plants, microorganisms, micro-, meso-, and macrofauna[12,13]. Land use, human activities, and climate changes induce widespread and rapid changes in the abundance, diversity, and activity of soil biota, altering functional connections and ecosystem-level processes in the terrestrial biosphere[14]. To understand, predict, and adapt to these changes, comprehensive knowledge about the global distribution of multiple soil biota components is urgently needed[15,16].

With a growing understanding of the biogeography of microorganisms[17], micro-[18] and macrofauna[19], a critical knowledge gap is the global distribution of soil mesofauna. Springtails (Collembola, Hexapoda) are among the most abundant groups of mesofauna and soil animals from the equator to polar regions[4,5]. They are mostly microbial feeders, but also graze on litter and are often closely associated with plant roots[3,20]. Through these trophic relationships, springtails affect the growth and dispersal of prokaryotes, fungi, and plants, thereby supporting nutrient cycling via the transformation, degradation, and stabilisation of organic matter[5,21]. Furthermore, springtails are a key food resource for soil- and surface-dwelling predators[3,5], thus occupying a central position in soil food webs and supporting multitrophic biodiversity.

To assess different functional facets of biological communities, metrics such as population density and biomass (reflecting carbon stocks), taxonomic and phylogenetic diversity

(ensuring multifunctionality and stability), and metabolic activity (quantifying energy fluxes and thus functional influence) are commonly used[6,22–24]. Soil biodiversity assessments have found unexpected global hotspots in temperate regions for microorganisms (fungi and prokaryotes)[17] and macrofauna (earthworms)[19], which are not in line with the common latitudinal biodiversity gradient found in aboveground organisms[25]. Functional complementarity principles[23] suggest that diverse soil communities in temperate ecosystems are able to support higher organismal densities and have a more efficient resource use (i.e., higher total activity) than at other latitudes. However, there are no global assessments of soil animal metabolic activities. In contrast to expectations of complementarity principles, previous studies on plants[26,27] and microbes[9,17] suggest that diversity and activity (represented by respiration) do not co-vary at the global scale, probably because strong environmental constraints limit this relationship. These discrepancies emphasize the need to investigate relationships of multiple metrics of soil animal communities. Springtails are an ideal model organism for exploring such relationships at a global scale, due to their ubiquity, functional diversity and high local species richness[3–5].

Current knowledge suggests that springtails are especially abundant and diverse in temperate coniferous forests and tundra, but less diverse in polar regions[24,28]. Many springtails are adapted to high and stable humidity, and sensitive to drought and temperature changes[29,30]. Consequently, springtail density and diversity is likely to decrease with future climate change, detrimentally affecting soil food webs and ecosystem functioning[31]. At the same time, springtail densities are relatively high in urban areas and in agricultural fields[32,33], so global springtail biomass may be moderately affected by land-use changes worldwide. Disentangling the roles of vegetation, climate, human disturbance, and other drivers of various springtail community metrics will be critical to understand their contribution to soil functioning under different global change scenarios[15,18].

Here, we report the joint projection of density, diversity, and metabolic activity of soil springtail communities at the global scale and test whether high species richness supports increased density and total activity across springtail communities globally, or whether this relationship is constrained by environmental and biotic controls. We further aimed (1) to assess whether the global distribution of springtail diversity matches that of aboveground biota or other soil animals; (2) to test how different metrics of springtail communities are affected by climate and human activities; and (3) to quantify the global biomass of springtails as a component of the global carbon stock. Using an extensive dataset of soil springtail communities collected within the framework of the #GlobalCollembola initiative[5] (2,470 sites and 43,601 samples across all continents; Fig. 1a), we show contrasting patterns across soil biodiversity metrics at a global scale and demonstrate that springtails are among the most functionally important and ubiquitous animals in the terrestrial biosphere.

Latitudinal gradient

To calculate total biomass and metabolism of each springtail community, we used recorded population densities together with estimated individual body masses and metabolic rates. Body masses and metabolic rates were derived from taxon-specific body lengths using mean annual soil temperature and allometric regressions (for calculations and parameter uncertainties see Methods). For the assessment of local species richness, we selected 70% of the sampling sites with taxonomically-resolved communities and calculated rarefaction curves to account for unequal sampling efforts. As such, our trends refer to local diversity (hundreds of meters), but may not be representative of regional-level diversity[34].

Springtail density varied c. 30-fold across latitudes (Fig. 1b), with maximum densities in tundra (median = 131,422 individuals m^{-2}) and minimum densities in tropical forests (5,831 individuals m^{-2}) and agricultural ecosystems (3,438 individuals m^{-2}; Fig. S2; n = 2,210).

Springtail dry biomass followed the same trend, with c. 20-fold higher biomass in tundra (median = 3.09 g m^{-2}) compared to tropical agricultural and forest ecosystems (c. 0.16 g m^{-2}), due to a lower average community body mass in polar as opposed to temperate and tropical ecosystems (Fig. 1d,f; Fig. S2; n = 2,053). These density and biomass estimates are in line with earlier studies[24] but cover wider environmental gradients. The difference in average community body mass may be explained by lower proportion of large surface-dwelling springtail genera in polar regions[35].

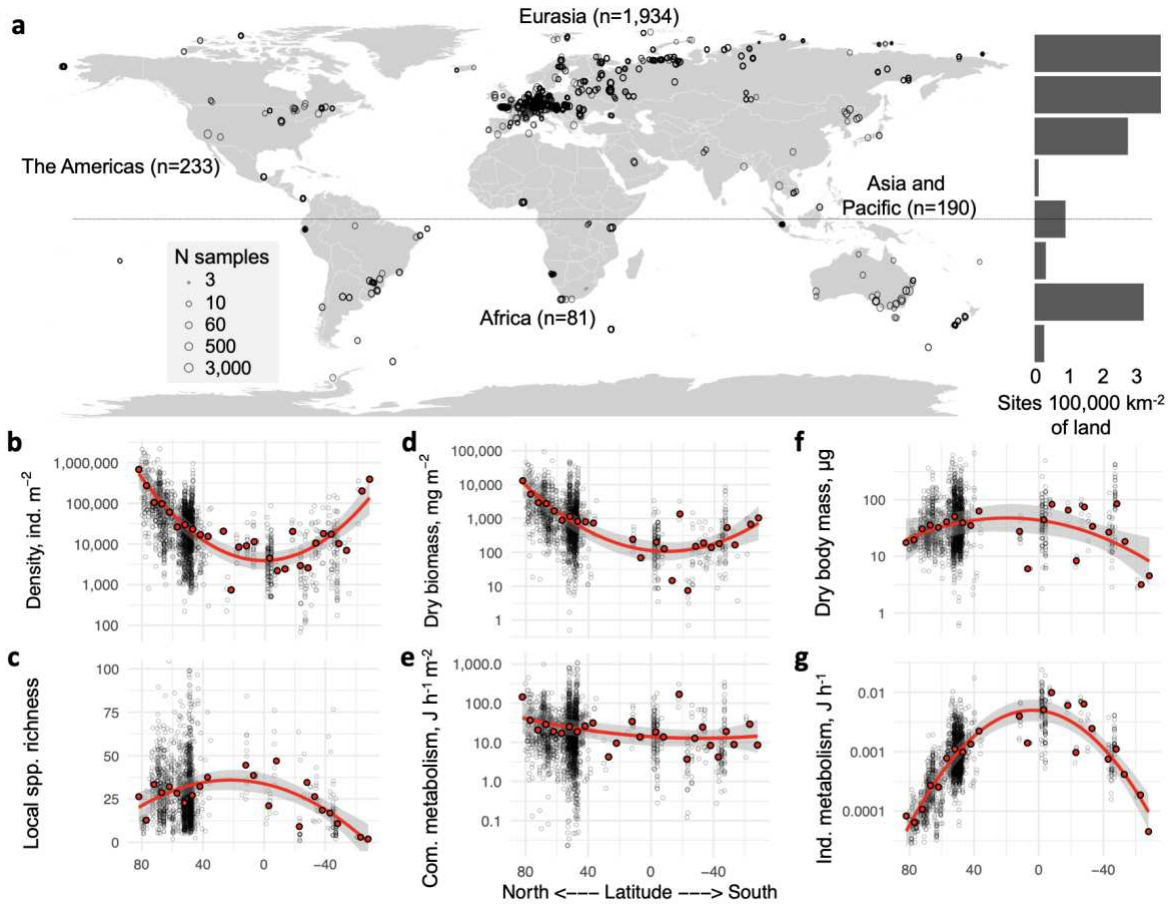


Fig. 1 | Sampling locations and latitudinal gradients in springtail community metrics. a, Distribution of the 2,470 sampling sites (43,601 soil samples). The histogram shows the number of sites in each 20-degree latitudinal belt, relative to the total land area in the belt. **b-g**, Variation in density (n = 2,210), local species richness (n = 1,735), biomass, community metabolism, average body mass and average individual metabolism (n = 2,053) with latitude.

Grey circles across panels show sampling sites; red points are averages for 5-degree latitudinal belts; trends are illustrated with a quadratic function based on 5-degree averages.

Being dependent on temperature and body mass, average individual metabolism was approximately 20 times higher in tropical than in polar ecosystems (Fig. 1g), which resulted in similar community metabolism across the latitudinal gradient (Fig. 1e; total n = 2,053). Hence, tropical springtail communities expend a similar amount of energy per unit time and area as polar communities, despite having 20-fold lower biomass. This striking pattern resembles aboveground ecosystem respiration, which also changes little across the global temperature gradient[27]. High metabolic rates but low densities of springtail communities are consistent with the high soil respiration rates and low litter accumulation in the tropics compared to biomes at higher latitudes[9,16]. Litter removal is facilitated by soil animals, which have to consume more food per unit biomass to meet their metabolic needs under high tropical temperatures[7] and thus enhance decomposition in wet and warm tropical ecosystems[10]. This suggests that soil animal communities in the tropics are under strong bottom-up control (by the amount and quality of litter), but also under strong top-down control by predators, which likewise have to feed more at high temperatures[7,8]. By contrast, polar communities have access to ample organic matter stocks[16], are under weaker top-down control[7,8], but their activity is constrained by the cold environment. The latitudinal gradient in environmental and biotic controls may explain why community metabolism did not increase as expected towards warm tropical ecosystems.

We found only weak latitudinal trends in local species richness, which was highest in tropical forests (mean = 36.6 species site^{-1}) and lowest in temperate agricultural (19.5 species site^{-1}) and grassland ecosystems (22.8 species site^{-1}; Fig. 1c; Fig. S2). Generally, the similar local diversity in different climates deviates from the latitudinal biodiversity gradients reported for

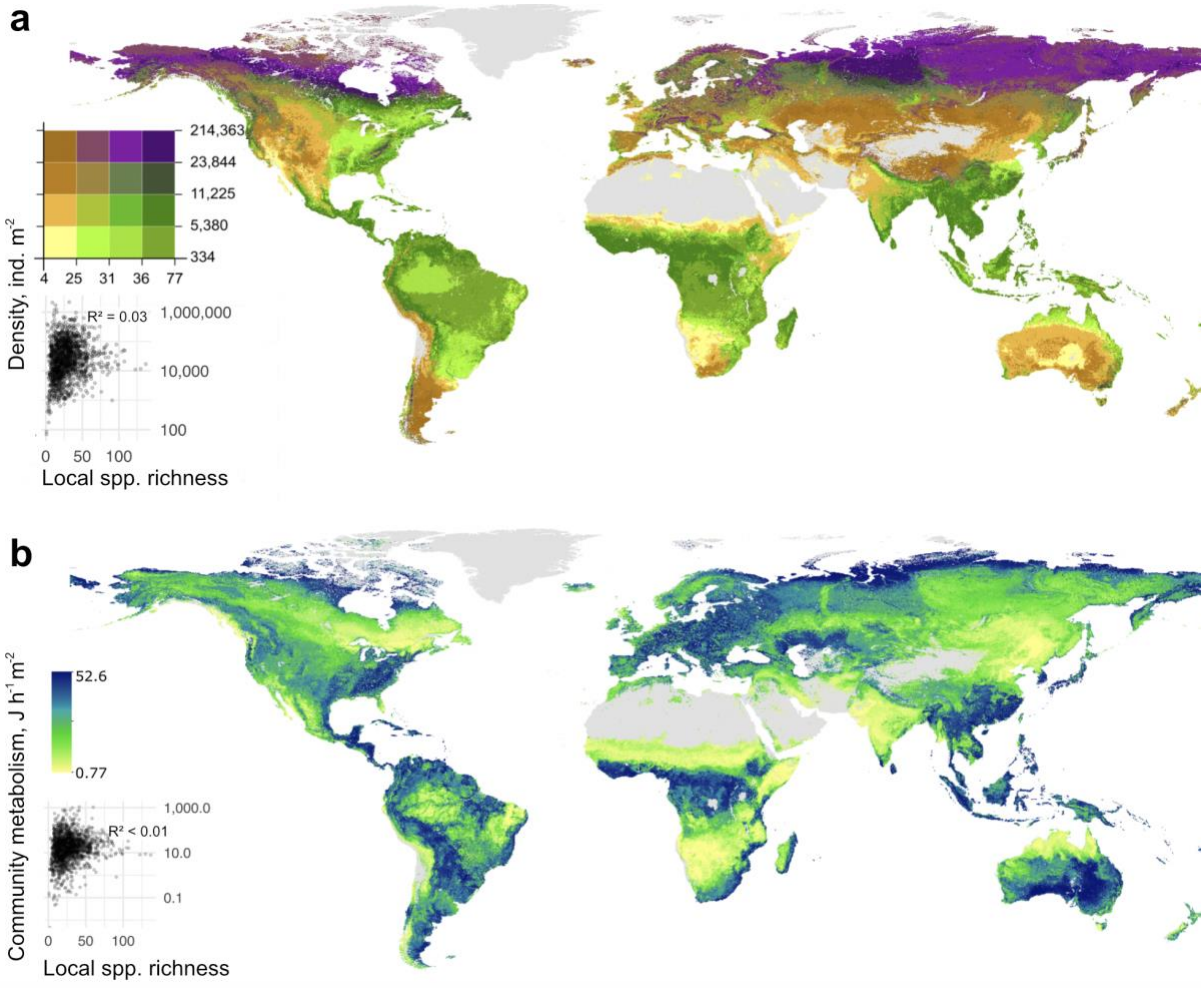
aboveground and aquatic taxa[25,26] and corroborates the hypothesized mismatch between above- and belowground biodiversity distributions[36]. This mismatch calls for explicit assessments of soil biodiversity hotspots for monitoring and conservation of soil organisms[15].

Global distribution and its drivers

To map the global distribution of springtail community metrics and uncover its drivers, we pre-selected climatic, vegetation, soil, topographic and anthropogenic variables with known ecological effects on springtails (Extended Data Fig. 9a). To perform a global extrapolation, we used 22 of the pre-selected variables that were globally available and applied a random forest algorithm to identify the strongest spatial associations of community parameters with environmental layers[18]. To reveal the key driving factors of springtail communities, we ran a path analysis with 12 non-collinear variables (Extended Data Fig. 9b). The European spatial clustering in our data distribution (Fig. 1a), was taken in consideration with a continental-scale validation in both analyses (see Methods).

At the global scale, species richness was not related to biomass (Pearson's $R^2 = 0.02$) or density (Pearson's $R^2 = 0.03$; Fig. 2a). Our extrapolations revealed at least five types of geographical areas with specific combinations of density and species richness patterns (Fig. 2a): (1) polar regions with very high densities and medium to high species richness such as the Arctic; (2) temperate regions with medium densities and high species richness such as mountainous and forested areas in Europe, Asia and North America; (3) temperate regions with medium to high densities but moderate species richness such as arid temperate biomes (e.g., dry grasslands); (4) temperate, subtropical and tropical arid ecosystems with low densities and species richness such as semi-deserts and other arid regions; (5) tropical areas with low densities but high species richness such as tropical forests and grasslands. Hotspots of springtail community metabolism were observed across a range of different latitudes (Fig.

253 2b), but were not associated with biodiversity hotspots (Pearson's $R^2 < 0.01$), emphasizing
254 that species richness is neither associated with higher density nor activity of springtail
255 communities at the global scale.



256

257 **Fig. 2 | Global maps overlapping modelled springtail density and local species richness**
258 **(a) and community metabolism (b) in soil.** In (a) colours distinguish areas with different
259 combinations of density and species richness, e.g., low density - low richness is given in
260 yellow and high density - high richness in violet. In (b) the colour gradient indicates
261 community metabolism, with potential hotspots shown in blue. All data were projected at the
262 30 arcsec (approximately 1 km^2) pixel scale. Pixels below the extrapolation threshold are
263 masked. Correlations between density or metabolism and species richness (inset graphs) are
264 based on site-level data.

Path analysis suggested that springtail density increases with latitude, NDVI (vegetation richness), aridity index and at high soil pH, but decreases with increasing mean annual temperature and elevation (Fig. 3). The positive global relationship of density with the aridity index was unexpected for physiologically moisture-dependent animals such as springtails[29], but was also observed in nematodes[18] and is probably due to the low amount of precipitation in circumpolar climates and very few data from desert sites. Density and biomass of springtails increased with precipitation within the tropical zone (Extended Data Fig. 8). Similar to patterns for earthworms[19], soil properties had less evident linear effects on springtail density than climate at the global scale. However, the relationships of density with soil pH and organic carbon content were hump-shaped, suggesting that intermediate values of these parameters are optimal for springtails (Extended Data Fig. 8), which is also observed for nematodes[18]. Existing evidence points to soil properties as key drivers of microfauna (nematodes)[6], climate as a key driver of macrofauna (earthworms)[7] and a combination of both as drivers of mesofauna (springtails) at the global scale.

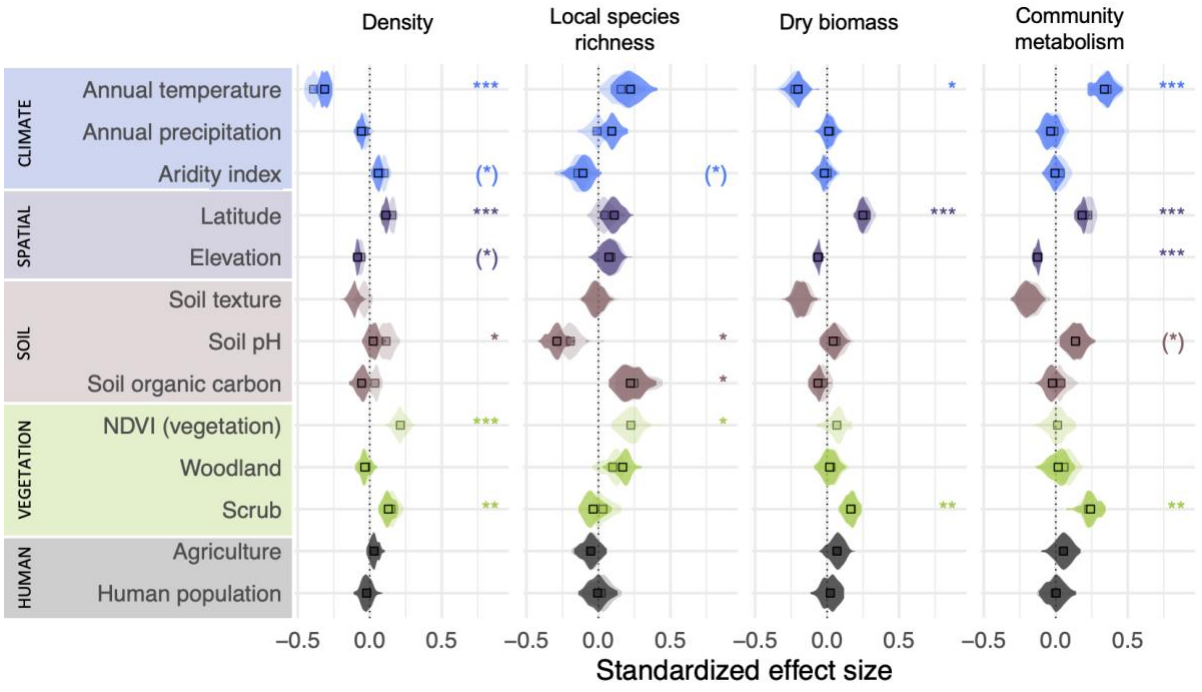


Fig. 3 | Environmental drivers of springtail communities at the global scale. Standardized effect sizes for direct (semi-transparent colour) and total (direct and indirect, solid colour) effects from path analysis are shown for density ($R^2 = 0.36 \pm 0.01$, n = 723 per iteration), local species richness ($R^2 = 0.20 \pm 0.02$, n = 352), biomass ($R^2 = 0.40 \pm 0.02$, n = 568) and community metabolism ($R^2 = 0.17 \pm 0.02$, n = 533). Mean values (squares) and data distribution (violins) are shown. Asterisks denote factors with a significant direct effect ($p < 0.05$) on a given springtail community metric for >25%$^{(*)}$, >50%*, >75%** and >95%*** of iterations.

Springtail density and biomass were lower in woodlands, grasslands and agricultural sites in comparison to scrub-dominated landscapes (Fig. 3). In contrast to previous global assessments of soil animal biodiversity[18,19], tundra was extensively sampled in our dataset (n = 253; Fig. 1a), and densities >1 million individuals per square meter were recorded at 12 independent sites. The high species richness of tundra communities (Fig. 2a), suggests a long evolutionary history of springtails in cold climates; indeed, they are currently the most taxonomically represented group of terrestrial arthropods in the Arctic[35] and the Antarctic[37]. Tundra remains under snow cover for most of the year, flourishing during summer when high springtail densities were recorded. During winter, springtails survive under the snow using remarkable adaptations to subzero temperatures (dehydration[38] and 'supercooling'[38]). Importantly, tundra soils contain a major proportion of the total soil organic matter and microbial biomass stored in the terrestrial biosphere[16]. As climate warming alters carbon cycling in the tundra[39], longer active periods of springtails could accelerate soil carbon release to the atmosphere in polar regions[40].

Across tropical ecosystems in the Amazon basin, equatorial Africa and Southeast Asia, low density and biomass of springtails were recorded and extrapolated (Fig. 2a, Extended Data

306 Figs. 4 and 6). Mesofauna in general have low abundances in tropical ecosystems, where the
307 litter layer is shallow and larger soil-associated invertebrates, such as earthworms, termites
308 and ants, play a more important role[24]. Our study supports this trend also found in recent
309 global assessments of other soil invertebrates[18,19,41]. However, considering the high mass-
310 specific metabolism of springtails and high predation rates in tropical communities[7,8,22], a
311 quantitative comparison of energy flows and stocks across latitudes and groups of soil fauna
312 is needed.
313 Interestingly, we found no pronounced influence of agriculture and human population on
314 springtail communities at the global scale; agriculture tended to have a positive impact on
315 biomass but a negative impact on species richness (Fig. 3). Agricultural sites had similar
316 springtail densities compared to woodlands and grasslands in the temperate zone (ca. 15-25k
317 individuals m^{-2}; Extended Data Fig. 3), which may be explained by large variation in
318 management within each of these habitat types. Some springtail species effectively survive in
319 agricultural fields[33], where they are involved in nutrient cycling and serve as biocontrol
320 agents by grazing on pathogenic fungi[42] and supporting arthropod predators[43]. Springtails are
321 also commonly found in urban areas[32]. However, the negative trend in species richness at
322 human-modified sites suggests that intensive land use may reduce springtail diversity, which
323 is indeed often recorded[32,33,44].
324 The only variable that was positively associated with both density and local species richness
325 of springtails, was NDVI (as a proxy for vegetation richness), reinforcing the close
326 connection between springtail communities and the vegetation[20]. Overall, high local species
327 richness was predicted in warm, acidic woodlands with high soil organic carbon stocks (Fig.
328 3) and geospatial extrapolation emphasized tropical regions and some boreal forests in North
329 America and Eurasia as springtail diversity hotspots (Extended Data Fig. 5). In our dataset,
330 sites with the highest extrapolated local species richness (i.e., >100 species) were located in

European woodlands (Czech Republic, Slovakia). However, this picture may be biased by the historical clustering of taxonomic expertise in Europe[5]. Outside Eurasia, species-rich sites (i.e. 60-80 species) were located in Vietnamese monsoon forests and some Brazilian rainforests, but 70-90% of species in tropical communities remain undescribed[45,46]. Thus, despite low springtail density, tropical forests contribute substantially to global springtail diversity but the full extent of this contribution is unknown.

Our extrapolations suggest that there are c. 2×10^{18} soil springtails globally and their total biomass comprises c. 29 Mt C (c. 200 Mt fresh weight), with respiration of c. 16 Mt C $month^{-1}$ (which is c. 0.2% of the global soil respiration[9]). Our biomass estimates are very similar to the global estimated biomass of nematodes (c. 31 Mt C[18]), but lower than that of earthworms (c. 200 Mt C[19]), and exceeding by far that of all wild terrestrial vertebrates (c. 9 Mt C)[6], demonstrating that springtails are among the most abundant and ubiquitous animals on Earth. Overall, our global dataset on soil springtail communities synthesized the work of soil zoologists across the globe. It presents another milestone towards understanding the functional composition of global soil biodiversity. Being highly abundant in polar regions and some human-modified landscapes, springtails are facing two main global change frontiers: warming in the polar regions, and land-use change and urbanization in temperate and tropical regions. While the global abundance and biomass of springtails may decline with climate warming in the coming decades, their global activity may remain unchanged. The global diversity of springtails will depend on the balance between anthropogenic transformations and conservation efforts of biomes worldwide.

References

1. FAO, ITPS, GSBI, SCBD & EC. *State of knowledge of soil biodiversity Status, challenges and potentialities, Report 2020.* (FAO, 2020). doi:10.4060/cb1928en.
2. Bardgett, R. D. & van der Putten, W. H. Belowground biodiversity and ecosystem functioning. *Nature* **515**, 505–511 (2014).
3. Rusek, J. Biodiversity of Collembola and their functional role in the ecosystem. *Biodiversity and Conservation* **7**, 1207–1219 (1998).
4. Hopkin, S. P. *Biology of springtails: (Insecta: Collembola)*. (Oxford Science Publications, 1997).
5. Potapov, A. *et al.* Towards a global synthesis of Collembola knowledge – challenges and potential solutions. *Soil Organisms* **92**, 161–188 (2020).
6. Bar-On, Y. M., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proceedings of the National Academy of Sciences* **115**, 6506–6511 (2018).
7. Rall, B. C. *et al.* Universal temperature and body-mass scaling of feeding rates. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 2923–2934 (2012).
8. Roslin, T. *et al.* Higher predation risk for insect prey at low latitudes and elevations. *Science* **356**, 742–744 (2017).
9. Huang, N. *et al.* Spatial and temporal variations in global soil respiration and their relationships with climate and land cover. *Science Advances* **6**, eabb8508 (2020).
10. Wall, D. H. *et al.* Global decomposition experiment shows soil animal impacts on decomposition are climate-dependent. *Global Change Biology* **14**, 2661–2677 (2008).
11. Handa, I. T. *et al.* Consequences of biodiversity loss for litter decomposition across biomes. *Nature* **509**, 218–221 (2014).
12. Wagg, C., Bender, S. F., Widmer, F. & van der Heijden, M. G. A. Soil biodiversity and soil community composition determine ecosystem multifunctionality. *Proceedings of the National Academy of Sciences* **111**, 5266–5270 (2014).
13. Delgado-Baquerizo, M. *et al.* Multiple elements of soil biodiversity drive ecosystem functions across biomes. *Nature Ecology & Evolution* **4**, 210–220 (2020).

14. Geisen, S., Wall, D. H. & van der Putten, W. H. Challenges andopportunities for soil biodiversity in the Anthropocene. *Current Biology* **29**, R1036–R1044 (2019).
15. Guerra, C. A. *et al.* Tracking, targeting, and conserving soil biodiversity. *Science* **371**, 239–241 (2021).
16. Crowther, T. W. *et al.* The global soil community and its influence on biogeochemistry. *Science* **365**, (2019).
17. Bahram, M. *et al.* Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237 (2018).
18. van den Hoogen, J. *et al.* Soil nematode abundance and functional group composition at a global scale. *Nature* (2019) doi:10.1038/s41586-019-1418-6.
19. Phillips, H. R. P. *et al.* Global distribution of earthworm diversity. *Science* **366**, 480–485 (2019).
20. Fujii, S., Saitoh, S. & Takeda, H. Effects of rhizospheres on the community composition of Collembola in a temperate forest. *Applied Soil Ecology* **83**, 109–115 (2014).
21. Filser, J. *et al.* Soil fauna: key to new carbon models. *Soil* **2**, 565–582 (2016).
22. Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. & West, G. B. Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789 (2004).
23. Hooper, D. U. *et al.* Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. *Ecological Monographs* **75**, 3–35 (2005).
24. Petersen, H. & Luxton, M. A comparative analysis of soil fauna populations and their role in decomposition processes. *Oikos* **39**, 288–388 (1982).
25. Hillebrand, H. On the Generality of the Latitudinal Diversity Gradient. *The American Naturalist* **163**, 192–211 (2004).
26. Kreft, H. & Jetz, W. Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences* **104**, 5925–5930 (2007).
27. Enquist, B. J., Kerkhoff, A. J., Huxman, T. E. & Economo, E. P. Adaptive differences in plant physiology and ecosystem paradoxes: insights from metabolic scaling theory. *Global Change Biology* **13**, 591–609 (2007).

28. Fierer, N., Strickland, M. S., Liptzin, D., Bradford, M. A. & Cleveland, C. C. Global patterns in belowground communities. *Ecology Letters* **12**, 1238–1249 (2009).
29. Kærsgaard, C. W., Holmstrup, M., Malte, H. & Bayley, M. The importance of cuticular permeability, osmolyte production and body size for the desiccation resistance of nine species of Collembola. *Journal of Insect Physiology* **50**, 5–15 (2004).
30. Janion-Scheepers, C. *et al.* Basal resistance enhances warming tolerance of alien over indigenous species across latitude. *Proceedings of the National Academy of Sciences* **115**, 145–150 (2018).
31. Peguero, G. *et al.* Fast attrition of springtail communities by experimental drought and richness–decomposition relationships across Europe. *Global Change Biology* **25**, 2727–2738 (2019).
32. Joimel, S. *et al.* Urban and industrial land uses have a higher soil biological quality than expected from physicochemical quality. *Science of The Total Environment* **584–585**, 614–621 (2017).
33. Filser, J., Mebes, K.-H., Winter, K., Lang, A. & Kampichler, C. Long-term dynamics and interrelationships of soil Collembola and microorganisms in an arable landscape following land use change. *Geoderma* **105**, 201–221 (2002).
34. Phillips, H. R. *et al.* Response to Comment on "Global distribution of earthworm diversity". *Science* **371**, (2021).
35. Babenko, A. B. The structure of springtail fauna (Collembola) of the Arctic. *Entomological Review* **85**, 878–890 (2005).
36. Cameron, E. K. *et al.* Global mismatches in aboveground and belowground biodiversity. *Conservation Biology* (2019).
37. Baird, H. P., Janion-Scheepers, C., Stevens, M. I., Leihy, R. I. & Chown, S. L. The ecological biogeography of indigenous and introduced Antarctic springtails. *Journal of Biogeography* **46**, 1959–1973 (2019).
38. Sørensen, J. G. & Holmstrup, M. Cryoprotective dehydration is widespread in Arctic springtails. *Journal of Insect Physiology* **57**, 1147–1153 (2011).
39. Box, J. E. *et al.* Key indicators of Arctic climate change: 1971–2017. *Environmental Research Letters* **14**, 045010 (2019).

40. Sørensen, L. I., Holmstrup, M., Maraldo, K., Christensen, S. & Christensen, B. Soil fauna communities and microbial respiration in high Arctic tundra soils at Zackenberg, Northeast Greenland. *Polar Biology* **29**, 189–195 (2006).
41. Johnston, A. S. A. & Sibly, R. M. Multiple environmental controls explain global patterns in soil animal communities. *Oecologia* **192**, 1047–1056 (2020).
42. Goncharov, A. A. *et al.* Detrital subsidy alters the soil invertebrate community and reduces infection of winter wheat seedlings by Fusarium wilt. *Applied Soil Ecology* **163**, 103914 (2021).
43. von Berg, K., Thies, C., Tscharntke, T. & Scheu, S. Changes in herbivore control in arable fields by detrital subsidies depend on predator species and vary in space. *Oecologia* **163**, 1033–1042 (2010).
44. Tsiafouli, M. A. *et al.* Intensive agriculture reduces soil biodiversity across Europe. *Global Change Biology* **21**, 973–985 (2015).
45. Shveenkova, Y. Springtail communities (Collembola, Hexapoda). in *Structure and Functions of Soil Communities of a Monsoon Tropical Forest (Cat Tien National Park, southern Vietnam)* (ed. Tiunov, A. V.) 131–147 (KMK Scientific Press, 2011).
46. Deharveng, L. & Bedos, A. Factors influencing diversity of soil Collembola in a tropical mountane forest (Doi Inthanon, Northern Thailand). in *Soil biota, nutrient cycling and farming systems* (eds. Paoletti, M. G., Foissner, W. & Coleman, D. C.) (Lewis Publishers, 1993).

Methods

Data reporting. The data underpinning this study is a compilation of existing datasets and therefore, no statistical methods were used to predetermine sample size, the experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. The measurements were taken from distinct samples, repeated measurements from the same sites were averaged.

Data acquisition. Data were primarily collected from individual archives of contributing co-authors. Both published and unpublished data were collected, using raw data whenever possible entered into a common template. In addition, data available from Edaphobase[47] was included. The following minimum set of variables was collected: collectors, collection method (including sampling area and depth), extraction method, identification precision and resources, collection date, latitude and longitude, vegetation type (generalized as grassland, scrub, woodland, agriculture and 'other' for the analysis), and abundances of springtail taxa found in each soil sample (or sampling site). Underrepresented geographical areas (Africa, South America, Australia and Southeast Asia) were specifically targeted by a literature search in the Web of Science database using the keywords 'springtail' or 'Collembola', 'density' or 'abundance' or 'diversity', and the region of interest; data were acquired from all found papers if the minimum information listed above was provided. In total, 363 datasets comprising 2,783 sites were collected and collated into a single dataset (Extended Data Fig. 1).

Calculation of community parameters. Community parameters were calculated at the site level. Here, we defined a site as a locality that hosts a defined springtail community, is covered by a certain vegetation type and has a maximum spatial extent (diameter) of several hundred meters, making species co-occurrence and interactions plausible. To calculate density, numerical abundance in all samples was averaged and recalculated per square meter using the sampling area. Springtail communities were assessed predominantly during active vegetation periods (i.e., spring, summer and autumn in temperate and boreal biomes, and summer in polar biomes). Our estimations of community parameters therefore refer to the most favourable conditions (peak yearly densities). This seasonal sampling bias is likely to have little effect on our conclusions, since most springtails survive during cold periods[38,48].

Finally, we used mean annual temperatures[49] to estimate the seasonal mean community metabolism (described below).

All data analyses were conducted in R v. 4.0.2[50] with RStudio interface v. 1.4.1103 (RStudio, PBC), unless otherwise mentioned. To calculate local species richness, we used data identified to species or morphospecies level. Since the sampling effort varied among studies, we extrapolated species richness using rarefaction curves based on individual samples with the Chao estimator[51] in the vegan package[52]. For some sites, sample-level data were not available in the original publications, but an extensive sampling effort was made. In such cases, we predicted extrapolated species richness based on the completeness (ratio of observed to extrapolated richness) recorded at sites where sample-level data were available (only sites with 5 or more samples were used for the prediction). We built a binomial model to predict completeness in sites where no sample-level data were available (435 sites in Europe, 15 in Australia, 6 in South America, 4 in Asia, and 3 in Africa) using latitude and the number of samples taken at a site as predictors.

To calculate biomass, we first cross-checked all taxonomic names with the collembola.org checklist[53] using fuzzy matching algorithms (*fuzzyjoin* R package[54]) to align taxonomic names and correct typos. Then we merged taxonomic names with a dataset on body lengths compiled from the BETSI database[55], a personal database of Matty P. Berg, and additional expert contributions. We used average body lengths for the genus level (body size data on 432 genera) since data at the species level were not available for many species and morphospecies, and species within most springtail genera had similar body size ranges. Dry and fresh body masses were calculated from body length using a set of group-specific length-mass regressions (Extended Data Table 1)[56,57] and the results of different regressions applied to the same morphogroup were averaged. Dry mass was recalculated to fresh mass using corresponding group-specific coefficients[56]. We used fresh mass to calculate individual

metabolic rates[58] and account for the mean annual topsoil (0-5 cm) temperature at a given site[59]. Group-specific metabolic coefficients for insects (including Collembola) were used for the calculation: normalization factor (i0) ln(21.972) [J h^{-1}], allometric exponent (a) 0.759, and activation energy (E) 0.657 [eV][58]. Community-weighted (specimen-based) mean individual dry masses and metabolic rates were calculated for each sample and then averaged by site after excluding 10% of maximum and minimum values as outlier samples with small sampling areas, which have a high probability of randomly including large individuals. To calculate site-level biomasses and community metabolism, we summed masses or metabolic rates of individuals, averaged them across samples, and recalculated them per unit area (m^2).

Parameter uncertainties. Our biomass and community metabolism approximations contain several assumptions and ignore latitudinal variation in body sizes within taxonomic groups[60]. Nevertheless, latitudinal differences in springtail density (30-fold), environmental temperature (from -17.0 to +27.6°C), and genus-level community compositions (there are only few common genera among polar regions and the tropics)[53] are higher than the uncertainties introduced by indirect parameter estimations, which allowed us to detect global trends. Although most springtails are concentrated in the litter and uppermost soil layers[24], their vertical distribution depends on the particular ecosystem[61]. Since sampling methods are usually ecosystem-specific (i.e. sampling is done deeper in soils with developed organic layers), we treated the methods used by the original data collectors as representative of a given ecosystem. Under this assumption, we might have underestimated the number of springtails in soils with deep organic horizons, so our global estimates are conservative and we would expect true global density and biomass to be slightly higher. To minimize these effects, we excluded sites where the estimations were likely to be unreliable (see data selection below).

Data selection. Only data collection methods allowing for area-based recalculation (e.g. Tullgren or Berlese funnels) were used for analysis. Data from artificial habitats, coastal ecosystems, caves, canopies, snow surfaces, and strong experimental manipulations beyond the bounds of naturally occurring conditions were excluded (Extended Data Fig. 1). To ensure data quality, we performed a two-step quality check: technical selection and expert evaluation. Collected data varied according to collection protocols, such as sampling depth and the microhabitats (layers) considered. To technically exclude unreliable density estimations, we explored data with a number of diagnostic graphs (see Supplementary Data Cleaning Protocol) and filtered it, excluding the following: (1) All woodlands where only soil or only litter was considered; (2) All scrub ecosystems where only ground cover (litter or mosses) was considered; (3) Agricultural sites in temperate zones where only soil with sampling depth <10 cm was considered. Additionally, 10% of the lowest values were individually checked and excluded if density was unrealistically low for the given ecosystem (outliers with density over three times lower than 1% percentile within each ecosystem type). In total, 237 sites were excluded from density, and 394 sites from biomass, and community metabolism analyses based on these criteria. For the local species richness estimates, we removed all extrapolations based on sites with fewer than three samples and no (morpho)species identifications (647 sites; Extended Data Fig. 1).

Data expert evaluation. We performed manual expert evaluation of every contributed dataset. Evaluation was done by an expert board of springtail specialists, each with extensive research experience in a certain geographic area. Each dataset was scored separately for density and species richness as either trustworthy, acceptable, or unreliable. Density estimation quality was assessed using information about the sampling and extraction method and the density estimation itself. Species richness estimation quality was assessed using information about the identification key, experience of the person who identified the material,

557 species (taxa) list, and the species richness estimation itself. Based on the expert opinions,
558 unreliable estimates of density (together with biomass and community metabolism) and
559 species richness were excluded (Extended Data Fig. 1). The resulting final dataset included
560 2,470 sites and 43,601 samples[62] with a median of six samples collected at each site. The
561 dataset comprised 2,210 sites with density estimation (69 - 2,181,600 individuals m^{-2}), 2,053
562 sites with mean fresh body mass (1.8 - 3,110 µg), mean metabolic rate (0.028 - 2.4 mJ h^{-1}),
563 dry biomass (0.5 - 92,943 mg m^{-2}), fresh biomass (1.6 - 277,028 mg m^{-2}) and community
564 metabolism estimations (0.03 - 999.68 J h^{-1}), and 1,735 sites with local species richness
565 estimation (1 - 136.7 species; Extended Data Figs. 1 and 2).
566 **Data transformation.** All parameters except for extrapolated local species richness were
567 highly skewed (e.g., density had a global median of 21,016 individuals m^{-2} and a mean of
568 60,454 individuals m^{-2}) and we applied $\log_{10}$-transformation prior to analysis. This greatly
569 improved the fit of all statistical analyses.
570 **Latitudinal and ecosystem trends.** To explore changes in springtail communities with
571 latitude, we sliced the global latitudinal gradient into 5-degree bins and calculated average
572 parameters across sites in each bin after trimming to ensure the same statistical weight for
573 each latitudinal bin while plotting the gradient. The latitudinal gradient was plotted with
574 *ggplot2*[63], and quadratic smoothers were used to illustrate trends. Mean parameters of
575 springtail communities were compared across ecosystem types using a linear model and
576 multiple comparisons with the Tukey HSD test using *HSD.test* in the *agricolae* package[64].
577 Habitats were classified according to the vegetation types. Climates were classified as polar
578 (beyond the polar circles, i.e., more than 66.5 and less than -66.5 degrees), temperate (from
579 the polar circles to the tropics of Capricorn/Cancer, i.e. to 23.5 and -23.5 degrees) and
580 tropical (in between 23.5 and -23.5 degrees). Habitats and climates were combined to
581 produce ecosystem types. For the analysis, only well-represented ecosystem types were

582 retained: polar scrub (n = 253), polar grassland (n = 39), polar woodland (n = 28), temperate
583 woodland (n = 907), temperate scrub (n = 104), temperate grassland (n = 445), temperate
584 agriculture (n = 374), tropical agriculture (n = 68) and tropical forest (n = 141; Extended Data
585 Fig. 3).

586 **Selection of environmental predictors**. To assess the drivers of global distributions of
587 springtail community metrics, we pre-selected variables with a known ecological effect on
588 springtail communities (based on expert opinions) and constructed a hypothetical relationship
589 diagram (Extended Data Fig. 9a). Environmental data were very heterogeneous across the
590 springtail studies, so we used globally available climatic and other environmental layers;
591 these included layers bearing climatic (mean annual temperature, temperature seasonality,
592 temperature annual range, mean annual precipitation, precipitation seasonality, precipitation
593 of the driest quarter[65], aridity index[66]), topographic (elevation, roughness[67]), vegetative and
594 land cover (aboveground biomass[68], tree cover[69], Net Primary Production, Normalized
595 Difference Vegetation Index [NDVI][70]), topsoil physicochemical (0-15 cm depth C to N
596 ratio, pH, clay, sand, coarse fragments, organic carbon, bulk density[71]) and human population
597 density[72].

598 **Geospatial global projections.** To create global spatial predictions of springtail density,
599 species richness, biomass, and community metabolism, we followed the approach previously
600 used for nematodes[18,73] that is based on spatial associations of community parameters with
601 global environmental information. A Random Forest algorithm was applied to identify the
602 spatial associations and extrapolate local observations to the global scale[18,73]. After retrieving
603 the environmental variable values for each location, we trained 18 model versions, each with
604 different hyperparameter settings, i.e., variables per split (range: 2 - 7); minimum leaf
605 population (range: 3 - 5). To minimize the potential bias of a single model, we used an
606 ensemble of the top 10 best-performing models, selected based on the coefficient of

determination (R^2), to create global predictions of each of the community parameters. Geographical regions with climatic conditions poorly represented by our sites and without NPP data were excluded from the extrapolation (e.g., Sahara, Arabian desert, Himalayas). We evaluated our extrapolation quality based on spatial approximations of interpolation versus extrapolation[73]. In this approach, we first determined the range of environmental conditions represented by the observations. Next, we classified all pixels to fall within or outside the training space, in univariate and multivariate space. For the latter, we first transformed the data into principal component space, and selected the first 11 PC axes, collectively explaining 90% of the variation. Finally, we classified pixels to fall within or outside the convex hulls drawn around each possible bivariate combination of these 11 PC axes; pixels that fell outside the convex hulls in >90% of cases were masked on the map.

To estimate spatial variability of our predictions while accounting for the spatial sampling bias in our data (Fig. 1a) we performed a spatially stratified bootstrapping procedure. We used the relative area of each IPBES[74] region (i.e., Europe and Central Asia, Asia and the Pacific, Africa, and the Americas) to resample the original dataset, creating 100 bootstrap resamples. Each of these resamples was used to create a global map, which was then reduced to create mean, standard deviation, 95% confidence interval, and coefficient of variation maps (Extended Data Figs. 4-7).

Global biomass, abundance, and community metabolism of springtails were estimated by summing predicted values for each 30 arcsec pixel[18]. Global community metabolism was recalculated from joule to mass carbon by assuming 1 kg fresh mass = 7×10^6 J[75], an average water proportion in springtails of 70%[56], and an average carbon concentration of 45% (calculated from 225 measurements across temperate forest ecosystems)[76].

Path analysis. To reveal the drivers of springtail communities at the global scale, we performed a path analysis. After filtering the selected environmental variables (see above)

according to their global availability and collinearity, 13 variables were used (Extended Data Fig. 9b): mean annual temperature, mean annual precipitation (CHELSA database[65]), aridity (CGIAR database[66]), soil pH, sand and clay contents combined (sand and clay contents were co-linear in our dataset), soil organic carbon content (SoilGrids database[71]), NDVI (MODIS database[70]), human population density (GPWv4 database[72]), latitude, elevation[67], and vegetation cover (woodland, scrub, or agriculture; grasslands were represented as the combination of woodland, scrub, and agriculture absent). Before running the analysis, we performed the Rosner's generalized extreme Studentized deviate test in the *EnvStats* package[77] to exclude extreme outliers and we z-standardized all variables (Supplementary R Code).

Separate piecewise structural equation models were run to predict density, dry biomass, community metabolism, and local species richness in the *lavaan* package[78]. To account for the spatial clustering of our data in Europe, instead of running a model for the entire dataset, we divided the data by the IPBES[74] geographical regions and selected a random subset of sites for Eurasia, such that only twice the number of sites were included in the model as the second most represented region. We ran the path analysis 99 times for each community parameter with different Eurasian subsets (density had n = 723 per iteration, local species richness had n = 352, dry biomass had n = 568, and community metabolism had n = 533). We decided to keep the share of the Eurasian dataset larger than other regions to increase the number of sites per iteration and validity of the models. The Eurasian dataset also had the best data quality among all regions and a substantial reduction in datasets from Eurasia would result in a low weight for high quality data. We additionally ran a set of models in which the Eurasian dataset was represented by the same number of sites as the second-most represented region, which yielded similar effect directions for all factors, but slightly higher variations and fewer consistently significant effects. In the paper, only the first version of analysis is

presented. To illustrate the results, we averaged effect sizes for the paths across all iterations and presented the distribution of these effect sizes using mirrored Kernel density estimation (violin) plots. We marked and discussed effects that were significant at $p < 0.05$ in more than a given number of iterations (arbitrary thresholds were set to 25%, 50%, 75% and 95% of iterations; Fig. 3).

Data availability statement.

The data that support the findings of this study are available under CC-BY 4.0 license from Figshare: https://doi.org/10.6084/m9.figshare.16850419; high-resolution maps can be assessed at https://doi.org/10.6084/m9.figshare.16850446.

Code availability statement

Programming code for the path analysis and the geospatial modelling is available under CC-BY 4.0 from Figshare: https://doi.org/10.6084/m9.figshare.16850419.

Methods references

47. Burkhardt, U. *et al.* The Edaphobase project of GBIF-Germany—A new online soil-zoological data warehouse. *Applied Soil Ecology* **83**, 3–12 (2014).
48. Sømme, L. Supercooling and winter survival in terrestrial arthropods. *Comparative Biochemistry and Physiology Part A: Physiology* **73**, 519–543 (1982).
49. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**, 4302–4315 (2017).
50. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2019).

681 51. Chao, A. & Jost, L. Coverage-based rarefaction and extrapolation: standardizing samples by
682 completeness rather than size. *Ecology* **93**, 2533–2547 (2012).
683 52. Oksanen, J. *et al.* Vegan: community ecology package. R package version 1.17-6. (2011).
684 53. Bellinger, P. F., Christiansen, K. A. & Janssens, F. Checklist of the Collembola of the World.
685 http://www.collembola.org/ (2020).
686 54. Robinson, D. *fuzzyjoin: Join Tables Together on Inexact Matching*. (2020).
687 55. Pey, B. *et al.* A Thesaurus for Soil Invertebrate Trait-Based Approaches. *PLoS ONE* **9**, e108985
688 (2014).
689 56. Petersen, H. Estimation of dry weight, fresh weight, and calorific content of various Collembolan
690 species. *Pedobiologia* **15**, 222–243 (1975).
691 57. Tanaka, M. Ecological studies on communities of soil Collembola in Mt. Sobo, southwest Japan.
692 *Japanese Journal of Ecology* **20**, 102–110 (1970).
693 58. Ehnes, R. B., Rall, B. C. & Brose, U. Phylogenetic grouping, curvature and metabolic scaling in
694 terrestrial invertebrates: Invertebrate metabolism. *Ecology Letters* **14**, 993–1000 (2011).
695 59. Lembrechts, J. *et al. Global maps of soil temperature*. https://osf.io/pksqw (2021)
696 doi:10.32942/osf.io/pksqw.
697 60. Bonfanti, J. *et al.* Intraspecific body size variability in soil organisms at a European scale:
698 implications for functional biogeography. *Functional ecology* **32**, 2562–2570 (2018).
699 61. Potapov, A. M. *et al.* Arthropods in the subsoil: Abundance and vertical distribution as related to
700 soil organic matter, microbial biomass and plant roots. *European Journal of Soil Biology* **82**, 88–
701 97 (2017).
702 62. Potapov, A. M., Guerra, C. A. & van den Hoogen, J. #GlobalCollembola: site-level database and
703 analyses. *Figshare.* Dataset. https://doi.org/10.6084/m9.figshare.16850419.v1
704 63. Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer-Verlag, 2009).
705 64. Mendiburu, F. de. *agricolae: Statistical Procedures for Agricultural Research*. (2020).
706 65. Karger, D. N. *et al.* Climatologies at high resolution for the earth's land surface areas. *Scientific*
707 *Data* **4**, 170122 (2017).

66. Zomer, R. J., Trabucco, A., Bossio, D. A. & Verchot, L. V. Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agriculture, Ecosystems & Environment* **126**, 67–80 (2008).

67. Amatulli, G. *et al.* A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific Data* **5**, 180040 (2018).

68. Santoro, M. GlobBiomass - global datasets of forest biomass. (2018) doi:10.1594/PANGAEA.894711.

69. Hansen, M. C. *et al.* High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**, 850–853 (2013).

70. Tuanmu, M.-N. & Jetz, W. A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Global Ecology and Biogeography* **23**, 1031–1045 (2014).

71. Hengl, T. *et al.* SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE* **12**, e0169748 (2017).

72. Center for International Earth Science Information Network - CIESIN - Columbia University. Gridded Population of the World, Version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision UN WPP Country Totals. (2016).

73. van den Hoogen, J. *et al. A geospatial mapping pipeline for ecologists*. http://biorxiv.org/lookup/doi/10.1101/2021.07.07.451145 (2021) doi:10.1101/2021.07.07.451145.

74. IPBES. *Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. (2019).

75. Peters, R. H. *The ecological implications of body size*. (Cambridge University Press, 1983).

76. Potapov, A. M., Tiunov, A. V. & Scheu, S. Uncovering trophic positions and food resources of soil animals using bulk natural stable isotope composition. *Biological Reviews* **94**, 37–59 (2019).

77. Millard, S. P. *EnvStats: An R Package for Environmental Statistics*. (Springer, 2013).

78. Rosseel, Y. lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software* **48**, 1–36 (2012).

Acknowledgements

The article is an outcome of the #GlobalCollembola community initiative that is voluntarily supported by researchers around the world. Data collection and analysis was supported by the Russian Science Foundation (19-74-00154 to A.P.) and by Deutsche Forschungsgemeinschaft (SFB990-EFForTS to S.S.). The following funding bodies provided support for individual contributors: ARC SRIEAS Grant SR200100005 Securing Antarctica's Environmental Future to S.L.C., Slovak Scientific Grant Agency VEGA 1/0346/18 to Ľ.K., RFBR 19-516-60002 to N.A.K., Carl Tryggers Stiftelse för Vetenskaplig Forskning and Qatar Petroleum to J.M.A., BIO 27 (2013-2014)-MAGyP and PICTO 2084 (2012)-ANPCyT to V.B., DAAD-19-10 and MSM200962001 to T.C., grant TE, PN-III-P1-1.1-TE-2019-0358 to C.F., NWO grant 821.01.015 to O.F., National Natural Sciences Foundation of China No 41471037 and 41871042 to M.G., BIO 27 (2013-2014), MAGyP; PICT 2084 (2012), FONCyT to D.F.G., NRF South African National Antarctic Programme grant 110734 to M.G., Natural Resources Canada (NRCan), EcoEnergy Innovation Initiative under the Office of Energy Research and Development, and the Natural Sciences and Engineering Research Council of Canada (NSERC) to I.T.H., L.A.V. and L.R., Independent Research Fund Denmark grant no. DFF-4002-00384 to M.H., Estonian Science Foundation G9145 to M.I., SA-France bilateral grant to C.J., SA (NRF) / Russia (RFBR) Joint Science and Technology Research Collaboration project no. 19-516-60002 (FRBR) and no. 118904 (NRF) to M.P. and C.J., European Research Council (ERC), European Union's Horizon 2020 research and innovation programme (grant agreement no. 677232; to N.E.); iDiv, German Research Foundation (DFG–FZT 118, 202548816) to M.J. and N.E., French National Agency of Research (ANR) (JASSUR research project; ANR-12-VBDU-0011), «Ministère de l'Agriculture et de la Pêche» and «Ministère de l'Education Nationale de la Recherche et de la Technologie» (ACTA programme), «Ministère de l'Aménagement du Territoire et de l'Environnement» (Pnetox programme), EU-funded project, ECOGEN QLK5-CT-2002-01666 (www.ecogen.dk), "Agence de l'Environnement et de la Maîtrise de l'Énergie" (BIOINDICATEUR 2, BIOTECHNOSOL), ANDRA and GISFI (www.gisfi.fr) to S.J., GRR SER-BIODIV (Région Normandie, France) to MCha, ESF9258, B02 to A.K., Fundamental Research Funds for the Central Universities (grant no. 2018CDXYCH0014) to D.L., DFG 316045089 to J.L., Massey University Research Fund grant to M.A.M., DFG SCHE 376/38-2 to M.M.P., grant from the Austria Academy of Science: Heritage_2020-043_Modeling-Museum to P.Q., Slovak Scientific Grant Agency: VEGA Nos. 1/0441/03 and 1/3267/06 to N.R., Higher Education Commission of Pakistan to M.I.R., RSF 21-74-00126 to R.A.S., Austrian Federal Government and European Union (Rural Development 2014-

765 2020) to J.S., AAAA-A17-117112850235-2 to A.A.T., Brazilian Council for Scientific and Technological
766 Development - CNPq (grant no. 152717/2016-1) to B.R.W., 309030/2018-8 to D.Z. and 305426/2018-4 to
767 B.C.B., National Natural Science Foundation of China (31970434, 31772491) to N.N.G., Research and
768 Innovation Support Foundation of Santa Catarina (FAPESC) (6.309/2011-6/FAPESC) and the CNPq
769 (563251/2010-7/CNPq) to L.C.I.O.F., O.K.-F., the Latvian Council of Science Grants no. 90.108, 93.140,
770 96.0110, 01.0344 to E.J., CNPq for the Research Productivity Grant (305939/2018-1) to D.B., FPI-MICINN
771 grant in the project INTERCAPA (CGL2014-56739-R) to P.H. Authors are grateful to Penelope Greenslade for
772 providing the literature on Australian Collembola communities. Authors are grateful to Frans Janssens for
773 providing the global checklist of Collembola.

774 Author contributions

775 A.M.P. designed the study, coordinated the collection, cleaning and standardization of data and wrote the first
776 draft of the manuscript. C.G. and A.M.P. designed and performed the path analysis. J.v.d.H. designed and
777 performed the geospatial modelling. A.B., B.C.B., L.D., Ľ.K., N.A.K., J.F.P. and M.B.Pot. evaluated the data
778 quality. M.P.B., S.L.C., J.F.P., D.J.R., T.C., N.E., S.S., M.Cha., J.F. and I.T.H. contributed to writing and
779 conceptualisation of the manuscript. A.M.P., A.B., M.P.B., S.L.C., L.D., Ľ.K., N.A.K., J.F.P., M.B.Pot., D.J.R.,
780 J.M.A., J.I.A., I.B., V.B., S.B., T.B., G.C., M.Cha., T.Che., M.C., A.T.C., J.C., P.Č., A.M.d.l.P, A.D., S.D.F.,
781 C.F., J.F., O.F., S.F., E.G., M.G., B.G., D.F.G., M.Gre., I.T.H., C.H., M.H., P.H., M.I., C.J., M.J., S.J., B.S.J.,
782 E.J., L.C.I.O.F., O.K.-F., D.B., E.J.K., A.K., E.A.L., D.L., J.L., M.J.L., M.T.M., M.M.M., M.A.M., T.N., I.N.,
783 R.O., J.G.P., M.M.P., P.Q., N.R., M.I.R., L.J.R., L.R., R.A.S., S.Sal., E.J.S., N.S., C.S., J.S., Y.B.S., S.K.S.,
784 M.S., X.S., W.I.S., A.A.T., M.P.T., M.A.T., M.S.T., M.N.T., A.V.U., L.A.V., L.A.W., B.R.W., D.W., D.Wu.,
785 Z.J.X., R.Y., D.Z., N.E. and S.S. contributed data. All authors contributed to editing of the paper.
786

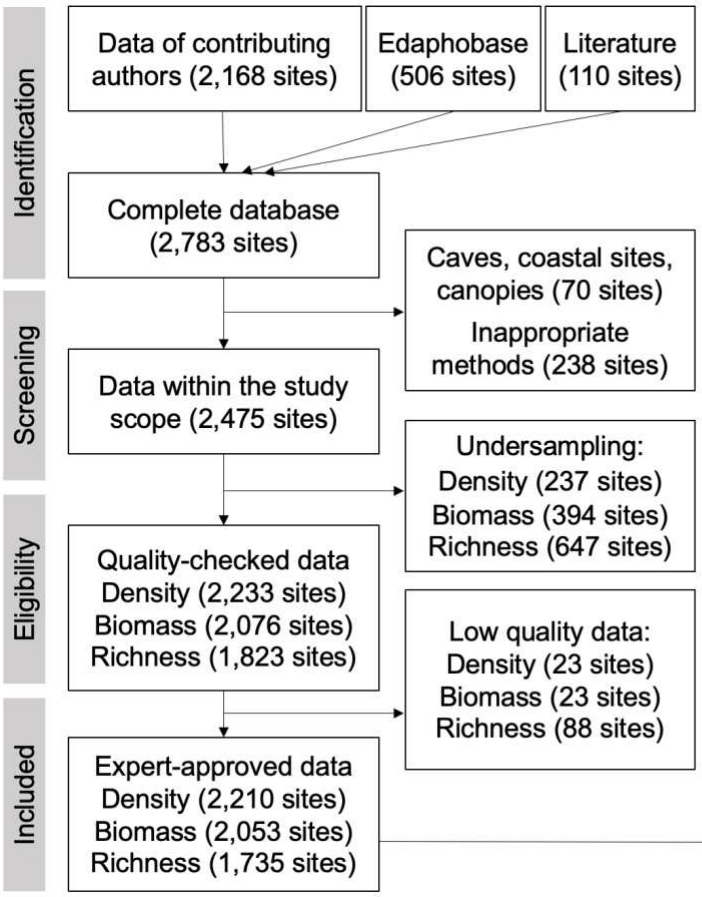
787 **Competing interests**. Authors have no competing interests to declare.
788

789 Supplementary Information is available for this paper.
790

791 **Materials & Correspondence**. Correspondence and requests for materials should be

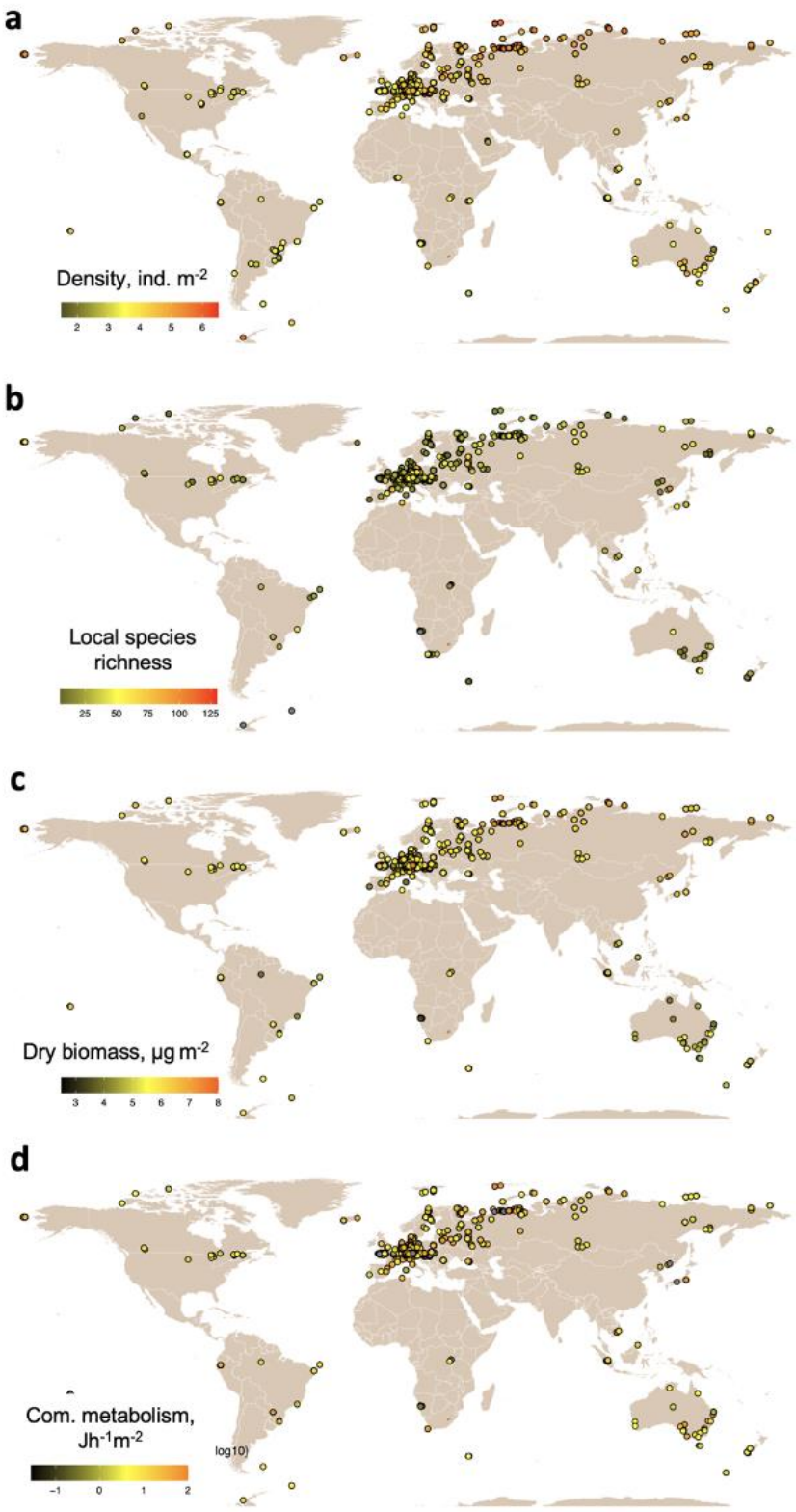
792 addressed to A.M.P.

Extended data



Major data providers	Sites	Samples
Edaphobase	488	77
Anatoly Babenko	332	2,667
Sophie Joimel, Angélique Renaud, Céline Pernin and Jérôme Cortet	139	4,407
Ľubomír Kováč, Natália Raschmanová and Peter Čuchta (collaborator: Juha Alatalo)	136	3,239
Natalia Kuznetsova (collaborators: Mikhail Potapov, Anna Bokova, Anna Saraeva)	115	10,507
Stef Bokhorst	84	893
Daniel Winkler	84	420
Pascal Querner	63	782
Jean-François Ponge	32	1,118
Viesturs Melecis and Edite Jucevica	1	3,300
Anton Potapov (collaborators: Anastasiya Stryuchkova, Stefan Scheu, Winda Ika Susanti) includes literature data	176	1,955

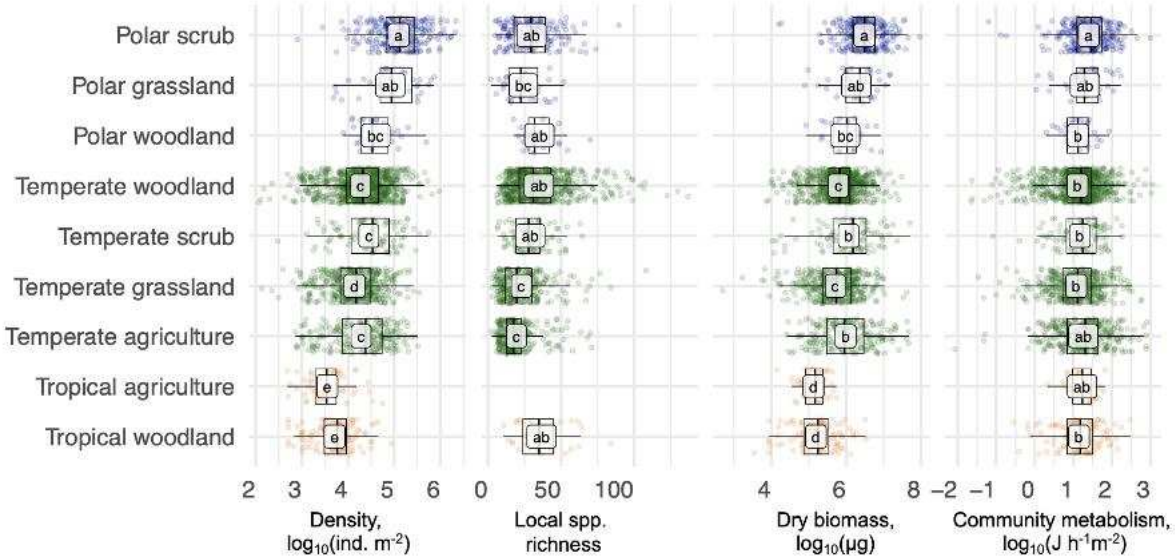
Extended Data Fig. 1 | Flow diagram of data compilation and selection. Major data providers of #GlobalCollembola whose data were used in the analysis are given in the shaded table on the right side. Providers are ordered based on the number of sites, but exemplar datasets with extensive sampling efforts (number of samples) are given to illustrate the available data.



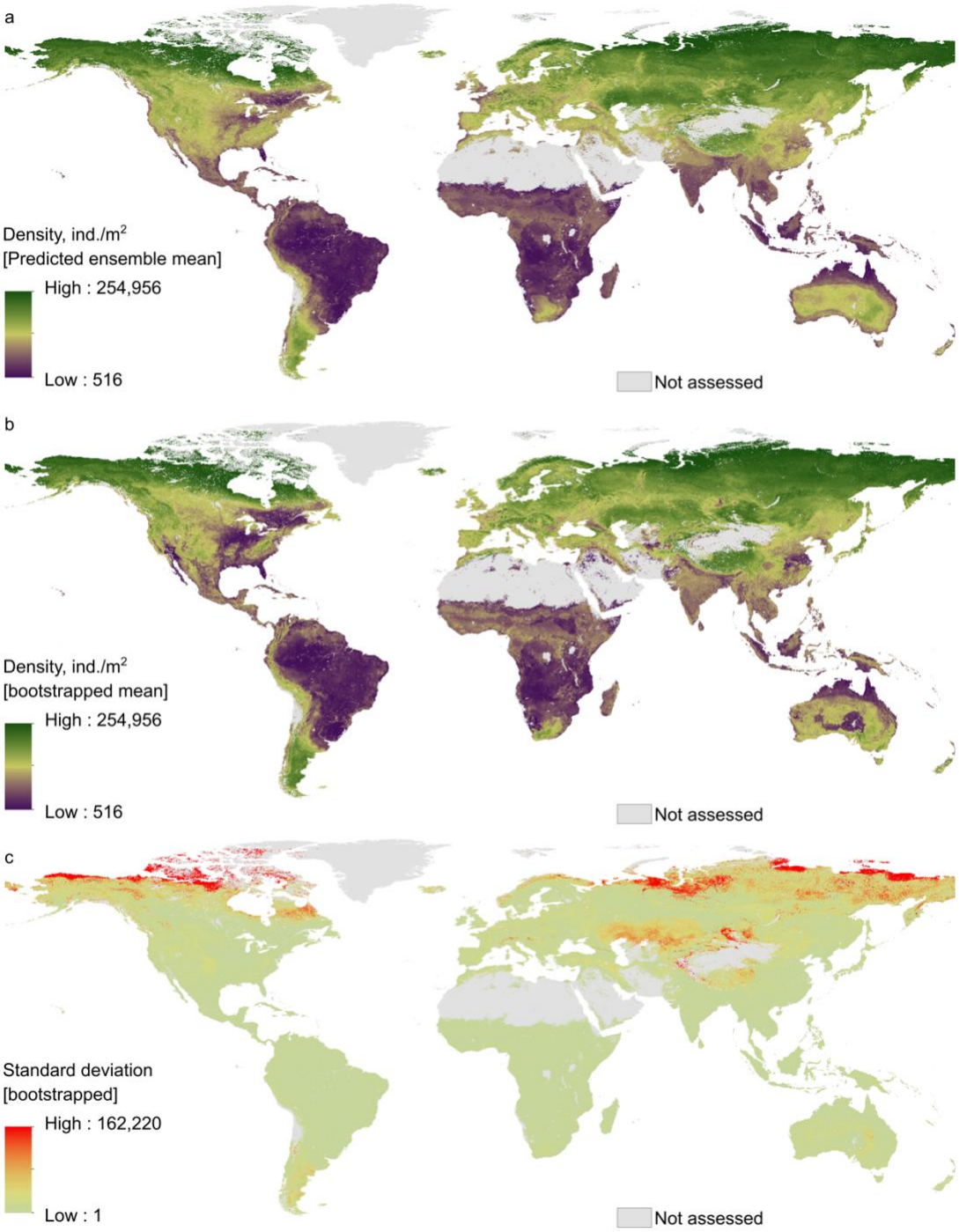
Extended Data Fig. 2 | Selected sampling sites that were used in the analysis. **a**, Density (n = 2210), **b**, Local species richness (n = 1735); **c**, Dry biomass (n = 2053); **d**, Community metabolism (n = 2053). Data scales are logarithmic except for local species richness.

Extended Data Table 1 | Regression coefficients used to estimate the dry and fresh body masses of springtail genera based on body lengths. For each genus, the average body mass (*M*) [µg dry weight] was calculated from the average body length (*L*) [mm] using the power equation: $M = a*L\hat{}b$, where *a* is the normalisation coefficient and *b* is the exponent. Abdomen length of Symphypleona was used in the original equations and was assumed to be 0.83 of the total body length. Two sets of coefficients coming from two independent studies[56,57] were used for each morphogroup (a_1, b_1 and a_2, b_2) and the two estimates of dry body mass were averaged. Fresh body mass was calculated from the resulting average by dividing it by the proportion of the dry weight.

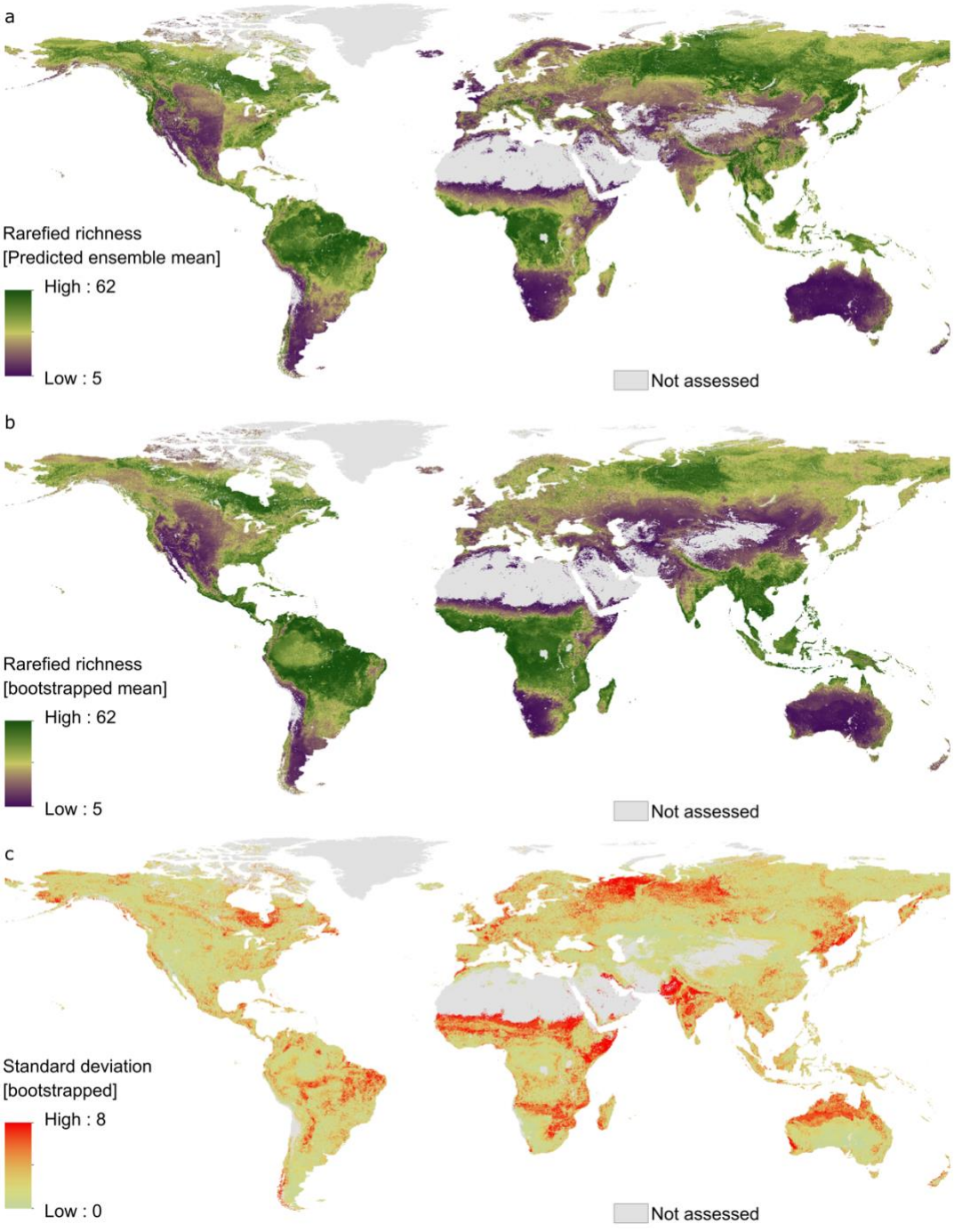
Morphogroup	Normalisation (a_1)	Exponent (b_1)	Normalisation (a_2)	Exponent (b_2)	Dry weight proportion
Entomobryidae	11.749	2.52	14.256	2.708	0.30
Isotomidae (small)	6.457	2.99	5.623	2.799	0.36
Isotomidae (large)	5.623	3.28	8.427	3.223	0.36
Onychiuridae	4.266	2.75	5.598	2.769	0.30
Poduromorpha (excl. Onychiuridae)	9.772	2.55	5.598	2.769	0.30
Symphypleona	190.546	3.627	39.628	3.796	0.21
Tomoceridae	9.204	2.744	14.256	2.708	0.25



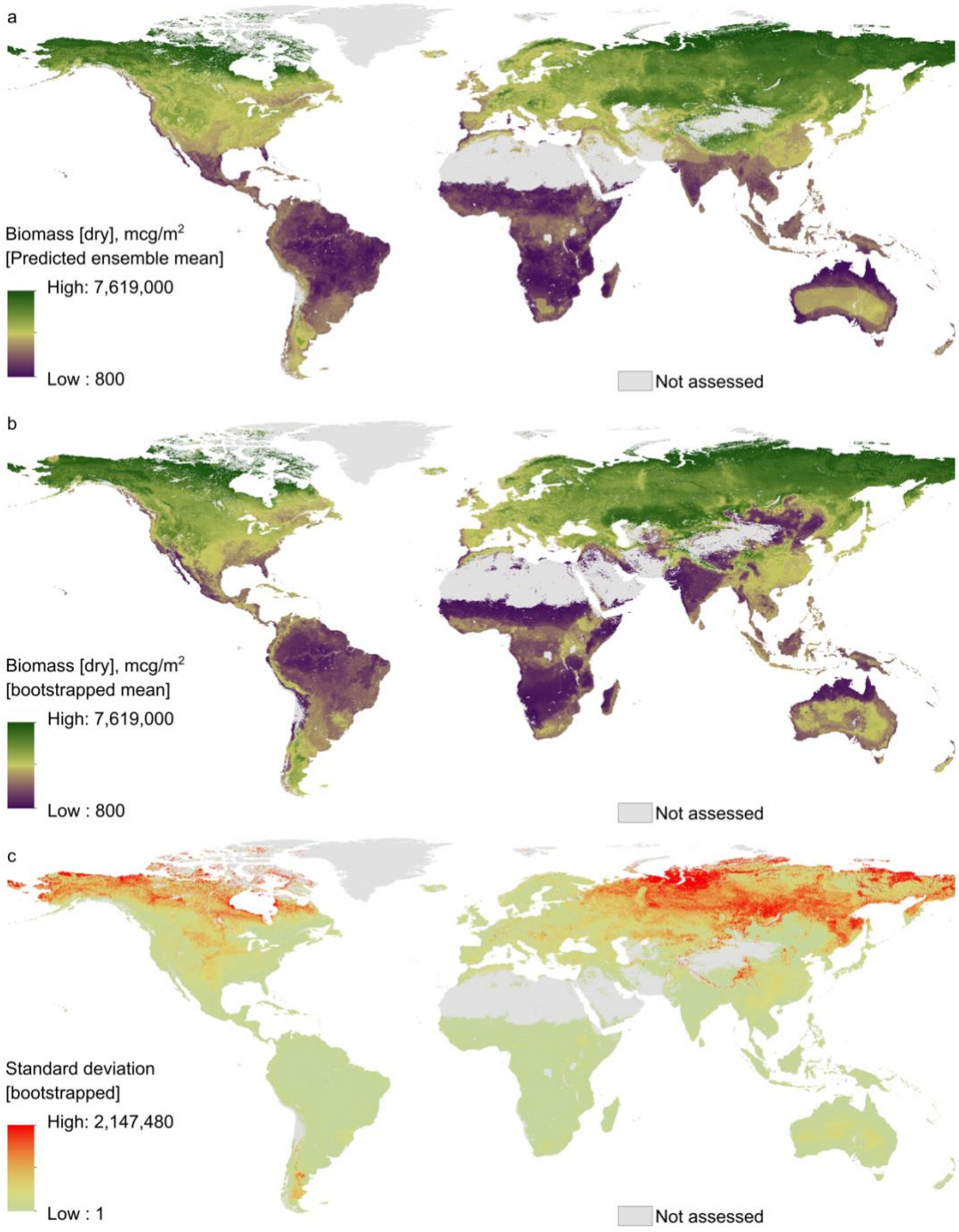
Extended Data Fig. 3 | Mean estimates for community parameters in different ecosystem types. Points represent sites, labels represent mean values, means sharing the same letter are not significantly different (Tukey's HSD test for multiple comparisons[64]). For ecosystem classification see Methods.



Extended Data Fig. 4 | Global projection of springtail density. Distribution was predicted with the random forest algorithm (a) based on the entire dataset and (b) using mean prediction after bootstrapping data by continents ($R^2 = 0.57 \pm 0.04$). Green colour identifies hot spots, violet colour cold spots. The bottom map (c) shows the standard deviation across the bootstrapped predictions (red – high, yellow – low). All data were projected at the 30 arcsec (approximately 1 km^2) pixel scale.

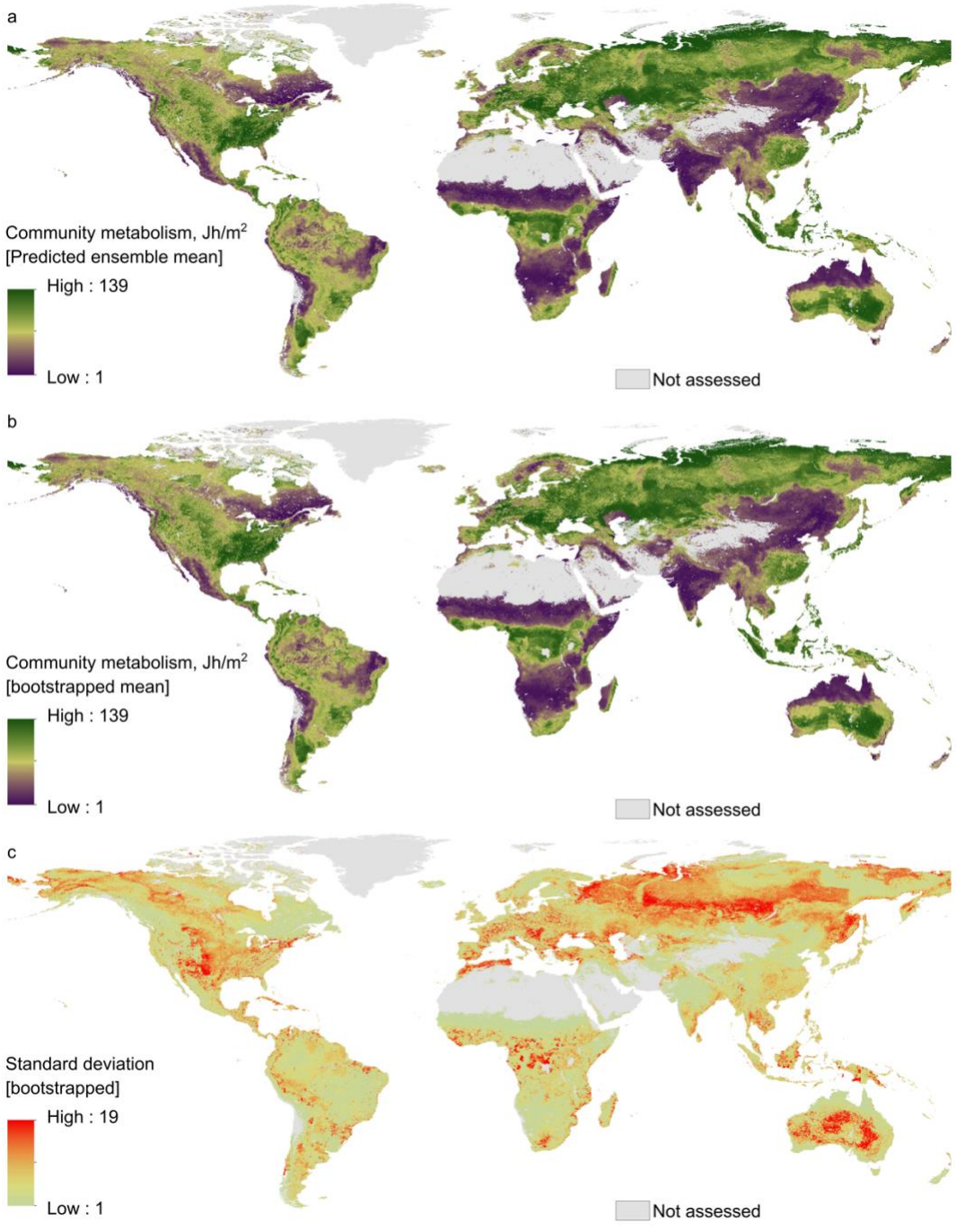


Extended Data Fig. 5 | Global projection of springtail local species richness. Distribution was predicted with the random forest algorithm (a) based on the entire dataset and (b) using mean prediction after bootstrapping data by continents ($R^2 = 0.31 \pm 0.06$). Green colour identifies hot spots, violet colour cold spots. The bottom map (c) shows the standard deviation across the bootstrapped predictions (red – high, yellow – low). All data were projected at the 30 arcsec (approximately 1 km^2) pixel scale.



835

836 **Extended Data Fig. 6 | Global projection of springtail biomass.** Distribution was predicted
837 with the random forest algorithm (a) based on the entire dataset and (b) using mean
838 prediction after bootstrapping data by continents ($R^2 = 0.47 \pm 0.05$). Green colour identifies
839 hot spots, violet colour cold spots. The bottom map (c) shows the standard deviation across
840 the bootstrapped predictions (red – high, yellow – low). All data were projected at the 30
841 arcsec (approximately 1 km^2) pixel scale.



842

843 **Extended Data Fig. 7 | Global projection of springtail community metabolism.**

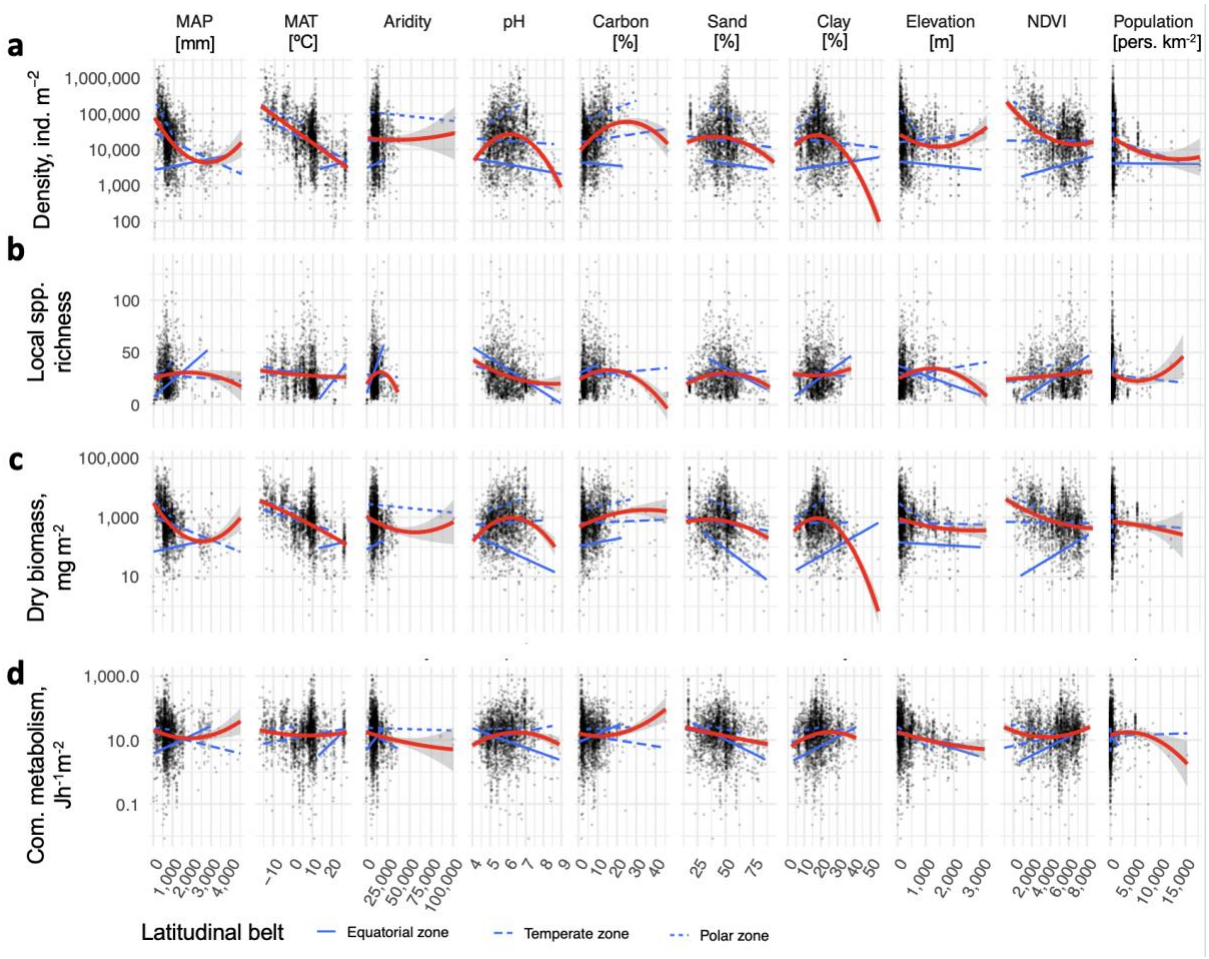
844 Distribution was predicted with the random forest algorithm (a) based on the entire dataset

845 and (b) using mean prediction after bootstrapping data by continents ($R^2 = 0.33 \pm 0.09$).

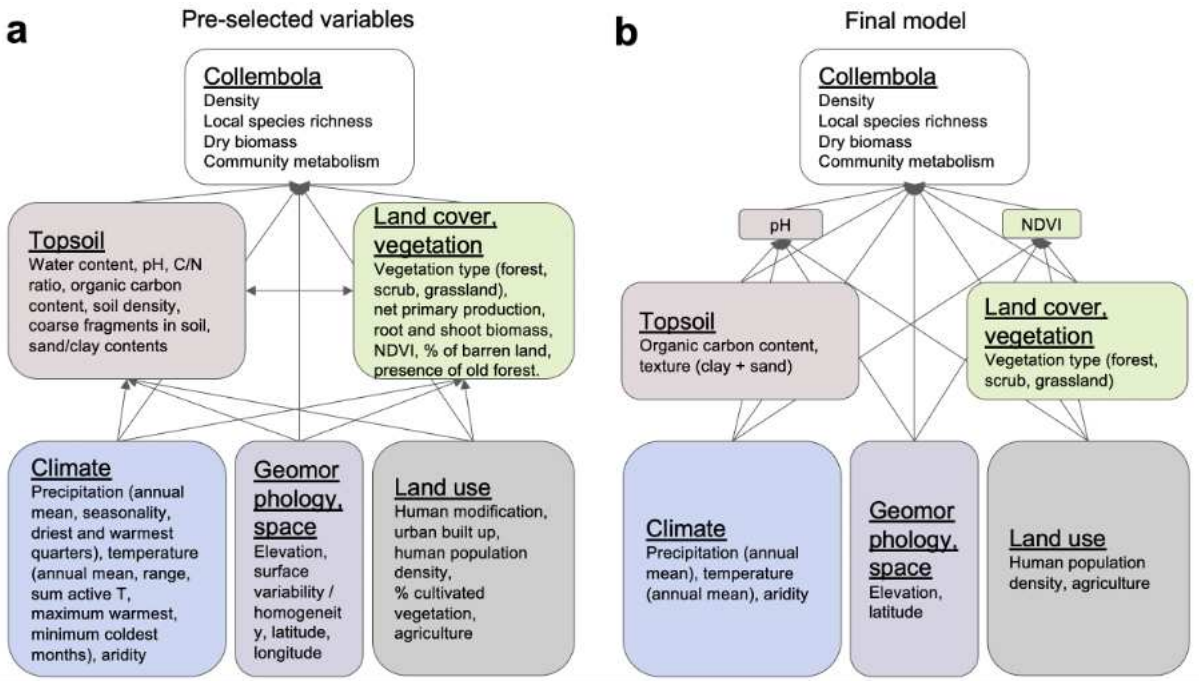
846 Green colour identifies hot spots, violet colour cold spots. The bottom map (c) shows the

847 standard deviation across the bootstrapped predictions (red – high, yellow – low). All data

848 were projected at the 30 arcsec (approximately 1 km^2) pixel scale.



Extended Data Fig. 8 | Associations of selected environmental variables with springtail density, local species richness, dry biomass and community metabolism. Quadratic function was used for approximation to illustrate global trends (red line). Blue lines show linear trends in equatorial (solid), temperate (long dash) and polar zones (short dash).



Extended Data Fig. 9 | Initial and final relationship diagram in the path analysis. Factors directly and indirectly affecting community parameters of springtails at the global scale were pre-selected based on expert opinion (a). Factors in the final model (b) were further selected according to their global availability and collinear factors were removed. The global distributions of pH and NDVI (Normalized Difference Vegetation Index) are initially modelled based on other factors, which was accounted for in the final model.