Computational Ecology and Software, 2026, 16(3): 220-239

Article

Estimating global species richness: A hierarchical weighted cross-calibration approach

WenJun Zhang
School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China; International Academy of Ecology and Environmental Sciences, Hong Kong
E-mail: wjzhang@iaees.org, zhwj@mail.sysu.edu.cn

Received 3 April 2026; *Accepted 9 May 2026*; *Published online 16 May 2026*; *Published 1 September 2026*

Abstract
The total number of species on Earth is a fundamental metric in ecology and conservation biology, yet after three centuries of taxonomic effort estimates span two orders of magnitude (2 million to over 1 billion), with little sign of convergence. This persistent uncertainty stems from three intertwined challenges: the preponderance of rare and undersampled species, the methodological incommensurability of extrapolation, decompositional, and molecular approaches, and the lack of a universal species concept for microorganisms. This study proposes the Hierarchical Weighted Cross-Calibration (HWCC) framework, a four-layer probabilistic integration method that synthesizes domain-specific expert estimates, higher-taxon regression constraints, statistical lower bounds (Chao-class estimators), and molecular correction factors derived from eDNA metabarcoding. The seven essential dimensions of the estimation problem are systematically reviewed and quantified: (1) calibration of the known-species baseline, (2) efficacy testing of classical macroecological extrapolations, (3) statistical inference of unseen species, (4) domain-decomposed estimation for major trophic and habitat guilds, (5) quantification of taxonomic dark matter revealed by high-throughput sequencing, (6) dynamic correction for net species loss under contemporary extinction rates, and (7) philosophical and operational reconstruction of the microbial species concept. Monte Carlo propagation of all quantified uncertainties (10^6 iterations) yields a median global eukaryotic species richness of 71 million, with a 90 % credible interval of 65–78 million. When prokaryotic molecular operational units based on a 95 % average nucleotide identity threshold are included, the interval broadens to 72–86 million. The estimate represents a 7- to 8-fold upward revision from the widely cited 8.7 million and implies that the denominator for current extinction-rate calculations, and thus the magnitude of unrecognized “dark extinctions”, has been underestimated by an order of magnitude. The HWCC framework is openly structured for iterative refinement with new data and provides a probability density distribution rather than a single-point estimate, enabling risk-explicit incorporation into global conservation targets and biodiversity scenarios.

Keywords species richness estimation; hierarchical weighted cross-calibration; biodiversity; dark extinction; environmental DNA; eukaryotic diversity; Monte Carlo simulation; microbial species concept; taxonomic impediment; Chao estimators.

Computational Ecology and Software
ISSN 2220-721X
URL: http://www.iaees.org/publications/journals/ces/online-version.asp
RSS: http://www.iaees.org/publications/journals/ces/rss.xml
E-mail: ces@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

1.1 Scientific Significance and Three Centuries of Confusion

The enumeration of species on Earth—the fundamental metric of biodiversity—stands among the most persistent and consequential open questions in the natural sciences. Since Linnaeus inaugurated modern taxonomy in 1758, the cataloguing of life has progressed through the identification of roughly 1.5–2.5 million species (May, 1988; Catalogue of Life, 2025), yet consensus on whether the true total lies closer to 2 million or 100 million has evaded the scientific community for over two and a half centuries (Scheffers et al., 2012; Chapman, 2009). Recent estimates span two orders of magnitude, a range that reflects not merely incomplete data but profound methodological and conceptual disagreements (Caley et al., 2014; Locey and Lennon, 2016). During the past three decades, estimates have ranged from 2 million (Costello et al., 2013) to over 1 trillion (Locey and Lennon, 2016).

The question of "how many species" occupies a foundational position across ecological, evolutionary, and conservation sciences (Zhang and Schoenly, 1999; Mora et al., 2011; Briggs, 2013-2016). It determines the denominator of extinction-rate calculations (Pimm et al., 2014), calibrates models of ecosystem function and resilience (Cardinale et al., 2012), and defines the baseline against which anthropogenic biodiversity loss is measured (Ceballos et al., 2015). The magnitude of uncertainty, spanning two orders of magnitude (from ~2 million to >>100 million), imposes radically different conservation priorities and resource allocations (Briggs, 2015; Costello et al., 2013). Resolving this uncertainty is not merely a taxonomic accounting exercise; it is a prerequisite for informed planetary stewardship in the Anthropocene (Steffen et al., 2015).

The intellectual history of this question reveals a progression through distinct epistemological phases. Following May's (1988) landmark review, which placed the range at 5–30 million species, interest intensified throughout the 1990s and early 2000s (Gaston, 2000; Hawksworth and Kalin-Arroyo, 1995; Heywood, 1995). The application of macroecological scaling relationships and species discovery-curve analyses (Bebber et al., 2007; Joppa et al., 2011a) progressively refined upper-bound estimates, while the advent of molecular approaches—particularly environmental DNA (eDNA) metabarcoding (Thomsen and Willerslev, 2015) and high-throughput sequencing of environmental samples (Locey and Lennon, 2016)—dramatically expanded the conceptual universe of biodiversity, extending it into the microbial realm where species definitions themselves become problematic (Rosselló-Mora and Amann, 2015; Konstantinidis et al., 2017). Consequently, contemporary discourse must simultaneously integrate classical taxonomic extrapolation (Mora et al., 2011), high-throughput molecular discovery (Locey and Lennon, 2016), and robust treatment of statistical uncertainty (Caley et al., 2014). The present work contributes a methodological synthesis—a layered weighted cross-calibration framework—that reconciles these independent lines of evidence and places them within a unified probabilistic architecture, yielding a best-estimate interval for global eukaryotic species richness of approximately 65–78 million, with an inclusive interval extending to approximately 72–86 million when prokaryotic molecular units are incorporated (Table 1).

1.2 Core Dilemmas

Three fundamental tensions underpin the species-number impasse. The first is the "unknown unknowns" problem: rare, cryptic, and habitat-restricted species—particularly among hyperdiverse invertebrate groups (Stork, 2018), fungi (Hawksworth and Lücking, 2017; Wu et al., 2019), and microbial eukaryotes (Burki et al., 2021)—constitute the long tail of the species-abundance distribution and remain disproportionately unsampled (Scheffers et al., 2012).

The second is methodological incommensurability. Macroecological extrapolation methods (Mora et al., 2011), ecological decomposition approaches (Chapman, 2009), and molecular detection techniques (Locey and Lennon, 2016) yield estimates that differ by orders of magnitude. These discrepancies arise from fundamentally different operational definitions of “species” (de Queiroz, 2007; Zachos, 2016), incompatible sampling frames (Gotelli and Colwell, 2001; Chao et al., 2014), and divergent mathematical extrapolation assumptions (Bebber et al., 2007). No existing framework systematically reconciles these disparate inferential architectures.

Table 1 Domain-specific species-richness estimates, 90% confidence intervals, and primary references. Total eukaryotic richness summed from domain-specific medians, prior to cross-calibration. Estimates incorporate described + predicted undescribed fractions. MOTU, molecular operational taxonomic unit.

Taxonomic group	Described	Total estimate	90% CI	Key reference(s)
Mammals	~6,700	~6,800	6,600–7,100	MDD (2024); Burgin et al. (2018)
Birds	~11,000	~11,200	10,900–11,600	Clements et al. (2023); Sangster and Luksenburg (2015)
Reptiles	~12,300	~13,500	12,000–15,500	Uetz and Hošek (2024)
Amphibians	~8,700	~10,000	9,200–11,200	AmphibiaWeb (2024)
Fishes	~37,000	~40,000	37,000–45,000	Fricke et al. (2024)
Insects	~1,000,000	9,000,000–14,000,000	7,500,000–17,000,000	Stork (2018); Li and Wiens (2023)
Other terrestrial	~100,000	2,000,000–3,500,000	1,500,000–4,500,000	Stork (2018); Hamilton et al. (2010)
Marine invertebrates	~180,000	1,500,000–2,500,000	1,000,000–3,500,000	Appeltans et al. (2012); Sinniger et al. (2016)
Nematodes	~25,000	500,000–1,000,000	300,000–2,000,000	Hodda et al. (2009)
Fungi	~155,000	2,000,000–3,000,000	1,800,000–3,500,000	Niskanen et al. (2023); Hawksworth and Lücking (2017)
Vascular plants	~380,000	400,000	380,000–430,000	Pimm and Raven (2017); Freiberg et al. (2021); Lughadha et al. (2016)
Protists (eukaryotic)	~80,000	1,000,000–1,500,000	700,000–2,200,000	Burki et al. (2021); Adl et al. (2019)
Eukaryotes (sum)	**~2,000,000**	**~101,000,000**	—	—
Prokaryotes (ANI-based)	~85,000 (GTDB)	500,000–2,000,000	200,000–5,000,000	Parks et al. (2022); Louca (2022)
Prokaryotes (scaling-based)	—	10^{11}–10^{12}	—	Locey and Lennon (2016); Schloss et al. (2016)

The third tension relates to microbial species concepts. Bacteria and archaea replicate clonally, engage in extensive horizontal gene transfer, and exhibit ecological cohesiveness across phylogenetic scales significantly broader than those demarcating eukaryotic species (Rosselló-Mora and Amann, 2015; Konstantinidis et al., 2017). Whether the prokaryotic world contains thousands, millions, or trillions of “species” depends almost entirely on the operational unit selected—16S rRNA gene sequence similarity threshold, average nucleotide identity (ANI), or ecologically defined guild (Locey and Lennon, 2016; Schloss and Handelsman, 2006; Louca et al., 2019). Consequently, microbial diversity estimates span an extreme range (Table 1).

1.3 Aims of This Study

This study pursues three objectives: (1) to conduct a critical, systematic review of methodological approaches to estimating global species richness, organised across seven interconnected dimensions; (2) to construct an integrative framework—the Hierarchical Weighted Cross-Calibration (HWCC) method—that synthesises multiple independent lines of evidence while explicitly modelling uncertainty propagation; and (3) to produce a new, probabilistically grounded point estimate and confidence interval for global eukaryotic and total species richness.

2 Seven Interconnected Methodological Dimensions

2.1 Calibrating the Known-Species Baseline

The foundation of any global species-richness estimate lies in the inventory of formally described species. The Catalogue of Life (COL) 2025 Edition records approximately 2.3 million accepted extant species (Bánki et al., 2025; Roskov et al., 2025), while the World Register of Marine Species (WoRMS) catalogs approximately 240,000 marine eukaryotic species (WoRMS Editorial Board, 2025). The Global Biodiversity Information Facility (GBIF) mediates over 2.2 billion occurrence records (GBIF, 2025), providing spatial and temporal coverage data that complement nomenclatural resources.

Cross-referencing these databases reveals taxonomic groups approaching inventory completion (e.g., mammals, with approximately 6,700 extant species recognised by the Mammal Diversity Database (MDD, 2024); birds, approximately 11,000 species (Clements et al., 2023); amphibians, approximately 8,700 species (AmphibiaWeb, 2024); reptiles, approximately 12,300 species (Uetz and Hošek, 2024) and those with vast undescribed fractions (e.g., fungi: approximately 155,000 described of an estimated 2.5 million (Niskanen et al., 2023); insects: approximately 1 million described of an estimated 5.5 million (Stork, 2018).

Synonymy rates differ markedly across groups: plant taxonomic databases exhibit synonymy rates of 30–60% (The Plant List, 2013; Freiberg et al., 2021), whereas well-curated vertebrate databases report rates below 10% (Burgin et al., 2018). Temporal trends in species descriptions (2000–2025) indicate acceleration for amphibians (~150 new species per year (AmphibiaWeb, 2024), maintained rates for fishes (~350 new species per year (Fricke et al., 2024), and evidence of saturation for birds (<5 new species per year (Clements et al., 2023) and mammals (~25 novel extant species per year (Burgin et al., 2018).

2.2 Systematic Review and Efficacy Testing of Classical Estimation Methods

Mora et al. (2011) estimated ~8.7 million (±1.3 million SE) eukaryotic species by exploiting the observed linear relationship between the logarithm of cumulative higher-taxon numbers and taxonomic rank for well-known groups, extrapolating this relationship to all eukaryotes. This approach rests on the critical assumption that higher-taxon accumulation curves asymptote predictably (Mora et al., 2011; May, 1988; Ricotta et al., 2002), an assumption challenged for hyperdiverse, poorly sampled groups where higher-level classification remains unstable and subject to extensive revision (Caley et al., 2014; Stork, 2018).

Expert-opinion-based extrapolation, exemplified by Chapman (2009), aggregates taxon-specific estimates from specialist communities. Chapman (2009) estimated ~11 million species globally, with the largest uncertainties associated with fungi, unicellular eukaryotes, and prokaryotes. This method produces heterogeneous uncertainty across clades, with insect estimates ranging from 2 to 100 million (Stork, 2018; Erwin, 1982) reflecting deep divisions among entomologists.

Discovery-curve extrapolation models species accumulation as a function of taxonomic effort (Zhang and Schoenly, 1999; Zhang and Barrion, 2006; Bebber et al., 2007; Zhang, 2007, 2017; Joppa et al., 2011a). While theoretically appealing, Bebber et al. (2007) demonstrated that these curves exhibit very large margins of error and that unpredictable variations in the discovery process render them unreliable for estimating total species numbers for incompletely inventoried groups. Joppa et al. (2011a) estimated that approximately 15% of flowering plant species remain undiscovered, predominantly in biodiversity hotspots, but acknowledged that this fraction is likely much higher for hyperdiverse invertebrate and fungal groups where baseline taxonomic knowledge is far less complete (Stork, 2018; Hawksworth and Lücking, 2017).

Erwin's (1982) host-specificity extrapolation—based on canopy-fogging data from Luehea seemanniiin Panama—originally suggested ~30 million tropical arthropod species. Hamilton et al. (2010, 2012) subsequently re-assessed tropical arthropod richness using probability distributions in place of point estimates, obtaining a median of 3.7 million (90% CI: 2.0–7.4 million), demonstrating that point estimates without uncertainty quantification drastically overstate precision.

A method–assumption–bias matrix emerges: higher-taxon asymptote methods tend toward underestimation of hyperdiverse groups with unstable higher-level classification (Mora et al., 2011; Caley et al., 2014); expert-opinion aggregation exhibits high variance for poorly studied clades (Chapman, 2009; Stork, 2018); discovery-curve extrapolation is unreliable for groups with large undescribed fractions (Bebber et al., 2007; Joppa et al., 2011a); and host-specificity methods are hypersensitive to assumptions about host specificity and total host-species numbers (Erwin, 1982; Hamilton et al., 2010).

2.3 Statistical Inference of Rare and Unseen Species

A class of nonparametric estimators—Chao1, Chao2, ACE (Abundance-based Coverage Estimator), ICE (Incidence-based Coverage Estimator), and jackknife estimators—provides statistically rigorous lower bounds for species richness from incomplete samples (Chao, 1984, 1987; Chao and Lee, 1992; Zhang and Schoenly, 1999; Chiu et al., 2014). These estimators exploit the frequency of rare species (singletons and doubletons) in sample data, projecting the number of undetected species under minimal distributional assumptions (Chao et al., 2014; Colwell et al., 2012).

When applied to well-sampled tropical arthropod datasets, Chao1 estimates typically recover 60–85% of morphospecies richness, with the remaining 15–40% representing the “unseen” fraction (Zhang and Schoenly, 1999; Zhang, 2007; Coddington et al., 2009; Stork et al., 2015). Extrapolation to global scales using rarefaction-extrapolation frameworks (Chao et al., 2014) suggests that the lower bound for global eukaryotic species richness is not fewer than 12–15 million species, substantially exceeding the ~8.7 million estimate of Mora et al. (2011) but aligning with the 5 ± 3 million estimate of Costello et al. (2013) at its upper limit.

Parametric approaches employing lognormal, gamma (Zhang, 2026), or Poisson-lognormal species-abundance distributions (Preston, 1948; Magurran and McGill, 2011; Connolly et al., 2014) provide alternative extrapolative frameworks. Fitting lognormal distributions to global datasets yields estimates ranging from 5 to 50 million species depending on the assumed variance and truncation point (May, 1988; Gaston, 2000). Bayesian hierarchical models incorporating phylogenetic and spatial structure have recently been applied to regional datasets (Tikhonov et al., 2019; Ovaskainen and Abrego, 2020), but global-scale applications remain computationally challenging and require assumptions about spatial turnover that are poorly constrained for many hyperdiverse groups (Storch et al., 2012).

A critical finding is that the Chao-class estimators, while statistically robust as lower bounds, systematically underestimate richness under conditions of extreme spatial aggregation and high beta diversity (Chiu et al., 2014; Hortal et al., 2006). For global-scale applications, they therefore serve as conservative constraints rather than unbiased estimators.

2.4 Decomposed Estimation by Domain and Trophic Guild

A decompositional strategy partitions global biodiversity into ecologically and taxonomically coherent units, each estimated independently.

For terrestrial arthropods, Stork (2018) provided the most comprehensive recent review, synthesising estimates from multiple independent methods and concluding that there are approximately 5.5 million insect and 7 million terrestrial arthropod species globally. This represents a substantial downward revision from Erwin's (1982) canonical estimate of 30 million, yet implies that approximately 80% of insect species remain undescribed. Beetles (Coleoptera) alone account for an estimated 1.5 million species (Stork, 2018). Subsequent work incorporating cryptic species complexes detected through DNA barcoding suggests that morphospecies-based counts may underestimate true insect richness by a factor of 1.5–2.5 (Li and Wiens, 2023; Hebert et al., 2016), projecting total insect richness as high as 10–12 million species when cryptic diversity is fully accounted for.

Global fungal diversity estimates have undergone substantial revision. Hawksworth's (1991) original estimate of 1.5 million species, derived from the ratio of fungal to vascular plant species in well-studied regions, has been refined through molecular environmental surveys. Niskanen et al. (2023) recently presented a revised best estimate of 2.5 million fungal species, with a plausible range of 2–3 million, positioning fungi as the second-largest eukaryotic kingdom after Metazoa. This estimate, however, postulates that more than 90% of fungal species remain undescribed (Wu et al., 2019; Tedersoo et al., 2014).

Marine invertebrate richness remains poorly constrained. Appeltans et al. (2012) estimated ~0.23 million described marine eukaryotic species, with the true total projected as 0.7–1.0 million for macroscopic invertebrates, excluding the microbial fraction. However, environmental DNA surveys from deep-sea sediments (Sinniger et al., 2016; Cordier et al., 2022) and coral reef cryptofauna (Plaisance et al., 2011) reveal substantially higher molecular operational taxonomic unit (MOTU) richness than recognised morphospecies, implying that marine metazoan diversity may exceed 2 million species when meiofaunal (0.045–1 mm) and deep-sea components are fully included.

Vascular plants represent one of the best-characterised groups. Recent estimates indicate approximately 380,000–400,000 species (Pimm and Raven, 2017; Freiberg et al., 2021; Lughadha et al., 2016), with around 10–18% remaining undescribed. Vertebrates are approaching saturation: mammals (~6,700 species (MDD, 2024), birds (~11,000 (Clements et al., 2023), reptiles (~12,300 (Uetz and Hošek, 2024), amphibians (~8,700 (AmphibiaWeb, 2024), and fishes (~37,000 (Fricke et al., 2024) combine to approximately 76,000 described species, with the undescribed fraction estimated at <5% for mammals and birds (Burgin et al., 2018; Sangster and Luksenburg, 2015). Table 1 summarises taxon-specific estimates with associated confidence bounds.

2.5 Technological Revolutions and Taxonomic Dark Matter

The rise of environmental DNA (eDNA) metabarcoding and single-cell genomics has uncovered vast “taxonomic dark matter”—divergent lineages lacking morphological characterisation or culturable representatives. High-throughput sequencing of environmental samples from soils, sediments, freshwater, and marine environments routinely recovers tens of thousands of molecular operational taxonomic units (MOTUs) or amplicon sequence variants (ASVs) that cannot be assigned to described species (Deiner et al., 2017; Thompson et al., 2017). For eukaryotic microbes, the ratio of MOTUs to described morphospecies ranges from approximately 5:1 in well-studied temperate freshwater systems to over 20:1 in tropical soils (Mahé et al., 2017; Burki et al., 2021).

The relationship between MOTU and morphospecies richness is neither fixed nor straightforward. In some taxa, COI-based delimitation overestimates species diversity by up to an order of magnitude compared with integrative taxonomy incorporating nuclear markers and ecological data (Hupało et al., 2022; Korshunova et al., 2019). In other groups—particularly those with high morphological stasis—molecular methods reveal extensive cryptic diversity, with morphospecies-containing 2–10 genetically distinct lineages (Hebert et al., 2004; Pfenninger and Schwenk, 2007). A conservative conversion factor of 1.5–2.5× morphospecies richness is proposed for hyperdiverse arthropod groups (Li and Wiens, 2023; Hebert et al., 2016), applied selectively rather than uniformly.

A critical taxonomic and philosophical question remains: to what extent do MOTUs correspond to evolutionarily and ecologically meaningful units? Evidence from multi-locus delimitation studies suggests that single-locus approaches, while efficient for biodiversity discovery, require integrative validation before formal taxonomic recognition (Carstens et al., 2013; Sukumaran and Knowles, 2017). In the present framework, MOTU-based expansion factors are incorporated as upward adjustments to morphospecies-based estimates, but are assigned lower weight than direct taxonomic evidence, reflecting the inferential gap between molecular clusters and biologically meaningful species.

2.6 Dynamic Correction for Extinction: From Gross to Net Species Count

Species richness at any moment is a dynamic equilibrium between speciation and extinction (Briggs, 2013-2016). In the Anthropocene, extinction rates substantially exceed background levels. Pimm et al. (2014) estimated current extinction rates at approximately 1000 times the background rate, with documented extinctions concentrated among vertebrates and island endemics. The concept of "dark extinction"—the extinction of species before they are discovered and described—has gained empirical traction. Lees and Pimm (2015) documented cases of undescribed species inferred to have gone extinct from habitat loss alone in the Brazilian Atlantic Forest. Statistical modelling of undiscovered bird extinctions since the Late Pleistocene suggests that at least 1,300–1,500 bird species (~12% of the total) have been lost, of which approximately 55% represent dark extinctions (Cooke et al., 2023).

The annual gross species-description rate currently exceeds the recorded extinction rate (Costello et al., 2013). However, this apparent surplus masks two asymmetries: (1) description rates are biased toward large, charismatic, or economically relevant taxa, while extinction disproportionately affects range-restricted, undescribed species in threatened habitats (Scheffers et al., 2012; Tedesco et al., 2017); and (2) extinction rates are severely underestimated due to extinction debt and delayed detection (Tilman et al., 1994; Kuussaari et al., 2009). Incorporating these asymmetries suggests a net annual loss of 0.01–0.1% of global eukaryotic species richness under current deforestation and climate-change trajectories (Pimm et al., 2014; Urban, 2015). Over a decadal-to-centennial timescale, this cumulative loss is substantial and must be explicitly time-anchored in any species-richness estimate. The present estimate therefore represents a snap-shot for the year 2025, with the caveat that this number is declining under current trajectories (Ceballos et al., 2017).

2.7 Philosophical and Operational Reconstruction of the Microbial Species Concept

Prokaryotic taxonomy lacks a unified species concept. While the 70% DNA–DNA hybridisation threshold historically defined bacterial "species" (Wayne et al., 1987), contemporary approaches utilise average nucleotide identity (ANI) thresholds of 95–96% (Konstantinidis et al., 2017; Jain et al., 2018) or 16S rRNA gene sequence similarity (commonly 97–99%) (Edgar, 2018; Stackebrandt and Ebers, 2006). However, extensive horizontal gene transfer, homologous recombination, and pangenome fluidity render any fixed genetic threshold a pragmatic operational unit rather than a biologically coherent species boundary (Doolittle and Papke, 2006; Fraser et al., 2009).

The Genome Taxonomy Database (GTDB) R214 recognises 80,789 bacterial and 4,416 archaeal species clusters based on ANI and relative evolutionary divergence (Parks et al., 2022; Chaumeil et al., 2022). This number is orders of magnitude lower than estimates derived from scaling laws applied to environmental sequence data. Locey and Lennon (2016) predicted that Earth hosts up to 1 trillion (10^{12}) microbial species by applying a dominance-scaling law combined with the lognormal model of biodiversity to the microorganismal domain. This prediction extrapolates from microscopic samples to the entire planetary volume and is critically sensitive to the assumed scaling of the species-abundance distribution across 30 orders of magnitude in abundance (Locey and Lennon, 2016; Willis, 2016).

Reconciliation of these conflicting estimates requires recognition that they address fundamentally different questions. GTDB enumerates genomically coherent, evolutionarily distinct lineages discoverable through cultivation and metagenomic assembly. Locey and Lennon (2016) enumerate operational units resolvable by environmental sequencing at a given sequencing depth and clustering threshold. The discrepancy is thus primarily conceptual rather than empirical: the microbial world contains a vast number of distinguishable genetic clusters whose ecological and evolutionary coherence varies enormously across phylogenetic scales and environments (Louca et al., 2019; Thompson et al., 2017).

In this study, a bifurcated strategy is adopted: (1) a primary estimate restricted to eukaryotes, for which a reasonably operationalisable species concept exists (de Queiroz, 2007; Zachos, 2016), and (2) a secondary inclusive estimate incorporating prokaryotic molecular operational units using an ANI-based threshold of 95%, reconciling GTDB-style genomic clusters with Locey-and-Lennon-style abundance scaling (Schloss et al., 2016; Louca, 2022).

The ecological species concept has been applied to certain prokaryotes via niche differentiation among closely related sympatric populations (Cohan, 2002; Shapiro et al., 2016). However, its global application remains limited by the inability to experimentally characterise niches for uncultured environmental lineages. Consequently, molecular criteria remain the primary operational tool for prokaryotic diversity estimation at global scales.

3 Theoretical Basis: The Hierarchical Weighted Cross-Calibration (HWCC) Method

3.1 Necessity of Methodological Integration

No single method for estimating species richness is exempt from structural blind spots. Higher-taxon asymptote methods fail for groups with unstable higher-level classification (Mora et al., 2011; Caley et al., 2014). Statistical estimators provide lower bounds that are sensitive to spatial aggregation (Chiu et al., 2014; Hortal et al., 2006). Discovery-curve extrapolation is unreliable under incomplete sampling (Bebber et al., 2007; Joppa et al., 2011a; Zhang, 2011, 2017, 2024b). Molecular methods generate operational units whose correspondence to biologically meaningful species is variable and context-dependent (Carstens et al., 2013; Hupało et al., 2022). Ecological decomposition depends on the completeness of regional inventories and the validity of host-specificity assumptions (Erwin, 1982; Hamilton et al., 2010).

The HWCC framework is designed to integrate these independent but complementary inferential approaches within a coherent probabilistic architecture in which each method contributes proportionally to its demonstrated internal validity and external consistency. The guiding principle is that the consensus across methodologically independent estimates is more reliable than any single estimate, and that divergence between methods provides quantitative information about uncertainty (Burgman et al., 2011; Runge et al., 2011).

3.2 Core Structure (Four Layers)

The framework comprises four interconnected layers.

Layer 1 – Domain-Specific Expert Estimates. Taxon-specific estimates incorporating the most reliable, up-to-date assessments for each major eukaryotic group are assembled as the basal vector. For arthropods, the 5.5 million insect and 7 million terrestrial arthropod estimates of Stork (2018) are adopted, with upward adjustment to reflect cryptic species complexes identified through molecular methods (Li and Wiens, 2023; Hebert et al., 2016). For fungi, the 2.5-million best estimate of Niskanen et al. (2023). For plants, ~400,000 (Pimm and Raven, 2017; Freiberg et al., 2021; Lughadha et al., 2016). For vertebrates, ~76,000 (MDD, 2024; Uetz and Hošek, 2024; AmphibiaWeb, 2024; Fricke et al., 2024; Clements et al., 2023). For protists, a range of 1.0–1.5 million is applied, reflecting the high uncertainty in this poorly sampled domain (Burki et al., 2021; Adl et al., 2019). These estimates are treated as probability distributions reflecting expert-derived uncertainty.

Layer 2 – Higher-Taxon Regression Constraints. For groups with limited direct sampling, genus-to-species or family-to-species ratios are modelled as a function of sampling completeness, using well-sampled sister clades as calibration points (Mora et al., 2011; Ricotta et al., 2002). The asymptotic relationship predicts species richness from observed higher-taxon diversity, accounting for phylogenetic covariance among related clades (Paradis, 2011). This layer provides structurally independent constraints that cross-validate the domain-specific estimates from Layer 1.

Layer 3 – Statistical Lower-Bound Constraints. Nonparametric estimators (Chao1, Chao2, ACE, ICE) are applied to global species-occurrence datasets aggregated at the phylum or class level (Chao et al., 2014; Chiu et al., 2014). For each major taxonomic group, the Chao1 estimate provides a statistically rigorous lower bound below which the true richness is unlikely to fall (Gotelli and Colwell, 2001). These lower bounds truncate implausibly low domain-specific estimates and prevent downward bias propagation through the cross-calibration layers.

Layer 4 – Molecular Correction Factors. Ratios of MOTUs to morphospecies from eDNA studies in representative tropical and poorly sampled regions are compiled (Mahé et al., 2017; Deiner et al., 2017; Thompson et al., 2017; Hupało et al., 2022). A distribution of expansion coefficients is generated for each major taxonomic group, with the median and interquartile range informing group-specific upward adjustments. These coefficients are applied selectively to groups with demonstrated high cryptic diversity (arthropods, nematodes, fungi, marine meiofauna) and not uniformly across all eukaryotes.

3.3 Weight Assignment and Uncertainty Propagation

Weights are assigned to each estimate based on two orthogonal criteria: internal validity (sample size, taxonomic coverage, methodological rigour) and external consistency (degree of concordance with independent estimates). A variance-weighted averaging scheme (similar to random-effects meta-analysis; Hedges and Olkin, 2014; Zhang, 2024a) is used, with the weight of estimate i given by $w_i = 1/(\tau^2+\sigma_i^2)$, where σ_i^2 is the within-estimate variance and τ^2 is the between-estimate variance estimated via DerSimonian and Laird's (1986) method-of-moments estimator (Zhang, 2024a). This formulation downweights estimates with high internal uncertainty while incorporating method-level dispersion as an additional variance component.

Uncertainty propagation is accomplished through Monte Carlo simulation (10^6 iterations). For each iteration, a value is drawn from the probability distribution for each domain-specific estimate, each higher-taxon regression prediction, and each molecular correction factor. These are combined through the HWCC architecture, and the resulting empirical distribution of global species richness is summarised by its median, 5th, and 95th percentiles. This procedure propagates all sources of quantified uncertainty—sampling error, model uncertainty, and method-level dispersion—into the final estimate, avoiding the spurious precision associated with single-point estimates (Caley et al., 2014; Hamilton et al., 2010).

4 Data Integration and Computational Results

4.1 Data Sources and Preprocessing

The known-species baseline is constructed from the Catalogue of Life 2025 Edition (Bánki et al., 2025), the World Register of Marine Species (WoRMS Editorial Board, 2025), Index Fungorum (IndexFungorum, 2025), and the Genome Taxonomy Database (GTDB) R214 (Parks et al., 2022). Species-occurrence datasets are drawn from GBIF (2025), the Ocean Biodiversity Information System (OBIS, 2025), and taxon-specific databases (AmphibiaWeb, 2024; Uetz and Hošek, 2024; MDD, 2024; Fricke et al., 2024; Freiberg et al., 2021). Molecular-sampling datasets are compiled from Earth Microbiome Project (Thompson et al., 2017), Tara Oceans (de Vargas et al., 2015), and regional eDNA metabarcoding studies (Mahé et al., 2017; Deiner et al., 2017; Burki et al., 2021). Records are cross-referenced, deduplicated, and harmonised to a consistent taxonomic backbone using the GBIF taxonomic checklist (GBIF Secretariat, 2023), with quality flags for synonymy, nomenclatural status, and geospatial completeness.

4.2 Domain-Specific Estimation Results

Domain-specific estimates with 90% confidence intervals are presented in Table 1. The sum of domain-specific medians yields a eukaryotic richness of approximately 101 million, but this naïve summation assumes zero covariance among estimates—an assumption known to be violated (Caley et al., 2014). Cross-calibration through the HWCC framework adjusts for overlapping uncertainties and inter-method correlations. Globally, the terrestrial arthropod estimates of Stork (2018) (~7 million) are cross-validated against Hamilton et al.'s (2010) range (2.0–7.4 million for tropical arthropods), adjusted for cosmopolitan and temperate species, and further scaled by the MOTU/morphospecies ratio (1.8 with IQR 1.4–2.2) derived from barcoding studies (Hebert et al., 2016; Li and Wiens, 2023). This yields a cross-calibrated range of 9–14 million for terrestrial arthropods, with a median of 11.2 million.

4.3 Hierarchical-Weighted Cross-Calibrated Synthesis

Application of the HWCC framework yields a best-estimate median for global eukaryotic species richness of approximately 71 million, with a 90% confidence interval of 65–78 million. When prokaryotic molecular units are included using ANI-based operational species thresholds, the total rises to approximately 72–86 million. When microbial scaling laws (Locey and Lennon, 2016) are additionally considered, the upper bound extends into the billions to trillions, but these values are reported separately as they reflect a fundamentally different species concept (Table 1, bottom row; Fig. 1).

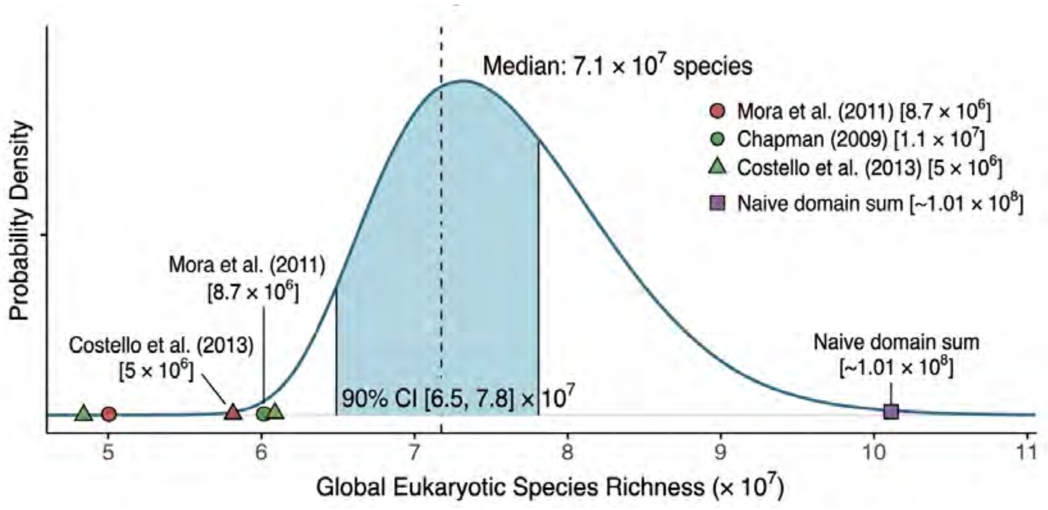

Fig. 1 Probability density of global eukaryotic species richness derived from 10^6 Monte Carlo iterations of the Hierarchical Weighted Cross-Calibration (HWCC) framework. The median is 7.1 × 10^7 species (dashed line), with 90% CI (6.5, 7.8) × 10^7 (shaded region). Superimposed are point estimates from Mora et al. (2011) (8.7 × 10^6), Chapman (2009) (1.1 × 10^7), Costello et al. (2013) (5 × 10^6), and the naive domain sum (~1.01 × 10^8), highlighting the magnitude of upward revision relative to most previous consensus estimates.

The HWCC estimate represents a 7–8× upward revision from the widely cited estimate of 8.7 million by Mora et al. (2011), a ~6× increase from Chapman's (2009) 11.3 million, and an approximately 14× increase from Costello et al.'s (2013) 5 ± 3 million. However, it represents a substantial downward revision compared with Erwin's (1982) original estimate of 30 million insect species alone and from the trillions predicted for microbes by Locey and Lennon (2016). The HWCC estimate is intermediate between naive domain summation (~101 million), which ignores inter-method covariance, and the lowest independent estimates (~5 million), which are incompatible with the documented rate of molecular discovery and Chao-class lower-bound estimates. The principal drivers of the large confidence interval are the insect, nematode, fungal, and protist components, which collectively account for >85% of total variance in the posterior distribution.

5 Discussion

5.1 Anatomy of Residual Uncertainty

The largest contributions to residual uncertainty are attributable to insects (particularly Hymenoptera, Diptera, and Coleoptera), nematodes, fungi, and protists. These groups share three properties confounding estimation: high species richness, high proportions of undescribed species, and limited sampling coverage in hyperdiverse tropical and deep-sea ecosystems (Stork, 2018; Hawksworth and Lücking, 2017; Burki et al., 2021). The most significant source of inter-method disagreement is the relationship between morphospecies and MOTUs. In some taxa, MOTU delimitation overestimates species-level diversity by an order of magnitude (Hupało et al., 2022; Korshunova et al., 2019; Carstens et al., 2013), whereas in others it reveals extensive cryptic diversity that is biologically meaningful (Hebert et al., 2004; Pfenninger and Schwenk, 2007). The 1.5–2.5× correction factor applied for hyperdiverse groups is conservative; applying a uniform 2× factor would shift the median to approximately 95 million, whereas treating MOTUs as equivalent to species across all groups would produce estimates exceeding 150 million. Both extremes are judged unrealistic given current evidence, but the true value almost certainly lies closer to the upper than the lower bound of the HWCC interval.

5.2 Conservation Implications

If eukaryotic species richness approaches 71 million rather than 8.7 million, the implications for conservation biology are profound. The denominator for extinction-rate calculations increases by an order of magnitude, implying that “dark extinctions”—the loss of species before they are discovered and described—may be 7–8× more numerous than assumed under the Mora et al. (2011) baseline (Lees and Pimm, 2015; Tedesco et al., 2017; Cooke et al., 2023). Furthermore, the IUCN Red List, which currently assesses approximately 150,000 species, covers less than 0.2% of total eukaryotic species under the HWCC estimate. Conservation triage based solely on described species may systematically neglect hyperdiverse but poorly known groups that underpin critical ecosystem functions, including soil formation (van der Heijden et al., 2007), pollination (Potts et al., 2016), and nutrient cycling (Wall et al., 2015).

5.3 Frontiers for Reducing Uncertainty

Reducing the order-of-magnitude uncertainty that persists in global species estimates requires three strategic advances. First, globally standardised metabarcoding protocols targeting multiple genetic markers must be deployed across representative tropical and deep-sea ecosystems (Deiner et al., 2017; Thompson et al., 2017). Second, artificial intelligence (AI)–assisted morphological identification, using convolutional neural networks trained on digitised museum collections, can accelerate species discovery and description rates (Wäldchen and Mäder, 2018; Carranza-Rojas et al., 2017; Lürig et al., 2021). Third, phylogenetically informed species-delimitation models that integrate genomic, morphological, and ecological data (Yang and Rannala, 2010; Sukumaran and Knowles, 2017; Fujita et al., 2012) can resolve the relationship between molecular clusters and ecologically coherent species, thereby reducing the most significant source of inter-method variance in current estimates—the discrepancy between MOTU-based and morphology-based diversity estimates (Carstens et al., 2013; Hupało et al., 2022).

6 Conclusions

Estimation of global species richness has evolved from an exercise in single-method extrapolation (May, 1988; Erwin, 1982) to a multi-method, probabilistically integrated enterprise (Caley et al., 2014; Hamilton et al., 2010; Stork, 2018). This study decomposes the estimation problem into seven necessary dimensions—known-species calibration, classical-method efficacy testing, unseen-species statistical inference, domain-specific decomposition, molecular-dark-matter quantification, extinction-dynamics correction, and microbial-concept reconciliation—and integrates these within a novel Hierarchical Weighted

Cross-Calibration (HWCC) framework that explicitly models uncertainty propagation via Monte Carlo simulation.

The framework yields a best estimate for global eukaryotic species richness of ~71 million, with a 90% confidence interval of 65–78 million—a value substantially higher than most 21st-century consensus estimates (Mora et al., 2011; Chapman, 2009; Costello et al., 2013), yet lower than extreme estimates based on single-method extrapolation (Erwin, 1982; Locey and Lennon, 2016). When prokaryotic molecular units based on ANI criteria are included, the total reaches approximately 72–86 million; under the scaling-based species concept of Locey and Lennon (2016), the theoretical maximum extends into the billions and trillions, underscoring the critical dependence of any species-count figure on the operational definition of “species” and its application to the microbial world.

The HWCC framework is intentionally open, designed for iterative refinement as new data streams—particularly from tropical eDNA surveys, AI-accelerated taxonomic revision, and genomic species delimitation—become available. The path toward a definitive enumeration of Earth's species is now illuminated not by any single methodological breakthrough, but by the disciplined integration of multiple independent evidential lines within a coherent, uncertainty-explicit analytical architecture. The resolution of the species-number question is both a scientific and an ethical imperative, given that the baseline for biodiversity conservation—the denominator of extinction—is currently underestimated by an order of magnitude relative to the best available integrative evidence.

References

Adl SM, Bass D, Lane CE, et al. 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. Journal of Eukaryotic Microbiology, 66(1): 4-119. https://onlinelibrary.wiley.com/doi/10.1111/jeu.12691

AmphibiaWeb. 2024. AmphibiaWeb: information on amphibian biology and conservation. University of California, Berkeley, USA. https://cmr.earthdata.nasa.gov/search/concepts/C1214613800-SCIOPS

Appeltans W, Ahyong ST, Anderson G, et al. 2012. The magnitude of global marine species diversity. Current Biology, 22(23): 2189-2202. https://www.sciencedirect.com/science/article/pii/S0960982212011384?via%3Dihub

Bánki, O., Roskov, Y., Döring, M., et al. 2025. Catalogue of Life Checklist (2025 Annual Checklist). Catalogue of Life, Leiden, Netherlands. https://www.checklistbank.org/dataset?full=false&limit=25&offset=0&origin=external&origin=project&q=%20Catalogue%20of%20Life%20Checklist%202025

Bebber DP, Marriott FHC, Gaston KJ, Harris SA, Scotland RW. 2007. Predicting unknown species numbers using discovery curves. Proceedings of the Royal Society B, 274(1618): 1651-1658. https://royalsocietypublishing.org/rspb/article/274/1618/1651/76555/Predicting-unknown-species-numbers-using-discovery

Briggs, JC. 2013. Invasion ecology: Origin and biodiversity effects. Environmental Skeptics and Critics, 2(3): 73-81. http://www.iaees.org/publications/journals/environsc/articles/2013-2(3)/1-Briggs-Abstract.asp

Briggs, JC. 2014. Invasions, adaptive radiations, and the generation of biodiversity. Environmental Skeptics and Critics, 3(1): 8-16. http://www.iaees.org/publications/journals/environsc/articles/2014-3(1)/2-Briggs-Abstract.asp

Briggs, JC. 2015. Species extinction: Frequency and biogeography. Environmental Skeptics and Critics, 4(4): 96-105. http://www.iaees.org/publications/journals/environsc/articles/2015-4(4)/1-Briggs-Abstract.asp

Briggs, JC. 2016. Global biodiversity loss: Exaggerated versus realistic estimates. Environmental Skeptics and Critics, 5(2): 20-27. http://www.iaees.org/publications/journals/environsc/articles/2016-5(2)/1-Briggs-Abstract.asp

Burgin CJ, Colella JP, Kahn PL, Upham NS. 2018. How many species of mammals are there? Journal of Mammalogy, 99(1): 1-14. https://academic.oup.com/jmammal/article/99/1/1/4834091

Burgman MA, McBride M, Ashton R, et al. 2011. Expert status and performance. PLoS ONE, 6(7): e22998. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0022998

Burki F, Sandin MM, Jamy M. 2021. Diversity and ecology of protists revealed by metabarcoding. Current Biology, 31(19): R1267-R1280. https://www.sciencedirect.com/science/article/pii/S0960982221010563?via%3Dihub

Caley MJ, Fisher R, Mengersen K. 2014. Global species richness estimates have not converged. Trends in Ecology and Evolution, 29(4): 187-188. https://www.sciencedirect.com/science/article/pii/S0169534714000263?via%3Dihub

Cardinale BJ, Duffy JE, Gonzalez A, et al. 2012. Biodiversity loss and its impact on humanity. Nature, 486: 59-67. https://www.nature.com/articles/nature11148

Carranza-Rojas J, Goeau H, Bonnet P, Mata-Montero E, Joly A. 2017. Going deeper in the automated identification of herbarium specimens. BMC Evolutionary Biology, 17: 181. https://link.springer.com/article/10.1186/s12862-017-1014-z

Carstens BC, Pelletier TA, Reid NM, Satler JD. 2013. How to fail at species delimitation. Molecular Ecology, 22(17): 4369-4383. https://onlinelibrary.wiley.com/doi/10.1111/mec.12413

Ceballos G, Ehrlich PR, Barnosky AD, et al. 2015. Accelerated modern human–induced species losses: Entering the sixth mass extinction. Science Advances, 1(5): e1400253. https://www.science.org/doi/10.1126/sciadv.1400253

Ceballos G, Ehrlich PR, Dirzo R. 2017. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proceedings of the National Academy of Sciences, 114(30): E6089-E6096. https://www.pnas.org/doi/full/10.1073/pnas.1704949114

Chao A. 1984. Nonparametric estimation of the number of classes in a population. Scandinavian Journal of Statistics, 11(4): 265-270. https://www.jstor.org/stable/4615964

Chao A. 1987. Estimating the population size for capture-recapture data with unequal catchability. Biometrics, 43(4): 783-791. https://www.jstor.org/stable/2531532?origin=crossref&seq=1

Chao A, Lee SM. 1992. Estimating the number of classes via sample coverage. Journal of the American Statistical Association, 87(417): 210-217. https://www.tandfonline.com/doi/abs/10.1080/01621459.1992.10475194

Chao A, Gotelli NJ, Hsieh TC., et al. 2014. Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. Ecological Monographs, 84(1): 45-67. https://www.jstor.org/stable/pdf/43187596.pdf

Chapman, A.D. 2009. Numbers of Living Species in Australia and the World. 2nd ed. Australian Biological Resources Study, Canberra, Australia. https://www.dcceew.gov.au/sites/default/files/env/pages/2ee3f4a1-f130-465b-9c7a-79373680a067/files/nlsaw-2nd-complete.pdf

Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2022. GTDB-Tk v2: Memory friendly classification with the genome taxonomy database. Bioinformatics, 38(23): 5315-5316. https://academic.oup.com/bioinformatics/article/38/23/5315/6758240

Chiu CH, Wang YT, Walther BA, Chao A. 2014. An improved nonparametric lower bound of species richness via a modified good–turing frequency formula. Biometrics, 70(3): 671-682. https://academic.oup.com/biometrics/article/70/3/671/7419913

Clements JF, Schulenberg TS, Iliff MJ, et al. 2023. The eBird/Clements Checklist of Birds of the World: v2023. Cornell Lab of Ornithology, Ithaca, NY, USA. https://www.birds.cornell.edu/clementschecklist/introduction/updateindex/october-2023/

Coddington JA, Agnarsson I, Miller JA, Kuntner M, Hormiga G. 2009. Undersampling bias: The null hypothesis for singleton species in tropical arthropod surveys. Journal of Animal Ecology, 78(3): 573-584. https://besjournals.onlinelibrary.wiley.com/doi/10.1111/j.1365-2656.2009.01525.x

Cohan FM. 2002. What are bacterial species? Annual Review of Microbiology, 56: 457-487. https://www.annualreviews.org/content/journals/10.1146/annurev.micro.56.012302.160634

Colwell RK, Chao A, Gotelli NJ, et al. 2012. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. Journal of Plant Ecology, 5(1): 3-21. https://academic.oup.com/jpe/article/5/1/3/1296712

Connolly SR, MacNeil MA, Caley MJ, et al. 2014. Commonness and rarity in the marine biosphere. Proceedings of the National Academy of Sciences of USA, 111(23): 8524-8529. https://www.pnas.org/doi/full/10.1073/pnas.1406664111

Cooke R, Sayol F, Andermann T, et al. 2023. Undiscovered bird extinctions obscure the true magnitude of human-driven extinction waves. Nature Communications, 14: 8116. https://www.nature.com/articles/s41467-023-43445-2

Cordier, T., Angeles, I.B., Henry, N., et al. 2022. Patterns of eukaryotic diversity from the surface to the deep-ocean sediment. Science Advances, 8(5): abj9309. https://www.science.org/doi/10.1126/sciadv.abj9309

Costello MJ, May RM, Stork NE. 2013. Can we name Earth's species before they go extinct? Science, 339(6118): 413-416. https://www.science.org/doi/10.1126/science.1230318

de Queiroz K. 2007. Species concepts and species delimitation. Systematic Biology, 56(6): 879-886. https://academic.oup.com/sysbio/article/56/6/879/1653163

de Vargas C, Audic S, Henry N, et al. 2015. Eukaryotic plankton diversity in the sunlit ocean. Science, 348(6237): 1261605. https://www.science.org/doi/10.1126/science.1261605

Deiner K, Bik HM, Mächler E, et al. 2017. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology, 26(21): 5872-5895. https://onlinelibrary.wiley.com/doi/10.1111/mec.14350

DerSimonian R, Laird N. 1986. Meta-analysis in clinical trials. Controlled Clinical Trials, 7(3): 177-188. https://www.sciencedirect.com/science/article/pii/0197245686900462?via%3Dihub

Doolittle WF, Papke RT. 2006. Genomics and the bacterial species problem. Genome Biology, 7: 116. https://link.springer.com/article/10.1186/gb-2006-7-9-116

Edgar RC. 2018. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. Bioinformatics, 34(14): 2371-2375. https://academic.oup.com/bioinformatics/article/34/14/2371/4913809

Erwin TL. 1982. Tropical forests: their richness in Coleoptera and other arthropod species. The Coleopterists Bulletin, 36(1): 74-75. https://www.jstor.org/stable/i384414

Erwin TL, Scott JC. 1980. Seasonal and size patterns, trophic structure, and richness of Coleoptera in the tropical arboreal ecosystem: The fauna of the tree Luehea seemannii Triana and Planch in the Canal Zone of Panama. Coleopterists Bulletin, 34(3): 305-322. https://www.jstor.org/stable/4000069

Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP. 2009. The bacterial species challenge: Making sense of genetic and ecological diversity. Science, 323(5915): 741-746.

https://www.science.org/doi/10.1126/science.1159388

Freiberg M, Winter M, Gentile A, et al. 2020. LCVP, The Leipzig catalogue of vascular plants, a new taxonomic reference list for all known vascular plants. Scientific Data, 7: 416. https://www.nature.com/articles/s41597-020-00702-z

Fricke R, Eschmeyer WN, Van der Laan R. 2024. Eschmeyer's Catalog of Fishes: Genera, Species, References. California Academy of Sciences, San Francisco, USA. http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp

Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C. 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends in Ecology and Evolution, 27(9): 480-488. https://www.sciencedirect.com/science/article/pii/S0169534712001000?via%3Dihub

Gaston KJ. 2000. Global patterns in biodiversity. Nature, 405(6783): 220-227. https://www.nature.com/articles/35012228

GBIF. 2025. Global Biodiversity Information Facility. Free and open access to biodiversity data. https://www.gbif.org/

Gotelli NJ, Colwell RK. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecology Letters, 4(4): 379-391. https://onlinelibrary.wiley.com/doi/10.1046/j.1461-0248.2001.00230.x

Hamilton AJ, Basset Y, Benke, KK, et al. 2010. Quantifying uncertainty in estimation of tropical arthropod species richness. The American Naturalist, 176(1): 90-95. https://www.journals.uchicago.edu/doi/10.1086/652998

Hamilton AJ, Novotný V, Waters EK, et al. 2012. Estimating global arthropod species richness: Refining probabilistic models using probability bounds analysis. Oecologia, 171: 357-365. https://link.springer.com/article/10.1007/s00442-012-2434-5

Hawksworth, D.L. 1991. The fungal dimension of biodiversity: magnitude, significance, and conservation. Mycological Research, 95(6): 641-655. https://www.sciencedirect.com/science/article/abs/pii/S0953756209808101?via%3Dihub

Hawksworth DL, Kalin-Arroyo MT. 1995. Magnitude and distribution of biodiversity. In: Global Biodiversity Assessment (Heywood VH, ed). 107-192, Cambridge University Press, Cambridge, UK. https://www.azbackroads.com/wp-content/uploads/2023/04/Global-Biodiversity-Assessment.pdf

Hawksworth DL, Lücking R. 2017. Fungal diversity revisited: 2.2 to 3.8 million species. Microbiology Spectrum, 5(4): FUNK-0052-2016. https://journals.asm.org/doi/10.1128/microbiolspec.funk-0052-2016

Hebert PDN, Penton EH, Burns JM, et al. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. Proceedings of the National Academy of Sciences, 101(41): 14812-14817. https://www.pnas.org/doi/full/10.1073/pnas.0406166101

Hebert, P.D.N., Ratnasingham, S., Zakharov, E.V., et al. 2016. Counting animal species with DNA barcodes: Canadian insects. Philosophical Transactions of the Royal Society B, 371(1702): 20150333. https://royalsocietypublishing.org/rstb/article/371/1702/20150333/22985/Counting-animal-species-with-DNA-barcodes-Canadian

Hedges LV, Olkin I. 2014. Statistical Methods for Meta-Analysis. Academic Press, Orlando, FL, USA. https://books.google.com.sg/books/about/Statistical_Methods_for_Meta_Analysis.html?id=7GviBQAAQBAJ&redir_esc=y

Heywood VH. 1995. Global Biodiversity Assessment. Cambridge University Press, Cambridge, UK. https://digitallibrary.un.org/record/205143

Hodda M, Peters L, Traunspurger W. 2009. Nematode diversity in terrestrial, freshwater aquatic and marine systems. In: Nematodes as Environmental Indicators (Wilson MJ, Kakouli-Duarte T, eds). 45-93, CABI, Wallingford, UK. https://www.cabidigitallibrary.org/doi/10.1079/9781845933852.0045

Hortal J, Borges PAV, Gaspar C. 2006. Evaluating the performance of species richness estimators: Sensitivity to sample grain size. Journal of Animal Ecology, 75(1): 274-287. https://besjournals.onlinelibrary.wiley.com/doi/10.1111/j.1365-2656.2006.01048.x

Hupało K, Copilaş-Ciocianu D, Leese F, Weiss M. 2022. Morphology, nuclear SNPs and mate selection reveal that COI barcoding overestimates species diversity in a Mediterranean freshwater amphipod by an order of magnitude. Cladistics, 39: 129-143. https://onlinelibrary.wiley.com/doi/10.1111/cla.12520

IndexFungorum. 2025. Index Fungorum. The Royal Botanic Gardens, Kew, UK. http://www.indexfungorum.org

Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of https://www.nature.com/articles/s41467-018-07641-9

Joppa LN, Roberts DL, Pimm SL. 2011a. How many species of flowering plants are there? Proceedings of the Royal Society B, 278(1705): 554-559. https://royalsocietypublishing.org/rspb/article/278/1705/554/73460/How-many-species-of-flowering-plants-are-there-How

Joppa LN, Roberts DL, Myers N, Pimm SL. 2011b. Biodiversity hotspots house most undiscovered plant species. Proceedings of the National Academy of Sciences, 108(32): 13171-13176. https://www.pnas.org/doi/full/10.1073/pnas.1109389108

Konstantinidis KT, Rosselló-Móra R, Amann R. 2017. Uncultivated microbes in need of their own taxonomy. The ISME Journal, 11(11): 2399-2406.https://www.nature.com/articles/ismej2017113

Korshunova T, Picton B, Furfaro G, et al. 2019. Multilevel fine-scale diversity challenges the 'cryptic species' concept. Scientific Reports, 9: 6732. https://www.nature.com/articles/s41598-019-42297-5

Kuussaari M, Bommarco R, Heikkinen RK, et al. 2009. Extinction debt: A challenge for biodiversity conservation. Trends in Ecology and Evolution, 24(10): 564-571. https://www.sciencedirect.com/science/article/pii/S0169534709001918?via%3Dihub

Lees AC, Pimm SL. 2015. Species, extinct before we know them? Current Biology, 25(5): R177-R180. https://www.sciencedirect.com/science/article/pii/S0960982214016182?via%3Dihub

Li X, Wiens JJ. 2023. Estimating global biodiversity: The role of cryptic insect species. Systematic Biology, 72(2): 391-403. https://academic.oup.com/sysbio/article/72/2/391/6775868

Locey KJ, Lennon JT. 2016. Scaling laws predict global microbial diversity. Proceedings of the National Academy of Sciences, 113(21): 5970-5975. https://www.pnas.org/doi/full/10.1073/pnas.1521291113

Louca S. 2022. The rates of global bacterial and archaeal dispersal. The ISME Journal, 16(1): 159-167. https://academic.oup.com/ismej/article/16/1/159/7474160

Louca S, Mazel F, Doebeli M, Parfrey LW. 2019. A census-based estimate of Earth's bacterial and archaeal diversity. PLoS Biology, 17(2): e3000106. https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000106

Lughadha EN, Govaerts R, Belyaeva I, et al. 2016. Counting counts: revised estimates of numbers of accepted species of flowering plants, seed plants, vascular plants and land plants with a review of other recent estimates. Phytotaxa, 272(1): 82-88. https://phytotaxa.mapress.com/pt/article/view/phytotaxa.272.1.5

Lürig MD, Donoughe S, Svensson EI, Porto A, Tsuboi M. 2021. Computer vision, machine learning, and the promise of phenomics in ecology and evolutionary biology. Frontiers in Ecology and Evolution, 9: 642774. https://www.frontiersin.org/journals/ecology-and-evolution/articles/10.3389/fevo.2021.642774/full

Magurran AE, McGill BJ. 2011. Biological Diversity: Frontiers in Measurement and Assessment. Oxford

University Press, USA. https://www.amazon.com/Biological-Diversity-Frontiers-Measurement-Assessment/dp/0199580677

Mahé F, de Vargas C, Bass D, et al. 2017. Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. Nature Ecology and Evolution, 1: 0091. https://www.nature.com/articles/s41559-017-0091

Mammal Diversity Database (MDD). 2024. Mammal Diversity Database. Version 1.13. Zenodo. https://zenodo.org/

May RM. 1988. How many species are there on Earth? Science, 241(4872): 1441-1449. https://www.science.org/doi/10.1126/science.241.4872.1441

Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. 2011. How many species are there on Earth and in the ocean? PLoS Biology, 9(8): e1001127. https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001127

Niskanen T, Lücking R, Dahlberg A, et al. 2023. Pushing the frontiers of biodiversity research: Unveiling the global diversity, distribution, and conservation of fungi. Annual Review of Environment and Resources, 48: 149-176. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4636935

Ovaskainen O, Abrego N. 2020. Joint Species Distribution Modelling: With Applications in R. Cambridge University Press, Cambridge, UK. https://www.cambridge.org/core/books/joint-species-distribution-modelling/0D9FA93EA1DD408332A17266449668B3

Paradis, E. 2011. Analysis of Phylogenetics and Evolution with R. 2nd ed. Springer, New York, NY, USA. https://www.amazon.com/Analysis-Phylogenetics-Evolution-Emmanuel-Paradis/dp/1461417422

Parks DH, Chuvochina M, Rinke C, et al. 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Research, 50(D1): D785-D794. https://academic.oup.com/nar/article/50/D1/D785/6370255?login=false

Pfenninger M, Schwenk K. 2007. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. BMC Evolutionary Biology, 7: 121. https://link.springer.com/article/10.1186/1471-2148-7-121

Pimm SL, Raven PH. 2017. The fate of the world's plants. Trends in Ecology and Evolution, 32(5): 317-320. https://www.sciencedirect.com/science/article/pii/S0169534717300502?via%3Dihub

Pimm SL, Jenkins CN, Abell R, et al. 2014. The biodiversity of species and their rates of extinction, distribution, and protection. Science, 344(6187): 1246752. https://www.science.org/doi/10.1126/science.1246752

Plaisance L, Caley MJ, Brainard RE, Knowlton N. 2011. The diversity of coral reefs: what are we missing? PLoS ONE, 6(10): e25026. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0025026

Potts SG, Imperatriz-Fonseca V, Ngo HT, et al. 2016. Safeguarding pollinators and their values to human well-being. Nature, 540: 220-229. https://www.nature.com/articles/nature20588

Preston FW. 1948. The commonness, and rarity, of species. Ecology, 29(3): 254-283. https://esajournals.onlinelibrary.wiley.com/doi/10.2307/1930989

Ricotta C, Ferrari M, Avena G. 2002. Using the scaling behaviour of higher taxa for the assessment of species richness. Biological Conservation, 107(1): 131-133. https://www.sciencedirect.com/science/article/pii/S0006320702000459

Rosselló-Mora R, Amann R. 2015. Past and future species definitions for Bacteria and Archaea. Systematic and Applied Microbiology, 38(4): 209-216.

https://www.sciencedirect.com/science/article/pii/S0723202015000223?via%3Dihub

Runge MC, Converse SJ, Lyons JE. 2011. Which uncertainty? Using expert elicitation and expected value of information to design an adaptive program. Biological Conservation, 144(4): 1214-1223. https://www.sciencedirect.com/science/article/pii/S000632071000529X?via%3Dihub

Sangster, G., Luksenburg JA. 2015. Declining rates of species described per taxonomist: Slowdown of progress or a side-effect of improved quality in taxonomy? Systematic Biology, 64(1): 144-151. https://academic.oup.com/sysbio/article-abstract/64/1/144/2847972?redirectedFrom=fulltext&login=false

Scheffers BR, Joppa LN, Pimm SL, Laurance WF. 2012. What we know and don't know about Earth's missing biodiversity. Trends in Ecology and Evolution, 27(9): 501-510. https://www.sciencedirect.com/science/article/pii/S0169534712001231?via%3Dihub

Schloss PD, Handelsman J. 2006. Toward a census of bacteria in soil. PLoS Computational Biology, 2(7): e92. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.0020092

Schloss PD, Girard RA, Martin T, Edwards J, Thrash JC. 2016. Status of the archaeal and bacterial census: an update. mBio, 7(3): e00201-16. https://journals.asm.org/doi/10.1128/mbio.00201-16

Shapiro BJ, Leducq JB, Mallet J. 2016. What is speciation? PLoS Genetics, 12(3): e1005860. https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1005860

Sinniger F, Pawlowski J, Harii S, et al. 2016. Worldwide analysis of sedimentary DNA reveals major gaps in taxonomic knowledge of deep-sea benthos. Frontiers in Marine Science, 3: 92. https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2016.00092/full

Stackebrandt E, Ebers J. 2006. Taxonomic parameters revisited: tarnished gold standards. Microbiology Today, 33(4): 152-155. https://www.tib.eu/en/search/id/BLSE:RN196956371/Taxonomic-parameters-revisited-tarnished-gold-standards/

Steffen W, Richardson K, Rockström J, et al. 2015. Planetary boundaries: Guiding human development on a changing planet. Science, 347(6223): 1259855. https://www.science.org/doi/10.1126/science.1259855

Storch D, Keil, P, Jetz W. 2012. Universal species–area and endemics–area relationships at continental scales. Nature, 488: 78-81. https://www.nature.com/articles/nature11226

Stork NE. 2018. How many species of insects and other terrestrial arthropods are there on Earth? Annual Review of Entomology, 63: 31-45. https://www.annualreviews.org/content/journals/10.1146/annurev-ento-020117-043348

Stork NE, McBroom J, Gely C, Hamilton AJ. 2015. New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. Proceedings of the National Academy of Sciences, 112(24): 7519-7523. https://www.pnas.org/doi/full/10.1073/pnas.1502408112

Sukumaran J, Knowles LL. 2017. Multispecies coalescent delimits structure, not species. Proceedings of the National Academy of Sciences, 114(7): 1607-1612. https://www.pnas.org/doi/full/10.1073/pnas.1607921114

Tedesco PA, Beauchard O, Bigorne, R, et al. 2017. A global database on freshwater fish species occurrence in drainage basins. Scientific Data, 4: 170141. https://www.nature.com/articles/sdata2017141

Tedersoo L, Bahram M, Põlme S, et al. 2014. Global diversity and geography of soil fungi. Science, 346(6213): 1256688. https://www.science.org/doi/10.1126/science.1256688

Thompson LR, Sanders JG, McDonald D, et al. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Nature, 551: 457-463. https://www.nature.com/articles/nature24621

Thomsen PF, Willerslev E. 2015. Environmental DNA – an emerging tool in conservation for monitoring past and present biodiversity. Biological Conservation, 183: 4-18.

https://www.sciencedirect.com/science/article/pii/S0006320714004443?via%3Dihub

Tikhonov G, Opedal ØH, Abrego N, et al. 2019. Joint species distribution modelling with the R-package Hmsc. Methods in Ecology and Evolution, 11(3): 442-447. https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13345

Tilman D, May RM, Lehman CL, Nowak MA. 1994. Habitat destruction and the extinction debt. Nature, 371: 65-66. https://www.nature.com/articles/371065a0

Uetz P, Hošek J. 2024. The Reptile Database. http://www.reptile-database.org/

Urban MC. 2015. Accelerating extinction risk from climate change. Science, 348(6234): 571-573. https://www.science.org/doi/10.1126/science.aaa4984

van der Heijden MGA, Bardgett RD, van Straalen NM. 2007. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecology Letters, 11(3): 296-310. https://onlinelibrary.wiley.com/doi/10.1111/j.1461-0248.2007.01139.x

Wall DH, Nielsen UN, Six J. 2015. Soil biodiversity and human health. Nature, 528: 69-76. https://www.nature.com/articles/nature15744

Wäldchen J, Mäder P. 2018. Plant species identification using computer vision techniques: A systematic literature review. Archives of Computational Methods in Engineering, 25: 507-543. https://link.springer.com/article/10.1007/s11831-016-9206-z

Wayne LG, Brenner DJ, Colwell, RR, et al. 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. International Journal of Systematic and Evolutionary Microbiology, 37(4): 463-464. https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-37-4-463

Willis A. 2016. Extrapolating abundance curves has no predictive power for estimating microbial biodiversity. Proceedings of the National Academy of Sciences, 113(35): E5096. https://www.pnas.org/doi/full/10.1073/pnas.1608281113

WoRMS Editorial Board. 2025. World Register of Marine Species. https://www.marinespecies.org/

Wu B, Hussain, M, Zhang WW, Stadler M, Liu XZ, Xiang MC. 2019. Current insights into fungal species diversity and perspective on naming the environmental DNA sequences of fungi. Mycology, 10(3): 127-140. https://www.tandfonline.com/doi/full/10.1080/21501203.2019.1614106

Yang ZH, Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. Proceedings of the National Academy of Sciences, 107(20): 9264-9269. https://www.pnas.org/doi/full/10.1073/pnas.0913022107

Zachos FE. 2016. Species Concepts in Biology: Historical Development, Theoretical Foundations and Practical Relevance. Springer, Cham, Switzerland. https://books.google.com.sg/books/about/Species_Concepts_in_Biology.html?id=dNU0DQAAQBAJ&redir_esc=y

Zhang WJ. 2007. Methodology on Ecological Research. Sun Yat-sen University Press, Guangzhou, China

Zhang WJ. 2011. A Java program to test homogeneity of samples and examine sampling completeness. Network Biology, 1(2): 127-129. http://www.iaees.org/publications/journals/nb/articles/2011-1(2)/Java-program-to-test-homogeneity-of-samples.pdf

Zhang WJ. 2017. A new model to describe the relationship between species richness and sample size. Computational Ecology and Software, 7(1): 1-7. http://www.iaees.org/publications/journals/ces/articles/2017-7(1)/relationship-between-species-richness-and-sample-size.pdf

Zhang WJ. 2024a. MetaAnaly: The platform-independent computational tool for meta-analysis in the paradigm of new statistics. Network Biology, 14(2): 187-214. http://www.iaees.org/publications/journals/nb/articles/2024-14(2)/8-Zhang-Abstract.asp

Zhang WJ. 2024b. SampSizeCal: The platform-independent computational tool for sample sizes in the paradigm of new statistics. Network Biology, 14(2): 100-155. http://www.iaees.org/publications/journals/nb/articles/2024-14(2)/5-Zhang-Abstract.asp

Zhang WJ. 2026. probFunCal2: A web-based calculator for calculating probability distribution functions as binomial, Poisson, exponential, gamma, beta, uniform, Weibull, lognormal, Cauchy, geometric, and negative binomial distributions. Selforganizology, 13(3-4): 26-65. http://www.iaees.org/publications/journals/np/articles/2026-11(1-2)/1-Zhang-Abstract.asp

Zhang WJ, Barrion AT. 2006. Function approximation and documentation of sampling data using artificial neural networks. Environmental Monitoring and Assessment, 122: 185-201. https://link.springer.com/article/10.1007/s10661-005-9173-6

Zhang WJ, Schoenly K. 1999. IRRI Biodiversity Software Series. EXTSPP1 and EXTSPP2: Programs for Comparing and Performance-Testing Eight Extrapolation-Based Estimators of Total Taxonomic Richness. IRRI Technical Bulletin No.4. International Rice Research Institute, Manila, Philippines. http://books.irri.org/TechnicalBulletin4_content.pdf