A Manuscript Microscope 'Sentence Audit' of:

Machine-learning predicts genomic determinants of meiosis-driven structural variation in a eukaryotic pathogen

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/10.1101/2020.10.23.352468v2

Manuscript Authors: Thomas Badet, Simone Fouché, Fanny E. Hartmann, Marcello Zala & Daniel Croll

Audit Date: 31/03/21 **Audit Identifier:** G13CCW6XF7DTL4T **Code Version:** 3.6

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.

Section No.	Headings	Sentences
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0
N/A	--- --- --- --- --- --- ---	0

The Sentence Audit Of The Research Paper

Title **Machine-learning predicts genomic determinants of meiosis-driven structural variation in a eukaryotic pathogen**

S0 [001] Abstract

Abstract

S0 [002] Species harbor extensive structural variation underpinning recent adaptive evolution and major disease phenotypes.

Species harbor extensive structural variation underpinning recent adaptive evolution ...

... and major disease phenotypes.

S0 [003] Most sequence rearrangements are generated non-randomly along the genome through non-allelic recombination and transposable element activity.

Most sequence rearrangements are generated non-randomly ...

... along the genome ...

... through non-allelic recombination ...

... and transposable element activity.

S0 [004] However, the causality between genomic features and the induction of new rearrangements is poorly established.

However, ...

... the causality ...

... between genomic features ...

... and the induction ...

... of new rearrangements is poorly established.

S0 [005] Here, we analyze a global set of telomere-to-telomere genome assemblies of a major fungal pathogen of wheat to establish a nucleotide-level map of structural variation.

Here, ...

... we analyze a global set ...

... of telomere-to-telomere genome assemblies ...

... of a major fungal pathogen ...

... of wheat ...

... to establish a nucleotide-level map ...

... of structural variation.

S0 [006] We show that the recent emergence of pesticide resistance has been disproportionally driven by rearrangements.

We show ...

... that the recent emergence ...

... of pesticide resistance has been disproportionally driven ...

... by rearrangements.

S0 [007] We used machine-learning to train a model on structural variation events based on 30 chromosomal sequence features.

We used machine-learning ...
... to train a model ...
... on structural variation events based ...
... on 30 chromosomal sequence features.

S0 [008] We show that base composition and gene density are the major determinants of structural variation.

We show ...
... that base composition ...
... and gene density are the major determinants ...
... of structural variation.

S0 [009] Low-copy LINE and Gypsy retrotransposons explain most inversion, indel and duplication events.

Low-copy LINE ...
... and Gypsy retrotransposons explain most inversion, ...
... indel ...
... and duplication events.

S0 [010] We retrain our model on Arabidopsis thaliana and show that our modelling approach can be extended to more complex genomes.

We retrain our model ...
... on Arabidopsis thaliana ...
... and show ...
... that our modelling approach can be extended ...
... to more complex genomes.

S0 [011] Finally, we analyzed complete genomes of haploid offspring in a four-generation pedigree.

Finally, ...
... we analyzed complete genomes ...
... of haploid offspring ...
... in a four-generation pedigree.

S0 [012] Meiotic crossover locations were enriched for newly generated structural variation consistent with crossovers being mutational hotspots.

Meiotic crossover locations were enriched ...
... for newly generated structural variation consistent ...
... with crossovers being mutational hotspots.

S0 [013] The model trained on species-wide structural variation predicted the position of >74% of the newly generated variants along the pedigree.

The model trained ...
... on species-wide structural variation predicted the position ...
... of >74% ...
... of the newly generated variants ...

... along the pedigree.

S0 [014] The predictive power highlights causality between specific sequence features and the induction of chromosomal rearrangements.

The predictive power highlights causality ...
... between specific sequence features ...
... and the induction ...
... of chromosomal rearrangements.

S0 [015] Our work demonstrates that training sequence-derived models can accurately identify regions of intrinsic DNA instability in eukaryotic genomes.

Our work demonstrates ...
... that training sequence-derived models can accurately identify regions ...
... of intrinsic DNA instability ...
... in eukaryotic genomes.

S0 [016] INTRODUCTION

INTRODUCTION

S0 [017] Structural variation including duplications, inversions, insertions and deletions is often the most common form of genetic variation within species 1–3.

Structural variation including duplications, ...
... inversions, ...
... insertions ...
... and deletions is often the most common form ...
... of genetic variation ...
... within species 1–3.

S0 [018] A wide range of phenotypes including human diseases, traits for crop improvements and pathogen virulence are associated with chromosomal rearrangements 4–7.

A wide range ...
... of phenotypes including human diseases, ...
... traits ...
... for crop improvements ...
... and pathogen virulence are associated ...
... with chromosomal rearrangements 4–7.

S0 [019] Breaks in genome collinearity can have important consequences for the faithful exchange of alleles during meiosis 8–10.

Breaks ...
... in genome collinearity can have important consequences ...
... for the faithful exchange ...
... of alleles ...
... during meiosis 8–10.

S0 [020] Hence, chromosomal rearrangements can also underpin adaptation stemming from co-adapted alleles including sex determination loci 11, 12.

Hence, ...

... chromosomal rearrangements can also underpin adaptation stemming ...

... from co-adapted alleles including sex determination loci 11, ...

... 12.

S0 [021] Structural variation arises either from non-allelic homologous recombination, homology-directed repair and microhomology-mediated end joining between repetitive regions in the genome or the activity of transposable elements 13, 14.

Structural variation arises either ...

... from non-allelic homologous recombination, ...

... homology-directed repair ...

... and microhomology-mediated end joining ...

... between repetitive regions ...

... in the genome ...

... or the activity ...

... of transposable elements 13, ...

... 14.

S0 [022] However, predicting the likelihood of sequence rearrangement events remains challenging.

However, ...

... predicting the likelihood ...

... of sequence rearrangement events remains challenging.

S0 [023] In most species, the repetitive sequence content is poorly resolved at the nucleotide-level leading to fuzzy predictions about fragile sites that are prone to rearrangements 15.

In most species, ...

... the repetitive sequence content is poorly resolved ...

... at the nucleotide-level leading ...

... to fuzzy predictions ...

... about fragile sites ...

... that are prone ...

... to rearrangements 15.

S0 [024] Furthermore, analyzing segregating structural variants at the population level only partially reflects sequence rearrangement events because selection tends to remove deleterious variants over generations.

Furthermore, ...

... analyzing segregating structural variants ...

... at the population level ...

... only partially reflects sequence rearrangement events ...

... because selection tends ...

... to remove deleterious variants ...

... over generations.

S0 [025] Most sequence rearrangements are generated through double-strand break repair 16.

Most sequence rearrangements are generated ...

... through double-strand break repair 16.

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
