

Species Tree Inference on Data with Paralogs is Accurate Using Methods Intended to Deal with Incomplete Lineage Sorting

What is the Manuscript Microscope Sentence Audit?

The Manuscript Microscope Sentence Audit is a research paper introspection system that parses the text of your manuscript into minimal sentence components for faster, more accurate, enhanced proofreading.

Why use a Sentence Audit to proofread your manuscript?

- **Accelerated Proofreading:** Examine long technical texts in a fraction of the usual time.
- **Superior Proofreading:** Detect subtle errors that are invisible to traditional methods.
- **Focused Proofreading:** Inspect each individual sentence component in isolation.
- **Reliable Proofreading:** Ensure every single word of your manuscript is correct.
- **Easier Proofreading:** Take the hardship out of crafting academic papers.

Bonus 1: **Improved Productivity:** Rapidly refine rough drafts to polished papers.

Bonus 2: **Improved Authorship:** Cultivate a clear, concise, consistent, writing style.

Bonus 3: **Improved Reputation:** Become known for rigorously precise publications.

Manuscript Source: <https://www.biorxiv.org/content/10.1101/498378v4>

Manuscript Authors: Zhi Yan, Megan L. Smith, Peng Du, Matthew W. Hahn & Luay Nakhleh

Features of the Sentence Audit:

The Sentence Audit combines two complementary proofreading approaches:

1. Each sentence of your text is parsed and displayed in isolation for focused inspection.
2. Each individual sentence is further parsed into Minimal Sentence Components for a deeper review of the clarity, composition and consistency of the language you used.

The Minimal Sentence Components shown are the smallest coherent elements of each sentence of your text as derived from it's conjunctions, prepositions and selected punctuation symbols (i.e. commas, semicolons, round and square brackets).

The combined approaches ensure easier, faster, more effective proofreading.

Comments and Caveats:

- The sentence parsing is achieved using a prototype natural language processing pipeline written in Python and may include occasional errors in sentence segmentation.
- Depending on the source of the input text, the Sentence Audit may contain occasional html artefacts that are parsed as sentences (E.g. "Download figure. Open in new tab").
- Always consult the original research paper as the true reference source for the text.

Contact Information:

To get a Manuscript Microscope Sentence Audit of any other research paper, simply forward any copy of the text to John.James@OxfordResearchServices.com.

All queries, feedback or suggestions are also very welcome.

Research Paper Sections:

The sections of the research paper input text parsed in this audit.

[illegible]

Title **Species Tree Inference on Data with Paralogs is Accurate Using Methods Intended to Deal with Incomplete Lineage Sorting**

S1 [001] Abstract

S1 [002] Many recent phylogenetic methods have focused on accurately inferring species trees when there is gene tree discordance due to incomplete lineage sorting (ILS).

Many recent phylogenetic methods have focused ...
... on accurately inferring species trees ...
... when there is gene tree discordance ...
... due to incomplete lineage sorting ...
... (ILS).

S1 [003] For almost all of these methods, and for phylogenetic methods in general, the data for each locus is assumed to consist of orthologous, single-copy sequences.

For almost all of these methods, ...
... and for phylogenetic methods ...
... in general, ...
... the data ...
... for each locus is assumed ...
... to consist ...
... of orthologous, ...
... single-copy sequences.

S1 [004] Loci that are present in more than a single copy in any of the studied genomes are excluded from the data.

Loci ...
... that are present ...
... in more than a single copy ...
... in any ...
... of the studied genomes are excluded ...
... from the data.

S1 [005] These steps greatly reduce the number of loci available for analysis.

These steps greatly reduce the number ...
... of loci available ...
... for analysis.

S1 [006] The question we seek to answer in this study is: What happens if one runs such species tree inference methods on data where paralogy is present, in addition to or without ILS being present?

The question we seek ...
... to answer ...

... in this study is: ...
... What happens ...
... if one runs ...
... such species tree inference methods ...
... on data ...
... where paralogy is present, ...
... in addition ...
... to or without ILS being present?

S1 [007] Through simulation studies and analyses of two large biological data sets, we show that running such methods on data with paralogs can still provide accurate results.

Through simulation studies ...
... and analyses ...
... of two large biological data sets, ...
... we show ...
... that running ...
... such methods ...
... on data ...
... with paralogs can still provide accurate results.

S1 [008] We use multiple different methods, some of which are based directly on the multispecies coalescent (MSC) model, and some of which have been proven to be statistically consistent under it.

We use multiple different methods, ...
... some of which are based directly ...
... on the multispecies coalescent ...
... (MSC) ...
... model, ...
... and some of which have been proven ...
... to be statistically consistent ...
... under it.

S1 [009] We also treat the paralogous loci in multiple ways: from explicitly denoting them as paralogs, to randomly selecting one copy per species.

We also treat the paralogous loci ...
... in multiple ways: ...
... from explicitly denoting them ...
... as paralogs, ...
... to randomly selecting one copy ...
... per species.

S1 [010] In all cases the inferred species trees are as accurate as equivalent analyses using single-copy orthologs.

In all cases the inferred species trees are ...
... as accurate ...
... as equivalent analyses ...
... using single-copy orthologs.

- S1 [011]** Our results have significant implications for the use of ILS-aware phylogenomic analyses, demonstrating that they do not have to be restricted to single-copy loci.
- Our results have significant implications ...
 - ... for the use ...
 - ... of ILS-aware phylogenomic analyses, ...
 - ... demonstrating ...
 - ... that they do not have ...
 - ... to be restricted ...
 - ... to single-copy loci.
- S1 [012]** This will greatly increase the amount of data that can be used for phylogenetic inference.
- This will greatly increase the amount ...
 - ... of data ...
 - ... that can be used ...
 - ... for phylogenetic inference.
- S1 [013]** Species tree inference often requires us to account for the fact that the evolutionary histories of different loci can disagree with each other, as well as with the phylogeny of the species.
- Species tree inference often requires us ...
 - ... to account ...
 - ... for the fact ...
 - ... that the evolutionary histories ...
 - ... of different loci can disagree ...
 - ... with each other, ...
 - ... as well ...
 - ... as with the phylogeny ...
 - ... of the species.
- S1 [014]** The reasons for this incongruence include biological causes such as incomplete lineage sorting (ILS) and introgression (broadly interpreted to include all biological processes involving genetic exchange), as well as technical causes such as the misidentification of paralogs as orthologs ("hidden paralogy"; Doolittle and Brown, 1994).
- The reasons ...
 - ... for this incongruence include biological causes ...
 - ... such as incomplete lineage sorting ...
 - ... (ILS) ...
 - ... and introgression ...
 - ... (broadly interpreted ...
 - ... to include all biological processes ...
 - ... involving genetic exchange), ...
 - ... as well ...
 - ... as technical causes ...
 - ... such as the misidentification ...
 - ... of paralogs ...
 - ... as orthologs ...
 - ... ("hidden paralogy"; ...
 - ... Doolittle ...
 - ... and Brown, 1994).

S1 [015] The inference of phylogenies can be carried out by concatenating all loci together or by treating each locus separately (reviewed in Bryant and Hahn, 2020).

The inference ...
... of phylogenies can be carried out ...
... by concatenating all loci together ...
... or by treating each locus separately ...
... (reviewed ...
... in Bryant ...
... and Hahn, 2020).

S1 [016] While concatenation ignores incongruence, gene tree-based methods allow each locus to take on its own topology.

While concatenation ignores incongruence, ...
... gene tree-based methods allow each locus ...
... to take ...
... on its own topology.

S1 [017] Some gene tree-based methods rely on a model for how these trees evolve within the species phylogeny (in addition to probabilistic models of sequence evolution on the gene trees).

Some gene tree-based methods rely ...
... on a model ...
... for how these trees evolve ...
... within the species phylogeny ...
... (in addition ...
... to probabilistic models ...
... of sequence evolution ...
... on the gene trees).

S1 [018] The multispecies coalescent (MSC) (Hudson, 1983; Takahata, 1989; Rannala and Yang, 2003; Degnan and Rosenberg, 2009) has emerged as the most commonly employed model of such gene genealogies.

The multispecies coalescent ...
... (MSC) ...
... (Hudson, 1983; ...
... Takahata, 1989; ...
... Rannala ...
... and Yang, 2003; ...
... Degnan ...
... and Rosenberg, 2009) ...
... has emerged ...
... as the most commonly employed model ...
... of such gene genealogies.

S1 [019] Indeed, in the last two decades a wide array of methods and computer programs have been developed for species tree inference under the MSC; see (Liu et al., 2009; Knowles and Kubatko, 2011; Nakhleh, 2013; Liu et al., 2015) for recent reviews and surveys of these methods.

Indeed, ...

End of Sample Audit

This is a truncated Manuscript Microscope Sample Audit.

To get the full audit of this text (or any other research paper),
forward a copy of the research paper to John James at
John.James@OxfordResearchServices.com
