Global distribution of earthworm diversity

Helen R. P. Phillips[1,2]*, Carlos A. Guerra[1,3], Marie L. C. Bartz[4], Maria J. I. Briones[5], George Brown[6], Thomas W. Crowther[7], Olga Ferlian[1,2], Konstantin B. Gongalsky[8,9], Johan van den Hoogen[7], Julia Krebs[1,2], Alberto Orgiazzi[10], Devin Routh[7], Benjamin Schwarz[11], Elizabeth M. Bach[12,13], Joanne Bennett[1,3], Ulrich Brose[1,14], Thibaud Decaëns[15], Birgitta König-Ries[1,16], Michel Loreau[17], Jérôme Mathieu[18], Christian Mulder[19], Wim H. van der Putten[20,21], Kelly S. Ramirez[20], Matthias C. Rillig[22,23], David Russell[24], Michiel Rutgers[25], Madhav P. Thakur[20], Franciska T. de Vries[26], Diana H. Wall[12,13], David A. Wardle[27], Miwa Arai[28], Fredrick O. Ayuke[29], Geoff H. Baker[30], Robin Beauséjour[31], José C. Bedano[32], Klaus Birkhofer[33], Eric Blanchart[34], Bernd Blossey[35], Thomas Bolger[36,37], Robert L. Bradley[31], Mac A. Callaham[38], Yvan Capowiez[39], Mark E. Caulfield[40], Amy Choi[41], Felicity V. Crotty[42,43], Andrea Dávalos[35,44], Darío J. Diaz Cosin[45], Anahí Dominguez[32], Andrés Esteban Duhour[46], Nick van Eekeren[47], Christoph Emmerling[48], Liliana B. Falco[49], Rosa Fernández[50], Steven J. Fonte[51], Carlos Fragoso[52], André L. C. Franco[12], Martine Fugère[31], Abegail T. Fusilero[53,54], Shaieste Gholami[55], Michael J. Gundale[56], Mónica Gutiérrez López[45], Davorka K. Hackenberger[57], Luis M. Hernández[58], Takuo Hishi[59], Andrew R. Holdsworth[60], Martin Holmstrup[61], Kristine N. Hopfensperger[62], Esperanza Huerta Lwanga[63,64], Veikko Huhta[65], Tunsisa T. Hurisso[51,66], Basil V. Iannone III[67], Madalina Iordache[68], Monika Joschko[69], Nobuhiro Kaneko[70], Radoslava Kanianska[71], Aidan M. Keith[72], Courtland A. Kelly[51], Maria L. Kernecker[73], Jonatan Klaminder[74], Armand W. Koné[75], Yahya Kooch[76], Sanna T. Kukkonen[77], Hmar Lalthanzara[78], Daniel R. Lammel[23,79], Iurii M. Lebedev[8,9], Yiqing Li[80], Juan B. Jesus Lidon[45], Noa K. Lincoln[81], Scott R. Loss[82], Raphael Marichal[83], Radim Matula[84], Jan Hendrik Moos[85,86], Gerardo Moreno[87], Alejandro Morón-Ríos[88], Bart Muys[89], Johan Neirynck[90], Lindsey Norgrove[91], Marta Novo[45], Visa Nuutinen[92], Victoria Nuzzo[93], Mujeeb Rahman P[94], Johan Pansu[95,96], Shishir Paudel[82], Guénola Pérès[97], Lorenzo Pérez-Camacho[98], Raúl Piñeiro[99], Jean-François Ponge[100], Muhammad Imtiaz Rashid[101,102], Salvador Rebollo[98], Javier Rodeiro-Iglesias[103], Miguel Á. Rodríguez[104], Alexander M. Roth[105,106], Guillaume X. Rousseau[58,107], Anna Rozen[108], Ehsan Sayad[55], Loes van Schaik[109], Bryant C. Scharenbroch[110], Michael Schirrmann[111], Olaf Schmidt[37,112], Boris Schröder[22,113], Julia Seeber[114,115], Maxim P. Shashkov[116,117], Jaswinder Singh[118], Sandy M. Smith[119], Michael Steinwandter[115], José A. Talavera[120], Dolores Trigo[45], Jiro Tsukamoto[121], Anne W. de Valença[122], Steven J. Vanek[51], Iñigo Virto[123], Adrian A. Wackett[124], Matthew W. Warren[125], Nathaniel H. Wehr[126], Joann K. Whalen[127], Michael B. Wironen[128], Volkmar Wolters[129], Irina V. Zenkova[130], Weixin Zhang[131], Erin K. Cameron[132,133]†, Nico Eisenhauer[1,2]†

Affiliations:

[1] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany
[2] Institute of Biology, Leipzig University, Deutscher Platz 5e, 04103 Leipzig, Germany
[3] Institute of Biology, Martin Luther University Halle-Wittenberg, Am Kirchtor 1, 06108, Halle (Saale), Germany
[4] Universidade Positivo, Rua Prof. Pedro Viriato Parigot de Souza, 5300, Curitiba, PR, Brazil, 81280-330
[5] Departamento de Ecología y Biología Animal, Universidad de Vigo, 36310 Vigo, Spain
[6] Embrapa Forestry, Estrada da Ribeira, km. 111, C.P. 231, Colombo, PR, Brazil, 83411-000
[7] Crowther Lab, Institute of Integrative Biology, ETH Zürich, 8092 Zürich, Switzerland
[8] A.N. Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Leninsky pr., 33, Moscow, 119071, Russia
[9] M.V. Lomonosov Moscow State University, Leninskie Gory, 1, Moscow, 119991, Russia
[10] European Commission, Joint Research Centre (JRC), Ispra, Italy
[11] Biometry and Environmental System Analysis, University of Freiburg, Tennenbacher Str. 4, 79106 Freiburg, Germany

[12] Department of Biology, Colorado State University, Fort Collins, CO, 80523, USA
[13] Global Soil Biodiversity Initiative and School of Global Environmental Sustainability, Colorado State University, Fort Collins, CO 80523, USA
[14] Institute of Biodiversity, Friedrich Schiller University Jena, Dornburger-Str. 159, 07743, Jena, Germany
[15] CEFE, UMR 5175, CNRS–Univ Montpellier–Univ Paul–Valéry–EPHE–SupAgro Montpellier–INRA–IRD, 1919 Route de Mende, 34293 Montpellier Cedex 5, France
[16] Institute of Computer Science, Friedrich Schiller University Jena, Ernst-Abbe-Platz 2, 07743 Jena, Germany
[17] Centre for Biodiversity Theory and Modelling, Theoretical and Experimental Ecology Station, CNRS, 09200 Moulis, France
[18] Sorbonne Université, CNRS, UPEC, Paris 7, INRA, IRD, Institut d'Ecologie et des Sciences de l'Environnement de Paris, 4 place Jussieu, F-75005, Paris, France
[19] Department of Biological, Geological and Environmental Sciences, University of Catania, Via Androne 81, 95124 Catania, Italy
[20] Department of Terrestrial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), 6700 AB Wageningen, The Netherlands
[21] Laboratory of Nematology, Department of Plant Sciences, Wageningen University & Research, 6708 PB, Wageningen, The Netherlands
[22] Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), Altensteinstraße 6, 14195 Berlin, Germany
[23] Institute of Biology, Freie Universität Berlin, Altensteinstraße 6, 14195 Berlin, Germany
[24] Senckenberg Museum for Natural History Görlitz, Department of Soil Zoology, 02826 Görlitz, Germany
[25] National Institute for Public Health and the Environment, Bilthoven, The Netherlands
[26] Institute of Biodiversity and Ecosystem Dynamics, University of Amsterdam, The Netherlands
[27] Asian School of the Environment, Nanyang Technological University, Singapore 639798
[28] Institute for Agro-Environmental Sciences, National Agriculture and Food Research Organization, Tsukuba, 305-8604, Japan
[29] Department of Land Resource Management and Agricultural technology (LARMAT), College of Agriculture and Veterinary Sciences, University of Nairobi, Kapenguria Road, P.O Box 29053, Nairobi, 00625, Kenya
[30] CSIRO Health & Biosecurity, GPO Box 1700, Canberra A.C.T. 2601, Australia
[31] Département de Biologie, Université de Sherbrooke, 2500 boulevard de l'Université, Sherbrooke J1K 2R1, Canada
[32] Geology Department, FCEFQyN, ICBIA-CONICET (National Scientific and Technical Research Council), National University of Río Cuarto, Ruta 36, Km. 601, X5804 BYA Río Cuarto, Argentina
[33] Department of Ecology, Brandenburg University of Technology, Konrad-Wachsmann-Allee 6, 03046 Cottbus, Germany
[34] Eco&Sols, University of Montpellier, IRD, CIRAD, INRA, Montpellier SupAgro, 34060 Montpellier, France
[35] Department of Natural Resources, Cornell University, Fernow Hall, Ithaca, NY, 14853, USA
[36] School of Biology and Environmental Science, University College Dublin, Belfield, Dublin 4, Ireland
[37] UCD Earth Institute, University College Dublin, Belfield, Dublin 4, Ireland
[38] USDA Forest Service, Southern Research Station, 320 Green Street, Athens, GA, 30602, USA
[39] UMR 1114 "EMMAH', INRA, Site Agroparc, 84914 Avignon, France
[40] Farming Systems Ecology, Wageningen University and Research, Droevendaalsesteeg 1, 6700 AK Wageningen, The Netherlands
[41] Faculty of Forestry, University of Toronto, 33 Willcocks Street, Toronto, Ontario, M5S3B3, Canada
[42] Institute of Biological, Environmental & Rural Sciences, Aberystwyth University, Plas Gogerddan, Aberystwyth, SY23 3EE, United Kingdom

[43] School of Agriculture, Food and Environment, Royal Agricultural University, Stroud Road, Cirencester, GL7 6JS, United Kingdom
[44] Department of Biological Sciences, SUNY Cortland, Bowers Hall 1215, Cortland, NY, 13045, USA
[45] Biodiversity, Ecology and Evolution, Faculty of Biology, Complutense University of Madrid, Calle Jose Antonio Novais, 12, 28040 Madrid, Spain
[46] Laboratorio de Ecología, Instituto de Ecología y Desarrollo Sustentable, Universidad Nacional de Luján, Avenida Constitución y Ruta 5, 6700 Luján, Argentina
[47] Louis Bolk Institute, Kosterijland 3-5, 3981 AJ Bunnik, The Netherlands
[48] Department of Soil Science, Faculty of Regional & Environmental Sciences, University of Trier, Campus II, 54286 Trier, Germany
[49] Ciencias Básicas, Instituto de Ecología y Desarrollo Sustentable -INEDES, Universidad Nacional de Lujan, Av. Constitución y Ruta N 5, 6700 Luján, Argentina
[50] Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37, 08003 Barcelona, Spain
[51] Department of Soil and Crop Sciences, Colorado State University, 305 University Ave, Fort Collins, CO, 80523, USA
[52] Biodiversity and Systematic Network, Instituto de Ecología A.C., Carretera Antigua a Coatepec 351, Xalapa, 91070, Mexico
[53] Department of Biological Science and Environmental Studies, University of the Philippines - Mindanao, Barangay Mintal, 8022 Davao City, Philippines
[54] Laboratory of Environmental Toxicology and Aquatic Ecology, Environmental Toxicology Unit (GhEnToxLab), Ghent University (UGent), Campus Coupure, Coupure Links 653, Ghent, Belgium
[55] Natural Resources Department, Razi University, Kermanshah, Iran
[56] Forest Ecology and Management, Swedish University of Agricultural Sciences, Skogsmarksgrand 17, 90183 Umeå, Sweden
[57] Department of Biology, J. J. Strossmayer University of Osijek, Cara Hadrijana 8a, 31000 Osijek, Croatia
[58] Agricultural Engineering, Postgraduate Program in Agroecology, Maranhão State University, Avenida Lourenço Vieira da Silva 1000, 65.055-310 São Luís, Brazil
[59] Faculty of Agriculture, Kyushu University, 949 Ohkawauchi, Shiiba, 883-0402, Japan
[60] Minnesota Pollution Control Agency, 520 Lafayette Road, St Paul, MN, 55155, USA
[61] Department of Bioscience, Aarhus University, Vejlsøvej 25, 8600 Silkeborg, Denmark
[62] Biological Sciences, Northern Kentucky University, 1 Nunn Drive, Highland Heights, KY, 41099, USA
[63] Agricultura Sociedad y Ambiente, El Colegio de la Frontera Sur, Av. Rancho Polígono 2-A, Ciudad Industrial, Lerma, Campeche, 24500, Mexico
[64] Soil Physics and Land Management degradation, Wageningen University & Research, Wageningen University, Droevendaalsesteeg 3, 6708 PB Wageningen, The Netherlands
[65] Department of Biological and Environmental Science, University of Jyväskylä, Box 35, 40014 Jyväskylä, Finland
[66] College of Agriculture, Environmental and Human Sciences, Lincoln University of Missouri, Jefferson City, MO, 65101, USA
[67] School of Forest Resources and Conservation, University of Florida, PO Box 110940, Gainesville, FL, 32611, USA
[68] Sustainable Development and Environment Engineering, Banat's University of Agricultural Sciences and Veterinary Medicine "King Michael the 1st of Romania", Calea Aradului 119, 300645, Timisoara, Romania
[69] Experimental Infrastructure Platform, Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Str. 84, 15374 Müncheberg, Germany
[70] Faculty of Food and Agricultural Sciences, Fukushima University, Kanayagawa 1, Fukushima City, Japan

[71] Department of Environmental Management, Faculty of Natural Sciences, Matej Bel University, Tajovského 40, Banská Bystrica, Slovakia
[72] Centre for Ecology & Hydrology, Library Avenue, Bailrigg, Lancaster, LA1 4AP, United Kingdom
[73] Land Use and Governance, Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Str. 84, 15374 Müncheberg, Germany
[74] Department of Ecology and Environmental Science, Climate Impacts Research Centre, Umeå University, 90187, Umeå, Sweden
[75] UR Gestion Durable des Sols, UFR Sciences de la Nature, Université Nangui Abrogoua, Abidjan, Côte d'Ivoire
[76] Faculty of Natural Resources & Marine Sciences, Tarbiat Modares University, 46417-76489, Noor, Mazandaran, Iran
[77] Production Systems, Horticulture Technologies, Natural Resources Institute Finland, Survontie 9 A, 40500 Jyväskylä, Finland
[78] Department of Zoology, Pachhunga University College, College Veng, Aizawl, 796001, India
[79] Soil Science, ESALQ-USP, Universidade de São Paulo, Av. Pádua Dias, 11, Piracicaba, 13418, Brazil
[80] College of Agriculture, Forestry and Natural Resource Management, University of Hawaii at Hilo, 200 W. Kawili Street, Hilo, HI, 96720, USA
[81] Tropical Plant and Soil Sciences, College of Tropical Agriculture and Human Resources, University of Hawai‘i at Mānoa, 3190 Maile Way, St. John 102, Honolulu, HI, 96822, USA
[82] Department of Natural Resource Ecology and Management, Oklahoma State University, 008C Ag Hall, Stillwater, OK, 74078, USA
[83] UR Systèmes de pérennes, CIRAD, Univ Montpellier, TA B-34 / 02, Avenue Agropolis, 34398 Montpellier, France
[84] Department of Forest Ecology, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Kamýcká 129, 165 21 Prague, Czech Republic
[85] Department of Soil and Environment, Forest Research Institute of Baden-Wuerttemberg, Wonnhaldestrasse 4, 79100 Freiburg, Germany
[86] Thuenen-Institute of Organic Farming, Trenthorst 32, 23847 Westerau, Germany
[87] Forestry School - INDEHESA, University of Extremadura, 10600 Plasencia, Spain
[88] Conservación de la Biodiversidad, El Colegio de la Frontera Sur, Av. Rancho, poligono 2A, Cd. Industrial de Lerma, 24500 Campeche, Mexico
[89] Department of Earth & Environmental Sciences, KU Leuven, Celestijnenlaan 200E, Box 2411, 3001 Leuven, Belgium
[90] Research Institute for Nature and Forest, Gaverstraat, 35, 9500 Geraardsbergen, Belgium
[91] School of Agricultural, Forest and Food Sciences, Bern University of Applied Sciences, Länggasse 85, 3052 Zollikofen, Switzerland
[92] Soil Ecosystems, Natural Resources Institute Finland (Luke), Tietotie 4, 31600 Jokioinen, Finland
[93] Natural Area Consultants, 1 West Hill School Road, Richford, NY, 13835, USA
[94] Department of Zoology, Pocker Sahib Memorial Orphanage College, Tirurangadi, Malappuram, Kerala, India
[95] CSIRO Ocean & Atmosphere, CSIRO, New Illawarra Road, Lucas Heights, 2234, Australia
[96] UMR7144 Adaptation et Diversité en Milieu Marin, Station Biologique de Roscoff, CNRS-Sorbonne Universite, Place Georges Teissier, 29688 Roscoff, France
[97] UMR SAS, INRA, Agrocampus Ouest, 65 rue Saint Brieuc, 35000 Rennes, France
[98] Ecology and Forest Restoration Group, Department of Life Sciences, University of Alcalá, 28801 Alcalá De Henares, Spain
[99] Computing, ESEI, Vigo, Edf. Politécnico - Campus As Lagoas, 32004 Ourense, Spain
[100] Adaptations du Vivant, CNRS UMR 7179, Muséum National d'Histoire Naturelle, 4 avenue du Petit Château, 91800 Brunoy, France
[101] Centre of Excellence in Environmental Studies, King Abdulaziz University, P.O. Box 80216, Jeddah 21589, Saudi Arabia

[102] Environmental Sciences, COMSATS University, Islamabad, Sub-campus, Vehari, Vehari, 61100, Pakistan
[103] Departamento de Informática, Escuela Superior de Ingeniería Informática, Universidad de Vigo, 36310 Vigo, Spain
[104] Life Sciences, Sciences Faculty, University of Alcalá, Science and Technology Campus, 28805 Alcalá De Henares, Spain
[105] Department of Forest Resources, University of Minnesota, 1530 Cleveland Avenue North, St. Paul, MN, 55101, USA
[106] Friends of the Mississippi River, 101 East Fifth Street Suite 2000, St Paul, MN, 55108, USA
[107] Postgraduate Program in Biodiversity and Conservation, Federal University of Maranhão, Avenida dos Portugueses 1966, 65080-805 São Luís, Brazil
[108] Institute of Environmental Sciences, Jagiellonian University, Gronostajowa 7, 30-087 Kraków, Poland
[109] Institute of Ecology, Technical University of Berlin, Ernst Reuter Platz 1, 10587 Berlin, Germany
[110] College of Natural Resources, University of Wisconsin - Stevens Point, 800 Reserve Street, Stevens Point, WI, 54481, USA
[111] Engineering for Crop Production, Leibniz Institute for Agricultural Engineering and Bioeconomy (ATB), Max-Eyth-Allee 100, 14469 Potsdam, Germany
[112] UCD School of Agriculture and Food Science, University College Dublin, Belfield, Ireland
[113] Landscape Ecology and Environmental Systems Analysis, Institute of Geoecology, Technische Universität Braunschweig, Langer Kamp 19c, 38106 Braunschweig, Germany
[114] Department of Ecology, University of Innsbruck, Technikerstrasse 25, 6020 Innsbruck, Austria
[115] Institute for Alpine Environment, Eurac Research, Drususallee 1, 39100 Bozen/Bolzano, Italy
[116] Laboratory of Ecosystem Modelling, Institute of Physicochemical and Biological Problems in Soil Sciences, Russian Academy of Science, Institutskaya 2, Pushchino, 142290, Russia
[117] Laboratory of Computational Ecology, Institute of Mathematical Problems of Biology RAS – the Branch of Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Professor Vitkevich 1, Pushchino, 142290, Russia
[118] Post Graduate Department of Zoology, Khalsa College Amritsar, Amritsar, 143002, India
[119] John H. Daniels Faculty of Architecture, Landscape and Design, University of Toronto, 33 Willcocks St., Earth Sciences Building, Toronto, Ontario, M5S 3B3, Canada
[120] Department of Animal Biology (Zoology area), Science Faculty, University of La Laguna, 38200 La Laguna, Spain
[121] Faculty of Agriculture, Kochi University, Monobe Otsu 200, Nankoku, 783-8502, Japan
[122] Food & Agriculture, WWF-Netherlands, Driebergseweg 10, 3708 JB Zeist, The Netherlands
[123] Dpto. Ciencias, IS-FOOD, Universidad Pública de Navarra, Edificio Olivos - Campus Arrosadia, 31006 Pamplona, Spain
[124] Soil, Water and Climate, University of Minnesota, 1991 Upper Buford Circle, St Paul, MN, 55108, USA
[125] Earth Innovation Institute, 98 Battery Street, Suite 250, San Francisco, CA, 94111, USA
[126] Department of Natural Resources & Environmental Management, University of Hawai'i at Mānoa, 1910 East-West Rd., Honolulu, HI, 96822, USA
[127] Natural Resource Sciences, McGill University, 21111 Lakeshore Road, Ste-Anne-de-Bellevue, H9X 3V9, Canada
[128] The Nature Conservancy, 4245 North Fairfax Drive Suite 100, Arlington, VA, 22203, USA
[129] Department of Animal Ecology, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany
[130] Laboratory of Terrestrial Ecosystems, Kola Science Centre, Institute of the North Industrial Ecology Problems, Akademgorodok, 14a, Apatity, 184211, Russia
[131] Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions (Henan University), Ministry of Education, College of Environment and Planning, Henan University, Kaifeng, 475004, China

[132] Department of Environmental Science, Saint Mary's University, Halifax, Nova Scotia, Canada
[133] Faculty of Biological and Environmental Sciences, Post Office Box 65, FI 00014, University of Helsinki, Finland

* Correspondence to: Helen R P Phillips (helen.phillips@idiv.de)
† These authors contributed equally

Abstract

Soil organisms, including earthworms, are a key component of terrestrial ecosystems. However, little is known about their diversity, distribution, and the threats affecting them. Here, we compiled a global dataset of sampled earthworm communities from 6928 sites in 57 countries to predict patterns in earthworm diversity, abundance, and biomass. We identified that local species richness and abundance typically peaked at higher latitudes, patterns opposite to those observed in aboveground organisms. However, diversity across the entirety of the tropics may be higher than elsewhere, due to high species dissimilarity across locations. Climate variables were more important in shaping earthworm communities than soil properties or habitat cover. These findings suggest that climate change may have serious implications for earthworm communities and therefore the functions they provide.

One sentence summary: Precipitation and temperature drive global earthworm diversity, abundance, and biomass, but latitudinal patterns differ from many aboveground taxa.

Main Text

Soils harbour high biodiversity, and are responsible for a wide range of ecosystem functions and services upon which terrestrial life depends (*1*). Despite calls for large-scale biogeographic studies of soil organisms (*2*), global biodiversity patterns remain relatively unknown, with most efforts focused on soil microbes (*3*–*5*). Consequently, the drivers of soil biodiversity, particularly soil fauna, remain unknown at the global scale.

Furthermore, our ecological understanding of global biodiversity patterns (e.g., latitudinal diversity gradients (*5*)) is largely based on the distribution of aboveground taxa. Yet, many soil organisms have shown global diversity patterns that differ from aboveground organisms (*3*, *7*–*9*), although the patterns often depend on the size of the soil organism (*10*).

Here, we analyse global patterns in earthworm diversity, total abundance, and total biomass (hereafter 'community metrics'). Earthworms are considered ecosystem engineers (*11*) in many habitats, and also provide a variety of vital ecosystem functions and services (*12*). The provisioning of ecosystem functions by earthworms likely depends on the abundance, biomass, and ecological group of the earthworm species (*13*, *14*). Consequently, understanding global patterns in community metrics for earthworms is critical for predicting how changes in their communities may alter ecosystem functioning.

Small-scale field studies have shown that soil properties such as pH and soil carbon influence earthworm diversity (*11*, *15*, *16*). For example, lower pH values constrain the diversity of earthworms by reducing calcium availability (*17*), and soil carbon provides resources that sustain earthworm diversity and population sizes (*11*). Alongside many interacting soil properties (*15*), a variety of other drivers can shape earthworm diversity, such as climate and habitat cover (*11*, *18*, *19*). However, to date, no framework has integrated a comprehensive set of environmental drivers of earthworm communities to identify the most important ones at a global scale.

Previous reviews suggested earthworms may have high diversity across the tropics due to high endemism (*10*). However, this high regional diversity may not be captured by local-scale metrics. Alternatively, in the temperate region, local diversity may be higher (*20*) but include fewer endemic species (*10*). We anticipate that earthworm community metrics (particularly diversity) will not follow global patterns seen aboveground, and instead, as seen across Europe (*15*), increase with latitude. This finding would be

consistent with previous studies at regional scales, which showed that the species richness of earthworms increases with latitude (*19*). Because of the relationship between earthworm communities, habitat cover, and soil properties on local scales, we expect soil properties (e.g., pH and soil organic carbon) to be key environmental drivers of earthworm communities.

Here, we present global maps predicting local (i.e., site-level; a location of one or more samples, which adequately captured the earthworm community): diversity (the number of species), abundance, and biomass. We collated 180 datasets from the literature and unpublished field studies (164 and 16, respectively) to create a dataset spanning 57 countries (all continents except Antarctica) and 6928 sites (Fig. 1A). We explore spatial patterns of earthworm communities, and determine the environmental drivers that shape earthworm biodiversity. We then used the relationships between earthworm community metrics and environmental drivers (Table S1) to predict local earthworm communities across the globe.

Three generalised linear mixed effects models were constructed, one for each of the three community metrics; species richness (calculated within a site), abundance per m^2, and biomass per m^2. Each model contained 12 environmental variables as main effects (Table S2), which were grouped into six themes; 'soil', 'precipitation', 'temperature', 'water retention', 'habitat cover', and 'elevation' (habitat cover and some soil variables were measured in the field, the remaining variables were extracted from global data layers using the geographic coordinates of the sites (*12*)). Within each theme, each model contained interactions between the variables. Following model simplification, all models retained most of the original variables, but some interactions were removed (Table S3).

Consistent with previous results (*20*), predictions based on global environmental data layers resulted in estimates of local earthworm diversity between 1 and 4 species per site across most of the terrestrial surface (Fig. 1B) (mean: 2.42 species; SD: 2.19). Most of the boreal/subarctic regions were predicted to have low values of species richness, which is in line with aboveground biodiversity patterns (*21*, *22*).

However, low local diversity also occurred in subtropical and tropical areas, such as Brazil, India and Indonesia, in contrast with commonly observed aboveground patterns, such as the latitudinal gradient in plant diversity (*22*). This pattern could be due to different relationships with climate variables. For example, while plant diversity increases with potential evapotranspiration (PET) (*22*), earthworm diversity tended to decrease with increasing PET (Table S3). In addition, soil properties, which are typically not included in models of aboveground diversity, can play a role in determining earthworm communities (*11*, *15*, *23*). For instance, litter availability and soil nutrient content are important regulators of earthworm diversity, with oligotrophic forest soils having more epigeic species, and eutrophic soils more endogeics (*23*). Furthermore, tropical regions with higher decomposition rates have fewer soil organic resources and lower local earthworm diversity (Fig. 1B & Table S3), dominated by endogeic species, that have specific digestion systems allowing them to feed on low quality soil organic matter (*11*, *14*, *20*).

High local species richness was found at mid-latitudes, such as the southern tip of South America, the southern regions of Australia and New Zealand, Europe (particularly north of the Black Sea) and northeastern USA. While this pattern contrasts with latitudinal diversity patterns found in many aboveground organisms (*6*, *24*), it is consistent with patterns found in some belowground organisms (ectomycorrhizal fungi (*3*, *22)*, bacteria (*23*)), but not all (arbuscular mycorrhizal fungi (*9)*, oribatid mites (*29*)). Such mismatches between above- and belowground biodiversity have been predicted (*1*, *7*) but not shown across the globe for soil fauna at the local scale.

The patterns seen here could in part be a result of glaciation in the last ice age, as well as human activities. Temperate regions (mid- to high latitudes) that were previously glaciated were likely re-colonised by earthworm species with high dispersal capabilities and large geographic ranges (*19*) and through human-mediated dispersal (‘anthropochorous’ earthworms (*16*)). Thus, temperate communities could have high local diversity, as seen here, but those species would be widely distributed resulting in

lower regional diversity relative to local diversity. In the tropics, which did not experience glaciation, the opposite may be true. Specific locations may have individual species that are highly endemic, but these species are not widely distributed (Table S4). This high local endemism would result in low local diversity (as found here) and high regional diversity (as suggested by (*10*)) relative to that low local diversity. When the number of unique species within latitudinal zones that had equal number of sites was calculated (i.e., a regional richness that accounted for sampling effort), there appeared to be a regional latitudinal diversity gradient (Fig. 2). Even with a sampling bias (Table S4), regional richness in the tropics was greater than the temperate regions, despite low local diversity. These results should be interpreted with caution though given the latitude span of the tropical zones, highlighting the need for additional sampling within this region. However, the underlying data suggests endemism of earthworms and beta diversity within the tropics (*28*) may be considerably higher than within the well-sampled temperate region (Table S4). Therefore, it is likely that the tropics harbour more species overall.

The predicted total abundance of the local community of earthworms typically ranged between 5 and 150 individuals per m^2 across the globe, in line with other estimates (*29*) (Fig. 1C; mean: 77.89 individuals per m^2; SD: 98.94). There was a slight tendency for areas of higher total abundance to be in temperate areas, such as Europe (particularly the UK, France and Italy), New Zealand, and part of the Pampas and surrounding region (South America), rather than the tropics. Lower total abundance occurred in many of the tropical and sub-tropical regions, such as Brazil, central Africa, and parts of India. Given the positive relationship between total abundance and ecosystem function (*30*), in regions with lower earthworm abundance functions may be reduced or carried out by other soil taxa (*1*).

The predicted total biomass of the local earthworm community (adults and juveniles) across the globe showed extreme values (>2 kg) in 0.3% of pixels, but biomass typically ranged (97% of pixels) between 1 g and 150 g per m^2 (Fig. 1D; median: 6.69; mean: 2772.8; SD: 1312782; see (*14*) for additional discussion of extreme values). The areas of high total biomass were concentrated in the Eurasian Steppe

and some regions of North America. The majority of the globe showed low total biomass. In northern North America, where there are no native earthworms (*13*), high density and, in some regions, higher biomass of earthworms likely reflects the earthworm invasion of these regions. The small invasive European earthworm species encounter an enormous unused resource pool, which leads to high population sizes (*31*). Based on previous suggestions (*29*), we expected that earthworms would decrease in body size towards the poles, showing low biomass relative to the total abundance in temperate/boreal regions. In contrast, in tropical regions (e.g., Brazil and Indonesia) that are dominated by giant earthworms that normally occur at low densities and low species richness (*32*), we expected high biomass but low abundance. However, these patterns were not found. This could be due to the relatively small number of sample points for the biomass model (n = 3296) compared to the diversity (n = 5416) and total abundance model (n = 6358), reducing the predictive ability of the model (Fig. S1C), most notably in large regions of Asia, and areas of Africa, particularly the boundaries of the Sahara Desert and the southern regions (which coincides with where samples are lacking). Additionally, the difficulty in consistently capturing such large earthworms in every sample may increase data variability, reducing the ability of the model to predict.

Overall, the three community metric models performed well in cross-validation (Fig. S3 & 4) with relatively high R^2 values (Table 1 A and C; see (*14*) for further details and caveats discussion). But, given the nature of such analyses, models and maps should only be used to explore broad patterns in earthworm communities and not at the fine scale, especially in relation to conservation practices (*33*).

For all three community metric models, climatic variables were the most important drivers ('precipitation' theme being the most important for both species richness and total biomass models, and 'temperature' for the total abundance model; Fig. 3). The importance of climatic variables in shaping diversity and distribution patterns at large scales is consistent with many aboveground taxa (e.g., plants (*20)*, reptiles, amphibians, and mammals (*32*)) and belowground taxa (bacteria and fungi (*3)*, nematodes (*22*, *23*)). This

suggests that climate-related methods and data, which are typically used by macroecologists to estimate aboveground biodiversity, may also be suitable for estimating earthworm communities. However, the strong link between climatic variables and earthworm community metrics is cause for concern, as climate will continue to change due to anthropogenic activities over the coming decades (*34*). Our findings further highlight that changes in temperature and precipitation are likely to influence earthworm diversity (*35*) and their distributions (*15*), with implications for the functions that they provide (*12*). Shifts in distributions may be particularly problematic in the case of invasive earthworms, such as in areas of North America, where they can considerably change the ecosystem (*13*). However, a change in climate will most likely affect abundance and biomass of the earthworm communities before diversity, as shifts in the latter depend upon dispersal capabilities, which are relatively low in earthworms.

We expected that soil properties would be the most important driver of earthworm communities, but this was not the case (Fig. 3), likely due to scale of the study. Firstly, the importance of drivers could change at different spatial scales. Climate is driving patterns at global scales but within climatic regions (or at the local scale) other variables may become more important (*36*). Thus, one or more soil properties may be the most important drivers of earthworm communities within each of the primary studies, rather than across them all. Secondly, for soil properties, the mismatch in scale between community metrics and soil properties taken from global layers (for sites where sampled soil properties were missing (*14*)) could also reduce the apparent importance of the theme. Habitat cover did influence the earthworm community (Fig. S5 A and B), especially the composition of the three ecological groups (epigeic, endogeics, and anecics, Fig. S6; (*14*)). Across larger scales, climate influences both habitat cover and soil properties, all of which affect earthworm communities. Being able to account for this indirect effect with appropriate methods and data may alter the perceived importance of soil properties and habitat cover (e.g., with pathway analysis (*36)* and standardised data). However, our habitat cover variable did not directly consider local management (such as land use or intensity).

By compiling a global dataset of earthworm communities, we show, the global distribution of earthworm diversity, abundance, and biomass, and identify key environmental drivers responsible for these patterns. Our findings suggest that climate change might have substantial effects on earthworm communities and the functioning of ecosystems; any climate change-induced alteration in earthworm communities is likely to have cascading effects on other species in these ecosystems (*13*, *29*). Despite earthworm communities being controlled by similar environmental drivers as aboveground communities (*22*, *37*), these relationships result in different patterns of diversity. We highlight the need to integrate belowground organisms into biodiversity research, despite differences in the scale of sampling, if we are to fully understand large-scale patterns of biodiversity and their underlying drivers (*7*, *8*, *38*), especially if processes underlying macroecological patterns differ between aboveground and belowground diversity (*38*). The inclusion of soil taxa may alter the distribution of biodiversity hotspots and conservation priorities. For example, protected areas (*7*) may not be protecting earthworms (*7*), despite their importance as ecosystem function providers (*12*) and soil ecosystem engineers for other organisms (*11*). By modelling both realms, aboveground/belowground comparisons are possible, potentially allowing a clearer view of the biodiversity distribution of whole ecosystems.

References

1. R. D. Bardgett, W. H. van der Putten, Belowground biodiversity and ecosystem functioning. *Nat. 515*. **505**, 505–511 (2014).
2. N. Eisenhauer, P. M. Antunes, A. E. Bennett, K. Birkhofer, A. Bissett, M. A. Bowker, T. Caruso, B. Chen, D. C. Coleman, W. de Boer, P. de Ruiter, T. H. DeLuca, F. Frati, B. S. Griffiths, M. M. Hart, S. Hättenschwiler, J. Haimi, M. Heethoff, N. Kaneko, L. C. Kelly, H. P. Leinaas, Z. Lindo, C. Macdonald, M. C. Rillig, L. Ruess, S. Scheu, O. Schmidt, T. R. Seastedt, N. M. va. Straalen, A. V. Tiunov, M. Zimmer, J. R. Powell, Priorities for research in soil ecology. *Pedobiologia (Jena).* **63**, 1–7 (2017).
3. L. Tedersoo, M. Bahram, S. Polme, U. Koljalg, N. S. Yorou, R. Wijesundera, L. V. Ruiz, A. M. Vasco-Palacios, P. Q. Thu, A. Suija, M. E. Smith, C. Sharp, E. Saluveer, A. Saitta, M. Rosas, T. Riit, D. Ratkowsky, K. Pritsch, K. Poldmaa, M. Piepenbring, C. Phosri, M. Peterson, K. Parts, K. Partel, E. Otsing, E. Nouhra, A. L. Njouonkou, R. H. Nilsson, L. N. Morgado, J. Mayor, T. W. May, L. Majuakim, D. J. Lodge, S. S. Lee, K.-H. Larsson, P. Kohout, K. Hosaka, I. Hiiesalu, T. W. Henkel, H. Harend, L. -d. Guo, A. Greslebin, G. Grelet, J. Geml, G. Gates, W. Dunstan, C. Dunk, R. Drenkhan, J. Dearnaley, A. De Kesel, T. Dang, X. Chen, F. Buegger, F. Q. Brearley, G. Bonito, S. Anslan, S. Abell, K. Abarenkov, Global diversity and geography of soil fungi. *Science*

(80-.). **346**, 1256688–1256688 (2014).

4. M. Delgado-Baquerizo, A. M. Oliverio, T. E. Brewer, A. Benavent-González, D. J. Eldridge, R. D. Bardgett, F. T. Maestre, B. K. Singh, N. Fierer, A global atlas of the dominant bacteria found in soil. *Science (80-.).* **359**, 320–325 (2018).
5. M. Bahram, F. Hildebrand, S. K. Forslund, J. L. Anderson, N. A. Soudzilovskaia, P. M. Bodegom, J. Bengtsson-Palme, S. Anslan, L. P. Coelho, H. Harend, J. Huerta-Cepas, M. H. Medema, M. R. Maltz, S. Mundra, P. A. Olsson, M. Pent, S. Põlme, S. Sunagawa, M. Ryberg, L. Tedersoo, P. Bork, Structure and function of the global topsoil microbiome. *Nature*. **560**, 233–237 (2018).
6. H. Hillebrand, On the Generality of the Latitudinal Diversity Gradient. *Am. Nat.* **163**, 192–211 (2004).
7. E. K. Cameron, I. S. Martins, P. Lavelle, J. Mathieu, L. Tedersoo, M. Bahram, F. Gottschall, C. A. Guerra, J. Hines, G. Patoine, J. Siebert, M. Winter, S. Cesarz, O. Ferlian, H. Kreft, T. E. Lovejoy, L. Montanarella, A. Orgiazzi, H. M. Pereira, H. R. P. Phillips, J. Settele, D. H. Wall, N. Eisenhauer, Global mismatches in aboveground and belowground biodiversity. *Conserv. Biol.*, 430 (2019).
8. N. Fierer, M. S. Strickland, D. Liptzin, M. A. Bradford, C. C. Cleveland, Global patterns in belowground communities. *Ecol. Lett.* **12**, 1238–1249 (2009).
9. J. van den Hoogen, S. Geisen, D. Routh, H. Ferris, W. Traunspurger, D. A. Wardle, R. G. M. de Goede, B. J. Adams, W. Ahmad, W. S. Andriuzzi, R. D. Bardgett, M. Bonkowski, R. Campos-Herrera, J. E. Cares, T. Caruso, L. de Brito Caixeta, X. Chen, S. R. Costa, R. Creamer, J. Mauro da Cunha Castro, M. Dam, D. Djigal, M. Escuer, B. S. Griffiths, C. Gutiérrez, K. Hohberg, D. Kalinkina, P. Kardol, A. Kergunteuil, G. Korthals, V. Krashevska, A. A. Kudrin, Q. Li, W. Liang, M. Magilton, M. Marais, J. A. R. Martín, E. Matveeva, E. H. Mayad, C. Mulder, P. Mullin, R. Neilson, T. A. D. Nguyen, U. N. Nielsen, H. Okada, J. E. P. Rius, K. Pan, V. Peneva, L. Pellissier, J. Carlos Pereira da Silva, C. Pitteloud, T. O. Powers, K. Powers, C. W. Quist, S. Rasmann, S. S. Moreno, S. Scheu, H. Setälä, A. Sushchuk, A. V. Tiunov, J. Trap, W. van der Putten, M. Vestergård, C. Villenave, L. Waeyenberge, D. H. Wall, R. Wilschut, D. G. Wright, J. Yang, T. W. Crowther, Soil nematode abundance and functional group composition at a global scale. *Nature*. **572**, 194–198 (2019).
10. T. Decaëns, Macroecological patterns in soil communities. *Glob. Ecol. Biogeogr.* **19** (2010), pp. 287–302.
11. C. A. Edwards, *Earthworm ecology* (2004).
12. M. Blouin, M. E. Hodson, E. A. Delgado, G. Baker, L. Brussaard, K. R. Butt, J. Dai, L. Dendooven, G. Peres, J. E. Tondoh, D. Cluzeau, J. J. Brun, A review of earthworm impact on soil function and ecosystem services. *Eur. J. Soil Sci.* **64**, 161–182 (2013).
13. D. Craven, M. P. Thakur, E. K. Cameron, L. E. Frelich, R. Beausï¿½jour, R. B. Blair, B. Blossey, J. Burtis, A. Choi, A. Dï¿½valos, T. J. Fahey, N. A. Fisichelli, K. Gibson, I. T. Handa, K. Hopfensperger, S. R. Loss, V. Nuzzo, J. C. Maerz, T. Sackett, B. C. Scharenbroch, S. M. Smith, M. Vellend, L. G. Umek, N. Eisenhauer, The unseen invaders: introduced earthworms as drivers of change in plant communities in North American forests (a meta-analysis). *Glob. Chang. Biol.* **23**, 1065–1074 (2017).
14. Supplementary Materials and Methods.
15. M. Rutgers, A. Orgiazzi, C. Gardi, J. Römbke, S. Jänsch, A. M. Keith, R. Neilson, B. Boag, O. Schmidt, A. K. Murchie, R. P. Blackshaw, G. Pérès, D. Cluzeau, M. Guernion, M. J. I. Briones, J. Rodeiro, R. Piñeiro, D. J. D. Cosín, J. P. Sousa, M. Suhadolc, I. Kos, P. H. Krogh, J. H. Faber, C. Mulder, J. J. Bogte, H. J. va. Wijnen, A. J. Schouten, D. de Zwart, Mapping earthworm communities in Europe. *Appl. Soil Ecol.* **97**, 98–111 (2016).
16. P. F. Hendrix, P. J. Bohlen, Exotic earthworm invasions in North America: Ecological and policy implications. *Bioscience*. **52**, 801–811 (2002).
17. T. G. Piearce, The calcium relations of selected lumbricidae. *J. Anim. Ecol.* **41**, 167 (1972).
18. D. J. Spurgeon, A. M. Keith, O. Schmidt, D. R. Lammertsma, J. H. Faber, Land-use and land-

management change: relationships with earthworm and fungi communities and soil structural properties. *BMC Ecol.* **13**, 46 (2013).
19. J. Mathieu, T. J. Davies, Glaciation as an historical filter of below-ground biodiversity. *J. Biogeogr.* **41**, 1204–1214 (2014).
20. P. Lavelle, C. Lattaud, D. Trigo, I. Barois, Mutualism and biodiversity in soils. *Plant Soil*. **170**, 23–33 (1995).
21. R. R. Dunn, D. Agosti, A. N. Andersen, X. Arnan, C. A. Bruhl, X. Cerdá, A. M. Ellison, B. L. Fisher, M. C. Fitzpatrick, H. Gibb, N. J. Gotelli, A. D. Gove, B. Guenard, M. Janda, M. Kaspari, E. J. Laurent, J.-P. Lessard, J. T. Longino, J. D. Majer, S. B. Menke, T. P. McGlynn, C. L. Parr, S. M. Philpott, M. Pfeiffer, J. Retana, A. V. Suarez, H. L. Vasconcelos, M. D. Weiser, N. J. Sanders, Climatic drivers of hemispheric asymmetry in global patterns of ant species richness. *Ecol. Lett.* **12**, 324–333 (2009).
22. H. Kreft, W. Jetz, Global patterns and determinants of vascular plant diversity. *Proc. Natl. Acad. Sci.* **104**, 5925–5930 (2007).
23. C. Fragoso, P. Lavelle, Earthworm communities of tropical rain forests. *Soil Biol. Biochem.* (1992), doi:10.1016/0038-0717(92)90124-G.
24. K. J. Gaston, T. M. Blackburn, *Pattern and process in macroecology* (2007).
25. D. Song, K. Pan, A. Tariq, F. Sun, Z. Li, X. Sun, L. Zhang, O. A. Olusanya, X. Wu, Large-scale patterns of distribution and diversity of terrestrial nematodes. *Appl. Soil Ecol.* **114**, 161–169 (2017).
26. M. Maraun, H. Schatz, S. Scheu, Awesome or ordinary? Global diversity patterns of oribatid mites. *Ecography (Cop.).* **30**, 209–216 (2007).
27. J. Davison, L. Ainsaar, S. Burla, a G. Diedhiou, I. Hiiesalu, T. Jairus, N. C. Johnson, A. Kane, K. Koorem, M. Kochar, C. Ndiaye, R. Singh, M. Vasar, M. Zobel, M. Moora, M. Öpik, A. Adholeya, L. Ainsaar, A. Bâ, S. Burla, a G. Diedhiou, I. Hiiesalu, T. Jairus, N. C. Johnson, A. Kane, K. Koorem, M. Kochar, C. Ndiaye, M. Pärtel, Ü. Reier, Ü. Saks, R. Singh, M. Vasar, M. Zobel, Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science (80-.).* **127**, 970–973 (2015).
28. T. Decaëns, D. Porco, S. W. James, G. G. Brown, V. Chassany, F. Dubs, L. Dupont, E. Lapied, R. Rougerie, J. Rossi, V. Roy, DNA barcoding reveals diversity patterns of earthworm communities in remote tropical forests of French Guiana. *Soil Biol. Biochem.* **92**, 171–183 (2016).
29. D. C. Coleman, D. A. Crossley, P. F. Hendrix, *Fundamentals of Soil Ecology* (Elsevier, 2004; http://www.sciencedirect.com/science/article/pii/B9780121797263500095).
30. J. W. Spaak, J. M. Baert, D. J. Baird, N. Eisenhauer, L. Maltby, F. Pomati, V. Radchuk, J. R. Rohr, P. J. Van den Brink, F. De Laender, Shifts of community composition and population density substantially affect ecosystem function despite invariant richness. *Ecol. Lett.* **20**, 1315–1324 (2017).
31. N. Eisenhauer, J. Schlaghamerský, P. B. Reich, L. E. Frelich, The wave towards a new steady state: Effects of earthworm invasion on soil microbial functions. *Biol. Invasions*. **13**, 2191–2196 (2011).
32. M. Drumond, A. Guimarães, H. El Bizri, L. Giovanetti, D. Sepúlveda, R. Martins, Life history, distribution and abundance of the giant earthworm Rhinodrilus alatus RIGHI 1971: conservation and management implications. *Brazilian J. Biol.* **73**, 699–708 (2013).
33. L. Santini, N. J. B. Isaac, L. Maiorano, G. F. Ficetola, M. A. J. Huijbregts, C. Carbone, W. Thuiller, Global drivers of population density in terrestrial vertebrates. *Glob. Ecol. Biogeogr.* **27**, 968–979 (2018).
34. Intergovernmental Panel on Climate Change, *Climate Change 2014 Synthesis Report Summary Chapter for Policymakers* (2014).
35. D. K. Hackenberger, B. K. Hackenberger, Earthworm community structure in grassland habitats differentiated by climate type during two consecutive seasons. *Eur. J. Soil Biol.* **61**, 27–34 (2014).
36. M. A. Bradford, G. F. Ciska, A. Bonis, E. M. Bradford, A. T. Classen, J. H. C. Cornelissen, T. W.

Crowther, J. R. De Long, G. T. Freschet, P. Kardol, M. Manrubia-Freixa, D. S. Maynard, G. S. Newman, R. S. P. Logtestijn, M. Viketoft, D. A. Wardle, W. R. Wieder, S. A. Wood, W. H. Van Der Putten, A test of the hierarchical model of litter decomposition. *Nat. Ecol. Evol.* **1**, 1836–1845 (2017).

37. A. Rice, P. Šmarda, M. Novosolov, M. Drori, L. Glick, N. Sabath, S. Meiri, J. Belmaker, I. Mayrose, The global biogeography of polyploid plants. *Nat. Ecol. Evol.* **3**, 265–273 (2019).
38. A. Shade, R. R. Dunn, S. A. Blowes, P. Keil, B. J. M. Bohannan, M. Herrmann, K. Küsel, J. T. Lennon, N. J. Sanders, D. Storch, J. Chase, Macroecology to unite all life, large and small. *Trends Ecol. Evol.* **33**, 731–744 (2018).
39. J. M. Anderson, J. S. I. Ingram, Tropical Soil Biology and Fertility: A handbook of methods. *Trop. Soil Biol. Fertil. A Handb. methods*. **2 Ed.**, 88–91 (1993).
40. ISO, "Soil quality - Sampling of soil invertebrates - Part 1: Hand-sorting and extraction of earthworms (ISO/FDIS 23611-1:2018)" (2018).
41. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, Fiji: an open-source platform for biological-image analysis. *Nat. Methods*. **9**, 676–682 (2012).
42. J. Koricheva, J. Gurevitch, K. Mengersen, *Handbook of meta-analysis in ecology and evolution* (2013; https://books.google.com.br/books?hl=pt-BR&lr=&id=l3oXBPrOkuYC&oi=fnd&pg=PP2&dq=handbook+of+meta-analysis+in+ecology+and+evolution&ots=GOIZ9RS2A_&sig=c1sBss0NniRZGyAhmryg8ZiL_iE#v=onepage&q&f=false).
43. M. D. Bartlett, M. J. I. Briones, R. Neilson, O. Schmidt, D. Spurgeon, R. E. Creamer, A critical review of current methods in earthworm ecology: From individuals to populations. *Eur. J. Soil Biol.* **46**, 67–73 (2010).
44. M. J. Crawley, *The R book* (John Wiley & Sons, 2012).
45. M. B. Bouché, Strategies lombriciennes. *Ecol. Bull.*, 122–132 (1977).
46. G. G. Brown, How do earthworms affect microfloral and faunal community diversity? *Plant Soil*. **170**, 209–231 (1995).
47. J. Seeber, G. U. H. Seeber, R. Langel, S. Scheu, E. Meyer, The effect of macro-invertebrates and plant litter of different quality on the release of N from litter to plant on alpine pastureland. *Biol. Fertil. Soils*. **44**, 783–790 (2008).
48. M. Blouin, Y. Zuily-Fodil, A. T. Pham-Thi, D. Laffray, G. Reversat, A. Pando, J. Tondoh, P. Lavelle, Belowground organism activities affect plant aboveground phenotype, inducing plant tolerance to parasites. *Ecol. Lett.* **8**, 202–208 (2005).
49. J. Boyer, G. Reversat, P. Lavelle, A. Chabanne, European Journal of Soil Biology Interactions between earthworms and plant-parasitic nematodes. *Eur. J. Soil Biol.* **59**, 43–47 (2013).
50. G. Loranger-Merciris, Y.-M. Cabidoche, B. Deloné, P. Quénéhervé, H. Ozier-Lafontaine, How earthworm activities affect banana plant response to nematodes parasitism. *Appl. Soil Ecol.* **52**, 1–8 (2012).
51. G. G. Brown, E. Soja, C. A. Edwards, L. Brussaard, in *Earthworm Ecology, Second Edition* (2004), pp. 13–49.
52. M. B. Bouché, F. Al-Addan, Earthworms, water infiltration and soil stability: Some new assessments. *Soil Biol. Biochem.* **29**, 441–452 (1997).
53. M. Joschko, H. Diestel, O. Larink, Assessment of earthworm burrowing efficiency in compacted soil with a combination of morphological and soil physical measurements. *Biol. Fertil. Soils*. **8**, 191–196 (1989).
54. T. Hengl, J. Mendes de Jesus, G. B. M. Heuvelink, M. Ruiperez Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, M. N. Wright, X. Geng, B. Bauer-Marschallinger, M. A. Guevara, R. Vargas, R. A. MacMillan, N. H. Batjes, J. G. B. Leenaars, E. Ribeiro, I. Wheeler, S. Mantel, B. Kempen, SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*. **12**, e0169748 (2017).

55. D. N. Karger, O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. E. Zimmermann, H. P. Linder, M. Kessler, Climatologies at high resolution for the earth's land surface areas. *Sci. Data*. **4**, 170122 (2017).
56. D. K. Hall, G. A. Riggs, MODIS/Terra Snow Cover Monthly L3 Global 0.05Deg CMG, Version 6. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: https://doi.org/10.5067/MODIS/MOD10CM.006. [Accessed in January 2018] (2015).
57. R. J. Zomer, A. Trabucco, D. A. Bossio, L. V. Verchot, Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agric. Ecosyst. Environ.* **126**, 67–80 (2008).
58. R. J. Zomer, D. A. Bossio, A. Trabucco, L. Yuanjie, D. C. Gupta, V. P. Singh, Trees and Water: Smallholder Agroforestry on Irrigated Lands in Northern India. *IWMI Res. Rep. 122*, 45 (2007).
59. J. Danielson, D. Gesch, "Global Multi-resolution Terrain Elevation Data 2010(GMTED2010)" (2011), , doi:citeulike-article-id:13365221.
60. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
61. A. F. Zuur, E. N. Ieno, C. S. Elphick, A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.* **1**, 3–14 (2010).
62. N. Eisenhauer, A. Stefanski, N. A. Fisichelli, K. Rice, R. Rich, P. B. Reich, Warming shifts "worming": Effects of experimental warming on invasive earthworms in northern North America. *Sci. Rep.* **4**, 4–10 (2014).
63. M. Nieminen, E. Ketoja, J. Mikola, J. Terhivuo, T. Siren, V. Nuutinen, Local land use effects and regional environmental limits on earthworm communities in Finnish arable landscapes. *Ecol. Appl.* **21**, 3162–3177 (2011).
64. A. F. Zuur, E. N. Ieno, A. A. Saveliev, *Mixed Effects Models and Extensions in Ecology with R* (Springer, 2009).
65. S. Dray, A.-B. Dufour, The ade4 Package: Implementing the Duality Diagram for Ecologists. *J. Stat. Softw.* **22**, 1–20 (2007).
66. L. Breiman, Random forests. *Mach. Learn.* **45**, 5–32 (2001).
67. A. Liaw, M. Wiener, Classification and regression by randomForest. *R news*. **2**, 18–22 (2002).
68. U. Grömping, Variable importance assessment in regression: Linear regression versus random forest. *Am. Stat.* **63**, 308–319 (2009).
69. C. Strobl, A. Boulesteix, A. Zeileis, T. Hothorn, Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*. **8**, 25 (2007).
70. B. H. Menze, B. M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, F. A. Hamprecht, A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics*. **10**, 213 (2009).
71. G. James, D. Witten, T. Hastie, R. Tibshirani, *An introduction to statistical learning* (Springer, 2013), vol. 112.
72. K. Barton, MuMIn: Multi-Model Inference. *R Packag. version 1.42.1. https//CRAN.R-project.org/package=MuMIn. Bates*, (2018).
73. N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, R. Moore, Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* (2017), doi:10.1016/j.rse.2017.06.031.
74. R Core Team, *R: A language and environment for statistical computing* (Vienna, Austria, 2016).
75. P. Lavelle, A. V Spain, *Soil ecology* (Springer Science & Business Media, 2001).
76. P. A. Sanchez, S. Ahamed, F. Carre, A. E. Hartemink, J. Hempel, J. Huising, P. Lagacherie, A. B. McBratney, N. J. McKenzie, M. d. L. Mendonca-Santos, B. Minasny, L. Montanarella, P. Okoth, C. A. Palm, J. D. Sachs, K. D. Shepherd, T.-G. Vagen, B. Vanlauwe, M. G. Walsh, L. A. Winowiecki, G.-L. Zhang, Digital Soil Map of the World. *Science (80-.).* **325**, 680–681 (2009).

Acknowledgments

We thank all the reviewers who provided thoughtful and constructive feedback on this manuscript. We thank Marten Winter and the sDiv team for their help in organizing the sWorm workshops, and the Biodiversity Informatics Unit (BDU) at iDiv for their assistance in making the data open access. In addition, the data providers would like to thank all supervisors, students, collaborators, technicians, data analysts, land owners/managers and anyone else involved with the collection, processing and/or publication of the primary datasets.

Funding: This work was developed during and following two 'sWorm' workshops. HRPP and the sWorm workshops were supported by the sDiv (Synthesis Centre of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig (DFG FZT 118)). HRPP, OF and NE acknowledge funding by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 677232 to NE). CAG and NE acknowledge funding by iDiv (DFG FZT118) Flexpool proposal 34600850. EKC acknowledges funding from the Academy of Finland (285882) and the Natural Sciences and Engineering Research Council of Canada (postdoctoral fellowship and RGPIN-2019-05758). TWC, JVDH and DR were supported by DOB Ecology. MR was supported by ERC-AdG 694368. ML was supported by the TULIP Laboratory of Excellence (ANR-10-LABX-41). KSR and WvdP were supported by (ERC-ADV grant 323020 to WvdP). In addition, data collection was funded by: Russian Foundation for Basic Research (12-04-01538-a, 12-04-01734-a, 14-44-03666-r_center_a, 15-29-02724-ofi_m, 16-04-01878-a 19-05-00245), Tarbiat Modares University, Aurora Organic Dairy, UGC(NERO) (F. 1-6/Acctt./NERO/2007-08/1485), Natural Sciences and Engineering Research Council (RGPIN-2017-05391), Slovak Research and Development Agency (APVV-0098-12), Science for Global Development through Wageningen University, Norman Borlaug LEAP Programme and International Atomic Energy Agency (IAEA), São Paulo Research Foundation - FAPESP (12/22510-8), Oklahoma Agricultural Experiment Station, INIA - Spanish Agency (SUM 2006-

00012-00-0), Royal Canadian Geographical Society, Environmental Protection Agency (Ireland) (2005-S-LS-8), University of Hawai'i at Mānoa (HAW01127H; HAW01123M), CAPES, European Union FP7 (FunDivEurope, 265171), U.S. Department of the Navy, Commander Pacific Fleet (W9126G-13-2-0047), Science and Engineering Research Board (SB/SO/AS-030/2013) Department of Science and Technology, New Delhi, India, Strategic Environmental Research and Development Program (SERDP) of the U.S. Department of Defense (RC-1542), Maranhão State Research Foundation (FAPEMA), Coordination for the Improvement of Higher Education Personnel (CAPES), Ministry of Education, Youth and Sports of the Czech Republic (LTT17033), Colorado Wheat Research Foundation, Zone Atelier Alpes, French National Research Agency (ANR-11-BSV7-020-01, ANR-09-STRA-02-01, ANR 06 BIODIV 009-01), Austrian Science Fund (P16027, T441), Landwirtschaftliche Rentenbank Frankfurt am Main, Welsh Government and the European Agricultural Fund for Rural Development (Project Ref. A AAB 62 03 qA731606), SÉPAQ, Ministry of Agriculture and Forestry of Finland, McKnight Foundation, Science Foundation Ireland (EEB0061), University of Toronto (Faculty of Forestry), National Science and Engineering Research Council of Canada, Haliburton Forest & Wildlife Reserve, NKU College of Arts & Sciences Grant, Österreichische Forschungsförderungsgesellschaft (837393 and 837426), Mountain Agriculture Research Unit of the University of Innsbruck, Higher Education Commission of Pakistan, Kerala Forest Research Institute, Peechi, Kerala, UNEP/GEF/TSBF-CIAT Project on Conservation and Sustainable Management of Belowground Biodiversity, Ministry of Agriculture and Forestry of Finland, Complutense University of Madrid / European Union FP7 project BioBio (FPU UCM 613520), GRDC, AWI, LWRRDC, DRDC, CONICET (National Scientific and Technical Research Council) and FONCyT (National Agency of Scientific and Technological Promotion) (PICT, PAE, PIP), Universidad Nacional de Luján y FONCyT (PICT 2293 (2006)), Fonds de recherche sur la nature et les technologies du Québec (131894), Deutsche Forschungsgemeinschaft (SCHR1000/3-1, SCHR1000/6-1, 6-2 (FOR 1598), WO 670/7-1, WO 670/7-2, & SCHA 1719 / 1-2), CONACYT (FONDOS MIXTOS TABASCO/PROYECTO11316), NSF (DGE-0549245, DGE-0549245, DEB-BE-0909452, NSF1241932), Institute for Environmental Science and Policy at the University of Illinois at Chicago,

Dean's Scholar Program at UIC, Garden Club of America Zone VI Fellowship in Urban Forestry from the Casey Tree Endowment Fund, J.E. Weaver Competitive Grant from the Nebraska Chapter of The Nature Conservancy, The College of Liberal Arts and Sciences at Depaul University, Elmore Hadley Award for Research in Ecology and Evolution from the UIC Dept. of Biological Sciences, Spanish CICYT (AMB96-1161; REN2000-0783/GLO; REN2003-05553/GLO; REN2003-03989/GLO; CGL2007-60661/BOS), Yokohama National University, MEXT KAKENHI (25220104), Japan Society for the Promotion of Science KAKENHI (25281053, 17KT0074, 25252026), ADEME (0775C0035), Ministry of Science, Innovation and Universities of Spain (CGL2017-86926-P), Syngenta Philippines, UPSTREAM, LTSER (Val Mazia/Matschertal), Marie Sklodowska Curie Postdoctoral Fellowship (747607), National Science & Technology Base Resource Survey Project of China (2018FY100306), McKnight Foundation (14-168), Program of Fundamental Researches of Presidium of Russian Academy of Sciences , Brazilian National Council of Research CNPq, French Ministry of Foreign and European Affairs

Author contributions: HRPP led the analysis, data curation, and writing of the original manuscript draft. CAG assisted in analyses, and writing the original manuscript draft. EKC and NE revised subsequent manuscript drafts. EKC, NE and MPT acquired funding for the project. JK, KBG, BS, MLCB, MJIB, and GB contributed to data curation. HRPP, CAG, MLCB, MJIB, GB, OF, AO, EMB, JB, UB, TD, FTDeV, BK-R, ML, JM, CM, WHvdP, KSR, MCR, DR, MR, MPT, DHW, DW, EKC and NE contributed to the project conceptualisation. All authors reviewed and edited the final draft manuscript. The majority of the authors provided data for the analyses.

Competing interests: The authors declare no competing interests.

Data and materials availability: Data and analysis code is available on the iDiv Data repository (DOI: 10.25829/idiv.1804-5-2593), and GitHub (https://github.com/helenphillips/GlobalEWDiversity; DOI: 10.5281/zenodo.3386456)

Supplementary Materials

Materials and methods

Supplementary text

Figures S1 – S6

Table S1 – S4

References 38 to 76

Captions

Fig. 1: Global distribution of earthworm diversity. (A) Black dots represent the centre of a 'study' used in at least one of the three models (species richness, total abundance, and total biomass). The size of the dot corresponds to the number of sites within the study. Opaqueness is for visualisation purposes only. (B-D): The globally predicted values of (B) species richness (within site), (C) total abundance (individuals per m^2), and (D) total biomass (grams per m^2). Yellow indicates high diversity dark purples low diversity. Grey areas are habitat cover categories which lacked samples.

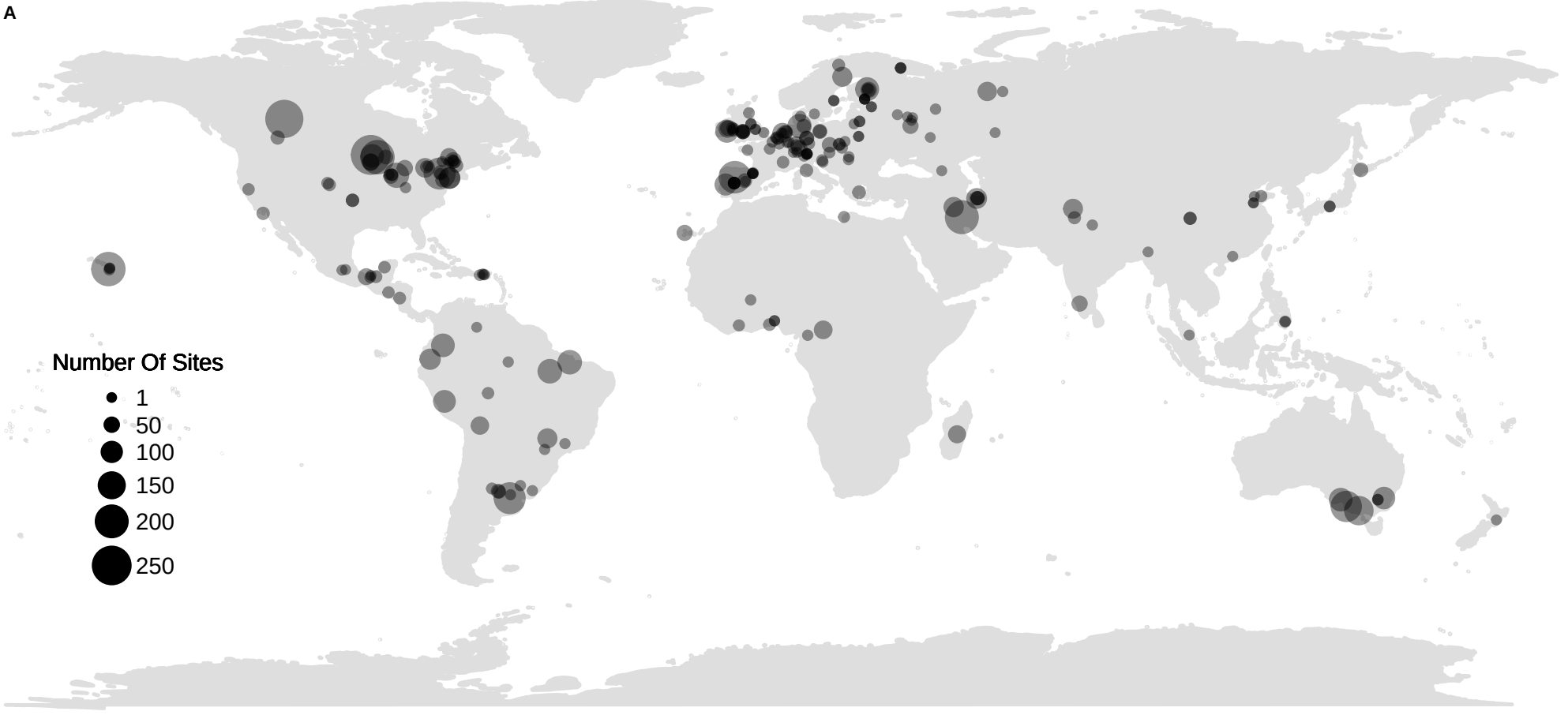
Fig. 2: The number of unique species within each latitudinal zone, when the number of sites within each zone was comparable. The width of the bar shows the latitude range of the sites/zones.

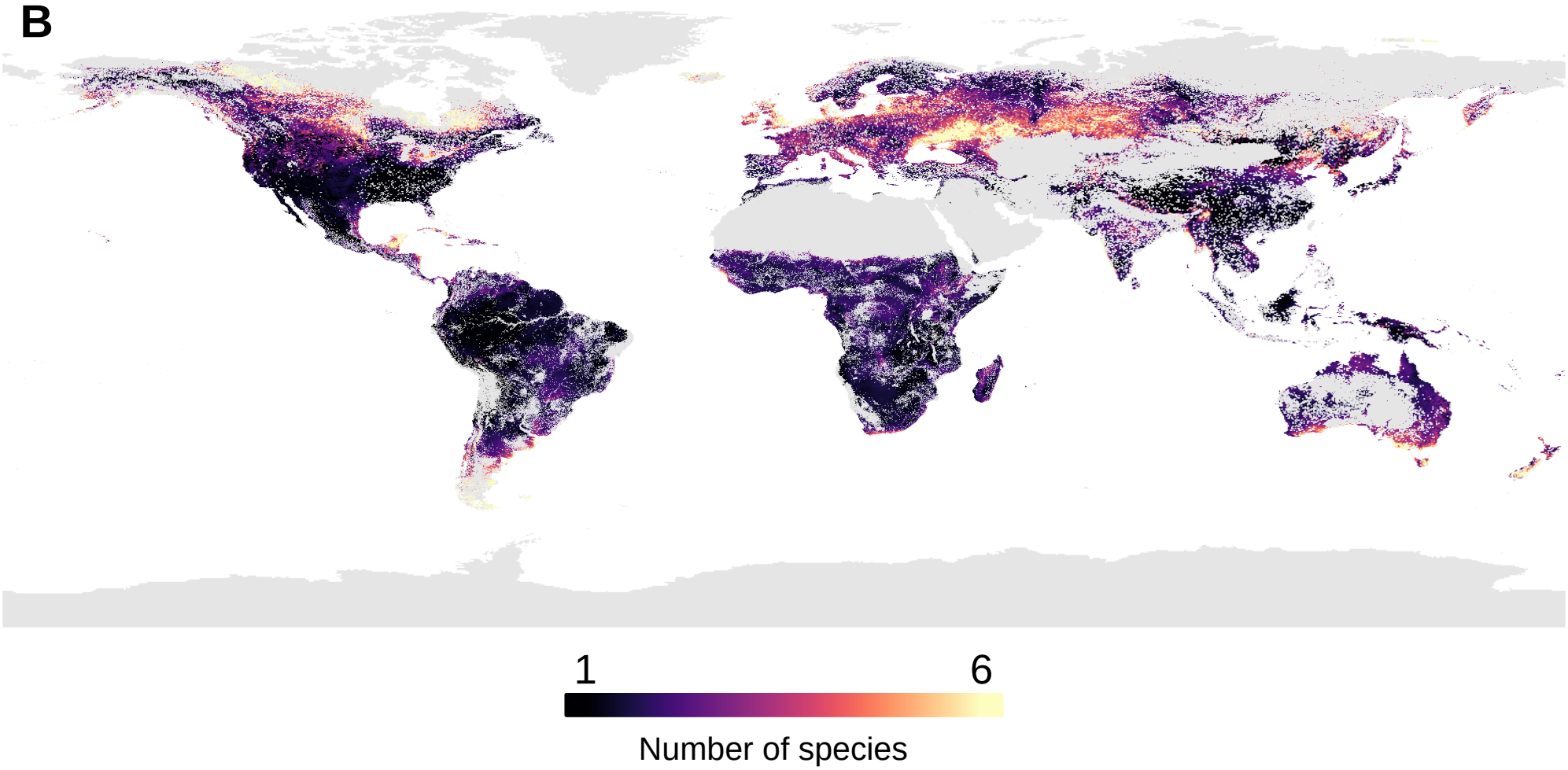
Fig. 3: T**he importance of the six variable themes from the three biodiversity models**. Rows show the results of each model (top: species richness, middle: abundance, bottom: biomass). Columns represent the theme of variables that was present in the simplified biodiversity model. The most important variable group has the largest circle. Within each row, the circle size of the other variable themes scale depending on the relative change in importance. The circle size should only be compared within a row.

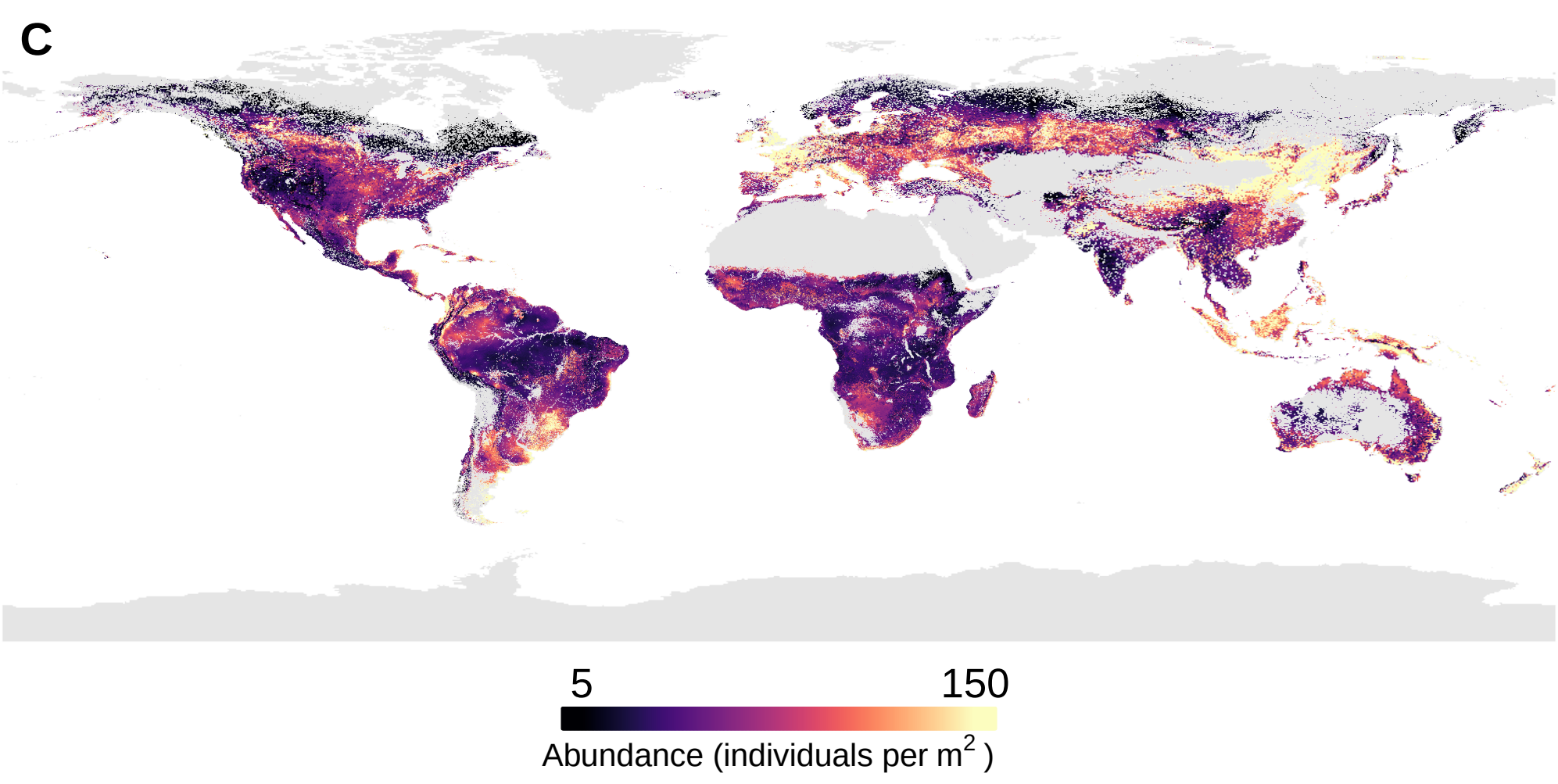
Table 1: **Model validation results**. Highlighted cells show the 'best' value when comparing between the main models (a mixture of sampled soil properties and SoilGrids data) and models containing only SoilGrids data. The mean square error (MSE) following 10-fold cross-validation of (A) the main models and (B) models containing only SoilGrids data. MSE was calculated

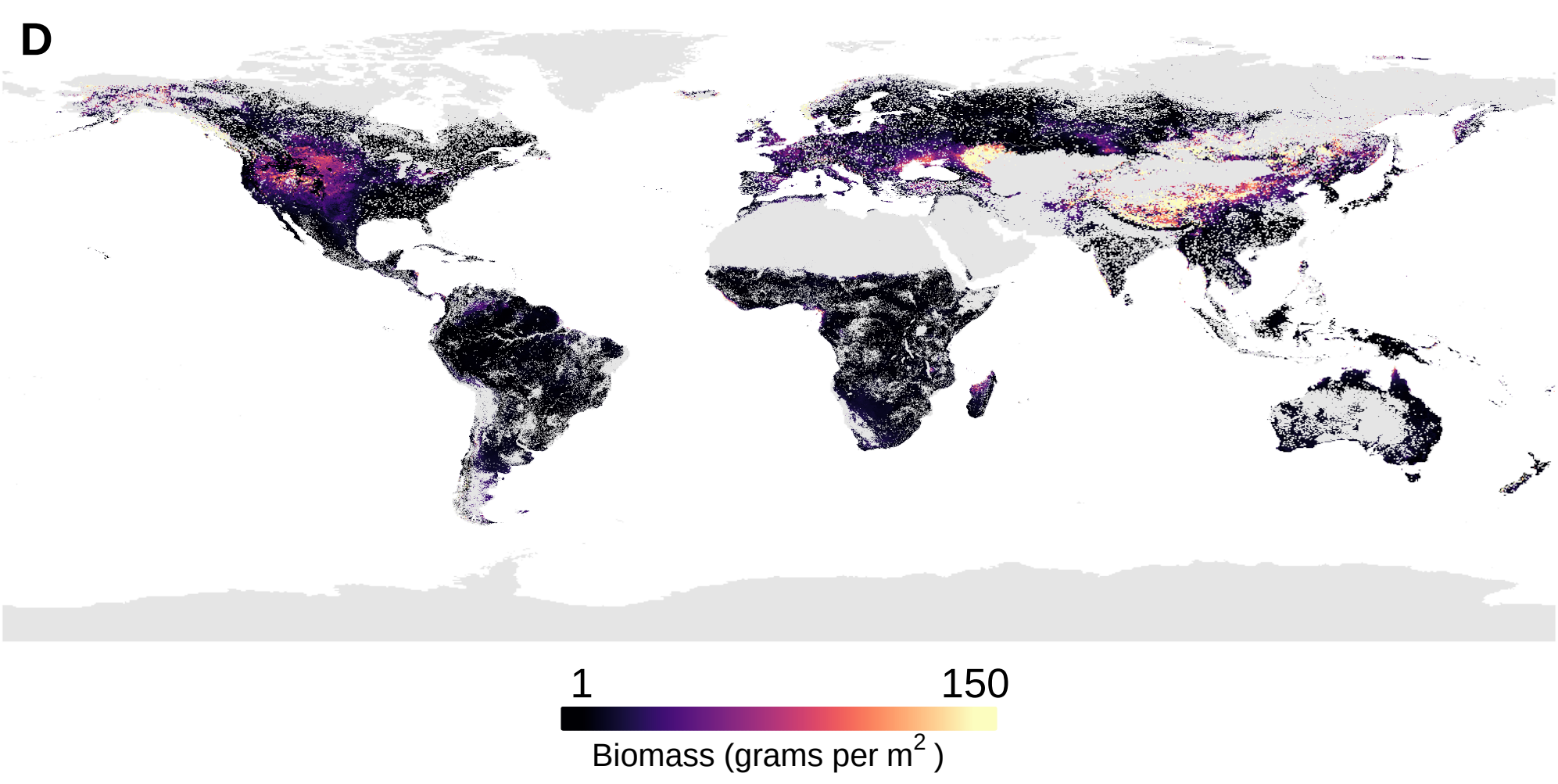
for all predicted data ('Total'), and for tertiles ('Low', 'Mid', 'High') of the observed data. In addition, the R^2 of (C) the main models and (D) SoilGrids-only models.

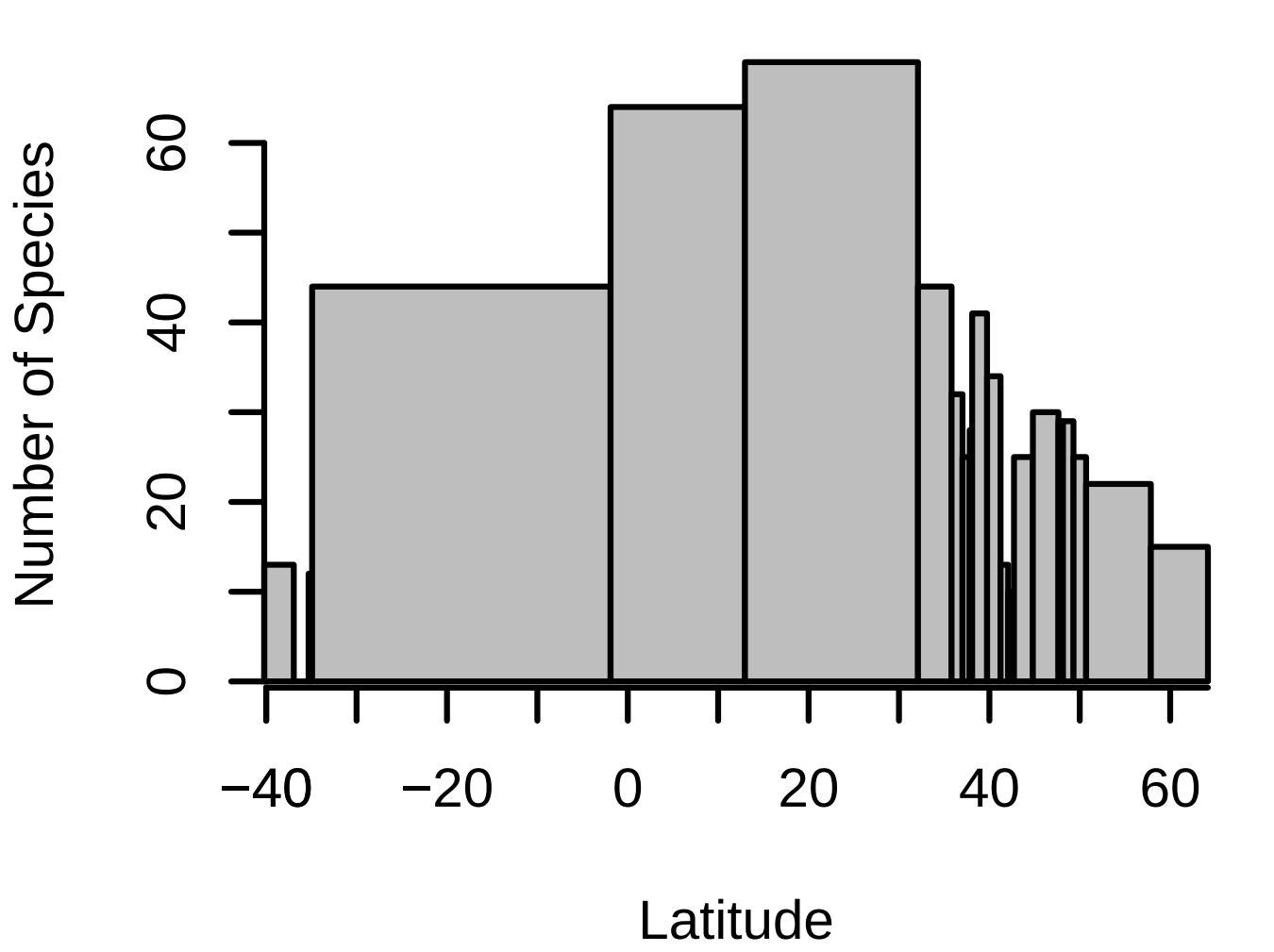
	Total	Low	Mid	High
A) MSE – Main Models				
Species Richness	1.376	0.917	0.812	3.561
Abundance	17977.42	1720.75	2521.25	48751.51
Biomass	3220.29	264.56	441.25	8783.77
B) MSE – SoilGrids Models				
Species Richness	1.385	0.887	0.793	3.716
Abundance	18775.81	1735.11	2516.13	51156.76
Biomass	3068.00	199.91	461.88	8380.81
	Marginal		Conditional	
C) R^2 – Main Models.				
Species Richness	0.132		0.748	
Abundance	0.176		0.626	
Biomass	0.201		0.612	
D) R^2 – SoilGrids Models				
Species Richness	0.142		0.745	
Abundance	0.234		0.643	
Biomass	0.242		0.650	

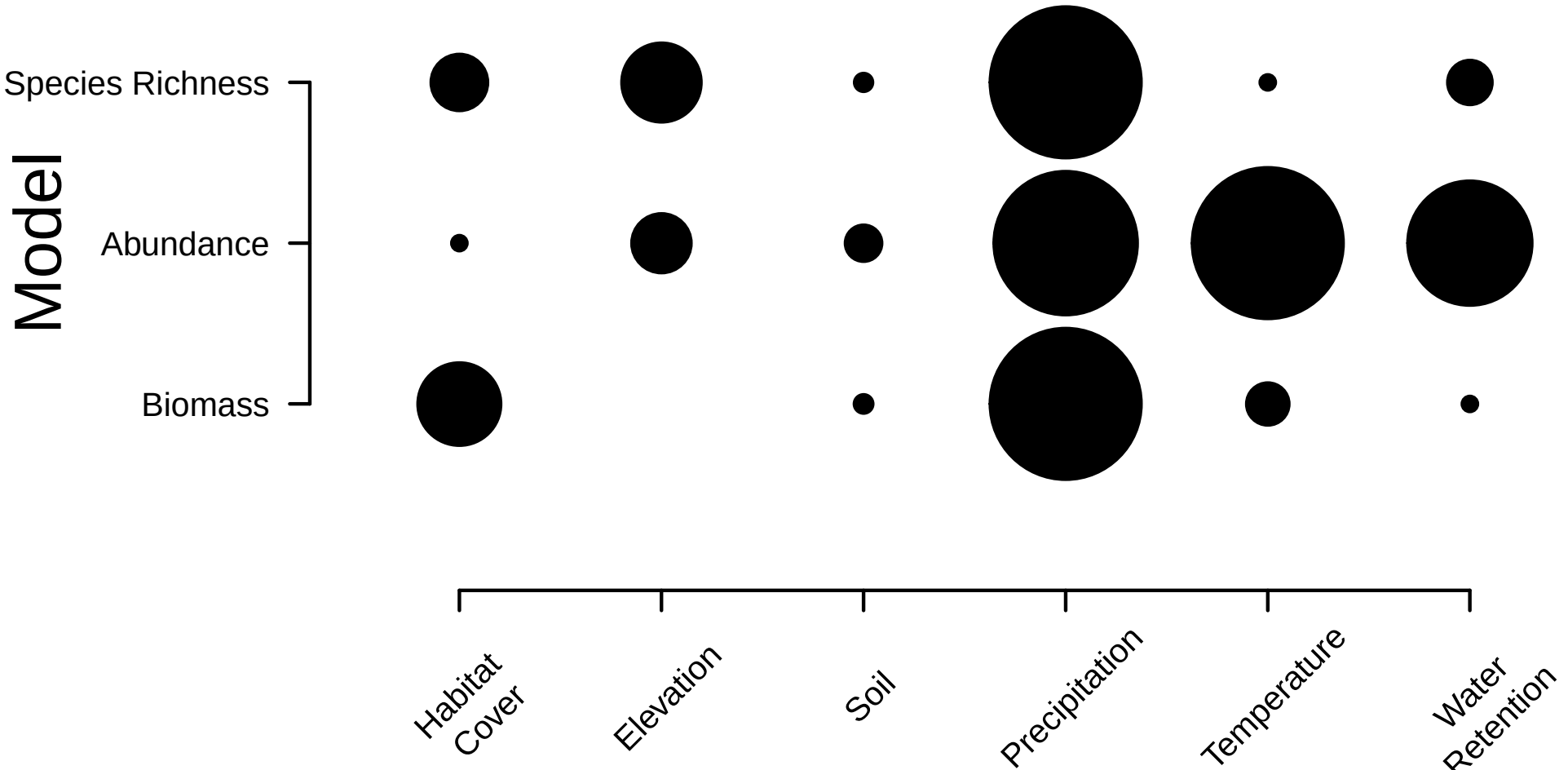

A
Number Of Sites
1
50
100
150
200
250


B
1
6
Number of species


C
5
150
Abundance (individuals per m2)


D
1
150
Biomass (grams per m2)


Number of Species
0
20
40
60
−40
−20
0
20
40
60
Latitude


Model
Species Richness
Abundance
Biomass
Habitat Cover
Elevation
Soil
Precipitation
Temperature
Water Retention

Science
AAAS

Supplementary Materials for

Global distribution of earthworm diversity

Helen R. P. Phillips, Carlos A. Guerra, Marie L. C. Bartz, Maria J. I. Briones, George Brown, Thomas W. Crowther, Olga Ferlian, Konstantin B. Gongalsky, Johan van den Hoogen, Julia Krebs, Alberto Orgiazzi, Devin Routh, Benjamin Schwarz, Elizabeth M. Bach, Joanne Bennett, Ulrich Brose, Thibaud Decaëns, Birgitta König-Ries, Michel Loreau, Jérôme Mathieu, Christian Mulder, Wim H. van der Putten, Kelly S. Ramirez, Matthias C. Rillig, David Russell, Michiel Rutgers, Madhav P. Thakur, Franciska T. de Vries, Diana H. Wall, David A. Wardle, Miwa Arai, Fredrick O. Ayuke, Geoff H. Baker, Robin Beauséjour, José C. Bedano, Klaus Birkhofer, Eric Blanchart, Bernd Blossey, Thomas Bolger, Robert L. Bradley, Mac A. Callaham, Yvan Capowiez, Mark E. Caulfield, Amy Choi, Felicity V. Crotty, Andrea Dávalos, Darío J. Diaz Cosin, Anahí Dominguez, Andrés Esteban Duhour, Nick van Eekeren, Christoph Emmerling, Liliana B. Falco, Rosa Fernández, Steven J. Fonte, Carlos Fragoso, André L. C. Franco, Martine Fugère, Abegail T. Fusilero, Shaieste Gholami, Michael J. Gundale, Mónica Gutiérrez López, Davorka K. Hackenberger, Luis M. Hernández, Takuo Hishi, Andrew R. Holdsworth, Martin Holmstrup, Kristine N. Hopfensperger, Esperanza Huerta Lwanga, Veikko Huhta, Tunsisa T. Hurisso, Basil V. Iannone III, Madalina Iordache, Monika Joschko, Nobuhiro Kaneko, Radoslava Kanianska, Aidan M. Keith, Courtland A. Kelly, Maria L. Kernecker, Jonatan Klaminder, Armand W. Koné, Yahya Kooch, Sanna T. Kukkonen, Hmar Lalthanzara, Daniel R. Lammel, Iurii M. Lebedev, Yiqing Li, Juan B. Jesus Lidon, Noa K. Lincoln, Scott R. Loss, Raphael Marichal, Radim Matula, Jan Hendrik Moos, Gerardo Moreno, Alejandro Morón-Ríos, Bart Muys, Johan Neirynck, Lindsey Norgrove, Marta Novo, Visa Nuutinen, Victoria Nuzzo, Mujeeb Rahman P, Johan Pansu, Shishir Paudel, Guénola Pérès, Lorenzo Pérez-Camacho, Raúl Piñeiro, Jean-François Ponge, Muhammad Imtiaz Rashid, Salvador Rebollo, Javier Rodeiro-Iglesias, Miguel Á. Rodríguez, Alexander M. Roth, Guillaume X. Rousseau, Anna Rozen, Ehsan Sayad, Loes van Schaik, Bryant C. Scharenbroch, Michael Schirrmann, Olaf Schmidt, Boris Schröder, Julia Seeber, Maxim P. Shashkov, Jaswinder Singh, Sandy M. Smith, Michael Steinwandter, José A. Talavera, Dolores Trigo, Jiro Tsukamoto, Anne W. de Valença, Steven J. Vanek, Iñigo Virto, Adrian A. Wackett, Matthew W. Warren, Nathaniel H. Wehr, Joann K. Whalen, Michael B. Wironen, Volkmar Wolters, Irina V. Zenkova, Weixin Zhang, Erin K. Cameron, Nico Eisenhauer

Correspondence to: helen.phillips@idiv.de

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S6
Tables S1 to S4

Materials and Methods

Literature Search

Web of Science was searched on 18th December 2016, using the following search term: *((Earthworm* OR Oligochaeta OR Megadril* OR Haplotaxida OR Annelid* OR Lumbric* OR Clitellat* OR Acanthodrili* OR Ailoscoleci* OR Almid* OR Benhamiin* OR riodrilid* OR Diplocard* OR Enchytraeid* OR Eudrilid* OR Exxid* OR Glossoscolecid* OR Haplotaxid* OR Hormogastrid* OR Kynotid* OR Lutodrilid* OR Megascolecid* OR Microchaetid* OR Moniligastrid* OR Ocnerodrilid* OR Octochaet* OR Sparganophilid* OR Tumakid*) AND (Diversity OR "Species richness" OR "OTU" OR Abundance OR individual* OR Density OR "tax* richness" OR "Number" OR Richness OR Biomass))*

This search returned 7783 papers. All titles and abstracts of papers post-2000 were screened (6140 papers), and were excluded if they did not reference data suitable for the analysis (suitability discussed below). Since it was anticipated that raw data would need to be requested, papers published before 2000 were not screened, as it was unlikely that available author contact details were up-to-date. We note however that earlier publications may be useful for future research, e.g., focusing on long-term monitoring and temporal analyses. After this initial screening, PDFs of all remaining papers (n = 986) were manually screened to determine whether data were suitable.

In order to be suitable for the analysis, papers had to present (or make reference to) the following information and data:

1. Sampled earthworm communities using standard earthworm extraction methodologies, which would adequately capture quantitative information of the earthworm community, such as hand-sorting of a sufficient soil volume (e.g., 39) or chemical expulsion from a quadrat (e.g., 40) at two or more sites. At a minimum, total fresh biomass and/or total abundance of the earthworms at each site had to be measured. Ideally, there was data on identification of all individuals to species level, with the abundance/biomass data of each species;
2. Available geographic coordinates for all sampled sites, or maps that could be georeferenced;
3. Measurements of at least one soil property at each site (see below);
4. Information on the habitat cover and/or land use;
5. Differences in land use/habitat cover or soil properties (see below for information on the land use/habitat cover and soil properties) across the sites.

Where possible, all suitable data were taken from the 477 papers that were identified as containing suitable data. Data were extracted from figures where necessary (using IMAGEJ (39)). If data were not provided in the text or the supplementary materials, authors were contacted to obtain the raw data from each site. As some datasets remain unpublished, or are yet to be published, individual earthworm researchers were also contacted to enquire as to whether they had suitable data. Including unpublished data helps to reduce publication bias (42).

Data collation

The data taken or requested from one publication or an unpublished field campaign was considered a 'dataset'. If a dataset contained data sampled using different methodologies, we split it into different 'studies' based on the methodology, as measured diversity of earthworms is highly dependent on the methods used (43). For datasets where sites were repeatedly sampled

over time, both within years and across years, we used only the first and the last sampling campaign and these were split into two studies. The modelling approach used (linear mixed-effects models, with random effects accounting for different studies) dealt with non-independence of such datasets (44).

Site level information

Sites were described as a location of one or more samples, which, when taken together, adequately captured the earthworm community. Sampling methodology, and therefore the number of samples per site, were determined by the original data collectors. But sampling effort was constant within a study. For each dataset, we collated the following information into a standardised data template: geographic coordinates for each of the sampled sites, start and end dates of sampling (month and year), and the sampling method used. For each dataset, we requested at least one soil property (pH, cation exchange capacity (CEC) or base saturation, organic carbon, soil organic matter, C/N ratio, soil texture, soil type, soil moisture) for each site, but only pH, CEC, organic carbon and soil texture (silt and clay) variables were used for this analysis. Most sites contained pH values (63.7%), 14% of sites contained organic carbon, 40% of sites contained silt and clay, but only 7.3% contained CEC. Any missing soil properties were filled with SoilGrids data, described below. If soil properties were given for different soil depths, then we calculated a weighted average (maximum soil depth = 1 m, but typically collected down to 30 cm). Using information within the published articles, and additional information provided by the data collectors, the habitat cover at each site was classified into categories based on the ESA CCI-LC 300m map (http://maps.elie.ucl.ac.be/CCI/viewer/index.php; Table S1).

Recorded community metrics

For each dataset, the following site-level community metrics were calculated where possible: total (adults and juveniles) abundance of earthworms at the site, total (adults and juveniles) fresh biomass of earthworms at the site, and number of species at the site. Using the area sampled at the site, both abundance and biomass were transformed to individuals per m^2 and grams per m^2, respectively, if they were not already given in that unit, to standardize the data into commonly used units. Species richness of each site was calculated from available species lists, if not already provided. Two issues arose when calculating species richness of earthworms. Firstly, many specimens were not identified to species level. Where data collectors identified a specimen as a unique morphospecies (species delineation based solely on morphological characteristics, typically identified to genus level with a unique ID differentiating from other species of the same genus, as determined by the original data collector), they were included in the species richness estimate as an additional species. Records that were not identified to species level, or identified as a morphospecies, were excluded. Secondly, typically only adult specimens of many earthworm species can be identified to species level (43), so juveniles were excluded from the calculation. Therefore, a more appropriate term would be 'number of identified adult (morpho-) species', but for brevity this will be referred to as 'species richness'. Species richness was not calculated per unit area (i.e., density), as within each study the sampled area was consistent. Thus, due to the modelling framework, issues of diversity increasing with sampled area were accounted for.

Species identity

For datasets where the earthworms had been identified to species level, all species names were checked for spelling errors and synonyms. Scientific names were standardised using expert opinion (MJIB, GB, MLCB) and DriloBASE (http://drilobase.org/drilobase). Following standardisation, earthworm species were categorised into the three main ecological groups:

epigeics, endogeics, and anecics (45), plus a fourth minor group, epi-endogeic (species which exhibit traits of both epigeics and endogeics). Earthworms provide a variety of ecosystem functions, for example, increasing crop yield by enhancing decomposition and nutrient minerialization rates (12), but each ecological group contributes in different ways, often on the basis of their feeding or habitat preferences (45). Epigeic species are typically found in the upper layers of the soil and litter, and, amongst other roles, are important in the first stages of decomposition through burial of the litter layer (11, 46, 47). Endogeic species live in the mineral soil layers, creating horizontal burrows (45). One function they have been shown to provide is a decrease in the density of root-pathogenic nematodes (48, 49), reducing nematode populations and disease incidence, which can contribute to increased crop yields (50, 51). Anecic species mix the litter and mineral soil via surface cast production (45, 46). In addition, the vertical burrows created by anecic species increase water infiltration into deeper soil layers, increasing water holding capacity (52, 53), and regulating water availability.

Data extraction and harmonisation across global layers

In order to predict earthworm communities across the globe, we required harmonised sets of spatially distributed variables. We collected 15 globally distributed layers that are described as predictors of earthworm distribution (Table S2). For the SoilGrids data (17; https://soilgrids.org; modelled global layers of soil properties based on soil profiles and remotely-sensed products), which provides soil properties for different layers within the soil profile, we calculated the weighted average for the values of the top four layers (corresponding to the top 30 cm of the soil profile, which matches the soil depth of the earthworm sampling techniques). For sites missing one or more sampled soil properties, the soil properties associated with the 1km pixel corresponding to the site's geographical coordinates were used in the analyses. For CEC, for all sites, values were taken from SoilGrids.

Where possible, the land cover global layer (ESA CCI-LC 300 m; https://www.esa-landcover-cci.org/) was re-categorised to amalgamate similar habitat cover categories matching the ones collected within the dataset (see Table S1). Where not possible, the categories were ignored (i.e., classified as NA) during later steps, as estimates could not be produced for unknown habitat cover categories.

No climate variables were taken from the papers or raw data provided, as there was little consistency in climate variables across the papers. Instead, five global climate layers (climatologies) obtained from the CHELSA climate dataset (55) were used (annual mean temperature, temperature seasonality, temperature annual range, annual precipitation, and precipitation seasonality) and, from other sources, the number of months of snow cover (56), and the aridity index and potential evapotranspiration (PET; 57, 58). The within-year standard deviation of PET (PETSD) was calculated as well. Finally, a globally distributed layer of elevation (59) was also included in the analysis. For all of these layers, the value within the 1 km pixel that matched the site's coordinates was used in the analyses.

For an initial harmonisation across all global layers, it was necessary to aggregate or disaggregate - when appropriate - the spatial resolution of the different layers to match a one-kilometre square grid. A nearest neighbour disaggregation algorithm was applied without changing the pixel values, but changing the pixel resolution using the one-kilometre square resolution from SoilGrids as a reference.

Following the spatial harmonisation, the global layers were matched with the collated dataset based on the geographic coordinates of the sampled sites. In the case of the climate layers, all variables were appended to the dataset. Soil variables were only appended if the sites were missing sampled measures, with all studies lacking at least one soil property.

To help prevent extrapolation, all global layers were truncated to values represented by each subset of data, i.e., the minimum and maximum values used in each of the three community metric models. The exception was the number of months of snow cover, which was truncated at four months, thus any sites or areas of the globe with a greater number of months than four were modelled and predicted (respectively) as four months. This ensured an even spread across the range of values (many sites were within 0-4 months, only 9% of sites were greater than four).

Mixed effects modelling

Earthworm species richness, abundance, and biomass models

Three (generalised) linear mixed effects models were constructed, using lme4 (60), one for each of the site-level community metrics: species richness, total abundance (individuals per m^2), and total biomass (grams per m^2). Prior to modelling, the full dataset was split into three subsets, based on the response variables (i.e., a dataset containing all sites with a species richness value). Within each dataset, we tested for multicollinearity between the elevation, climate, and soil variables using Variance Inflation Factors (VIFs) and removing the variable with highest VIF in turn until all remaining variables were below the predetermined threshold of 3 (61).

Abundance and biomass were log transformed ($\log(x + 1)$) prior to modelling and were then modelled using a Gaussian error structure. Species richness was not log transformed, but instead modelled with a Poisson error structure. All models had random effects that accounted for variation between each of the different studies, using an intercept only structure. Fixed effects included habitat cover, elevation, soil properties, and climate variables. All continuous variables (i.e., elevation, all soil variables, and most climate variables) were centred and scaled (variables were centred on the mean value and divided by the standard deviation) to aid model fitting and interpretability. Number of months of snow cover was modelled as a categorical variable (and therefore not centred and scaled) to allow for a non-linear relationship. As it is expected that earthworm diversity will peak with some snow cover, due to increased precipitation, and soil protection during freezing months (62), but be restricted by prolonged snow cover (63). This also improved the modelling process, as sites were skewed towards the lower number of months, with not enough data at the higher latitudes to fit a non-linear regression.

For each of the three models, the structure of the fixed effects in the maximal model was the same. Habitat cover and elevation were included as additive effects with no interactions. The other variables were grouped into four themes: ‘soil’, ‘precipitation’, ‘temperature’ and ‘water retention’ (Table S3). For example, all precipitation variables that remained (i.e., were not removed due to their VIF score) were grouped together. Within the soil and two climate groups, all two-way interactions were considered. The water retention group contained specific two-way interactions between soil structure variables (clay and silt percentage) and climate variables relating to water availability that were present in the two climate themes (annual precipitation, precipitation seasonality, PET, PETSD, and aridity). These specific interactions were to account for soil moisture and how quickly moisture might leave the soil.

Each maximal model was then simplified using Akaike information criterion (AIC) values. All interactions were tested first, and removed if AIC values were reduced compared to the more

complicated model. Any main effects that were not involved with interactions were tested, and removed if AIC values were reduced (44, 64) (Table S3).

Ecological group responses

The same modelling approach was used to investigate changes in earthworm ecological groups across the different habitat types. For each site, the diversity, abundance and biomass of the three main ecological groups (epigeic, endogeic, and anecic) and one minor ecological group (epi-endogeic) were calculated, based on the category assigned following species name standardisation. Three (generalised) linear mixed effects models with diversity, abundance and biomass as response variables were constructed as detailed above, with the exception that habitat cover interacted with the ecological group (i.e., the biomass of epigeics, endogeics, and anecics at each site). The model was simplified following details given above.

The community metrics of each ecological group in each habitat cover was then predicted, using the 'predict' function in 'lme4' (when all other variables were at zero, i.e., the mean). The predicted values for the three main ecological groups (epigeic, endogeic and anecic, which had sufficient underlying data. Epi-endogeics were modelled but did not have enough underlying data for robust predictions) were plotted using the 'triangle.plot' function in 'ade4' (65). The predicted total biomass, i.e., the total of the predicted biomass of the three main ecological groups, was used to determine size of the points within the triangle plot.

Creating maps of earthworm communities

The global patterns of earthworm communities (species richness, abundance, and biomass) were predicted using each of the three models. The values from the relevant global layers (i.e., those corresponding to the variables that remained in each model following simplification) were used in the 'predict' function in the 'lme4' package, being predicted based on the coefficients of the final models.

A global layer of predicted values was then presented as maps of local communities of earthworms. Although all global layers had been capped at values represented in the underlying dataset, extrapolation still occurred during the prediction (there were instances where grid cells in multiple layers were at the extreme values, and such combinations were not represented in the underlying data, most evident in the predictions of earthworm biomass, see 'Interpreting the model validation'). To prevent outliers skewing the visualization of results, the colour of maps were curtailed at the extreme low and high values. Curtailing was based on where the majority of values laid. Thus, values lower or higher than the number marked on the scale are coloured the same but may represent a large range of values.

Variable Importance

In order to determine which themes (soil, elevation, habitat cover, precipitation, temperature, water retention) were the most influential in driving earthworm communities, Variable Importance was performed using random forest models (66, 67).

For each of the three community metrics, random forest models were constructed (67) with all the variables that were present in the final (i.e. simplified) model. Random forest models use multiple regression trees to classify data (67). This method was chosen as these models can handle non-linear data, whilst interactions are not specified but can be learnt from the data (68). Random forest models are an ensemble of individual regression (or classification) trees (66, 67). Each tree is created using around two-thirds of the available data, i.e., "out-of-bag" regression, and the process is repeated until the 'forest' is complete (ntree default = 500 trees). At each node

in the tree, the subset of response variables is split using the best predictor variable. Unlike regression trees, where at each node the best predictor is used from all available predictor variables, random forest models use only a random sample of the predictor variables ("Mtry") to determine the best predictor to split the response variable at each node (66, 67). The default Mtry value was used (number of predictors divided by 3), so in our case of 10 to 12 predictor variables Mtry = 3 (biomass model) and 4 (species richness and abundance models). The "out-of-bag" data is then predicted using the average prediction of all trees (67).

In addition, random forest models can be used to assess the importance of individual variables (66). One such measure is the mean decrease in node impurity calculated from the decrease in the residual sum of squares for the variable that was used at the node. The average decrease for each variable is averaged across all the trees to create the node impurity (67). An alternative importance variable is the mean decrease in accuracy. For each tree, when the "out-of-bag" data (~one-third of the data) is being predicted, a single predictor variable is permuted, and the increase in prediction error calculated (67). This mean decrease in accuracy is often considered the best of those available (69), but results between the decrease in node impurity and mean decrease in accuracy often correlate well (70).

For each of the three random forest models, the mean decrease in node impurity and mean decrease in accuracy was calculated (using the 'importance' function in 'randomForest') for each predictor variable in the random forest. In order to determine which theme of variables (habitat cover, elevation, soil, temperature, precipitation, and water retention) was most important in driving patterns in earthworm communities, the mean decrease in node impurity was averaged for all variables within each theme and weighted by the number of times each variable was used in the random forest compared to the other variables in the same theme.

Model validation and sensitivity analysis

A number of additional analyses were performed to determine the robustness of the models and the ability to predict new values. Firstly, the influence of combining both measured soil properties and soil properties from SoilGrids was tested. For the three response variables (species richness, abundance, and biomass) models were created that only included data from SoilGrids. The same modelling process was used as described above (using VIFs to determine starting variables, then simplification). Once the final model had been identified for each of the three community metrics, 10-fold cross-validation was performed (71).

Cross-validation was performed in two ways. Firstly by randomly splitting the dataset underlying each of the three models into 10 nearly-equal size groups. Using the model structures produced following simplification, the model was built using 9 of the groups of data. The 10th group of data was predicted from the re-built model. The predicted data was plotted against the observed data. This process was repeated until all 10 groups of data had been predicted. This process was done for the models that contained only SoilGrids data, and the main models (that used a mixture of soil property data, Fig. S3). Secondly, by splitting the dataset into 10 near-equal sized groups based on study. Thus, the site-level community metrics for a 10th of the studies were predicted by the remaining 9/10 of the data. This process was only performed on the main models (Fig. S4).

For the site-level cross-validated models, the mean squared error (MSE) was calculated from the results of the cross-validation. MSE measures the ability of the model to predict new data, and the result are easily interpretable as they are on the same scale as the original data. MSE was calculated for the total of all models, but as the models may be better at predicting

certain ranges of values, MSE was also calculated for the tertiles of the observed data (i.e., the ability of the model to predict the low, medium, and high values of earthworm communities).

Secondly, the R^2 values of all models (the main models, and the models with only SoilGrids data) were calculated using the MuMIn package (72). The R^2 values describe the fit of the model to the data. The $R^2_{marginal}$ is the variance explained by the fixed effects, whilst the $R^2_{conditional}$ is the variance explained by the fixed and random effects.

To determine the confidence of the globally predicted values, we followed methods of (9). For each of the biodiversity models, we investigated how well the underlying data represented the full multivariate environmental covariate space of the global layers. We performed a Principal Components based approach on each of the datasets. The centering values, scaling values, and eigenvectors were then used to transform all global layers into the same PCA spaces. Then, we created convex hulls for each of the bivariate combinations from the first 6 (total biomass data) and 7 (richness and total abundance data) i.e., half of the number of variables within the model, principal components to cover more than 90% of the sample space variation. Using the coordinates of these convex hulls, we classified whether each pixel of each global layer falls within or outside each of these convex hulls. Therefore, if all global layers within a pixel were within the convex hull, the interpolation percentage would be 100%, while if only half of the layers were within the convex hull, the interpolation percentage would be 50%. This analysis was performed in Google Earth Engine (73).

Regional latitudinal diversity gradients

In order to ascertain whether there is a species diversity gradient with latitude, the site-level diversity data underlying the species richness model (i.e., contained sites with species level or morphospecies identification) was used. The sites were split into latitude zones that contained roughly equal numbers of sites. Sites were assigned to a zone based on their latitude, with the intention that each zone would contain close to 250 sites. However, all sites with the same coordinates were kept within the same band, so the number of sites within a zone did vary (min = 209, max = 341, mean = 267.6). The number of unique species, based on species binomials, across all sites within each zone was calculated. Within each zone, it was also assumed that each uniquely named morphospecies was different from any of the named species (number of morphospecies across zones, min = 0, max = 21, mean = 3.05). Some of the sites also contained genus-level only identification. When this was the case, a genus was included as one additional species if the genus was unique within the zone (i.e., no named species belonged to that genus). As the number of morphospecies was biased with latitude (i.e., greater taxonomic expertise in the temperate regions, Table S4), the analysis was repeated excluding morphospecies (Fig. S2). The two methods resulted in similar patterns, but reduced richness in some of the zones in the tropics.

All statistics, data manipulation and processing of global data layers was implemented in R (version 3.3.1; 73).

Supplementary Text

Interpreting the model validation

The results of the biomass model highlighted an issue with the modelling technique used. All of the global data layers were cut at values represented by the underlying datasets. However, during the prediction, it was often the case that multiple data layers were at the extreme ends of the possible range of values. This led to, especially in the case of the biomass model, and to

some extent the abundance model, unrealistically high values being predicted. This issue could only be fixed with additional data, but does not affect the visual maps produced in this study. For the global predictions of biomass, values greater than 2 kg per m^2 were deemed to be unrealistic. This threshold is over 4 times the maximum recorded biomass of earthworm communities (75), and thus is highly unlikely to be realistic. 98.9% of pixels were less than 400 g per m^2 (the maximum recorded earthworm community biomass recorded in the temperate region (75)).

Overall, the models had reasonably good fit to the data, assessed using the R^2 values (Table 1C). However, the predictive power of the models was variable. With all models, the total MSE (Table 1A) increased mainly due to the ill-fitting of the sites with higher values. It is unclear why high values cannot be fitted well with the models; however, it is highly likely that increasing the number of sites would help either identify the issue or improve model fit.

For the majority of the datasets (182 out of 228 studies), the models contained the measured soil properties for some of the variables. Where this was missing, we used the SoilGrids data. Models which contained only SoilGrids data had a better fit to the data (Table 1C) and were typically better at predicting during cross-validation (lower MSE values; Table 1B). However, in most instances, the change in MSE was negligible between the different types of models (Table 1 A versus B). Despite the models that contained only SoilGrids data performing slightly better in terms of R^2 and MSE, there are other reasons why using a mixture of the measured variables and the SoilGrids variables is the best option in the modelling process. Firstly, modelled global estimates of the soil properties may not accurately depict site-level conditions (76), which could result in the variables appearing less important than they would be if they matched the measured communities. Secondly, some of the coordinates within a study were identical which would result in identical SoilGrids data (for these datasets, often small-scale field experiments, the measured soil properties variables were not identical). Using only SoilGrids data would reduce the gradient of soil properties within each study, reducing the number of available gradient comparisons across all datasets. And given that a number of studies (106 out of 228 studies) had identical climate variables across all sites, having variety in all other variables prevented this being an issue within the modelling framework. We call on soil ecologists to collect data on soil properties when they measure diversity of soil taxa, as this permits more robust modelling at both the small scale, and across larger scales.

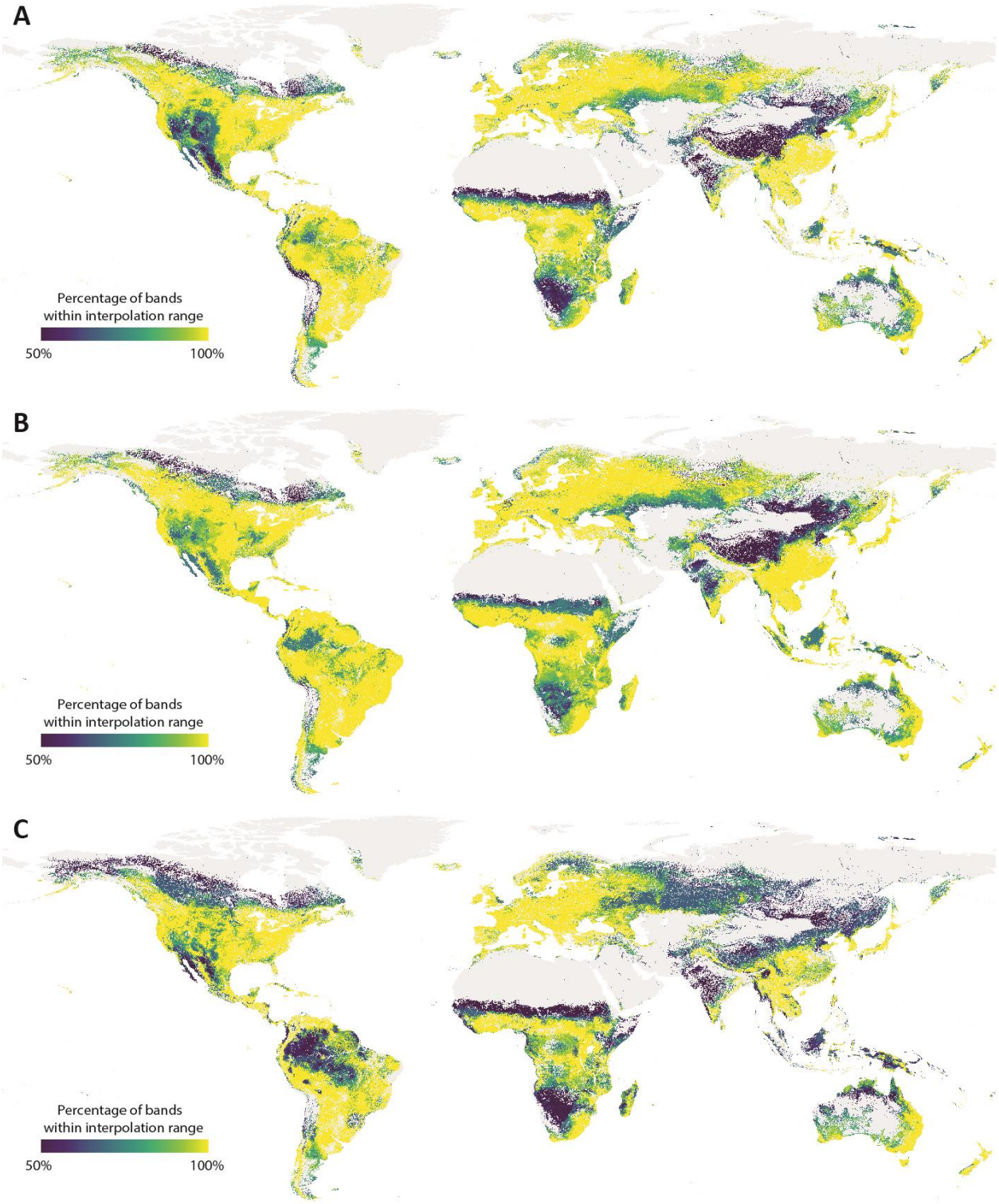
Regardless of whether the model contained measured soil properties or only SoilGrids data, the models were consistently worse at predicting when observed values were high (Table 1). This is likely due to the small number of studies where sampled values were high. Only 5 studies had more than 10 species of earthworms in at least one site, and only 6 studies had more than 300 grams per m^2 of earthworm biomass in at least one site. There were a greater number of studies that contained high abundance of earthworms, with 34 studies having at least one site that contained more than 600 individuals per m^2. Increasing the number of studies and sites would help identify whether this, or another cause, is responsible. Ideally, this would improve the predictive power of the models. It is hoped that efforts will continue to collate earthworm diversity data from across the globe.

When cross-validation was performed at the study level (Fig. S4) the predictions were not scattered around the 1:1 line. However, this is to be expected, as when sites are randomly selected and predicted, the study level random-effect is most likely still present in the model. This ensures that the community metrics of each sites can be predicted using the variance from the study it is within. When an entire study is removed, and so no random-effect level exists for

it in the model, all study-level random effects are averaged in order to produce the prediction. Thus, the prediction error is increased, and more concentrated around the overall mean.

For the species richness (Fig. S1A) and total abundance data (Fig. S1B), the interpolation percentage across the globe was relatively high (i.e., the underlying datasets adequately captured the majority of the multivariate environmental conditions). Regions surrounding the Eurasian Steppe, and the Himalayas were some of the most extrapolated regions, with arid regions in Africa and boreal regions also having lower interpolation percentages. For the total biomass data, more regions of the globe had low interpolation percentages (Fig. S1C). These low-value regions were spread across the tropics, particularly Brazil and Indonesia, and large parts of Africa, the sub-tropics, such as India, and temperate regions, including northern China and Russia. Overall, we would expect the globally predicted values of the biomass model to be more extrapolated, than the diversity and total abundance models.

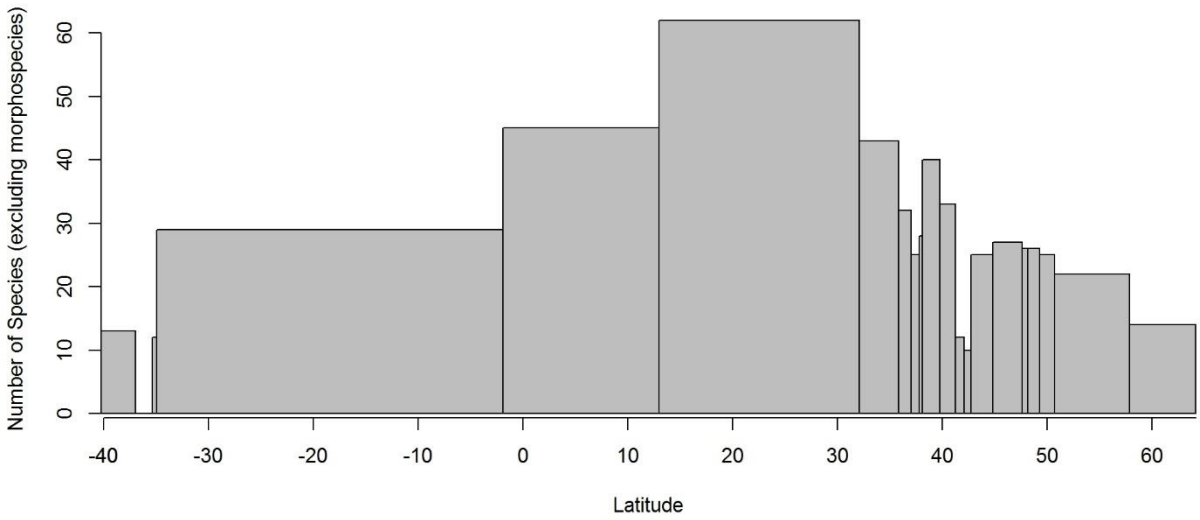
Fig. S1.



Assessment of global extrapolation and interpolation for the (A) species richness data, (B) total abundance data, and (C) total biomass data. Scale shows the percentage of pixels (from each of

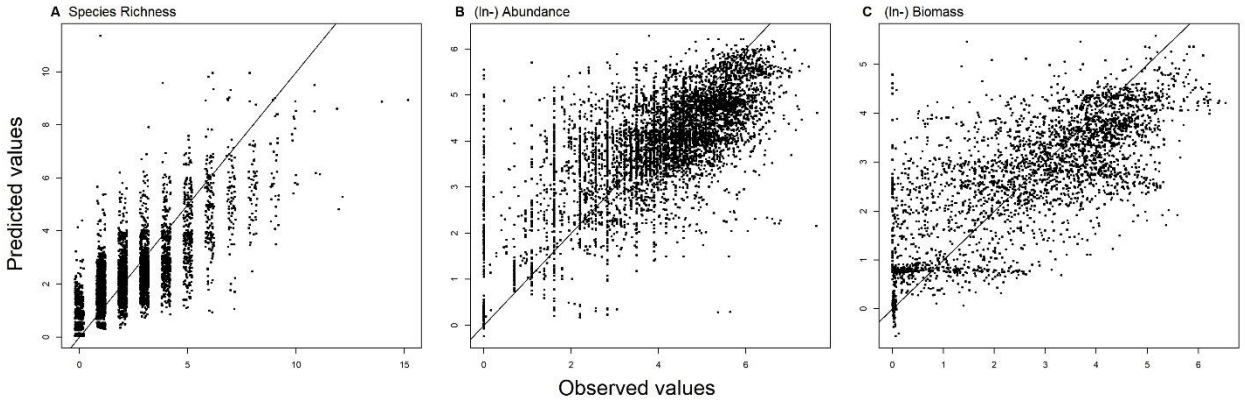
the global layers) falling within the convex hull spaces of the first 6 (biomass) and 7 (richness, abundance) Principal Components collectively explaining >90% of the variation. Low interpolation percentage values (in blue) indicate that few global layers were represented by data, thus extrapolation would have occurred during prediction, whilst high interpolation percentages (in yellow) indicated that many or all global layers were represented by data, thus interpolation would have occurred during prediction.

Fig. S2.



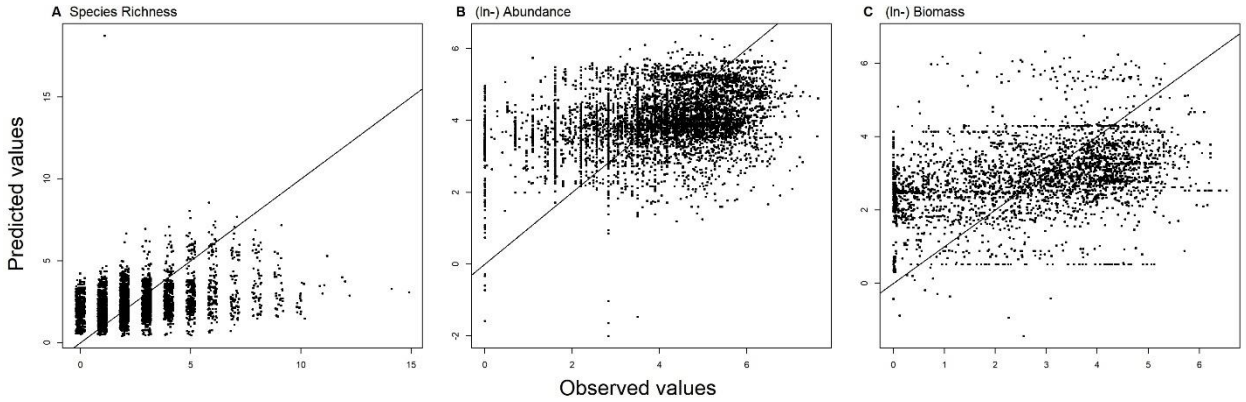
The number of unique species within each latitudinal zone, when the number of sites within each zone was kept relatively equal. The height of the bar indicates the number of unique species across all sites. The width of the bar shows the latitude range the sites cover. Within each zone only the species with binomials, or genera with no other identified species, were included in the calculations (morphospecies were excluded).

Fig. S3.



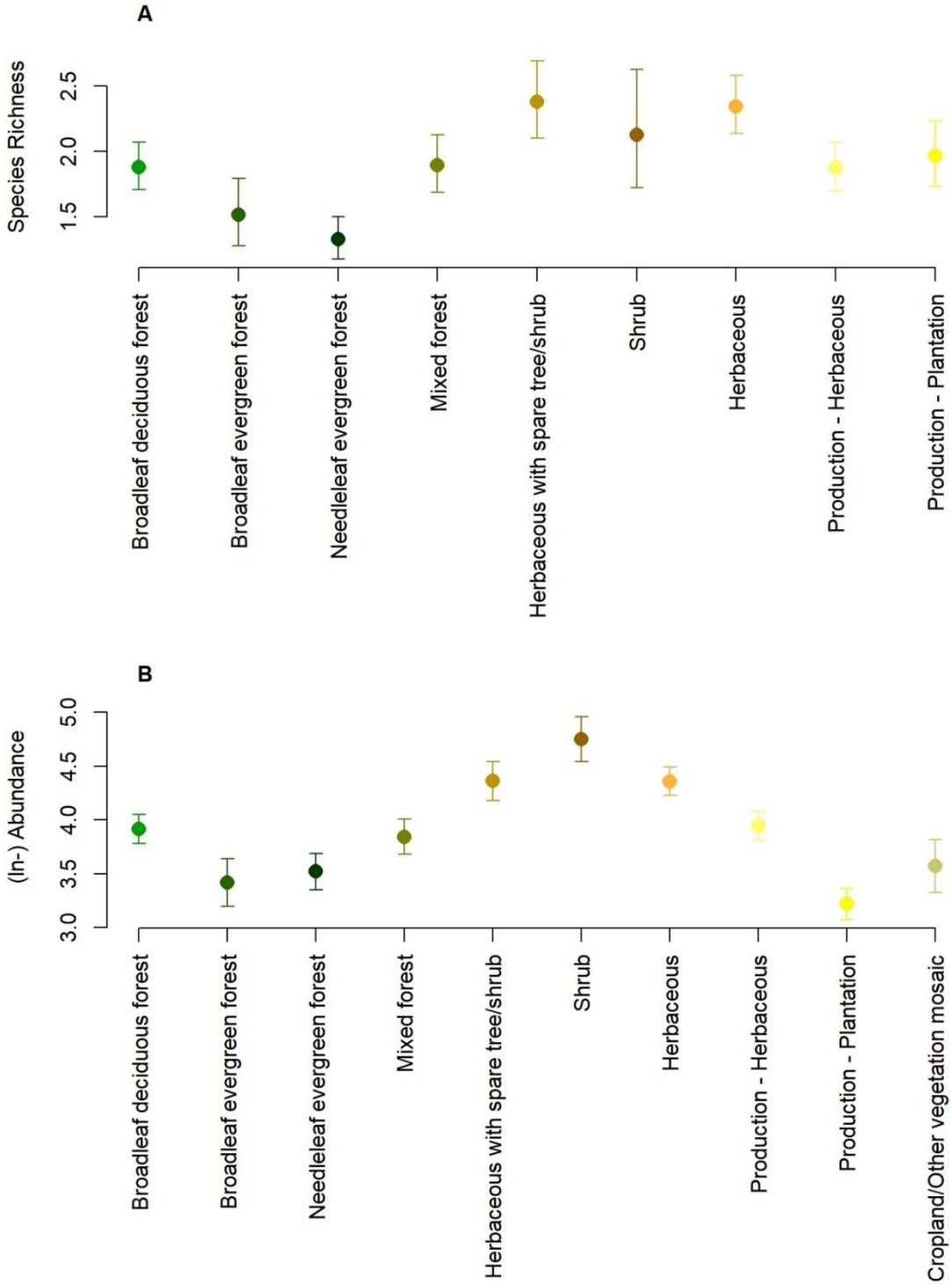
10-fold cross validation of the three main community metric models, (A) species richness, (B) ln-abundance, and (C) ln-biomass. X-axis shows the observed value, and Y-axis the predicted value, black line is the 1:1 line. The underlying dataset of each model was randomly split into 10 nearly-equal size groups. Using the model structures produced following simplification, the model was built using 9 of the groups of data. The 10th group of data was predicted from the re-built model. This process was repeated until all 10 groups of data had been predicted. The predicted data was plotted against the observed data.

Fig. S4.



10-fold cross validation of the three main community metric models, (A) species richness, (B) ln-abundance, and (C) ln-biomass. X-axis shows the observed value, and Y-axis the predicted value, black line is the 1:1 line. The underlying dataset of each model was randomly split into 10 nearly-equal size groups, so that each group contained all the data of a tenth of the studies. Using the model structures produced following simplification, the model was built using 9 of the groups of studies. The 10th group of studies was predicted from the re-built model. This process was repeated until all 10 groups of studies had been predicted. The predicted data was plotted against the observed data.

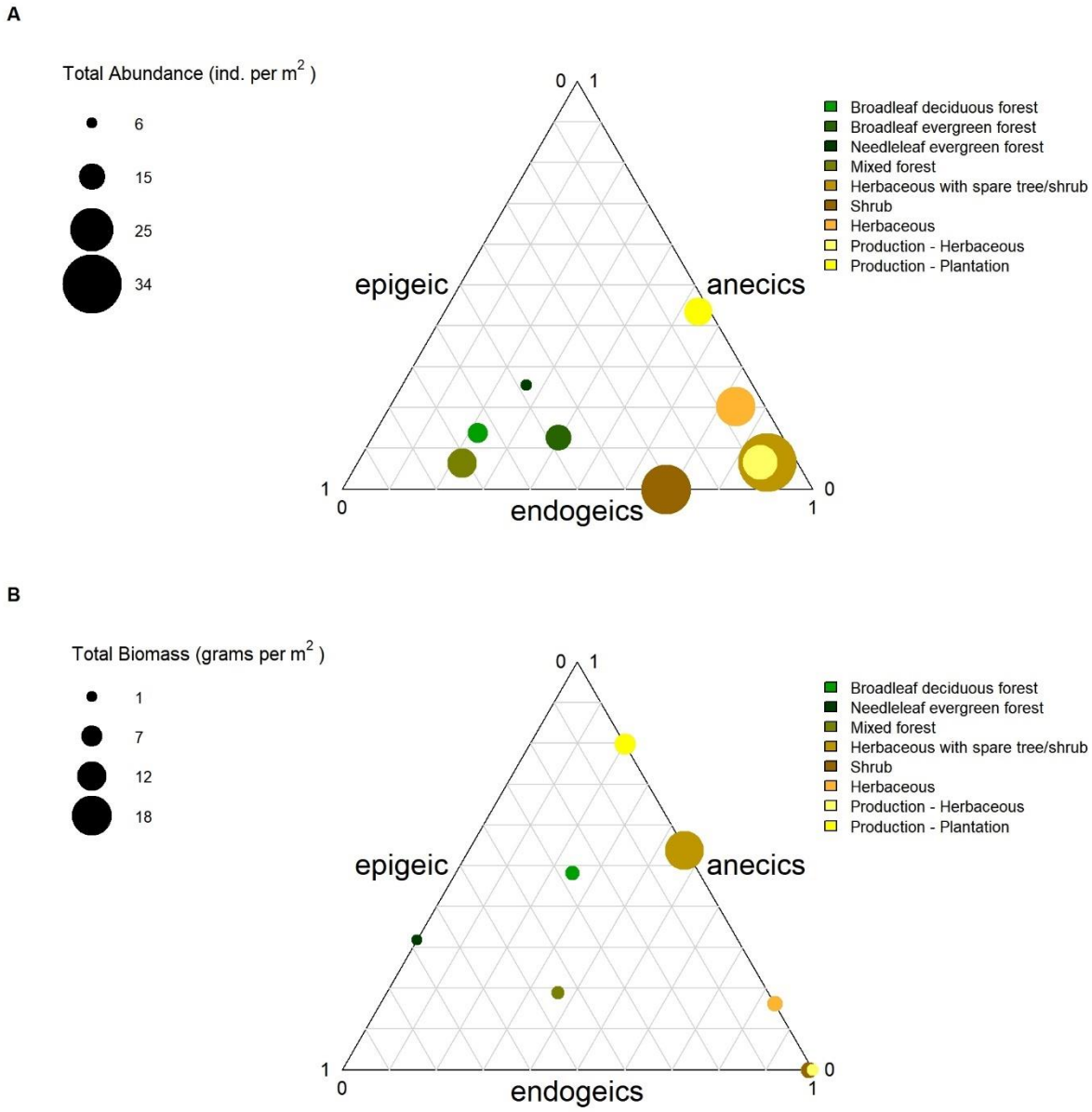
Fig. S5.



Changes in (A) species richness and (B) ln-abundance across the different habitat cover categories (+/- SD). Values of species richness and abundance are predicted from the main

models when all other variables are at zero, i.e., the mean. Not all habitat cover categories had sampled estimates (i.e., species richness could not be estimated for ‘Cropland/Other vegetation mosaic’).

Fig. S6.



The (A) total abundance and (B) total biomass of the three ecological groups (epigeic, endogeic and anecic earthworms) within each habitat cover category based on modelled estimates. Circle size is relative to the total biomass predicted for the habitat cover, and circle colour indicates the habitat cover. Position within the three axes indicates the proportion of each of the three ecological groups within the community, based on the interaction term between habitat cover and ecological group. During simplification, the interaction term between habitat cover and ecological group was removed in the species richness model, thus those results are not shown.

Not all habitat cover categories had sampled estimates (i.e., biomass could not be estimated for ‘Broadleaf evergreen forest’ or ‘Cropland/Other vegetation mosaic’). This figure shows, for example, that “Broadleaf deciduous forests” have a rather even predicted biomass distribution across the three ecological groups (but low total biomass), while “Production sites” (“Plantation” and “Herbaceous”) have high total earthworm abundance, but are dominated by endogeic species.

Table S1.

Original habitat cover variable	Reclassified habitat cover
No Data	NA
Cropland, rainfed	Production - Herbaceous
Cropland - herbaceous cover	Production - Herbaceous
Cropland - Tree or shrub cover	Production - Plantation
Cropland, irrigated or post-flooding	NA
Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%)	Cropland/Other vegetation mosaic
Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%)	Cropland/Other vegetation mosaic
Tree cover, broadleaved, evergreen, closed to open (>15%)	Broadleaf evergreen forest
Tree cover, broadleaved, deciduous, closed (>40%)	Broadleaf deciduous forest
Tree cover, broadleaved, deciduous, open (15-40%)	Broadleaf deciduous forest
Tree cover, needleleaved, evergreen, closed (>40%)	Needleleaf evergreen forest
Tree cover, needleleaved, evergreen, open (15-40%)	Needleleaf evergreen forest
Tree cover, needleleaved, deciduous, closed (>40%)	Needleleaf deciduous forest
Tree cover, needleleaved, deciduous, open (15-40%)	Needleleaf deciduous forest
Tree cover, mixed leaf type (broadleaved and needleleaved)	Mixed forest
Mosaic tree and shrub (>50%) / herbaceous cover (<50%)	Tree open
Mosaic herbaceous cover (>50%) / tree and shrub (<50%)	Herbaceous with spare tree/shrub
Shrubland	Shrub
Grassland	Herbaceous
Lichens and mosses	NA
Sparse vegetation (tree, shrub, herbaceous cover) (<15%)	Sparse vegetation
Tree cover, flooded, fresh or brackish water	NA
Tree cover, flooded, saline water	NA
Shrub or herbaceous cover, flooded, fresh/saline/brackish water	NA
Urban areas	Urban
Bare areas - consolidated	Bare area (consolidated)
Bare areas - unconsolidated	Bare area (unconsolidated)
Water bodies	Water bodies
Permanent snow and ice	NA

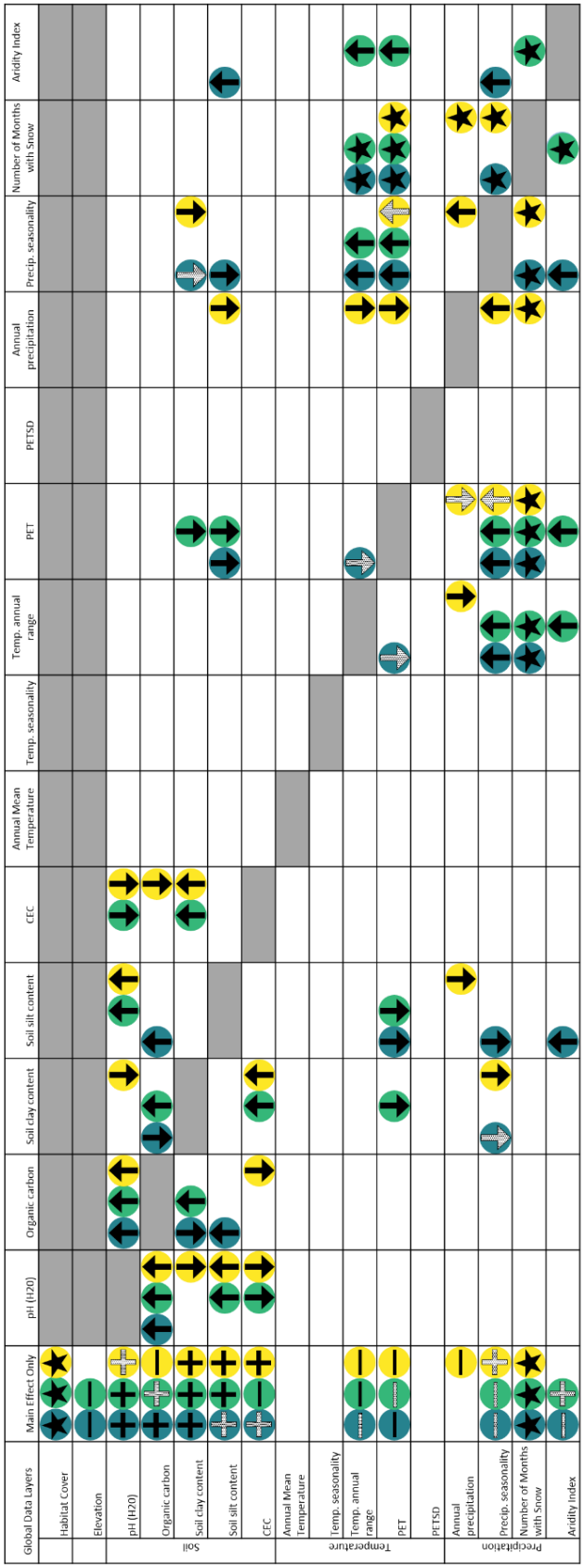
The re-categorisation of the ESA habitat cover variable. Habitat cover at a sampled site was classified based on the ‘Reclassified habitat cover’ column. As not all categories of habitat were available in the data (i.e., due to too detailed categories, or in habitats typically devoid of sampling), some of the categories of the original habitat cover variable (left-hand column) were reclassified (right-hand column). Usually, this meant that categories were grouped together (i.e., to reduce the categories based on ‘openness’).

Table S2.

	Variable	***Source***	***Original Spatial Resolution***
	Habitat Cover	ESA CCI-LC	300 m
	Elevation	(*59*)	1 km
Soil Parameters	pH (H20)	SoilGrids (*54*)	1 km
	Organic carbon		1 km
	Soil clay content		1 km
	Soil silt content		1 km
	CEC		1 km
Temperature	Annual Mean Temperature	CHELSA (*55*)	1 km
	Temp. seasonality		1 km
	Temp. annual range		1 km
	PET	(*57*, *58*)	1 km
	PETSD		1 km
Precipitation	Annual precipitation	CHELSA (*55*)	1 km
	Precip. seasonality		1 km
	Number of Months with Snow	(*56*)	1 km
	Aridity Index	(*57*, *58*)	1 km

Information for each of the 15 global layers detailed in the methods. Abbreviations: CEC = Cation exchange capacity, Temp. = Temperature, Precip. = Precipitation, PET = Potential evapotranspiration, PETSD = within year standard deviation of PET

Table S3.


Global Data Layers
Main Effect Only
pH (H20)
Organic carbon
Soil clay content
Soil silt content
CEC
Annual Mean Temperature
Temp. seasonality
Temp. annual range
PET
PETSD
Annual precipitation
Precip. seasonality
Number of Months with Snow
Aridity Index
Habitat Cover
Elevation
Soil
Temperature
Precipitation

Results following model simplification of the three community metric models. ‘Main Effect Only’ column shows the slope for the main effect of each variable in the final species richness (turquoise), total abundance (green) and total biomass (yellow) models. ‘+’ indicates the slope was positive, ‘-’ indicates a negative slope, and ‘*’ indicates that the variable was categorical (with intercepts and slopes depending on the category). The remaining columns show the interactions between the variables. An upwards arrow indicates that the slope of one variable would become more positive as the other variable is increased. A downwards arrows indicates that the slope of the one variable would become more negative as the other variable is increased. However, it may not necessarily indicate that the slope changes direction. Black symbols indicate that the coefficient was significant ($p < 0.05$) within the model, and grey/hatched symbols indicate they were not significant [NB. P-values are for illustrative purposes only, as models were simplified based on AIC values]. Habitat cover and elevation were only in the models as main effects. Also noted is the variable theme in which the variable was grouped. Variables that interacted within the ‘water retention’ theme are not shown explicitly, but can be deduced based on interactions between a climate variable and soil property variable. Abbreviations: CEC = Cation exchange capacity, Temp. = Temperature, Precip. = Precipitation, PET = Potential evapotranspiration, PETSD = within year standard deviation of PET.

Table S4.

Latitude	*Number of sites*	*Number of named species*	*Mean Latitude Range*	*Number of Morphospecies*	*% Native*	*% Non-native*	*% Unknown*
(65,70]	55	11	63.66	0	21.02	57.32	21.66
(60,65]	255	14	56.91	1	0	0	100
(55,60]	157	18	55.11	0	1.57	0	98.43
(50,55]	960	35	37.47	3	11.74	2.85	85.41
(45,50]	1136	38	32.53	1	1	54.42	44.58
(40,45]	1080	54	29.63	2	7.5	13.7	78.8
(35,40]	308	47	34.49	1	6.05	7.74	86.22
(30,35]	113	18	28.15	3	0	0	100
(25,30]	47	18	56.81	0	0	0	100
(20,25]	9	12	10.78	3	22.86	0	77.14
(15,20]	30	11	6.17	2	16.87	45.78	37.35
(10,15]	26	4	14.18	3	0	0	100
(5,10]	40	27	5.53	3	27.11	19.88	53.01
(0,5]	127	10	14.08	12	0	2.17	97.83
(-5,0]	146	10	14.21	14	0	0	100
(-10,-5]	22	1	46.08	6	0	0	100
(-15,-10]	5	0	NA	1	0	0	100
(-20,-15]	0	NA	NA	NA	NA	NA	NA
(-25,-20]	5	8	14.9	4	40	43.33	16.67
(-30,-25]	0	NA	NA	NA	NA	NA	NA
(-35,-30]	150	12	86.6	1	39.66	37.93	22.41
(-40,-35]	679	16	62.45	0	24.93	74.45	0.62
(-45,-40]	3	4	105.78	0	0	0	100

Details of the number of sites, and composition of the earthworm species for each 5 degree latitudinal band. Based on the geographical coordinates, each site was classified into 5 degree latitudinal bands. The number of species is based on the binomial names given by the original data collector, then revised for consistency. For each named species in each band the latitudinal range (the difference between the minimum latitude and the maximum latitude, based on all sites within the dataset that the species occurred) was calculated, and the average taken from all species within the band. Morphospecies are individuals that were identified to genus level, and identified by the data collectors as morphologically distinct from other (morpho-) species, but were not identified to species-level. The percentage of native and non-native species is based on information provided by the original data collectors, and is therefore often incomplete (depicted in the '% unknown' column).