The rise of disease ecology

Ellen E. Brandell[1], Daniel J. Becker[2,3], Laura Sampson[1], Kristian M. Forbes[4]

1. Center for Infectious Disease Dynamics & Department of Biology, Huck Institute of the Life Sciences, Pennsylvania State University, University Park, PA 16802
2. Department of Biology, Indiana University, Bloomington, IN 47405
3. Center for the Ecology of Infectious Disease, University of Georgia, Athens GA 30602
4. Department of Biological Sciences, University of Arkansas, Fayetteville AR 72701

***Corresponding author:**
Ellen E. Brandell
2327 University Way #2
Bozeman, MT 59715
Phone: 248-568-2151
Email: ebrandell08@gmail.com

Keywords: host-pathogen interactions, infectious disease ecology, machine learning, questionnaire, research trends

ABSTRACT

Disease ecology is an interdisciplinary field that has recently rapidly grown in size and influence. We described the composition and educational experiences of disease ecology practitioners and identified changes in research foci. We combined a global survey with a literature synthesis involving machine-learning topic detection. Disease ecology practitioners have diversified in the last decade in terms of gender identity and institution, with weaker diversification in terms of race and ethnicity. Topic detection analysis of over 18,500 research articles revealed research foci that have declined (e.g., HIV), increased (e.g., infectious disease in bats), and have remained common (e.g., malaria ecology, influenza). The steady increase in topics such as climate change, and emerging infectious diseases, superspreaders indicate that disease ecology as a field of research will continue advancing our understanding of complex host-pathogen interactions and forms a critical and adaptable component of the global response to emergent health and environmental threats.

INTRODUCTION

Infectious diseases are one of the leading sources of human, domestic animal, and wildlife mortality, killing an estimated 17 million people each year (Brand 2013; World Health Organization 1996, 2018), threatening economic security through crop and production animal losses (Benavides et al. 2017; Ahmed et al. 2019), and causing declines of endangered species (Scheele et al. 2019). Moreover, infectious disease outbreaks are predicted to be exacerbated by contemporary issues such as climate change, high human population density, and fragmentation of natural environments (Daszak et al. 2001; Patz et al. 2004; Altizer et al. 2013). It is vital to establish a strong foundation and specialization for research on infectious diseases in their ecological and evolutionary context in order to maintain our standard of living and enhance wildlife and ecosystem health; for example, we continue to face challenges such as

emerging pathogens (e.g. SARS-CoV-2; Andersen et al. 2020), pathogen evolution (Van Boeckel et al. 2019), and the need for innovative interventions (Sokolow et al. 2019).

Disease ecology is the study of how pathogens and parasites move through host species, populations, and landscapes, considering both abiotic and biotic factors, as well as the consequences of their infections. It is a relatively new and rapidly expanding research focus within ecology and evolutionary biology that draws heavily on early foundations in population biology (Anderson & May 1979; May & Anderson 1979), and integrates many biological science fields including parasitology, epidemiology, microbiology, and immunology (Grenfell et al. 1995; Hudson et al. 2002; Wilson et al. 2019). Disease ecologists investigate a range of practical and fundamental questions relevant to humans, other animals and plants, ranging from the natural origins of disease outbreaks; to heterogeneities in pathogen susceptibility, transmission, and impact; and the effectiveness of intervention strategies (Hudson et al. 1998; Condeso & Meentemeyer 2007; Joseph et al. 2013; VanderWaal & Ezenwa 2016; Olival et al. 2017).

Disease ecology, in part, adapted and developed population biology theory to address societal needs (Scheiner & Rosenthal 2004; Johnson et al. 2015; Koprivnikar & Johnson 2016). Key among these is the urgency to understand and address novel disease threats, which are rooted in natural systems but are often exacerbated by societal inequalities (Carlson & Mendenhall 2019). For example, the impacts of habitat degradation on pathogen spillover is an expanding area of research that can be used to guide risk assessments and environmental policy (Patz et al. 2004). At the same time, infrastructure has developed around disease ecology, such as a specialized National Science Foundation and National Institutes of Health funding program and conference series (Scheiner & Rosenthal 2004), which have helped to direct research effort and create networks amongst researchers.

Still, many questions remain as to the composition of disease ecology practitioners, core research foci, and how research trends have changed to meet societal needs. Answering these questions will help to improve training pathways and prioritize research emphases for future research. However, understanding these complex and interrelated factors as they apply to an interdisciplinary research topic requires diverse and innovative approaches.

Here we characterize both the practitioners and field of disease ecology by addressing the following questions:

(1) Who comprises the field in terms of education, demographics, and research foci?
(2) Which are the most influential scientific articles and journals?
(3) And significantly, how have research trends emerged and changed over time? For example, do they follow global health priorities such as disease outbreaks?

To answer these questions, we surveyed self-declared disease ecologists globally and conducted a literature synthesis with machine-learning topic detection (Loper & Bird 2002; Bird et al. 2009; Blei 2012). Systematic and quantitative approaches to literature syntheses are increasingly favored over narrative-based reviews (Lajeunesse 2010; Hedges & Olkin 2014; Haddaway & Watson 2016). However, increasingly high volumes of published research across scientific fields raise difficulties with conducting such systematic syntheses and require novel approaches (Lajeunesse 2016; Nunez-Mir et al. 2016). Following recent adoptions of data mining approaches to systematic reviews (Han & Ostfeld 2019), we apply topic detection using non-negative matrix factorization to characterize the research core and trajectory of disease ecology. More broadly, our approach can provide a quantitative synthesis framework to examine the composition and trends in other fields that cross traditional disciplinary boundaries.

MATERIALS AND METHODS

Survey

We developed a survey instrument (Pennsylvania State University Institutional Review Board Study 00010582; Supporting Information) to quantify the demographics and research core of disease ecology. The survey was disseminated on disease ecology email listservs such as conference attendees (e.g., the past five years of Ecology and Evolution of Infectious Disease conferences, Ecological Society of America), specific research groups (e.g., VectorBiTE), scientific organizations (e.g., American Society of Parasitologists, Ecological Society of America disease ecology section), and institutional research centers (e.g., Pennsylvania State University Center for Infectious Disease Dynamics, University of Georgia Center for the Ecology of Infectious Diseases). We also distributed the survey to prominent non-USA research groups (in e.g. South America, Europe, Australia) to diversify our survey respondents and therefore better represent the true population of disease ecologists. Survey deployment was necessarily biased towards established research centers and active researchers in the disciplinary community, and likely missed some individual scientists and institutional groups.

The survey was open from November 2018 until January 2019, closing once the response rate dropped below two new responses per day for one consecutive week. All survey respondents were self-declared disease ecology practitioners, who were informed about the potential use of results in a consent statement. The survey asked respondents questions on their demographics, institution, education, types and topics of current research, opinion on the importance of different topics in the field, and influential scientific articles and journals. It included a combination of multiple choice and short answer response questions. A full copy of the survey and description of the data cleaning procedure is available in the Supporting Information.

Literature search

Our objective was to compile an extensive corpus robustly representative of publications in disease ecology, not to include every article *per se*. To reduce bias, both search terms and included journals were based on survey results. We included journals that were listed by at least four survey respondents as significant to the field (n = 42), as well as *Nature* and *Science*. To generate the original list of search terms, we compiled a set of papers that cited the foundational paper in disease ecology (Anderson & May 1979), and then used topic detection algorithms to generate the key word list that was used to inform the final literature search criteria. Topic detection was performed using the *nltk* library in Python 2.7 (Bird et al. 2009).

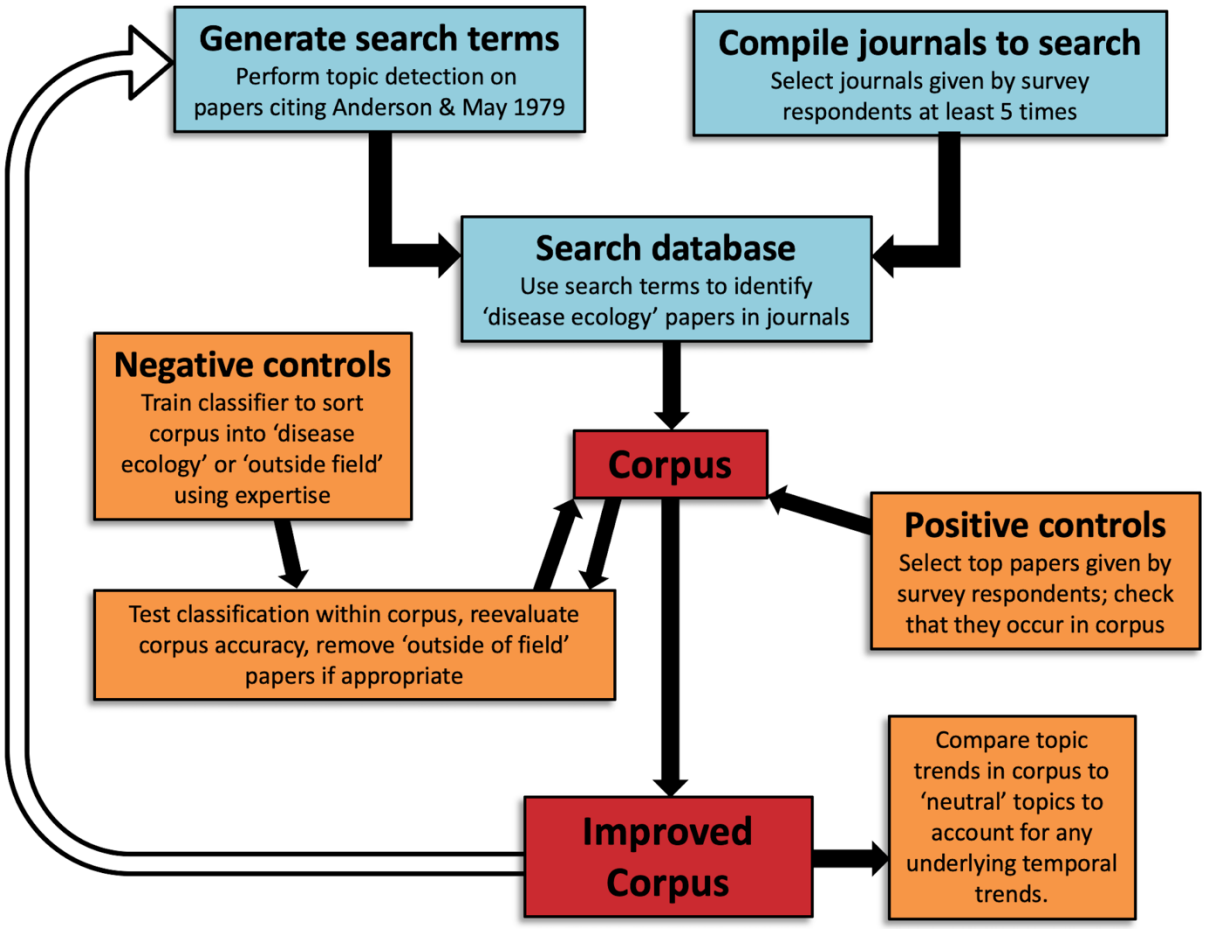

Figure 1. Workflow of systematic literature search and corpus development. Box coloration denotes different stages of corpus development: literature compilation (blue), corpus

assessment and validation (orange), and corpus completion (red). The unfilled arrow denotes repeating of the workflow to optimize corpus accuracy.

The literature search was conducted in Web of Science for the years 1975 to 2018. Each article had to meet specific criteria using Boolean filters, including a focus on studying a pathogen or parasite, host infections (to distinguish from solely environmental persistence of microorganisms), and individual-level or higher-order dynamics (e.g. not cellular processes, with the exception of those analyzed as a population-level process). The full list of search terms is provided in the Supporting Information, alongside a set of exclusionary terms to remove similar but non-disease ecology articles. Web of Science categories were used to narrow our search and also reduce false-positive inclusions. Finally, articles with fewer than four citations were removed as a form of quality control.

To evaluate false-positives, two authors (DJB and KMF) independently evaluated the same 100 randomly selected articles and classified them as ‘disease ecology’ or ‘outside the field’. Papers that fell outside the field predominantly described bacterial communities, within-host behavior or adaptations, or genetics/genomics. Over 75% of the articles in the final corpus were classified as disease ecology, and consensus was strong among evaluators (94% agreement, Cohen’s $\kappa=0.84$). Within-host studies were accepted if they focused on population-level processes (e.g., Cressler et al. 2014).

To evaluate false-negatives, we cross-validated our corpus using our survey data. Specifically, we assessed whether articles that were identified by at least two respondents as influential were present in our corpus. We calculated the proportion of papers that were included in our corpus out of the list of such articles, with the requirement that at least 70% of papers had to be

included. Of the influential articles identified by survey respondents (written ≥2 times) restricted to journals used in building the corpus, approximately 71% (50/70) were present in the corpus. Yet the 'most influential' articles had a higher probability of being included: the corpus included 85% of articles written four or more times, 75% of articles written three times, and 63% of articles written twice. We adjusted the search and exclusion terms twice using the workflow described in Figure 1 (unfilled arrow) to obtain a corpus with high classification and cross-validation success.

Literature analysis

We conducted topic detection on the validated corpus using non-negative matrix factorization. Topic clusters represent a set of co-occurring words that can be used to define an area of research. The number of topic clusters (*i*) and words per topic (*j*) were the only parameters imposed on the literature analysis. To select appropriate values for *i* and *j*, we ran topic detection for a range of values and combinations of *i* and *j* and assessed outputs. If *i* was too small or too large, we were unable to detect temporal variation in that topic. If *j* was too small or too large, the topics were not clearly defined. For example, a topic with only five words may not be interpretable; similarly, a topic with 30 words may be too broad to assign meaning. We used $i = 15$ and $j = 15$, so our corpus was analyzed for 15 topics with 15 words each.

We used K-means clustering from the *nltk* Python library to construct topics, where each topic comprised 15 commonly co-occuring words. We assigned a name to each topic to describe its theme. For example, we named a topic containing immunodeficiency, HIV, patient, therapy, drug, AIDS, background, treatment, and risk, as an *HIV topic*. We gave each topic name a 'confidence' measurement of 1–3, from high to low confidence in identifying the topic. In addition to topics that emerged from the literature, we also generated and assessed our own topic lists

based on key research areas, such as climate change, dilution effect, superspreaders, network analysis, EIDs, bovine tuberculosis, infectious diseases in bats and rodents, and chytrid fungus (Fig. 4). To ensure topic trends were not confounded by an increase in the total number of published articles through time, we constructed a baseline topic using neutral words that should be in all disease ecology articles: analysis, study, and paper. We evaluated temporal trends in publications for each theme using generalized additive models (GAMs) fit using the *mgcv* package in R (Wood 2006). The proportion of words in each topic relative to all words was modeled as a binomial response using thin plate splines with shrinkage for publication year. Lastly, to assess covariation among topics, we estimated Spearman's rank correlation coefficients (ρ) at the zero-year lag.

RESULTS

Survey

A total of 413 self-declared disease ecologists participated in the survey. The average respondent was 36.1 years old (21-76, median: 34; n=348). 76.7% of participants (n=408) considered at least half of their research to fall within disease ecology. Participants that considered ≥75% of their research to be disease ecology were concentrated from ages 25–40, and most self-identified as women (60%) (n=344). More broadly, 56.1% of participants identified as women (n=231), 42.5% as men (n=175), 0.7% as other (n=3), and 0.7% preferred not to say (n=3). We report on participants that chose to disclose a gender identity for results regarding gender.

Most respondents identifying as women were younger (age ≤25–35) than most respondents identifying as men (age 26–50). The youngest age category (≤25 years) was 68.9% women (n=45), and the oldest age category (60+ years) was 85.0% men (n=20). Current positions held by survey participants were: undergraduate student (1.2%, n=5), Master's student (2.9%; n=12),

192 PhD student (24.5%; n=100), post-doctoral researcher (21.1%; n=86), faculty (39.5%; n=161),
193 researcher (9.1%; n=37), and other (1.7%; n=7). Respondents identifying as women comprised
194 most of each academic position except Master's student and faculty (Table S1). In general,
195 most PhD students and post-doctoral researchers were young and identified as women. Most
196 Masters' students were young and identified as men, and most faculty were middle-aged and
197 identified as men (Tables S1–S3, Fig. S2C). Participants that did not identify with a strict gender
198 binary were distributed across age (≤25–50) and position categories.
199
200 We identified clear trends in education and demographics of survey respondents. 92.7% of
201 participants had completed their undergraduate degree by 2018 (n=382), 50.0% had completed
202 their Master's degree by 2018 (n=206), and 73.3% had completed their PhD by 2018 (n=302).
203 8.3% of participants had also earned an additional degree, most commonly as a Doctor of
204 Veterinary Medicine (n=18). Nearly half of respondents had a PhD but not a Master's degree
205 (45.4%; n=187). The total number of graduate degrees earned per year among disease ecology
206 researchers has dramatically increased since the year 2000 (Fig. 2A).
207
208 Most participants identifying as non-white were less than 36 years old, indicating a recent,
209 though minor, diversification of disease ecology practitioners. This was especially pronounced
210 when analyzed by education: the average proportion of participants identifying as non-white
211 who have earned a Master's degree has nearly doubled from pre-2000 to post-2000 (10.2% to
212 19.1%), and had risen to 24.3% in the last decade (Fig. 2D). The proportion of participants
213 identifying as non-white who have earned a PhD increased from pre-2000 to post-2000 but has
214 since remained stable (16-19%; Fig. 2D).
215
216 The proportion of Master's and PhD degrees earned in low- and middle-income countries has
217 also recently increased (Fig. 2B). The proportion of participants identifying as women who have

earned Master's degrees has substantially increased from a mean of 24.5% pre-2000 to 53.8% post-2000. From 2000-2018, this proportion was approximately stable at around 55%. Similarly, the proportion of participants identifying as women who have earned PhDs substantially increased from a mean of 30.5% pre-2000, to 52.4% post-2000; this proportion has continued to increase to 64.2% since 2010 (Fig. 2C).

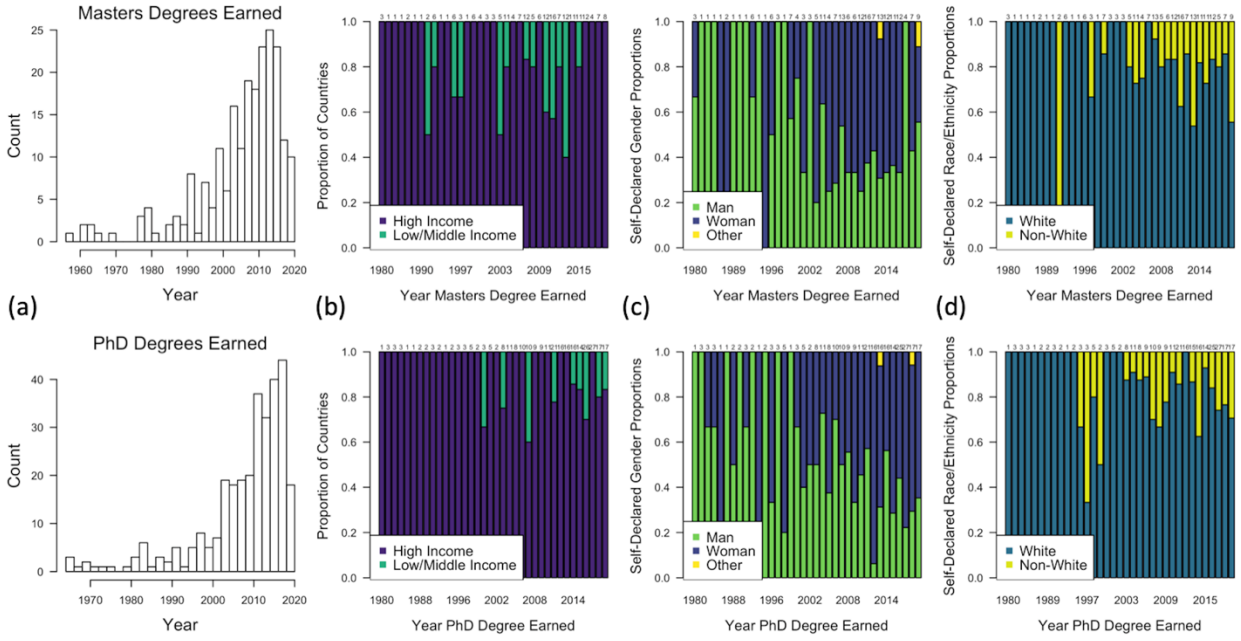

Figure 2. Plots showing demographic trends in survey participants that indicated they earned a Masters (top row) or PhD (bottom row) degree by 2018**. (A)** Displays the number of degrees earned by year, and the stacked bar plots show the proportion of degree earners by **(B)** high or low/middle income countries, **(C)** self-declared gender identity, and **(D)** self-declared race/ethnicity. Sample sizes are displayed above their respective bars and years included 1980-2019; some years may be excluded if no degree earners indicated the information of interest.

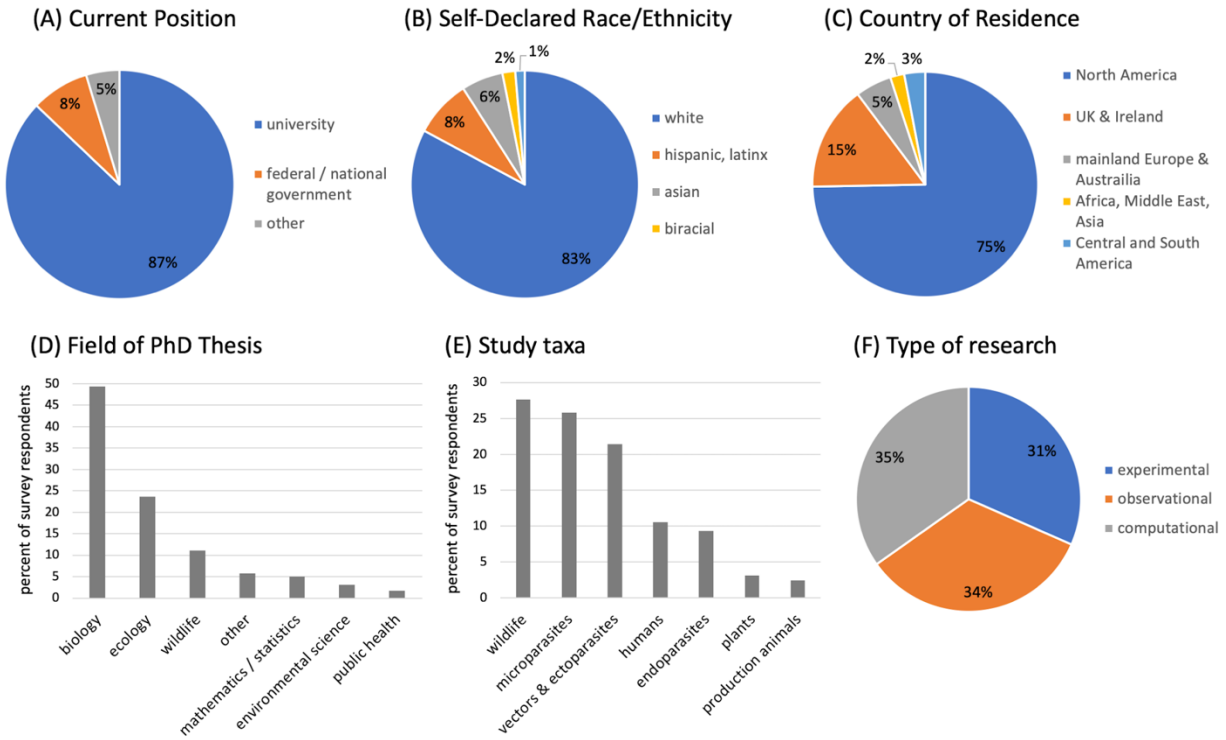

Figure 3. Summary of disease ecology demographics and research topics/types from survey participants. Pie charts display **(A)** current position or institution type, **(B)** self-declared race or ethnicity, and **(C)** country of residence; research topics and types are categorized by **(D)** field of PhD thesis, **(E)** current study taxa, and **(F)** type of primary research.

The most popular areas of research were ranked as: epidemiology, mathematical modeling, population ecology, wildlife ecology/management, parasitology, community ecology, and infectious disease evolution/life history (up to 5 responses were included per person; n participants=410, n responses=1739). The least common areas (<60 responses) included behavioral ecology, bioinformatics, field and laboratory techniques, movement ecology, virology, landscape ecology, and zoology.

Participants could rank up the ten aggregated areas of research (e.g. all types of ecology were combined) by importance (n=152). The 'most important' area was by far ecology, then infectious

disease evolution/life history, epidemiology, and field/laboratory techniques. The 'least important' areas were parasitology, immunology, field/laboratory techniques, microbiology/virology, and genetics/genomics/bioinformatics. Ecology was also listed as least important, suggesting that most participants considered ecology as either the most or least important area of research; however, the number of responses for the former was nearly three times greater than for the latter.

Survey respondents were asked to write in scientific journals and articles that they believed were the most influential in disease ecology (Table 1). Influential articles (written in at least twice, n=76) were most often published in *Science* (n=17), *Nature* (n=10), *Trends in Ecology & Evolution* (n=9), *Ecology Letters* (n=8), and *Proceedings of the National Academy of Sciences* (n=6), generally matching the influential journal rankings. See Supporting Information for full lists of both articles and journals.

Table 1. The five highest ranking scientific journals (left) and articles (right) based on survey responses. Survey respondents were asked to pick journals other than *Science* or *Nature*.

Influential journals	**n**	**Influential articles**	**n**
Proceedings of the Royal Society - Biology	98	Hudson et al. 1998. Prevention of Population Cycles by Parasite Removal. *Science.*	18
Ecology Letters	90	Anderson & May. 1979. Population biology of infectious diseases: Part I. *Nature.*	13
Proceedings of the National Academy of Sciences of the United States of America	85	Anderson & May. 1978. Regulation and Stability of Host-Parasite Population Interactions: I. Regulatory Processes. *Journal of Animal Ecology.*	12
Ecology	72	Lloyd-Smith et al. 2005. Superspreading and the effect of individual variation on disease emergence. *Nature.*	10

Journal of Animal Ecology	60	Keesing et al. 2006. Effects of species diversity on disease risk. *Ecology Letters*.	9

Literature search & analyses

We compiled a list of 42 journals that at least four survey participants said were the most important in disease ecology, plus *Science* and *Nature*. We searched these 44 journals for relevant articles in the field using the algorithm described below, and our final corpus comprised 18,695 articles. Our validation processes demonstrated that approximately 75% of these articles were properly classified. Articles span from 1975 to 2018, with most published after 2000, indicating a rapid and considerable expansion of the field. However, some journals were not available in Web of Science until the 1980s-1990s, so article availability may slightly bias our corpus in the early years.

Topic clusters are classified into two categories: those that emerged and were identifiable from the literature, and those that we deliberately searched for using key term searches. Of emergent topics, malaria and mosquito-borne pathogens appeared most frequently in the topic clusters (3/15), followed by experimental infection trials (2/10). Other clear topics included HIV, influenza, vaccine research, host-pathogen coevolution, and infectious disease modeling. Some topics were more ambiguous but still identifiable, such as wildlife pathogens, tick-borne pathogens with rodent hosts, and serological analyses. Overall, we had high confidence assigning names to topic clusters emerging from the literature, indicating defined areas of research in the corpus (for more, see Supporting Information).

Many of the topics that emerged from the disease ecology literature, such as malaria, influenza, and vaccination, have remained constant over time (Fig. 4B). Others, such as HIV and serology,

287 have declined over time, and host-pathogen coevolution has instead steadily increased. These
288 emergent topics comprised a notable portion of the disease ecology literature and were more
289 prominent than author-selected topics. We constructed a neutral topic for comparison, which
290 was constant through time (Fig. 4B, gray line in panels), thus validating the observed temporal
291 changes in these topics.
292
293 Using key term searches, we next explored select topic trends: climate change, emerging
294 infectious diseases (EIDs), the dilution effect, superspreaders, network analysis, pathogens in
295 rodents and bats, bovine tuberculosis, and chytrid fungus in amphibians (Fig. 4B). As with
296 emergent topics, our topic detection was sensitive to detecting changes in frequency over time,
297 identifying global and local peaks and troughs. For instance, research on pathogens in bats had
298 a small peak around the time of the first SARS epidemic (2002-2004), and bat disease research
299 has steadily increased since 1979; however, overall, research on bat pathogens has been a
300 small proportion of all disease ecology literature. Research on rodent pathogens was greatest in
301 the 1990s and has generally declined, although recent years have also seen an increase in
302 rodent-related disease publications. Many topics have steadily gained recognition in the
303 literature, and continue to grow, such as EIDs, climate change, the dilution effect, network
304 analyses, and superspreaders. Bovine tuberculosis and chytrid fungus literature, on the other
305 hand, appear to have declined or plateaued in more recent years.
306

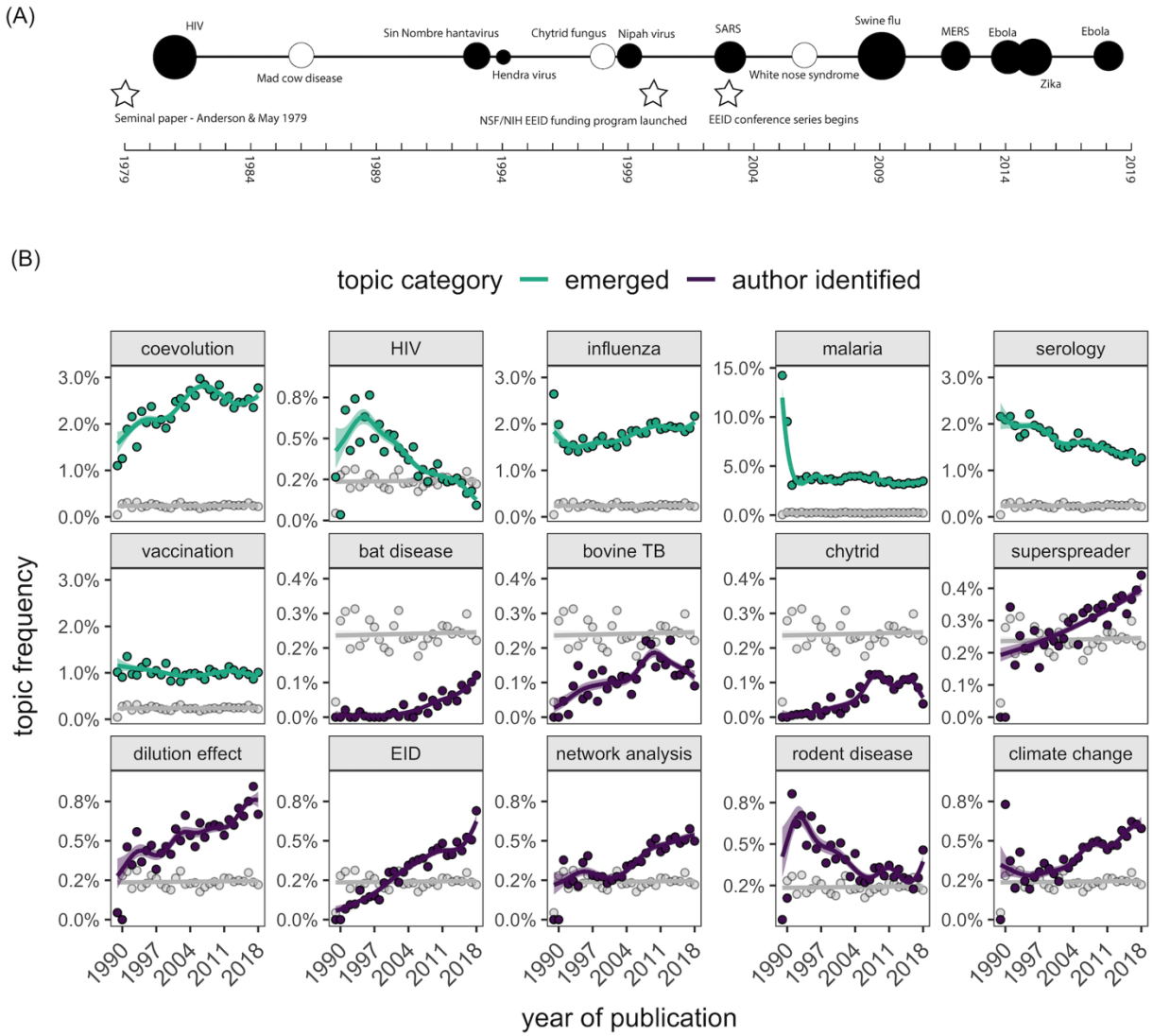

Figure 4. Timelines of events relevant to disease ecology and research trends. **(A)** A timeline of select human (filled circles) and wildlife (open circles) infectious disease events from 1979-2019. The approximate number of humans infected is represented by circle size (scaled by log/4). Analogous estimates are rare for wildlife diseases, thus these circles are an equivalent arbitrary size. Stars denote notable events for the development of the field of disease ecology. **(B)** Frequency of selected topics (green = emerged from the literature corpus, purple = selected by authors) compared with the neutral topic (gray) from 1976-2018; 1976-1989 are binned in the

first data point. Thick lines and ribbons show the fitted values and 95% confidence intervals from the GAMs.

Topic trends displayed strong degrees of cross-correlation, with both positive and negative covariation in annual trends (Spearman's rank correlation coefficient ρ=-0.89-0.88; Fig. S6). Particularly strong positive correlations were observed for superspreaders and network analyses (ρ=0.88), superspreaders and the dilution effect (ρ=0.85), EIDs and bats (ρ=0.83), EIDs and the dilution effect (ρ=0.83), EIDs and network analysis (ρ=0.82), superspreaders and bat pathogens (ρ=0.80). Especially strong negative correlations were observed for EIDs and serology (ρ=-0.89), serology and the dilution effect (ρ=-0.87), serology and bat disease (ρ=-0.81), influenza and HIV (ρ=-0.77), climate change and HIV (ρ=-0.76), network analysis and serology (ρ=-0.70), influenza and rodent disease (ρ=-0.68), climate change and serology (ρ=-0.68), bovine tuberculosis and serology (ρ=-0.67), malaria and network analysis (ρ=-0.65), and malaria and rodent disease (ρ=-0.65).

DISCUSSION

Interdisciplinary research fields often rapidly grow to address important societal needs, and retrospective analysis of their evolution can help improve their future trajectory and growth. Disease ecology in particular is a growing field of research that formed from the application of population dynamics to parasite and pathogen biology to provide dynamical insights and identify intervention strategies (Anderson & May 1979; May & Anderson 1979; Grenfell et al. 1995; Wilson et al. 2019). By combining a survey with a unique and powerful quantitative literature synthesis, we here demonstrate the increasing demographic and institutional diversity of disease ecology practitioners alongside the breadth of research activities. Certain topical themes that emerged from our literature corpus, such as influenza, malaria, and vaccination, have remained prominent foci of disease ecology, whereas an increase in *a priori* selected

341 topics such as emerging infectious diseases, climate change, and effects of biodiversity loss
342 emphasize how this expanding field has grown to meet global health crises and priorities.
343 Further, addressing diversity and allocating resources toward these growing topics could
344 promote equity within disease ecology, improve training programs, designate funding
345 opportunities, and provide the infrastructure for concentrated advancements.
346
347 Self-declared disease ecology practitioners are becoming more diverse in terms of country of
348 education, gender identity, and institution (Fig. 1). This echoes similar trends in conservation
349 (Jones & Solomon 2019), engineering (National Science Foundation 2020), and ecology as a
350 whole (Fox et al. 2018, 2019). Diversity in the workplace and educational institutions is
351 fundamentally important but also increases performance, cooperation, problem-solving, and
352 student retention (Milem 2003; Roberge & Van Dick 2010; Drury et al. 2011). Demographic
353 expansions also facilitate an influx of new ideas and perspectives into a field. The highest
354 demographic and institutional diversity in disease ecology existed in younger age groups (<36
355 years old), graduate programs, and post-doctoral positions (Table S2,S3). This may be due to
356 increasing levels of education globally (Group of Eight 2013, Unesco Institute for Statistics), or
357 targeted programs to increase diversity in science and mathematics, particularly focused on
358 recruiting women (Burke & Mattis 2007; Huntoon & Lane 2007). Another non-mutually exclusive
359 driver of these trends could be the failed retention of minorities and women in later career
360 stages (Blickenstaff 2005; Diekman et al. 2010; Shaw & Stanton 2012). Although we identified
361 some increases in diversity within disease ecology, concerted efforts to improve equity must
362 continue and explicitly address retention. This is especially salient in terms of racial and ethnic
363 diversity.
364
365 The second part of our study comprised an extensive literature synthesis. Literature reviews can
366 be compromised by author bias when search terms are subjectively selected (reviewed in Okoli

367 2015). We organically constructed a literature corpus that reduced the likelihood of author bias
368 by utilizing key terms derived through topic detection and including journals identified as
369 important to the field by a large group of self-declared disease ecologists. Additional validations
370 of papers mistakenly included or excluded enabled us to reach a final corpus of over 18,500
371 papers that represent the majority of all disease ecology publications. Ecological literature
372 reviews often fail to account for search bias and do not provide transparency in inclusion criteria
373 or type I and type II error rates (Haddaway & Watson 2016). Additionally, ecological literature
374 reviews usually consist of a <2,000 article corpus and often much fewer (Poff & Zimmerman
375 2010; Wortley et al. 2013; Lowry et al. 2013; Han & Ostfeld 2019). Using our robust corpus
376 formulation and validation (Fig. 1), we are confident we were able to identify true, broad patterns
377 across the literature at an unprecedented scale.
378
379 Topic detection of publications revealed research priorities and their changes over time. For
380 example, HIV research output peaked in the 1990s but in recent years has declined to be lower
381 than most other topics in the field. While a direct association is difficult to demonstrate, the
382 decrease in HIV-related publications has roughly coincided with advances in HIV treatments
383 (hiv.gov). Likewise, although we also observed significant temporal fluctuations in publications
384 related to malaria, their frequency has remained remarkably constant since the late 1990s. This
385 likely reflects ongoing and continued efforts to reduce public health burdens of this disease and
386 to understand complex interactions between mosquito vectors, human hosts, and the
387 environment (Suh et al. 2020), especially in the context of emerging human pathogens (e.g.,
388 *Plasmodium knowlesi*) (Lee et al. 2011). We also identified broader concept-based trends in
389 disease ecology literature. In particular, research on the dilution effect has undergone several
390 spikes following key findings (Keesing et al. 2006; Civitello et al. 2015) and a steady increase in
391 publication rate. Similarly, predicting how climate changes may alter pathogen spread continues
392 to be of growing research interest (Altizer et al. 2013; Ryan et al. 2019), as does research on

393 superspreaders (Lloyd-Smith et al. 2005). More broadly, these temporal patterns in publications
394 suggest that work related to biodiversity and infectious disease, climate change, pathogen
395 spillover, heterogeneity in pathogen transmission, and new tools to analyze epidemiological
396 data will all continue to be active areas of inquiry, and fruitful for future funding and graduate
397 training.
398
399 Research on emerging infectious diseases and bats has increased through time, consistent with
400 bats being established reservoir hosts for pathogens such as Nipah virus, SARS, Marburg virus,
401 and Hendra virus, and with infection-related population declines in bats through white-nose
402 syndrome (Calisher et al. 2006; Frick et al. 2010; Fig. 4A). However, although disease ecology
403 publications on bat pathogens have increased markedly in recent years, they still remain
404 relatively understudied compared to our neutral term and rodent pathogens, and comprise only
405 a small proportion of emerging infectious disease research in general. It is worth noting that our
406 search does not include the current SARS-CoV-2 pandemic. Due to its global impacts, we
407 expect a large spike in disease ecology research on pathogens with their evolutionary origins in
408 bats and the role of intermediate hosts in spillover, as evidenced by recent publications (e.g.,
409 Andersen et al. 2020).
410
411 In general, research on epidemics tended to be responsive rather than anticipatory, such that
412 we observed an immediate spike in publications on high-profile pathogens followed by a decline
413 or plateau (e.g., bovine tuberculosis and chytrid fungus). Emergent topics were remarkably
414 stable through time, with the exception of HIV and host-pathogen coevolution, which have
415 respectively decreased and increased. Research focusing on concepts (e.g., dilution effect,
416 superspreaders, coevolution) or approaches (e.g., network analyses) rather than specific hosts
417 or pathogens tended to rise more gradually and remain a notable proportion of the literature. On
418 the other hand, mosquito-borne pathogens and influenza have been defining topics over the

419 entire time series, which we expect to persist for the foreseeable future. Although our analysis of
420 cross-correlation between the topic time series is associative, we observed several especially
421 interesting relationships. In particular, publications on bat disease, chytrid fungus, climate
422 change, the dilution effect, superspreaders, and emerging infectious diseases were all positively
423 correlated over time, suggesting a general increase in research on emergent disease risks to
424 wildlife and humans in relation to anthropogenic change and heterogeneities in pathogen
425 transmission (Daszak et al. 2001; Lloyd-Smith et al. 2005; Jones et al. 2008). From another
426 perspective, publications on serology have steadily declined over time and were negatively
427 associated with topics such as bat disease and emerging infectious diseases, possibly
428 indicating advances in rapid and affordable sequencing efforts to quantify pathogen diversity
429 (Lipkin 2013). Lastly, negative cross-correlations also show areas where research overlap may
430 be low but could be insightful, such as applying network analyses to studies of malaria or how
431 climate change will affect HIV (Wesolowski et al. 2012; Talman et al. 2013).
432
433 Using a global survey and large-scale quantitative literature synthesis, we here demonstrated
434 that disease ecology is a rapidly growing field that is becoming more diverse, although
435 continued efforts are critically needed to enhance retention. Our analysis identified trends and
436 patterns in research topics addressed by disease ecologists that can help prioritize future
437 training, funding, and research efforts. Together, these findings indicate that disease ecology as
438 a field is well equipped to continue increasing our understanding of the complex and diverse
439 interactions among pathogens, hosts, and the environment. As such, its practitioners form a
440 rapidly growing and critical component of the global response to emerging health priorities and
441 environmental threats. More broadly, our quantitative synthesis framework used here could help
442 examine the composition and trends of other major research topics that cross traditional
443 disciplinary boundaries.
444

Data Availability Statement

Pending acceptance, appropriate data and code will be published in a Dryad Repository and listed here.

Author Contributions

EEB designed, disseminated, and analyzed the survey, conducted the literature search in Web of Science and drafted most of the initial manuscript; LS conducted the topic detection; DJB and KMF conducted the article classification and wrote sections of the manuscript. All authors contributed to conceptual development of the study, editing, and approved the final version of the manuscript.

Acknowledgements

Our survey was considered exempt under the Pennsylvania State University Institutional Review Board (STUDY00010582). We thank all participants who completed the online survey as well as helpful feedback from participants at the 2019 Ecology and Evolution of Infectious Diseases conference. We are grateful to Vanessa Ezenwa and Peter Hudson for constructive comments on an earlier version of this report.

References

Ahmed, H., J. Yoder, W. A. de Glanville, A. Davis, T. J. Kibona, B. T. Mmbaga, F. Lankester, E. S. Swai, and S. Cleaveland. 2019. Economic burden of livestock disease and drought in Northern Tanzania. Journal of Development and Agricultural Economics 11:140–151.

Altizer, S., R. S. Ostfeld, P. T. J. Johnson, S. Kutz, and C. D. Harvell. 2013. Climate change and infectious diseases: From evidence to a predictive framework. Science 341:514–519.

471 Andersen, K. G., A. Rambaut, W. I. Lipkin, E. C. Holmes, and R. F. Garry. 2020. The
472 proximal origin of SARS-CoV-2. Nature medicine 26:450–452.
473 Anderson, R. M., B. Anderson, and R. M. May. 1992. Infectious diseases of humans:
474 dynamics and control. Oxford University Press.
475 Anderson, R. M., and R. M. May. 1979. Population biology of infectious diseases: Part I.
476 Nature 280:361–367.
477 Benavides, J. A., E. Rojas Paniagua, K. Hampson, W. Valderrama, and D. G. Streicker.
478 2017. Quantifying the burden of vampire bat rabies in Peruvian livestock. PLoS
479 Neglected Tropical Diseases 11:1–17.
480 Bird, S., E. Klein, and E. Loper. 2009. Natural language processing with Python: analyzing
481 text with the natural language toolkit. O'Reilly Media, Inc.
482 Blei, D. M. 2012. Probabilistic topic models. Communications of the ACM 55:77–84.
483 Blickenstaff, J. C. 2005. Women and science careers: leaky pipeline or gender filter?
484 Gender and education 17:369–386.
485 Van Boeckel, T. P., J. Pires, R. Silvester, C. Zhao, J. Song, N. G. Criscuolo, M. Gilbert, S.
486 Bonhoeffer, and R. Laxminarayan. 2019. Global trends in antimicrobial resistance in
487 animals in low-and middle-income countries. Science 365:eaaw1944.
488 Brand, C. J. 2013. Wildlife mortality investigation and disease research: Contributions of
489 the USGS National Wildlife Health Center to endangered species management and
490 recovery. EcoHealth 10:446–454.
491 Burke, R. J., and M. C. Mattis, editors. 2007. Women and minorities in science, technology,
492 engineering, and mathematics: Upping the numbers. Edward Elgar Publishing.
493 Calisher, C., J. Childs, H. Field, K. Holmes, and T. Schountz. 2006. Bats: important
494 reservoir hosts of emerging viruses. Clinical microbiology reviews 19:531–545.
495 Carlson, C. J., and E. Mendenhall. 2019. Preparing for emerging infections means
496 expecting new syndemics. The Lancet 394:297.

Civitello, D. J., J. Cohen, H. Fatima, N. T. Halstead, J. Liriano, T. A. McMahon, C. N. Ortega, E. L. Sauer, T. Sehgal, S. Young, and J. . Rohr. 2015. Biodiversity inhibits parasites: broad evidence for the dilution effect. Proceedings of the National Academy of Sciences 112:8667–8671.

Condeso, T. E., and R. K. Meentemeyer. 2007. Effects of landscape heterogeneity on the emerging forest disease sudden oak death. Journal of Ecology 95:364–375.

Cressler, C. E., W. A. Nelson, T. Day, and E. Mccauley. 2014. Disentangling the interaction among host resources, the immune system and pathogens. Ecology Letters 17:284–293.

Daszak, P., A. A. Cunningham, and A. D. Hyatt. 2001. Anthropogenic environmental change and the emergence of infectious diseases in wildlife. Acta Tropica 78:103–116.

Diekman, A. B., E. R. Brown, A. M. Johnston, and E. K. Clark. 2010. Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. Psychological science 21:1051–1057.

Drury, B. J., J. O. Siy, and S. Cheryan. 2011. When do female role models benefit women? The importance of differentiating recruitment from retention in STEM. Psychological Inquiry 22:265–269.

Fox, C. W., M. A. Duffy, D. J. Fairbairn, and J. A. Meyer. 2019. Gender diversity of editorial boards and gender differences in the peer review process at six journals of ecology and evolution. Ecology and Evolution 9:13636–13649.

Fox, C. W., J. P. Ritchey, and C. E. T. Paine. 2018. Patterns of authorship in ecology and evolution: First, last, and corresponding authorship vary with gender and geography. Ecology and Evolution 8:11492–11507.

521 Frick, W. F., J. F. Pollock, A. C. Hicks, K. E. Langwig, D. S. Reynolds, G. G. Turner, C. M.
522 Butchkoski, and T. H. Kunz. 2010. An emerging disease causes regional population
523 collapse of a common North American bat species. Science 329:679–682.
524 Grenfell, B. T., A. P. Dobson, and H. K. Moffatt. 1995. Ecology of infectious diseases in
525 natural populations. Volume 7. Cambridge University Press.
526 Group of Eight. 2013. The Changing PhD.
527 Haddaway, N. R., and M. J. Watson. 2016. On the benefits of systematic reviews for
528 wildlife parasitology. International Journal for Parasitology: Parasites and Wildlife
529 5:184–191.
530 Han, B. A., and R. S. Ostfeld. 2019. Topic modeling of major research themes in disease
531 ecology of mammals. Journal of Mammalogy 100:1008–1018.
532 Hedges, L. V., and I. Olkin. 2014. Statistical methods for meta-analysis. Academic press.
533 hiv.gov. 2019. A timeline of HIV/AIDS. https://www.hiv.gov/sites/default/files/aidsgov-
534 timeline.pdf.
535 Hudson, P. J., A. P. Dobson, and D. Newborn. 1998. Prevention of Population Cycles by
536 Parasite Removal. Science 282:2256–2258.
537 Hudson, P. J., A. Rizzoli, B. T. Grenfell, H. Heesterbeek, and A. P. Dobson. 2002. Ecology
538 of wildlife diseases. Oxford University Press.
539 Huntoon, J. E., and M. J. Lane. 2007. Diversity in the geosciences and successful
540 strategies for increasing diversity. Journal of Geoscience Education 55:447–457.
541 Johnson, P. T., J. C. De Roode, and A. Fenton. 2015. Why infectious disease research
542 needs community ecology. Science 349:1259504.
543 Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P.
544 Daszak. 2008. Global trends in emerging infectious diseases. Nature 451:990–993.
545 Jones, M. S., and J. Solomon. 2019. Challenges and supports for women conservation
546 leaders. Conservation Science and Practice 1:e36.

547 Joseph, M. B., J. R. Mihaljevic, A. L. Arellano, J. G. Kueneman, D. L. Preston, P. C. Cross,
548 and P. T. Johnson. 2013. Taming wildlife disease: bridging the gap between science
549 and management. Journal of Applied Ecology 50:702–712.
550 Keesing, F., R. D. Holt, and R. Ostfeld. 2006. Effects of species diversity on disease risk.
551 Ecology Letters 9:485–498.
552 Koprivnikar, J., and P. T. J. Johnson. 2016. The Rise of Disease Ecology and Its
553 Implications for Parasitology— A Review. Journal of Parasitology 102:397–409.
554 Lajeunesse, M. J. 2010. Achieving synthesis with meta-analysis by combining and
555 comparing all available studies. Ecology 91:2561–2564.
556 Lajeunesse, M. J. 2016. Facilitating systematic reviews, data extraction and meta-analysis
557 with the metagear package for R. Methods in Ecology and Evolution 7:323–330.
558 Lee, K. S., P. C. Divis, S. K. Zakaria, A. Matusop, R. A. Julin, D. J. Conway, J. Cox-Singh,
559 and B. Singh. 2011. Plasmodium knowlesi: reservoir hosts and tracking the
560 emergence in humans and macaques. PLoS Pathogens 7.
561 Lipkin, W. I. 2013. The changing face of pathogen discovery and surveillance. Nature
562 Reviews Microbiology 11:133–141.
563 Lloyd-Smith, J. O., S. J. Schreiber, P. E. Kopp, and W. M. Getz. 2005. Superspreading and
564 the effect of individual variation on disease emergence. Nature 438:355–359.
565 Loper, E., and S. Bird. 2002. NLTK: the natural language toolkit. arXiv preprint cs/0205028.
566 Lowry, E., E. J. Rollinson, A. J. Laybourn, T. E. Scott, M. E. Aiello-Lammens, S. M. Gray, J.
567 Mickley, and J. Gurevitch. 2013. Biological invasions: a field synopsis, systematic
568 review, and database of the literature. Ecology and Evolution 3:182–196.
569 May, R. M., and R. M. Anderson. 1979. Population biology of infectious diseases: Part II.
570 Nature 280:455–461.

Milem, J. F. 2003. The educational benefits of diversity: Evidence from multiple sectors. Pages 126–169 Compelling interest: Examining the evidence on racial dynamics in higher education.

National Science Foundation. 2020. Women, Minorities, and Persons with Disabilities in Science and Engineering. https://www.nsf.gov/statistics/2017/nsf17310/digest/fod-women/engineering.cfm.

Nunez-Mir, G. C., B. V. Iannone, B. C. Pijanowski, N. Kong, and S. Fei. 2016. Automated content analysis: addressing the big literature challenge in ecology and evolution. Methods in Ecology and Evolution 7:1262–1272.

Okoli, C. 2015. A guide to conducting a standalone systematic literature review. Communications of the Association for Information Systems 37.

Olival, K. J., P. R. Hosseini, C. Zambrana-Torrelio, N. Ross, T. L. Bogich, and P. Daszak. 2017. Host and viral traits predict zoonotic spillover from mammals. Nature 546:646–650.

Patz, J. A., P. Daszak, G. M. Tabor, A. A. Aguirre, M. Pearl, J. Epstein, N. D. Wolfe, A. M. Kilpatrick, J. Foufopoulos, D. Molyneux, D. J. Bradley, F. P. Amerasinghe, R. W. Ashford, D. Barthelemy, R. Bos, D. J. Bradley, A. Buck, C. Butler, E. S. Chivian, K. B. Chua, G. Clark, R. Colwell, U. E. Confalonieri, C. Corvalan, A. A. Cunningham, J. Dein, A. P. Dobson, J. G. Else, J. Epstein, H. Field, P. Furu, C. Gascon, D. Graham, A. Haines, A. D. Hyatt, A. Jamaluddin, E. F. Kleinau, F. Koontz, H. S. Koren, S. LeBlancq, S. Lele, S. Lindsay, N. Maynard, R. G. McLean, T. McMichael, D. Molyneux, S. S. Morse, D. E. Norris, R. S. Ostfeld, M. C. Pearl, D. Pimentel, L. Rakototiana, O. Randriamanajara, J. Riach, J. P. Rosenthal, E. Salazar-Sanchez, E. Silbergeld, M. Thomson, A. Y. Vittor, L. Yameogo, and V. Zakarov. 2004. Unhealthy landscapes: Policy recommendations on land use change and infectious disease emergence. Environmental Health Perspectives 112:1092–1098.

Poff, N. L., and J. K. Zimmerman. 2010. Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. Freshwater Biology 55:194–205.

Roberge, M.-É., and R. Van Dick. 2010. Recognizing the benefits of diversity: When and how does diversity increase group performance? Human Resource management review 20:295–308.

Ryan, S. J., C. J. Carlson, E. A. Mordecai, and L. R. Johnson. 2019. Global expansion and redistribution of Aedes-borne virus transmission risk with climate change. PLoS Neglected Tropical Diseases 13.

Scheele, B. C., F. Pasmans, L. F. Skerratt, L. Berger, A. Martel, W. Beukema, A. A. Acevedo, P. A. Burrowes, T. Carvalho, A. Catenazzi, and I. De la Riva. 2019. Amphibian fungal panzootic causes catastrophic and ongoing loss of biodiversity. Science2 363:1459–1463.

Scheiner, S. M., and J. P. Rosenthal. 2006. Ecology of infectious disease: Forging an alliance. EcoHealth 3:204–208.

Shaw, A. K., and D. E. Stanton. 2012. Leaks in the pipeline: Separating demographic inertia from ongoing gender differences in academia. Proceedings of the Royal Society B: Biological Sciences 279:3736–3741.

Sokolow, S. H., N. Nova, K. M. Pepin, A. J. Peel, J. R. Pulliam, K. Manlove, P. C. Cross, D. J. Becker, R. K. Plowright, H. McCallum, and G. A. De Leo. 2019. Ecological interventions to prevent and manage zoonotic pathogen spillover. Philosophical Transactions of the Royal Society B: Biological Sciences 374:20180342.

Suh, E., M. K. Grossman, J. L. Waite, N. L. Dennington, E. Sherrard-Smith, T. S. Churcher, and M. B. Thomas. 2020. The influence of feeding behaviour and temperature on the capacity of mosquitoes to transmit malaria. Nature Ecology & Evolution:1–12.

Talman, A., S. Bolton, and J. L. Walson. 2013. Interactions between HIV/AIDS and the environment: Toward a syndemic framework. American Journal of Public Health 103:253–261.

World Health Organization. 1996. Press release: Infectious diseases kill over 17 million people a year: WHO warns of global crisis. https://www.who.int/whr/1996/media_centre/press_release/en/.

World Health Organization. 2018. The top 10 causes of death. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

UNESCO Institute for Statistics (UIS) Higher Education. 2020. http://data.uis.unesco.org/, http://uis.unesco.org/en/topic/higher-education.

VanderWaal, K. L., and V. O. Ezenwa. 2016. Heterogeneity in pathogen transmission: mechanisms and methodology. Functional Ecology 30:1606–1622.

Wesolowski, A., N. Eagle, A. J. Tatem, D. L. Smith, A. M. Noor, R. W. Snow, and C. O. Buckee. 2012. Quantifying the impact of human mobility on malaria. Science 338:267–270.

Wilson, K., A. Fenton, and D. Tompkins. 2019. Wildlife Disease Ecology: Linking Theory to Data and Application. Cambridge University Press.

Wood, S. N. 2006. Generalized Additive Models: An Introduction with R. Edition 1. Chapman & Hill/CRC.

Wortley, L., J. M. Hero, and M. Howes. 2013. Evaluating ecological restoration success: a review of the literature. Restoration Ecology 21:537–543.