Completing Linnaeus's Inventory of the Swedish Insect Fauna: Only 5000 Species Left

Fredrik Ronquist[1*], Mattias Forshage[2], Sibylle Häggqvist[2], Dave Karlsson[3], Rasmus Hovmöller[2,], Johannes Bergsten[2], Kevin Holston[1], Tom Britton[4], Johan Abenius[5], Bengt Andersson[6], Peter Neerup Buhl[7], Carl-Cedric Coulianos[8], Arne Fjellberg[9], Carl-Axel Gertsson[10], Sven Hellqvist[11], Mathias Jaschhof[3], Jostein Kjærandsen[12], Seraina Klopfstein[13], Sverre Kobro[14], Andrew Liston[15], Rudolf Meier[16], Marc Pollet[17], Marko Prous[15], Matthias Riedel[18], Jindřich Roháček[19], Meike Schuppenhauer[20], Julia Stigenberg[2], Ingemar Struwe[21], Andreas Taeger[15], Sven-Olof Ulefors[3], Oleksandr Varga[22], Phil Withers[23] & Ulf Gärdenfors[24]

[1]Dept. Bioinformatics and Genetics, Swedish Museum of Natural History, Box 50007, SE-104 05 Stockholm, Sweden. [2]Dept. Zoology, Swedish Museum of Natural History, Box 50007, SE-104 05 Stockholm, Sweden. [3]Station Linné, Ölands Skogsby 161, SE-386 93 Färjestaden, Sweden. [4]Dept. Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden. [5]Vårdkasstigen 10, SE-149 41 Nynäshamn, Sweden. [6]Borgvägen 10, SE-382 45 Nybro, Sweden. [7]Dept. Entomology, Natural History Museum of Denmark, Universitetsparken 15, DK-2100 Copenhagen, Denmark. [8]Kummelnäsvägen 90, SE-132 37 Saltsjö-Boo, Sweden. [9]Mågeröveien 168, NO-3145 Tjöme, Norway. [10]Murarevägen 13, SE-227 30 Lund, Sweden. [11]Älvtået 4, SE-903 60 Umeå, Sweden. [12]Tromsø University Museum, UiT—The Arctic University of Norway, PO Box 6050, Langnes, NO-9037 Tromsø, Norway. [13]Natural History Museum Basel, Augustinergaße 2, CH-4051 Basel, Switzerland. [14]Brattbakken 2, NO-3159 Melsomvik, Norway. [15]Senckenberg German Entomological Institute, Eberswalder Straße 90, DE-15374 Müncheberg, Germany. [16]Lee Kong Chian Natural History Museum and Dept. Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore. [17]Research Institute for Nature and Forest, Kliniekstraat 25, BE-1070 Bruxelles, Belgium; Research Group Terrestrial Ecology, Ghent University, K. L. Ledeganckstraat 35, BE-9000 Ghent, Belgium; and Entomology Unit, Royal Belgian Institute for Natural Sciences, Vautierstraat 29, BE-1000 Bruxelles, Belgium. [18]Bavarian State Collection of Zoology, Münchhausenstraße 21, DE-81247 München, Germany. [19]Dept. Entomology, Silesian Museum, Nádražní okruh 31, CZ-746 01 Opava, Czech Republic. [20]Dept. Soil Zoology, Senckenberg Museum of Natural History Görlitz, Am Museum 1, DE-02826 Görlitz, Germany. [21]Stenbrohultsvägen 111, SE-757 58 Uppsala, Sweden. [22]Schmalhausen Institute of Zoology, National Academy of Sciences of Ukraine, B. Khmelnytskyi 15, Kyiv 01030, Ukraine. [23]40 Montée du Cimetière, Sainte Euphémie, 01600, France. [24]Swedish Species Information Centre, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden.

*Corresponding author: fredrik.ronquist@nrm.se

Abstract

Despite more than 250 years of taxonomic research, we still have only a vague idea about the true size and composition of the faunas and floras of the planet [1–4]. Many biodiversity inventories provide limited insight because they focus on a small taxonomic subsample or a tiny geographic area [5, 6]. Here, we report on the size and composition of the Swedish insect fauna, representing roughly half the macroscopic diversity of one of the largest European countries, based on more than a decade of data from the Swedish Taxonomy Initiative and a massive inventory of the country's insect fauna [7, 8]. The fauna is considered one of the best known in the world, but the inventory has nevertheless revealed a surprising amount of hidden diversity: more than 3,000 new species (301 new to science) have been documented so far. We show that three independent extrapolation methods converge on similar estimates of the true size and composition of the true fauna, suggesting that it comprises around 33,000 species. Of those, 8,600 (26%) were unknown at the start of the inventory and 5,500 (17%) still await discovery. Most of the new species belong to Hymenoptera and Diptera groups that are decomposers or parasitoids. Thus, current knowledge of the Swedish insect fauna is strongly biased taxonomically and ecologically, and we argue that this is likely true for most insect faunas. Addressing these biases is critical in understanding insect biomes and the ecosystem services they provide.

The existence of uncharted organism diversity in tropical regions and biodiversity hotspots has long been recognized, and recent work has highlighted the gaping holes in our knowledge of microscopic diversity [9, 10]. But how much do we actually know about the macroscopic floras and faunas of the most intensely studied corners of the planet? This is the question we address in the present paper.

Specifically, we focus on Sweden, one of the largest countries in Europe with respect to area, and its insect fauna, believed to comprise roughly half the diversity of macroscopic organisms in the country. Thanks to the Linnaean legacy, the Swedish flora and fauna are among the best known in the world. Since 2002, the knowledge of Swedish organismal diversity has also increased substantially thanks to the Swedish Taxonomy Initiative, the aim of which is to completely chart the macroscopic flora and fauna of the country.

Insects have received a considerable amount of attention from the initiative. In addition to supporting taxonomic research projects on the most poorly known insect groups, the initiative has also funded a massive countrywide inventory, the Swedish Malaise Trap Project [11]. Malaise traps [12] are particularly effective in collecting the insect orders Diptera (mosquitoes, gnats, midges and flies) and Hymenoptera (sawflies, wasps, ants and bees), to which most of the poorly-known insect groups belong.

At the start of the initiative, the known Swedish insect fauna was estimated to comprise 24,300 species [13]. Since then, using a combination of traditional and molecular methods, taxonomists have added 3,097 species, 301 of which have been described as new to science (Supplementary Table S1). It is clear that many more species remain to be discovered and described, challenging biologists to reexamine questions that were long thought to have been satisfactorily answered: how large is the Swedish insect fauna actually, and what is its true taxonomic and ecological composition?

Here, we address these questions using three different methods. First, we asked experts involved in the Swedish Taxonomy Initiative and the Malaise trap inventory to provide informed guesses about the number of Swedish species for their group(s). Specifically, the insect fauna was divided into families, with the three largest families

(Ichneumonidae, Braconidae and Staphylinidae) divided into subfamilies, resulting in a total of 663 family-group taxa, for each of which we obtained an expert estimate (Table S1). The experts estimated that the Swedish insect fauna consists of approximately 33,000 species. This exceeds the 2003 estimate of the known fauna by 8,600 species and suggests that there are still around 5,500 species that remain to be discovered. Those species are concentrated, however, to a small number of taxa. Only ten of the 663 family-group taxa are expected to contain more than 100 of these missing species; only two of these taxa (Ichneumonidae: Cryptinae and Pteromalidae, both Hymenoptera) still await attention from taxonomists funded through the Swedish Taxonomy Initiative.

For critical groups, we checked the expert assessments against two independent species richness estimates based on data from the Swedish Malaise Trap Project. The project ran Malaise traps at 73 sites spread throughout the country and covering a wide range of habitats; most traps were run for three consecutive years, 2003–2006 (Fig. 1; Table S2). The total material is estimated to comprise around 20 million specimens. To date, more than 90% of the material has been sorted to taxonomic fractions suitable for treatment by experts, mostly at the family or subfamily level. Over 100 taxonomic experts from around the world have been involved in identifying this material. Here, we analyze data from the first 165,000 specimens that have been identified, representing about 1 % of the catch. The identifications fall into 3,916 species, of which 609 have been putatively identified as new to science. Unlike data from most other insect inventories, the identified material is focused to a large extent on poorly-known insect groups.

We analyzed the inventory data in 79 subsets, each corresponding to a taxon or assemblage of related taxa examined by a particular expert or group of experts (Tables S3–S4). The majority of the data (35 groups; 120,530 specimens) consists of abundance records for poorly known groups, where taxonomists have found new Swedish species in the material. Together, these groups account for 29 % (2,525 of 8,655 species) of the estimated fauna that had not been recorded in 2003 (Table 1). Using the ratio of new to previously known species in the catch of these groups to estimate total species richness (analogous to mark-recapture approximation of

population size) produces results that are generally in line with expert guesses for these groups (Table 1; Table S5).

We also used statistical estimators of the total species pool (see Methods and Table S5 for details). These are based on abundance or incidence data of the identified species, and do not use information about whether or not the species were known previously. Specifically, we focused on the non-parametric estimators commonly known as Chao1 and Chao2, together with a new estimator developed for this paper, the combined non-parametric estimator (CNE). Chao1 [14–16] is based on the accumulation of new species as individual specimens are added to the total sample, while Chao2 [15–17] focuses instead on the accumulation of new species across subsamples (different Malaise trap sites). The CNE estimator combines the Chao1 and Chao2 extrapolations, giving a reasonable estimate of the total species pool under some assumptions concerning the interaction between the spatial and temporal accumulation of diversity (see Methods).

Estimating species richness is difficult because this statistic is as sensitive to rare and hard-to-detect species as it is to common ones [18], and most estimators tend to underestimate actual diversity [15, 16]. We tested the accuracy of the chosen estimators with abundance data for well-known groups, and for poorly-known groups using the species pool known in 2003 (Table S5). The tests show that estimates tend to become more accurate as the number of specimens, sites, or specimens per site increases. However, it is the coverage (proportion of the species pool that has been sampled) that largely determines accuracy: it is only at relatively high coverage (50–60% or more of the total number of species in the pool) that the estimates become reasonably robust. Chao1 and Chao2 consistently underestimate true species diversity unless the coverage is very high, while CNE does better in terms of bias at the cost of increased variance (Extended Data Figs. 1–2).

The poorly-known groups in the inventory are represented by six times as many specimens as the well-known groups, and the inventory sample covers 50% or more of the estimated fauna in many cases (Table S5; Extended Data Fig. 3), suggesting that species richness estimates may be reliable or at least indicative of true species richness for these groups. The CNE estimates (Table 1; see also Table S5) also tend to be in line with expert guesses and mark-recapture estimates, although there is

considerable variation across groups. Conspicuous outliers are associated primarily with groups where the inventory sample does not cover the expected fauna well (Table S5).

A geographic breakdown of the inventory data for some key taxa shows that many of the new discoveries, including species new to science, are made in mixed forests in the boreal and boreonemoral zones (Fig. 1; Extended Data Fig. 4). Including species identified as new to science but not described yet, approximately 28,000 insect species are known from Sweden, leaving 5,000 species (15 %) expected to be discovered from the fauna. This possibly represents the smallest gap between the known and true species stock of any major insect fauna studied to date. Given this, and the consistency across the three different extrapolation methods, we may be justified in drawing some tentative conclusions concerning the true size and composition of the fauna.

Knowledge of the fauna has always been worse than biologists realized at the time. Less than 5 % of the true fauna was known to Linnaeus [19], mainly conspicuous insects living on plants, such as butterflies, moths and herbivorous beetles (Extended Data Fig. 5). During the 20th century [13, 20], as the fauna became better known, the proportion of Hymenoptera and Diptera increased dramatically, as well as the proportion of decomposers and parasitoids. These trends are quite noticeable even when comparing the fauna known in 2003 with the true fauna estimated here (Fig. 2).

Given the current estimate, we can, for the first time, answer questions concerning the taxonomic and ecological composition of the fauna with some precision. For instance, our results indicate that species of plant feeders in the Swedish fauna on average host nearly one parasitoid species each, while the decomposer and predator guilds are attacked by much fewer parasitoids (Fig. 2). Taxonomically, the Swedish fauna is dominated by Hymenoptera and Diptera (Fig. 3), both orders dwarfing the famous insect radiations in the Coleoptera (beetles) and Lepidoptera (butterflies and moths). In fact, there are more than twice as many species of parasitic wasps as there are beetle species.

Many of the patterns described for European insect faunas [21] could be affected by biases similar to the ones that have historically thwarted our understanding of the

Swedish fauna. For instance, when controlling for area, there is no latitudinal gradient in the richness of known European insect faunas (Extended Data Fig. 6) but there is a significant northern increase in the proportion of Hymenoptera and Diptera species, and in the proportion of decomposers and parasitoid species (Extended Data Fig. 7). However, these patterns could be explained largely or partly by northern faunas being better known than southern ones. The Swedish inventory data for the nemoral to boreal transition suggest that there is indeed an increase in the proportion of Diptera and possibly also decomposers with increasing latitudes, but there is no such trend for Hymenoptera or parasitoids (Extended Data Fig. 8).

A recent study of the Canadian insect fauna based on DNA barcoding of Malaise trap catches [22] agrees with our study in finding extensive undocumented Diptera diversity, more than doubling previous estimates, but fails to detect a similar effect for Hymenoptera. Thus, the Canadian fauna is supposedly much richer in Diptera than in Hymenoptera, in contrast to the inferred Swedish fauna. It is possible that the faunas are indeed different, but we consider it more likely that the Canadian results are misleading because of the low coverage of many critical Hymenoptera taxa and the low success in barcoding hymenopterans [see also 23]. The Canadian study also covers a larger geographic area, analyzes a larger and less studied fauna, and lacks the focus on critical taxa characterizing the data reported here.

It appears unlikely that the apparent difference in taxonomic composition between the Swedish and Canadian insect faunas is due to undetected cryptic diversity in Sweden. DNA barcoding tends to be consistent with traditional morphology-based taxonomy when sufficient resources have gone into the latter [24–27]. Barcoding of critical Swedish Diptera taxa (Phoridae, Cecidomyiidae, Chironomidae, and Mycetophilidae) have revealed 0–20 % cryptic diversity [28–30] but similar results have been obtained for parasitic Hymenoptera [31].

Tropical insect faunas are famous for their richness in beetles, but recent studies suggest that the poor representation of Diptera and Hymenoptera in these faunas may not be real, as has been assumed in the past [32–33], but due to insufficient study [34–40]. What is abundantly clear is that our current understanding of the size and composition of insect faunas is seriously biased because of the lack of taxonomic study of Diptera and Hymenoptera. This impedes our understanding of how insect

faunas are composed, and our ability to monitor the ecosystem services they provide. Addressing this knowledge gap is all the more urgent as we may possibly be facing a rapid decline in the abundance and diversity of insects [41–42].

1. Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. 2011. How Many Species Are There on Earth and in the Ocean? *PLoS Biology* 9: e1001127.
2. Novotny V, Basset Y, Miller SE, Weiblen GD, Bremer B, Cizek L, Drozd P. 2002. Low host specificity of herbivorous insects in a tropical forest. *Nature* 416: 841–844.
3. Hamilton AJ et al. 2010. Quantifying uncertainty in estimation of tropical arthropod species richness. *American Naturalist* 176: 90–95.
4. Stork NE, McBroom J, Gely C, Hamilton AJ. 2015. New approaches narrow global species estimates for beetles, insects and terrestrial arthropods. *Proceedings of the National Academy of Sciences of the USA* 112: 7519–7523.
5. Bassett Y et al. 2012. Arthropod diversity in a tropical forest. *Science* 338: 1481–1484.
6. Wirta H et al. 2015. Establishing a community-wide DNA barcode library as a new tool for arctic research. *Molecular Ecology Resources* 16: 809–822.
7. Ronquist F, Gärdenfors U. 2003. Taxonomy and biodiversity inventories: Time to deliver. *Trends in Ecology and Evolution* 18: 269–270.
8. Miller G. 2005. Linnaeus's legacy carries on. *Science* 307: 1038–1039.
9. Venter C et al. 2004. Environmental shotgun sequencing of the Sargasso Sea. *Science* 304: 66–74.
10. Tedersoo L et al. 2014. Fungal biogeography. Global diversity and geography of soil fungi. *Science* 346: 1078.
11. Karlsson D, Pape T, Johanson KA, Liljeblad J, Ronquist F. 2005. Svenska Malaisefälleprojektet, eller hur många arter steklar, flugor och myggor finns i Sverige? *Entomologisk Tidskrift* 126: 43–53.
12. Townes H. 1972. A light weight Malaise trap. *Entomological News* 83: 239–247.
13. Gärdenfors U, Hall R, Hallingbäck T, Hansson HG, Hedström L. 2003. *Djur, svampar och växter i Sverige. Förteckning över antal arter per familj.* Uppsala: SLU Publikationsservice. 89 pp.
14. Chao A. 1984. Non-parametric estimation of classes in a population. *Scandinavian Journal of Statistics* 11: 265–270.
15. Colwell RK, Coddington JA. 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* 345: 101–118.
16. Magurran AE. 2003. *Measuring Biological Diversity*. Oxford: Blackwell. 264 pp.
17. Chao A. 1987. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 43: 783–791.
18. Chao A, Jost L. 2012. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* 93: 2533–2547.
19. Linnaeus C. 1761. *Fauna Svecica*. Stockholm: Laurentii Salvii. 578 pp.
20. Tullgren A, Wahlgren E. 1922. *Svenska insekter. En orienterande handbok vid studiet av vårt lands insektfauna*. Stockholm, Norstedt. 812 pp.
21. de Jong Y et al. 2014. Fauna Europaea – all European animal species on the

web. *Biodiversity Data Journal* 2: e4034.
22. Hebert PD, Ratnasingham S, Zakharov EV, Telfer AC, Levesque-Baudin V, Milton MA, Pedersen S, Jannetta P, deWaard JR. 2016. Counting animal species with DNA barcodes: Canadian insects. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 371: 20150333.
23. Langor DW. 2019. The diversity of terrestrial arthropods in Canada. *ZooKeys* 819: 9–40.
24. Schmidt S, Schmidt-Egger C, Moriniere J, Haszprunar G, Hebert PDN. 2015. DNA barcoding largely supports 250 years of classical taxonomy: identifications for Central European bees (Hymenoptera: Apoidea partim). *Molecular Ecology Resources* 15: 985–1000.
25. Hendrich L, Moriniere J, Haszprunar G, Hebert PDN, Hausmann A, Kohler F, Balke M. 2015. A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3500 identified species in BOLD. *Molecular Ecology Resources* 15: 795–818.
26. Hausmann A, Godfray HCJ, Huemer P, Mutanen M, Rougerie R, van Nieukerken EJ, Ratnasingham S, Hebert PDN. 2013 Genetic patterns in European geometrid moths revealed by the Barcode Index Number (BIN) system. *PLoS ONE* 8: e84518.
27. Pentinsaari M, Hebert PDN, Mutanen M. 2014. Barcoding beetles: a regional survey of 1872 species reveals high identification success and unusually deep interspecific divergences. *PLoS ONE* 9, e108651.
28. Häggqvist S, Ulefors SOU, Ronquist F. 2015. A new species group in *Megaselia*, the *lucifrons* group, with description of a new species (Diptera: Phoridae). *ZooKeys* 512: 89–108.
29. Kjærandsen J. 2017. Building the DNA barcode library of Holarctic Mycetophilidae. Genome 60: 74.
30. Jaschhof M. 2017. A taxonomic review of Palaearctic *Peromyia* Kieffer (Diptera: Cecidomyiidae, Micromyinae), with descriptions of thirty-eight new species mostly from Sweden. Studia dipterologica 23: 51–116.
31. Stigenberg J, Ronquist F. 2011. Revision of the Western Palearctic Meteorini (Hymenoptera, Braconidae), with a molecular characterization of hidden Fennoscandian species diversity. *Zootaxa* 3084: 1–95.
32. Gauld I. 1987. Some factors affecting the composition of tropical ichneumonid faunas. *Biological Journal of the Linnean Society* 30: 299–312.
33. Janzen DH. 1981. The peak in North American ichneumonid species richness lies between 38° and 42°N. *Ecology* 62: 532–537.
34. Noyes JS. 1989. The diversity of Hymenoptera in the tropics with special reference to Parasitica in Sulawesi. *Ecological Entomology* 14: 197–207.
35. Jones OR, Purvis A, Quicke DLJ. 2012 Latitudinal gradients in taxonomic overdescription rate affect macroecological inferences using species list data. *Ecography* 35: 333–340.
36. Quicke DLJ. 2012. We know too little about parasitoid wasp distributions to draw any conclusions about latitudinal trends in species richness, body size and biology. *PLoS ONE* 7: e32101.
37. Rodriguez JJ, Fernández-Triana, JL, Smith MA, Janzen, DH, Hallwachs W, Erwin TL, Whitfield JB. 2012. Extrapolations from field studies and known faunas converge on dramatically increased estimates of global microgastrine parasitoid wasp species richness (Hymenoptera: Braconidae). *Insect Conservation and Diversity* 6: 530–536.

38. Veijalainen A, Wahlberg N, Broad GR, Erwin TL, Longino JT, Sääksjärvi IE. 2012. Unprecedented ichneumonid parasitoid wasp diversity in tropical forests. *Proceedings of the Royal Society B: Biological Sciences* 279: 4694–4698.
39. Borkent A et al. 2018. Remarkable fly (Diptera) diversity in a patch of Costa Rican cloud forest: Why inventory is a vital science. *Zootaxa* 4402: 53–90.
40. Forbes AA, Bagley RK, Beer MA, Hippee AC, Widmayer HA. 2018. Quantifying the unquantifiable: why Hymenoptera, not Coleoptera, is the most speciose animal order. *BMC Ecology* 18: 21.
41. Hallmann CA et al. 2017. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLoS ONE* 12: e0185809.
42. Lister BC, Garcia A. 2018. Climate-driven declines in arthropod abundance restructure a rainforest food web. *Proceedings of the National Academy of Sciences of the USA* 115: E10397–E10406.

Acknowledgements. We would like to thank Kajsa Glemhorn, Pelle Magnusson, and Mareike Kiupel for invaluable assistance in coordinating the Malaise trap inventory. Anders Göthberg, Bengt-Åke Bengtsson, Gerhard Bächli, Gösta Gillerfors (†), Aurel I. Lozan, Emilia Nartshuk and many additional experts and volunteers contributed in various ways to the inventory. FR was supported by the Swedish Research Council (grant 2014-05901), DK by Formas (grant 217-2004-2101) and JR by the Ministry of Culture of the Czech Republic (grant MK000100595). The Swedish Taxonomy Initiative funded the Malaise trap inventory and many of the taxonomic research projects contributing to the charting of the Swedish insect fauna since 2003.

Author contributions: FR designed and coordinated the study, supervised work, ran analyses and drafted the first version of the manuscript; MF, SiH, DK, RH, JB, KH assembled and cleaned data, collected expert estimates, contributed to analyses, and helped prepare figures, tables and supplementary information; TB developed the new species richness estimator and assisted with statistical analyses; DK organized the Malaise trap inventory; JA, BA, PNB, CCC, AF, MF, CAG, SvH, KH, RH, MJ, JK, SeK, SvK, AL, RM, MPo, MPr, MR, JR, MS, JS, IS, AT, SOU, OV, PW determined inventory material and provided expert estimates; UG supported and supervised the Swedish Taxonomy Initiative and the Malaise trap inventory; the final manuscript was a joint effort involving all authors.

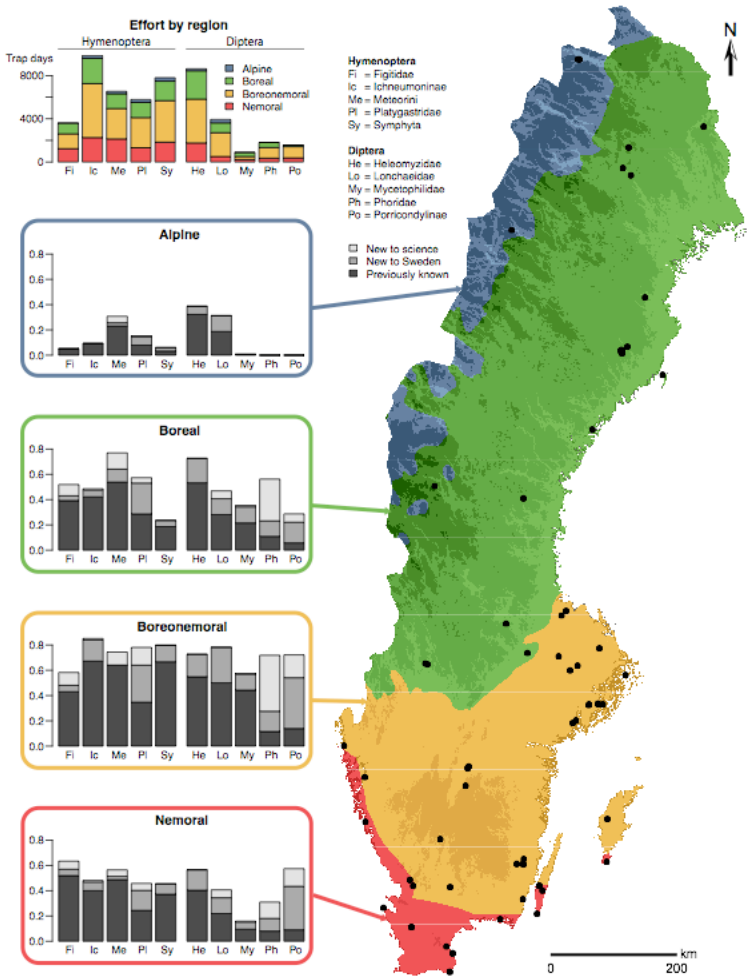


Figure 1 | Biogeographic breakdown of the results of the Malaise trap inventory for ten poorly known groups of Hymenoptera and Diptera. The species new to science, new to Sweden, and known previously in each biogeographic region are shown as a fraction of the total species pool of the group encountered in the inventory. The dots on the map mark trap sites. Many of the new species discovered in the inventory are found in the boreal and boreonemoral zones. For details on the circumscription of each group, see Table S4.

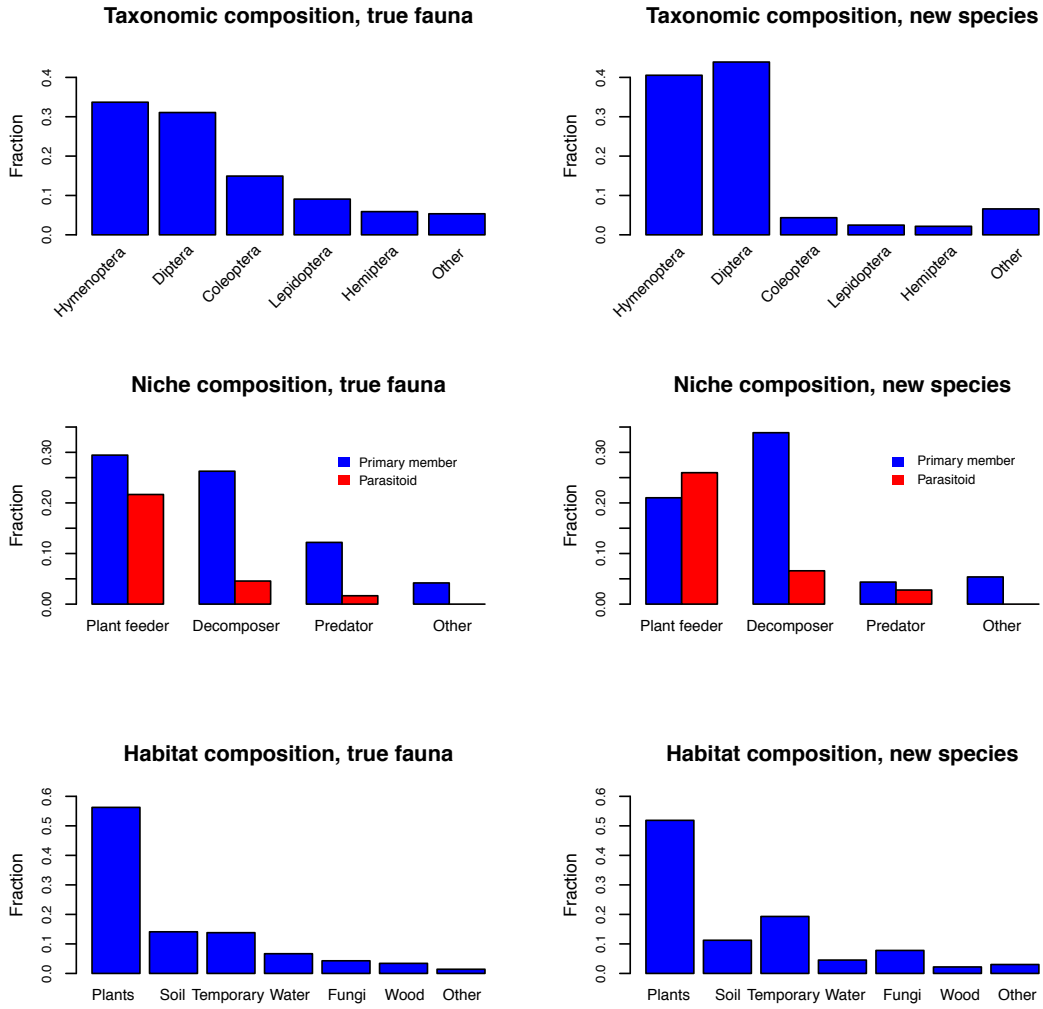


Figure 2 | Changes in our understanding of the composition of the Swedish insect fauna brought about by the inventory. The taxonomic and ecological composition of the true fauna, as estimated here (33,000 spp.), is given to the left, and the composition of the species that were still unknown in 2003 (8,600 spp., many of which still await discovery) to the right. The new species belong to a large extent to the Hymenoptera and Diptera. Based on their phylogenetic affinities, many of them are predicted to be decomposers or parasitoids, and more often associated with fungi or temporary habitats than other insect species. The new species substantially change our understanding of the composition of the fauna (see also Extended Data Figure 5).

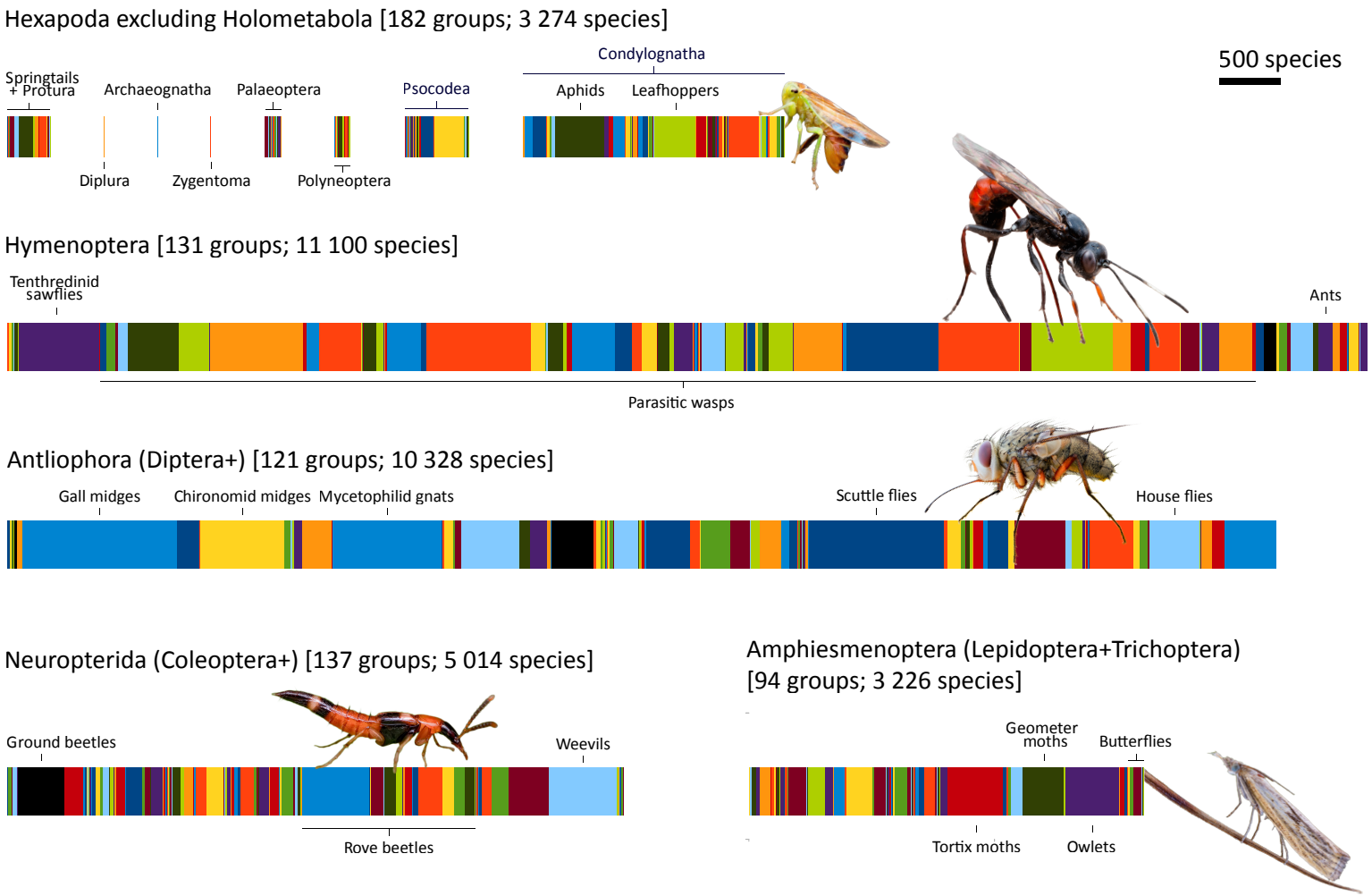


Figure 3 | Detailed view of the taxonomic composition of the true Swedish insect fauna, as estimated here. Each colored fragment represents a different family (subfamily for Ichneumonidae, Braconidae and Staphylinidae); the width is proportional to the number of species. The families are grouped into monophyletic higher clades according to our current understanding of insect relationships. The inventory has shown that the true fauna is much richer in Hymenoptera (sawflies, wasps, ants and bees) and Diptera (midges, gnats, mosquitoes and flies) than expected. These groups vastly outnumber Coleoptera (beetles), Lepidoptera (butterflies and moths) and other insect orders in terms of species diversity.

Table 1. Species richness estimates based on abundance data for poorly known insect groups from the Malaise trap inventory. "Known spp 2003" is the number of species known from Sweden prior to the inventory. For the inventory material, we give the proportion of the catch (Prop) and the number of specimens (N) processed, as well as the total number of species found (Spp), and the number of those species that were not recorded as Swedish prior to the inventory (New). Species richness estimates include expert guesses (Expert), mark-recapture (MR), Chao's estimator based on the entire material (Chao1) or on occurrences of species across sites (Chao2); and the new combined non-parametric estimator introduced here (CNE). See Supplementary Information Table S5 for more detailed results and additional species richness estimates.

Taxon	**Known spp 2003**	**Malaise trap inventory catch**				**Species richness estimates**				
		Prop	**N**	**Spp**	**New**	**Expert[1]**	**MR**	**Chao1**	**Chao2**	**CNE**
Dolichopodidae	314	38%	43011	204	31	356-360	370	219	235	246
Phoridae	182	6%	33797	863	743	1100-1100	1306	1171	1639	1982
Drosophilidae	59	33%	7405	39	4	66-80	66	46	63	106
Ichneumoninae (part)[1]	325	36%	5659	235	50	480-490	413	307	312	373
Heleomyzidae + Odiniidae	54	24%	4911	62	21	62-80	82	71	73	84
Sepsidae	34	16%	3870	16	1	32-38	36	16	16	16
Platygastridae (s. str.)	124	22%	2971	164	92	250-270	282	186	218	267
Symphyta	584	30%	2912	183	34	875-799	717	238	256	318
Diplazontinae	49	37%	2788	63	24	80-90	79	68	69	73
Thysanoptera	119	7%	1528	34	1	117-186	123	54	65	113
Pimplinae	127	28%	1521	64	9	150-150	148	112	97	121
Euphorinae: Meteorini	37	27%	1304	39	10	60-53	50	42	42	43
Piophilidae	17	24%	1063	11	1	17-25	19	11	11	11
Sciomyzidae	77	25%	1021	58	3	85-83	81	61	93	140
Other groups	888	2-70%	6769	474	148	1358-1681	1289	627	887	1268
TOTAL SAMPLE	*2990*	*2-70%*	*120530*	*2509*	*1172*	*5088-5515*	*5610*	*3234*	*NA*	*NA*
GROUP TOTALS	**2990**	**2-67%**	**120530**	**2509**	**1172**	**5088-5515**	**5057**	**3228**	**4076**	**5159**

[1]First estimate from 2007, prior to inventory material being processed; second estimate current.
[2]Excludes the tribe Phaeogenini.

Methods

Current knowledge of the Swedish insect fauna. To assess the current knowledge of the Swedish insect fauna, we used the content of DynTaxa (http://dyntaxa.se), the official checklist of the Swedish flora and fauna. Data were pulled from DynTaxa on February 8, 2017 (provided in the Data Package). We included all species recorded as reproducing ("Reproducerande"), accidental and reproducing ("Tillfälligt reproducerande"), uncertain ("Osäker förekomst"), accidental and not reproducing ("Tillfällig förekomst, ej reproducerande"), nationally extinct ("Nationellt utdöd"), no longer reproducing ("Ej längre reproducerande") or possibly nationally extinct ("Möjligen nationellt outdöd"). Due to lack of knowledge, very few insect species occurrence records in DynTaxa are currently classified into any of these detailed categories; the species are simply assumed to be "reproducing". In well-known insect groups, a handful of species are known to be accidentals or belong to one of the other categories, but we chose to include them in the species counts to avoid systematic bias towards higher counts in poorly known groups.

Five species lacked occurrence status information in Dyntaxa: *Dixella martini* (Dixidae), *Aphis violae* (Aphididae), *Diaspis boisduvalii* (Diaspididae), *Cosmopterix pulchrimella* (Cosmopterigidae) and *Haplothrips vuilleti* (Phlaeothripidae). They were nevertheless included in our analysis as Swedish based on other data. *Sciarosoma borealis* (Diptera) is not classified to family in Dyntaxa, only to superfamily; it was included by us in the Diadocidiidae, following the classification in Fauna Europaea [21]. Five species in DynTaxa lack information about family classification, namely *Dimorphopterus spinolae* and *Ischnodemus sabuleti* in Blissidae, and three species of *Cymus* in Cymidae: *C. claviculus*, *C. glandicolor*, and *C. melanocephalus*. The family information was added for these records. For a few taxa, the number of Swedish species given in the DynTaxa 2017 checklist is obviously wrong and was corrected (details in Table S1).

Historical knowledge of the Swedish insect fauna. Information was extracted from the three most comprehensive surveys published before the start of the Swedish Taxonomy Initiative. The first [19] lists all Swedish insect species known to Linnaeus

in 1761. These species were mapped to currently valid species so that they could be placed in the classification used by us. The second [20] and third [13] do not provide species lists. Instead, they summarize the knowledge of the fauna at the time (1920–1922 and 2003, respectively) at higher taxonomic levels, mostly at the family level, and give the estimated number of known species of each taxon. We mapped these numbers onto the higher taxa in our classification as detailed in the "Taxonomic composition" section below and in Table S1. In addition, the 2003 numbers were corrected for some groups where they are obviously wrong (see Table S1 for details).

For our analyses of data from the Swedish Malaise Trap Project, we needed explicit species lists from 2003 of the target taxa. These were reconstructed with the help of experts and the literature on each group. The resulting check lists from 2003 are provided in the Data Package of the Supplementary Material.

Known European insect faunas. Data on the European insect fauna were obtained from the Fauna Europaea database (version 2.6, distribution 3) kindly provided by Yde de Jong. This version of the database was current as of February 2017 according to Falko Glöckler at the Museum für Naturkunde in Berlin (pers. comm.). The dump includes both taxonomic data and distribution data, detailing the occurrence of each species in each of the European countries and geographic regions recognized in Fauna Europaea. The dump is provided in the Data Package of the Supplementary Material.

For the latitudinal analysis, we excluded all areas outside of Europe, as well as European areas of countries that mostly lie outside of Europe (Russia and Turkey). We also excluded some exotic areas with faunas that are not typical of the country they belong to: Madeira, the Selagen Islands, the Azores, Canary Islands, and Gibraltar. Country latitude data were taken from a file provided by Google [43]. Land area data were taken from the EuroStat NUTS survey in 2013 [44] and complemented with data from Wikipedia [45]. The raw data are provided in the Data Package.

Taxonomic composition. Taxonomic composition of the Swedish fauna was analyzed in terms of the families and orders recognized in the DynTaxa classification as of February 8, 2017. However, at the family level, we divided the three most species-rich insect families (Ichneumonidae, Braconidae and Staphylinidae) into subfamilies to make the family-level units more comparable in size. When data on the

Swedish fauna were analyzed from older sources using a less detailed classification [13, 20], we split the listed species numbers based on the available literature from the time (see also Table S1). The analysis of the European data, and of taxa expected to occur in Sweden, was based on the Fauna Europaea taxonomy [21].

Ecological composition. In the analysis of life-history traits, we focused on two traits that are conservative enough that they can be reasonably assumed, in most cases, to be homogeneous within the family-level groups we used: the main feeding niche and the main feeding (micro-)habitat. This is the niche and habitat of the immature stages (the main feeding stages), and may or may not be the same as the niche and habitat of more short-lived adult stages. Data were taken from standard works [20, 46–50] complemented with data from relevant taxonomic specialists and the primary literature.

Specifically, we defined the feeding niches as follows:
Parasite. This includes bloodsuckers, endoparasites (botflies) and exoparasites (lice).
Phytophage (plant feeder). This includes both chewers and sap suckers, as well as stem borers, leaf miners, root feeders and gall inducers.
Phytophage-parasitoid. This is restricted to all primary parasitoids of plant feeders.
Predator. This is restricted to free-living predators, it does not include parasitoids.
Predator-parasitoid. This includes both primary parasitoids of predators and all hyperparasitoids (parasitoids of parasitoids).
Saprophage (decomposer). This includes scavengers, decomposers, fungivores and microflora/bacterial grazers.
Saprophage-parasitoid. This is restricted to all primary parasitoids of decomposers.

The (micro-)habitats of the main feeding stages (larvae, nymphs) were defined as follows:
Fungi. Within fruiting bodies of macrofungi.
Homeothermic animals. On homeothermic animals (birds, mammals). The habitat of parasites (bloodsuckers, endoparasites and exoparasites).
Plants. In or on aerial parts of living plants, excluding wood.
Soil. In soil, including ground litter as well as living roots.
Temporary habitats. In or on ephemeral resources of rich nutrients, such as bird nests, carrion, dung and fermenting sap. Also includes indoors habitats.

Water. In or on water.
Wood. In wood.

Groups are coded for the dominant trait; notable exceptions and other comments are given in the tables. Because of different taxon or group circumscriptions, the traits differ in a few cases between the Swedish fauna (Table S1), the groups used in analyzing inventory data (Table S4), and the European fauna (Table S6).

Expert estimates of the true fauna. Expert estimates of the total size of the Swedish fauna were based on the opinions of taxonomic specialists with intimate knowledge of the Swedish insect fauna. They were obtained for all family-level groups in the fauna at two occasions: in 2007, before the Malaise trap inventory material was processed, and in the spring of 2017 (current estimate). The 2007 estimates were based on less extensive surveys of experts, and we only give them here for the target groups used in the species richness analysis of the inventory data (Table S5). The estimates from 2017 are detailed in Table S1, together with the experts responsible for them.

For a few groups, we were not able to find local specialists willing to estimate the total size of the fauna. In those cases, MF, RH and FR estimated the total size of the Swedish fauna as follows. For small well-known groups (Diplura, Archaeognatha, Zygentoma, Dermaptera), the current number of known species was accepted as the estimate of the true fauna except in the case of Dermaptera, where we added one expanding species known from nearby countries (*Labidura riparia*). Protura are relatively well-known but rarely studied from a faunistic or systematic viewpoint, and are thus likely to contain undiscovered species. We estimated the true fauna to be 30 % larger than the currently known one (an increase by a factor of 1.3).

For the two main groups of vertebrate parasites, Phthiraptera and Siphonaptera, we based our estimates on each family's host preferences. Families of parasites exclusively or mainly on man and domestic mammals were considered well-known and we accepted the current number as final. For families of parasites on wild rodents, insectivores, bats and carnivores, plus siphonapteran parasites on birds, we calculated with an increase of current numbers by a factor of 1.4. For phthirapteran families including parasites of birds we used a factor of 1.6. This seemed reasonable considering the patterns of host range, the fauna of hosts in Sweden, the fauna of

these families in better studied neighboring countries, and the relatively poor global taxonomic knowledge of phthirapteran bird parasites. In two species-poor families where these considerations suggested no addition, we nevertheless added expected species that are known from neighboring countries (one each in Enderleinellidae and Hoplopleuridae).

Malaise trap inventory. The Swedish Malaise Trap Project is the first systematic inventory of the Swedish insect fauna, particularly targeting the most poorly known groups in the orders Diptera and Hymenoptera [11]. The material was collected using 73 Malaise traps, deployed at different sites throughout the country and representing a wide range of habitats (Fig. 1). Most of the traps were run continuously from 2003–2006, but some traps were run for only part of this period, and a couple of traps were run in the period 2007–2009. Detailed data on the trap sites and the samples collected (1919 samples in total) are provided in Table S2 (see also http://www.stationlinne.se/en/research/the-swedish-malaise-trap-project-smtp/). The trap sites were classified into four biogeographic regions (Fig. 1) and six habitat classes to simplify analyses of the data (Table S2).

The inventory material (unsorted material and identified specimens) is deposited at the Swedish Museum of Natural History (http://www.nrm.se) through its collaboration with the Station Linné Field Station (http://stationlinne.se). The material is continuously sorted into more than 300 taxonomic fractions (see http://www.stationlinne.se/en/research/the-swedish-malaise-trap-project-smtp/taxonomic-units-in-the-smtp/), which are sent out to experts for identification. The returned data are validated and cleaned, and identifications are matched to the national checklist (DynTaxa) as far as possible. Data on the sorted fractions is available in a separate dataset at the Global Biodiversity Information Facility portal (https://www.gbif.org) and at the Swedish natural history collection portal (https://naturarv.se); the latter version is updated on a daily basis. For access to the material and for questions about the data, contact DK or JB.

The data analyzed here represent all the cleaned data files that were available in the spring of 2017 (Table S3). In total, there are 127 files, most of which contain abundance data (117 files), but some of which only have incidence data (10 files). The incidence data include several groups (Braconidae: Cheloninae, Braconidae:

Rogadinae, Cecidomyiidae: Porricondylinae (*sensu lato*) and Psylloidea) for which abundance data were not reported consistently for all samples and species. The files are provided in the Data Package in the Supplementary Material.

We divided the data into 79 datasets, each corresponding to a set of samples of a particular taxon or assemblage of taxa ("analysis taxon") processed by a different expert or group of experts (Table S4). The data are available in the data files provided as Supplementary Material. The analysis taxa mostly correspond to families or subfamilies of Diptera and Hymenoptera, and to orders or major portions of orders for other insect groups.

The analyzed information is estimated to represent from 2 to 67 % of the entire catch of each analysis taxon (Table S4). In total, the abundance data comprise 157,225 specimens (30,643 records) identified to 3,916 taxa. The incidence data represent another 1,516 observations of 279 taxa. The incidence data roughly corresponds to 7,800 specimens, if this material contained the same average number of specimens per species and sample as the material covered by the abundance data.

Specimens or records that were determined only to genus level (marked by "sp." in the Species field in the data file provided in the Supplementary Material) were removed prior to analyses of species richness and compositional trends. This affected 14,034 specimens (161 taxa) in the abundance data (most of these are females of Phoridae, which cannot at this point be identified to species) and 36 records (8 taxa) in the incidence data. In a few cases, determinations were to unspecified species that may or may not be distinct (affecting 40 records of abundance data marked by "sp. indet." in the Author field); these records were included in the analyses under the assumption that they represent distinct species.

Species richness estimates. To estimate the total Swedish fauna of the target groups from SMTP inventory data, we used the R package vegan, version 2.4-3 [51]. There are two types of non-parametric estimators provided by vegan. The first set is based on abundance data, and include the estimator commonly referred to as Chao1 [16] (called abundance-based Chao in the vegan documentation) and ACE (the abundance-based coverage estimator). These estimators are provided by the vegan function 'estimateR'.

The second set of estimators is available through the vegan function 'specpool', and include the estimator commonly refer to as Chao2 [16], as well as the first-order jackknife (Jack1), the second-order jackknife (Jack2) and the bootstrap (Boot) estimators. The second set of estimators is based on the incidence of species at the sample sites and can be used both for abundance and incidence data, whereas the first set of estimators is only applicable to abundance data. The vegan package also provides a parametric species richness estimator (Preston) based on the Preston model through the functions 'prestonfit' and 'veiledspec'. It requires abundance data, which are fitted to a lognormal model. See the vegan documentation for a discussion of these estimators and how they are implemented in vegan, with references to the original literature.

We complemented the vegan estimators with two additional estimators. One is analogous to mark-recapture estimation of population size. In the simplest form, the mark-recapture estimator (MR) simply uses the fraction of new ('unmarked') species encountered in a sample to extrapolate the number of known species (prior to sampling) to the total number of species in the species pool. However, this estimator is biased when the sample size is small, and it is undefined when no species of the group are known prior to sampling.

These problems are addressed by the Chapman version of the MR estimator [52], which is defined as

$$\hat{S}_{MR} = \frac{S_{obs} + 1}{S_{obs,known} + 1} \times (S_{known} + 1) - 1,$$

where

S_{obs} = number of observed species,

$S_{obs,known}$ = number of observed species known previously, and

S_{known} = number of known species.

Finally, we developed a new non-parametric estimator for this paper, which we call the **combined non-parametric estimator** (**CNE)**. It is based on the idea of combining the Chao1 and Chao2 extrapolations. The Chao1 extrapolation is based on

the number of singleton and doubleton specimens found in the total sample. This might be expected to estimate the fauna of the sample sites more accurately than the total fauna, which is particularly problematic when trying to estimate the fauna of a large and heterogeneous area from a sample coming from a small number of sites. In contrast, the Chao2 extrapolation is based on the number of species that occur at only one sampling site (singleton species) or at two sampling sites (doubleton species). Thus, Chao2 explicitly takes the heterogeneity across sites into account, which should result in a better estimate of the total fauna under these conditions.

In the Chao2 case, however, it is possible that we have underestimated the number of singleton and doubleton species because of undersampling of the selected sites. That is, more extensive sampling at the same sites would potentially result in the discovery of more of the highly localized singleton and doubleton species occurring there. Of course, more extensive sampling might also reduce the number of singleton and doubleton species because some species may turn out to be more widespread than indicated by the original sample. If many of the true singleton and doubleton species are rare at the sites where they occur, then it might be reasonable to assume that the proportion of singleton and doubleton species observed at a site would remain the same even with more extensive sampling there.

This leads to an estimator applying a Chao1-type extrapolation to obtain an estimate of the number of singleton and doubleton species one would have observed with more extensive sampling of the selected sites. The estimated numbers of singleton and doubleton species can then be used to estimate the size of the total fauna using Chao2-type extrapolation.

Formally, define

$S^{(i)}$ = number of species at site *i*,

$S_{obs}^{(i)}$ = number of species observed at site *i*,

$S_1^{(i)}$ = number of species that only occur at site *i*,

$S_2^{(i)}$ = number of species that only occur at site *i* and one other site,

$S_{obs}^{(i)}(1)$ = number of species observed at site *i* and only there,

$S_{obs}^{(i)}(2)$ = number of species observed at site *i* and one other site,

$F_1^{(i)}$ = number of observed species at site *i* that are only observed once there,

$F_2^{(i)}$ = number of observed species at site *i* that are only observed twice there,

Q_1 = number of species that only occur at one site, and

Q_2 = number of species that only occur at two sites.

We now have the following estimators:

$$\hat{S}^{(i)} = S_{obs}^{(i)} + \frac{F_1^{(i)}(F_1^{(i)} - 1)}{2F_2^{(i)} + 1}$$

estimates the number of species at site *i* (this is the same as the Chao1 estimator),

$$\hat{S}_1^{(i)} = S_{obs}^{(i)}(1) + \frac{F_1^{(i)}(F_1^{(i)} - 1)}{2F_2^{(i)} + 1} \times \frac{S_{obs}^{(i)}(1)}{S_{obs}^{(i)}}$$

estimates the number of unique species at site *i*, and

$$\hat{S}_2^{(i)} = S_{obs}^{(i)}(2) + \frac{F_1^{(i)}(F_1^{(i)} - 1)}{2F_2^{(i)} + 1} \times \frac{S_{obs}^{(i)}(2)}{S_{obs}^{(i)}}$$

estimates the number of species occurring there and at one other site.

Since

$$Q_1 = \sum_{i=1}^{n} S_1^{(i)}$$

and

$$Q_2 = \frac{1}{2}\sum_{i=1}^{n} S_2^{(i)},$$

we obtain the estimators

$$\hat{Q}_1 = \sum_{i=1}^{n} \hat{S}_1^{(i)}$$

and

$$\hat{Q}_2 = \frac{1}{2}\sum_{i=1}^{n} \hat{S}_2^{(i)}.$$

This gives us the final species richness estimator (with correction for small samples)

$$\hat{S}_{CNE} = S_{obs} + \frac{\hat{Q}_1(\hat{Q}_1 - 1)}{2\hat{Q}_2 + 1}.$$

Performance of species richness estimators. To assess the performance of the species richness estimators, we used test data from two sources. First, we used well-

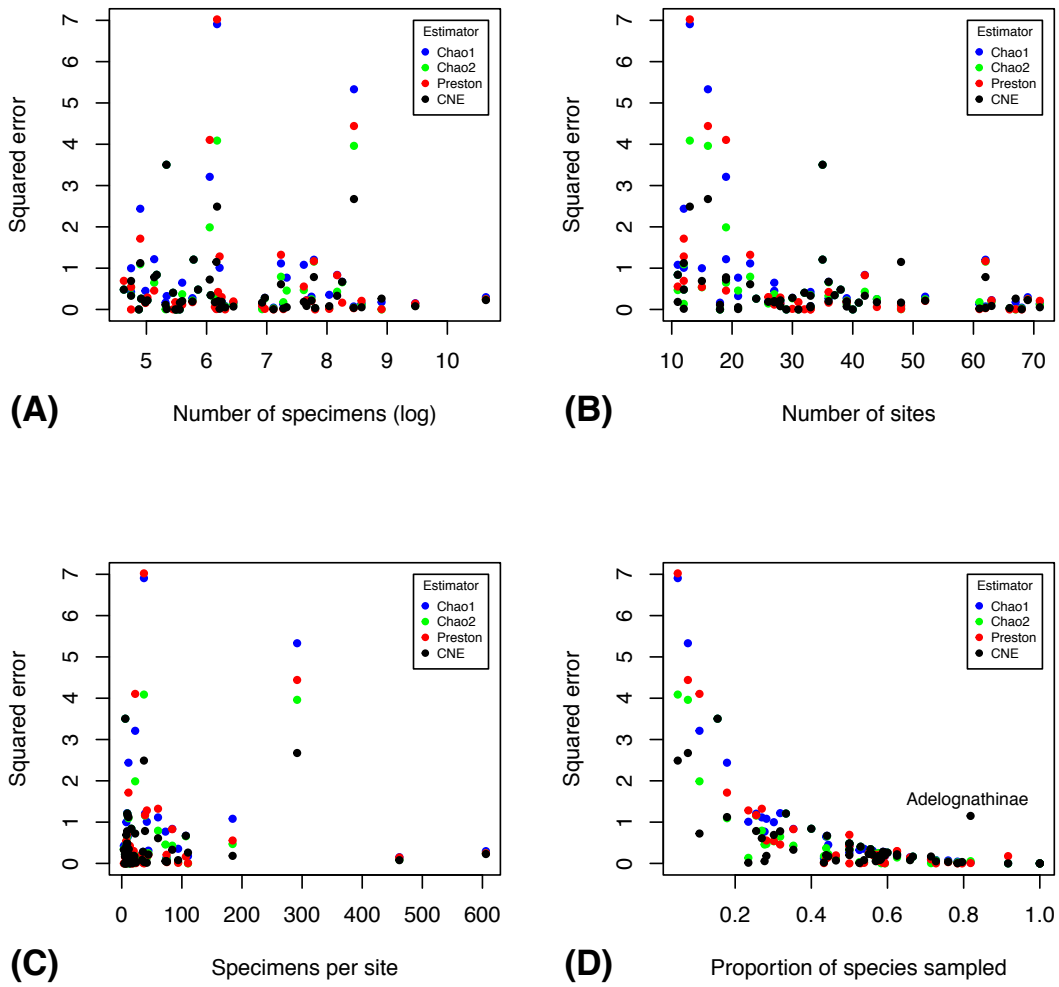
studied groups, defined as those groups where the SMTP inventory has not discovered any new species (Table S4). As the 'true' diversity for these groups, we used the number of species in the current expert estimate, which was typically very close to if not identical to the number of known species (Table S4). As a second source of test data, we used the subset of species known in 2003 in the more poorly studied groups, that is, those groups where the SMTP inventory did discover new species (Table S4). For these species pools, the 'true' diversity is known as the pool is completely defined a priori. We detected no qualitative difference between these two data sources, except that the first set typically included a smaller number of specimens and represented a much smaller fraction of the expected species pool. We only included groups with more than 100 specimens observed from more than 5 sites in the analyses. The accuracy of the estimators was measured using the squared error and bias of the log estimates, that is, $(\log \hat{S} - \log S)^2$ and $\log \hat{S} - \log S$, respectively, where $\hat{S}$ is the estimate and S the true value (Extended Data Figs. 1–2).

Analyses of compositional trends. For analyses of compositional trends, we used standard functions in R, version 3.3.1 [53]. The R scripts are available in the Supplementary Material.

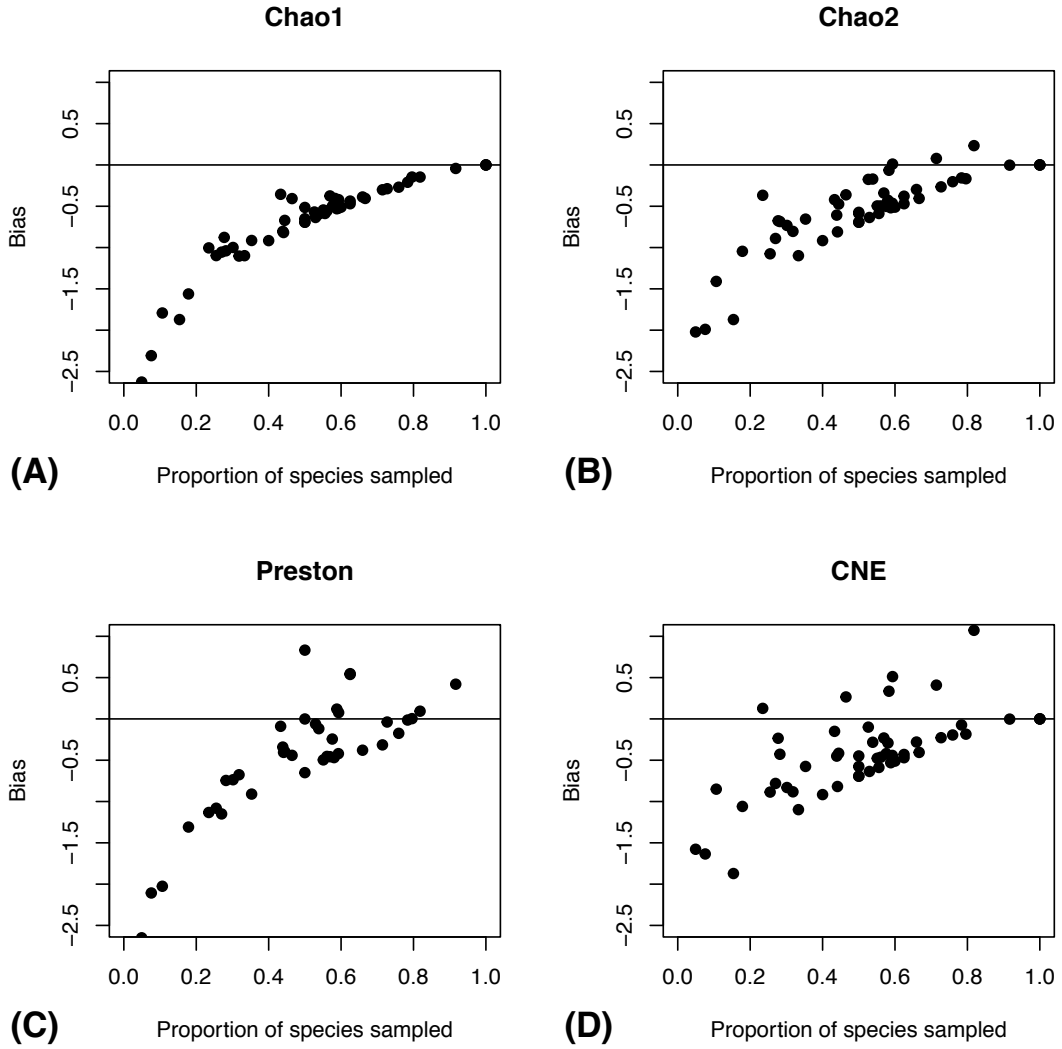
43. Google Dataset Publishing Language: geo canonical concept dataset, countries table. URL: https://developers.google.com/public-data/docs/canonical/countries_csv (accessed on June 29, 2017).
44. EuroStat: NUTS Area 2013. URL: http://ec.europa.eu/eurostat/documents/205002/6786255/AreaSize_20161109.xlsx (accessed on June 29, 2017).
45. List of countries and dependencies by area. URL: https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_area (accessed on June 29, 2017).
46. Landin B-O. 1967. *Fältfauna – Insekter. Del 1*. Stockholm: Natur & Kultur. 394 pp.
47. Landin B-O. 1970. *Fältfauna – Insekter. Del 2:1*. Stockholm: Natur & Kultur. Pp. 1–380.
48. Landin B-O. 1971. *Fältfauna – Insekter. Del 2:2*. Stockholm: Natur & Kultur. Pp. 381–1053.
49. Goulet H, Huber JT (eds). 1993. *Hymenoptera of the world: An identification guide to families*. Ottawa: Agriculture Canada. 680 pp.
50. Oosterbroek P. 2006. *The European families of the Diptera: identification, diagnosis, biology*. Zeist, The Netherlands: KNNV Uitgevereij. 205 pp.
51. Oksanen JF et al. 2017. vegan: Community Ecology Package. R package version 2.4-3. URL: http://CRAN.R-project.org/package=vegan (accessed on July 1, 2017).
52. Chapman DG. 1951. *Some properties of the hypergeometric distribution with*

applications to zoological sample censuses. University of California publications in statistics, Volume 1, Number 7. Berkeley: University of California Press.

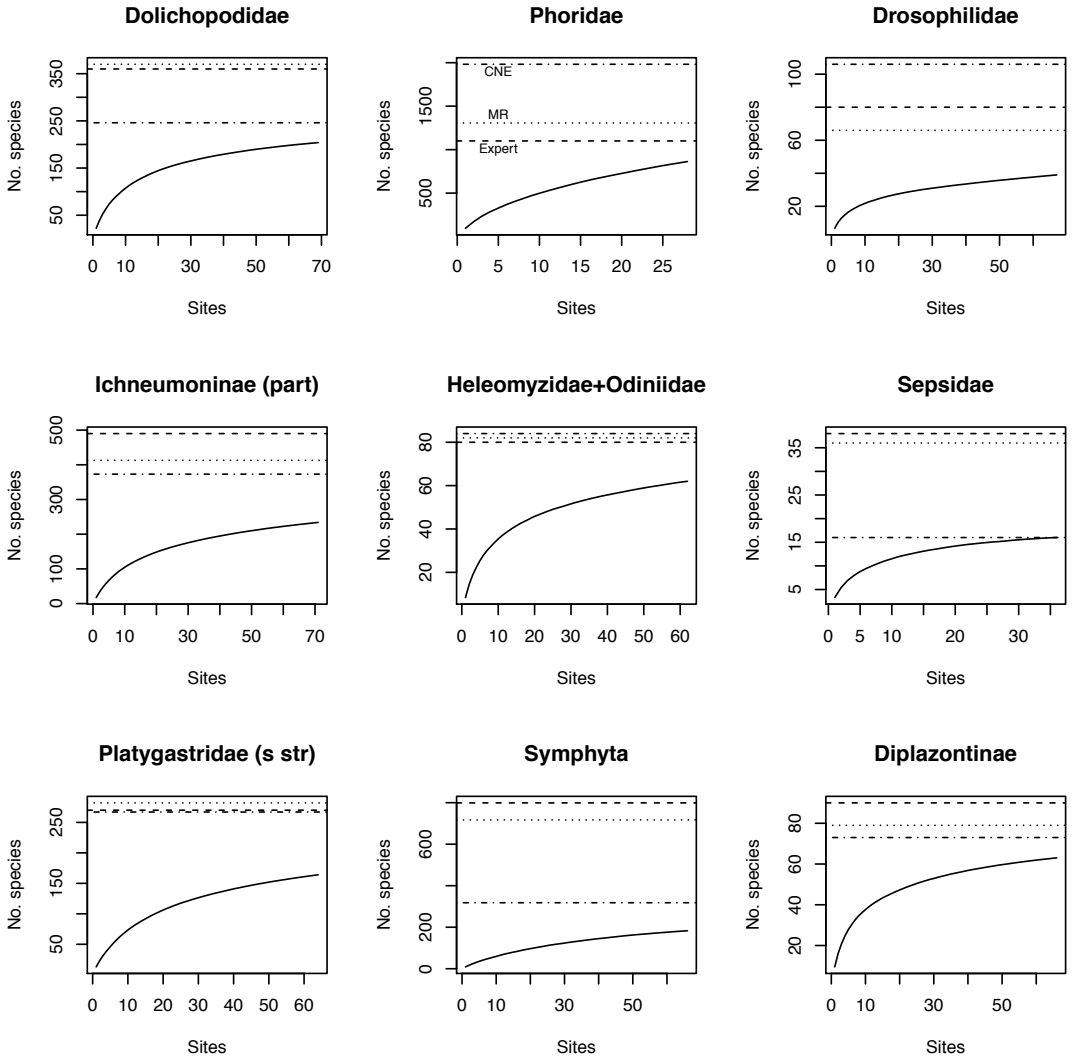
53. R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.



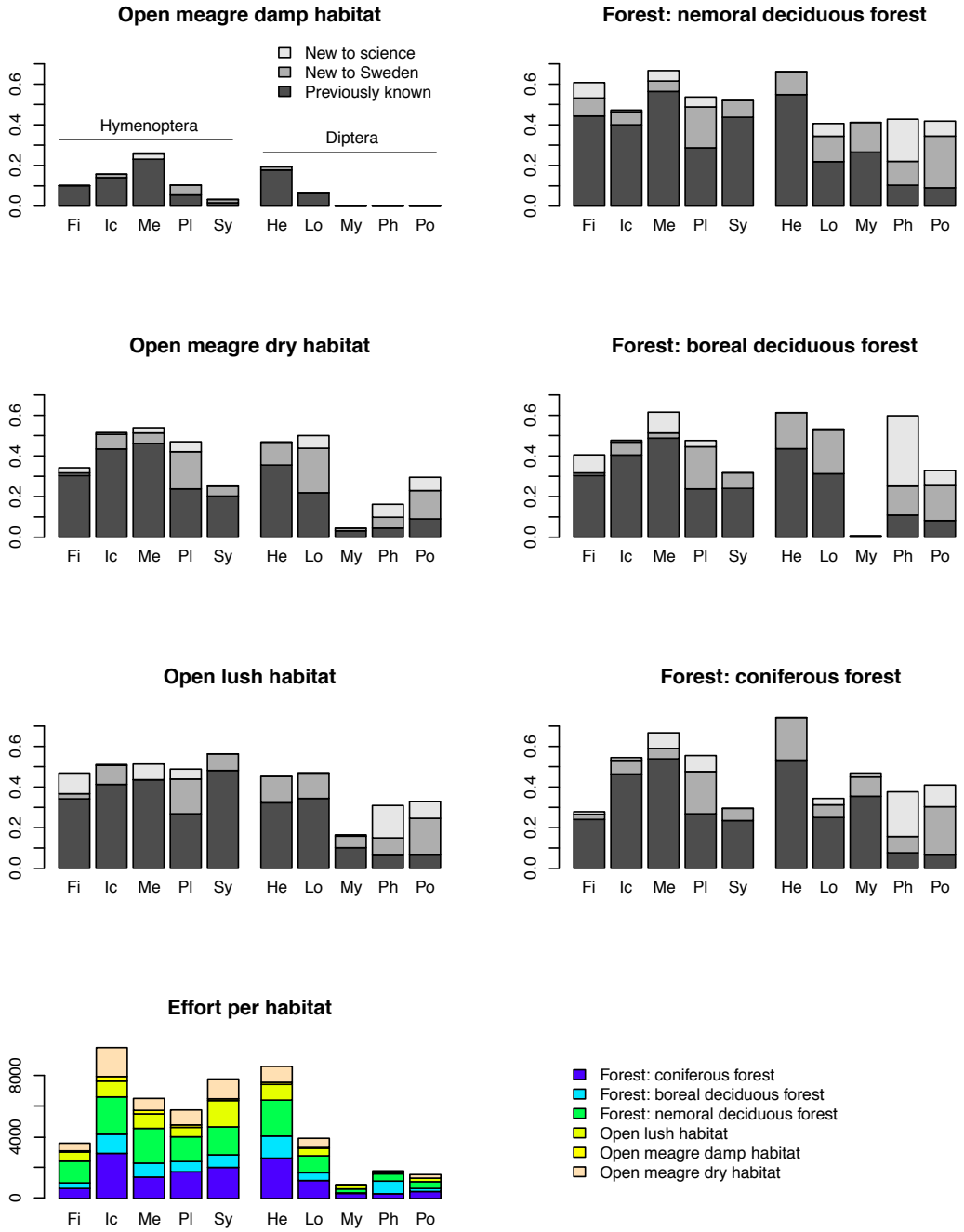
Extended Data Figure 1 | Accuracy of different estimators in predicting the species richness of well-studied groups, and the species pool known in 2003 of more poorly studied groups. The accuracy (measured as squared error of log estimates) is improved by increasing the number of specimens (A), the number of sites (B), or the number of specimens per site (C). However, the most important factor influencing the accuracy is the fraction of the species pool sampled (D). When half the species have been sampled, one can expect a worst-case squared error of the log estimates of approximately 0.4, which means that the estimate may be off by a factor of 1.9 (close to half or double the real species number). The average estimate is considerably better, however, especially for estimators other than Chao1. The CNE outlier in (D) is Ichneumonidae: Adelognathinae, one of the groups with the smallest number of specimens per site.



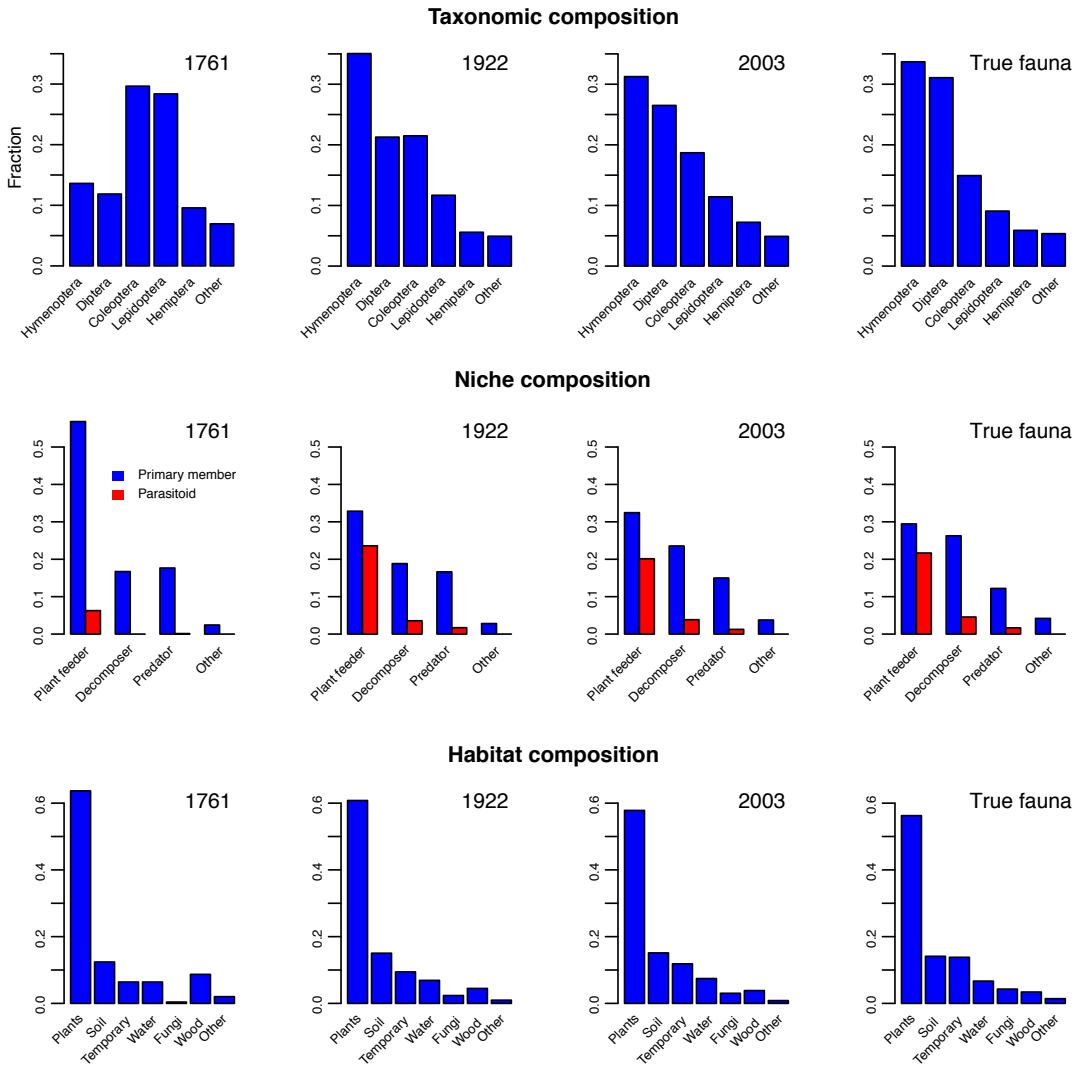
Extended Data Figure 2 | Bias in species richness estimators as a function of the fraction of the species pool sampled. Bias is measured on the log scale; the horizontal line represents absence of bias. All estimators are biased downwards, that is, they underestimate the size of the true fauna, especially when a small proportion of the species pool is sampled. The bias is less severe for the Preston and CNE estimators, but this comes at the expense of increased variance. Note the improved performance of the estimators that take heterogeneity across sites into account (Chao2 and CNE) when a small fraction of the species pool is sampled.



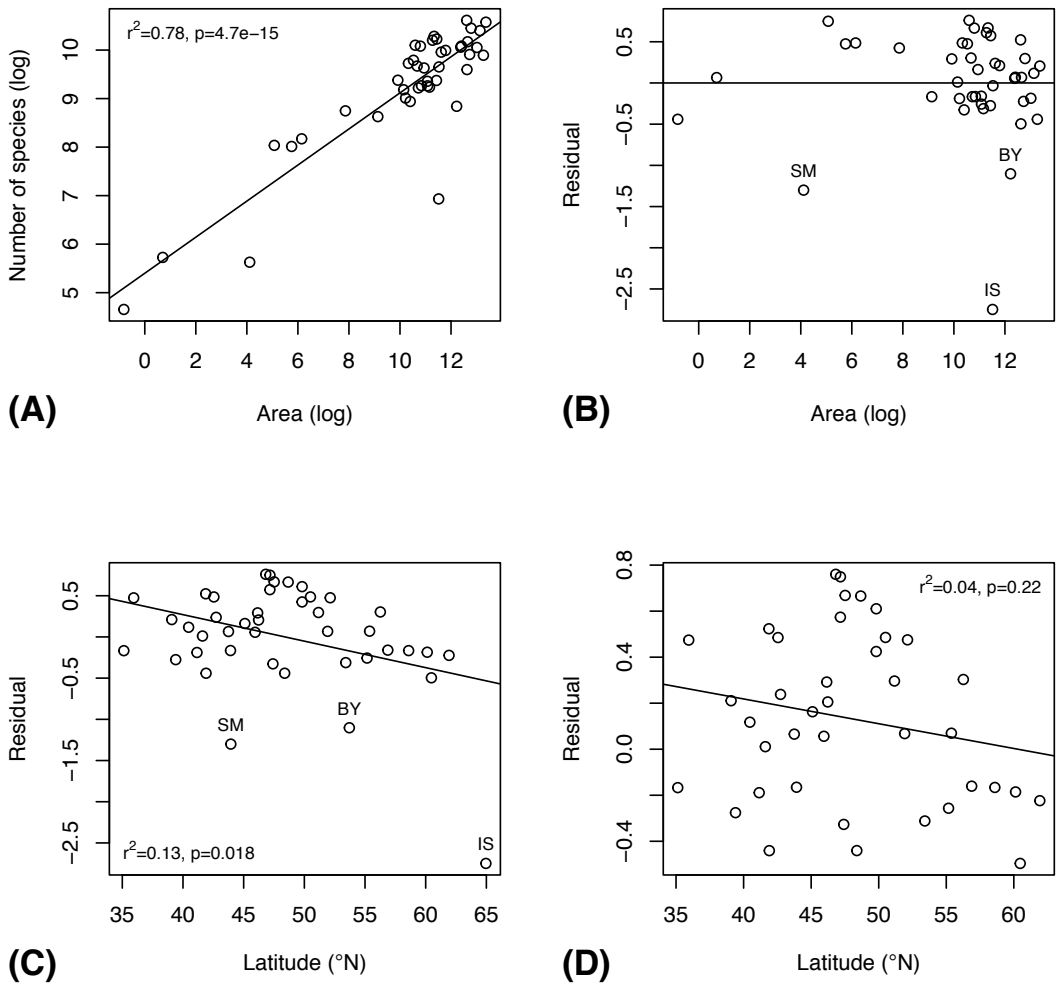
Extended Data Figure 3 | Species accumulation curves for some groups studied in the Malaise trap inventory. Accumulation curves are shown for the nine groups with the largest number of identified specimens (Tables 1, S5). Accumulation curves are based on 10,000 random perturbations of the accumulation order of sites. The horizontal lines correspond to the expert guess from 2017 (Expert, dashed line), the estimate from the mark-recapture method (MR, dotted line) and the estimate from the combined nonparametric method (CNE, dashed and dotted line). Major disagreement among estimates tends to be associated with groups that are not covered well by the sample. In the Phoridae, the low coverage may be partly due to the small fraction (6%) of the material processed; in other cases, like the Symphyta (30 % of the material processed), it appears instead that it is due to the inefficiency of Malaise traps in collecting these groups. DNA barcoding indicates that the accumulation curve for Phoridae is too steep because aberrant male-genitalia variants are sometimes misclassified by taxonomists as representing separate species [28]; this is likely to partly explain the poor performance of species richness estimators for this group.



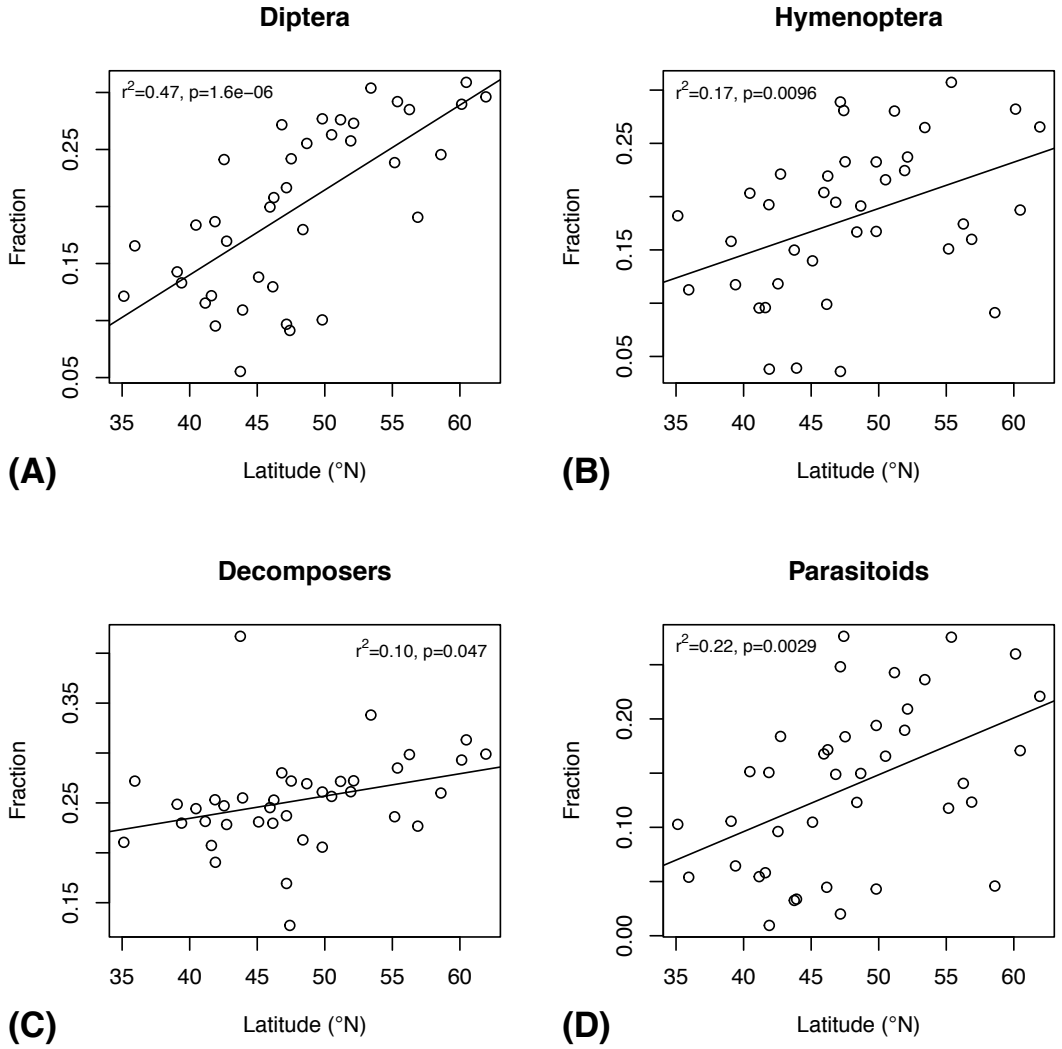
Extended Data Figure 4 | Habitat breakdown of the species found in the Malaise trap inventory. Data are given for the same taxonomic groups as those in Figure 1. See Table S4 for detailed circumscription of each analysis group. Many of the new species are found in forest habitats.



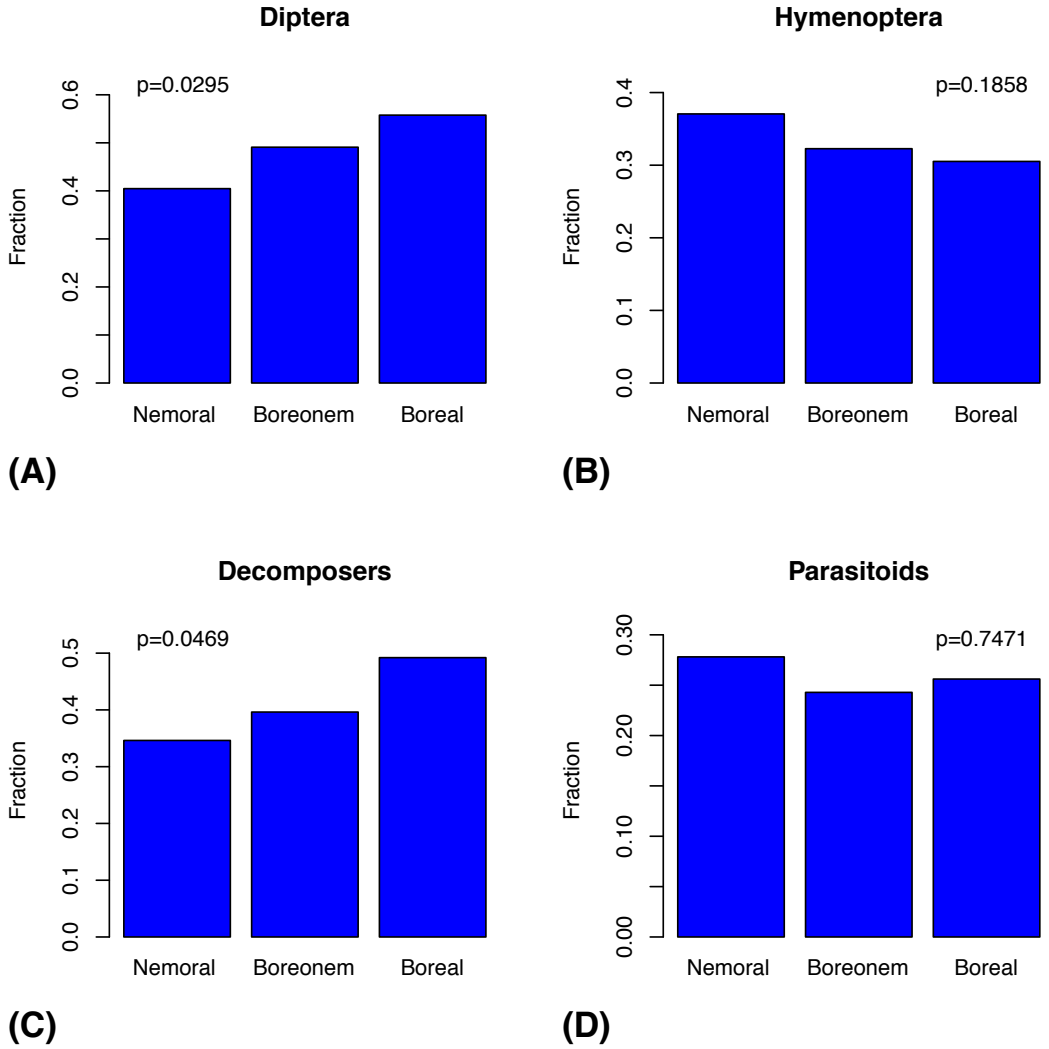
Extended Data Figure 5 | Changes in apparent composition of the Swedish insect fauna over time as it has become better known. Over time, the apparent fraction of Hymenoptera and Diptera species has increased dramatically. The same is true for the community of decomposers and parasitoids (red bars). Changes in apparent microhabitat composition have been less pronounced; most insects live on or in plants. However, the proportion of species found in other microhabitats has increased over time.



Extended Data Figure 6 | Trends in overall species richness of national European insect faunas, as documented by the Fauna Europaea database. The species richness of national faunas is determined to a large extent by the land area of the country, as shown by a log–log regression analysis (A). The residuals from this regression (B) reveal that there are three outliers — Belarus (BY), San Marino (SM) and Iceland (IS) — that all have unexpectedly small insect faunas given their land areas. This may be because the fauna is little known (or poorly recorded in Fauna Europaea), or because of other factors. After the area effect is removed, there is a significant latitudinal gradient in species richness (C) when all data are included, but the effect disappears when the three outliers are removed (D), or when only Iceland is removed (plot not shown; r^2=0.03, p=0.31).



Extended Data Figure 7 | Latitudinal trends in the composition of national European insect faunas, as documented by the Fauna Europaea database. The data suggest that there are significant latitudinal trends in the fraction of Diptera (A), Hymenoptera (B), decomposers (C) and parasitoids (D). However, some of these trends are more probably due to knowledge biases (see main text). Three outliers (see Extended Data Figure 6B) were removed from these analyses.



Extended Data Figure 8 | Compositional changes in the Swedish insect fauna over the nemoral to boreal transition, as indicated by the Malaise trap inventory data. There is a significant increase in the proportion of Diptera species and in the proportion of decomposers with increasing latitude, while there is no such trend for Hymenoptera or parasitoids. The *p* values were computed by comparing the observed values with a reference distribution generated using 10,000 random permutations of the trap assignments to biogeographic region. The arctic zone was excluded from this comparison because there is an order of magnitude less inventory data for this zone, and data are lacking for key groups. Sample sizes: nemoral zone 18 traps, 1,557 species; boreonemoral zone 29 traps, 2,693 species; and boreal zone 22 traps, 1,691 species.

Supplementary Material

Appendix S1. Supplementary information, with description of the Data Package, and explanations of Supplementary Tables.

Data Package. A compressed file (gzipped tarball) containing the data used in the paper, as well as the R scripts used in the analyses and to generate tables and plots.

Supplementary Tables S1–S6. Supplementary tables in Excel (xlsx) format. The data are also provided in the data package in several formats.