Research

Complete, accurate, mammalian phylogenies aid conservation planning, but not much

Ana S. L. Rodrigues[1,*]**, Richard Grenyer**[2]**, Jonathan E. M. Baillie**[3]**,
Olaf R. P. Bininda-Emonds**[5]**, John L. Gittlemann**[6]**,
Michael Hoffmann**[7,8,9]**, Kamran Safi**[4,10]**, Jan Schipper**[11]**,
Simon N. Stuart**[7,8,9,12,13] **and Thomas Brooks**[14,15,16]

[1]*Centre d'Ecologie Fonctionnelle et Evolutive, CNRS UMR5175, 1919 Route de Mende,
34293 Montpellier, France*
[2]*School of Geography and the Environment, Oxford University Centre for the Environment,
University of Oxford, South Parks Road, Oxford OX1 3QY, UK*
[3]*Zoological Society of London, and* [4]*Institute of Zoology, Zoological Society of London,
Regent's Park, London NW1 4RY, UK*
[5]*AG Systematik und Evolutionsbiologie, IBU—Fakultät V, Carl von Ossietzky Universität Oldenburg,
Carl von Ossietzky Strasse 9-11, 26111 Oldenburg, Germany*
[6]*Odum School of Ecology, University of Georgia, Athens, GA 30602, USA*
[7]*IUCN Species Survival Commission, Rue Mauverney 28, 1196 Gland, Switzerland*
[8]*United Nations Environment Programme World Conservation Monitoring Centre, 219 Huntington Road,
Cambridge CB3 0DL, UK*
[9]*Conservation International, 2011 Crystal Drive Suite 500, Arlington, VA 22202, USA*
[10]*Max Planck Institute for Ornithology, Vogelwarte Radolfzell, Schlossallee 2, 78315 Radolfzell, Germany*
[11]*Pacific Cooperative Studies Unit, University of Hawai'i, Big Island Invasive Species Committee,
23 East Kawili Street, Hilo, HI 96720, USA*
[12]*Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK*
[13]*Al Ain Wildlife Park and Resort, PO Box 45553, Abu Dhabi, United Arab Emirates*
[14]*NatureServe, 1101 Wilson Boulevard, 15th Floor, Arlington, VA 22209, USA*
[15]*World Agroforestry Center (ICRAF), University of the Philippines Los Baños, Laguna 4031, Philippines*
[16]*School of Geography and Environmental Studies, University of Tasmania, Hobart,
Tasmania 7001, Australia*

In the face of unprecedented global biodiversity loss, conservation planning must balance between refining and deepening knowledge versus acting on current information to preserve species and communities. Phylogenetic diversity (PD), a biodiversity measure that takes into account the evolutionary relationships between species, is arguably a more meaningful measure of biodiversity than species diversity, but cannot yet be applied to conservation planning for the majority of taxa for which phylogenetic trees have not yet been developed. Here, we investigate how the quality of data on the taxonomy and/or phylogeny of species affects the results of spatial conservation planning in terms of the representation of overall mammalian PD. The results show that the better the quality of the biodiversity data the better they can serve as a basis for conservation planning. However, decisions based on incomplete data are remarkably robust across different levels of degrading quality concerning the description of new species and the availability of phylogenetic information. Thus, given the level of urgency and the need for action, conservation planning can safely make use of the best available systematic data, limited as these data may be.

Keywords: complementarity; conservation planning; evolutionary history; mammals;
phylogenetic diversity; surrogacy

1. INTRODUCTION

Effective biodiversity conservation planning—deciding where to focus conservation efforts to maximize the persistence of biodiversity—relies on detailed, spatially explicit data about each of the biodiversity features of

interest [1]. Yet real data are inevitably incomplete and perpetually changing as new knowledge is incorporated. There is a trade-off in the cost-effectiveness of conservation action between investing in refining existing information and action based on existing data [2–4]. Surrogacy tests can shed light on this trade-off, by quantifying how the adequacy of conservation planning results varies as data quality improves. Such tests can be performed by degrading existing high-quality data in order to simulate more incomplete (and more realistic) datasets, and then testing the surrogacy value of the latter using the former as benchmarks. For example, high-quality species distributional data can be degraded into simulated incomplete or biased datasets, and the effectiveness of conservation planning based on the degraded datasets assessed against the complete data (e.g. [3–5]). Similarly, multi-taxon distribution data can be degraded into single taxon datasets to investigate whether conservation planning based on just one taxon adequately represents broader, multi-taxon, diversity (see [6] for a review). Here, we analyse how data can improve over time as knowledge increases for a particular taxonomic group being studied (e.g. because new species are described; [7]; because of taxonomic changes affecting already recognized taxa (e.g. [8]); or through better understanding of the relationship between species (e.g. [9]). We focus on mammals because they are among the best-known taxa, and by degrading the available taxonomic and phylogenetic information in various ways we can simulate the more imperfect state of knowledge of less-known groups.

Surrogacy tests require a common biodiversity metric to assess the quality of the results obtained from conservation planning based on datasets of variable quality. The measure most commonly used is the total number of species represented in the selected areas [6], but this assumes species are equivalent, ignoring the fact that they differ in the amount of evolutionary history they represent [10–12]. For example, the Platypus *Ornithorhynchus anadinus* is the sole species in the family Ornithorhynchidae, and one of only five species in the order Monotremata, whereas the Cactus Mouse *Peromyscus eremicus* is one of 56 species in its genus, one of 692 species in the family Cricetidae and one of 2280 species in the order Rodentia [13]. Phylogenetic diversity (PD) [14,15] is a biodiversity measure that takes into account the evolutionary relationships between species, and can be calculated for mammals because a phylogenetic super-tree is already available [16,17]. The PD in a particular site can be measured as the total branch length of the phylogenetic tree, which includes only those taxa present at the site. PD is arguably a more meaningful measure of biodiversity than simple species richness [18] because differences among genotypes are the raw material on which evolutionary processes operate. It is also more robust to idiosyncratic variation in the species concept because splits between closely related species add little diversity to a phylogenetic tree [19,20]. Furthermore, studies show that extinction is phylogenetically non-random and that we are losing evolutionary history faster than expected from species loss [21–23], suggesting that evolutionary history

needs to be targeted directly as part of conservation strategies. In this analysis, PD is the biodiversity metric employed to evaluate the outcome of conservation planning based on data of various quality.

Taking into account evolutionary history clearly makes a difference when prioritizing species for conservation [12,24,25], by distinguishing those that are phylogenetically most distinct from those with many living close relatives (e.g. Platypus and Cactus Mouse, as mentioned above). However, this is not necessarily the case when targeting areas for conservation, because when conservation is done spatially rather than species-by-species (as it usually is), the possible combinations of species (and corresponding PD) are limited by the variety of assemblages that exist in nature [18]. Previous simulations suggest that, in most circumstances, selecting networks of protected areas by maximizing overall species richness is likely to capture overall PD efficiently; that is, that species richness is a good surrogate of PD [18]. But there have been few empirical tests of these predictions, and common conclusions have not yet emerged [26–28]. Here, we contribute to this discussion by testing whether species data are an appropriate surrogate for the representation of global mammalian PD for spatial conservation planning purposes. We further extend these tests by analysing the surrogacy value of other types of data in representing PD: data on genera (a coarser taxonomic resolution than species); using taxonomy as a proxy for phylogeny; and simulating earlier, more incomplete, states of knowledge on species diversity.

Testing for surrogacy involves evaluating the outcomes of systematic conservation planning, yet there are countless ways in which such planning can be done. The methods originally proposed aimed simply to ensure representation of all biodiversity features (such as species) in the smallest possible area [29,30]. Current approaches have moved very substantially from these simple 'minimum sets' to increase the level of socio-economic realism and of ecological pertinence of conservation planning, for example, by taking into account economic costs, the effects of climate change, variability in species' conservation requirements and the spatio-temporal dynamics of threats (see [31] for a review). In the surrogacy tests presented in these analyses, we nonetheless use minimum representation sets as the outputs of conservation planning, in order to simplify the interpretation of the results, and to ensure comparability with previous surrogacy tests [6].

2. DATA AND METHODS
(a) *Species distribution data*
We obtained data on the spatial distribution of 5258 land mammal species from the International Union for Conservation of Nature (IUCN) Red List of Threatened Species [32,33], mapped as polygons representing species' extent of occurrence. These are coarse generalizations of species' distributions, generally obtained as 'envelopes' including original records and through interpolation (but not extrapolation) from original records [34]. They may include relatively extensive areas from which the species is absent (e.g. freshwater habitats within a terrestrial species' range) and are therefore likely to overestimate the

species' true area of occupancy (see the supporting online material in Hoffmann *et al.* [35] for further details). Individual polygons are coded according to species' presence (1 extant; 2 probably extant; 3 possibly extant; 4 possibly extinct; 5 extinct), origin (1 native; 2 reintroduced; 3 introduced; 4 vagrant; 5 origin uncertain) and seasonality (1 resident; 2 breeding season; 3 non-breeding season; 4 passage; 5 seasonal occurrence uncertain). In this analysis, we included only parts of each species' range coded as 1 or 2 in presence and in origin (thereby excluding extinct or possibly extinct species and areas where species are not native).

(b) *Spatial units*
The world's land area was divided into equal-area (approx. 23 322 km^2), equal-shape (hexagonal) cells, the spatial units used in all analyses. These were obtained from a geodesic discrete global grid system, defined on an icosahedron and projected to the sphere using the inverse Icosahedral Snyder Equal Area (ISEA) Projection [36]. A species was considered present in a given hexagon whenever its mapped range overlaps the cell. There were 7316 hexagons with at least one species.

(c) *Species taxonomy*
The taxonomic classification for mammals follows that used in the IUCN Red List (itself largely based on Wilson & Reeder [13], with minor adaptations; see the supporting online material of Schipper *et al.* [32] for further details).

(d) *Species phylogeny*
We followed the method described by Schipper *et al.* [32] (see their supporting online material) to expand and modify the mammalian super-tree published by Fritz *et al.* [17], by: adding recently described or recognized species currently listed in the IUCN Red List and not included in the original tree, using their taxonomy to infer their relative phylogenetic position; removing species that are no longer recognized, and those that are extinct or domesticated. Polytomies—unresolved nodes in a phylogeny with more than two descendent nodes—distort the length of branches within a phylogeny. Since only one descendent branch of a polytomy is of its true length, we used a correction factor to reduce the length of terminal branches descending from polytomies (corollary 3 (ii) in Steel & Mooers [37]). To date there is no recognized way to perform this correction for internal polytomous nodes. The final taxonomic tree included all 5258 species for which distribution data were available, corresponding to a total PD of 64 102.1 million years before present (MYBP).

(e) *Maximizing phylogenetic diversity*
As a working scenario, we considered that the ultimate conservation goal was to maximize the representation of overall mammalian PD in a hypothetical global set of protected areas. The maximum PD that can be represented in a given number of sites can be known exactly by solving a maximal covering location problem [27,38]. We used the Solving Constraint Integer Programs (*SCIP*) solver [39] in the free online NEOS Server [40,41] to obtain an optimal solution to maximal covering location problems that maximized the total PD in sets of sites of variable size (up to 597 cells, the minimum set needed for representing all total PD).

(f) *Surrogate datasets*
For the purpose of these analyses, we considered that the current data on species' distribution and phylogeny (henceforth, dataset A) corresponds to complete knowledge, and that the values of PD that can be obtained from these data were the ideal measure of mammalian diversity. We then obtained 10 types of surrogate datasets, by degrading these original data (table 1; datasets B to K). Some of these datasets included information on the evolutionary relationships between species, others did not. Dataset B included the same species information as dataset A, but without the phylogenetic data, thereby assuming that all currently known species are strictly equivalent. We then considered three pairs of datasets—with and without phylogenetic data (respectively, subsets of A and of B)—by including the following subsets of the currently known species:

— Species described pre-1993 (datasets C and D). To simulate an advanced state of taxonomic knowledge (90% of all species). The year 1993 was chosen because it corresponds to the Second Edition of *Mammal Species of the World* [42], although the data obtained in this way are not exactly the same as those that were available then (for example, the West African Linsang *Poiana leightoni* is a currently recognized species with a pre-1992 description date; it was then considered a sub-species of the Central African Linsang *Poiana richardsonii*).
— Species described pre-1970 (datasets E and F). To simulate scenarios where a group is relatively well-known (83% of all species), but where significant numbers of species remain to be described (the 1970 cut-off date is arbitrary).
— Species described pre-1908 (datasets G and H). To simulate scenarios where only a small fraction of the group's diversity is known (22% of all species). The 1908 cut-off was chosen because it resulted in the same number of species as there are currently recognized genera. This allowed us to compare (together with datasets I and J, see below), two types of datasets that have the same number of taxa but different taxonomic resolutions.

A fourth pair of datasets was obtained by subsuming all constituent species into one of 1163 genera, either including (dataset I) or not (dataset J) information on the phylogenetic relationships between genera. The 1163 genera then became the conservation targets. These scenarios may occur in situations where phylogenetic and/or the distributional data are not resolved to the species level [26–28], a plausible scenario for hyperdiverse taxa (such as plants or insects, particularly in tropical countries) for which identification is often done to a higher taxon level rather than resolved to species [43,44].

Finally, we considered a dataset (K) where a taxonomic tree was used as an approximation to the phylogenetic tree, with four levels: order, family, genera and species. Branches between levels were assumed to

Table 1. Overview of the 11 types of dataset used in this analysis, including the complete dataset (A) and the 10 surrogates derived from it (B to K).

dataset designation	taxonomic information	evolutionary information
A phylogenetic diversity	complete species data ($n = 5258$ species)	yes. Complete phylogenetic tree (PD = 64 102.1 MYBP)
B species diversity	complete species data ($n = 5258$ species)	no
C phylogenetic diversity pre-1993	species described pre-1993 ($n = 4724$ species)	yes. Tree pruned to pre-1993 species (PD = 58 537.6 MYBP)
D species diversity pre-1993	species described pre-1993 ($n = 4724$ species)	no
E phylogenetic diversity pre-1970	species described pre-1970 ($n = 3828$ species)	yes. Tree pruned to pre-1970 species (PD = 50 317.9 MYBP)
F species diversity pre-1970	species described pre-1970 ($n = 3828$ species)	no
G phylogenetic diversity pre-1908	species described pre-1908 ($n = 1163$ species)	yes. Tree pruned to pre-1908 species (PD = 21 518.9 MYBP)
H species diversity pre-1908	species described pre-1908 ($n = 1163$ species)	no
I genus phylogenetic diversity	taxonomy subsumed to genera ($n = 1163$ genera)	yes. Tree subsumed to genera (PD = 30 669.6 MYBP)
J genus diversity	species described pre-1908 ($n = 1163$ genera)	no
K taxonomic diversity	complete species data ($n = 5258$ species)	yes. Taxonomic tree used as an approximation to phylogenetic tree (TD = 6589 units)

have a length of one unit. The taxonomic diversity (TD) of a set of species was then measured in a similar way to PD, as the total branch length of the tree including only those species.

For each of these datasets, there was an associated surrogate measure of mammalian diversity: variations of species richness in B, D, F and H; variations of PD in C, E, G and I; genus richness in J; and TD in K (figure 2). We obtained, for each of the surrogate datasets, sets of sites of variable size (up to 597 cells, or the minimum set needed to represent the full diversity in each case) that maximized the corresponding diversity measures, and quantified how much overall PD was incidentally captured in each of these sets.

(g) Random site selection

We quantified how much PD is incidentally captured in randomly selected sets of sites of variable size (up to 597 cells, the size of the minimum set representing all species, and corresponding to *ca* 8% of the total area), by obtaining in each case 100 replicates and calculating the mean PD and limits of the 95% confidence interval. This corresponds in practice to what would be expected in terms of PD representation in the absence of biological data, a measure of the effectiveness of simply using area as a surrogate [6].

(h) Quantifying surrogacy

For each dataset, we quantified its surrogacy in representing PD by following the protocol proposed by Ferrier & Watson ([45]; described in detail in Rodrigues & Brooks [6]). This compared: the surrogacy value (S), defined as the total PD that is incidentally represented when conservation planning is based on the surrogate data (datasets B to K); the optimum value (O), defined as the maximum possible PD that can be represented in the same area
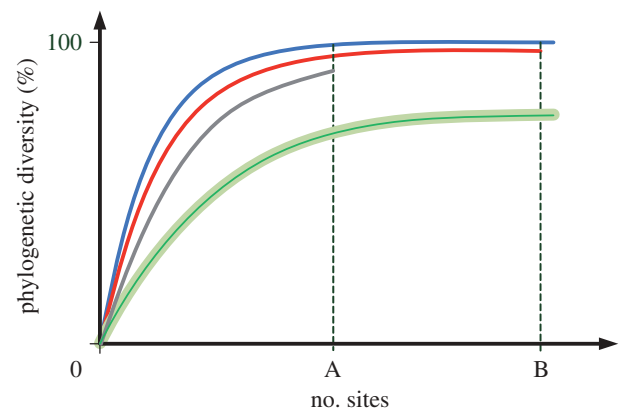
Figure 1. Schematic of the method followed for quantifying surrogacy value using the SAI index. $SAI = (S - R)/(O - R)$, where O is the area under the optimal curve (the maximum phylogenetic diversity, PD, that can be represented for any given number of sites), S is the area under the surrogate curve (the PD incidentally represented when maximizing a surrogate measure of diversity), R is the area under the random curve (the PD incidentally represented in sets selected at random). Values of S, O and R should be obtained for areas of the graph covered by all three curves (for example, SAI can be calculated for surrogate 1 to the left of point B, but for surrogate 2 only to the left of point A). Comparison between two surrogate curves with different endpoints must focus on the area of the graph common to both (here, to the left of point A). Blue line, optimal curve; red line, surrogate curve 1; grey line, surrogate curve 2; green line, random curve (mean + confidence intervals).

given perfect knowledge (i.e. based on dataset A); and the random value (R), defined as the expected PD that is incidentally represented if sites are selected at random. Each of these values can be measured as the area under the surrogate, optimal and random curves that are obtained for a range of solutions of variable size (figure 1). These three values are combined into an
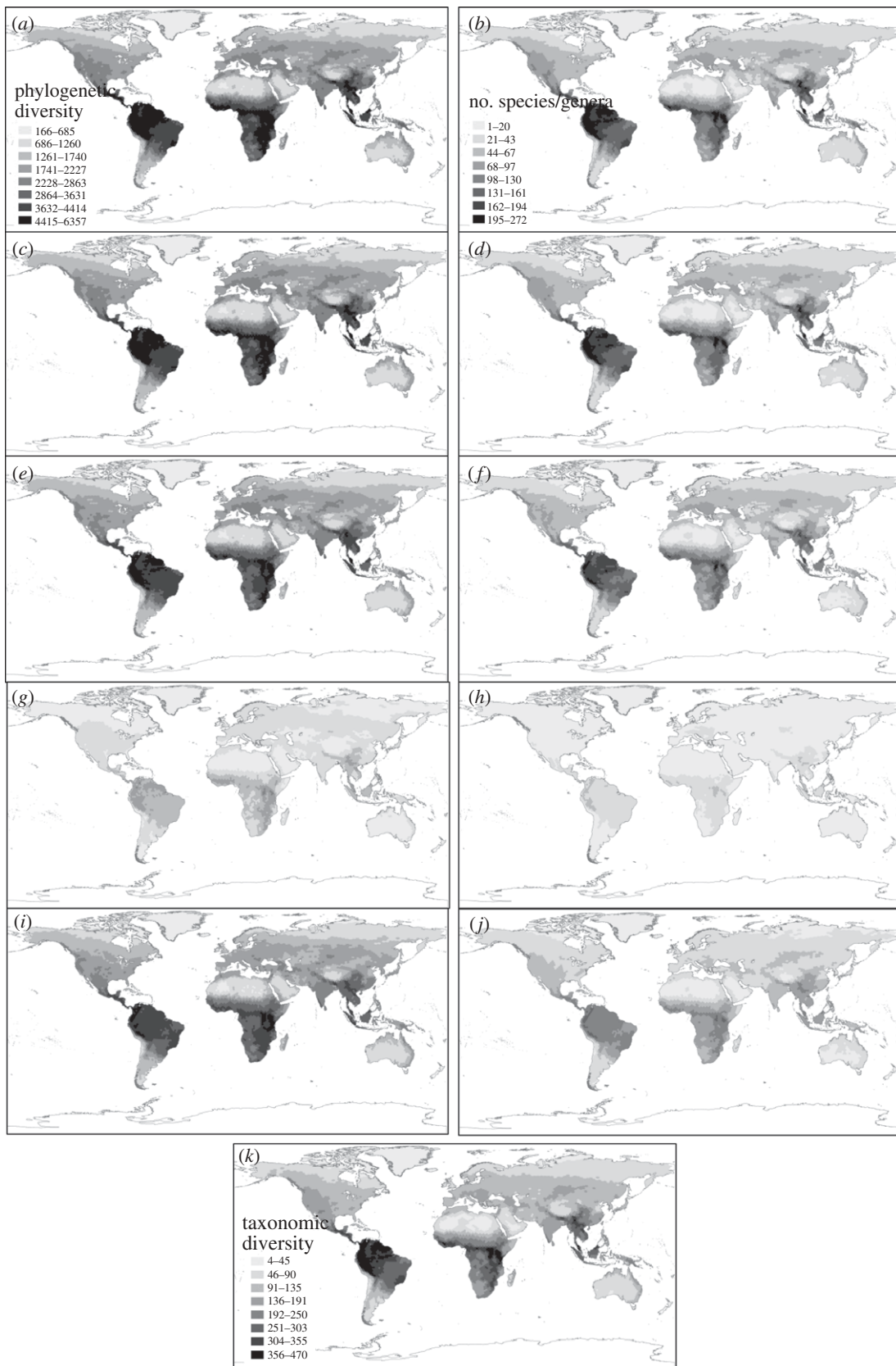
Figure 2. Spatial patterns of distribution of terrestrial mammals according to 11 measures of diversity: (*a*) PD of all currently described mammals (millions of years before present, MYBP); (*b*) species richness of all currently described mammals; (*c*) PD for species described pre-1993; (*d*) species richness for species described pre-1993; (*e*) PD for species described pre-1970; (*f*) species richness for species described pre-1970; (*g*) PD for species described pre-1908; (*h*) species richness for species described pre-1908; (*i*) PD for currently described genera; (*j*) genus richness for currently described mammals; (*k*) taxonomic diversity (measured in arbitrary units of branch length, see text for details). The spatial units are equal-area hexagons. The same legend is used across panels (*a,c,e,g,i*), and then across panels (*b,d,f,h,j*), and so maps in each group are directly comparable.

index $SAI = (S - R)/(O - R)$ [45]. SAI (which originally stood for 'species accumulation index') equals one if there is perfect surrogacy (i.e. if the surrogate data produce results as good as the perfect data), zero if there is null surrogacy (using the surrogate data is as good as a random selection), and less than zero if there is negative surrogacy (using the surrogate data is worse than a random selection).

SAI can be calculated for the mean random curve or the upper 95% confidence interval [6]. For simplicity, and given that the confidence intervals are narrow, we report SAI values based on the mean random curve.

For context, we have also quantified the degree of relationship between the local values (i.e. per cell) of each of the 10 surrogate diversity measures considered and PD, by calculating the respective correlation coefficient.

3. RESULTS AND DISCUSSION

We found high levels of surrogacy for all 10 datasets tested: distinctly superior to a random selection of sites, and in some cases even approaching optimality (figure 3 and table 2). The observed values of SAI (median = 0.94, Q1 = 0.88, Q3 = 0.96) are very high within the context of previously published surrogacy tests (see Rodrigues & Brooks [6] for a review: median = 0.12, Q1 = 0.03, Q3 = 0.28; n = 575). There were also very high correlations between each of the 10 surrogate biodiversity measures and PD (figure 2 and table 2), even if these are likely to be somewhat inflated by spatial autocorrelation.

Our results support the predictions from simulated scenarios that species data are generally likely to be good surrogates of PD for spatial conservation planning purposes [18]. Nonetheless, all datasets that used information on the evolutionary relationship between species (phylogenetic or taxonomic trees) did even slightly better as surrogates for overall PD than their counterparts that treated species or genera as equivalent (table 1). There is, therefore, a marginal benefit to using phylogenetic information in conservation planning, and it should be used whenever available. Future studies will need to confirm whether the very slight improvement in surrogacy obtained when taking into account the taxonomic relationships between species (dataset K) is reproducible in other situations. If so, this could be a simple first approximation to incorporating evolutionary history in conservation planning, as basic taxonomic information is typically already available for most described species even if phylogenetic information is not.

Predictably, surrogacy effectiveness declined as we degraded the datasets, by reducing the information they contained in terms of number of species and corresponding detail of the phylogenetic tree (figure 3 and table 2). Real data are typically not only incomplete but also biased spatially [46] and taxonomically [7], unlike randomly selected sets of the same number of species [47,48] which are incomplete but not biased; the former are therefore expected to perform worse as surrogates [6]. The approach we used to prune our dataset used the chronological sequence of species description to replicate some of the taxonomic and geographical biases of real data (for example, large,
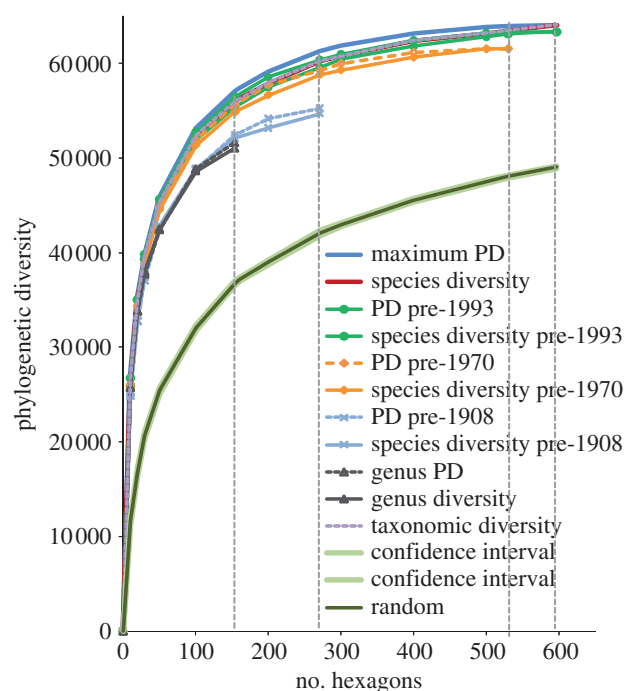

Figure 3. Surrogacy value of 10 measures of mammalian diversity in representing PD. The closer the surrogate curve is to the optimal curve, the higher the surrogacy value. The vertical lines mark the upper limit of the intervals used to calculate SAI (table 2).

conspicuous species tend to be described first) [7]. It is all the more surprising that surrogacy levels were nonetheless high, even for datasets based on the less than one-fourth of all original species described by 1908. However, our method of simulating datasets by simply pruning the current phylogenetic tree does not replicate all types of imperfection that existed in the real earlier data, for example, taxa already described but taxonomically misclassified (e.g. [49]). Furthermore, our species diversity datasets (B, D, F, H) were assumed to be equivalent to their PD counterparts (A, C, E, G) except for the absence of phylogenetic data. In practice, the species data would not be the same without all the phylogenetic studies that underpinned the creation of the mammalian super-tree [16]. Finally, here we have only addressed changes in taxonomic and phylogenetic knowledge, but real data change in other ways, such as the quality of species' distribution data, that may also affect the results of conservation planning [46]. Our results, therefore, underestimate changes in real datasets as knowledge improves, and probably overestimate the surrogacy value of earlier datasets.

It has been proposed that collating data at a higher taxon level may be a useful shortcut for conservation planning of highly diverse taxa/regions [43,44]. From a practical perspective, not only are there inevitably (and usually markedly) fewer higher taxa of a given rank to be counted in an area than there are species, but it is also typically much easier to distinguish between higher taxa than it is among their constituent species. Indeed, trees are usually highly imbalanced, meaning that most higher taxa contain relatively few species and a few contain very many [22]. It is generally easier to distinguish between, say, 100 specimens from different

Table 2. As measured by the SAI index and the coefficient of determination R^2, surrogacy value of each of the 11 biodiversity measures considered in representing phylogenetic diversity. Each column corresponds to an interval (e.g. from 0 to 271 cells) for which the SAI was calculated (the upper limit of each interval is marked by a vertical line in figure 3). Comparisons of SAI value across datasets should be done within each column.

	SAI				R^2
	[0–597]	[0–530]	[0–271]	[0–154]	
A phylogenetic diversity	1.00	1.00	1.00	1.00	1.00
B species diversity	0.95	0.95	0.95	0.95	0.96
C phylogenetic diversity pre-1993	0.96	0.97	0.97	0.98	1.00
D species diversity pre-1993	0.93	0.93	0.93	0.94	0.96
E phylogenetic diversity pre-1970	—	0.88	0.93	0.94	0.99
F species diversity pre-1970	—	0.88	0.90	0.92	0.95
G phylogenetic diversity pre-1908	—	—	0.78	0.81	0.92
H species diversity pre-1908	—	—	0.77	0.82	0.90
I genus phylogenetic diversity	—	—	—	0.81	0.99
J genus diversity	—	—	—	0.81	0.98
K taxonomic diversity	0.96	0.96	0.96	0.97	0.98

genera than 100 specimens from different species because disproportionate effort is expended on differentiating among species within the speciose genera, which are often morphologically very similar [44]. Likewise, one could also predict that higher taxon datasets should be better surrogates for PD than datasets of similar numbers of species. Indeed, it should be expected that, everything else being equal, a set of 100 specimens of different genera should have higher overall PD than a set of 100 species, because the latter should include some closely related species that would contribute little to overall PD representation, while the former approach would ensure representation is better spread across the tree of life. Accordingly, we found that our genus-level phylogenetic tree (dataset I) had a total PD = 30 669.6 MYBP, whereas our tree of pre-1908 species (dataset G) had a PD = 21 518.9 MYBP (table 1), and trees of 1163 randomly selected species had a mean PD = 22 634.1 MYBP (standard deviation = 309.9, $n = 100$).

The higher-taxon approach is, therefore, an indirect way of incorporating the evolutionary relationships between species in conservation planning. Yet, we found that our datasets based on genus information did not perform better than those datasets based on the same number of species (if anything, the pre-1908 data performed better; figure 3 and table 2). As discussed above, when conservation planning is done spatially, rather than by selecting individual species/genera from a tree, results are constrained by the variety of assemblages that exist in nature [18]. Hence, the total PD obtained by, for example, selecting a minimum set of sites representing all pre-1908 species is not just the PD of the corresponding tree (21 518.9 MYBP) but that of all species co-occurring with those species (we found PD = 54 702.8 MYBP). The PD of a set of sites is phylogenetically non-random: it is affected not only by the structure of the phylogenetic tree, but also by the structure of the species spatial distributions, and the relationship between the two (they are not independent because the evolutionary processes of speciation and extinction that shape phylogenetic trees are themselves spatially explicit, simultaneously affecting and being affected

by species' distributions) [18]. For this reason, it is not straightforward to predict that genera should be better surrogates of PD than species in spatial conservation planning, just because genera are better spread across the phylogenetic tree. Future studies are needed to understand if results generalize, and if so whether there is something in the relationship between the structure of the phylogenetic tree and the structure of spatial distributions that may explain it. In any case, given the practical advantages of distinguishing between genera mentioned above, the higher-taxon approach may still be a more cost-effective way of collecting data for conservation planning aiming at maximizing PD, something which should also be addressed in future studies.

Estimates suggest that more than 7000 living species of mammals will eventually be recognized [7], and so the dataset we used and considered to be perfect (i.e. dataset A) is far from so. However, our results suggest that basing decisions on data on the distribution and phylogeny of already known species is likely to produce reliable results for the conservation of broader mammalian diversity. However, as recognized above, the conservation planning problem we considered throughout—maximizing the representation of diversity in a given set of sites—is extremely simplified in relation to the approaches currently developed for the selection of real networks of protected areas, neglecting both ecological dynamics and socioeconomic constraints that affect real-life conservation. Whether this affects the levels of surrogacy tested here remains to be seen, although Bode *et al.* [50] found that incorporating socio-economic considerations into conservation planning at the global scale increased the level of agreement between results based on different taxa.

There are also other reasons why species may be considered non-equivalent in conservation planning besides phylogeny: species that are threatened and/or with very small distributions may require a more targeted conservation investment in order to keep conservation options open [51]. It remains to be tested whether our results hold in more complex conservation scenarios. Even if they do, there are other good reasons for continuing to

refine the existing biodiversity information, for example, by obtaining better information on species' ecology, life-history, population trends and threats in order to support species-based conservation [25] and for the long-term monitoring of change in species status [52]. Here, we have only tested the importance of phylogenetic and taxonomic information to spatial mapping of priority conservation areas, and not the importance of biodiversity datasets in general for various other conservation purposes.

Overall, our results confirm that the better the quality of biodiversity data, the better they can serve as a basis for spatial conservation planning. But we also found that decisions based on incomplete data are remarkably robust to the improvement in quality that inevitably takes place over time in biodiversity data, both as new species are described and as new phylogenetic information becomes available. Furthermore, using simple species data proved nearly as reliable to the representation of mammalian PD as more complex approaches that integrated information on the evolutionary relationships between species (either directly, through PD, or indirectly, through TD and the higher-taxon approach). This is good news for conservation practitioners, because species data are the most common type of information hitherto employed in conservation planning, and will probably remain the case for many taxa that are yet lacking phylogenetic data despite distribution data being available. Given the urgency of conservation action in most parts of the world, our results suggest that conservation planning should make use of the best available systematic data, limited as they may be, instead of waiting for the future availability of better data on taxonomy and phylogeny. This is not to say that the collection of biodiversity data should stop, but that conservation action cannot wait.

REFERENCES

1 Margules, C. R. & Pressey, R. L. 2000 Systematic conservation planning. *Nature* **405**, 243–253. (doi:10.1038/35012251)

2 Balmford, A. & Gaston, K. J. 1999 Why biodiversity surveys are good value. *Nature* **398**, 204–205. (doi:10.1038/18339)

3 Grantham, H. S., Moilanen, A., Wilson, K. A., Pressey, R. L., Rebelo, T. G. & Possingham, H. P. 2008 Diminishing return on investment for biodiversity data in conservation planning. *Conserv. Lett.* **1**, 190–198. (doi:10.1111/j.1755-263X.2008.00029.x)

4 Grantham, H. S., Wilson, K. A., Moilanen, A., Rebelo, T. & Possingham, H. P. 2009 Delaying conservation actions for improved knowledge: how long should we wait? *Ecol. Lett.* **12**, 293–301. (doi:10.1111/j.1461-0248.2009.01287.x)

5 Gaston, K. J. & Rodrigues, A. S. L. 2003 Reserve selection in regions with poor biological data. *Conserv. Biol.* **17**, 188–195. (doi:10.1046/j.1523-1739.2003.01268.x)

6 Rodrigues, A. S. L. & Brooks, T. M. 2007 Shortcuts for biodiversity conservation planning: the effectiveness of surrogates. *Annu. Rev. Ecol. Evol. Syst.* **38**, 713–737. (doi:10.1146/annurev.ecolsys.38.091206.095737)

7 Reeder, D. M., Helgen, K. M. & Wilson, D. E. 2007 Global trends and biases in new mammal species discoveries. *Occ. Papers Museum Texas Tech. Univ.* **269**, 1–35.

8 Helgen, K. M., Kock, D., Gomez, R. K. S. C., Ingle, N. R. & Sinaga, M. H. 2007 Taxonomy and natural history of the Southeast Asian fruit-bat genus *Dyacopterus*. *J. Mammal.* **88**, 302–318. (doi:10.1644/06-MAMM-A-276R.1)

9 Teeling, E. C., Springer, M. S., Madsen, O., Bates, P., O'Brien, S. J. & Murphy, W. J. 2005 A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* **307**, 580–584. (doi:10.1126/science.1105113)

10 May, R. M. 1990 Taxonomy as destiny. *Nature* **347**, 129–130. (doi:10.1038/347129a0)

11 Vane-Wright, R. I., Humphries, C. J. & Williams, P. H. 1991 What to protect—systematics and the agony of choice. *Biol. Conserv.* **55**, 235–254. (doi:10.1016/0006-3207(91)90030-D)

12 Collen, B., Turvey, S. T., Waterman, C., Meredith, H. M. R., Kuhn, T. S., Baillie, J. E. M. & Issac, N. J. B. 2011 Investing in evolutionary history: implementing a phylogenetic approach for mammal conservation. *Phil. Trans. R. Soc. B* **366**, 2611–2622. (doi:10.1098/rstb.2011.0109).

13 Wilson, D. E. & Reeder, D. M. 2005 *Mammal species of the world: a taxonomic and geographic reference*, 3rd edn. Baltimore, CA: The Johns Hopkins University Press.

14 Faith, D. P. 1992 Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10. (doi:10.1016/0006-3207(92)91201-3)

15 Faith, D. P. 1994 Phylogenetic diversity: a general framework for the prediction of feature diversity. In *Systematics and conservation evaluation* (eds P. I. Forey, C. J. Humphries & R. I. Vane-Wright), pp. 251–268. Oxford, UK: Claredon Press.

16 Bininda-Emonds, O. R. P. *et al.* 2007 The delayed rise of present-day mammals. *Nature* **446**, 507–512. (doi:10.1038/nature05634)

17 Fritz, S. A., Bininda-Emonds, O. R. P. & Purvis, A. 2009 Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecol. Lett.* **12**, 538–549. (doi:10.1111/j.1461-0248.2009.01307.x)

18 Rodrigues, A. S. L., Brooks, T. M. & Gaston, K. J. 2005 Integrating phylogenetic diversity in the selection of priority areas for conservation: does it make a difference? In *Phylogeny and conservation* (eds A. Purvis, J. L. Gittleman & T. M. Brooks), pp. 101–119. Cambridge, UK: Cambridge University Press.

19 Mace, G. M., Gittleman, J. L. & Purvis, A. 2003 Preserving the tree of life. *Science* **300**, 1707–1709. (doi:10.1126/science.1085510)

20 Isaac, N. J. B., Mallet, J. & Mace, G. M. 2004 Taxonomic inflation: its influence on macroecology and conservation. *Trends Ecol. Evol.* **19**, 464–469. (doi:10.1016/j.tree.2004.06.004)

21 Purvis, A., Agapow, P.-M., Gittleman, J. L. & Mace, G. M. 2000 Nonrandom extinction and the loss of evolutionary

history. *Science* **288**, 328–330. (doi:10.1126/science.288.5464.328)

22 von Euler, F. 2001 Selective extinction and rapid loss of evolutionary history in the bird fauna. *Proc. R. Soc. Lond. B* **268**, 127–130. (doi:10.1098/rspb.2000.1340)

23 Purvis, A. 2008 Phylogenetic approaches to the study of extinction. *Annu. Rev. Ecol. Evol. Syst.* **39**, 301–319. (doi:10.1146/annurev.ecolsys.063008.102010)

24 Redding, D. W. & Mooers, A. O. 2006 Incorporating evolutionary measures into conservation prioritization. *Conserv. Biol.* **20**, 1670–1678. (doi:10.1111/j.1523-1739.2006.00555.x)

25 Isaac, N. J. B., Turvey, S. T., Collen, B., Waterman, C. & Baillie, J. E. M. 2007 Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS ONE* **2**, e296. (doi:10.1371/journal.pone.0000296)

26 Polasky, S., Csuti, B., Vossler, C. A. & Meyers, S. M. 2001 A comparison of taxonomic distinctness versus richness as criteria for setting conservation priorities for North American birds. *Biol. Conserv.* **97**, 99–105. (doi:10.1016/S0006-3207(00)00103-8)

27 Rodrigues, A. S. L. & Gaston, K. J. 2002 Maximising phylogenetic diversity in the selection of networks of conservation areas. *Biol. Conserv.* **105**, 103–111. (doi:10.1016/S0006-3207(01)00208-7)

28 Forest, F. *et al.* 2007 Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* **445**, 757–760. (doi:10.1038/nature05587)

29 Kirkpatrick, J. B. 1983 An iterative method for establishing priorities for the selection of nature reserves—an example from Tasmania. *Biol. Conserv.* **25**, 127–134. (doi:10.1016/0006-3207(83)90056-3)

30 Margules, C. R., Nicholls, A. O. & Pressey, R. L. 1988 Selecting networks of reserves to maximise biological diversity. *Biol. Conserv.* **43**, 63–76. (doi:10.1016/0006-3207(88)90078-X)

31 Moilanen, A., Wilson, K. A. & Possingham, H. (eds) 2009 *Spatial conservation prioritization: quantitative methods and computational tools.* Oxford: Oxford University Press.

32 Schipper, J. *et al.* 2008 The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science* **322**, 225–230. (doi:10.1126/science.1165115)

33 IUCN 2010 IUCN Red List of Threatened Species. Version 2010.3. See http://www.iucnredlist.org/.

34 Gaston, K. 1994 *Rarity.* London, UK: Chapman & Hall.

35 Hoffmann, M. *et al.* 2010 The impact of conservation on the status of the world's vertebrates. *Science* **330**, 1503–1509. (doi:10.1126/science.1194442)

36 Sahr, K., White, D. & Kimerling, A. J. 2003 Geodesic discrete global grid systems. *Cartogr. Geogr. Inform. Sci.* **30**, 121–134.

37 Steel, M. & Mooers, A. 2010 The expected length of pendant and interior edges of a Yule tree. *Appl. Math. Lett.* **23**, 1315–1319. (doi:10.1016/j.aml.2010.06.021)

38 Church, R. L., Stoms, D. M. & Davis, F. W. 1996 Reserve selection as a maximal covering location problem. *Biol. Conserv.* **76**, 105–112. (doi:10.1016/0006-3207(95)00102-6)

39 Achterberg, T. 2009 SCIP: solving constraint integer programs. *Math. Prog. Comp.* **1**, 1–41. (doi:10.1007/s12532-008-0001-1)

40 Czyzyk, J., Mesnier, M. P. & Moré, J. J. 1998 The NEOS Server. *Comp. Sci. Eng.* **5**, 68–75.

41 Wisconsin Institutes for Discovery. 2011 NEOS Server. See http://www.neos-server.org.

42 Wilson, D. E. & Reeder, D. M. 1993 *Mammal species of the world: a taxonomic and geographic reference*, 2nd edn. Washington, DC: Smithsonian Institution Press.

43 Gaston, K. J. & Williams, P. H. 1993 Mapping the world's species—the higher taxon approach. *Biodiv. Lett.* **1**, 2–8. (doi:10.2307/2999642)

44 Gaston, K. J. 2000 Biodiversity: higher taxon richness. *Prog. Phys. Geogr.* **24**, 117–127.

45 Ferrier, S. & Watson, G. 1997 *An evaluation of the effectiveness of environmental surrogates and modelling techniques in predicting the distribution of biological diversity.* Armidale NSW: Environment Australia.

46 Grand, J., Cummings, M. P., Rebelo, T. G., Ricketts, T. H. & Neel, M. C. 2007 Biased data reduce efficiency and effectiveness of conservation reserve networks. *Ecol. Lett.* **10**, 364–374. (doi:10.1111/j.1461-0248.2007.01025.x)

47 Andelman, S. J. & Fagan, W. F. 2000 Umbrellas and flagships: efficient conservation surrogates or expensive mistakes? *Proc. Natl Acad. Sci. USA* **97**, 5954–5959. (doi:10.1073/pnas.100126797)

48 Manne, L. L. & Williams, P. H. 2003 Building indicator groups based on species characteristics can improve conservation planning. *Anim. Conserv.* **6**, 291–297. (doi:10.1017/S1367943003003354)

49 Rueness, E. K., Asmyhr, M. G., Sillero-Zubiri, C., Macdonald, D. W., Bekele, A., Atickem, A. & Stenseth, N. C. 2011 The cryptic African wolf: *Canis aureus lupaster* is not a golden jackal and is not endemic to Egypt. *PLoS ONE* **6**, e16385. (doi:10.1371/journal.pone.0016385)

50 Bode, M., Wilson, K. A., Brooks, T. M., Turner, W. R., Mittermeier, R. A., McBride, M. F., Underwood, E. C. & Possingham, H. P. 2008 Cost-effective global conservation spending is robust to taxonomic group. *Proc. Natl Acad. Sci. USA* **105**, 6498–6501. (doi:10.1073/pnas.0710705105)

51 Rodrigues, A. S. L., Gray, C. L., Crowter, B. J., Ewers, R. M., Stuart, S. N., Whitten, T. & Manica, A. 2010 A global assessment of amphibian taxonomic effort and expertise. *BioScience* **60**, 798–806. (doi:10.1525/bio.2010.60.10.6)

52 Butchart, S. H. M. *et al.* 2007 Improvements to the Red List Index. *PLoS ONE* **2**, e140. (doi:10.1371/journal.pone.0000140)