Comprehensive analysis of repetitive extragenic palindrome sequences identified in bacteria and archaea using a new webbased tool, RepRanger

Oleg N. Murashko,Connor Morgan-Lang, Chen-Hsin Albert Yu,Hsin-Nan Lin, Anna Chao Kaberdina,Shin-Yu Kung,Vladimir R.Kaberdin, Sue Lin-Chao

Abstract

This study performed a comprehensive analysis of repetitive extragenic palindromic sequences (REPs) in bacteria and archaea and developed a web-based platform, RepRanger, for rapid identification and annotation. Using this tool, more than 4,000 REPs were identified in the *E. coli* MG1655 genome, with ~52% of small noncoding RNAs (sRNAs) containing REPs. Ten consensus motifs were characterized, showing potential roles in translational regulation and environmental adaptation. Comparative analysis revealed that REP motifs are more similar in pathogenic and environmental strains than in commensal or laboratory strains, and that REPs are widely distributed across bacteria and archaea.

Winpact Model

Introduction

REP elements were first discovered in *E. coli* and **Salmonella** and later identified in many bacterial species. These short sequences form secondary structures that influence gene expression and RNA stability. However, their conservation, biological functions, and evolutionary roles remain unclear. As previous search tools are no longer available, this study established the RepRanger platform to enable systematic exploration of REP distribution and function across species.

Materials and Methods

The Winpact Parallel Fermentation System FS-05-220 (Major Science Inc., USA) was used to culture *E. coli* MG1655 in M9 medium at 37°C, 200 rpm, and pH 7.0, with continuous aeration at 0.4 L/min. Cultures were harvested for RNA isolation and Northern blot analysis.

For bioinformatics, RepRanger (PHP/Python) processed genome sequences (FASTA) and annotations (GFF) to identify REPs, combined with MEME, FIMO, and TomTom for motif analysis. Novel sRNAs were identified from RNA-seq datasets, with Integrative Genomics Viewer (IGV) and ShinyGO used for functional analysis.

Results

RepRanger predicted 5,755 candidate elements in *E. coli* MG1655, of which 4,072 were REPs. These REPs displayed 10 consensus motifs, with motif 1 resembling previously reported sequences and others newly identified. About 11% of REPs were located within 15 nt downstream of stop codons, suggesting a role in translational repression. Over half of the sRNAs contained REP elements, targeting genes involved in metabolic rewiring and stress responses. REP motifs were highly conserved among pathogenic and environmental strains but differed from commensal or laboratory strains. Comparative analysis confirmed widespread presence of REPs in bacterial and archaeal genomes.

References

Comprehensive analysis of repetitive extragenic palindrome sequences identified in bacteria and archaea using a new webbased tool, RepRanger

https://doi.org/10.1128/msphere.001