The diversity of arthropods in homes across the United States as determined by environmental DNA analyses

ANNE A. MADDEN,*†‡ ALBERT BARBERÁN,* MATTHEW A. BERTONE,§
HOLLY L. MENNINGER,¶ ROBERT R. DUNN†‡** and NOAH FIERER*††
*Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, USA, †Department of Applied Ecology, North Carolina State University, Raleigh, NC 27695, USA, ‡W.M. Keck Center for Behavioral Biology, North Carolina State University, Raleigh, NC 27695, USA, §Department of Entomology, North Carolina State University, Raleigh, NC 27695, USA, ¶Department of Biological Sciences, North Carolina State University, Raleigh, NC 27695, USA, **Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, Universitetsparken 15, Building 3, 2nd floor, Copenhagen DK-2100, Denmark, ††Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO 80309, USA

Abstract

We spend most of our lives inside homes, surrounded by arthropods that impact our property as pests and our health as disease vectors and producers of sensitizing allergens. Despite their relevance to human health and well-being, we know relatively little about the arthropods that exist in our homes and the factors structuring their diversity. As previous work has been limited in scale by the costs and time associated with collecting arthropods and the subsequent morphological identification, we used a DNA-based method for investigating the arthropod diversity in homes via high-throughput marker gene sequencing of home dust. Settled dust samples were collected by citizen scientists from both inside and outside more than 700 homes across the United States, yielding the first continental-scale estimates of arthropod diversity associated with our residences. We were able to document food webs and previously unknown geographic distributions of diverse arthropods – from allergen producers to invasive species and nuisance pests. Home characteristics, including the presence of basements, home occupants and surrounding land use, were more useful than climate parameters in predicting arthropod diversity in homes. These noninvasive, scalable tools and resultant findings not only provide the first continental-scale maps of household arthropod diversity, but our analyses also provide valuable baseline information on arthropod allergen exposures and the distributions of invasive pests inside homes.

Keywords: allergens, arthropod diversity, cytochrome *c* oxidase, DNA barcoding, insect diversity

Received 17 June 2016; revision received 16 September 2016; accepted 17 October 2016

Introduction

As soon as humans started building homes, they began to provide, inadvertently, shelter to other animals, including scavengers, but also parasites and associated pathogens. Arthropods likely represent the majority of these animal species (Bertone *et al.* 2016; Leong *et al.* 2016). The most conspicuous arthropods in houses tend to be those that are unwanted, either because they damage our homes (e.g. termites), and products (e.g. pantry pests), are viewed as unsightly (e.g. house centipedes), irritate our pets (e.g. fleas), or negatively affect our health. Some arthropods disperse pathogens and antibiotic-resistant microorganisms through homes and similar built environments as they move from faeces, sewage, carrion and waste to our food, beds and bodies (Graczyk *et al.* 2001; Olsen *et al.* 2001; Ahmad *et al.* 2011; Zurek & Ghosh 2014). Other arthropods, including dust mites and cockroaches, affect us more indirectly, by shedding sensitizing allergens that can lead to

Correspondence: Anne A. Madden and Noah Fierer,
Emails: madden.anne@gmail.com; Noah.Fierer
@colorado.edu

allergic rhinitis, allergic asthma and atopic eczema (Arlian 2002; Calderón et al. 2015), with the frequency and severity of these disorders having increased in the United States in recent decades (Bousquet et al. 2001; Moorman et al. 2007). Our general understanding of the types of arthropods we are exposed to inside our homes, and their geographic distributions, remains limited despite the potential impacts of these arthropods on our health and the overall quality of our living spaces.

Most studies investigating indoor arthropods have focused exclusively on a select few pest species that are most conspicuous, such as cockroaches and ants (Smith 1965; Schal & Hamilton 1990), or those species that are most damaging to building materials, such as termites (Lax & Osbrink 2003). Yet, recent in-depth studies of arthropods in houses in a single city, Raleigh, North Carolina, suggest that these relatively well-studied species represent a miniscule fraction of household arthropod diversity (Bertone et al. 2016; Leong et al. 2016). Bertone et al. (2016) found species from an average of 62 arthropod families per home, with many hundreds of morphospecies identified across the fifty homes surveyed. How these arthropods might be distributed within or among cities and geographic regions remains largely undetermined. As approximately 56% of United States' households use insecticides (Grube et al. 2011), arthropods are almost certainly common in many homes, but the full extent of this arthropod diversity across the United States remains unknown. One reason these knowledge gaps persist is that surveying arthropod diversity using standard methods is difficult. Collecting and accurately identifying the broad range of arthropod taxa that can be found in homes is costly, labour-intensive, highly invasive and logistically challenging. For example, just to survey arthropod diversity in 50 homes in North Carolina required six person hours per home to first collect samples, and then at least 2 h per home by trained entomologists to morphologically identify the arthropods via microscopy. Even with such exhaustive techniques, some species remain impossible to identify because of their life stage, sex or sample degradation.

Recently, DNA sequencing-based methods have been developed with the potential to circumvent some of the limitations of traditional morphology-based identifications and, in doing so, accelerate the process of surveying arthropod diversity from environmental samples. Typically, these DNA-based methods rely on PCR amplification and high-throughput sequencing of a fragment of the cytochrome c oxidase subunit 1 gene (Cox1/Co1/COI) using arthropod-specific primers that should amplify this gene region from nearly all arthropods (Hebert et al. 2003). The targeted gene region is advantageous as it is sufficiently variable to permit identification of most arthropods down to the taxonomic level of genus or species (Meusnier et al. 2008). Moreover, the targeted gene region is sufficiently short to permit analyses of arthropod diversity from environmental samples even if the arthropod DNA has been degraded. Such techniques have been used previously to quantify arthropod diversity in taxonomically unresolved communities (Burgar et al. 2014), animal diets (Bohmann et al. 2011), artificially mixed communities (Yu et al. 2012) and tree canopies (Gibson et al. 2014), but, to our knowledge, such molecular techniques have not previously been used to survey arthropod diversity inside homes.

We used such a DNA-based approach to rapidly assess arthropod diversity in settled home dust collected from inside and outside ~730 homes across the continental United States. Our goals were to characterize the diversity of arthropods found inside homes, determine which arthropods are repeatedly found inside homes vs. those which are transient, outdoor taxa and identify those factors that predict the continental-scale distributions of individual taxa (such as dust mites). Specifically, we investigated some of the building, environmental and occupant factors that predict what homes harbour a greater diversity of arthropods. We also assembled some of the first maps of arthropods detected in homes across the United States to describe how the distributions of specific taxa of medical, economic or ecological interest vary across regions.

Materials and methods

Sampling

Dust samples were collected as part of the Wild Life of Our Homes citizen science project, using methods described previously (Barberán et al. 2015a, b). Briefly, citizen scientists used sterile, dry swabs to sample the small ledge of the top of the door trim from the main living area inside homes and the door trim on the outside of the main entrance to the home. Swabs were then shipped to Boulder, CO, and stored at −20 °C until they were further processed. These samples were collected as part of a previous investigation of the biogeography of bacteria and fungi in the United States, where participants completed an online questionnaire about home occupants, design and habits wherein. All participants were provided a written informed consent form approved by NC State University's Human Subjects Research Committee (Approval No. 2177), as well as a sampling kit and instructions for sampling their home. Our study included 1462 of these settled dust samples collected from every state in the continental United

States, with 732 collected from inside homes and 730 from outside homes (Fig. S1, Supporting information). We identified characteristics of the sample geographic location and associated household information using occupant surveys and location information as described in more detail previously (Barberán *et al.* 2015b).

Molecular analyses

DNA was extracted from swabs using MoBio Power-Soil-htp kits, using the modified protocol described previously (Fierer *et al.* 2008). We amplified a ~157-bp region of the mitochondrial cytochrome *c* oxidase subunit I gene using arthropod-specific primers (Zeale *et al.* 2011; as used for high-throughput sequencing such as Bohmann *et al.* 2011; Table S1, Supporting information). Primers were modified to permit multiplex sequencing on the Illumina MiSeq platform as both the forward and reverse primers included the appropriate Illumina adapters and the reverse primers included a unique error-correcting 12-base pair barcode for each dust swab sample per sequencing run (Hamady *et al.* 2008). The 25 μL PCR cocktail included 3 μL gDNA, 12.5 μL Master Mix (Promega Biotech Co.) and 1 μM each of the forward and reverse primers. Cycling parameters were modified from Pinol *et al.* (2014) (94 °C for five min, followed by 45 cycles of 94 °C 30 s, 45 °C 45 s, 72 °C 45 s and a final extension at 72 °C for 10 min). Duplicate PCRs for each sample and the associated controls were pooled, cleaned and normalized using SequalPrep Normalization Plates (Thermo Fisher Scientific). Samples were pooled and sequenced on three Illumina MiSeq runs using 2 × 250 chemistry at the University of Colorado BioFrontiers Institute Next-Gen Sequencing Core Facility. To increase sequencing run quality with such an AT-rich amplicon, we increased the standard amount of PhiX used during the sequencing runs from 15% to 30% as suggested by the manufacturer.

Bioinformatic analyses

We generated approximately 43 million sequences. Sequences were demultiplexed using a custom python script 'prep_fastq_for_uparse.py' (https://github.com/leffj/helper-code-for-uparse), and the reverse reads were used for downstream analyses due to higher quality scores. Sequences were filtered, and OTU (operational taxonomic unit) picking was performed using the UPARSE pipeline (USEARCH version 7; Edgar 2013). Briefly, sequences were initially quality-filtered in UPARSE (Edgar 2013), sequences were trimmed to the expected amplicon length (158 bp) with those sequences removed that were of insufficient quality (max *e* value <1.5), below the minimum expected amplicon length (90 bp), or singletons (i.e. unique sequences that appeared only once across the entire data set). Sequences were clustered de novo at 99% similarity. Taxonomy assignments were performed in QIIME (Caporaso *et al.* 2010), using the hierarchical naïve Bayesian classifier from the Ribosomal Database Project at 99% similarity and with a confidence of 0.5 (Wang *et al.* 2007), retrained with a custom reference database curated from the BAR CODE OF LIFE Database (version 3; Ratnasingham & Hebert 2007). Due to the variable homology of the Co1 gene across taxa, we agree with others (Porter *et al.* 2014) that it is critical to use such a confidence-based hierarchical taxonomy classifier when assigning identity to these sequences. As expected, the differential variability in the sequenced Co1 gene region, as well as potential sequencing errors, led to multiple OTUs per genus for certain taxa, so all unique OTUs were collapsed into their respective genera for downstream analyses (see Fig. S2, Supporting information for more information on the sequence processing pipeline).

We validated this molecular and bioinformatics pipeline by processing and sequencing the targeted Co1 gene region from a collection of 58 arthropod taxa that were morphologically identified by a trained entomologist (Table S2, Supporting information). Based on this validation, identified genera in our sample set with singleton sequences were further removed and sequences were transformed into presence-only data. Three taxa were identified in a control sample of our sequencing run and were therefore conservatively removed from the analysis (these taxa made up less than 0.4% of that identified taxa). Sequences identified as originating from marine and aquatic arthropods (e.g. crabs, shrimp and crayfish) were removed from downstream analyses (except for mapping in Fig. S3, Supporting information), as they likely originated from pet or human food. Given that arthropods can vary dramatically in size and DNA can be shed through frass (faeces), sequence abundances are not directly proportional to biomass and all data were therefore analysed using presence/absence metrics alone (Elbrecht & Leese 2015; Barnes & Turner 2016).

Data analyses

All analyses were carried out in the R environment (R Core Team 2013; See Fig. S1, Supporting information for sample sizes used in different analyses.)

To determine how the results from our DNA-based survey compare to results obtained with surveys conducted using traditional sampling and morphology-based identification methods, we compared our findings with those of a recent study that investigated the arthropods of 50 homes in Raleigh, North Carolina

(Bertone *et al.* 2016). For these analyses, we focused just on those indoor dust samples collected from 87 North Carolina houses included in this study. The Bertone *et al.* (2016) survey involved exhaustive sampling of all rooms in each of the 50 houses within close proximity to the city of Raleigh, NC, where ours involved door trim swabs from a nonoverlapping set of 87 homes collected within the same state.

To investigate what factors might be driving differences in diversity, we ran generalized linear models (GLM) with Poisson errors. Variables were selected using forward and backward stepwise model selection by the Akaike information criterion (AIC) as implemented in the R function stepAIC (Venables & Ripley 2002). Three variables (urbanization, presence of basement and presence of pets) explained 9.3% of the variability of indoor arthropod diversity. To determine whether the greater diversity trend with pet ownership was consistent across broader taxonomic groups, we compared diversity across the factor of interest within the top five orders in homes (Lepidoptera, Diptera, Hemiptera, Coleoptera and Araneae) with false discovery rate (fdr) corrections. All variables [outdoor and indoor variables; see Barberán *et al.* (2015b) for the complete list of variables explored] explained 15.4% of the variability of indoor arthropod diversity.

To determine what arthropods were preferentially found in indoor vs. outdoor home samples, we determined what taxa had significant fidelity to each of these locations. This was calculated using Pearson's phi coefficient of association for each genus and each habitat using the multipatt function (De Cáceres *et al.* 2010) in the R package indicspecies (De Cáceres & Legendre 2009). We adjusted the *P* values for multiple comparisons using an fdr correction. For taxa significantly correlated with indoor or outdoor environments, we visualized the correlation strengths using a heat map.

For select taxonomic groups of interest, we mapped detection in indoor samples over the continental United States.

Results and discussion

Based on swab collections of settled dust and the arthropod DNA found therein, we noted the presence of more than 600 unique arthropod genera inside the 732 homes included in this study. Across all homes, we detected 28 orders of terrestrial arthropods. The taxa identified range from those that are well-studied inhabitants of the home (e.g. dust mites, cockroaches and carpet beetles) to taxa that are not commonly considered to be household inhabitants (e.g. aphids and certain parasitic wasps; Table S3, Supporting information). Most of these unexpected taxa, however, appear to be associated with homes as highlighted by the fact that many of these same unexpected taxa were also found in homes where arthropod diversity was surveyed using a more traditional approach (Bertone *et al.* 2016).

Most arthropod taxa were restricted in their distributions and only detected in a few homes (Fig. 1). Surprisingly, the genera that were most ubiquitous (i.e. detected in at least 10% of those homes with arthropods) included mosquitos (*Aedes* spp.) and plant-associated aphids (*Aphis* spp.), along with those taxa more commonly associated with the home environment such as dust mites (*Dermatophagoides* spp.), carpet beetles (*Anthrenus* spp.) and Indian meal moths (*Plodia* spp.). More broadly, taxa from the orders Lepidoptera (moths and butterflies), Hemiptera (true bugs) and Diptera (true flies) were found in a majority of indoor samples (Fig. 1). These three orders also represent the majority of the arthropod diversity detected across the homes (Fig. S4, Supporting information). Many of the taxa that were frequently detected inside homes include light-bodied, flying arthropod species (Figs 1 and 2). The abundance and diversity of such taxa may reflect the extent to which insects are attracted to and subsequently trapped inside, homes, either by home lighting (Somers-Yeates *et al.* 2013), elevated carbon dioxide inside homes (van Loon *et al.* 2015) or household organic waste (Karimifar *et al.* 2011).

Comparisons of DNA-based methods to more traditional survey methods

Despite the known limitations of DNA-based surveys – articulated well elsewhere (e.g. 14, 18) – many of the dominant taxa identified with our DNA-based approach were also identified using more traditional survey methods (Figs S5, S6, Supporting information). Our results compared favourably, despite sampling different homes within North Carolina, and using very different sampling approaches. The two methods captured many of the same families and orders of arthropods. Moreover, the most diverse orders identified were similar in the two studies (Fig. S7, Supporting information). Although many of the arthropod families identified from morphology-based collections were also found in the molecular survey, there were some exceptions (Table S4, Supporting information). We did not detect certain taxa such as ants (Formicidae), termites (Isoptera) or book lice (Liposcelididae) at the genus level (Table S4, Supporting information) likely due to known primer biases (Clarke *et al.* 2014; Elbrecht & Leese 2015), limitations of the reference databases or inherent limitations in sampling dust. Conversely, the DNA-based survey was better able to identify the presence of moths and butterflies (Lepidoptera) as we were not
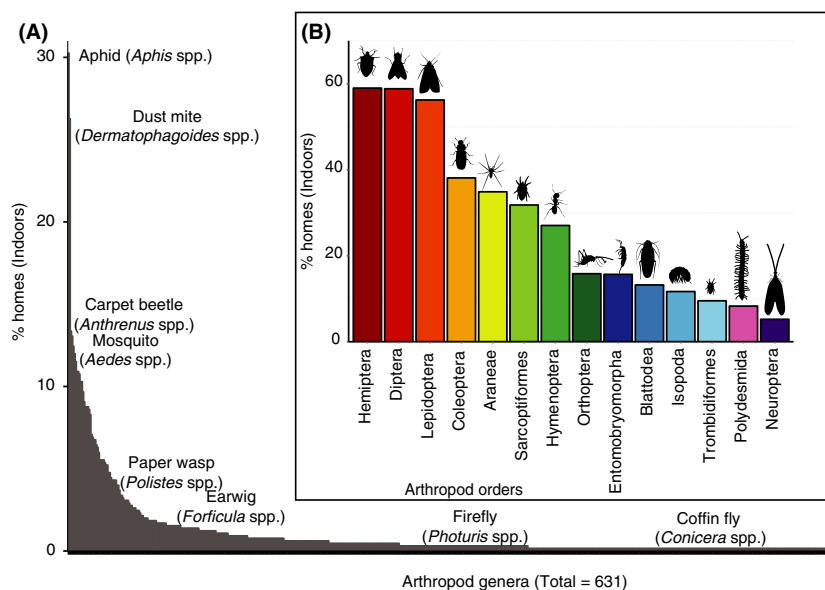
Fig. 1 Ubiquity of arthropods detected in settled dust samples collected from inside homes. (A) Genera detected. (B) Orders detected in at least 5% of homes. Y-axes indicate the percentage of homes (of 651 homes with arthropods detected) where those arthropods were detected.

limited by requiring intact specimens, identifiable parts of Lepidoptera (e.g. scales) or certain life stages (Fig. S7, Supporting information).

Our approach has some limitations when compared to traditional approaches. For one, it does not yield collections of specimens, or species-level resolution, necessary for additional studies of taxonomy or life history. Nor is it easy to link the proportional abundance of taxa in our samples to true abundances in homes. But our DNA-based survey approach also has a number of key advantages over more standard arthropod diversity assessment methods. Our sampling was not an invasive procedure requiring trained individuals to spend hours combing through a home to conduct intensive surveys. Rather, dust samples could be collected by the citizen scientist volunteers – requiring mere minutes per home. A swab of settled dust from door trim revealed arthropod diversity in the home. Furthermore, our DNA-based approach could identify arthropods from their faeces alone or remnant, miniscule parts in dust. Thus, despite the known limitations of using environmental DNA to survey arthropod diversity (see Gibson et al. (2014), and the fact that we were using a single dust sample to characterize arthropod diversity found throughout a home, our study allows us to describe arthropod diversity across an unprecedented number of homes located across the United States, thus yielding our first insight into the continental-scale diversity and distributions of arthropods inside homes.

Factors predicting arthropod diversity in homes

Arthropod diversity in dust varied among homes. We were able to detect greater than 40 arthropod genera in some individual homes with 72% of the homes having fewer than five detectable genera. We expected the diversity of arthropods found indoors to be positively associated with outdoor temperature (Dunn et al. 2009) and precipitation (Jenkins et al. 2011), as these factors are known to influence arthropod diversity outdoors. However, we found that outdoor climatic conditions were not reliable predictors of indoor arthropod diversity. Thus, given the relative homogeneity of environmental conditions inside homes and the presence of many arthropods living inside homes, it is difficult to use the large body of literature focused on outdoor arthropod diversity patterns to infer indoor arthropod diversity patterns.

We did find significant effects of surrounding land-use type on indoor arthropod diversity, with rural homes harbouring, on average, 50% more arthropod diversity than suburban or urban homes (Fig. 3; $P < 0.001$, Wilcoxon test). Likewise, the diversity of arthropods in homes was predictable from other home characteristics. Specifically, we found elevated arthropod diversity in homes with a basement (Fig. 3; $P < 0.05$, Wilcoxon test), and in homes with either dogs or cats (Fig. 3; $P < 0.001$, Wilcoxon test). The increased arthropod diversity found in homes with basements could be explained by the additional, cave-like habitat a basement offers, harbouring uniquely adapted taxa not found elsewhere in homes (Epps et al. 2014; Bertone et al. 2016). The presence of more diverse arthropod assemblages – across taxonomic orders – in homes with pets may result from the pets acting as dispersal agents – either dispersing arthropods directly associated with their bodies into homes (e.g. fleas and ticks) or by pets inadvertently bringing arthropods, or arthropod DNA,
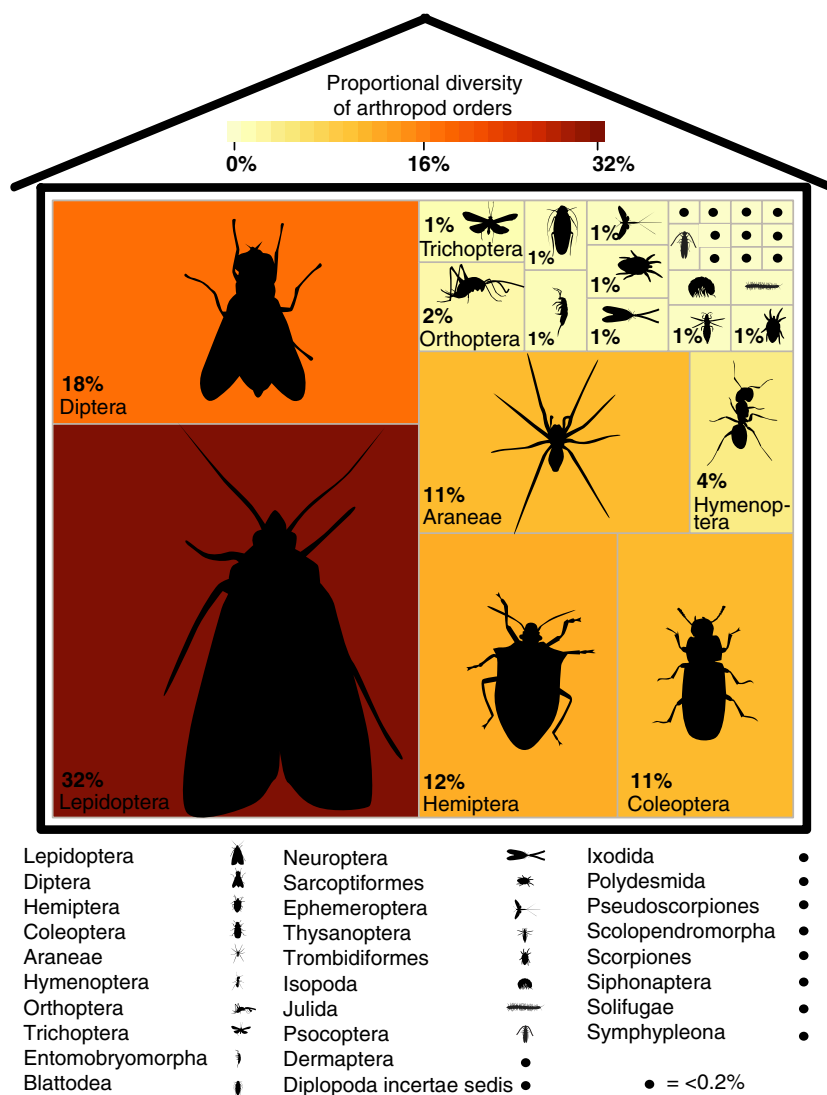
Fig. 2 Proportional arthropod diversity – number of unique taxa per order – detected across all indoor samples with each box representing a unique arthropod order and the size and colour of the box scaled to reflect the number of unique taxa within that order.

indoors from the outside (Fig. S8, Supporting information). These arthropod diversity patterns parallel those reported previously for microbial diversity; the presence of dogs and cats also increases the diversity of bacteria and fungi found inside homes (Barberán *et al.* 2015a). When you welcome a pet into your home, you are also introducing a diversity of other organisms – and their sensitizing allergens – an observation that may explain why children raised in homes with dogs or cats frequently have reduced susceptibilities to allergenic disorders (Ownby 2002).

Indoor vs. outdoor arthropod assemblages

The dust samples collected from door trim inside homes consistently harboured greater arthropod diversity than did outdoor samples collected from exterior door trim, an observation consistent with patterns observed for both home bacterial and fungal communities (Barberán *et al.* 2015a; Fig. 3; $P < 0.001$, Wilcoxon test, paired). For those 477 homes for which we had samples from both inside and outside the home (permitting direct comparisons), we found 2.5 times more arthropod genera, on average, inside the home than outside the home. This may be due to a greater level of DNA degradation in outdoor samples. In addition, outdoor dust samples have DNA from arthropods found outside the home, while those dust samples collected from inside homes have arthropod DNA both from those organisms that originated outdoors and those arthropods actually living, and originating from, the interior of the home itself.

We distinguished which of these arthropods are transitory visitors dispersed into the home from outdoors, and which are likely synanthropic (living out their life cycles with humans and originating from inside homes)
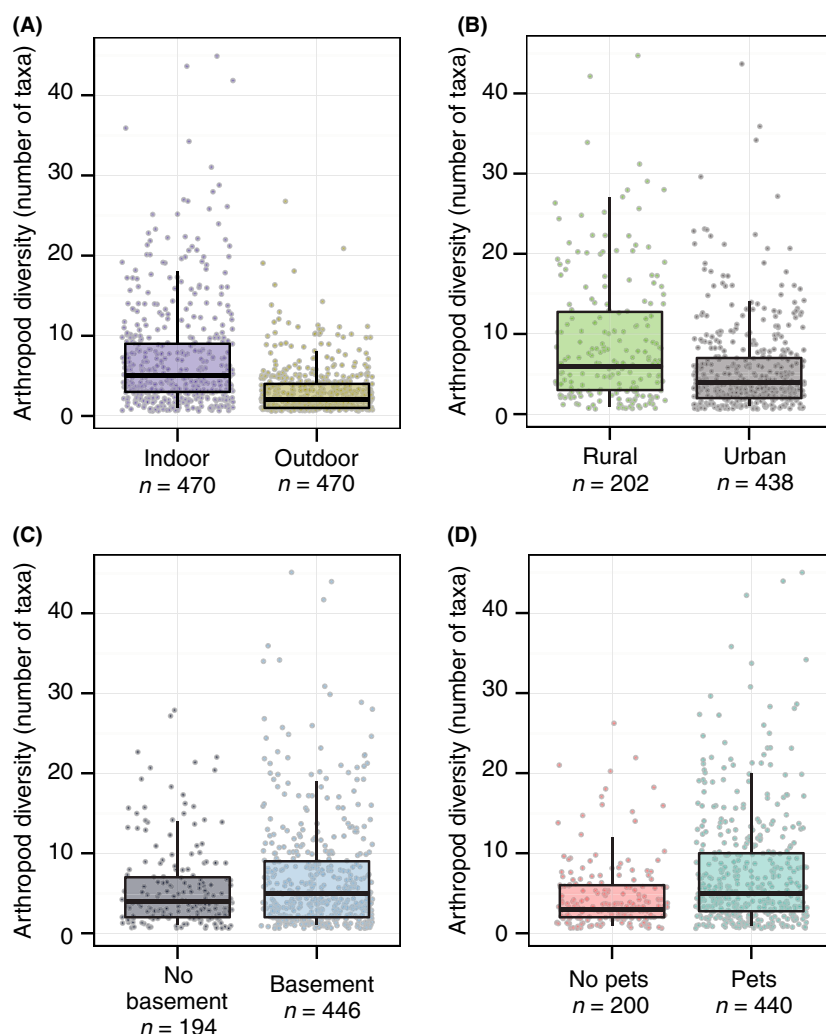
Fig. 3 Significant predictors of arthropod diversity inside home dust. (A) Indoor vs. outdoor samples (*P* < 0.001, Wilcoxon test, paired). (B) Indoor dust samples in relation to whether the home was in a rural vs. a suburban or urban area (*P* < 0.001, Wilcoxon test). (C) Indoor dust samples in homes with or without basements (*P* < 0.05, Wilcoxon test) or (D) with or without pets (*P* < 0.001, Wilcoxon test).

by comparing the presence of taxa found inside vs. outside homes. We identified 22 taxa (of 703 indoor and outdoor arthropod taxa detected) that were more likely to be identified from inside homes vs. outside (Fig. S9, Supporting information). These 'indoor' taxa span a wide range of arthropod classes from arachnids (mites) to Insecta (moths, flies, fleas, cockroaches, wasps, etc.), and Collembola (springtails). Taxa more likely to be found indoors could be more common because they are truly synanthropic, or because they preferentially accumulate inside homes. Dust mites (*Dermatophagoides* spp.), carpet beetles (*Anthrenus* spp.), Indian meal moths (*Plodia* spp.), fruit/vinegar flies (*Drosophila* spp.) and fleas (*Ctenocephalides* spp.) are common indoor inhabitants that we expected to find inside homes, as their diets are strongly associated with resources found in many homes (Bertone *et al.* 2016). While fleas feed on us and our pets, other taxa innocuously spend their lives consuming human skin cells or skin products (e.g. sebum), textiles, microbes and each other. We also found some taxa to be far more abundant indoors than outdoors which we did not expect to be living inside homes, including diverse herbivores: aphids (Aphididae), owlet moths (Noctuidae), planthoppers (Delphacidae) and parasitoid wasps (Rhopalosomatidae and Braconidae). We were more likely to find evidence of aphids in home dust than we were cockroaches, despite the latter receiving more attention as a pest and source of allergens. It may be that these diverse, often tiny, flying insects are lured into homes by unintentional attractants generated by humans in built spaces, or via passive collection (as in Malaise traps). Additionally, human occupants may be directly transporting some of these arthropods into the home as hitchhikers on ornamental plants.

Given the high frequency of aphids detected inside homes, we investigated this group in more detail to determine whether known aphid parasitoids and predators tended to co-occur with their prey. If they do co-occur, it would suggest that we can reconstruct known

biological interactions from the arthropod DNA left in the collected dust samples. We found that parasitoid wasps (Braconidae: Aphidiinae) tended to co-occur with aphids (Aphididae) more than expected by chance. In other words, when we find homes with aphid parasitoids, we are far more likely than expected by chance to also find aphid DNA in the same sample (*P* < 0.001, chi-square; Fig. S10, Supporting information). Likewise, when we looked for DNA from known predators of aphids (e.g. *Coccinella* spp., *Coleomegilla* spp., *Harmonia* spp., *Hippodamia* spp., *Hyperaspis* spp., *Scymnus* spp., and *Aphidoletes* spp.), they were also more likely to be found in homes with aphid DNA (*P* < 0.001, chi-square; Fig. S10, Supporting information).

The biogeography of key arthropod taxa

Some of the more frequently identified arthropods from indoor dust samples are considered to be of critical importance to human health, for example allergenic dust mites and cockroaches, or are perceived as nuisance pests of homes, for example pestiferous ladybugs, boxelder bugs and invasive cockroaches. For these taxa,

we used our analyses of the indoor dust samples to identify where these taxa are distributed throughout the United States (Fig. 4, see Fig. S11, Supporting information for additional maps).

No continental-scale maps exist for most of the 600 + taxa we identified in this study. However, because of the prevalence and significant economic costs associated with dust mite allergies, considerable effort has gone into building maps of predicted dust mite distributions (Portnoy *et al.* 2013). We found that dust mites were more likely to be detected in homes located in more humid regions of the United States (Fig. 4), patterns that overlap well with the predicted distributions of dust mites reported in Portnoy *et al.* (2013). Thus, even though we surveyed a small amount of settled dust from door trim, we were still able to reconstruct the expected geographic distributions of dust mites. This is particularly surprising, as dust mite allergens are not typically found on such hard surfaces (Portnoy *et al.* 2013). As many arthropod taxa shed potentially sensitizing allergens, maps of dust mites and other potentially allergenic arthropods – including cockroaches, aphids, ladybugs and even various shellfish
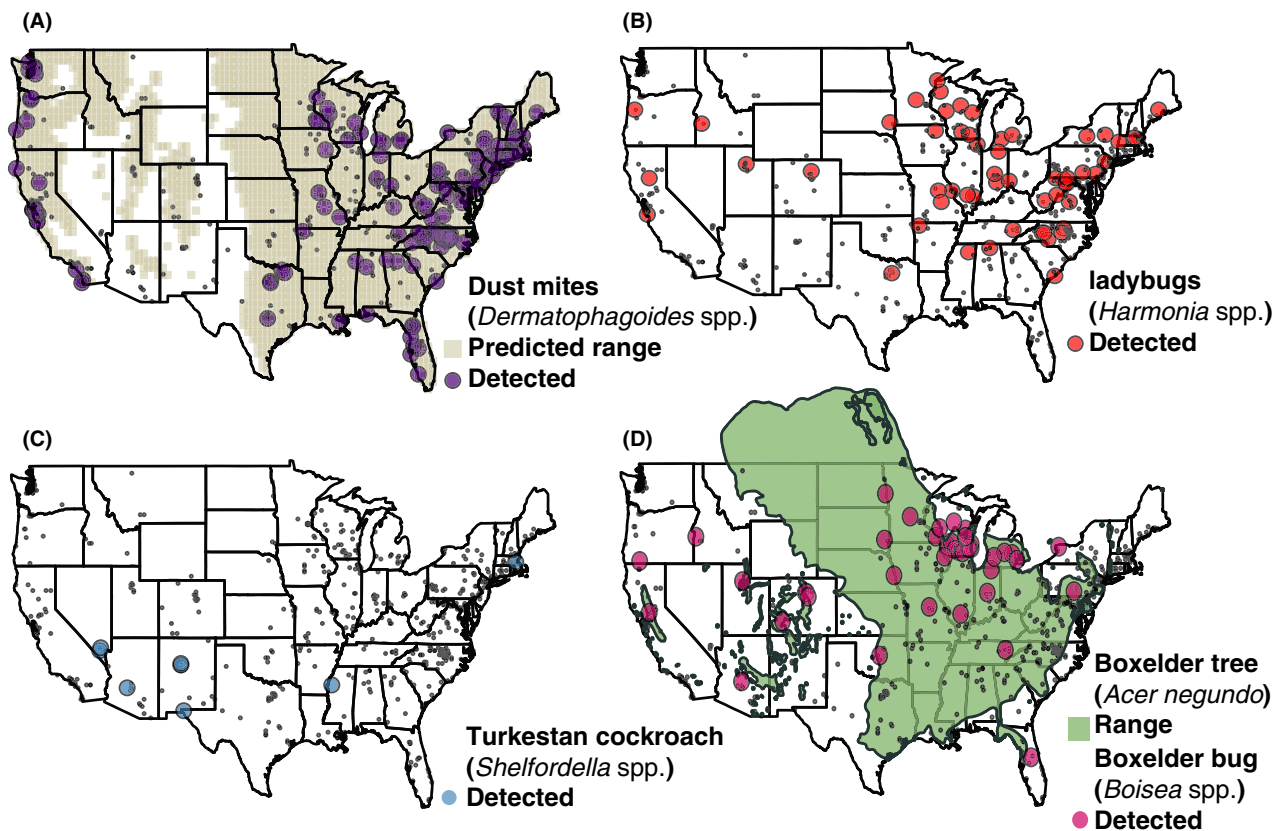

Fig. 4 Distributions of select allergenic, pestiferous and invasive arthropods in indoor dust. (A) Predicted allergenic dust mite (*Dermatophagoides* spp.) range vs. detected range, (B) alien ladybugs (*Harmonia* spp.), (C) invasive cockroach (*Shelfordella* spp.), (D) boxelder bugs (*Boisea* spp.) vs. the range of its primary host plant, boxelder trees (*Acer negundo* L.).

(Fig. 4, Figs. S3, S10, S11, Supporting information) – provide baseline information on the distributions of known, or likely, allergenic arthropods across the United States (Gaillard 1950; Esch *et al.* 2001; Arlian 2002). Such information could be integrated with data on antigen sensitivities to better understand allergenic arthropod exposures across the United States and their potential health ramifications.

Our geographic distribution maps also reflect the local outdoor ecology of other arthropod taxa. For example, the detected northern distribution of ladybugs (*Harmonia* spp.) in homes is consistent with this allergenic pests' cold intolerance, which leads them to seek shelter in homes during the winter (Goetz 2009; Fig. 4). Similarly, the nuisance pest, boxelder bugs (*Boisea* spp.), also overwinters in homes (Cranshaw 2011). As expected from other surveys of house arthropods, boxelder bugs were detected more often in homes that were generally at higher latitudes and were within the range of their main host plant boxelder trees (*Acer negundo* L.; U.S. Geological Survey 1999; Cranshaw 2011; Fig. 4).

Many of the taxa we found indoors are not native to the United States, imported as biocontrol agents (e.g. ladybugs like *Harmonia* spp.), or accidentally introduced, such as certain mosquito species (*Stegomyia* spp.). The value of our maps in visualizing the ranges of emerging invasive arthropod taxa is demonstrated in the distributions of the Turkestan cockroach (*Shelfordella* spp.; Fig. 4). This cockroach is an invasive species known to inhabit the south and south-western United States, but predicted to spread north and east due to its availability for Internet purchase as 'feeder' cockroaches (Kim & Rust 2013). We detected this species as far north as Massachusetts (Fig. 4) and in doing so reveal that this cockroach has already spread, undetected.

Conclusion

We show that noninvasive DNA-based surveys of settled dust can reveal many of the arthropods present inside homes. As a result, we were able to conduct the first continental-scale survey of arthropods in our homes. Indoor arthropod diversity is, to some degree, predictable from home design, degree of urbanization and the presence of pets, while the distributions of individual taxa often reflect the outdoor environment. These findings demonstrate the value of dust-based arthropod analyses for reconstructing arthropod food webs in homes and mapping the distributions of important arthropod taxa, including emerging and known arthropod allergens and invasive pest species. Moreover, the results presented here provide baseline information on the arthropods found inside our homes at the continental scale and demonstrate the utility of this DNA-based approach for future work focused on documenting how changes in climate, housing patterns or building design may alter the arthropods we come into contact with as we go about our daily lives.

Acknowledgements

References

Ahmad A, Ghosh A, Schal C, Zurek L (2011) Insects in confined swine operations carry a large antibiotic resistant and potentially virulent enterococcal community. *BMC Microbiology*, **11**, 23.

Arlian LG (2002) Arthropod allergens and human health. *Annual Review Entomology*, **47**, 395–433.

Barberán A, Dunn RR, Reich BJ *et al.* (2015a) The ecology of microscopic life in household dust. *Proceedings of the Royal Society B: Biological Sciences*, **282**, 20151139.

Barberán A, Ladau J, Leff JW *et al.* (2015b) Continental-scale distributions of dust-associated bacteria and fungi. *Proceedings of the National Academy of Sciences*, **112**, 5756–5761.

Barnes MA, Turner CR (2016) The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, **17**, 1–17.

Bertone MA, Leong M, Bayless KM *et al.* (2016) Arthropods of the great indoors: characterizing diversity inside urban and suburban homes. *PeerJ*, **4**, e1582.

Bohmann K, Monadjem A, Lehmkuhl Noer C *et al.* (2011) Molecular diet analysis of two African free-tailed bats (Molossidae) using high throughput sequencing (C Lalueza-Fox, Ed.). *PLoS ONE*, **6**, e21441.

Bousquet J, van Cauwenberge P, Khaltaev N (2001) Allergic rhinitis and its impact on asthma. *Journal of Allergy and Clinical Immunology*, **108**, S147–S334.

Burgar JM, Murray DC, Craig MD *et al.* (2014) Who's for dinner? High-throughput sequencing reveals bat dietary differentiation in a biodiversity hotspot where prey taxonomy is largely undescribed. *Molecular Ecology*, **23**, 3605–3617.

Calderón MA, Linneberg A, Kleine-Tebbe J *et al.* (2015) Respiratory allergy caused by house dust mites: what do we really know? *Journal of Allergy and Clinical Immunology*, **136**, 38–48.

Caporaso JG, Kuczynski J, Stombaugh J *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Clarke LJ, Soubrier J, Weyrich LS, Cooper A (2014) Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, **14**, 1160–1170.

Cranshaw W (2011) A review of nuisance invader household pests of the United States. *American Entomologist*, **57**, 165–169.

De Cáceres M, Legendre P (2009) Associations between species and groups of sites: indices and statistical inference. *Ecology*, **90**, 3566–3574.

De Cáceres M, Legendre P, Moretti M (2010) Improving indicator species analysis by combining groups of sites. *Oikos*, **119**, 1674–1684.

Dunn RR, Agosti D, Andersen AN *et al.* (2009) Climatic drivers of hemispheric asymmetry in global patterns of ant species richness. *Ecology Letters*, **12**, 324–333.

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, **10**, 996–998.

Elbrecht V, Leese F (2015) Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol (M Hajibabaei, Ed.). *PLoS ONE*, **10**, e0130324.

Epps MJ, Menninger HL, LaSala N, Dunn RR (2014) Too big to be noticed: cryptic invasion of Asian camel crickets in North American houses. *PeerJ*, **2**, e523.

Esch RE, Hartsell CJ, Crenshaw R, Jacobson RS (2001) Common allergenic pollens, fungi, animals, and arthropods. *Clinical Reviews in Allergy and Immunology*, **21**, 261–292.

Fierer N, Hamady M, Lauber CL, Knight R (2008) The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proceedings of the National Academy of Sciences*, **105**, 17994–17999.

Gaillard GE (1950) The aphid—an insect allergen. *Journal of Allergy*, **21**, 386–399.

Gibson J, Shokralla S, Porter TM *et al.* (2014) Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences*, **111**, 8007–8012.

Goetz DW (2009) Seasonal inhalant insect allergy: *Harmonia axyridis* ladybug. *Current Opinion in Allergy and Clinical Immunology*, **9**, 329–333.

Graczyk TK, Knight R, Gilman RH, Cranfield MR (2001) The role of non-biting flies in the epidemiology of human infectious diseases. *Microbes and Infection*, **3**, 231–235.

Grube A, Donaldson D, Kiely T, Wu L (2011) *Pesticides Industry Sales and Usage: 2006 and 2007 Market Estimates*. U.S. Environmental Protection Agency, Washington, D.C.

Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature Methods*, **5**, 235–237.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 313–321.

Jenkins CN, Sanders NJ, Andersen AN *et al.* (2011) Global diversity in light of climate change: the case of ants: global diversity of ants in light of climate change. *Diversity and Distributions*, **17**, 652–662.

Karimifar N, Gries R, Khaskin G, Gries G (2011) General food semiochemicals attract omnivorous German Cockroaches, *Blattella germanica*. *Journal of Agricultural and Food Chemistry*, **59**, 1330–1337.

Kim T, Rust MK (2013) Life history and biology of the invasive Turkestan cockroach (Dictyoptera: Blattidae). *Journal of Economic Entomology*, **106**, 2428–2432.

Lax AR, Osbrink WL (2003) United States Department of Agriculture-Agriculture Research Service research on targeted management of the Formosan subterranean termite *Coptotermes formosanus* Shiraki (Isoptera: Rhinotermitidae). *Pest Management Science*, **59**, 788–800.

Leong M, Bertone MA, Bayless KM, Dunn RR, Trautwein MD (2016) Exoskeletons and economics: indoor arthropod diversity increases in affluent neighbourhoods. *Biology Letters*, **12**, 20160322.

van Loon JJA, Smallegange RC, Bukovinszkiné-Kiss G *et al.* (2015) Mosquito attraction: crucial role of carbon dioxide in formulation of a five-component blend of human-derived volatiles. *Journal of Chemical Ecology*, **41**, 567–573.

Meusnier I, Singer GA, Landry J-F *et al.* (2008) A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, **9**, 214.

Moorman JE, Rudd RA, Johnson CA *et al.* (2007) National surveillance for asthma—United States, 1980–2004. *MMWR Surveillance Summaries*, **56**(1–14), 18–54.

Olsen AR, Gecan JS, Ziobro GC, Bryce JR (2001) Regulatory action criteria for filth and other extraneous materials V. Strategy for evaluating hazardous and nonhazardous filth. *Regulatory Toxicology and Pharmacology*, **33**, 363–392.

Ownby DR (2002) Exposure to dogs and cats in the first year of life and risk of allergic sensitization at 6 to 7 years of age. *JAMA*, **288**, 963–972.

Pinol J, San Andres V, Clare EL, Mir G, Symondson WOC (2014) A pragmatic approach to the analysis of diets of generalist predators: the use of next-generation sequencing with no blocking primers. *Molecular Ecology Resources*, **14**, 18–26.

Porter TM, Gibson JF, Shokralla S *et al.* (2014) Rapid and accurate taxonomic classification of insect (class Insecta) cytochrome *c* oxidase subunit 1 (COI) DNA barcode sequences using a naïve Bayesian classifier. *Molecular Ecology Resources*, **14**, 929–942.

Portnoy J, Miller JD, Williams PB *et al.* (2013) Environmental assessment and exposure control of dust mites: a practice parameter. *Annals of Allergy, Asthma & Immunology*, **111**, 465–507.

R Core Team (2013) *R: A Language and Environment for Statistical Computing*.

Ratnasingham S, Hebert PDN (2007) BARCODING: bold: the barcode of life data system (http://www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.

Schal C, Hamilton RL (1990) Integrated suppression of synanthropic cockroaches. *Annual Review of Entomology*, **35**, 521–551.

Smith MR (1965) *House-Infesting Ants of the Eastern United States. Their Recognition, Biology and Economic Importance*. United State Department of Agriculture, Washington, D.C. Technical Bulletin No. 1326.

Somers-Yeates R, Hodgson D, McGregor PK, Spalding A, ffrench-Constant RH (2013) Shedding light on moths: shorter wavelengths attract noctuids more than geometrids. *Biology Letters*, **9**, 20130376–20130376.

U.S. Geological Survey (1999) *Digital representation of "Atlas of United States Trees" by Elbert L. Little, Jr.*

Venables WN, Ripley BD (2002) *Modern Applied Statistics with S*. Springer, New York.

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, **73**, 5261–5267.

Yu DW, Ji Y, Emerson BC *et al.* (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring: biodiversity soup. *Methods in Ecology and Evolution*, **3**, 613–623.

Zeale MRK, Butlin RK, Barker GLA, Lees DC, Jones G (2011) Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces: DNA Barcoding. *Molecular Ecology Resources*, **11**, 236–244.

Zurek L, Ghosh A (2014) Insects represent a link between food animal farms and the urban environment for antibiotic resistance traits. *Applied and Environmental Microbiology*, **80**, 3562–3567.

Data accessibility

Data are publicly available in the Figshare digital repository doi: 10.6084/m9.figshare.3 824 448.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Sampling locations.

Fig. S2 Abbreviated flow chart of sequence processing pipeline.

Fig. S3 Aquatic arthropod distributions in indoor dust, including those arthropods responsible for shellfish allergies.

Fig. S4 Arthropod families detected in greater than 5% of indoor samples.

Fig. S5 Rank comparisons of the 20 most frequently detected families in North Carolina homes in a morphology-based survey (Bertone *et al.* 2016) versus those detected in a DNA-based study (this study).

Fig. S6 Most frequently detected families in comparative arthropod surveys.

Fig. S7 Comparison of arthropod families per order detected in a morphology-based survey of North Carolina homes ($n = 50$) (Bertone *et al.* 2016), versus our DNA-based study of the indoor trim of North Carolina homes ($n = 87$).

Fig. S8 Arthropod diversity for the top five orders in houses with, and without, pets.

Fig. S9 Arthropod taxa significantly more likely to be found in dust samples collected from inside homes than outside homes.

Fig. S10 Co-occurrence maps of taxa showing potential aphid-based food webs.

Fig. S11 Distribution of arthropods in indoor samples.

Table S1 Amplification and sequencing primers used.

Table S2 Arthropod taxa used to validate primers and bioinformatics pipeline.

Table S3 List of arthropod taxa detected in indoor dust trim.

Table S4 Total arthropod families detected across a morphology-based survey of arthropods in North Carolina homes (Bertone *et al.* 2016), and our DNA-based survey of indoor home samples in North Carolina.