Parasite biodiversity faces extinction and redistribution in a changing climate


Climate change is a well-documented driver of both wildlife extinction and disease emergence, but the negative impacts of climate change on parasite diversity are undetermined. We compiled the most comprehensive spatially explicit data set available for parasites, projected range shifts in a changing climate, and estimated extinction rates for eight major parasite clades. On the basis of 53,133 occurrences capturing the geographic ranges of 457 parasite species, conservative model projections suggest that 5 to 10% of these species are committed to extinction by 2070 from climate-driven habitat loss alone. We find no evidence that parasites with zoonotic potential have a significantly higher potential to gain range in a changing climate, but we do find that ectoparasites (especially ticks) fare disproportionately worse than endoparasites. Accounting for host-driven coextinctions, models predict that up to 30% of parasitic worms are committed to extinction, driven by a combination of direct and indirect pressures. Despite high local extinction rates, parasite richness could still increase by an order of magnitude in some places, because species successfully tracking climate change invade temperate ecosystems and replace native species with unpredictable ecological consequences.

INTRODUCTION

The biotic footprint of global climate change, quantified in the shifting distributions and extinctions of animal and plant taxa, has been a subject of intense research since the turn of the century (1–3). Although some species can track shifting climates (4), many species likely face extinction at rates projected a decade ago to be as high as 15 to 37% (5, 6). Despite recent refinements of that estimate, the overall rate of climate change–driven extinction is likely closer to 8% (7); others suggest that if current extinction rates (from climate change and other anthropogenic impacts) persist for hundreds to thousands of years, then total extinction could cross the 75% threshold that defines a geological mass extinction event (8). However, previous work has focused nearly exclusively on free-living biodiversity (especially vertebrates), and many important functional or taxonomic groups remain undescribed or are only now being included in extinction research (9). Particularly poorly profiled are commensalists, mutualists, and parasites (10, 11), which should exhibit an atypically high extinction rate because of their dependence on other species for survival (12).

Despite substantial research on parasite coextinction risk (10, 12) and an emerging body of theoretical work predicting the potential adverse impacts of climate change on parasites (13), climate change–driven extinction rates have never been estimated for parasitic groups, perhaps because the long-term data needed to detect extinctions in progress had not been previously collated (14). A recent study predicts that one of the most reliable benchmarks of parasite extinction risk should be their loss of suitable habitat but notes that distributional data are lacking for most parasites (15). For species with available data, two frequently cited studies contrastingly predict either local range loss (16) or global range increases for ixodid ticks (17). Even for zoonotic parasites, which are closely monitored compared to most of the parasitic organisms on Earth [especially in the context of climate change research (15)], the net relationship between climate change and disease emergence is uncertain (18). Early work argued that a warming climate facilitates range expansion (19), although others have predicted that range shifts will be accompanied by little expansion (20). Further evidence suggests that some zoonoses—like the nematode that causes angiostrongyliasis in humans—could lose suitable habitat as a result of climatic changes (21).

If parasites face severe extinction risk in a changing climate, then the cascading impacts on ecosystems are likely to be profound. Many parasites play an important immunoregulatory role in host populations, and some studies have found that a higher diversity of parasites can act as a partial buffer against the emergence of a virulent pathogen (22). Previous work has also pointed to the merits of parasites as regulators and connectors in resource-consumer webs, in which they can sometimes constitute more than 75% of the total links and in which their occasional role in altering host behavior can be critical to the flow of biomass between trophic levels (23). Despite their many hidden benefits, parasites are a difficult subject for conservation research, because parasites can come at an economic (for example, crop pathogens) or a health (for example, emerging infectious diseases) cost to wildlife and human populations. In the context of climate change research, the balance between parasite extinction and emergence is uncertain, and although some work has suggested these could be complementary processes (24), the net impact of climate change on parasite biodiversity is still unresolved.
Here, we use an updated and improved set of climate envelope models, conceptually based on those of Thomas et al. (6), and apply them to a data set we have assembled that contains the most comprehensive multispecies occurrence data set available for clades containing terrestrial macroparasites or groups historically treated as parasites, all of which have often been overlooked in extinction risk estimation. We focus on eight major clades: Acanthocephala (spiny-headed worms), Astigmata (two superfamilies of primarily ectosymbiont feather mites), Cestoda (tapeworms), Ixodida (ticks), Nematoda (roundworms), Phthiraptera (lice), Siphonaptera (fleas), and Trematoda (flukes). Our data combine and refine existing online repositories and newly digitized museum collections. More than 170,000 unique candidate points were reduced through strict data cleaning, quality control, and sample size limits to 53,133 georeferenced parasite occurrence records, cataloging the distributions of 457 species (fig. S1). From these data, we constructed maximum entropy species distribution models and projected each species’ distributional shift for the year 2070 under an ensemble of climate change models and scenarios (Fig. 1). As in the original Thomas et al. study, we forecast loss of “native” range (that is, loss of areas currently occupied: “0% dispersal”) and compare it to overall changes in suitable range (including areas not previously occupied: “100% dispersal” and “global”). The resulting maps of current and projected ranges, forecasts of extinction risk and habitat loss, measures of zoontic potential, and model accuracy metrics are available through the Parasite Extinction Assessment and Red List (PEARL), an atlas that is available online at pearl.berkeley.edu (fig. S2), and the data are available as supplementary files on request.

RESULTS

We found that most parasites, like most free-living species (6, 7), face an existential threat from substantial habitat loss in a changing climate. Changes in habitat loss were affected slightly by differences in general circulation models (GCMs), but more pronouncedly by different climate scenarios, with a contrasting average native range loss of 20.2% in the most optimistic Representative Concentration Pathway (RCP) scenario for greenhouse gases (RCP2.6) and 37.4% in the pessimistic (RCP8.5) scenario (see fig. S3). Although no species lost its entire suitable range across every climate scenario, species lost an average of 29.0% of total habitat without dispersal, 86 species lost more than 50%, and 8 species lost more than 80% of their range. Even allowing for dispersal, 202 species still lost range by 2070, and 32 species lost more than half of their global suitable range; despite those losses, species gained an average of 16.2% suitable habitat, 29 species at least doubled the extent of their range, and 7 species at least tripped it (none of those 7 have any evidence of zoonotic potential or human infection records).

Results highlight divergent outcomes of native range loss and range expansion for different species within the same clade (fig. S4). Although previous literature has suggested that climate change may increase disease emergence through range expansion (19), we found that strictly wildlife parasites experienced 17% more range gain, on average, than human infectious or biting species. This result could likely be explained by life history differences: In a two-way analysis of variance (ANOVA) accounting for endoparasites versus ectoparasites, the signal of human infection was insignificant, although endoparasites gained 39% more range than ectoparasites with dispersal (P < 0.001) and lost 10% less native range (P < 0.001). One potential explanation is that the range limits of endoparasites are more commonly at disequilibrium as a result of dispersal limitations, whereas ectoparasites may already fill most of the suitable habitat, relatively stabilizing their range size over time; similar assortative processes have proven important for shifting non-native plant species (25).

Clade projections of total suitable range expansion with dispersal were significantly different (F = 15.441, P < 0.001), but no universal trends were visible across taxa. Lice had both the highest average native range loss and the highest average global range gain, although this conflicting result was likely a product of a small sample size. Fleas and ticks consistently fared the worst, with both having average net loss even allowing for dispersal. The Thomas et al. (6) method for extinction rate estimation followed a similar pattern, with extinction rates for all species projected at 5.7 to 9.2% without dispersal and 1.7 to 4.0% with it (see Table 1). A recent meta-analysis refined the global extinction risk estimate to 7.9% of free-living species, placing our estimate of parasites’ extinction rates in a comparable place to any other group (7). Using a simplified version of the Thomas et al. habitat loss categories for International Union for Conservation of Nature (IUCN) red listing (which are, themselves, far reduced from actual IUCN listing criteria), 6.3% of species in our study are “endangered” and 0.7% are “critically
endangered” with dispersal, and 17% of species are endangered and 1.8% are critically endangered without dispersal (Fig. 2). Clade differences should be interpreted cautiously; although expansive for current data, our 457 species are a meager subset of likely 300,000+ species of free-living vertebrates (45). An even more conservative analysis that only includes species with 50 or more points (rather than 20 or more) is available in the Supplementary Materials (see text S3). That analysis reduces the sample size from 457 species down to 196; the key results discussed above are markedly similar, although some specific extinction rates are predicted to be slightly higher (likely because extinction rate estimators converge on more conservative values as sample size increases).

In a 2008 study, Dobson et al. (26) projected a coextinction rate of 3 to 5% for helminths in the next 50 to 100 years based on host IUCN status, a far lower estimate than comparable projections for free-living vertebrates (6, 7). A rough calculation using the same data shows that, if 15 to 37% of hosts were threatened [as Thomas et al. (6) predicted], then coextinction alone would be responsible for the loss of 8 to 24% of parasite species (text S1)—a far higher rate than has ever been observed empirically. The “paradox of missing coextinctions” is a significant problem on its own (27), but compounded with the baseline extinction rates from habitat loss that we estimate here (see eq. S1 in text S1), we suggest that a much higher fraction of species might be committed to primary and/or secondary extinction without dispersal by 2070: 5.6 to 15.4% of acanthocephalans, 11.9 to 29.0% of trematodes, 12.8 to 29.1% of cestodes, and 12.5 to 29.5% of nematodes (Fig. 3). Therefore, the loss of parasite biodiversity could make a significant contribution to the sixth mass extinction, especially compared to the 7.9% baseline extinction rate suggested by Urban’s recent meta-analysis (7).

**DISCUSSION**

We adopt a more conservative methodology than Thomas et al. (6), likely producing lower estimated extinction rates. Our study used only species with 20 or more occurrences (thereby selecting disproportionately for cosmopolitan, generalist, and human-hosted parasite species with the largest ranges), ran maximum entropy models rather than using the BIOCLIM algorithm [the latter representing a method known to predict higher extinction rates compared to other niche modeling methods (20)], and selected particular GCMs (see the Supplementary Materials). Species with smaller ranges, which are disproportionately poorly sampled in our data set, are likely subject to greater than average extinction risk. In addition, dispersal capacity will have a profound effect in determining which parasite ranges expand or contract, because some ectoparasites are likely to shift independent of host distributions (29) [leading to novel evolutionary opportunities—see the Stockholm Paradigm in evolutionary parasitology (22, 23)], but in other cases, hosts’ shifting ranges may “escape” their parasites (30, 31). Endoparasites with aquatic stages—like tapeworm coracidia, trematode cercariae, or the drifting planktonic stages of copepods

---

**Table 1. Habitat loss and projected extinction risk by dispersal scenario and clade.** Values are averaged across all general circulation models (GCMs) and RCP scenarios (46), and the percentage of species committed to extinction is calculated using the three Thomas et al. (6) SAR methods. Percentiles are calculated from species-level averages of GCMs and RCPs (that is, all variance is interspecific).

<table>
<thead>
<tr>
<th>Clade</th>
<th>Nspecies</th>
<th>Habitat loss (mean)</th>
<th>Habitat loss range (5th to 95th percentile)</th>
<th>% Committed to extinction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0% dispersal</td>
<td>100% dispersal</td>
<td></td>
</tr>
<tr>
<td>Acanthocephala</td>
<td>14</td>
<td>-16.6%</td>
<td>-50.6%, -0.2%</td>
<td>3.8%/4.4%/4.9%</td>
</tr>
<tr>
<td>Astig mata</td>
<td>18</td>
<td>-19.0%</td>
<td>-43.6%, -4.0%</td>
<td>4.4%/5.1%/5.3%</td>
</tr>
<tr>
<td>Cestoda</td>
<td>25</td>
<td>-13.6%</td>
<td>-29.1%, -2.9%</td>
<td>4.0%/3.6%/3.7%</td>
</tr>
<tr>
<td>Ixodida</td>
<td>141</td>
<td>-31.9%</td>
<td>-57.0%, -1.9%</td>
<td>8.1%/9.2%/9.8%</td>
</tr>
<tr>
<td>Nematoda</td>
<td>147</td>
<td>-28.0%</td>
<td>-74.4%, -2.6%</td>
<td>5.4%/7.9%/9.3%</td>
</tr>
<tr>
<td>Phthiraptera</td>
<td>5</td>
<td>-55.8%</td>
<td>-71.5%, -34.4%</td>
<td>10.5%/18.5%/19.3%</td>
</tr>
<tr>
<td>Siphonaptera</td>
<td>67</td>
<td>-40.6%</td>
<td>-69.5%, -11.0%</td>
<td>10.0%/12.2%/12.9%</td>
</tr>
<tr>
<td>Trematoda</td>
<td>40</td>
<td>-17.8%</td>
<td>-47.4%, -0.4%</td>
<td>3.8%/14.8%/6.0%</td>
</tr>
</tbody>
</table>

---

(or of long-distance dispersing hosts, such as birds)—may have greater dispersal capacity and ultimately fare better than average.

Range loss is also only one aspect of how parasites will experience climate change, and estimates of their vulnerability based only on range loss are likely to be fairly conservative. Even within a climatically suitable range, any given site may be missing hosts necessary for parasites to complete multistage life cycles; some parasites may be capable of plastic switching to truncated life cycles, but many will likely fail to persist in environments that ecological niche models would likely classify as a “suitable habitat” (15). Even when host ranges shift in concert, phenological mismatch could prevent transmission from one host species to another, even for parasites that otherwise might appear to experience an overall gain in suitable range (32). Finally, the transmission of parasites—and the interplay of virulence and host immunity—is often temperature-sensitive, and although it will have a critical role in determining parasite vulnerability to extinction, it is also essentially impossible to predict using the types of models we present here (15). In that way, parasite transmission ecology at a site-specific level could potentially produce even greater range losses than our models predict, once again making our models relatively conservative.

Even accounting for differences in sampling intensity between different continents, parasite species richness is far from being evenly
distributed at a global scale, potentially representing real underlying patterns or merely illustrating the sampling bias of parasite collections (for example, substantially greater data completeness in North America) (33). Independent of potential biases, our results strongly support the hypothesis that climate change will drive a major redistribution of parasite biodiversity through habitat gains, losses, and shifts. Following a similar approach to previous work by Cumming and Van Vuuren (17), we find that, in some cases, extinction is concentrated in the regions where our maps of parasite richness indicate diversity is the highest, as in the Gulf Coast of the United States and in most of Western Europe (Fig. 1). However, at latitudes closer to the poles, where measured richness is lower, our simulations project that species richness will double, triple, or increase even more. Many previous studies on Arctic parasite diversity have predicted and documented this developing ecological cascade (34–37), but this phenomenon had yet to be predicted by broad-scale biogeographic models before this study. This result poses an alarming question: What will the consequences of a shifting wave of new parasite species, augmenting and possibly replacing native diversity, be for ecosystem stability, wildlife communities, and human health? Most parasites are not agents of emerging disease, but destabilized host-parasite networks might nevertheless create opportunities for new patterns of emergence. The redistribution of species ranges on a global scale is likely to create opportunities for the origin and evolution of new host-parasite pairings, as well as to change the regional balance of parasite diversity in different ecosystems, thereby allowing different parasitic taxa to become ecologically dominant and potentially changing eco-evolutionary dynamics in the long term (38, 39).

While our study is the first to bring together this volume of data across parasite clades at the global scale, it is also a first step in a methodological progression that is 10 years behind the cutting edge of extinction analyses in nonparasitic animals (40, 41). Current best-practice research on climate change impacts on biodiversity depends on a cycle of data acquisition and model improvement that has progressed from basic biogeographic estimation methods to sophisticated biophysical mechanistic modeling that can encompass genetics, physiology, and dispersal. In the last decade, species-area relationship (SAR) models have been criticized as a method of assessing extinction risk, with strong evidence pointing to overestimation of extinction rates (42, 43). Future analyses are critically needed to verify the stability of the species-area curve at continental scales for helminths and ectoparasites.

The possibility for parasites to experience compounded range loss with shifting host ranges, and their increased vulnerability relative to their hosts, is robustly independent of the SAR. However, parasite ecology desperately requires updated methods that can distinguish between likely winners and losers. The vast majority of climate-driven species extinctions may only proximally be a result of range loss, whereas population size, species interactions, presence of free-living life cycle stages, number of intermediate host species, and plastic and genetic components of climate tolerance and adaptation may be better predictors (44, 45). Our study unambiguously deepens the need for experimental work, local long-term ecological experiments, and physiological mechanistic modeling to more accurately describe the threats parasites face than can be achieved with global distributional data.

Parasite conservation, as an applied discipline at the intersection of wildlife research and human health concerns, is in its infancy. Although parasite conservation is a topic of significant interest (11, 23) and has been for at least two decades (46, 47), most parasitic biodiversity is unrecorded in ecological databases. Our study, together with the release of PEARL and the associated data sets, offers a foundational framework for including parasites in conservation ecoinformatics and biogeography. Threatened parasites, as well as symbionts as a whole (48), require a specialized conservation approach, tailored to their unique life history, tremendous diversity, and the complex ecosystem services they provide. Some species in the most threatened clades may not even be parasites per se, such as the vane-dwelling feather mites in our study, for which, as for many other symbiont species, the line between parasitism, commensalism, and mutualism is dynamic and unresolved (48, 49); parasitic groups of nematodes are polyphyletic, and some of the nematodes in our data set are similarly nonparasitic and free-living (50). We include these species in our assessment nonetheless, because “parasitic clades” are markedly understudied across the board, and their vulnerability was equally unassessed compared to their parasitic counterparts (48). Similarly, protelean organisms, in which juvenile life stages are parasitic but adults are free-living (for example, mites in the Parasitengona or twisted-wing insects in the family Myrmelecidae, in which the sexes are juvenile specialists separately on ants and orthopterans), are also likely to require an extra conservation focus in a similar framework to ours.

Specialized, complex life cycles like these may face the most significant hidden vulnerability to habitat fragmentation or phenological mismatch (32) and may even experience a more severe extinction threat than our study predicts. Whereas clade-level assessments may help identify some of these highly vulnerable groups, a finer-grained study of parasite extinction risk will require a more integrated perspective. Recent research has highlighted how parasites’ range loss is likely to have a synergistic interaction with host specificity and host functional traits [like thermal ecology and body size (15)]. Combining the models presented here with Red List efforts for hosts, existing trait data, and network-based models of host parasite associations could substantially increase the resolution of parasite vulnerability assessments (including potential future iterations of PEARL). However, collecting the same big data that are revolutionizing wildlife and plant conservation will undoubtedly be a challenging next step for parasite conservationists.

MATERIALS AND METHODS

Data collection and georeferencing

Accurately describing the diversity and distribution of global parasite biodiversity is a nearly insurmountable task and, for a study of this nature, would be essentially impossible without taking advantage of existing data infrastructure, especially from natural history collections (51). Globally, we project extinction risk using a patchwork of regionally and taxonomically specialized data sets representing the best available distributional data sets in parasitology, including a published database on ixodid ticks in Africa (52, 53), a comprehensive database of all published records of feather mite occurrences (FeatherMites) recently published as a data paper in Ecology (54), data from the Global Biodiversity Informatics Facility (GBIF; www.gbif.org), the University of Michigan Museum of Zoology (UMMZ)’s database of bee mite occurrences, flea data from the VectorMap project (vectormap.si.edu), and, most significantly, the U.S. National Parasite Collection (USNPC) (fig. S1). The data sets included in this study represent some of the only true “big data” for wildlife parasites; a number of the largest data sets do not include any spatially explicit data, such as the London Natural History Museum database (55) and the FishPEST database (56). Others, like the Global Mammal Parasite Database, were not made available during the duration of our study, and some sources, like the University of Connecticut’s Global Cestode Database, do not have spatially explicit...
data that met the sample size requirements delineated below (57). Finally, a set of researchers noted in the Acknowledgments donated their data, but geographically biased focus led to their exclusion from the final data set.

In cases within these databases where occurrences lacked coordinate data, we georeferenced all specimen locality data using established guidelines in the literature (58) and the georeferencing software GEOLocate (59). For the remaining databases with existing georeferenced occurrences, we largely used the data in their original form [that is, the Cumming tick database (33,989 points) and UMMZ bee mite data set (1160 points)]. The VectorMap data set includes mites, ticks, and fleas, but the mite data did not include any species with 20 or more occurrences, and because the ticks were identical to GBIF and Cumming data sets, only fleas were included—adding an additional 8482 points to the data set. GBIF data were downloaded for eight clades [Acanthocephala, Astigmata (Astigmata), Cestoda, Ixodida, Nematoda, Phthiraptera, Siphonaptera, and Trematoda] and clipped to terrestrial points only. A number of other parasitic clades were also considered in the process of data collection, but GBIF data for the vast majority were limited and provided 10 or fewer suitable species. These required specific cleaning to remove country-level data sets that would have produced biased niche models from data limited by administrative boundaries (60)—for example, the Edaphobase data set and D. Sturhan’s German Nematode Collection–Terrestrial Nematodes (DNST) data set, both on nematodes in Germany. Records without species identification were eliminated, as were records without any coordinates. A total of 100,295 occurrences remained after the application of these guidelines (Acanthocephala, 2013; Astigmata, 4826; Cestoda, 3048; Ixodida, 17,695; Nematoda, 32,170; Phthiraptera, 3675; Siphonaptera, 23,573; and Trematoda, 13,294).

The georeferencing Web-based software GEOLocate (59) was used to assemble a spatially explicit database for the USNPC and for the feather mite database. Although the USNPC has more than 70,000 records, only some have locality data digitized, and of those, many have single or double records. To consolidate our efforts, we only georeferenced data for species with 20 or more records, a threshold chosen on the basis of both the distribution of sample sizes within the larger data set and the literature evidence about the effect of sample size on model accuracy. Although accuracy consistently stabilizes around 50 points for the most frequently used methods (61–63), 20 or more is often used as a threshold in the literature especially for maximum entropy (MaxEnt) (which particularly excels with small sample sizes), and a recent publication shows that 15 points often suffices for narrow-ranged species, and as few as 25 are sufficient for the most globally distributed species (64). The 20–or-more rule reduced the data set down to 31,212 specimens—many of which had shared locality information (that is, multiple species were collected from the same locality, often by the same collector). In the Supplementary Materials, we present an additional analysis based only on species with 50 or more points (text S3), reducing the data set from 457 species to 196. As mentioned in Results, the main analyses are essentially unaltered by the reduction, with some extinction estimators producing slightly higher values (especially for clades with already limited sample sizes). Consequently, in the main text, we have elected to present the most inclusive analysis, both maximizing the species coverage of our study and producing a slightly more conservative estimate of extinction rates.

Points attributed to townships and provinces were marked at the political center, and the uncertainty radius was the minimum to encompass the entire region. When political markers were not available for countries in the developing world, we used satellite imagery to identify the rough boundaries of cities and townships. In select cases, if the political marker was too far from the actual center, then our final point would be corrected to the center of the region of uncertainty. Broad geographical regions like Siberia, entire stretches of river or continental coasts, and points with too many candidate options of equal likelihood (for example, “Red Mountain, CA, United States,” which could be 17 localities, or “La Junta, Mexico,” which had 33 candidates) were all excluded from the final data set.

A total of 5507 specimens from the USNPC were skipped by virtue of incomplete information or because the locality data were insufficiently detailed, leaving 25,705 specimens georeferenced to 7373 unique localities. Of those remaining entries, the data were reduced to species with enough occurrence points for inclusion. A final verification against the Smithsonian’s collection management database revealed 21 inaccuracies that were subsequently corrected, and one nematomorph and five botflies were removed from the data set, as were two monogeneans that had been formerly recorded as trematodes. The final data set contained 15,741 unique entries for species with at least 20 points.

Uncertainty greater than 10 km has been shown to potentially negatively affect the accuracy of distribution models (65). However, we used climate data with a resolution of 10 arc min for our models, roughly an average of 20 km (variable owing to the shape of Earth), a coarser resolution than many niche modeling efforts, and one that absorbed much of the error associated with the data. Consequently, we set a threshold of 40 km (two cells) for acceptable maximum uncertainty in our georeferenced points and subsequently reapplied the 20–or-more-point threshold, eliminating some species in the process. Geolocation data for the FeatherMites data set were collected using the same methodology. From each data set, we removed all occurrence data that did not meet our criteria for a 40-km uncertainty radius. We then excluded all species that did not have at least 20 high-quality occurrence points.

In the final data set aggregated across every data source (available as table S12), we provided results for a total of 457 species across the compiled data sets with a total of unique 53,133 points (an average of 116 unique points per species, far above the 20-point minimum we set and on par with some of the more comprehensive distribution modeling published for individual species in the literature; fig. S5). Of those 457 species, these were broken down as follows: Acanthocephala, 14 species; Astigmata, 18 species; Cestoda, 25 species; Ixodida, 141 species; Nemato-oda, 147 species; Phthiraperta, 5 species; Siphonaptera, 67 species; and Trematoda, 40 species. Models were presented for all but one species in that set, because models ran unsuccessfully in that case, but locality data were still made available (the nematode Teratocephalus terrestris).

Climate data
Distribution models were run using the WorldClim v1.0 climate data set at 10–arc min resolution, which included 19 bioclimatic variables (Bioclim) that captured global trends and variability in precipitation and temperature (66). On the basis of models that have been previously published that forecast species distributional shifts, we selected a set of five of the most widely used GCMs at four RCPs that account for different global responses to mitigate climate change. These are the Beijing Climate Center Climate System Model (BCC-CSM1.1), the Commonwealth Scientific and Industrial Research Organisation (Australia’s GCM) (ACCESS-1.0), the Hadley GCM (HadGEM2-CC and HadGEM2-ES), and the National Center for Atmospheric Research Community Climate System Model (CCSM4). Each of these can be represented at RCP2.6, RCP4.5, RCP6.0, and RCP8.5, capturing a range of scenarios.
that species could experience (except ACCESS-1.0, which only exists for RCP4.5 and RCP8.5). Covariance between predictors and the definition of the accessible area are both significant problems with environmental predictors used in niche modeling efforts. To address these issues, models were trained on a data subset to continents with known parasite occurrences, and regularization procedures in MaxEnt accounted for collinearity in predictor variables (see below).

**Distribution modeling**

Our study was designed to reproduce the approach underlying the seminal Thomas et al. (6) paper, which projected that 15 to 37% of terrestrial species likely faced imminent extinction from climate change. The overall order of operations is conserved: (i) "Climate envelope" models are constructed using current best practices in ecological niche modeling, (ii) species range shifts are forecasted in response to climate change, and (iii) macroecological inference is made with respect to the consequences of that habitat loss for species extinction rates.

Since the publication of the study by Thomas et al., the climate envelope method for estimating extinction rates has drawn some criticism. Although climate envelopes (now more commonly termed ecological niche models or species distribution models) are one of the most commonly used statistical methods in ecology, and hundreds of papers and several books outline best practices for their implementation (67–69), many researchers are still skeptical of the methodology. Niche concepts (foundational to climate envelope models) are contentious in ecology, and significant literature has been devoted over the years to basic assessment and debate about the utility and validity of niche theory as an approach to ecology (70–72). More practically, climate envelope models fail to account for the roles that biotic interactions, source-sink dynamics, and dispersal play in determining range shifts (73). Moreover, model performance can be challenging to accurately and honestly assess, especially given the nonindependence inherent in the collection of many occurrence data sets (74).

However, for species that are poorly documented in situ, the climate envelope approach remains popular as a tool for inference regarding geographic distributions. Parasites especially qualify for such an approach, given that mounting a field survey on the scale of our study would be nearly impossible, compared to the efficiency with which existing data sources can be used in the Thomas et al. framework. More broadly, in the absence of data that could parameterize more detailed methods like integral projection models, climate envelope models are still widely regarded as a powerful and popular method for studying climate change impacts on species ranges. Empirical work has even shown that climate envelopes validate well against real range shifts, especially as a "first-order approximation" (74, 75).

Where the Thomas et al. study is outdated from 12 years of updated scientific literature, we have correspondingly updated the methodology to improve accuracy and theoretical validity. Foundational to that methodology is the assumption that most or all species have a fundamental niche, measurable in multidimensional climate space, which constrains their geography. For parasitic species, the relationship between bioclimatic variables and geography may be less intuitive than for plants or most animals, and in some cases, parasites with a high $R_0$ and low environmental sensitivity may have ranges predominantly driven by their hosts. However, numerous cases exist where parasites are directly sensitive to climate. Among the many examples, humidity and aridity define geographic boundaries between feather lice on a shared host (76); exposure to salt spray (77), altitude, and extreme cold negatively affect feather mites (78, 79); precipitation and soil type can have a profound effect on free-living stages of helminths (80); and further diverse support for parasitic niches independent of hosts comes from ticks (29), the plague bacterium (*Yersinia pestis*) (81), and even parasitic mistletoe (82).

In response to the critique by Thuiller et al. (28) that highlights the significance of modeling method on forecasting range loss in response to the Thomas *et al.* paper in 2004, we have replaced the BIOClim algorithm (more commonly referred to in the literature now as "surface range envelopes") with MaxEnt regression, as developed by Phillips et al. (83, 84) and refined over the past decade. MaxEnt is widely considered to be one of the best-performing nonensemble approaches to ecological niche modeling (62, 85) and is the most widely used method for the analysis of presence-only data in the literature to date (86). MaxEnt models are frequently used to forecast species range shifts in response to climate change (87, 88) and have been found to often successfully predict the realized shift in species distributions (89). However, MaxEnt allows fitting with up to five feature classes simultaneously (linear/L, quadratic/Q, hinge/H, product/P, and threshold/T) and, without careful tuning, has the propensity to overfit models more severely than many other comparable methods (90). Consequently, using MaxEnt effectively relies on an approach that is sensitive to some major methodological pitfalls.

Sampling bias is a key problem in MaxEnt studies, and correction for it can vastly improve predictive performance (86, 91), but some of the most direct solutions—like spatial bias filters—rely on knowledge that would be inconsistent across 457 species aggregated from different sources, with different relative biases. However, other approaches like cross-validation within models, feature class and variable set reduction (to reduce unnecessary complexity), and adjustment of the regularization parameter have all been shown to vastly strengthen MaxEnt models (90, 92, 93). These processes are all automated in ENMeval, an R package that performs cross-validation and model tuning (94). We used ENMeval to automate analyses for all 457 species, with the selection of the regularization parameter and feature classes based on an analysis using AIC$_C$, a version of the Akaike information criterion that is optimized for smaller sample sizes. Models with an $\Delta$AIC$_C$ of 2 or lower were considered to be "strongly supported" (95); our automated process selected the model with the lowest AIC$_C$ ($\Delta$AIC$_C = 0$), and the feature classes and regularization multiplier selected in that analysis are available in table S13 for each species, although alternative selections minimally affect the net results (fig. S6). This process substantially penalizes overfitting and overly complex models and, because variable set reduction is also automated within the MaxEnt process, similarly reduces problems arising from covariance between Bioclim variables. Cross-validation was performed using the "checkerboard2" method, which executes geographic cross-validation at a relatively broad scale, helping to account for some degree of sampling bias within data sets. Although the jackknifing approach is more strongly recommended for species with small sample sizes [under 20 to 25 points (37, 96)], our data set already excluded species with small sample sizes by this definition, and having a single standardized model selection method across all species likely reduced the amount of noise in the final results (that is, differences between species are more likely to predict different outcomes under climate change rather than inconsistency in methods).

For every analysis at the species level, a PDF is available on request that plots $\Delta$AIC$_C$, mean AUC (area under the receiver-operator characteristic curve), and training-minus-test AUC difference against the regularization multiplier $\beta$ for six considered feature class sets: L, Q, H, LQH, LQHP, and LQHTP. These tuning aspects had no consistent effect on habitat loss projections (fig. S6). In cases where two models...
had the same ΔAICc because features were included but not used, the more minimal-feature model (of the identical pair) was selected. The AUC and true skill statistic (TSS) were calculated subsequently for each species, and as a result of problems inherent in metric inflation with AUC that have been previously discussed in the niche modeling literature (97, 98), especially for data sets with spatial bias in their collection (99), we instead relied on TSS to measure final model validity. In ensemble approaches, modelers often exclude runs under a certain TSS to maintain quality, with values ranging from 0.3 to 0.85 (100, 101). According to Coetzee et al. (102), interpretation of the TSS can be roughly conceived of as “values from 0.2 to 0.5 [are] poor, values from 0.6 to 0.8 [are] useful, and values larger than 0.8 [are] good to excellent.” On the basis of that and other work using a 0.6 threshold (103), we present a comparison of habitat loss for all models and for those with TSS > 0.6 (figs. S3 and S6), noting that the pattern is essentially unchanged. However, all models of range shifts are made available on the PEARL server with AUC and TSS information presented alongside the models to improve transparency and allow further work the option to be more selective.

Final models, selected on the basis of the lowest ΔAICc, were projected onto our set of 18 GCM/RCP combinations (fig. S3). Non-logistic outputs of current and future projections were cut off by a threshold chosen to simultaneously maximize sensitivity and specificity (that is, maximizing the TSS), turning outputs into binary geographic ranges. Differences in area between the two were most easily analyzed in R by using the BIOMOD_RangeSize function from the BIOMOD2 package (104). Differences in suitable area were calculated with 0 and 100% dispersal (that is, change in native and global habitat; fig. S4).

Likely, outcomes between 0 and 100% dispersal will be a product of a species’ own dispersal ability and the dispersal ability of their hosts. Hosts at the leading edge of range shifts may escape some parasitic infection (30, 31), but in the case of some ectoparasites, their ranges may be less constrained by host distributions (29). Further analyses meant to better optimize accuracy and discriminate extinction risk between species should likely simulate the simultaneous shifts of host and parasite ranges and thereby refine intermediate forecast scenarios for habitat loss. These will likely produce less conservative estimates of habitat loss, based on evidence that every life cycle stage can compound host-parasite spatial mismatch when climate change drives range shifts (33), although for some generalist groups like ticks, hosts may be a minimal constraint on geographic range size (29).

**Extinction rate estimation**

Distributional shifts modeled for parasites in each regional data set were used as input to the same three species-area curves given by Thomas et al. (6) to estimate regional extinction risk

\[
E_1 = 1 - \left( \frac{\sum A_{\text{new}}}{\sum A_{\text{old}}} \right)^{0.25}
\]

\[
E_2 = 1 - \left( \frac{1}{n_{\text{species}}} \sum \frac{A_{\text{new}}}{A_{\text{old}}} \right)^{0.25}
\]

and

\[
E_3 = \frac{1}{n_{\text{species}}} \left[ 1 - \left( \frac{A_{\text{new}}}{A_{\text{old}}} \right)^{0.25} \right]
\]

The SAR with an exponent of 0.25 has come under methodological criticism by authors like Harte et al. (105), Kitzes and Harte (43), and He and Hubbell (42), among others (see also text S2), but remains a widely used method in the literature from the last 1 to 2 years, often in combination with a similar RCP-structured approach to the one we used here (106). We retain the method with the disclaimer that it is, in the current literature, more an “index of extinction” than a quantitatively strict prediction. We further suggest that future studies investigate the applicability of dynamically scaling SAR methods (107), which require data on population size and aggregation that are unavailable for most, if not all, parasites in this study. Calculation of compound risk was done using extinction estimates with and without dispersal, in combination with the coextinction risks presented in text S1. This was done by assuming the two probabilities—\( p_{\text{direct}} \), due directly to climate change and \( p_{\text{coextinction}} \) due to extinction of hosts—to be independent, thereby yielding

\[
P_{\text{extinction}} = 1 - (1 - p_{\text{direct}})(1 - p_{\text{coextinction}})
\]

Numbers presented in the main text were calculated as the product of values from Table 1 and text S1 for the four main worm clades in our study.

**Species richness mapping**

A global map of species richness was constructed by stacking each distribution model included in the study and counting the total number of species predicted to be present in each cell. We also present the turnover in each cell, like many previously published studies of climate change impacts (108, 109). However, the role of sampling bias in the spatial structure of species richness cannot be overlooked; we devised an analysis based loosely on the quantile subtraction method used by Hopkins and Nunn (110) for gap analysis to correct for sampling bias. Whereas Hopkins and Nunn’s analysis compared parasite occurrences to known patterns of mammal richness, our analyses divided our parasite richness maps and point density into 10 quantiles and took the difference between them to find hot and cold spots of parasite diversity based on a prediction from sampling point density (fig. S7). We also present a breakdown of our observed hot spots between species with and without human health relevance, showing the effect of sampling bias on parasite collections in Africa (fig. S8).

We discourage the unqualified interpretation of our results as an estimate of the underlying global patterns of parasite diversity. Compared to the 457 species of all parasites included in our study, there are an estimated 300,000 helminths alone, many to most of which have yet to be described by systematists. In future work, a reasonable assessment of parasite biodiversity hot spots might not be impossible, but it requires two major shifts in parasite open data. First, researchers must begin the process of georeferencing the major museum parasitology collections, including the full 70,000+ records of the USNPC (which we have georeferenced halfway in this study and intend to complete by Fall 2017) and, more significantly, the 200,000+ records of the London Museum of Natural History (a more challenging task, because this analysis will require retrieving geospatial information from each published paper in the database—hence why that analysis was not conducted during the time frame of this study). Second, targeted long-term work to profile the entire parasitic diversity of small, regional ecosystems is critically needed, especially in high-biodiversity systems that are proportionally neglected in our data set. Included in such systems are highly diverse regions like the Western Cape Province of...
South Africa, where parasitology work has been limited but is likely to discover high levels of uncataloged diversity (111), and conventional biodiversity hot spots like the Amazon basin that are hot spots of parasitological work and also assumed to be hot spots of parasite biodiversity, but where data are still too limited (see fig. S5) (112, 113). Finally, expanding all these analyses to freshwater systems and more significantly to the tremendous diversity of oceanic parasites (particularly for the speciose Cestoda, of elasmobranchs especially, and for unique specialists like Ozobranchus turtle leeches and cyamid whale lice that may be of special conservation interest) is a critical step forward.

Open data, model presentation, and PEARL
We present the final and forecasted distribution models for every species on an online server under the working title The PEARL version 0.1 (pearl.berkeley.edu; fig. S2). In the process of doing so, we set a new precedent for data storage in studies that do mass modeling of species range shifts, by making available a set of honest quantitative and qualitative metrics of data quality and model confidence:

Data coverage: Sample size by species is broken down into four quantities and classified as “weak” (0 to 25%; 0 to 28 points), “fair” (25 to 50%; 29 to 42 points), “good” (50 to 75%; 43 to 80 points), and “excellent” (75 to 100%; 81 to 3289 points).

Data uncertainty: The uncertainty radius (in meters) attached to each manually georeferenced point is available in the published form of the data set, but many species include a combination of data from manually georeferenced sources (the USNPC and FeatherMites) and nongeoreferenced sources (primarily GBIF). We set the uncertainty radius of nongeoreferenced sources as zero and then classified species’ average uncertainty radius across zero and nonzero values into five categories: “perfect” (0), “excellent” (0 to 5.1 km), “good” (5.1 to 6.5 km), “fair” (6.5 to 7.9 km), and “weak” (7.9 to 21.3 km).

Model predictive accuracy: The TSS (range, from −1 to +1, where 0 is a total lack of predictive power) and AUC (range, from 0 to 1, with predictions greater than 0.5 considered better than random chance) for each final, weighted ensemble model.

Although the last category is generally considered standard for publication-quality niche models, very few studies publish as many niche models simultaneously as this one, and presentation of results on online servers seldom includes these technical elements of model evaluation alongside the actual mapping efforts. With the quality of parasitological collections’ geospatial information, this could not be more important to an honest and open scientific method. In niche modeling studies for poorly documented clades like most parasitic groups, models can only be as good as the underlying data; rather than letting incomplete data preclude analysis (and thereby slow down parasite conservation and control efforts by at least 5 to 10 years), we instead recommend future work expanding our assessment’s focus on similarly engaging in honest post hoc evaluation of model accuracy and quality and recommend that this information be made available to the public. Doing so further helps highlight which species require the most thorough reassessment in future incarnations of our tentative “Red List.”

We also use our models to forecast the conservation status of each parasite in our study, using IUCN criteria to classify species into categories ranging from “least concern” to “critically endangered” based on average habitat loss rates. We adopt the breakdowns used by Thomas et al. (6) in terms of net habitat loss but drop the criteria involving absolute area, because a suitable area for a parasite likely operates on very different dimensions compared to free-living species, and their persistence via transmission is even more strongly linked to landscape connectivity. All our IUCN categorizations are also available at pearl.berkeley.edu. Expansion of PEARL beyond the 457 pilot species will follow the completion of the USNPC georeferencing project in late 2017, enabling higher-accuracy biodiversity mapping and parasite conservation planning in collaboration with other parasitology laboratories.

SUPPLEMENTARY MATERIALS
Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/3/9/e1602422/DC1

text S1. Primary, secondary, and compounded extinction rates.
text S2. Canonical and MaxEnt approaches to the SAR.
text S3. A more restrictive analysis based on 50+ point-per-species models.
table S1. Host specificity and parasite species richness.
table S2. Thomas best-case scenario coextinction rates.
table S3. Thomas worst-case scenario coextinction rates.
table S4. Thomas total coextinction rates.
table S7. Urban total coextinction rates.
table S8. Cumulative extinctions with dispersal.
table S9. Cumulative extinctions with dispersal.
table S10. 50+ points subanalisis-based habitat loss rates.
table S11. Source data (Excel).
table S12. Source data (Excel).
table S13. Source data (Excel).
fig. S1. Final data set breakdown by source and clade.
fig. S2. Example presentation of species distribution and conservation status on PEARL.
fig. S3. Loss of native habitat broken down by RCP and GCM.
fig. S4. Trade-offs between biodiversity loss and emergence across parasite clades.
fig. S5. Sources and distribution of occurrence data.
fig. S6. Loss of native habitat broken down by feature classes and regularization multiplier.
fig. S7. Visualizing spatial bias in species richness gradients.
fig. S8. parasite richness gradients by human health concern.

REFERENCES AND NOTES

10 of 12

11 of 12
the feather mite project and G.S.C. collected African tick data. C.J.C., K.R.B., C.F.C., E.R.D., and O.C.M. contributed to model design. C.J.C. ran all models and led the data analysis. V.M.B. compiled human infection data for parasites included in the final data set. O.C.M. and C.J.C. conceived the PEARL database, and O.C.M. and G.C. created the Web infrastructure. All authors contributed to the writing of the manuscript and approved the final submission draft.

**Competing interests:** The authors declare that they have no competing interests.

**Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Parasite biodiversity faces extinction and redistribution in a changing climate


Sci Adv 3 (9), e1602422.
DOI: 10.1126/sciadv.1602422