Achieving the Convention on Biological Diversity's Goals for Plant Conservation
L. N. Joppa et al.
*Science* 341, 1100 (2013);
DOI: 10.1126/science.1241706

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of September 5, 2013):

Updated information and services, including high-resolution figures, can be found in the online version of this article at: http://www.sciencemag.org/content/341/6150/1100.full.html

Supporting Online Material can be found at: http://www.sciencemag.org/content/suppl/2013/09/04/341.6150.1100.DC1.html

This article cites 22 articles, 9 of which can be accessed free: http://www.sciencemag.org/content/341/6150/1100.full.html#ref-list-1
Achieving the Convention on Biological Diversity’s Goals for Plant Conservation

L. N. Joppa, P. Visconti, C. N. Jenkins, S. L. Pimm

Identifying which areas capture how many species is the first question in conservation planning. The Convention on Biological Diversity (CBD) aspires to formal protection of at least 17% of the terrestrial world and, through the Global Strategy for Plant Conservation, 60% of plant species. Are these targets of protecting area and species compatible? We show that 67% of plant species live entirely within regions that comprise 17% of the land surface. Moreover, these regions include most terrestrial vertebrates with small geographical ranges. However, the connections between the CBD targets of protecting area and species are complex. Achieving both targets will be difficult because regions with the most plant species have only slightly more land protected than do those with fewer.

Protected areas are broadly effective (1–6) and thus usually necessary, if not always sufficient, to protect species. Their effective allocation is vital to slow present extinction rates that are 100 to 1000 times faster than the natural background rate (7). The Convention on Biological Diversity’s (CBD’s) 20 Aichi Targets—agreed in October 2010 in Nagoya, Japan—extend to 2020 an international commitment to halt biodiversity loss (www.cbd.int/sp/targets/). Target to 2020 an international commitment to halt biodiversity loss (www.cbd.int/sp/targets/). Target 11 seeks formal protection of >17% of the terrestrial world. One can use the most taxonomically complete and spatially best-resolved data on species distributions (8) —birds (9), mammals (10), and amphibians (10) —to assess this target (11). Unfortunately, these are a taxonomically limited subset of ~23,000 terrestrial species from more than a million described animal species with many more as-yet unknown (12). This raises concerns about their representativeness for setting global conservation priorities. By contrast, plants form a large taxonomic sample with >350,000 described species and ~15% awaiting description (13). Important in themselves, plants influence the diversity of insects (14) and other animals. In 2010, the CBD updated the Global Strategy for Plant Conservation (GSPC) (11), seeking protection for 60% of plant species as a critical indicator toward CBD goals. Are these targets of protecting area and species compatible?

Satisfying the joint aspirations of the CBD’s Aichi Target 11 and the GSPC will be difficult. First, by 2009, the world had protected ~13% of global land area (15), but half of the world’s major habitat divisions—ecoregions (16)—did not meet a target of 10% coverage. Some ~75% of them had <10% of their area strictly protected (15). Present conservation efforts bias toward lands that are high, cold, dry, or otherwise far from people—often a mismatch with where conservation needs are pressing (17). These statistics show that protected areas are not representative of terrestrial environments, but they do not address species targets directly.

Second, for plants, as for most taxa, unresolved issues of taxonomy generate uncertainties in how many species there are amid existing catalogues of described species and how many are still missing from them (13, 18). Target 1 of the GSPC is to complete “a widely accessible working list of all known plant species, as a step toward a complete world flora” (19). Major international botanical gardens responded in 2010 with “The Plant List” (www.theplantlist.org)—a working list of all known plant species. Here, we analyze a subset of ~109,000 species taken from the World Checklist of Selected Plant Families (WCSP) (20). For this subset, literature compilers and taxonomists have attempted a consensus overview of the current state of knowledge of select families, including correct names for currently accepted species and their synonyms. For this quarter of the world’s plant species, we previously predicted where as-yet undescribed species likely live (18).

Third, species’ distributional data are imperfect (21). Elsewhere, we map birds, mammals, and amphibians on a scale of 10 km by 10 km (8). We compare them to plant distributions below. The details of these animal distributions are exceptional; plant distributional data are coarser. Flowering plant species in the WCSP are tagged to one or more of the 369 countries or geographic regions delineated by the International Taxonomic Database Working Group (22). Fur details on the species and regions are in the supplementary materials (23). Nonetheless, this spatial scale captures the essential first step of comparing targets of area protected to species protected.

More problematic is that plant distributional data are species lists from regions where the largest region is 2 million times the area of the smallest. For the biodiversity hotspots of Myers et al. (24), the ratio is ~130. The relationships of numbers of species (S) to area (A) are well described by S = cA^z; c and z are parameters. Because z is <1, species densities, S/A, generally decline with increasing area. This makes objective comparisons of areas—and the designation of conservation priorities—challenging. On the basis of species’ totals alone, apparent priorities tend to be the largest regions; those based on species’ densities, the smallest ones. Further complicating matters, z depends on circumstance: Islands, continuous areas within continents, and biogeographically unrelated regions have different characteristic values (25). This fact dashes hopes of a single, global correction of species’ numbers by a simple function of area to permit regional comparisons.

We can address this issue directly, because our results fall from quantitative databases and not the expert opinions used by Myers et al., which are impossible to replicate or update. Our solution uses a greedy algorithm to accumulate species found only within a progressively larger set of regions (“endemic densities”; Fig. 1 and table S1b). We scale our results to 100,000 plant species and to 1000 km^2.

Regions with the highest densities enter first, followed by those adding progressively fewer new species to the aggregated total. Thus, the first 43 regions to enter are all islands, followed by Costa Rica. In the data that we consider, Costa Rica has 791 endemics and adds all of these, reducing the accumulated endemic density to 29. Panama enters next. It has 775 endemics, but adds

Acknowledgments: This paper is dedicated to Helmut Schwarz on the occasion of his 70th birthday. We thank D. Avni, T. Isaaq, and F. Lichtenhaler for discussion; M. Meckel for help with some figures; and A. Czasch for support concerning data analysis. This work was supported by the State Initiative for the Development of Scientific and Economic Excellence (LOEWE) in the LOEWE-Fokus ECH. The momentum data used to draw the figures are provided in the supplementary materials. Raw data are archived at the University of Frankfurt and are available upon request.

Supplementary Materials

www.sciencemag.org/cgi/content/full/341/6150/1096/DC1

*Corresponding author. E-mail: stuartpimm@me.com

References (25)

Movies S1 to S5

10 May 2013; accepted 2 August 2013
10.1126/science.1240362

© 2013 by The American Association for the Advancement of Science

Downloaded from www.sciencemag.org on September 5, 2013
1309 species to the accumulated total because additional species are endemic to Panama and Costa Rica combined. The accumulated endemic density of these 45 regions is now 25.

We obtain a broadly similar map when optimizing for species richness instead of number of endemic species (“richness densities”; fig. S1 and table S2).

As much as the discrete and variable areas of the plant regions permit, the advantage of greedy algorithms is that they create a continuous curve of increasing areas and species. Figure 2 shows the accumulation curve optimized to capture species only found within an accumulated set of regions (endemics) and those incidentally captured (all species) by this set.

Unlike the biodiversity hotspot approach in which areas and species are either included or not, our accumulations provide a continuous ranking on which the Aichi targets are entirely arbitrary benchmarks. Nonetheless, considering the 17% target—an area of ~24.3 million km²—allows useful comparisons.

First, the regions that we select encompass part of the ranges of 81% of the plant species and all of the ranges of 67%. We are not suggesting that we could protect all of these regions and nothing else, but this hypothetical 17% sets the bar for achieving the GSPC goal via CBD Target 11, as shown by the red crossed lines in Fig. 2. (By optimizing just for species richness, a similar set of regions captures 86% of species in the same total area; see fig. S2 and table S1a).

Second, for any benchmark, we must ask whether our greedy accumulation algorithm is optimal. Moreover, to what extent do other almost-as-good solutions surround this optimum? The supplementary materials detail our use of a genetic algorithm to answer these questions.

Third, our selected regions are important for terrestrial vertebrates. We find that 89% of bird species, 80% of amphibians, and 74% of...
mammals live within them. Percentages for species with ranges smaller than the median geographical size—those with a much higher risk of being threatened—are 88, 82, and 73%, respectively (Fig. 3).

Fourth, biodiversity hotspots cover only ~17.4 million km$^2$ and capture 44% of endemic species (24). Not surprisingly, our formal optimization performs better, capturing 59% in the same amount of area. It also captures 74% of all plant species, a number Myers et al. could not estimate. What is notable is not the better performance of our approach, but the similarities and differences in the regions chosen between the two approaches (Fig. 4).

The areas that we omit have much to do with spatial resolution. Hotspots include montane forest of East Africa and the Western Ghats of India, plus southwestern Australia, and the coastal moist forests of West and East Africa. These areas are not included in our selections because they are embedded in much larger regions. The finer-scale vertebrate data also show the importance of these regions, plus others including the forests along Australia’s eastern seaboard (Fig. 3). Conversely, Myers et al. considered levels of habitat loss and so excluded

**Fig. 3. Plant regions and endemic vertebrates.** The distribution of species of birds, mammals, and amphibians (color-coded) with smaller than the median geographical range (8) against the regions with 17% of the terrestrial surface and 67% of plant species as endemics (dark gray).

**Fig. 4. A comparison of this paper’s identified regions (blue) and biodiversity hotspots (24).**
New Guinea, which we do not. Our inclusion of Turkey and the Middle East, Japan, Korea, and more of China, northern South America, and Southern Africa reflects our estimates of their numbers of endemic species.

Our results have five consequences.

1) Broadly, tropical and subtropical islands, moist tropical and subtropical forests (especially those in mountains), and Mediterranean ecosystems hold concentrations of plant endemics. The majority of as-yet undescribed plant species also live in these regions (18). Further discoveries would likely enhance their importance.

2) Our figures suggest the achievement of Aichi goals by concentrating protected areas in regions of highest endemism. Had nations already implemented this strategy, we would see proportionally greater protection rates where species densities are high. Figure 2 shows that, within the regions we select, strict protected areas (IUCN I to VI plus indigenous territories), the most important areas include Costa Rica and Panama, which have >10% of their land in IUCN classes I and II. This is a weak trend, however. When considering all the categories of protection (IUCN I to VI plus indigenous territories), the total protected is much higher, but the trend similar.

3) The ability of protected areas to protect depends upon the nature and location of threats (2, 3). For example, within our 17% set, 9.4% of plant species are endemic to a total of ~1.9 million km² of islands. Island plants suffer greatly from introduced species (27), something the establishment of protected areas does not completely address.

4) Figure 2 (and see table S1b) show a hitherto poorly appreciated effect of indigenous areas. In tropical South America, these protect large areas of tropical moist forest with high plant richness.

5) Overall, the global land area currently protected, ~13%, is close to Aichi’s 17%. This seems encouraging. However, of the 17% that contains the entire range of 67% of the world’s plant species, only 14% is protected in some way, barely more than the global average.

The total area protected imperfectly measures species’ protection, however. Even with perfect data on species’ distributions, the “Noah’s Ark effect” (28) renders simple optimal allocation of priority areas meaningless. A small total area—a metaphorical “ark”—can capture many species but ignore long-term viability. Numerous protected areas of large aggregate size may house many species, but be individually too small to maintain viable populations. How small is “too small” depends on the species—tigers demand more area than tiger lilies—as well as the distribution of habitat fragments (29) and levels of threat (2, 5). How much area countries should protect—and where—are ecological questions. Political priorities dominate actions, as the Aichi target of 17% testifies.

The spatial resolution of presently available data is inadequate to address this key concern at the spatial scales at which conservation actions are taken and protected areas established. Nonetheless, we show that how protected areas are allocated within and across regions constrains how efficiently plant diversity can be sustained. Understanding this is necessary to achieve the Convention on Biological Diversity’s conservation goals.

Caffeoyl Shikimate Esterase (CSE)

Is an Enzyme in the Lignin Biosynthetic Pathway in Arabidopsis

Ruben Vanholme,1,2† Igor Cesarino,1,2† Katarzyna Rataj,3 Yuguo Xiao,5 Lisa Sundin,1,2 Geert Goeminne,1,2 Hoon Kim,4 Joanna Cross,3,2 Kris Morreel,1,2 Pedro Araujo,3,2 Lydia Welsh,3 Jurgen Haustraete,5 Christopher McClellan,7 Bartel Vanholme,1,2 John Ralph,5 Gordon G. Simpson,3,6 Claire Halpin,5,6† Wout Boerjan1,2,†§

Lignin is a major component of plant secondary cell walls. Here we describe caffeoyl shikimate esterase (CSE) as an enzyme central to the lignin biosynthetic pathway. Arabidopsis thaliana cse mutants deposit less lignin than do wild-type plants, and the remaining lignin is enriched in p-hydroxyphenyl units. Phenolic metabolite profiling identified accumulation of the lignin pathway intermediate caffeoyl shikimate in cse mutants as compared to caffeoyl shikimate levels in the wild type, suggesting caffeoyl shikimate as a substrate for CSE. Accordingly, recombinant CSE hydrolyzed caffeoyl shikimate into caffeate. Associated with changes in lignin, the conversion of cellulose to glucose in cse mutants increased up to fourfold as compared to that in the wild type upon saccharification without pretreatment. Collectively, these data necessitate the revision of currently accepted models of the lignin biosynthetic pathway.

The evolutionary emergence of lignin, a phenolic polymer deposited in the secondary cell wall, allowed the development of vascular land plants. The hydrophobic and strengthening nature of lignin enables conducting xylem vessels to transport water and nutrients from the
Supplementary Materials for

Achieving the Convention on Biological Diversity’s Goals for Plant Conservation

L. Joppa, P. Visconti, C. N. Jenkins, S. L. Pimm*

*Corresponding author. E-mail: stuartpimm@me.com

Published 6 September 2013, Science 341, 1100 (2013)
DOI: 10.1126/science.1241706

This PDF file includes:

Materials and Methods
Figs. S1 to S3
References

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/341/6150/1100/DC1

Table S1 (Microsoft Excel)
Supplementary Materials

Materials and Methods

The Dataset

We analyze a subset of the species found within the online database ‘The Plant List’ www.theplantlist.org. This is a working list of all known plant species first released in 2010 to coincide with the completion of the first phase of the GSPC. We analyzed 108,718 species from The Plant List, taken from the World Checklist of Selected Plant Families (WCSP) (20). For this subset of families, literature compilers and taxonomic specialists have attempted to present a consensus overview of the current state of knowledge of the family, including correct names for currently accepted species and their synonyms, in order to facilitate conservation action. This is a major but necessary undertaking to avoid the serious problem of counting a single species multiple times. This dataset is the same that we analyzed previously in an effort to estimate how many missing species there are and where those species are likely to be discovered (13,18).

Important for our purposes, species in the WCSP are tagged to one or more of the 369 Level 3 geographic regions delineated by the International Taxonomic Database Working Group’s (TDWG) world geographic scheme (22). Three of these regions (BOU, DSV, SSA) had no species associated with them in our dataset and were removed from the analysis, leaving 366 unique regions. Additionally, 38 records in our dataset (corresponding to 37 unique species) had 24 region codes absent from
the GIS layer for the geographic regions. We interpreted these as coding errors and removed the entries from the dataset. This did not affect the species count, as all implicated species occurred in other regions.

We took all data on protected areas from the most current available data on protected areas (26).

*How representative are the species we include in our analyses?*

First, consider whether this sample of plant species represents the roughly twice as numerous plant species for which their process of consensus overview is not yet complete. To assess this, we recorded the total number of genera and families reported for each of the 52 TDWG Level 2 regions, such as “Southwestern Europe”, that cluster the 366 TDWG Level 3 regions we describe above. Then, we compare these to the total numbers of genera and families in these major regions from the Kew Gardens summaries [http://www.kew.org/science/family-genera/](http://www.kew.org/science/family-genera/). Figure S1 shows the strong correlations between the numbers of families and genera in our samples and the totals. In both cases, there are roughly three times more genera and families in total than in the samples.

One could easily imagine otherwise: some regions might have fewer families and genera in the sample than others, relative to the total numbers known to occur there. This would likely lead to their having fewer species as well, making them less likely to be a conservation priority than in fact they should be. The strong correlations in Figure S1 reject this, at least at the level of these 52 major regions. To our
knowledge, there are no data that would allow us to check the actual numbers of species in the 366 regions against the numbers in our samples.

Second, we can ask whether our sample represents where we expect as-yet undescribed species to live. Clearly, if predictions of missing species were in regions that currently have low estimates of endemic plants, then our selections would be incorrect. In fact, the converse is true. Previous analyses predict that these missing species are likely to be in the same places that this paper shows are priorities for conservation (18).

The Algorithm

We accumulate species with increasing land area using a greedy algorithm that approximates the maximum number of species for any given percentage of total area. Our analysis for endemics started with the region that maximized the ratio of endemic species (those found entirely within a region) divided by the area of that region. For our endemics solution that region was Norfolk Island, with 41 endemic species within 63 km$^2$. We next chose the region that, when added to the regions already selected, added the most combined species with the least amount of combined area. The method continued until all species were included in our set of regions. Crucially, the number of species added was often not just the number of species endemic to the region itself, but endemic to it plus the set of regions previously selected. We present an example below.

Our richness algorithm mirrored that for endemics, only we did not require species
to be entirely constrained within any given set of regions, but simply to occur in at least one of them. In Figure S2, we map the results for species richness in an identical manner to Figure 1 in the main text.

We report the results for each step of our algorithm for both species richness (Table S1a) and endemic richness (Table S1b). Column A reports the iteration of the algorithm at which each region entered. The regions enter in similar, but not identical sequences in the tables of richness (1a) and endemics (1b).

In both tables, Columns B – D report the names and other geographic considerations of each region with respect to the TDWG standards, and Column E is our assessment of whether they are islands or mainland areas. Column F is the area of each region, column G the cumulative area of the set, and column H the cumulative percentage area of the land surface. Columns O – U report the area (in \( \text{km}^2 \)) of each region occurring in different classes of IUCN protected area, or without any protection at all (column N).

In Table S1a (richness), column I shows the number of species found in each region, column J shows the number of new species added to the accumulation, and column K shows the accumulated total. Thus, Norfolk Island has 145 species, and adds 144 new species to the 12 species found on Selvagens, for a total of 156. One species is common to both places. Column L shows the percentage of all species accumulated.

In Table S1b (endemics), column I shows the numbers of endemics found in each region, column J shows the number of new species added to the accumulated set
with the addition of a region, column K the cumulative number of endemics, and column L the percentage of all species accumulated. Thus, Norfolk Island has 41 endemic species, while Selvagens has one.

For Table S1b, the minimum value that column J can take is equal to the value in column I, because one always adds the region endemics. However, column J can be considerably greater than column I because some species may not be endemic to a region while still being endemic to several regions combined. Thus, the fifth region to enter the endemics’ solution is Kazan-retto Island, with three endemics. Its addition brings in a total of 10 endemics, however, because it shares seven endemics with the nearby Ogasawarashotto (also called the Bonin Islands), which is already in the accumulated set of regions.

In both tables, Column M reports the scaled density measure (scaled to 100,000 plant species and to 1,000 $\text{km}^2$) of the region considering all other regions previously included by the algorithm. In Table S1b (endemics), columns V – Z report the richness results discussed in the main text, where we used the set of regions chosen when optimizing for endemic species.

The accumulation curves

The difficulty in comparing the numbers of plant species and endemics in each region arises from their widely different areas. Species numbers increase with increasing area and do so in a nonlinear way, as the main text explains. Species ($S$) to area ($A$) relationships are well described by $S = cA^z$, offering the hope that one
might correct $S$ for $A$ in some way and so compare areas of very different sizes. The
difficulty, however, is that there are several species to area curves with different
values of $z$. Islands within the same archipelago have typical values of 0.25,
contiguous continental areas values of $< 0.1$, and unrelated regions close to 1.0. The
regions we consider fall into all three of these broad classes.

Because of this, we chose accumulation curves for they allow comparisons across a
range of values of total area. As table S1b (endemics) shows, the first 43 regions that
enter are islands. These 43 regions mostly bring just their region specific endemics
to the total. A significant counter-example involves islands in the Indian Ocean —
Mauritius, Reunion, Aldabra, and islands in the Mozambique Channel — that enter
at positions 16, 17, 18, and 20. Most obviously, Reunion has 144 endemics, but its
addition adds another 51 species that are shared only with Mauritius, 2 shared only
with Rodrigues, 5 shared with Rodrigues and Mauritius, and 1 shared with
Mauritius and the Seychelles.

The first two continental regions to enter the table for endemics are Costa Rica and
Panama in positions 44 and 45 respectively. They have 791 and 775 endemic
species respectively, but the addition of Panama adds another 534 on top of this.
That is, Panama adds 1,309 species total to the accumulated set because of its own
endemics (775), plus the species that it shares only with Costa Rica (533), and one
additional species found in these two countries and a region already in the
accumulated set.

Table S1a (richness) also chooses Selvagens and Norfolk Islands first, but then
selects Washington, DC, USA. Rhode Island follows as the next mainland choice in position 33. Neither region has endemics and their high ranking for species simply reflects their small size.

The values we map in Figure 1 (main text) and Figure S2 show accumulated species densities. This is the accumulated species divided by the accumulated total area, but with a correction. For example, Costa Rica and Panama have 791 and 775 endemics in areas of 4,600 and 5,909 km² respectively. Costa Rica thus has 0.172 species per km² (= 791 / 4600). We could use this value to compare it with other regions, such as Panama, which has of 0.131 species per km² (= 775 / 5909), but there is a price.

First, no other region has the same area, and we have already discussed the problems that this creates. Second, as we have shown, adding Panama to Costa Rica creates a region with many species endemic to the combined countries but not to either region independently.

Table S1b (endemics) shows a scaled endemic density of 28.8 for Costa Rica. To reach this value we take the number of species endemic to Costa Rica and all 43 regions before it on the list (4,600 species) divided by the total area accumulated (147,164 km²). We then standardize the resulting value to a total of 100,000 species and 1000 km². Thus,

\[
\left( \frac{4,600 \text{ endemic species}}{147,164 \text{ km}^2} \right) \times \left( \frac{100,000 \text{ species}}{108,718 \text{ species}} \right) \times 1,000 \text{ km}^2 = 28.8
\]

When Panama adds then its scaled endemic density will be 24.5, a value that
considers Costa Rica and the previous 43 regions. In this process, one loses the ability to compare single regions of different sizes (and the problems that causes!) and gains the ability to compare species accumulation curves.

**Cross checking our results with a genetic algorithm**

Accumulating regions in an orderly sequence based on maximizing the numbers of species found in the minimum amount of area produces an optimal accumulation curve. It does so in a simple, intuitive way. It does not guarantee the optimal number of species for any chosen amount of area, however. For example, progressive accumulation might add regions in the sequence A, B, C, D. It does not allow backwards elimination that might remove region A once C and D are in the mix to produce a greater ratio of species to total area.

To evaluate the efficiency of our simple “greedy algorithm”, we used a Genetic Algorithm approach (R package genalg, function ‘rbga.bin’) to seek an optimal solution for each target. While the genetic algorithm itself does not guarantee to find the optimal solution, it is likely to have found a solution that is at least close to it, provided it consistently converges to the same set when starting from different random starting points. As we shall discuss presently, it also generates a cloud of next-best solutions that contain useful information about conservation priorities. Genetic algorithms are a familiar approach for obtaining optimal solutions, so our discussion here is brief. The methods involve several steps.

1. We start with a population of 400 samples that the genetic algorithm jargon
calls “chromosomes.” Each of these has 366 regions, called “alleles.” For each chromosome, some 10% of these are chosen at random to be one, meaning we have selected that region, while the rest are zero.

2. “Fitness” is the sum of the endemic species in the combined regions selected on each chromosome. However, if the total area of all regions represented on the ‘chromosome’ is > 17%, we applied a penalty of one species for each 300 km² in excess of the target area.

3. We then select the 20% of the chromosomes with the highest fitness, and then pass these 80 “elite” chromosomes unchanged to the next “generation”.

4. To create the remaining 320 of the next generation, we pick two chromosomes, with replacement, from the pool of 400. We chose pairs of chromosomes randomly but with a probability that increases with each chromosome’s fitness. Next, we pick from a uniform, random variable, to decide where the chromosomes will “cross over”. The single “offspring” might then get, for example, alleles 1 to 99 from one chromosome and 100 to 366 from the other. We repeat this process 320 times.

5. We then mutate 0.5% of the total alleles, which can flip either zero to one, or one to zero.

6. We repeat steps 3 to 5 for 100 generations.

We now have 400 chromosomes — optimized solutions. However, only 113 were
between 16.5 and 17%. One of these is the best.

*Are our results optimal?* We compared targets of 5%, 10%, 15%, 17%, 20%, 25%, and 30% of land area and optimized species captured by those areas using this method just described. The solutions were not exactly comparable given the differences in actual area covered by the selected regions. For example, at the 5% target, the optimal set of regions selected by the genetic algorithm covered 5.5% while the greedy algorithm set covered 5.02% of the total land area.

In Figure S3, we plot the optimal solutions on top of the accumulation curves from the global solution originally shown in Figure 1. Reassuringly, optimal solutions are indistinguishable between the two approaches. The actual sets of regions were also extremely similar, with Jaccard similarity values between the two approaches (greedy and genetic algorithm from 5% to 30% targets, in 5% increments, including our 17% target) of 0.91, 0.95, 0.99, 0.99, 0.97, 0.99, and 0.98 for species richness solutions, and 0.99, 0.95, 0.99, 0.97, 1, 1, and 1 for endemic solutions. We defined Jaccard similarity as the number of shared regions between the two solutions divided by the number of unique regions included in both solutions.

*How special is our 17% optimum?* Selection of a single optimum set as “the priority” — such as the Myers et al. hotspots — inevitably raises concerns for the areas excluded. For the areas excluded by our ranking: by how much did they fail?

We answered this problem with two approaches. First, we produce maps, associated figures, and tables that show *progressive* inclusions. As it were, in figure 1, one
would start with the red areas as the highest priorities, then add on the orange ones, and proceed down. We drew a line at 17\% by coloring areas up to that point. Importantly, one can proceed to accumulate more species by working from dark grey through lighter shades of grey. Likewise, figure 2, of course, shows the complete accumulation. In short, we know what the optimal sets are for an increasing fraction of total area protected.

Second, we can ask how special is the optimum at a given point — say 17\% of total area. By analogy, we have found the solution with the “highest fitness” in the jargon of natural selection. It is “the peak”, in the language of fitness landscapes, the very best set of regions to grab the maximum number of endemic species.

We would like to know how pointed this “peak” is — is it a Matterhorn, sharp and distinct with nothing as high nearby — or a hill, equally high in this landscape metaphor, but surrounded by many other solutions that are almost as good.

Because regions are discrete, it is impossible to get a combination of areas that is exactly 17\%. As table S1 shows, the accumulated areas jump in discrete steps. So we retain only those solutions that encompass between 16.5\% and 17\% of the total area. We then report in what fraction of the final sample a particular region appears. We report these data in table S1b in column AA.

The Cape Floristic Province appears in every solution, for example. The top 100 sites in our greedy accumulation solution are picked better than 90\% of the time by the genetic algorithm, with the exception of a few places, such as the Balearic Islands.
that have small areas — and high densities of endemics per area.

Our stopping point for <17% of land area comes out as 16.54 percent in our greedy algorithm. Of the places just outside this solution, the genetic algorithm generally picks them <13% of the time. The exceptions are Western Australia (picked 65% of the time), Gabon (94%), Equatorial Guinea (66%), Cameroon (70%), Rwanda (69%), Tanzania (70%), and the Australia provinces of Victoria (69%) and New South Wales (71%). These regions are all shaded in the darkest shade of grey in figure 1. Were we to have taken 25% of the total land area, not 17%, then all these regions would have been in the solution.

So, in the metaphor of landscapes, these exceptions are high peaks — areas that sometimes fall in excellent, if not quite the best, selections of priorities. Setting 25% of the land surface as a priority would encompass these and only rarely pick others.

_Different Approaches – Different Regions: A Regional Explanation_

The correspondence between Figure 1 and Figure S2 highlights the broad similarities between the regions chosen as priorities for maximizing endemic richness and those chosen for simple species richness. There are differences, however. Our optimal solution for the 17% target uses 166 regions when prioritizing richness and these regions cover 24,532,093 km² and contain 86.2% of the species. When prioritizing endemics, one includes only 155 regions. They cover 24,292,910 km² and capture 66.6% species entirely within these regions. About 95% of the species included in the solution for endemics are shared with that for
richness, but only 73% of the species in the solution for richness are shared with those in that for endemics. These results point to the fact that a simple decision — to contain some of a species range (as in traditional gap-analysis exercises) or all of a species range — will necessarily change some priority regions.
Legends for supplementary figures

**Fig. S1. Checking the representativeness of the species sampled.** (A) Numbers of families in the sample of species analyzed and total numbers of families present in each of 52 major regions and (B) comparable plots for genera.

**Fig. S2. Species richness accumulation map.** Colors show the numbers of unique species they add to the total, given the inclusion of higher priority regions within the accumulated set, scaled by the accumulated area. Protecting all the areas with > 3.5 species added per 1000 km² (all colored areas) — and no others — would include 86% of flowering plant species.

**Fig. S3. Checking the whether the greedy algorithm is optimal.** Species accumulation curves as seen in Figure 2 of the main text (black: richness, grey: endemic richness). The optimal fraction of species as determined using a genetic algorithm, solved for 5%, 10%, 15%, 17%, 20%, 25%, and 30% of the total land area (red diamonds).
References and Notes


23. Materials and methods are available as supplementary materials on *Science* Online.


