Directional Preference Alignment Enhances Robustness in Code Generation Benchmarks

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: Does the adoption of directional preference alignment improve robustness against diverse user preference shifts in code generation benchmarks without degrading model efficiency. Methods for detecting nucleotide substitution rates that are faster or slower than expected under neutral drift are widely used to identify candidate functional elements in genomic sequences. However, most existing methods consider either reductions (conservation) or increases. 10 claims were extracted from source literature; 8 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 7.4/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Detection of nonneutral substitution rates on mammalian phylogenies. Research question: Does the adoption of directional preference alignment improve robustness against diverse user preference shifts in code generation benchmarks without degrading model efficiency?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.4/10.

3 Results

14 papers retrieved. 10 claims extracted; 8 independently verified. Quality review score: 7.4/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
Most existing methods for detecting nucleotide substitution rates consider either reductions (conservation) or increases	✓	0.32
Most existing methods assume that selection acts uniformly across the branches of a phylogeny.	✓	0.21
The study considers four statistical, phylogenetic tests: a likelihood ratio test, a score test, a test based on exact d	✓	0.35
All four tests described have been implemented in a program called phyloP.	×	0.12
The program phyloP is freely available.	×	0.11
Based on extensive simulation experiments, the four tests (likelihood ratio, score, exact distribution, GERP) are remark	✓	0.22
With 36 mammalian species, the tests are capable of detecting strong selection at individual nucleotides with fairly goo	✓	0.28
With 36 mammalian species, the tests are capable of detecting moderate selection in 3-bp elements with fairly good sensi	✓	0.28
With 36 mammalian species, the tests are capable of detecting weaker or clade-specific selection in longer elements with	✓	0.32
The study applied phyloP to mammalian multiple alignments from the ENCODE project.	✓	0.17

References

- `https://doi.org/10.1109/tcomm.2020.2973976`
- `https://doi.org/10.1101/gr.097857.109`

- `https://doi.org/10.1007/s11704-026-60308-3`