# Charisma counts: the presence of great apes affects the allocation of research effort in the paleotropics

Andrew J Marshall[1*], Erik Meijaard[2], Eric Van Cleave[3], and Douglas Sheil[4]

Scientific research in the biodiversity-rich paleotropics (the tropical portions of the Old World) provides multiple benefits, yet we know little about the distribution of research effort in this region and the factors that determine it. We used Google Scholar to assess paleotropical research effort – defined here as the number of published studies that reference one of 565 terrestrial protected areas in the 23 African and Asian countries that contain great apes (*Gorilla* spp, *Pan* spp, and *Pongo* spp). We found that research effort is strongly skewed toward a small number of sites and that most protected areas are inadequately represented in the scientific literature. Scientists conducting paleotropical research often focus their efforts on protected areas that are large, that are designated as national parks, and that contain non-human great apes. Our results highlight important gaps in research effort, and indicate that current understanding of tropical protected areas is limited and heavily biased toward specific sites, many of which may not be representative of existing protected areas in the paleotropics.

Please see Erratum attached at end of pdf

Although the tropics contain most of the world's biodiversity (Ceballos and Ehrlich 2006; Kreft and Jetz 2007), and despite much of this diversity being at risk (Grenyer *et al.* 2006; Bradshaw *et al.* 2009), our understanding of tropical regions lags behind that of other areas of the world (Stuart *et al.* 2004; Butchart and Bird 2010). Our knowledge of tropical species and their distributions is limited for many taxonomic groups (Platnick 1991; Collen *et al.* 2008; Schipper *et al.* 2008), and we lack the information necessary to protect and manage many vulnerable species (Sitas *et al.* 2009; Meijaard *et al.* 2012). Although more research is clearly needed in the tropics, important progress has been made. In addition to providing broadly applicable information (eg on ecological processes, species occurrences), scientific research can provide direct benefits to the locations where it is conducted (Wrangham 2008). Consider protected areas, widely perceived as bastions for the preservation of terrestrial tropical biodiversity (Bradshaw *et al.* 2009; Laurance *et al.* 2012): researchers contribute by informing protected area management and policy, deterring illegal activities, participating in education and awareness campaigns, building capacity, and providing alternative forms of income for local people (Campbell *et al.* 2011; Meijaard *et al.* 2012; Laurance 2013).

Despite the broad importance of scientific research and the potential benefits of researcher presence, little is known about the distribution of research effort across tropical protected areas or the factors that determine it. Identifying patterns in the current distribution of scientific research could valuably inform both the allocation of future research efforts and the interpretation of existing knowledge about tropical protected areas (Kier *et al.* 2005; Martin *et al.* 2012). For instance, if most research is concentrated in a handful of protected areas that differ substantially from other locations, then the majority of published scientific studies might apply only to a relatively small subset of areas and extrapolation beyond those areas might be invalid or of limited value.

We assessed research effort in protected areas in tropical Asia and Africa using systematic searches of Google Scholar (http://scholar.google.com). Analysis of online patterns and trends has become a valuable tool for gauging interest, investment, and the actions of scientists and the lay public in environmental and conservation issues (Sitas *et al.* 2009; Mccallum and Bury 2013; Meijaard *et al.* 2015). We hypothesized that research effort would be greater in protected areas that were larger (hypothesis $H_1$), older ($H_2$), and gazetted (ie officially designated) as national parks ($H_3$), as opposed to receiving an International Union for

[1]*Department of Anthropology, Program in the Environment, and School for Natural Resources and Environment, University of Michigan, Ann Arbor, MI* [*](ajmarsha@umich.edu); [2]*Borneo Futures, Tangerang Selatan, Indonesia;* [3]*Animal Behavior Graduate Group, University of California, Davis, Davis, CA;* [4]*Department of Ecology and Natural Resource Management (INA), Norwegian University of Life Sciences (NMBU), Ås, Norway*

Conservation of Nature (IUCN) designation mandating a lower level of protection. We also hypothesized that the presence of charismatic species would increase research interest ($H_4$). As a test case to examine the role of charismatic taxa, we used non-human great apes (hereafter "great apes") because they are iconic and threatened, and because their distribution is relatively well known (Caldecott and Miles 2005; Junker *et al.* 2012).

## Methods

### Information on protected areas

We compiled a list of all terrestrial protected areas ($n = 565$) – in the 21 African countries (Angola, Burundi, Cameroon, Central African Republic, Republic of the Congo, Democratic Republic of the Congo, Equatorial Guinea, Gabon, Ghana, Guinea, Guinea Bissau, Côte d'Ivoire, Liberia, Mali, Nigeria, Rwanda, Senegal, Sierra Leone, Sudan [which includes North and South Sudan], Tanzania, and Uganda) and two Asian countries (Indonesia and Malaysia) in which great apes are found (Caldecott and Miles 2005) – from the World Database of Protected Areas (WDPA; www. protectedplanet.net) and other online and print databases. We gathered data on protected area size (in square kilometers), IUCN management category (eg National Park, Species Management Area; www.iucn. org/about/work/programmes/gpap_home/gpap_quality/ gpap_pacategories), and the year the protected area was gazetted in its current designation from the WDPA. When data were unavailable from the WDPA database, we used data from published scientific papers or reports by governmental and non-governmental organizations.

### Determining great ape presence

We divided great apes into four groups: gorillas (*Gorilla* spp), chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*), and orangutans (*Pongo* spp). We determined the presence of great apes in each protected area using a variety of published sources (eg Caldecott and Miles 2005), personal knowledge, and correspondence with experienced scientists. If great ape presence could not be positively confirmed within a given area, we scored them as absent. We acknowledge that there are uncertainties about past and present great ape distribution, and that some great apes range in and out of protected areas, ensuring that presence/absence data will be imperfect. Nevertheless, we were able to confirm the presence/absence of great apes from multiple sources for most of the protected areas. We believe remaining errors are few and unlikely to influence the results presented here, especially as such uncertain presences are unlikely to have a major influence on researchers.

### Assessing research effort

We used the number of returned citations (hereafter "hits") on Google Scholar as a proxy for research effort in each protected area. As compared with other popular databases of scholarly work (eg the Institute for Scientific Information [ISI] Web of Science), Google Scholar returns a wider array of citable material, including conference proceedings, non-ISI-indexed journals, and studies published in international journals that are written in languages other than English (Meho and Yang 2007). We used an exact phrase search operator to return only works in which the title or text contained the full, official WDPA name; references to different or alternative spellings of protected area names therefore would not have been recorded. Although this approach almost certainly reduced the number of hits for some areas, any error introduced is likely to be random with respect to great ape presence or protected area designation and therefore is unlikely to have biased our results. In addition, because the full texts of some older (ie pre-1980) articles are probably inaccessible to the Google Scholar web crawlers, we may have underestimated the number of publications for older protected areas, but we expect this error to also be random with respect to great ape presence or protected area designation.

We conducted all searches within a 2-day period (1st and 2nd of May 2013) from Davis, California. We spot-checked the results by using multiple computers and multiple web browsers in Davis, as well as by repeating the searches from computers in Jakarta and Ketapang, both in Indonesia. Hit counts were virtually identical in all cases. To test the stability of the Google Scholar results, we compared the data from May 2013 to data collected on all parks using the same methods in June 2012 and found them to be highly correlated ($r = 0.996$), with a slope $>1$, as expected when hit counts increase over time.

### Statistical modeling

Because the distribution of hit counts was highly skewed, we modeled them using negative binomial regression with the glmmADMB package version 11.2 (Fournier *et al.* 2012), in R version 3.1.2 (R Development Core Team 2014). We constructed a series of negative binomial models to represent all plausible combinations of predictor variables and compared these using Akaike's Information Criterion (AIC) and Akaike model weights (Burnham and Anderson 2002). Our full dataset ($D_{FULL}$) contained information on 565 protected areas, although data on protected area size ($D_{SIZE}$, $N_{SITES} = 426$) were available for only a subset of these sites. For a smaller subset of protected areas ($D_{SIZE \cdot AGE}$, $N_{SITES} = 321$), we obtained data on both the size and age (defined as 2013 minus the year of gazetting at its current status). Because different amounts of information were available

           

for these three datasets, each dataset was analyzed separately. For $D_{SIZE \bullet AGE}$, we built 36 models, which included combinations of variables on protected area size, age, IUCN status, presence/absence of great apes, number of great ape taxa, great ape taxon present, and a random effect for country (given that there are substantial differences among countries in terms of the ease and cost of access and obtaining permits, political stability, infrastructure, and scientific capacity; WebTable 1). For $D_{SIZE}$, we assessed the same models as for $D_{SIZE \bullet AGE}$, with the exception of models that included age (27 models; WebTable 2). For $D_{FULL}$, we assessed the same models as for $D_{SIZE}$, with the exception of models that included protected area size (10 models; WebTable 3).

For each dataset, we reported model-averaged (using AIC weights) β coefficients and standard errors (SE) for each predictor across the top models (ie for which summed AIC weight >0.99). Because some predictors for the effects of great apes were redundant (eg great ape presence, number of great ape taxa), we also presented weighted β coefficients ± SE averaged across only models that included these predictors (WebTable 4).

### Assessing taxonomic focus

For a random sample (20% of $D_{SIZE}$, $n = 113$) of protected areas, we used simple decision rules to assess the taxonomic focus of all returned hits based on their titles. We classified hits based on presence of specific words in the title, as follows: (1) great apes: the title contained the word(s) "ape(s)", or the common or scientific name of a great ape taxon (with the exception of titles mentioning "lesser ape(s)" or "small ape(s)", which were excluded); (2) primates: the title contained "primate(s)", or the common or scientific name of any non-great ape, non-human primate taxon; (3) mammals: the title contained "mammal(s)", or the common or scientific name of any non-primate mammal; (4) birds: the title contained "bird(s)", or the common or scientific name of any bird taxon; (5) plants: the title contained "plant(s)" or "vegetation", or the common or scientific name of any plant taxon, or the name of any plant growth form (eg "liana(s)", "vine(s)", "tree(s)"); (6) other taxa: the title contained specific reference to any non-human taxon, with the exception of mammals, birds, or plants; and (7) other: none of the above. We scored papers as fulfilling multiple categories if their titles contained multiple relevant keywords. Although Google Scholar reported the total number of hits for a search term, the maximum number of results listed for searches was 1000. For four of the sites that we randomly selected to assess content, hit counts were >1000 (1110, 1150, 1230, and 1310), and therefore not all hits were listed. For these four sites we based content analysis on the 1000 hits listed. We modeled the hit counts for each taxonomic group

separately, using the same methods applied to our main analyses.

### ◼ Results

Research effort varied non-randomly across the 565 protected areas in our sample. The 17 sites with the highest hit counts provided more than 50% (26 899 of 52 502) of total hits returned; the top five sites in Africa and Asia returned 33.8% and 44.3%, respectively, of the total hits for their region. No hits were returned for 36% ($n = 185$ of 512) of African and 21% ($n = 11$ of 53) of Asian protected areas, and 57% ($n = 295$) of African and 38% ($n = 20$) of Asian sites returned fewer than five hits (Figure 1).

### Model results

For all three datasets, a few top models (ie models for which summed AIC weight >0.99) emerged; these are listed in WebTable 4. All top models contained predictors reflecting great ape presence, protected area status, and a random effect for country.

### Testing of the four hypotheses

#### $H_1$

Protected area size appeared in all of the top models for both datasets in which it was included ($D_{SIZE}$, $D_{SIZE \bullet AGE}$). As hypothesized, the effect of protected area size was positive, although the magnitude of the effect was much reduced in models that excluded national park age (WebTable 4). Figure 2a shows the raw hit counts for $D_{SIZE}$.

#### $H_2$

We hypothesized that older parks would receive more hits; although age had a positive effect in some of the top models for the dataset in which it was included, as predicted ($D_{SIZE \bullet AGE}$; WebTable 4) the magnitude of the effect was small and the β ± SE included zero, indicating weak support for $H_2$.

#### $H_3$

IUCN category had a strong effect in all top models, as predicted. Across datasets, national parks received consistently higher numbers of hits than did other types of protected areas ($N_{NATIONALPARKS} = 152$, $N_{OTHER PAS} = 274$; Figure 2c compares the raw hit counts for $D_{SIZE}$). In the top model for $D_{SIZE}$ (m19), national parks returned 15.2 times more hits than did the other types of protected areas (a result that was consistent across top models; Figure 2d; WebTable 4).
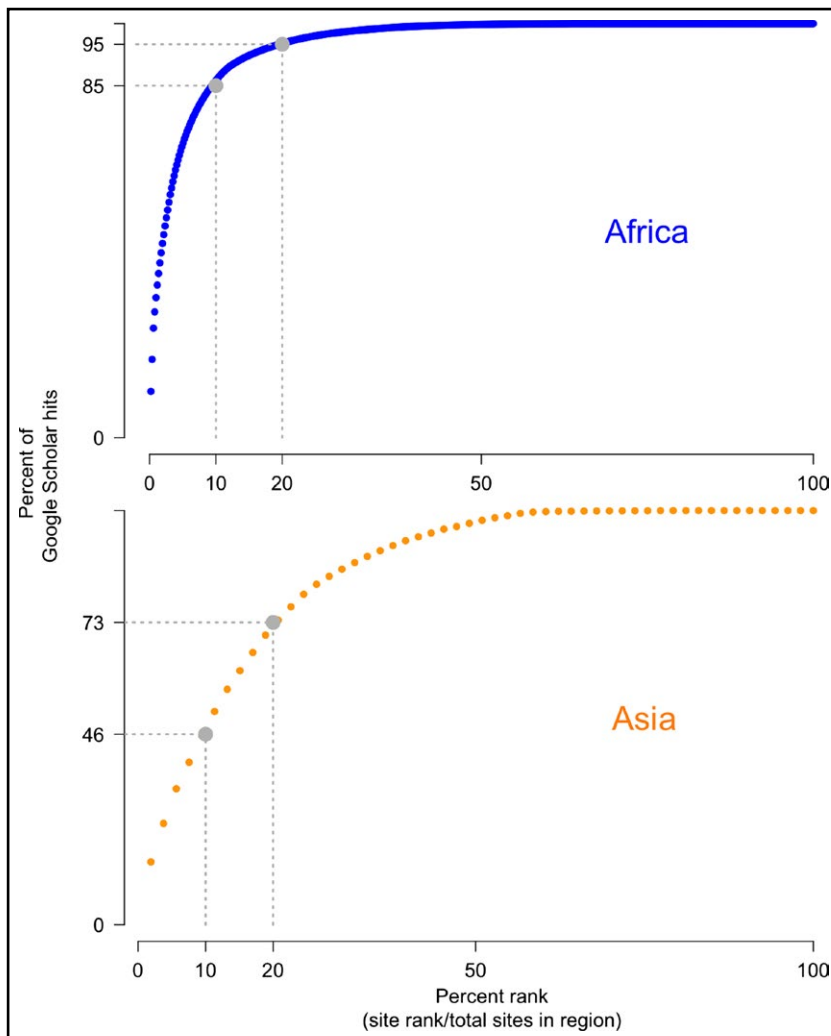
*Figure 1.* Accumulation curves depicting the percentage of total Google Scholar hits for African (n = 512 protected areas, 43 625 hits) and Asian (n = 53 protected areas, 8368 hits) sites as a function of the number of protected areas. The x-axis lists sites in decreasing order by their number of hits (ie the site with the most hits is listed farthest to the left). In Africa, the top 10% of sites account for 85% of Google Scholar hits; in Asia, the top 10% of sites account for 46% of all hits. The top 20% of sites comprise 95% and 73% of all Google Scholar hits for Africa and Asia, respectively.

$H_4$

Great ape presence had consistent and strong positive effects in all top models for all three subsets of our data (WebTable 4). For instance, for $D_{SIZE}$, sites with great apes returned 3.10 times more hits than did protected areas where they are absent ($N_{SITES \, (WITHAPES)}$ = 160, $N_{SITES \, (NOAPES)}$ = 266; see m5 in WebTable 4). The best models for each dataset included separate terms for the effects of each great ape taxon, and the magnitude and rank order of effects of specific great ape taxa were consistent across them (WebTable 4; Figure 2b depicts raw counts as a function of great ape presence for $D_{SIZE}$). In the top model for $D_{SIZE}$, the presence of gorillas had the strongest

effect: gorilla (only) presence was associated with a 35.2-fold increase in the number of hits ($N_{SITES}$ = 4 protected areas), followed by orangu-tans ($N_{SITES}$ = 17, 3.85-fold increase), chimpanzees ($N_{SITES}$ = 100, 2.34-fold increase), and bonobos ($N_{SITES}$ = 2, 1.42-fold increase; Figure 2d). Parks with both chimpanzees and gorillas ($N_{SITES}$ = 37) showed an increase that was intermediate between the effects of either species alone (3.0-fold in-crease; Figure 2d).

### Taxonomic focus of returned hits

The most common taxonomic content of returned hits was great apes (30.7% of the taxon-related hits in our random sample of sites), an unexpectedly high number given the few taxa in this group. This fact – coupled with the result that the next most common subjects were non-primate mammals and non-ape, non-human primates (23.6% and 16.8% of taxon hits, respectively) – indicates that most tropical research is focused on a very small subset of taxa: overall 71.1% of hits were related to mammals, as compared with 5.9% for birds, 11.3% for plants, and 11.7% for all other taxa (Figure 3). When we modeled hit counts for each taxon individually, the best model for each taxon contained terms for protected area size and IUCN desig-nation. The presence of great apes in a protected area was not a predictor in the top model for any other taxon. These results were consistent regardless of whether the four sites with hit counts >1000 were included or excluded from the analysis.

### Discussion

Our results show that research effort differs markedly among protected areas in the paleotropics, and that it increases with protected area size, is substantially higher in national parks than in other types of pro-tected areas, is higher in protected areas containing great apes, and varies greatly depending on which great ape taxon is present. Analysis of the taxonomic focus of all hits for a randomly selected subset of protected areas suggests that research published on great apes follows the same patterns seen in our broader analysis; research on non-ape taxa is also biased toward large national parks but is
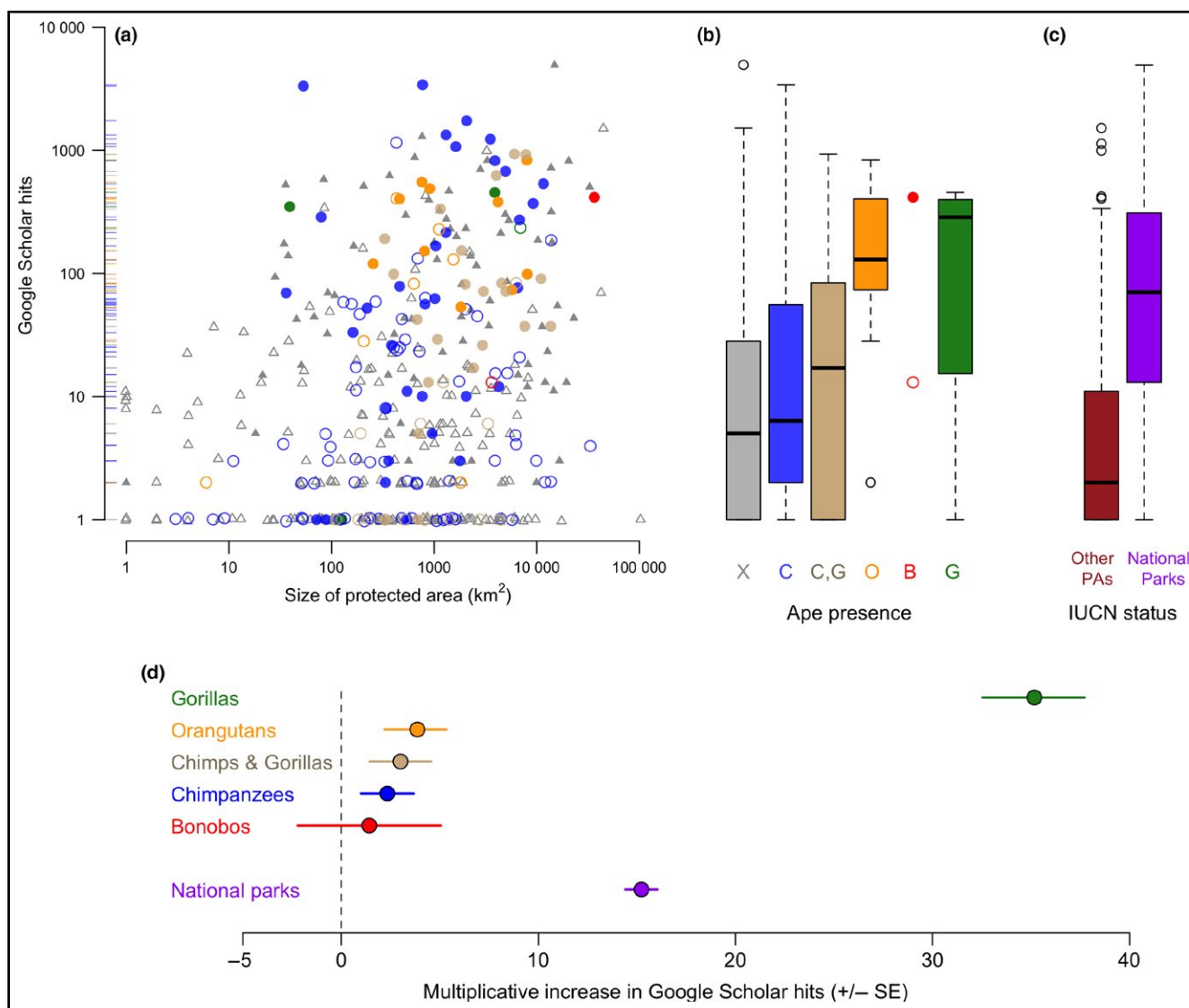
*Figure 2. Google Scholar hits as a function of protected area size, great ape presence, and protected area designation, using dataset $D_{SIZE}$. (a) Google Scholar hits as a function of protected area size (square kilometers, n = 426 protected areas). Solid symbols represent national parks; open symbols represent other protected areas. Gray triangles represent protected areas with no great apes; circles indicate protected areas containing gorillas (green), orangutans (orange), bonobos (red), chimpanzees (blue), and both chimpanzees and gorillas (brown). Tick marks to the right of the y-axis show hits for all protected areas with great apes, following the same color scheme. (b) Boxplots of Google Scholar hits, ordered from left to right by increasing median hit counts; X = no apes, C = chimpanzees, C,G = chimpanzees and gorillas, O = orangutans, B = bonobos (points are plotted as n = 2), G = gorillas. (c) Boxplots of Google Scholar hits for national parks and other protected areas (PAs). (d) Dot plots of back-transformed model β coefficients and their standard errors. Plotted values indicate the multiplicative increase in predicted number of hits for a park as compared with a non-national park protected area containing no great apes (controlling for protected area size).*

not predicted by the presence of great apes. Unexpectedly, this finding indicates that the research stations and associated infrastructure often built to support research on great apes rarely facilitate research on other taxa.

Hit counts reflect the amount of scientific information available for a given protected area, but we suggest that they can be used as a reasonable proxy for research effort more broadly. The correlation between on-the-ground research presence and research effort, as measured by Google Scholar hits, is likely to be imperfect due to potential publication biases (eg taxonomic bias, favoring reports of species presence versus absence), limited availability of older publications, and/or differences in incentives for scientific publication between countries, research teams, or scientific disciplines. Nevertheless, we are confident that these potential biases are insufficient to explain the broad, consistent effects we report here, and thus we interpret our results as a reasonable representation of the distribution of
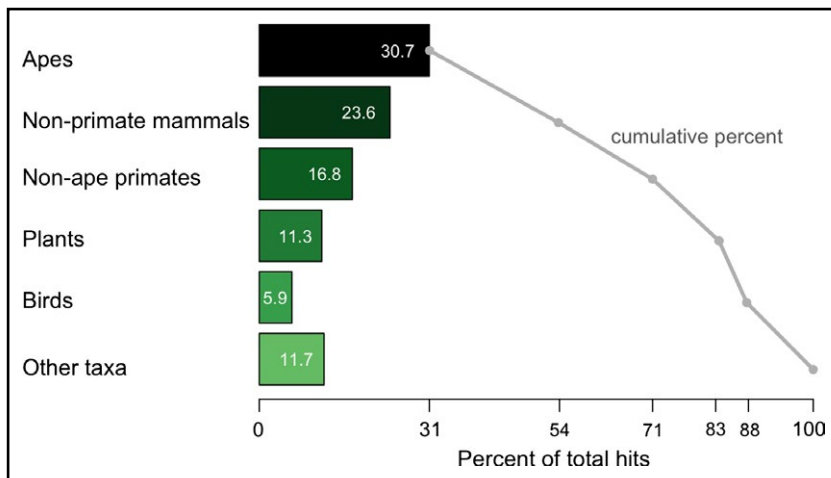
**Figure 3.** *Taxonomic content of Google Scholar hits for a randomly selected 20% sample of protected areas (n = 113 protected areas, 10 148 hits). Horizontal bars indicate the percentage of hits for each taxonomic group (eg 23.6% of hits are related to non-primate mammals). The gray line indicates the cumulative overall percentage of hits (eg mammals – apes, non-primate mammals, and non-ape primates – together comprise 71.1% of all hits).*

research effort across protected areas in tropical Asia and Africa.

We used great apes to assess the effect of the presence of a charismatic taxon, but we do not claim that the taxonomic bias we report here is confined to great apes or would apply everywhere. Other charismatic taxa will likely elicit focused interest; for instance, conservation efforts are generally biased toward large-bodied mammals (Magin *et al.* 1994; Sitas *et al.* 2009). Nevertheless, great apes represent an excellent test case, given that they often serve as a key justification for the establishment of long-term research stations and are a major focus of conservation fundraising efforts. We note also that in a small number of protected areas, the presence of great apes may be the result of researcher presence, not merely the cause (eg where research stations act as refugia for great apes; Campbell *et al.* 2011). Finally, we acknowledge that the number of sites containing bonobos and containing (only) gorillas was small, so the specific effects of these taxa should be interpreted with caution.

Our results are important for at least three related reasons. First, they indicate that the distribution of research effort in protected areas in the paleotropics is highly skewed toward a very small set of sites. Disproportionate or unequal sampling is suboptimal when trying to uncover broad patterns or trends (Martin *et al.* 2012); when interpreting broad statements about the tropics, the associated data sources should always be considered.

Second, our findings highlight major gaps in the allocation of research effort and, consequently, important knowledge gaps regarding biodiversity in tropical protected areas. Indeed, one-third of African and one-fifth of Asian protected areas did not return a single hit on Google Scholar, and the majority did not return enough

hits to fill one browser page (Figure 1). Increased archiving of "gray literature" on websites searchable by Google Scholar would surely improve the availability of information on many areas, but new research efforts targeting poorly known areas are also required. In addition to filling knowledge gaps, scientific investment at sites currently lacking a research presence could inform management and promote protection of these areas (Ahrends *et al.* 2011). The mere presence of researchers may enhance conservation at such unstudied sites (eg Campbell *et al.* 2011; Meijaard *et al.* 2012; Laurance 2013), but in order for the additional research effort to have maximal conservation benefit it would need to bridge well-known gaps between conservation science and direct, on-the-ground conservation action (Knight *et al.* 2008; Habel *et al.* 2013). For example, researchers should address questions that have concrete relevance for conservation policy and work directly with protected area managers to ensure that recommendations emerging from conservation research are implemented.

Third, our results raise the possibility that scientists' understanding of the tropics is not only limited but also biased. Most of our understanding of tropical ecology derives from in-depth and often long-term research focused on a relatively small number of well-studied sites. We tacitly assume that principles learned and processes identified in these sites characterize other areas, but this may not be so. We have shown that research effort in paleotropical protected areas is disproportionately focused on extensive national parks that support populations of a large, charismatic taxon, the great apes. These sites may differ systematically from most protected areas in other ways as well; for instance, national parks generally receive more funding and are better managed than other protected areas (McQuistan *et al.* 2006). Extrapolation from current paleotropical knowledge, which is based almost entirely on research conducted in only a handful of protected areas that are probably not representative of protected areas more generally, requires caution.

## References

Ahrends A, Burgess ND, Gereau R, *et al.* 2011. Funding begets biodiversity. *Divers Distrib* **17**: 191–200.

Bradshaw CJA, Sodhi NS, and Brook BW. 2009. Tropical turmoil – a biodiversity tragedy in progress. *Front Ecol Environ* **7**: 79–87.

Burnham KP and Anderson DR. 2002. Model selection and multimodal inference: a practical information–theoretic approach. New York, NY: Springer-Verlag.

Butchart SH and Bird JP. 2010. Data deficient birds on the IUCN Red List: what don't we know and why does it matter? *Biol Conserv* **143**: 239–47.

Caldecott JO and Miles L. 2005. World atlas of great apes and their conservation. Berkeley, CA, and Cambridge, UK: University of California Press and UNEP-WCMC.

Campbell G, Kuehl H, Diarrassouba A, *et al.* 2011. Long-term research sites as refugia for threatened and over-harvested species. *Biol Lett* **7**: 723–26.

Ceballos G and Ehrlich PR. 2006. Global mammal distributions, biodiversity hotspots, and conservation. *P Natl Acad Sci USA* **103**: 19374–79.

Collen B, Ram M, Zamin T, *et al.* 2008. The tropical biodiversity data gap: addressing disparity in global monitoring. *Trop Conserv Sci* **1**: 75–88.

Fournier DA, Skaug HJ, Ancheta J, *et al.* 2012. AD model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optim Method Softw* **27**: 233–49.

Grenyer R, Orme CDL, Jackson SF, *et al.* 2006. Global distribution and conservation of rare and threatened vertebrates. *Nature* **444**: 93–96.

Habel JC, Gossner MM, Meyer ST, *et al.* 2013. Mind the gaps when using science to address conservation concerns. *Biodivers Conserv* **22**: 2413–27.

Junker J, Blake S, Boesch C, *et al.* 2012. Recent decline in suitable environmental conditions for African great apes. *Divers Distrib* **18**: 1077–91.

Kier G, Mutke J, Dinerstein E, *et al.* 2005. Global patterns of plant diversity and floristic knowledge. *J Biogeogr* **32**: 1107–16.

Knight AT, Cowling RM, Rouget M, *et al.* 2008. Knowing but not doing: selecting priority conservation areas and the research–implementation gap. *Conserv Biol* **22**: 610–17.

Kreft H and Jetz W. 2007. Global patterns and determinants of vascular plant diversity. *P Natl Acad Sci USA* **104**: 5925–30.

Laurance WF, Useche LD, Rendeiro J, *et al.* 2012. Averting biodiversity collapse in tropical forest protected areas. *Nature* **489**: 90–94.

Laurance WF. 2013. Does research help to safeguard protected areas? *Trends Ecol Evol* **28**: 261–66.

Magin CD, Johnson TH, Groombridge B, *et al.* 1994. Species extinctions, endangerment and captive breeding. In: Olney PJS, Mace GM, and Feistner ATC (Eds). Creative conservation: interactive management of wild and captive animals. London, UK: Chapman and Hall.

Martin LJ, Blossey B, and Ellis E. 2012. Mapping where ecologists work: biases in the global distribution of terrestrial ecological observations. *Front Ecol Environ* **10**: 195–201.

Mccallum ML and Bury GW. 2013. Google search patterns suggest declining interest in the environment. *Biodivers Conserv* **22**: 1355–67.

McQuistan CI, Fahmi Z, Leisher C, *et al.* 2006. Protected area funding in Indonesia. Jakarta, Indonesia: Ministry of Environment of the Republic of Indonesia and The Nature Conservancy.

Meho L and Yang K. 2007. Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *J Am Soc Inf Sci Tec* **58**: 2105–25.

Meijaard E, Wich SA, Ancrenaz M, and Marshall AJ. 2012. Not by science alone: why orangutan conservationists must think outside the box. *Ann NY Acad Sci* **1249**: 29–44.

Meijaard E, Cardillo M, Meijaard EM, and Possingham HP. 2015. Geographic bias in citation rates of conservation research. *Conserv Biol* **29**: 920–25.

Platnick NI. 1991. Patterns of biodiversity: tropical vs temperate. *J Nat Hist* **25**: 1083–88.

R Development Core Team. 2014. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Schipper J, Chanso JS, Chiozza F, *et al.* 2008. The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science* **322**: 225–30.

Sitas N, Baillie JEM, and Isaac NJB. 2009. What are we saving? Developing a standardized approach for conservation action. *Anim Conserv* **12**: 231–37.

Stuart SN, Chanson JS, Cox NA, *et al.* 2004. Status and trends of amphibian declines and extinctions worldwide. *Science* **306**: 1783–86.

Wrangham RW. 2008. Why the link between long-term research and conservation is a case worth making. In: Wrangham RW and Ross E (Eds). Science and conservation in African forests. Cambridge, UK: Cambridge University Press.

## Supporting Information

Additional, web-only material may be found in the online version of this article at http://onlinelibrary.wiley.com/doi/10.1002/14-0195.1/suppinfo

**AJ Marshall *et al*. – Supplemental information**

**WebPanel 1. Author contributions**

AJM, EM, and DS conceived of and designed the study; AJM, EM, and EVC collected the data; AJM analyzed the data, produced the figures, and wrote the paper; EM, EVC, and DS provided substantive editorial input.

**WebTable 1. Full list of 36 models run on dataset $D_{SIZE \cdot AGE}$ ($n$ = 321 sites, all protected areas for which data on size and age were available); models ordered by ΔAIC**

| Model | Model description | Model syntax (# Google Scholar hits ~) | df | Log-likelihood | AIC | ΔAIC | AIC weight |
|---|---|---|---|---|---|---|---|
| m19.x | ape taxon + PA size + PA status + country (random effect) | ~ whatape + area + status + (1\|country) | 10 | -1389.42 | 2798.8 | 0.0 | 0.572 |
| m19.y | ape taxon + PA size + PA status + country (random effect) + age | ~ whatape + area + status + (1\|country) + age | 11 | -1389.09 | 2800.2 | 1.3 | 0.293 |
| m5.x | ape presence + PA size + PA status + country (random effect) | ~ AnyApe + area + status + (1\|country) | 6 | -1395.29 | 2802.6 | 3.7 | 0.088 |
| m5.y | ape presence + PA size + PA status + country (random effect) + age | ~ AnyApe + area + status + (1\|country) + age | 7 | -1395.23 | 2804.5 | 5.6 | 0.034 |
| m12.x | # of apes + PA size + PA status + country (random effect) | ~ Napes + area + status + (1\|country) | 6 | -1397.67 | 2807.3 | 8.5 | 0.008 |
| m12.y | # of apes + PA size + PA status + country (random effect) + age | ~ Napes + area + status + (1\|country) + age | 7 | -1397.62 | 2809.2 | 10.4 | 0.003 |
| m22.x | PA size + PA status + country (random effect) | ~ area + status + (1\|country) | 5 | -1401.45 | 2812.9 | 14.1 | < 0.001 |
| m24.y | ape taxon + PA status + country (random effect) + age | ~ whatape + status + (1\|country) + age | 10 | -1396.78 | 2813.6 | 14.7 | < 0.001 |
| m22.y | PA size + PA status + country (random effect) + age | ~ area + status + (1\|country) + age | 6 | -1401.36 | 2814.7 | 15.9 | < 0.001 |
| m24.x | ape taxon + PA status + country (random effect) | ~ whatape + status + (1\|country) | 9 | -1399.37 | 2816.7 | 17.9 | < 0.001 |
| m25.y | ape presence + PA status + country (random effect) + age | ~ AnyApe + status + (1\|country) + age | 5 | -1404.89 | 2821.8 | 22.9 | < 0.001 |
| m25.x | ape presence + PA status + country (random effect) | ~ AnyApe + status + (1\|country) | 5 | -1406.33 | 2822.7 | 23.8 | < 0.001 |
| m26.y | # of apes + PA status + country (random effect) + age | ~ Napes + status + (1\|country) + age | 5 | -1407.93 | 2827.9 | 29.0 | < 0.001 |
| m26.x | # of apes + PA status + country (random effect) | ~ Napes + status + (1\|country) | 5 | -1409.38 | 2828.8 | 29.9 | < 0.001 |
| m18.x | ape taxon + PA size + PA status | ~ whatape + area + status | 9 | -1442.66 | 2903.3 | 104.5 | < 0.001 |
| m18.y | ape taxon + PA size + PA status + age | ~ whatape + area + status + age | 10 | -1442.1 | 2904.2 | 105.4 | < 0.001 |
| m23.x | PA size + PA status | ~ area + status | 4 | -1449.73 | 2907.5 | 108.7 | < 0.001 |
| m23.y | PA size + PA status + age | ~ area + status + age | 4 | -1449.16 | 2908.3 | 109.5 | < 0.001 |
| m4.x | ape presence + PA size + PA status | ~ AnyApe + area + status | 5 | -1449.32 | 2908.6 | 109.8 | < 0.001 |
| m11.x | # of apes + PA size + PA status | ~ Napes + area + status | 5 | -1449.69 | 2909.4 | 110.6 | < 0.001 |

| Model | Model description | Model syntax (# Google Scholar hits ~) | df | Log-likelihood | AIC | ΔAIC | AIC weight |
|---|---|---|---|---|---|---|---|
| m7.x | ape presence + PA size + country (random effect) | ~ AnyApe + area + (1\|country) | 5 | -1450.73 | 2911.5 | 112.7 | < 0.001 |
| m14.x | # of apes + PA size + country (random effect) | ~ Napes + area + (1\|country) | 5 | -1451.55 | 2913.1 | 114.3 | < 0.001 |
| m21.x | ape taxon + PA size + country (random effect) | ~ whatape + area + (1\|country) | 9 | -1449.47 | 2916.9 | 118.1 | < 0.001 |
| m6.x | ape presence + country (random effect) | ~ AnyApe + (1\|country) | 4 | -1474.83 | 2957.7 | 158.9 | < 0.001 |
| m13.x | # of apes + country (random effect) | ~ Napes + (1\|country) | 4 | -1475.38 | 2958.8 | 160.0 | < 0.001 |
| m20.x | ape taxon + country (random effect) | ~ whatape + (1\|country) | 8 | -1473.81 | 2963.6 | 164.8 | < 0.001 |
| m3.x | ape presence + PA size + interaction | ~ AnyApe + area + AnyApe * area | 5 | -1486.51 | 2983 | 184.2 | < 0.001 |
| m2.x | ape presence + PA size | ~ AnyApe + area | 4 | -1487.69 | 2983.4 | 184.6 | < 0.001 |
| m16.x | ape taxon + PA size | ~ whatape + area | 8 | -1484.87 | 2985.7 | 186.9 | < 0.001 |
| m17.x | ape taxon + PA size + interaction | ~ whatape + area + whatape * area | 8 | -1484.87 | 2985.7 | 186.9 | < 0.001 |
| m10.x | # of apes + PA size + interaction | ~ Napes + area + Napes * area | 5 | -1488.3 | 2986.6 | 187.8 | < 0.001 |
| m9.x | # of apes + PA size | ~ Napes + area | 4 | -1489.85 | 2987.7 | 188.9 | < 0.001 |
| m1.x | ape presence | ~ AnyApe | 3 | -1500.7 | 3007.4 | 208.6 | < 0.001 |
| m0.x | intercept | ~ 1 | 2 | -1502.35 | 3008.7 | 209.9 | < 0.001 |
| m8.x | # of apes | ~ Napes | 3 | -1501.72 | 3009.4 | 210.6 | < 0.001 |
| m15.x | ape taxon | ~ whatape | 7 | -1499.46 | 3012.9 | 214.1 | < 0.001 |

**WebTable 2. Full list of 27 models run on dataset D$_{SIZE}$ ($n$ = 426 sites, all protected areas for which data on size were available)**

| Model | Model description | Model syntax (# Google Scholar hits ~) | df | Log-likelihood | AIC | ΔAIC | AIC weight |
|---|---|---|---|---|---|---|---|
| m19 | ape taxon + PA size + PA status + country (random effect) | ~ whatape + area + status + (1\|country) | 10 | -1698.31 | 3416.6 | 0.0 | 0.798 |
| m5 | ape presence + PA size + PA status + country (random effect) | ~ AnyApe + area + status + (1\|country) | 6 | -1703.77 | 3419.5 | 2.9 | 0.185 |
| m12 | # of apes + PA size + PA status + country (random effect) | ~ Napes + area + status + (1\|country) | 6 | -1706.44 | 3424.9 | 8.3 | 0.013 |
| m24 | ape taxon + PA status + country (random effect) | ~ whatape + status + (1\|country) | 9 | -1704.83 | 3427.7 | 11.1 | 0.003 |
| m25 | ape presence + PA status + country (random effect) | ~ AnyApe + status + (1\|country) | 5 | -1710.61 | 3431.2 | 14.6 | < 0.001 |
| m26 | # of apes + PA status + country (random effect) | ~ Napes + status + (1\|country) | 5 | -1713.51 | 3437 | 20.4 | < 0.001 |
| m22 | PA size + PA status + country (random effect) | ~ area + status + (1\|country) | 5 | -1715.13 | 3440.3 | 23.7 | < 0.001 |
| m18 | ape taxon + PA size + PA status | ~ whatape + area + status | 9 | -1759.01 | 3536 | 119.4 | < 0.001 |
| m4 | ape presence + PA size + PA status | ~ AnyApe + area + status | 5 | -1764.85 | 3539.7 | 123.1 | < 0.001 |
| m11 | # of apes + PA size + PA status | ~ Napes + area + status | 5 | -1767.87 | 3545.7 | 129.1 | < 0.001 |
| m23 | PA size + PA status | ~ area + status | 4 | -1769.74 | 3547.5 | 130.9 | < 0.001 |
| m14 | # of apes + PA size + country (random effect) | ~ Napes + area + (1\|country) | 5 | -1776.85 | 3563.7 | 147.1 | < 0.001 |
| m7 | ape presence + PA size + country (random effect) | ~ AnyApe + area + (1\|country) | 5 | -1777.96 | 3565.9 | 149.3 | < 0.001 |
| m21 | ape taxon + PA size + country (random effect) | ~ whatape + area + (1\|country) | 9 | -1775.14 | 3568.3 | 151.7 | < 0.001 |
| m13 | # of apes + country (random effect) | ~ Napes + (1\|country) | 4 | -1787.91 | 3583.8 | 167.2 | < 0.001 |
| m6 | ape presence + country (random effect) | ~ AnyApe + (1\|country) | 4 | -1788.64 | 3585.3 | 168.7 | < 0.001 |
| m20 | ape taxon + country (random effect) | ~ whatape + (1\|country) | 8 | -1786.36 | 3588.7 | 172.1 | < 0.001 |
| m2 | ape presence + PA size | ~ AnyApe + area | 4 | -1815.59 | 3639.2 | 222.6 | < 0.001 |
| m3 | ape presence + PA size + interaction | ~ AnyApe + area + AnyApe * area | 5 | -1814.96 | 3639.9 | 223.3 | < 0.001 |
| m16 | ape taxon + PA size | ~ whatape + area | 8 | -1813.19 | 3642.4 | 225.8 | < 0.001 |
| m9 | # of apes + PA size | ~ Napes + area | 4 | -1818.83 | 3645.7 | 229.1 | < 0.001 |
| m10 | # of apes + PA size + interaction | ~ Napes + area + Napes * area | 5 | -1818.24 | 3646.5 | 229.9 | < 0.001 |
| m17 | ape taxon + PA size + interaction | ~ whatape + area + whatape * area | 13 | -1811.49 | 3649 | 232.4 | < 0.001 |
| m1 | ape presence | ~ AnyApe | 3 | -1826.59 | 3659.2 | 242.6 | < 0.001 |
| m8 | # of apes | ~ Napes | 3 | -1828.3 | 3662.6 | 246.0 | < 0.001 |
| m15 | ape taxon | ~ whatape | 7 | -1825.56 | 3665.1 | 248.5 | < 0.001 |
| m0 | intercept | ~ 1 | 2 | -1830.57 | 3665.1 | 248.5 | < 0.001 |

Notes: these are the models run on D$_{SIZE•AGE}$, with the exception that models including age were excluded, as this information was lacking for many sites in D$_{SIZE}$. Models ordered by ΔAIC.

**WebTable 3. Full list of 10 models run on dataset D$_{FULL}$ ($n = 565$ sites, all protected areas in sample)**

| Model | Model description | Model syntax (# Google Scholar hits ~) | df | Log-likelihood | AIC | ΔAIC | AIC weight |
|---|---|---|---|---|---|---|---|
| m.24 | ape taxon + PA status + country (random effect) | ~ whatape + status + (1\|country) | 9 | -2062.73 | 4143.5 | 0.0 | 0.730 |
| m.25 | ape presence + PA status + country (random effect) | ~ AnyApe + status + (1\|country) | 5 | -2067.71 | 4145.4 | 1.9 | 0.270 |
| m.26 | # of apes + PA status + country (random effect) | ~ Napes + status + (1\|country) | 5 | -2076.21 | 4162.4 | 18.9 | < 0.001 |
| m.6 | ape presence + country (random effect) | ~ AnyApe + (1\|country) | 4 | -2173.19 | 4354.4 | 210.9 | < 0.001 |
| m.13 | # of apes + country (random effect) | ~ Napes + (1\|country) | 4 | -2174.65 | 4357.3 | 213.8 | < 0.001 |
| m.20 | ape taxon + country (random effect) | ~ whatape + (1\|country) | 8 | -2171.35 | 4358.7 | 215.2 | < 0.001 |
| m.1 | ape presence | ~ AnyApe | 3 | -2210.61 | 4427.2 | 283.7 | < 0.001 |
| m.15 | ape taxon | ~ whatape | 7 | -2208.94 | 4431.9 | 288.4 | < 0.001 |
| m.8 | # of apes | ~ Napes | 3 | -2212.95 | 4431.9 | 288.4 | < 0.001 |
| m.0 | intercept | ~ 1 | 2 | -2215.46 | 4434.9 | 291.4 | < 0.001 |

Notes: these are the models run on D$_{SIZE}$, with the exception that models including protected area size were excluded, as this information was lacking for many sites in D$_{FULL}$. Models ordered by ΔAIC.

# WebTable 4. Top models for each dataset (ie summed AIC weight >0.99)

| Dataset | Model | Model description | AIC | ΔAIC | AIC model weight | Gorillas | Orangutans | Chimpanzees & Gorillas | Chimpanzees | Bonobos | Any ape (β ± SE) | # of apes (β ± SE) | PA status (β ± SE) | Country (var ± SD) | PA size (log (area)) (β ± SE) | PA age (β ± SE) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **D_SIZE•AGE** (N= 321 sites) | m19.x | ape taxon + PA size + PA status + country | 2799 | 0.0 | 0.572 | 2.98 ± 0.88 | 1.13 ± 0.50 | 0.61 ± 0.45 | 0.56 ± 0.28 | -0.48 ± 1.65 | | | 2.55 ± 0.22 | 1.85 ± 1.4 | 0.23 ± 0.05 | |
| | m19.y | ape taxon + PA size + PA status + country + age | 2800 | 1.3 | 0.293 | 3.05 ± 0.89 | 1.18 ± 0.50 | 0.61 ± 0.45 | 0.54 ± 0.28 | -0.48 ± 1.65 | | | 2.59 ± 0.22 | 1.80 ± 1.3 | 0.22 ± 0.05 | 3.87E-03 ± 4.78E-03 |
| | m5.x | ape presence + PA size + PA status + country | 2803 | 3.7 | 0.088 | | | | | | 0.82 ± 0.24 | | 2.43 ± 0.22 | 1.74 ± 1.3 | 0.25 ± 0.05 | |
| | m5.y | ape presence + PA size + PA status + country + age | 2805 | 5.6 | 0.034 | | | | | | 0.82 ± 0.24 | | 2.45 ± 0.22 | 1.72 ± 1.3 | 0.25 ± 0.05 | 1.62E-03 ± 4.79E-03 |
| | m12.x | # of apes + PA size + PA status + country | 2807 | 8.5 | 0.008 | | | | | | | 0.53 ± 0.20 | 2.40 ± 0.22 | 1.88 ± 1.4 | 0.26 ± 0.05 | |
| | | weighted β ± SE across all top models | | | | 2.61 ± 0.76 | 1.00 ± 0.43 | 0.53 ± 0.39 | 0.48 ± 0.25 | -0.42 ± 1.43 | 0.10 ± 0.03 | 0.004 ± 0.002 | 2.54 ± 0.22 | | 0.23 ± 0.05 | 1.19E-03 ± 1.57E-03 |
| | | weighted β ± SE across top models containing predictor | | | | 3.00 ± 0.88 | 1.15 ± 0.50 | 0.61 ± 0.45 | 0.55 ± 0.28 | -0.48 ± 1.65 | 0.82 ± 0.24 | 0.53 ± 0.20 | 2.54 ± 0.22 | | 0.23 ± 0.05 | 3.63E-03 ± 4.78E-03 |
| **D_SIZE** (N= 426 sites) | m19 | ape taxon + PA size + PA status + country | 3417 | 0.0 | 0.798 | 3.56 ± 0.97 | 1.35 ± 0.51 | 1.1 ± 0.44 | 0.85 ± 0.29 | 0.35 ± 1.29 | | | 2.72 ± 0.21 | 1.85 ± 1.4 | 5.34E-05 ± 1.83E-05 | NA |
| | m5 | ape presence + PA size + PA status + country | 3420 | 2.9 | 0.185 | | | | | | 1.13 ± 0.24 | | 2.66 ± 0.21 | 1.71 ± 1.3 | 5.71E-05 ± 1.95E-05 | NA |
| | m12 | # of apes + PA size + PA status + country | 3425 | 8.3 | 0.013 | | | | | | | 0.8 ± 0.2 | 2.59 ± 0.21 | 1.95 ± 1.4 | 5.99E-05 ± 2.03E-05 | NA |
| | | weighted β ± SE across all top models | | | | 2.85 ± 0.78 | 1.08 ± 0.41 | 0.88 ± 0.35 | 0.68 ± 0.23 | 0.28 ± 1.03 | 0.21 ± 0.04 | 0.10 ± 0.002 | 2.71 ± 0.21 | | 5.42E-05 ± 1.85E-05 | NA |
| | | weighted β ± SE across top models containing predictor | | | | 3.56 ± 0.97 | 1.35 ± 0.51 | 1.1 ± 0.44 | 0.85 ± 0.29 | 0.35 ± 1.29 | 1.13 ± 0.24 | 0.8 ± 0.2 | 2.71 ± 0.21 | | 5.42E-05 ± 1.85E-05 | NA |
| **D_FULL** (N= 565 sites) | m.24 | ape taxon + PA status + country | 4144 | 0.0 | 0.730 | 3.76 ± 0.96 | 1.51 ± 0.56 | 1.13 ± 0.39 | 1.59 ± 0.26 | 1.39 ± 1.39 | | | 3.13 ± 0.21 | 1.85 ± 1.36 | NA | NA |
| | m.25 | ape presence + PA status + country | 4145 | 1.9 | 0.270 | | | | | | 1.62 ± 0.22 | | 3.08 ± 0.21 | 192 ± 1.38 | NA | NA |
| | | weighted β ± SE across all top models | | | | 2.73 ± 0.70 | 1.10 ± 0.41 | 0.82 ± 0.28 | 1.16 ± 0.19 | 1.01 ± 1.01 | 0.44 ± 0.06 | | 3.11 ± 0.21 | | NA | NA |
| | | weighted β ± SE across top models containing predictor | | | | 3.76 ± 0.96 | 1.51 ± 0.56 | 1.13 ± 0.39 | 1.59 ± 0.26 | 1.39 ± 1.39 | 1.62 ± 0.22 | | 3.11 ± 0.21 | | NA | NA |

Notes: cells marked "NA" indicate predictors that could not be included in models for that dataset due to missing data; empty cells indicate predictors not present in the model; shaded cells indicate predictors with estimated effects (β ± SE) that did not include zero. Below the list of top models for each dataset are weighted β ± SE for predictors across all top models and across only models containing that predictor.

# Erratum

In Marshall *et al.* (2016; **14**: 13–19, doi.org/10.1002/14-0195.1), due to errors discovered and reported by the authors, selected text and one figure require correction. The revised text and figure are as follows:

In the Methods, the subsection "Assessing taxonomic focus" should instead begin with the following sentence…

"For a random sample (10% of $D_{SIZE}$, $n$ = 62) of protected areas, we used simple decision rules to assess the taxonomic focus of all returned hits based on their titles."

…and end with the following three sentences:

"For one of the sites that we randomly selected to assess content, hit counts were >1000 (1120), and therefore not all hits were listed. For this site we based content analysis on the 1000 hits listed. We modeled the hit counts for each taxonomic group separately, using the same methods applied to our main analyses."

In the Results, the subsection "Taxonomic focus of returned hits" should instead read as:

"The most common taxonomic content of returned hits was non-primate mammals (30.9% of the taxon-related hits in our random sample of sites). Apes and non-ape primates comprised an additional 15.3% and 11.2% of hits, respectively. This indicates that most tropical research is focused on a small subset of taxa: overall 57.4% of hits related to mammals, as compared with 7.8% for birds, 16.4% for plants, and 18.4% for all other taxa (Figure 3). When we modeled hit counts for each taxon individually, the best model for each taxon contained terms for great ape presence (for apes, non-ape primates, non-primate mammals, plants, and other taxa) or which ape was present (birds). Protected area size appeared in the top models for apes, non-primate mammals, birds, plants, and other taxa, but not for non-ape primates. IUCN status appeared in the top model for birds and non-primate mammals only."

In the Discussion, the last two sentences of the first paragraph should instead read as:

"Analysis of the taxonomic focus of all hits for a randomly selected subset of protected areas suggests that research published on great apes follows the same general patterns seen in our broader analysis; research on non-ape taxa is also typically biased toward large national parks that contain great apes. This finding indicates that the research stations and associated infrastructure often built to support research on great apes may facilitate research on other taxa."

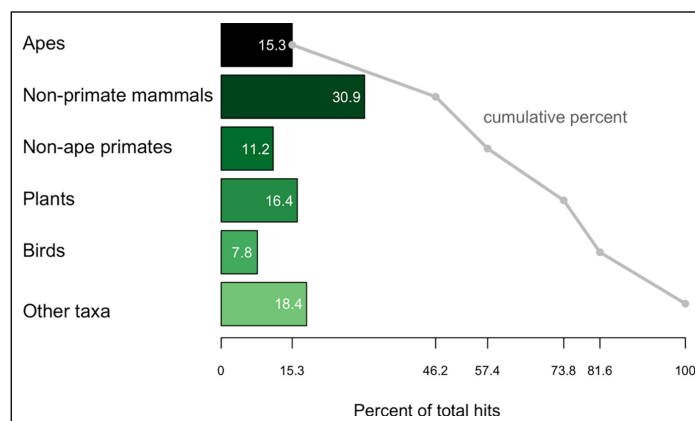The revised version of Figure 3 and its caption should appear as follows:



**Figure 3.** Taxonomic content of Google Scholar hits for a randomly selected 10% sample of protected areas (*n* = 62 protected areas, 4785 hits). Horizontal bars indicate the percentage of hits for each taxonomic group (eg 30.9% of hits related to non-primate mammals). The gray line indicates the cumulative overall percentage of hits (eg mammals – apes, non-primate mammals, and non-ape primates – together comprise 57.4% of all hits).