Predicting the distribution of a parasite using the ecological niche model, GARP

Terry R. Haverkost¹, Scott L. Gardner¹* and A. Townsend Peterson²

¹Harold W. Manter Laboratory of Parasitology, University of Nebraska-Lincoln, W529 Nebraska Hall, Lincoln, NE 68588-0514, USA.
²Natural History Museum and Biodiversity Research Center, The University of Kansas, Lawrence, KS 66045, USA.
*Correspondent: slg@unl.edu

Abstract. The ecological niche of a parasite exists only at the nexus of certain abiotic and biotic conditions suitable for both the definitive and intermediate hosts. However, the life cycles of most parasites are not known, or are poorly known, and using known ranges of hosts to find endemic parasitic infections has been difficult. However, with ecological niche modeling, we can create potential range maps using known localities of infection. Testing the validity of such maps requires knowledge of the localities of other parasites with common history. Here, we find that the ecological niche of a tapeworm parasite of voles, Paranoplocephala macrocephala (Cestoda: Anoplocephalidae), allows prediction of the presence (in ecological and geographic space) of 19 related parasite species from 3 genera in 23 different hosts throughout the Nearctic. These results give credence to the idea that this group shares similar life cycle requirements despite phylogenetic distance. This work further validates ecological niche modeling as a means by which to predict occurrence of parasites when not all facets of the life cycle are confirmed. Such inductive methods create the opportunity for deducing potential reservoir or intermediate hosts, and complementing studies of parasite biodiversity and community ecology.

Key words: phylogenetics, ecology, geographical distribution, cestodes, Paranoplocephala, biodiversity.

Introduction

The most thorough procedure for accurately defining the geographic range of a parasite would require 2 main pieces of information. Researchers would first have to know the realized and potential intermediate and definitive hosts used by the parasite. Researchers would also need a phylogeny including the parasite of interest and its relatives. The phylogeny would be used to identify closely-related congeneres and compare the known distribution and host usage of those congeneres to get a better picture of the occurrence and potential host usage of the original target species. While logical and intuitive, these pieces of information are rarely available for any given parasite. Studies aimed at elucidating intermediate hosts of parasites can take decades (e.g. Zelmer and Esch, 1998), and studies show that parasites utilize hosts with similar trophic tendencies despite their phylogeny (Hoberg, 1996) leading to incongruent host vs. parasite phylogenies (Kimura et al., 2006).

Ecological Niche Models (ENMs) have become a powerful tool that can be used to understand the potential distribution of parasites and diseases when not all facets of that disease are known (Peterson et al., 2002; Peterson et al., 2004). This approach is enticing to parasitologists since researchers can get a broad generalization of a parasite’s
geographic distribution without knowing the complexities of the parasite’s transmission dynamics. However, a map of the parasite’s distribution would allow researchers to make more accurate predictions of potential intermediate or reservoir hosts (Peterson et al., 2002; Peterson et al., 2007), if one of these hosts is not known.

The purpose of the present study was to create a potential distribution of Paranoplocephala macrocephala, (Douthitt, 1915) an anoplocephalid cestode, and use the resulting map to answer 2 separate questions. We were interested in 1) how well the modeled distribution could predict the presence of related species and 2) how well that same distribution could predict the presence of P. macrocephala in unsampled areas. We chose GARP as the ENM for this study, as it is well suited to answer both questions. Peterson et al. (1999) showed that predicted distributions made by the GARP algorithm of 1 species could accurately predict the presence of its sister species. Also, in a comparison of GARP and Maxent, Peterson et al. (2007a) showed that GARP outperformed Maxent when predicting species’ distributions into broad unsampled areas.

Materials and methods

The parasite, Paranoplocephala macrocephala is an anoplocephalid cestode. The cestode family Anoplocephalidae has a large number of described species, and includes forms that have no hooks on the scolex (holdfast), no apical organ or rostellum, a saccate or reticular uterus, and proglottids (segments) that are generally much wider than long. Species classified in this family occur in a wide variety of reptiles, birds, and mammals, with most species occurring in mammals. Stunkard (1937) was the first to work out the complete life cycle for a species in this family; it is now widely accepted that many members of the Anoplocephalinae use free-living grass mites (Oribatoidea) as intermediate hosts (Kates and Runkel 1948; Freeman, 1952). Paranoplocephala macrocephala has a Nearctic distribution, and is common in small intestines of voles (Rodentia: Arvicolidae) throughout their ranges (Spasskii, 1951; Haukisalmi and Henttonen, 2003).

The Ecological Niche Model. We used the Genetic Algorithm for Rule-set Production (GARP) model (Stockwell and Peters, 1999) to create the ENM for this analysis. Software is available free from the DesktopGARP website (http://www.lifemapper.org/desktopgarp). This method has seen extensive testing and validation (Stockwell and Peterson, 2002, 2003). Although GARP did not rank highly among methods for distribution modeling (Elith et al., 2006), these results have been seen to depend to an unknown degree on methodological artifact (Peterson et al., 2008). Regardless, the challenge of estimating ecological niches is distinct, and GARP has seen considerable success in niche modeling applications (Peterson, 2003).

To create a predictive model within GARP processing, available unique locality points are divided randomly and evenly into groups for model training and testing. The program uses a set of methods (logistic regression, bioclimatic rules, atomic rules, range rules) to make initial hypotheses of non-random associations between the training locality points, a population of randomly generated localities where the species has not been recorded, and the environmental layers provided. These rules are selected, evaluated, tested, and incorporated into or rejected from the final model through an iterative process that resembles chromosomal evolution (crossing over, additions, deletions, etc.): hence the term “genetic algorithm”. The testing points are used to assess model accuracy via comparing the percentage of points correctly predicted present or absent by the current rule. The change in predictive accuracy determines whether a particular rule gets incorporated into the final rule set. The algorithm either completes 1000 iterations or runs until further iterations no longer increase predictive accuracy. The final rule set, defined in ecological dimensions, is projected back into geographic space to produce a map.

Because of the random-walk nature of the GARP algorithm, results of replicate runs based on the same input data vary, if subtly. As a result, we filtered the replicate solutions based on their error characteristics to create a best-subset model following Anderson et al. (2003). In particular, spatial predictive models can show error of 2 types: omission (prediction of absence at sites where the species is present) and commission (prediction of presence at sites where the species is believed absent). Because the 2 error components differ in important ways (i.e., commission error is much less serious than omission error), the best-subsets approach emphasizes minimization of omission error. Hence, from 400 random replicate models created by the program, we selected the 20 that showed no omission error as measured based on independent testing data; from this group, we selected the 10 models that deviated least from the median commission index (area predicted as suitable). We summed these 10 models pixel by pixel to produce a final prediction of the potential distribution of P. macrocephala. We present this prediction as a lowest presence threshold map (Pearson et al., 2007); a conservative estimate of the minimum predicted area for P. macrocephala with no omission error.

Environmental data sets used included 14 layers: topographic data (elevation, slope, aspect, topographic
index, flow accumulation, and flow direction) from the United States Geological Survey’s Hydro-1K data set (http://edcdaac.usgs.gov/gtopo30/hydro/) and climate data (annual mean of daily temperature range, frost days, wet days, vapor pressure, precipitation, and maximum, minimum, and mean temperatures for 1961-1990) from New et al. (1999). All environmental data sets were resampled to 0.1° spatial resolution for analysis to match the approximate spatial precision of the occurrence data available.

To obtain occurrence data for *P. macrocephala*, we collected the geographic information associated with published literature of this species over the past 100 years (Erickson, 1938; Rankin, 1945; Rausch and Tiner, 1949; Hansen, 1950; Kuns and Rausch, 1950; Hall and Sonnenberg, 1955; Leiby, 1961; Kinsella, 1967) and searched specimen databases associated with the U. S. National Parasite Collection (USNPC) in Beltsville, Maryland, and the Harold W. Manter Laboratory of Parasitology (HWML) at the University of Nebraska - Lincoln. Synonymous names were rectified based on Rausch (1976). Specimens or records without specific latitude/longitude coordinates were geo-referenced to the nearest 0.01° using the GeoLocate geo-referencing program (Tulane University, Belle Chasse, Louisiana). We found 33 references to this species, 23 of which were spatially unique.

**Predicting potential distribution of related species.** Peterson et al. (1999) found that GARP was able predict distributions of sister species of their target species, and that this predictive ability was inversely proportional to time since divergence. We performed a similar test using museum records with locality data for closely related congeneric and confamilial cestodes from across North America. Wickström et al. (2005) found strong support for a close phylogenetic relationship among species of *Anoplocephaloides*, *Paranoplocephala*, and *Diandrya*. They called this assemblage the ‘arvicoline clade’ because of the common association with an arvicoline (vole) definitive host. Their phylogeny is simplified and redrawn as our Figure 1. As such, we gathered all occurrence data available for all species of *Anoplocephaloides*, *Diandrya*, and *Paranoplocephala* from the literature (see Rausch 1946, Hansen 1947, and Kamiya et al. 1979, in addition to the references listed above) and the USNPC and HWML databases. Entries with missing spatial data were geo-referenced as described above. In total, we found 83 spatially-unique localities of 24 parasite species occurring in 29 different definitive host species. Those parasite and host species are listed in Table 1.

![Figure 1. Reproduction of the phylogeny of the ‘arvicoline clade’ from Wickström et al., (2005). Only species included in the present study were redrawn in the phylogeny.](image)

To validate the ability of the model to predict the spatial arrangement of these new points, we used a one-tailed chi-square test to compare observed frequencies of successful prediction of test points across the range of thresholds of model agreement (0-10) with expectations if points and predictions were to be assorted at random with respect to one another. Specifically, we calculated null expectations as the proportion of the area predicted present multiplied by the total number of testing points. Observed and expected values were compared using a chi-square statistic with 1 degree of freedom.

**Results**

Each of the 10 models that were summed to create the best-subset model predicted independent testing points more accurately than would be expected were testing points and predictions to be at random with respect to one another (all *P* < 0.001). The best-subset model presented (Fig. 2) shows the lowest presence threshold of models (6) needed to incorporate all locality points for *P. macrocephala* without omission error. As such, we are confident that this ENM has significant predictive power regarding the geographic distribution of *P. macrocephala*.

When the *P. macrocephala* ENM is compared with localities of the additional parasite species, the best-subset model was able to predict the spatial arrangement
Table 1. Parasites used in this study and their associated hosts. Hosts numbered 1-28, parasites listed as being found in host with corresponding number. Parasite names are derived from the taxonomy of Rausch (1976); mammal nomenclature follows Wilson and Reeder (2005). Parasites and hosts only identified to genus were included in the study but not in this table.

<table>
<thead>
<tr>
<th>Parasite species</th>
<th>found in host(s)</th>
<th>host species</th>
<th>host ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andrya neotomae</td>
<td>21</td>
<td>Arborimus albipes</td>
<td>1</td>
</tr>
<tr>
<td>Anoplocephaloides cf. dentata</td>
<td>12</td>
<td>Arborimus longicaudus</td>
<td>2</td>
</tr>
<tr>
<td>Anoplocephaloides infrequens</td>
<td>5,17,20</td>
<td>Dicrostonyx groenlandicus</td>
<td>3</td>
</tr>
<tr>
<td>Anoplocephaloides kontrimavichusi</td>
<td>26</td>
<td>Geomys breviceps</td>
<td>4</td>
</tr>
<tr>
<td>Anoplocephaloides lemmi</td>
<td>7</td>
<td>Geomys bursarius</td>
<td>5</td>
</tr>
<tr>
<td>Anoplocephaloides romerolagi</td>
<td>22</td>
<td>Gaumomys sabrinus</td>
<td>6</td>
</tr>
<tr>
<td>Anoplocephaloides septentrionalis</td>
<td>16,17</td>
<td>Lemmus trimucronatus</td>
<td>7</td>
</tr>
<tr>
<td>Anoplocephaloides troeschi</td>
<td>13,17</td>
<td>Marmota caligata</td>
<td>8</td>
</tr>
<tr>
<td>Anoplocephaloides variabilis</td>
<td>5,28</td>
<td>Marmota flaviventris</td>
<td>9</td>
</tr>
<tr>
<td>Anoplocephaloides wigginsi</td>
<td>25</td>
<td>Microtus californicus</td>
<td>10</td>
</tr>
<tr>
<td>Diandrya composita</td>
<td>19,8,9</td>
<td>Microtus chrotorhinus</td>
<td>11</td>
</tr>
<tr>
<td>Paranoplocephala arctica</td>
<td>3</td>
<td>Microtus miurus</td>
<td>12</td>
</tr>
<tr>
<td>Paranoplocephala bairdi</td>
<td>1,2,11</td>
<td>Microtus mognolonensis</td>
<td>13</td>
</tr>
<tr>
<td>Paranoplocephala batzii</td>
<td>12</td>
<td>Microtus montanus</td>
<td>14</td>
</tr>
<tr>
<td>Paranoplocephala borealis</td>
<td>15</td>
<td>Microtus ochrogaster</td>
<td>15</td>
</tr>
<tr>
<td>Paranoplocephala etholeni</td>
<td>17</td>
<td>Microtus oeconomus</td>
<td>16</td>
</tr>
<tr>
<td>Paranoplocephala freeman</td>
<td>18,20,21</td>
<td>Microtus pennsylvanicus</td>
<td>17</td>
</tr>
<tr>
<td>Paranoplocephala kirbyi</td>
<td>10</td>
<td>Microtus xanthognathus</td>
<td>18</td>
</tr>
<tr>
<td>Paranoplocephala krebsi</td>
<td>3</td>
<td>Myodes gapperi</td>
<td>19</td>
</tr>
<tr>
<td>Paranoplocephala omphalodes</td>
<td>12</td>
<td>Neotoma cinerea</td>
<td>20</td>
</tr>
<tr>
<td>Paranoplocephala primordialis</td>
<td>12,14,23,24</td>
<td>Neotoma fuscipes</td>
<td>21</td>
</tr>
<tr>
<td>Paranoplocephala sciuri</td>
<td>6</td>
<td>Ondatra zibethicus</td>
<td>21</td>
</tr>
<tr>
<td>Paranoplocephala serrata</td>
<td>3</td>
<td>Romerolagus diazi</td>
<td>22</td>
</tr>
<tr>
<td>Paranoplocephala troeschi</td>
<td>17</td>
<td>Sciurus hudsonius</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sigmodon hispidus</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spermophilus undulatus</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Synaptomys borealis</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tamiasciurus hudsonius</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thomomys talpoides</td>
<td>28</td>
</tr>
</tbody>
</table>

Our results indicate that the ecological niche of *P. macrocephala* is similar to the ecological niches of 19 other parasite species included in the Anoplocephalidae. Put another way, the ecological conditions suitable for completion of the life cycle of *P. macrocephala* are variable, but overlap the conditions observed among the other species tested broadly (Fig. 3). This is not surprising, considering the proposed evolutionary relationships among microtine rodents (Conroy and Cook, 1999) and, similarly, among “arvicoline” cestodes (Wickström et al., 2005). The lack of resolution, i.e., polytomies, observed in the anoplocephalid phylogeny found by Wickström et al. (2005) (Fig. 1) appears to mirror the ‘pulses of speciation’ suggested for the Arvicoline by Conroy and Cook (1999) and the 2 microtine invasions of North America from Asia around 1.3 million years ago (Conroy and Cook,
The rapid diversification of *Microtus* during their Pleistocene invasion of the Nearctic region would be consistent with little time and opportunity for evolution of diverse ecological niches among their parasites, which likely accompanied the voles as they crossed Beringia.

Omission error among the other parasite species included in our test occurred mainly along the extremes of the climatic parameters in Figure 3. Ten of the points omitted correspond to cold and dry conditions above the Arctic Circle (~66° 30’ N latitude), where the possibility of anoplocephalid life cycle completion on the surface might seem unlikely, given the extreme environment during winter months. At these sites, however, cestodes were found in lemmings (*Dicrostonyx* sp. and *Lemmus* sp.), arvicoline rodents known to maintain subnival burrow systems in winter months (Wooding, 1982). The burrow systems allow these rodents to exist under the snow in a microclimate suitable for the intermediate host, offering opportunities for maintenance of the life cycle of the parasites. Since the maintenance of the life cycle in this area probably does not occur above ground during winter months, our predictive model based on broad scale climatic variables would be less likely to be able to predict their occurrence. The model also does not predict appropriate conditions to occur in the southeastern United States. However, it is known that confamilial species of *Cittotaenia* occur in cottontail rabbits (*Sylvilagus* spp.) in Florida, Georgia, and Alabama, and that this parasite also utilizes oribatid mites as intermediate hosts (Stunkard, 1941). As such, it would seem that the geographic extent of suitable habitat for this life cycle (as defined by *P. macrocephala*) may be limited by the definitive and not the intermediate host. This conclusion is supported by work on the historical ecology of other parasite taxa (Hoberg, 1996) and from the life cycle experiments that show that an anoplocephalid cestode not included in this study, *Moniezia expansa*, can utilize 73 species of oribatid mites as intermediate hosts and an oribatid mite, *Scheloribates laevigatus*, can serve as an intermediate host to 14 species of anoplocephalids (Denegri, 1993).

Our model predicted that the anoplocephalid life cycle would be present in 2 areas where, to our knowledge, there have been very few published reports of surveys for small-
mammal parasites in decades. The first area predicted as suitable for the transmission of anoplocephalids is along the eastern edge of the Sierra Madre Mountains in northwest Mexico. To our knowledge there are no records of anoplocephalids along this mountain range. In the area predicted to have anoplocephalids present just south of Mexico City, Kamiya et al. (1979) reported on the presence of *Anoplocephaloides romerolagi* in the volcano rabbit, *Romerolagus diazi*. These results support the use of GARP when determining likely areas for targeted sampling of anoplocephalid cestodes during biodiversity surveys.

It has been suggested that *P. macrocephala* may actually be a complex of species, although this complex has not yet been separated and will likely require molecular data to understand the true species limits (Haukisalmi and Henttonen, 2003). If such is the case, the model may overestimate the distributional area of any one species. The likelihood of an overestimation of the niche does not seem high, since the present study shows conserved life cycle characteristics across the subfamily.

Parasites offer a unique problem in niche modeling as parasites from highly vagile hosts may be collected in areas far from the areas in which the parasite infection was obtained (see Hobberg, 1996). We feel that this element does not greatly affect the current study, as movement by hosts in this study is unlikely to exceed the spatial resolution of the data set (0.1° pixels), with any movement between adjacent areas having minimal impact on the coarse environmental variables used. Consideration of this point must be made when making generalizations as to the geographic extent of parasite transmission.

Of the 2 methods outlined by Peterson (2006) for analyzing disease transmission, the present study treats the complex species interactions leading to parasite transmission as a ‘black box,’ (inductive approach). The alternative (deductive approach)—modeling each host in the life cycle separately—is not feasible in this case since the geographic range of the intermediate host (or hosts) of *P. macrocephala* are not well characterized. The deductive approach would also not consider incidental transmission of the parasite among suitable, but rare, hosts. Both systems may have merit, depending on the parasite species of interest in the study. For example, the inductive approach may be enlightening when facets of a life cycle are unknown or the parasite is known to show plasticity in its life cycle, e.g., utilizing multiple definitive and intermediate hosts, truncating or bypassing life cycle stages. The deductive approach would be beneficial if that parasite is known to show high levels of host specificity (such as pocket gophers and their ectoparasitic lice, see Hafner et al., 2003).

The creation of such inductive maps in the future may lead to a better understanding of the biogeography of parasites and other host/symbiont associations. Combination of maps across multiple species may prove useful when analyzing geographic patterns of parasite community ecology. Since detailed investigations of ecological characteristics of parasite life cycles of non-human hosts are rare, the inductive approach may prove more fruitful. For many groups or species of parasites, comprehensive life cycle studies, taxonomic and systematic studies, and rigorous field sampling are incomplete or lacking. Although this information is required for creating reliable range maps of parasites, the expertise required to evaluate the quality of these studies critically is receding. However, biogeographic studies of free-living animals and plants are relying less on expertly derived maps and more on existing museum records, bioinformatics, and GIS technologies to provide information relative to conservation of biodiversity. Thus far, parasitic taxa have been largely excluded from these analyses, even though they constitute a dominating presence in the environment (Lafferty et al., 2006). We present this suite of analyses as an exploration of the potential of ecological niche modeling of parasites, to capture their distributions and perhaps contribute more prominently to evaluations of biodiversity.

**Acknowledgments**

Thanks to Dave Tinnin and Gabor Racz for assistance.
during this project. This work was partially supported by grants (DEB-0717214, BSR-9024816, DEB-0097019, and DBI-0646356) from the National Science Foundation to S.L. Gardner.

Literature cited


Peterson, A. T., M. Papes and M. Eaton. 2007a. Transferability


