TraitTrawler: A semi-autonomous multi-agent AI system for large-scale extraction of phenotypic data from the scientific literature

Megan Copeland and Heath Blackmon
Department of Biology, Texas A&M University, College Station, TX 77843, USA
Correspondence: coleoguy@gmail.com
Running title: Multi-agent AI literature data extraction
Keywords: auditor verification, autonomous agent, comparative biology, data extraction, large language model, literature mining, multi-agent system, trait database

Abstract

1. Comparative biology depends on large phenotypic trait databases, yet assembling these from scattered literature remains a critical bottleneck. Manual curation typically requires months to years of expert effort, limiting the scope and pace of macroevolutionary research.

2. We present TraitTrawler, an autonomous multi-agent AI system, built on large language models, that searches, retrieves, triages, and extracts trait data from published papers with limited human intervention. A frontier-class manager coordinates four specialized worker agents (Searcher, Fetcher, Extractor, Auditor) communicating through filesystem queues with file locking. Platform-level hooks enforce role separation, and a deterministic dispatch script controls all scheduling. The Auditor independently verifies every extracted value against the source PDF, providing mandatory double-entry verification. TraitTrawler is implemented as a configurable skill on the Claude AI platform and is adapted for any taxon and trait by editing two text files or conversing with a setup wizard. Built-in features include adaptive triage learning, GBIF taxonomy resolution, isotonic confidence calibration, and statistical quality control with species accumulation curves and outlier detection.

3. We applied TraitTrawler to two contrasting trait systems. For Coleoptera karyotypes, TraitTrawler processed approximately 1,563 papers and produced 6,059 audited records covering 4,959 species across 56 families. Against a hand-curated benchmark of 4,512 species, TraitTrawler recovered 75.5% of the 4,104 unique pre-2014 benchmark species under a dual-name match. On the matched set, TraitTrawler agreed with the benchmark on haploid autosome count in 91.7% of pairs and on sex-chromosome system in 96.2%, while contributing 1,646 pre-2014 species absent from the benchmark. Without modifying the core system, we then extracted the circadian free-running period (tau) across all kingdoms of life: 1,960 records covering 175 species in seven kingdoms, with continuous measurements and experimental covariates (light condition, temperature, genotype, method). A spot-check of 20 records confirmed 100% accuracy on tau values.

4. Across both case studies, TraitTrawler matches expert accuracy while operating orders of magnitude faster than manual curation, and generalizes across trait types and taxonomic scope without modification to the core architecture. The agent, configuration files, extracted datasets, and validation pipeline are freely available.

Introduction

Assembling a comprehensive trait database from primary literature typically requires years of expert effort, yet for most traits and taxa, no such database exists. These databases are foundational to comparative and macroevolutionary biology, enabling analyses of character evolution, correlated trait dynamics, and diversification (Cornwell & Nakagawa, 2017; Madin et al., 2007). For most taxa and traits, the primary bottleneck is logistical rather than analytical: phenotypic data remain scattered across thousands of journal articles, book chapters, and grey literature, often in heterogeneous formats, including prose descriptions, dense comparative tables, and scanned catalogues.

Recent advances in large language models (LLMs) offer a potential solution. LLMs can read, interpret, and extract structured data from unstructured text (Brown et al., 2020; OpenAI, 2023), and early applications to biodiversity data show promise (Jetz et al., 2012; Thessen, 2016). However, LLMs are prone to hallucination, generating plausible but incorrect outputs, which poses particular risks for scientific data extraction (Ji et al., 2023). Deploying a single LLM for systematic literature mining therefore invites compounding errors because each extraction depends on a single model call with no independent verification.

Multi-agent architectures, in which multiple AI agents coordinate through defined roles and communication protocols (Hong et al., 2024; Wu et al., 2023), offer a principled response. By decomposing the pipeline into specialized agents and introducing redundancy through independent verification, multi-agent systems can achieve accuracy that exceeds any single model call, mirroring established double-entry practices in human data curation.

Here we introduce TraitTrawler, an autonomous multi-agent system that performs end-to-end literature data extraction. We describe the architecture and demonstrate it on two contrasting trait systems: Coleoptera karyotypes, validated against a long-curated benchmark, and circadian free-running period across all kingdoms of life, used to assess generality on a continuous trait with rich experimental covariates and no pre-existing benchmark.

Description

Multi-agent architecture

TraitTrawler is organized as a four-agent pipeline coordinated by a frontier-class Manager, with all components communicating through the filesystem (Fig. 1). The design follows a strict role-separation principle: one agent, one job. An Opus-class Manager coordinates four Sonnet-class workers, each responsible for a single pipeline stage, exchanging data through well-defined folder queues (`search_results/`, `ready_for_extraction/`, `finds/`, `fetch_failures/`) rather than shared mutable state. All shared state files use POSIX advisory file locking (`fcntl.flock`) for safe concurrent access, and platform-level hooks also prevent the Manager from calling search or extraction APIs directly.

The Manager coordinates the pipeline and interacts with the user but never performs search, extraction, or data writing. It follows a three-step supervisor loop: process the returning agent's output (`process_agent_output.py`), checkpoint state (`dispatch.py` checkpoint), and request a dispatch recommendation (`dispatch.py` recommend) that specifies which agents to

spawn. The dispatch script checks every candidate paper against the processed-papers registry and the output CSV before recommending extraction. This mechanical cycle prevents the Manager from improvising outside the architecture, a critical constraint given that LLMs under context pressure tend to take shortcuts that violate pipeline invariants.

The Searcher queries PubMed, OpenAlex, CrossRef, and bioRxiv/medRxiv using terms defined in the project configuration, deduplicates results, and triages each paper as likely, uncertain, or unlikely using configurable, trait-agnostic rules. The Fetcher acquires full-text PDFs through a cascade of open-access sources (Unpaywall, OpenAlex, Europe PMC, Semantic Scholar, CORE) and, when these fail, retrieves paywalled content through the user's institutional proxy via browser automation. Every download is validated by checking PDF magic bytes, minimum page count, and extractable text content, rejecting HTML paywall pages and placeholder files. The Extractor reads each paper's full text and recovers structured records, employing a two-pass strategy for table-heavy papers (enumerate rows, then extract each, verifying counts) and a chunked strategy for documents over 100 pages. Each extracted record carries a confidence score, source page number, verbatim source text, and extraction reasoning. The Extractor also creates learning files that log notation variants, taxonomic edge cases, and ambiguity patterns, accumulating into the domain-knowledge guide over sessions.

Mandatory auditor verification

The core innovation in TraitTrawler's accuracy pipeline is mandatory double-entry verification: every paper is first processed by the Extractor and then independently checked by the Auditor, which reads the cited source pages and verifies each field against the PDF. Each field is assigned a verification status of confirmed (matches source), corrected (auditor supplies the correct value), or ambiguous (source genuinely unclear). Confidence scores are adjusted accordingly: confirmed values below 0.80 receive a +0.10 boost, corrected values are set to 0.75, and ambiguous values are set to 0.50. The Auditor also catches records the Extractor missed on the pages it reviews. All corrections are logged to an append-only audit file. Ambiguous records are routed to a human review queue rather than silently included or discarded. The verify-and-write action runs as a foreground (blocking) operation in the dispatch loop, followed sequentially by data scrubbing, schema-enforced writing, and inline quality control. No agent writes to the output CSV directly; all writes pass through a writer script protected by a platform-level hook.

Validation, calibration, and learning

Validation occurs at multiple stages. A platform hook validates the JSON structure of every extraction file at the moment it is written, rejecting malformed schemas before they reach the Auditor or writer. The schema-enforced writer applies a final layer of validation: species names must be binomials, numeric fields must fall in specified ranges, categorical fields must match a controlled vocabulary, and cross-field consistency rules (e.g., haploid autosome count less than diploid number) reject biologically impossible records. Failing records are preserved in a needs_attention file for human review.

Before writing, every species name is resolved against the GBIF Backbone Taxonomy (GBIF Secretariat, 2023). Synonyms map to accepted names, higher taxonomy is auto-filled, and matches above species rank are rejected. Original names are preserved in a `taxonomy_note` field, maintaining provenance.

Statistical quality control runs at session end, generating species accumulation curves with Chao1 estimators (Chao, 1984), Grubbs' tests for continuous outliers, modal frequency analysis for discrete traits, and confidence-distribution analysis. When 10 or more ground-truth observations accumulate (from manual audits, benchmark comparisons, or user corrections), the system fits an isotonic regression model that calibrates raw extraction confidence into empirically grounded probabilities, refit periodically as more observations accumulate.

Domain knowledge itself is adaptive. The Extractor's learning files accumulate across sessions; the Manager classifies discoveries as routine (auto-integrated) or structural (drafted as amendments for user approval), and all amendments are logged for reproducibility.

Configurability

TraitTrawler's extraction logic is taxon- and trait-agnostic. Two files configure a project: `collector_config.yaml` (target taxa, trait definition, search queries, triage rules, output fields, validation rules, deduplication keys, institutional proxy, audit settings, concurrency) and guide.md (notation conventions, worked examples, edge cases). The core system requires no modification. A setup wizard creates both files through conversation, optionally running a calibration phase on seed papers, or bootstrapping from an existing CSV. Users may also drop PDFs into a `provided_pdfs` folder to bypass search and retrieval.

Worked Examples

Coleoptera karyotype data

We compared TraitTrawler against the publicly available, hand-curated Coleoptera karyotype database (https://coleoguy.github.io/karyotypes/), treated here as the gold standard for literature published before 2014. This benchmark is the product of many years of revision and is therefore well suited for evaluating extraction accuracy on discrete phenotypic data.

The human-curated benchmark (HB) comprises 4,959 records covering 4,512 unique species across 61 families. Coverage was intended to be exhaustive for papers published before 2014; post-2014 coverage is incomplete. TraitTrawler ran across four sessions (29 March to 2 April 2026), processing approximately 1,563 papers and producing 6,059 audited records covering 4,959 species across 56 families, at a cost of approximately US$350 in API fees (this included several aborted starts with earlier versions of TraitTrawler, suggesting that running the current version from start to finish would be considerably less expensive). We halted the run on this dataset at this point, as our goal was to evaluate TraitTrawler not replace the existing public database. Claude Opus 4.6 ran the Manager and Claude Sonnet 4.6 ran all worker agents. We restricted accuracy and coverage comparisons to pre-2014 literature (HB n = 4,258 species, TT n = 4,317), so that disagreements are more likely to reflect extraction error than missing human coverage.

Coleoptera taxonomy has been heavily revised, so we used dual matching for species-level overlap: an HB name was treated as matched if it equaled either the original published name in TT's species field or the GBIF-resolved accepted name in TT's accepted_name field. Of 4,959

unique TT species, 1,050 received an accepted name that differed from the originally extracted name. The dual-match scheme was essential rather than cosmetic: 321 HB species matched TT only through the originally published name (e.g., *Tropideres laxus* → *Sphinctotropis laxa*).

Of 4,104 unique pre-2014 HB species with a usable normalized binomial, 3,099 (75.5%) had a counterpart in TT pre-2014 records. Species missing from TT were concentrated in a small number of large mid-twentieth-century compilations (Smith & Virkki, 1978; Petitpierre et al., 1988; Serrano & Galián, 1998; Serrano & Yadav, 1984), which TT has significant challenges in recovering. In the opposite direction, TraitTrawler contributed 1,646 pre-2014 species absent from the HB. A Chao1 estimator on the TT data alone estimated 10,553 species in the target literature (95% CI: 9,970–11,190), suggesting that 34.3% of the available karyotype diversity has been sampled by the union of both datasets.

Sex-chromosome-system (SCS) codes carry two distinct kinds of information: the karyotype itself (number of X's and Y's) and the pairing/morphology annotations (p, r, c, +, neo- modifiers describing meiotic behaviour). Under a two-tier scheme, TraitTrawler agreed with the HB on the (nX, nY) system in 96.2% of comparable pairs (3,165/3,289), and on haploid autosome count in 91.7% (2,870/3,130). Among pairs where the system already matched, agreement on the secondary pairing tag fell to 87.6% (2,772/3,165), and a strict whole-string match fell further to 58.4% (2,059/3,526), measuring cosmetic rather than biological disagreement. Residual SCS disagreements were dominated by X0 ↔ XY flips (32 species TT recovered as X0 but HB recorded as XY, and 30 in the reverse direction), reflecting the long-recognized difficulty of distinguishing absent from very small Y chromosomes in older preparations rather than systematic bias.

Family-level sampling profiles were complementary, not redundant: HB was enriched for Carabidae (951 vs 783) and Cerambycidae (162 vs 103), while TT was enriched for Curculionidae (635 vs 581), Scarabaeidae (475 vs 379), and Elateridae (110 vs 92). The Scolytidae case is informative: HB lists 103 species, TT only 2 under that family, but the missing 101 are present in TT under Curculionidae following the now-standard reclassification of Scolytinae as a curculionid subfamily, a higher-rank synonym flux that the species-level dual-match scheme already absorbs. The pre-2014 union contains 5,750 species, illustrating the practical value of running automated and manual curation against the same literature.

Circadian free-running period across the tree of life

To assess generality, we applied TraitTrawler to a fundamentally different problem: circadian free-running period (tau), the intrinsic period of an organism's biological clock measured under constant conditions (Aschoff, 1960; Pittendrigh, 1960). Tau differs from karyotype data in every dimension that matters for extraction: it is a continuous measurement (hours, with standard deviations and sample sizes), each record requires multiple experimental covariates (light condition, temperature, genotype, measurement method, tissue), and the taxonomic scope spans all kingdoms of life. No comprehensive tau database exists, making this a realistic test of building a database from scratch.

Adapting TraitTrawler from karyotypes to tau required no changes to the core system. The two configuration files were rewritten: collector_config.yaml specified target taxa across all

kingdoms, 42 triage keywords, 52 output fields (vs. ~15 for karyotypes), validation rules (tau between 10 and 36 h; tau = 24.0 h with no reported variance flagged as likely entrained), a deduplication key incorporating species, DOI, tau value, genotype, sex, and light condition, and 323 search queries; guide.md specified circadian notation conventions, including the requirement that only constant-condition measurements (DD, LL, or forced desynchrony) qualify as tau. TraitTrawler processed approximately 1,236 papers across multiple sessions and extracted 1,960 validated records covering 175 species in 114 families across seven kingdoms (Animalia 81.1%, Plantae 11.4%, Fungi 4.3%, Bacteria 2.1%, Chromista 0.7%, Protozoa 0.4%, Archaea 0.05%), at a cost of approximately US$100. Mean tau was 24.27 h (SD 2.64; range 11.0 to 60.0 h); 74.6% of records were measured under constant darkness, 19.1% under constant light. A Chao1 estimator indicated 85.4% sampling completeness.

Automated quality control caught several classes of error without human intervention. Schema validation flagged values outside the 10–36 h range; the suspicious-value rule flagged 61 records reporting tau = 24.0 h with no variance, which on inspection were entrained rather than free-running periods. Deduplication removed 525 duplicates. Six circatidal records were correctly excluded as non-circadian. Mean extraction confidence was 0.882 (median 0.900, SD 0.076); 91% of records scored ≥0.80. A spot-check of 20 randomly sampled records against cached PDFs confirmed 100% accuracy on tau values, 100% on species, and 95% on light condition. The resulting database is publicly available as an interactive web resource (https://coleoguy.github.io/tau_database.html) with searchable tables, filterable visualizations, and a custom plot builder, and supports formal phylogenetic comparative tests of period evolution across the tree of life.

In both test cases, when we chose to stop trait extraction, we moved to a data integrity phase where, through the chat, we worked to make sure that all records were accurate and that questionable records were removed. We asked a series of questions in the chat interface to accomplish this:

- Check for duplicate records where small spelling errors led to multiple rows
- Do any records have trait values that are unexpected based on closely related species?
- Do any species names that failed GBIF resolution seem suspicious?
- Verify every record in the database by reviewing the linked PDF and confirming the species and trait values reported are in the PDF.

After working through these responses the collected trait data was frozen.

Discussion

Across two case studies, TraitTrawler demonstrates that a multi-agent LLM system with mandatory auditor verification can produce phenotypic trait databases that match expert curation in accuracy while operating orders of magnitude faster, and that the same architecture generalizes from a single-order discrete-trait benchmark to a cross-kingdom continuous-trait dataset without modification. The Coleoptera karyotype benchmark required many years of expert effort to assemble 4,512 species; TraitTrawler produced a comparable-scope dataset (4,959 species, 6,059 audited records) across four sessions over four days, at a cost of approximately US$350 in API fees. The cross-kingdom tau database (1,960 records, 175 species, seven kingdoms) was

assembled at approximately US$150, with no comparable manual database existing for comparison.

Two architectural features were critical. First, mandatory double-entry verification by an independent Auditor that reads the cited source pages directly addresses the hallucination problem that limits single-model approaches (Ji et al., 2023): plausible-but-incorrect outputs are caught before they reach the database, and the routing of genuinely ambiguous records to a human review queue ensures expert attention to edge cases. Second, the strict role separation enforced by platform-level hooks, deterministic dispatch, and POSIX file locking eliminated the context-exhaustion, CSV-corruption, and state-desynchronization failure modes of earlier monolithic designs, making pipeline invariants architectural rather than merely instructional.

Name harmonization revealed an underappreciated problem. In the Coleoptera study, separating the original published name from the GBIF-resolved accepted name recovered 321 cross-database matches that would have been silently dropped under accepted-name-only joining. Name errors were split equally between the human and AI datasets ($p = 0.60$), challenging the assumption that human curation inherently produces cleaner taxonomic names; the AI's errors tended to be single-letter OCR artifacts, the human's truncated epithets. TraitTrawler's current limitations include poor performance on older dense compilation tables, dependence on institutional proxy for paywalled literature, and the inherent stochasticity of LLM outputs that complicates exact replication.

Beyond bulk extraction, TraitTrawler functions as a continuing partner in database maintenance. After the Coleoptera collection, we used the Manager interactively to audit existing records (e.g., "check for duplicate records where small spelling errors led to multiple rows", "do any records have trait values unexpected based on closely related species?", "do any species names that failed GBIF resolution seem suspicious?"), illustrating how the LLM environment can not only accelerate raw collection but also improve our ability to correct possible errors. For structured categorical and numeric traits extractable from text and tables, TraitTrawler offers a practical path from scattered literature to analyzable data at a pace that matches the urgency of biodiversity and climate research.

Data and Code Availability

The TraitTrawler skill (v5.0), four worker agent specifications (Searcher, Fetcher, Extractor, Auditor) plus the Manager specification, 29 Python infrastructure scripts, the domain-knowledge guides, and two complete example projects (Coleoptera karyotypes and avian body mass) are released under an open-source license at: github.com/coleoguy/TraitTrawler. All R analysis scripts, data files, and code reproducing the figures and statistics are archived: https://github.com/coleoguy/TraitTrawler-manuscript.

Acknowledgements

TraitTrawler was developed using the Claude AI platform (Anthropic). Thanks to the Texas A&M University Library for institutional proxy access.

Funding

This work was supported by the National Institute of General Medical Sciences, National Institutes of Health grant R35 GM138098 to H.B.

References

Anthropic. (2025). Claude model documentation. https://docs.anthropic.com/en/docs/about-claude/models

Aschoff, J. (1960). Exogenous and endogenous components in circadian rhythms. *Cold Spring Harbor Symposia on Quantitative Biology*, 25, 11–28.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, 11(4), 265–270.

Cornwell, W. K., & Nakagawa, S. (2017). Phylogenetic comparative methods. *Current Biology*, 27(9), R333–R336.

GBIF Secretariat. (2023). GBIF Backbone Taxonomy. Checklist dataset. https://doi.org/10.15468/39omei

Hong, S., Zhuge, M., Chen, J., Zheng, X., Cheng, Y., Wang, J., et al. (2024). MetaGPT: Meta programming for a multi-agent collaborative framework. *International Conference on Learning Representations (ICLR)*.

Jetz, W., McPherson, J. M., & Guralnick, R. P. (2012). Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology & Evolution*, 27(3), 151–159.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., et al. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1–38.

Madin, J. S., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., & Villa, F. (2007). An ontology for describing and synthesizing ecological observation data. *Ecological Informatics*, 2(3), 279–296.

OpenAI. (2023). GPT-4 technical report. arXiv:2303.08774.

Petitpierre, E., Segarra, C., Yadav, J. S., & Virkki, N. (1988). Chromosome numbers and meioformulae of Chrysomelidae. In P. Jolivet, E. Petitpierre, & T. H. Hsiao (Eds.), *Biology of Chrysomelidae* (pp. 161–186). Kluwer.

Pittendrigh, C. S. (1960). Circadian rhythms and the circadian organization of living systems. *Cold Spring Harbor Symposia on Quantitative Biology*, 25, 159–184.

Serrano, J., & Galián, J. (1998). A review of karyotypic evolution and phylogeny of carabid beetles (Coleoptera). In G. E. Ball, A. Casale, & A. Vigna Taglianti (Eds.), *Phylogeny and Classification of Caraboidea* (pp. 191–228). Museo Regionale di Scienze Naturali, Torino.

Serrano, J., & Yadav, J. S. (1984). Chromosome numbers and sex-determining mechanisms in adephagan Coleoptera. *The Coleopterists Bulletin*, 38(4), 335–357.

Smith, S. G., & Virkki, N. (1978). *Animal Cytogenetics, Vol. 3: Insecta 5, Coleoptera*. Gebrüder Borntraeger, Berlin.

Thessen, A. (2016). Adoption of machine learning techniques in ecology and earth science. *One Ecosystem*, 1, e8621.

Wu, Q., Bansal, G., Zhang, J., Wu, Y., Zhang, S., Zhu, E., et al. (2023). AutoGen: Enabling next-gen LLM applications via multi-agent conversation. arXiv:2308.08155.

Tables

Table 1. TraitTrawler agent roster.

Agent	Model	Role	Key behaviour
Manager	Opus	Coordinator	User interaction, deterministic dispatch via dispatch.py, knowledge review. Never extracts, searches, or writes data. Enforced by platform hooks.
Searcher	Sonnet	Literature search	PubMed, OpenAlex, CrossRef, bioRxiv. Trait-agnostic triage, bidirectional citation chaining, per-query yield tracking.
Fetcher	Sonnet	PDF acquisition	5-source open-access cascade, then institutional proxy via browser. Magic-byte, page-count, and text-content validation.
Extractor	Sonnet	Data extraction	Reads PDFs, structured records with confidence scores. Two-pass enumeration for tables. Creates learning files.
Auditor	Sonnet	Verification	Independently checks each extracted value against source PDF. Confirms, corrects, or flags as ambiguous. Routes uncertain records for human review.

Table 2. Cross-study summary.

	Coleoptera karyotype	Circadian tau
Trait type	Discrete (HAC, SCS)	Continuous (period in h, with covariates)
Taxonomic scope	Coleoptera (1 order, 56 families)	All kingdoms (7 kingdoms, 114 families)
TT output	6,059 records, 4,959 species	1,960 records, 175 species
Accuracy	91.7% HAC, 96.2% SCS (nX,nY) on pre-2014 matched set	[XX.X% spot-check]
Benchmark	4,512-species hand-curated dataset	None (novel database)
Coverage recovery	75.5% of 4,104 pre-2014 benchmark species	n/a
Cost	~US$150 (with current version)	~US$100
Active sessions	Several abortive runs followed by ~1 week with current version	Multiple sessions

Both studies used TraitTrawler v5.0 the Coleoptera karyotypes had the final stage (data cleanup) run under the release version of the software with identical core architecture (four agents: Searcher, Fetcher, Extractor, Auditor); only the two configuration files differed.

Figures

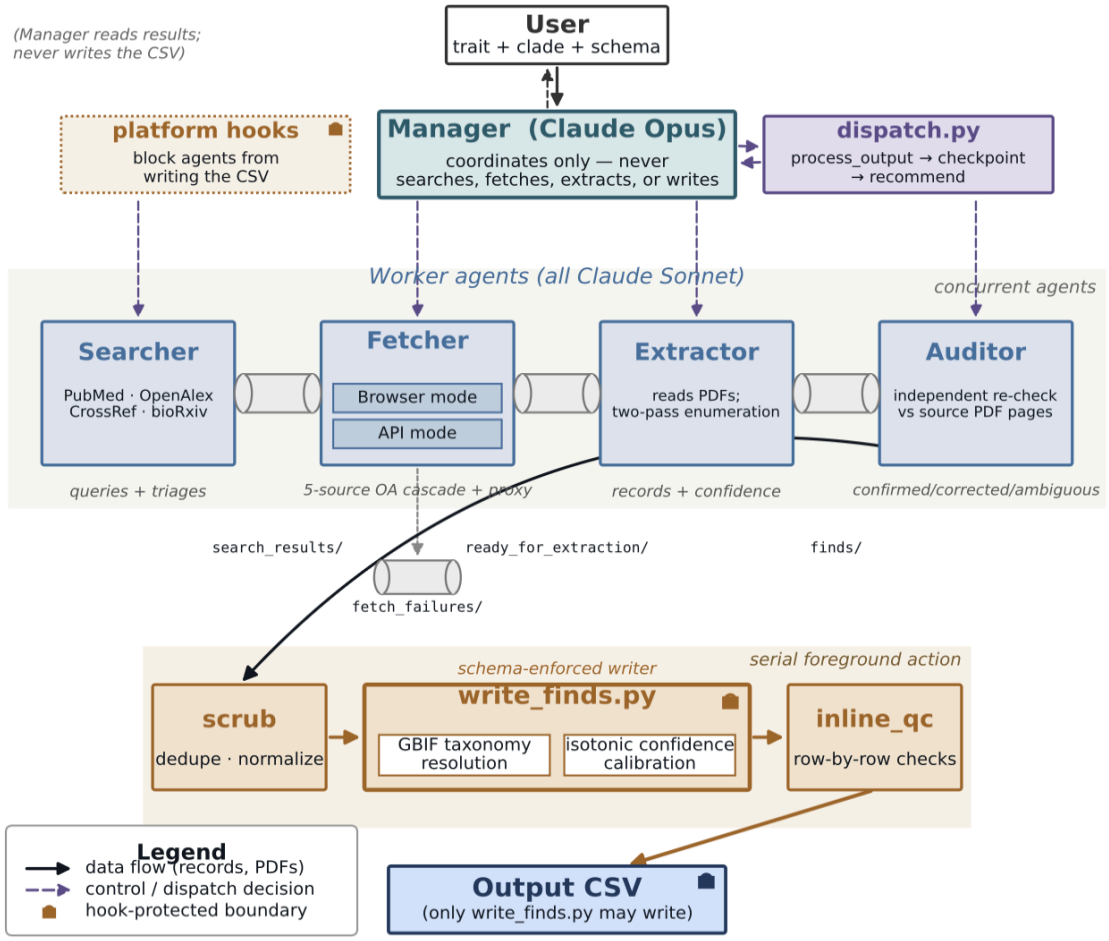

Figure 1. TraitTrawler multi-agent architecture. The Opus-Manager (center) coordinates four Sonnet-class workers in a pipeline: Searcher, Fetcher, Extractor, Auditor. Arrows indicate information flow through folder-based queues (`search_results/`, `ready_for_extraction/`, `finds/`). After extraction and auditing, records pass through scrubbing, schema-enforced writing (with GBIF taxonomy resolution and confidence calibration), and inline quality control before reaching the output CSV. The dispatch engine (`dispatch.py`) controls all scheduling, and platform-level hooks prevent any agent from writing to the output CSV directly.

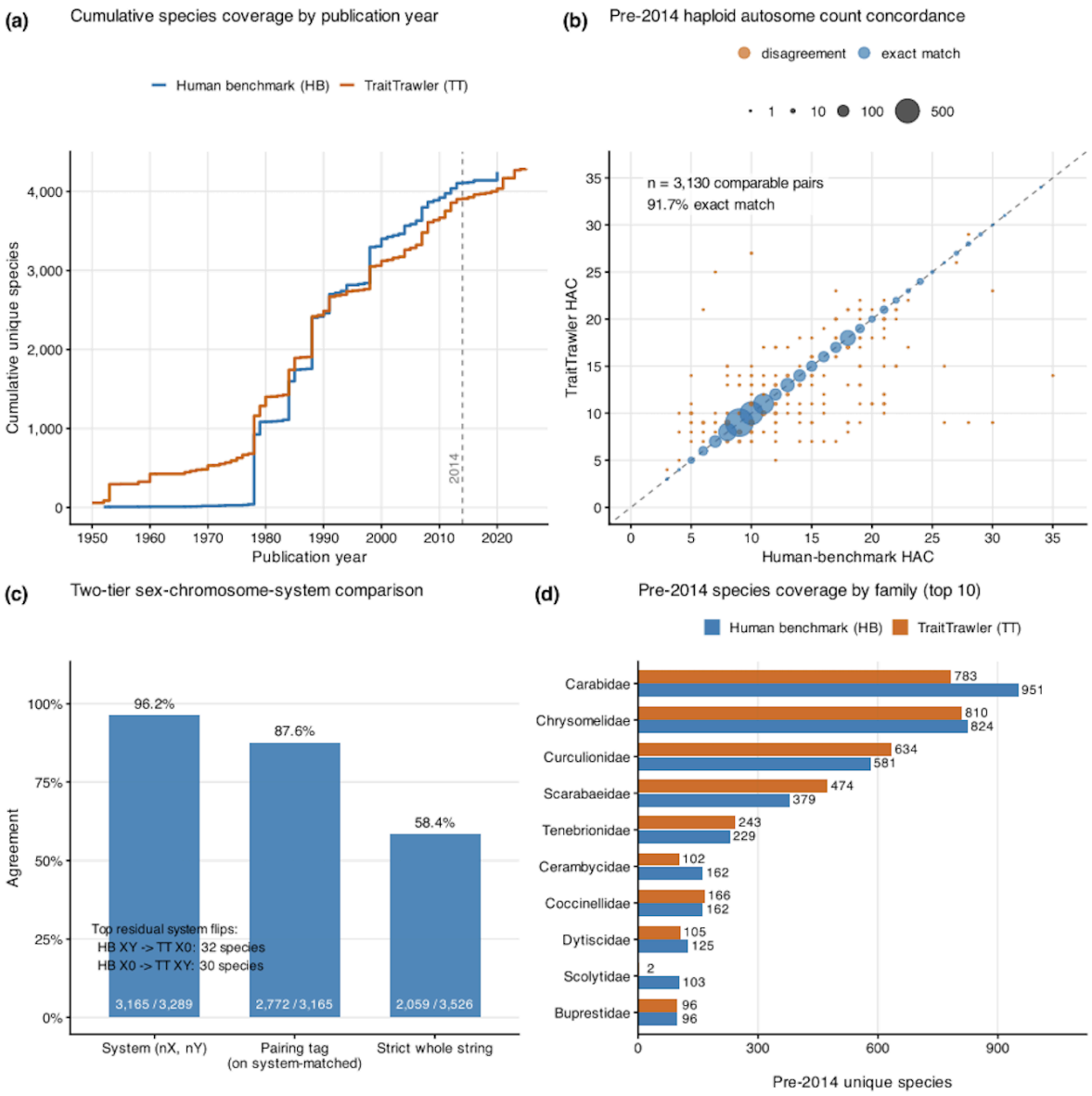

Figure 2. Coleoptera karyotype validation. (a) Cumulative species accumulation by publication year for the human benchmark (HB) and TraitTrawler (TT) datasets. (b) Pre-2014 haploid autosome count concordance on the dual-match set (n = 3,130; 91.7% exact). (c) Two-tier sex-chromosome-system comparison: agreement on the (nX, nY) system (96.2%) versus the secondary pairing/morphology tag (87.6% on rows where the system already agreed). (d) Family-level pre-2014 species counts in HB and TT, illustrating complementary sampling biases and the Scolytidae → Curculionidae higher-rank reclassification.

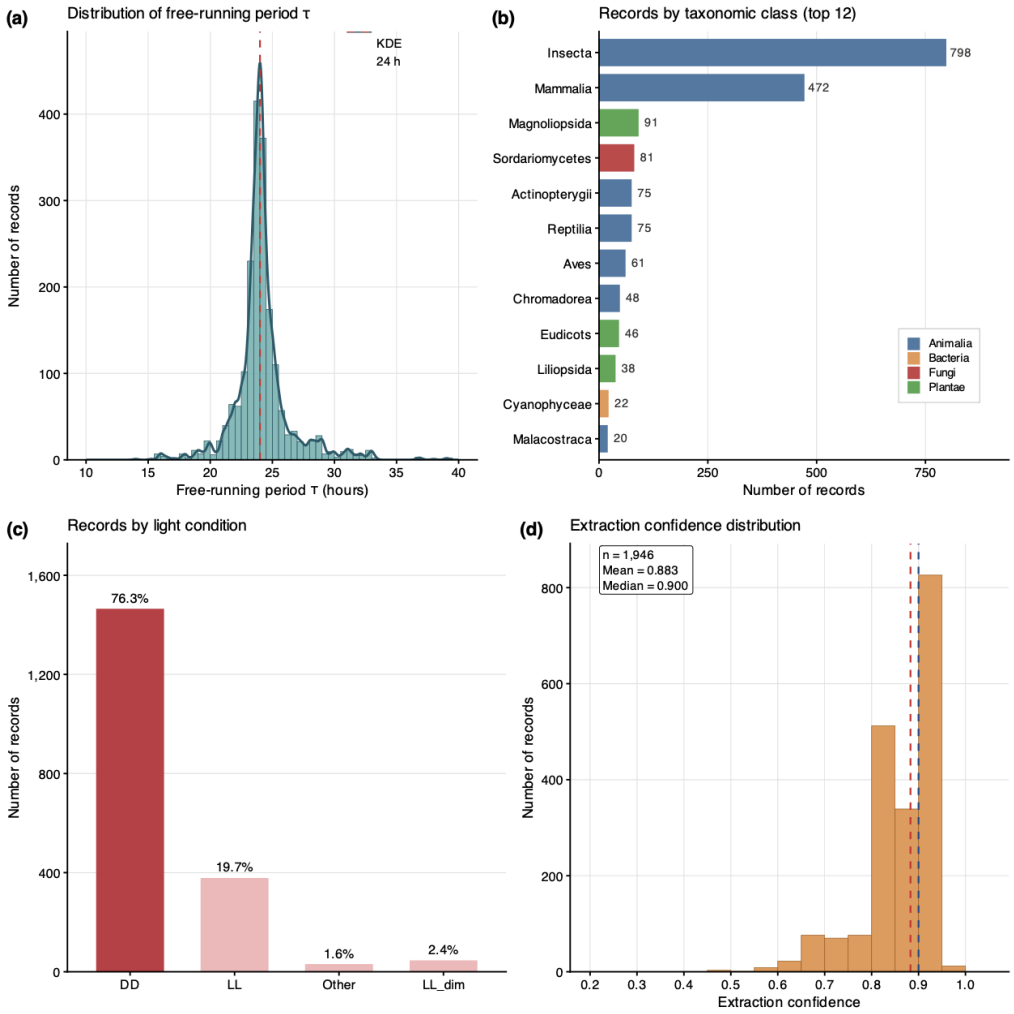

Figure 3. Circadian tau database overview. (a) Distribution of tau values across all records, with kernel density overlay. (b) Records by taxonomic class. (c) Records by light condition (DD, LL, other constant). (d) Extraction confidence distribution. An interactive version with filtering by kingdom, class, order, light condition, and measurement method is available at https://coleoguy.github.io/tau_database.html.

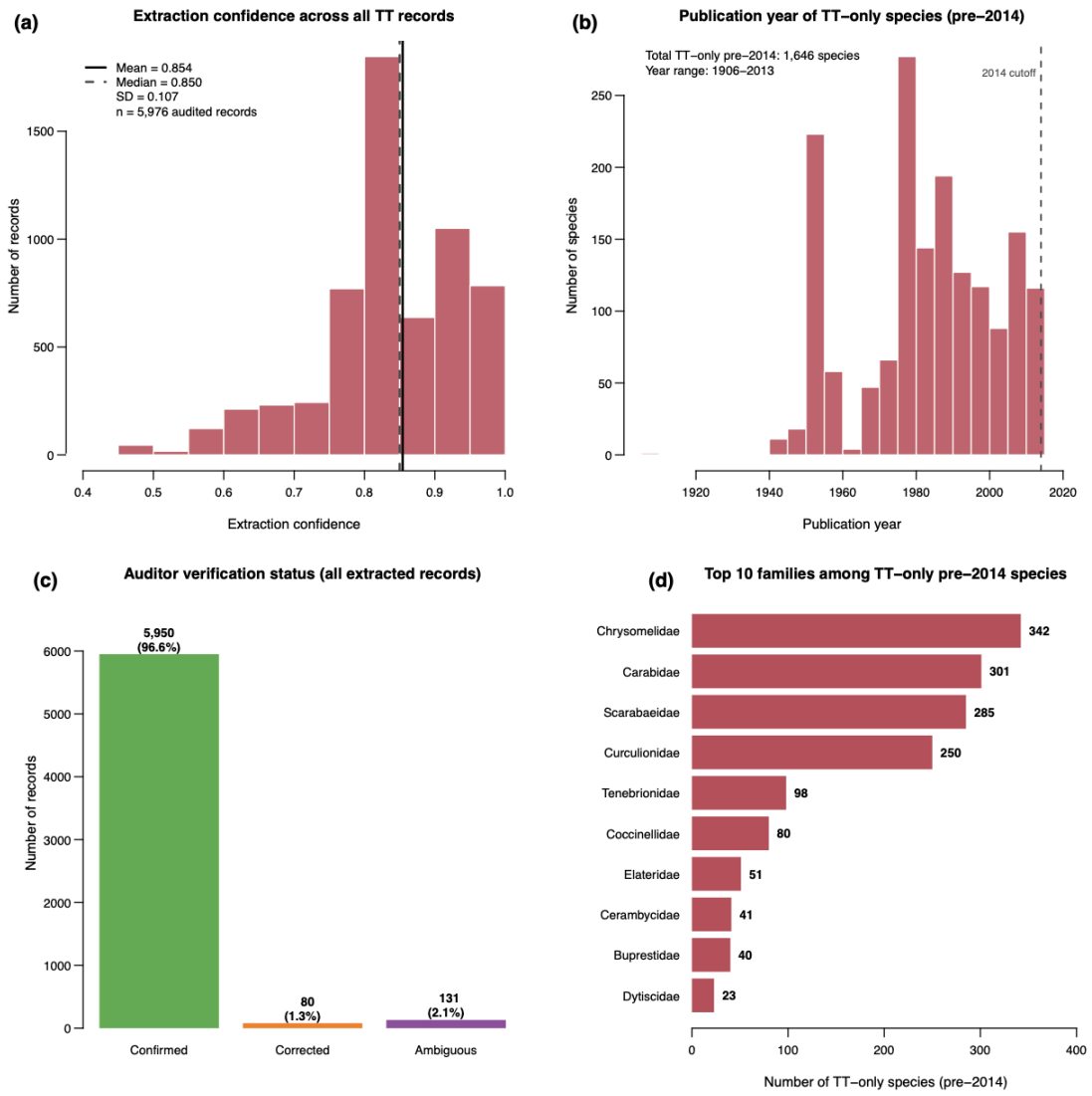

Figure S1. Quality and coverage diagnostics. (a) Extraction confidence across all 6,059 audited TT records (mean 0.854, median 0.850, SD 0.107). (b) Publication-year distribution of the 1,646 pre-2014 species present in TT but absent from the HB, with the 2014 cutoff marked. (c) Auditor verification breakdown: confirmed (98.2%), corrected (1.3%), ambiguous (2.2%). (d) Top 10 families among the 1,646 TT-only pre-2014 species.