

# Constructing phylogenetic relationship based on the independent selection law of genome sequences

---

## 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/10.1101/2021.03.20.436244v1>

**Manuscript Authors:** Li Xiaolong, Li Hong, Yang Zhenhua & Zhang Zefeng

### 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](mailto: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      **Constructing phylogenetic relationship based on the independent selection law of genome sequences**

### **S1 [001]      ABSTRACT**

**S1 [002]**      Exploring the composition and evolution regularity of genome sequences and constructing phylogenetic relationship by alignment-free method in genome level are high-profile topics.

Exploring the composition ...  
... and evolution regularity ...  
... of genome sequences ...  
... and constructing phylogenetic relationship ...  
... by alignment-free method ...  
... in genome level are high-profile topics.

**S1 [003]**      Our previous researches discovered the CG and TA independent selection law s existed in genome sequences by analysis on the spectral features of 8-mer subsets of 920 eukaryote and prokaryote genomes.

Our previous researches discovered the CG ...  
... and TA independent selection law s existed ...  
... in genome sequences ...  
... by analysis ...  
... on the spectral features ...  
... of 8-mer subsets ...  
... of 920 eukaryote ...  
... and prokaryote genomes.

**S1 [004]**      We found that the evolution state of genomes is determined by the intensity of the two independent selections and the degree of the mutual inhibition between them.

We found ...  
... that the evolution state ...  
... of genomes is determined ...  
... by the intensity ...  
... of the two independent selections ...  
... and the degree ...  
... of the mutual inhibition ...  
... between them.

**S1 [005]**      In this study, the two independent selection patterns of 22 primate and 28 insect genome sequences were analyzed further.

In this study, ...  
... the two independent selection patterns ...  
... of 22 primate ...  
... and 28 insect genome sequences were analyzed further.

- S1 [006]** The two complete 8-mer motif sets containing CG or TA dinucleotide and their feature of relative frequency are proposed.
- The two complete 8-mer motif sets containing CG ...  
... or TA dinucleotide ...  
... and their feature ...  
... of relative frequency are proposed.
- S1 [007]** We found that the two 8-mer sets and their feature are related directly to sequence evolution of genomes.
- We found ...  
... that the two 8-mer sets ...  
... and their feature are related directly ...  
... to sequence evolution ...  
... of genomes.
- S1 [008]** According to the relative frequency of two 8-mer sets, phylogenetic trees were constructed respectively for the given primate and insect genomes.
- According ...  
... to the relative frequency ...  
... of two 8-mer sets, ...  
... phylogenetic trees were constructed respectively ...  
... for the given primate ...  
... and insect genomes.
- S1 [009]** Through analysis and comparison, we found that our phylogenetic trees are more consistent with the known conclusions.
- Through analysis ...  
... and comparison, ...  
... we found ...  
... that our phylogenetic trees are more consistent ...  
... with the known conclusions.
- S1 [010]** The two kinds of phylogenetic relationships constructed by CG 8-mer set and TA 8-mer set are similar in insect genomes, but the phylogenetic relationship constructed by CG 8-mer set reflect the evolution state of genomes in current age and phylogenetic relationship constructed by TA 8-mer set reflect the evolution state of genomes in a slight earlier period.
- The two kinds ...  
... of phylogenetic relationships constructed ...  
... by CG 8-mer set ...  
... and TA 8-mer set are similar ...  
... in insect genomes, ...  
... but the phylogenetic relationship constructed ...  
... by CG 8-mer set reflect the evolution state ...  
... of genomes ...  
... in current age ...  
... and phylogenetic relationship constructed ...  
... by TA 8-mer set reflect the evolution state ...  
... of genomes ...  
... in a slight earlier period.

**S1 [011]** We thought it is the result that the TA independent selection is repressed by the CG independent selection in the process of genome evolution.

We thought it is the result ...  
... that the TA independent selection is repressed ...  
... by the CG independent selection ...  
... in the process ...  
... of genome evolution.

**S1 [012]** Our study provides a theoretical approach to construct more objective evolution relationships in genome level.

Our study provides a theoretical approach ...  
... to construct more objective evolution relationships ...  
... in genome level.

## **S2 [013] INTRODUCTION**

**S2 [014]** With the explosive growth of the complete genome sequencing data of species, it has become a study focus to discuss the evolutionary relationship of species in genome levels.

With the explosive growth ...  
... of the complete genome sequencing data ...  
... of species, ...  
... it has become a study focus ...  
... to discuss the evolutionary relationship ...  
... of species ...  
... in genome levels.

**S2 [015]** The genome sequence contains all of the information about the sequence composition and evolution.

The genome sequence contains all of the information ...  
... about the sequence composition ...  
... and evolution.

**S2 [016]** The occurrence frequency of k-mers in genome sequence is considered to be an ideal feature set to reveal genome information.

The occurrence frequency ...  
... of k-mers ...  
... in genome sequence is considered ...  
... to be an ideal feature set ...  
... to reveal genome information.

**S2 [017]** It has been found that the frequency of k -mers in DNA sequences is non-random.

It has been found ...  
... that the frequency ...

... of k -mers ...  
... in DNA sequences is non-random.

**S2 [018]** Based on this feature, functional fragments and functional regions in DNA sequences were analyzed and predicted.

Based ...  
... on this feature, ...  
... functional fragments ...  
... and functional regions ...  
... in DNA sequences were analyzed ...  
... and predicted.

**S2 [019]** Chan and Kibler used 6-mers to predict cis-regulatory motifs (Chan et al. 2005).

Chan ...  
... and Kibler used 6-mers ...  
... to predict cis-regulatory motifs ...  
... (Chan et al. 2005).

**S2 [020]** Li and Lin used the k-mers ( $2 \leq k \leq 9$ ) to predict the promoter region (Li et al. 2006; Lin et al. 2011).

Li ...  
... and Lin used the k-mers ...  
... ( $2 \leq k \leq 9$ ) ...  
... to predict the promoter region ...  
... (Li et al. 2006; ...  
... Lin et al. 2011).

**S2 [021]** Zhang et al. used k-mers to predict the binding site (Zhang et al. 2011) of locust DNA.

Zhang et al. used k-mers ...  
... to predict the binding site ...  
... (Zhang et al. 2011) ...  
... of locust DNA.

**S2 [022]** Hariharan et al. found that different k -mers were closely related to the diversity of functional fragments (Hariharan et al. 2013).

Hariharan et al. found ...  
... that different k -mers were closely related ...  
... to the diversity ...  
... of functional fragments ...  
... (Hariharan et al. 2013).

**S2 [023]** Guo et al. used k-mers to study the nucleosome localization (Guo et al. 2014).

Guo et al. used k-mers ...  
... to study the nucleosome localization ...  
... (Guo et al. 2014).

## **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](mailto:John.James@OxfordResearchServices.com)

---