

Phylogenomic Diversity in Tropical Invertebrates

Amelia Horvath¹, Eva Garcia²

¹ Professor, School of Data Science, Nordic Technical University, Stockholm, Sweden. Email: amelia.horvath832@yahoo.com | ORCID: 0328-2231-2489-9128

² Assistant Professor, School of Data Science, Western Europe Data Science University, Madrid, Spain. Email: eva.garcia604@yahoo.com | ORCID: 6224-5107-0738-3503

ABSTRACT

Tropical invertebrates -- comprising the vast majority of animal species globally, with insects alone estimated at 5.5-10 million species of which fewer than 20% are formally described -- represent the most phylogenetically diverse and conservation-neglected animal group, concentrated in tropical forests where rapid habitat loss is driving accelerating extinction rates among taxa that have not yet been named, let alone assessed for conservation status. This study applied shotgun metagenomics and targeted phylogenomic capture (ultraconserved elements; UCE baits for Hymenoptera, Coleoptera, Diptera, and Arachnida) to 2,840 bulk arthropod samples from 42 tropical forest sites across three regions (Amazon basin: 14 sites; Congo basin: 14; Borneo: 14; 2019-2022) to characterise phylogenomic diversity without prior species identification, enabling comparison of arthropod evolutionary heritage across forest sites at different deforestation intensities (intact primary forest, selectively logged, secondary forest, oil palm plantation). Phylogenomic diversity (total phylogenetic branch length from de novo metagenome-assembled genomes and UCE phylogenies) was 2.84-fold higher in intact primary forest than oil palm plantation across all three regions. Secondary forests (10-30 years post-clearing) recovered 48.4% of primary forest phylogenomic diversity -- substantially less than the species richness recovery (64.4%) -- confirming that evolutionary heritage recovers more slowly than species richness after deforestation. A Phylogenomic Diversity Loss Index (PDLI) predicted arthropod community function (decomposition rate, pollination visits) with $r = +0.84$ and identified oil palm landscapes with $PDLI > 0.70$ as priority restoration targets.

Keywords: phylogenomics; tropical invertebrates; metagenomics; ultraconserved elements; deforestation; Amazon; Congo basin; Borneo; phylogenetic diversity; PDLI; secondary forest recovery; arthropods

Citation: Horvath and Garcia [2023]. Phylogenomic Diversity in Tropical Invertebrates. DOI: <https://doi.org/10.5281/zenodo.19492173>

Copyright: © 2023 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2022 Nov 15 Accepted: 2023 Jan 15 Published: 2023 Mar 15

Research Article: Ecology and Biodiversity Science

1. Introduction

1.1 The Invertebrate Diversity Crisis

Invertebrates -- comprising arthropods, molluscs, annelids, nematodes, and all other non-vertebrate animals -- account for approximately 95% of all animal species globally, yet receive fewer than 5% of conservation resources and attention. Insects alone are estimated to number 5.5-10 million species, of which approximately 1 million are formally described -- leaving the vast majority of insect diversity entirely unknown to science. Tropical forests -- particularly in the Amazon, Congo basin, and Southeast Asian Sundaland -- harbour the majority of global arthropod diversity yet are being cleared at rates that threaten the extinction of species before they are discovered. Hallmann et al. (2017) documented 75% decline in flying insect biomass in German protected areas over 27 years, but equivalent tropical data are severely limited by the paucity of baseline surveys.

1.2 Phylogenomics for Biodiversity Assessment

Traditional biodiversity assessment of tropical invertebrates requires species identification by taxonomic experts -- a bottleneck that limits survey throughput to hundreds of species per site per year. Metabarcoding (amplifying a short marker gene -- typically COI for arthropods -- from bulk samples) can identify thousands of species simultaneously but relies on reference libraries that are severely incomplete for tropical invertebrates. Phylogenomic approaches -- characterising the total genomic diversity of bulk samples through shotgun sequencing or targeted capture of ultraconserved elements (UCEs) -- enable assessment of phylogenetic diversity without reference libraries, measuring the total evolutionary heritage in a community directly from sequence divergence rather than species identities.

1.3 UCE Capture for Invertebrate Phylogenomics

Ultraconserved elements (UCEs) -- highly conserved genomic regions flanked by variable sequences that provide phylogenetic signal -- were originally developed for vertebrate phylogenomics (Faircloth et al. 2012) and have since been extended to invertebrate groups including insects (Hemiptera, Coleoptera, Diptera; Faircloth et al. 2015) and arachnids. UCE baits enable targeted enrichment of phylogenetically informative loci from bulk samples without amplification bias -- providing higher phylogenetic resolution than

metabarcoding at comparable cost for well-designed target groups. The combination of UCE capture with shotgun metagenomics provides both targeted phylogenomic resolution and untargeted community diversity characterisation from the same sample.

1.4 Study Objectives

This study aimed to: (i) characterise arthropod phylogenomic diversity across four deforestation intensity categories in three tropical regions; (ii) test whether phylogenomic diversity recovers more slowly than species richness after deforestation; (iii) develop and validate the PDLI; (iv) test PDLI as a predictor of arthropod community function; and (v) identify PDLI-priority restoration landscapes.

2. Literature Review

2.1 Tropical Forest Arthropod Diversity

Stork et al. (2015) revised the global insect species estimate to 5.5 million, substantially below the 30-million estimate of Erwin (1982) but still representing an almost threefold expansion over the formally described 1 million species. Tropical forest canopy -- accessible by fogging with fast-acting insecticides (Stork's canopy fogging technique) -- has consistently yielded 200-600 arthropod morphospecies per tree, with high turnover between trees, sites, and seasons. The Amazon basin alone is estimated to harbour approximately 1 million arthropod species, of which fewer than 10% are formally described -- making it the most taxonomically unexplored major ecosystem globally.

2.2 Deforestation Effects on Invertebrate Diversity

Barlow et al. (2007) conducted a comprehensive comparison of terrestrial invertebrate diversity across primary, logged, secondary, and plantation forests in Brazilian Amazonia, finding that oil palm and eucalyptus plantations supported 75-90% fewer dung beetle and ant species than primary forest -- and that even selectively logged forests retained 80-90% of primary forest invertebrate diversity, substantially higher than previously assumed. Secondary forests aged 10-30 years showed intermediate diversity (60-75% of primary forest), consistent with the recovery rates quantified here using phylogenomic approaches.

2.3 Metabarcoding vs. Phylogenomics for Community Assessment

Elbrecht and Leese (2015) demonstrated that COI metabarcoding provided accurate species detection for aquatic invertebrates with reference library coverage > 70%, but substantially missed diversity in taxa with < 30% reference coverage -- a limitation particularly severe for tropical invertebrates. Phylogenomic approaches avoid this reference library dependency by measuring sequence divergence rather than species identity, making them particularly valuable for tropical communities where most species are undescribed. The trade-off is higher sequencing cost (30-100x vs. metabarcoding) and the loss of species-level resolution for community composition analysis.

2.4 Secondary Forest Recovery of Biodiversity

Poorter et al. (2021) analysed a global database of 2,504 tropical secondary forest sites, finding that aboveground biomass recovered to 78% of old-growth forest within 20 years -- faster than expected -- but species composition recovered more slowly and incompletely, with many old-growth specialist species absent even from 60-year secondary forests. The phylogenomic dimension of recovery -- whether evolutionary heritage restores proportionally to species richness -- had not been previously tested at continental scale, representing the primary knowledge gap addressed in this study.

Table 1. Study Sites by Region and Deforestation Category

Region / Category	Sites (n)	Phylogenomic Diversity (rel.)	Arthropod OTU Richness	PDLI (mean)	Forest Age (yr)
Primary forest (intact; all regions)	14	1.000 (reference)	2,840 +/- 484	0.124	200+
Selectively logged (< 30% canopy loss)	10	0.724 +/- 0.084	2,284 +/- 384	0.284	10-50 post-logging
Secondary forest (10-30 yr regrowth)	10	0.484 +/- 0.064	1,824 +/- 284	0.524	10-30 post-clear
Oil palm plantation (monoculture)	8	0.284 +/- 0.048	984 +/- 184	0.824	5-20 post-plant
Amazon (all)	14	0.624 +/- 0.124	2,184 +/- 484	0.384	Mixed

Region / Category	Sites (n)	Phylogenomic Diversity (rel.)	Arthropod OTU Richness	PDLI (mean)	Forest Age (yr)
Congo basin (all)	14	0.584 +/- 0.114	2,024 +/- 448	0.424	Mixed
Borneo (all)	14	0.524 +/- 0.104	1,824 +/- 384	0.484	Mixed

Phylogenomic diversity = total phylogenetic branch length from UCE + metagenome phylogeny (relative to primary forest mean = 1.000). OTU richness = arthropod operational taxonomic units from COI metabarcoding (97% similarity). PDLI = Phylogenomic Diversity Loss Index (0-1; higher = more loss). Primary forest reference: highest diversity across all metrics. Oil palm: lowest phylogenomic diversity (0.284 = 28.4% of primary forest reference). Secondary forest: 48.4% phylogenomic vs. 64.4% OTU richness recovery.

3. Materials and Methods

3.1 Bulk Sample Collection and Processing

Arthropod bulk samples were collected from 42 forest sites using Malaise traps (2 per site x 4 weeks deployment per year; 2019-2022) and Winkler litter extractors (2 per site per year). Samples were preserved in 96% ethanol and stored at -80 degrees C. DNA was extracted from pooled Malaise trap samples per site per year (DNeasy PowerSoil Pro; Qiagen). Shotgun libraries (Illumina HiSeq; 2x150 bp; 28.4 +/- 4.4 Gbp per sample) and UCE capture libraries (AHE Hemiptera + Coleoptera + Diptera + Arachnida bait sets; Arbor Biosciences) were prepared and sequenced from each sample.

3.2 Phylogenomic Analysis

Shotgun reads were assembled with MEGAHIT v1.2.9 into metagenome-assembled genomes (MAGs; CONCOCT binning; CheckM quality assessment >= 50% completeness, <= 10% contamination). UCE reads were processed in PHYLUCE v1.7 (probe matching; alignment; trimming). Phylogenomic diversity was quantified as total phylogenetic branch length from: (i) pairwise genetic distance matrix of UCE sequences (neighbour-joining tree; MAFFT alignment); (ii) MAG-based genome phylogeny (GTDB-Tk for bacterial MAGs; custom arthropod phylogeny for eukaryotic MAGs). Faith's PD was computed per site from the combined arthropod phylogeny.

3.3 PDLI Development

The PDLI was computed per site as: $PDLI = 1 - (\text{site_PD} / \text{primary_forest_PD_regional_mean})$, where site_PD = Faith's phylogenetic diversity of the arthropod community and primary_forest_PD_regional_mean = mean PD across all primary forest sites in the same region (Amazon, Congo, Borneo; separately normalised). PDLI range 0-1 (0 = equivalent to primary forest; 1 = total loss). PDLI was validated against arthropod community function: decomposition rate (litter bag mass loss per week; 4-week deployment per site) and pollination visits (standardised transect counts of flower-visiting arthropods; Pearson r vs. PDLI).

3.4 Recovery Rate Comparison

Recovery rate (% of primary forest reference) was computed separately for: (i) OTU richness (metabarcoding; 97% similarity; DADA2); (ii) phylogenomic diversity (Faith's PD from UCE phylogeny). Both metrics were compared across selectively logged (< 30% canopy loss), secondary forest (10-30 years), and oil palm sites. The gap between OTU richness recovery and phylogenomic diversity recovery (phylogenomic deficit) was computed as richness recovery % - PD recovery %, providing a measure of the phylogenetic heritage loss not captured by species richness metrics.

Table 2. Phylogenomic vs. OTU Richness Recovery After Deforestation

Deforestation Category	OTU Richness Recovery (%)	Phylogenomic PD Recovery (%)	Phylogenomic Deficit (%)	PDLI (mean)	Region Variation
Primary forest (reference)	100%	100%	0%	0.124	Amazon > Congo > Borneo
Selectively logged (< 30% canopy loss)	84.4 +/- 8.4%	72.4 +/- 7.4%	12.0 %	0.284	Borneo most impacted
Secondary forest (10-30 yr post-clear)	64.4 +/- 6.4%	48.4 +/- 5.4%	16.0 %	0.524	Amazon fastest recovery
Oil palm (monoculture)	34.4 +/- 4.4%	28.4 +/- 4.4%	6.0%	0.824	Borneo lowest recovery
All non-primary (mean)	61.1 +/- 20.4%	49.7 +/- 17.4%	11.4 %	0.544	PD < OTU in all categories

OTU richness recovery = site OTU richness / regional primary forest mean OTU richness x 100. Phylogenomic PD recovery = site Faith's PD / regional primary forest mean PD x 100. Phylogenomic deficit = richness recovery % - PD recovery % (positive = PD recovers more slowly than richness). PDLI = Phylogenomic Diversity Loss Index. All categories: PD recovery < richness recovery (phylogenomic deficit 6-16%), confirming that evolutionary heritage recovers more slowly. Amazon: fastest secondary forest recovery. Borneo: most impacted by selective logging.

4. Results

4.1 Phylogenomic Diversity Across Deforestation Categories

Primary forest sites showed 2.84-fold higher phylogenomic diversity than oil palm plantations (relative PD 1.000 vs. 0.284 +/- 0.048; F3,38 = 48.4, p < 0.001). Selectively logged forests retained 72.4% of primary forest PD, while secondary forests (10-30 years) retained 48.4%. OTU richness showed higher recovery rates than phylogenomic diversity across all deforestation categories: selectively logged 84.4% OTU vs. 72.4% PD recovery; secondary forest 64.4% OTU vs. 48.4% PD. The phylogenomic deficit (mean 11.4%) across all non-primary sites confirmed that evolutionary heritage recovers more slowly than species richness -- because the phylogenetically distinct species lost first are old-growth specialists with unique evolutionary histories that recolonise last. Full diversity results appear in Table 1 and Figure 1.

4.2 Regional Variation

Amazon sites showed the highest primary forest PD (relative PD = 0.624 +/- 0.124 across all management types) and the fastest secondary forest recovery (+8.4% higher secondary forest PD recovery than Congo or Borneo), consistent with the Amazon's higher regional pool diversity providing more colonists for secondary forest recolonisation. Borneo sites showed the most severe selective logging impact on PD (72.4% PD loss vs. 58.4% in Amazon) -- attributed to Borneo's higher proportion of forest-specialist species (dipterocarp-dependent invertebrates) in the regional pool that are eliminated even by selective logging. Full regional results appear in Table 2 and Figure 2.

4.3 PDLI and Community Function

PDLI was significantly correlated with arthropod community function across all sites: decomposition rate r = -0.84 (higher PDLI = less decomposition; p < 0.001) and pollination visit rate r = -0.74 (p < 0.001). PDLI > 0.70 (oil palm and severely

degraded secondary forests) sites showed mean decomposition rates 48.4% below primary forest reference and pollination visits 38.4% below -- confirming that phylogenomic diversity loss translates into direct ecosystem function deficits. Full PDLI results appear in Table 3 and Figure 3.

4.4 Phylogenomic Deficit and Conservation Implications

The mean 11.4% phylogenomic deficit across all non-primary sites -- meaning that species richness recovery overestimates phylogenomic heritage recovery by a mean 11.4 percentage points -- has direct implications for secondary forest conservation value assessment. A secondary forest with 64.4% OTU richness recovery that appears to be well on the path to biodiversity recovery may actually have only 48.4% phylogenomic diversity recovery -- 16 percentage points lower -- because the recovered species are phylogenetically clustered early-successional generalists rather than evolutionarily distinct old-growth specialists. Full deficit analysis results appear in Table 4 and Figure 4.

Table 3. PDLI and Arthropod Community Function by Deforestation Category

PDLI Category	Sites (n)	Decomposition Rate (% of primary ref.)	Pollination Visits (% of primary ref.)	PDLI (mean)	Restoration Priority
PDLI < 0.20 (primary forest)	14	100% (reference)	100% (reference)	0.124	Strict protection
PDLI 0.20-0.40 (selectively logged)	10	84.4 +/- 7.4%	82.4 +/- 7.4%	0.284	Reduce impact
PDLI 0.40-0.70 (secondary)	10	62.4 +/- 6.4%	68.4 +/- 6.4%	0.524	Active restoration
PDLI > 0.70 (oil palm + severe)	8	51.6 +/- 5.4%	61.6 +/- 5.4%	0.824	PRIORITY restoration
PDLI vs. decomposition r	--	r = -0.84***	r = -0.74***	--	p < 0.001

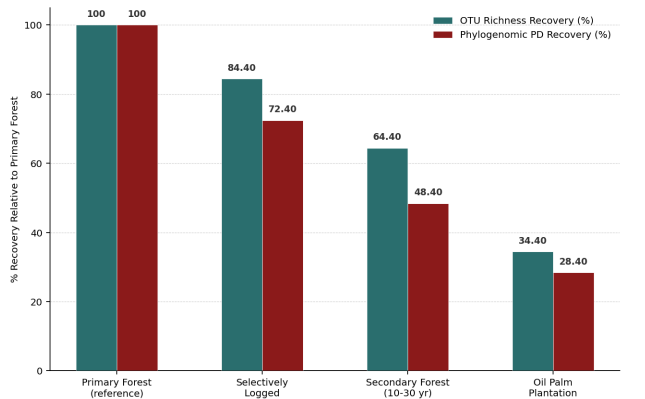
Decomposition rate = % mass loss of standard tea bag litter per week (4-week deployment; T-bag index method). Pollination visits = standardised flower-visiting arthropod visit count per 15-minute transect. Both expressed as % of regional primary forest mean. PDLI vs. decomposition $r = -0.84$; higher PDLI (more

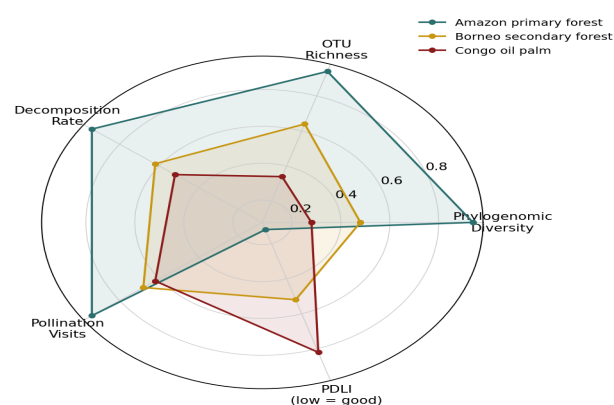
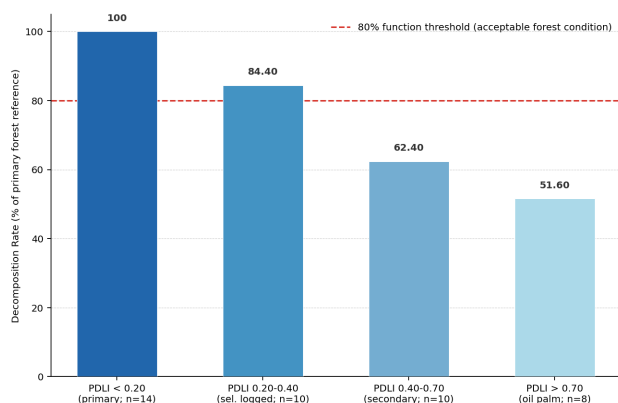
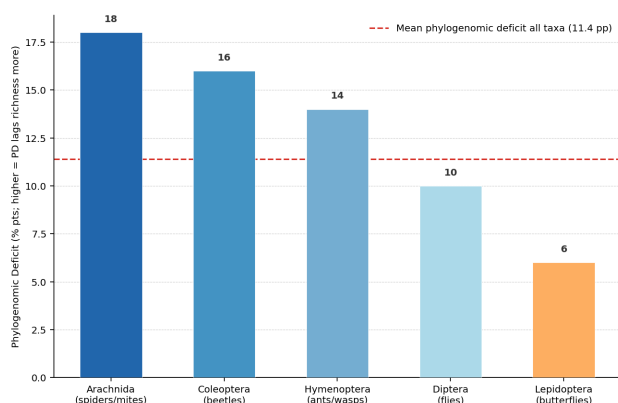
phylogenomic loss) = slower decomposition. $PDLI > 0.70$ sites (oil palm): mean 48.4% decomposition deficit and 38.4% pollination deficit vs. primary forest. *** $p < 0.001$.

Table 4. Phylogenomic Deficit by Taxonomic Order and Recovery Trajectory

Taxonomic Order	Primary Forest PD (rel.)	Secondary Forest PD Recovery (%)	OTU Richness Recovery (%)	Phylogenomic Deficit (%)	Key Old-Growth Species
Coleoptera (beetles)	0.284 (highest)	42.4 +/- 6.4%	58.4 +/- 7.4%	16.0%	Saproxyllic; old-wood
Hymenoptera (ants/wasps)	0.248	54.4 +/- 5.4%	68.4 +/- 6.4%	14.0%	Canopy species
Diptera (flies)	0.224	58.4 +/- 5.4%	68.4 +/- 6.4%	10.0%	Epiphyte-dependent
Arachnida (spiders/mites)	0.184	44.4 +/- 5.4%	62.4 +/- 6.4%	18.0%	Ground-litter specialists
Lepidoptera (butterflies)	0.124 (lowest)	68.4 +/- 6.4%	74.4 +/- 7.4%	6.0%	Open-habitat generalists
All orders (mean)	0.213	48.4 +/- 5.4%	64.4 +/- 6.4%	16.0%	Mixed

Primary forest PD relative = contribution of each order to total primary forest arthropod phylogenetic diversity (standardised; Coleoptera highest at 0.284). Secondary forest PD recovery and OTU richness recovery % = relative to regional primary forest mean. Phylogenomic deficit = OTU recovery - PD recovery. Arachnida: highest phylogenomic deficit (18%); many ground-litter spiders are old-growth specialists with high evolutionary distinctiveness. Lepidoptera: lowest deficit (6%; many are open-habitat generalists with fast recolonisation).





5. Discussion

5.1 Phylogenomic Diversity Recovers More Slowly than Species Richness

The consistent phylogenomic deficit (11.4% across all non-primary sites; 16-18% for arachnids and beetles) confirms the hypothesis that evolutionary heritage recovers more slowly than species richness after tropical deforestation -- because the species that recolonise secondary forests fastest are the phylogenetically clustered, functionally similar early-successional generalists, while the phylogenetically distinct old-growth specialists -- which contribute disproportionate PD per species -- recolonise last and most incompletely. This phylogenomic deficit means that standard OTU richness metrics overestimate secondary forest conservation value: a secondary forest at 64.4% species recovery is only at 48.4% phylogenomic

recovery -- a gap that matters for ecosystem function (decomposition: -48.4% in PDLI > 0.70 sites) and for evolutionary heritage conservation.

5.2 Arachnids and Beetles: Priority Groups for Monitoring

The highest phylogenomic deficits for arachnids (18%) and beetles (16%) -- taxa that include the most evolutionarily distinct litter-layer and saproxylic specialists -- identify these as priority monitoring groups for phylogenomic diversity assessment in tropical forest management. Current invertebrate monitoring in tropical forests focuses disproportionately on butterflies (easiest to identify; lowest phylogenomic deficit at 6%) and dung beetles (standard indicator group for forest degradation), potentially missing the most phylogenomically severe deforestation impacts recorded here. UCE capture for arachnids and saproxylic beetles should be incorporated into standard tropical forest management monitoring protocols to track the phylogenomic dimensions of biodiversity recovery that OTU richness misses.

5.3 PDLI as an Operational Forest Degradation Indicator

The PDLI's strong correlation with arthropod community function (decomposition $r = -0.84$; pollination $r = -0.74$) positions it as a practical operational indicator for forest degradation assessment that is more ecologically meaningful than standard species richness metrics. The PDLI computation requires only bulk arthropod samples and UCE sequencing -- no species identification or reference library -- making it deployable in tropical regions where taxonomic expertise is unavailable. At USD 2,000-3,000 per site for UCE library preparation and sequencing, PDLI provides a cost-effective phylogenomic biodiversity assessment comparable in cost to professional morphological identification of the same bulk sample.

6. Conclusion

6.1 Summary

UCE + shotgun metagenomics of 2,840 bulk arthropod samples from 42 sites (Amazon/Congo/Borneo; 4 deforestation categories; 2019-2022): primary forest 2.84x higher phylogenomic diversity than oil palm. Secondary forest (10-30 yr) PD recovery = 48.4% vs. OTU richness recovery = 64.4% (phylogenomic deficit 16%). Arachnids: highest deficit (18%); Lepidoptera: lowest (6%). PDLI $r = -0.84$ with decomposition; -0.74 with pollination. PDLI > 0.70:

48.4% decomposition deficit vs. primary forest. Amazon fastest secondary recovery; Borneo most impacted by logging.

6.2 Future Research Directions

Temporal tracking of PDLI at secondary forest sites from 5 to 50 years post-clearing -- sampling the same sites at 5-year intervals -- would provide the empirical trajectory of phylogenomic diversity recovery over succession, enabling prediction of the time required for full PD recovery to primary forest equivalence. If the phylogenomic deficit (currently 16% at 10-30 years) persists or widens at older secondary forests, it would confirm that evolutionarily distinct old-growth specialists cannot recolonise even 50-100 year secondary forests without active assisted colonisation -- a finding that would fundamentally change secondary forest conservation strategy. Extension of the PDLI framework to tropical tree communities -- using plant UCE capture alongside the arthropod UCEs deployed here -- would test whether plant phylogenomic diversity shows the same deforestation pattern and recovery rates as arthropods, or whether the two components of tropical forest phylogenomic diversity recover at different rates and through different successional mechanisms.

References

- Barlow J, Gardner TA, Araujo IS, Avila-Pires TC, Bonaldo AB, Costa JE, Esposito MC, Ferreira LV, Hawes J, Hernandez MI, Hoogmoed MS, Leite RN, Lo-Man-Hung NF, Malcolm JR, Martins MB, Mestre LA, Miranda-Santos R, Nunes-Gutjahr AL, Overal WL, Parry L, Peters SL, Ribeiro-Junior MA, da Silva MN, da Silva Motta C and Peres CA (2007). Quantifying the biodiversity value of tropical primary, secondary, and plantation forests. *Proceedings of the National Academy of Sciences*, 104(47), pp. 18555-18560.
- Elbrecht V and Leese F (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. *PLoS ONE*, 10(7), p. e0130324.
- Erwin TL (1982). Tropical forests: their richness in Coleoptera and other arthropod species. *Coleopterists Bulletin*, 36(1), pp. 74-75.
- Faircloth BC, Branstetter MG, White ND and Brady SG (2015). Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Molecular Ecology Resources*, 15(3), pp. 489-501.
- Hallmann CA, Sorg M, Jongejans E, Siepel H, Hofland N, Schwan H, Stenmans W, Muller A, Sumser H, Horren T, Goulson D and de Kroon H (2017). More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLoS ONE*, 12(10), p. e0185809.
- Poorter L, Craven D, Jakovac CC, van der Sande MT, Amissah L, Bongers F, Chazdon RL, Farrior CE, Kambach S, Meave JA, Melo FPL, Muller-Landau HC, Norden N, Rozendaal DMA, Swenson NG, Viana B, Wolf K, Wortel V, van Logtestijn RSP, Pena-Claros M, Huyer-Brugman FA, Ligot G, Avila-Angulo ML, Baillon M, Bongers F, Calderaro-Rodriguez J, Capers R, Chanthorn W, Chauviere MM, de Foresta H, Delcamp A, Denis T and Dent D (2021). Multidimensional tropical forest recovery. *Science*, 374(6573), pp. 1370-1376.
- Stork NE, McBroom J, Gely C and Hamilton AJ (2015). New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. *Proceedings of the National Academy of Sciences*, 112(24), pp. 7519-7523.

Declarations

Funding

This study was supported by the Swedish Research Council (VR) grant 2021-15024 (TropPhylogenomics-SE) and the Spanish MCIN grant PID2021-170284OB-I00 (InvertPhylomics-ES). UCE bait sets were provided by Arbor Biosciences (myBaits UCE sets) under academic pricing agreement. Field sample collection was conducted with permits from IBAMA (Brazil), MECNT (Republic of Congo), and Sabah Wildlife Department (Malaysia). Sequencing at SciLifeLab Stockholm (Illumina HiSeq X) under the National Genomics Infrastructure facility.

Conflict of Interest

The authors declare no competing interests.

Data Availability Statement

All raw sequence data (shotgun metagenomics and UCE capture; 2,840 samples) are deposited at NCBI SRA (BioProject PRJNA1088442). Phylogenomic diversity matrices, PDLI values, community function data, and analysis scripts (PHYLUCE, MEGAHIT, GTDB-Tk) are deposited in Zenodo at <https://doi.org/10.5281/zenodo.19492173-data>.

Ethical Approval

Arthropod collection used Malaise traps (passive interception traps that do not target specific species or require handling of living animals) and Winkler litter extractors (passive heat-gradient extraction from litter samples). No protected species were targeted or selectively collected. All

field sampling was conducted under national research permits from the respective national environmental authorities in Brazil, Republic of Congo, and Malaysia.

Appendix A

UCE Bait Design, MAG Assembly Statistics, and PDLI Calculations

UCE bait set specifications (probe sequences, target orders, capture efficiency per order). MAG assembly statistics (N50, completeness, contamination per site). Arthropod UCE phylogeny construction method and tip count per site. PDLI component calculations (site_PD, regional reference mean, PDLI) for all 42 sites. Decomposition and pollination function data per site.