Supplemental Methods

We searched Web of Science for fungal diversity studies using next generation sequencing platforms, such as 454, Illumina and Ion Torrent. Our primary search was conducted for the years 2009-2015 using the terms “fung*” and “communit*” and “454” or “Illumina” or “next generation” or “high throughput” within the Environmental Sciences Ecology research area. This resulted in 276 studies. We carried out additional searches using the terms “mycobiome” (32 studies), “marine fung*” (46) and “freshwater fung*” (29) along with the same sequencing related search terms (note there was some overlap across searches). From this list we included studies from four primary habitats, aquatic (freshwater and marine), terrestrial soils, plant associated (roots, leaves), and animal associated (internal and external). We excluded studies that were primarily methodological, agricultural, or used loci other than the internal transcribed spacer (ITS) region of the nuclear ribosomal RNA genes, the official taxonomic barcode for Fungi.

While the ITS is the most widely used taxonomic locus for general fungal surveys, as with any marker gene it underrepresents some taxonomic groups. For example, some basal fungal lineages and arbuscular mycorrhizal fungi are recovered infrequently from soils using ITS primers at least in part due to primer biases. Thus, while ITS has proved highly useful for broad comparisons of fungal communities, results using marker genes should always be interpreted with caution due to potential taxonomic biases.

From these studies, we recorded basic information on the habitat sampled and fungal diversity (Table S1). As studies varied tremendously in metrics reported, we used the total number of fungal sequences generated and the number of non-singleton OTUs as our primary diversity metric. If these were not reported for the entire study, we used data from the largest sampling unit where this was reported. If only the total number of species was reported we estimated the number of non-singletons using the median proportion of singletons from our data (10 studies). If
multiple loci were analyzed we recorded data from the ITS1, which appeared to be the most common target locus. To provide a more fine scale analysis of community structure, we reanalyzed datasets from representative studies across these habitats using the same bioinformatics approach. In total we were able to include data from 74 studies.

To analyze the data we divided assigned studies to four basic habitats, soils, plants, aquatic (marine and terrestrial) and animals. $\log_{10}$ fungal OTU richness was modeled as a function of $\log_{10}$ sequence depth with separate intercepts fit for the different substrates (i.e. $\log_{10}(\text{OTUs}) \sim \log_{10}(\text{Sequences}) + \text{Substrate}$). Adjusted $R^2$ was 0.59 and both substrate ($F_{3,69}=25.289$, $P < 0.001$) and sequence depth ($F_{1,69}=36.538$, $P < 0.001$) were significant. While data was insufficient to divide habitats more finely (e.g. pelagic, benthic and animal associated within marine ecosystems) this should be a priority for future studies. While there are of course major limitations to this type of analysis, i.e. the lack of standardized sampling units, spatial scale, sequencing protocols and data processing, our analysis indicates strong patterns and highlights opportunities for future research.