BMC Evol Biol. 2018 Feb 12;18(1):21. doi: 10.1186/s12862-018-1134-0.

A novel short L-arginine responsive protein-coding gene (laoB) antiparallel overlapping to a CadC-like transcriptional regulator in Escherichia coli O157:H7 Sakai originated by overprinting.

Hücker SM[1,2], Vanderhaeghen S[1], Abellan-Schneyder I[1,3], Wecko R[1], Simon S[4], Scherer S[1,5], Neuhaus K[6,7].
Author information

Abstract

BACKGROUND:
Due to the DNA triplet code, it is possible that the sequences of two or more protein-coding genes overlap to a large degree. However, such non-trivial overlaps are usually excluded by genome annotation pipelines and, thus, only a few overlapping gene pairs have been described in bacteria. In contrast, transcriptome and translatome sequencing reveals many signals originated from the antisense strand of annotated genes, of which we analyzed an example gene pair in more detail.

RESULTS:
A small open reading frame of Escherichia coli O157:H7 strain Sakai (EHEC), designated laoB (L-arginine responsive overlapping gene), is embedded in reading frame -2 in the antisense strand of ECs5115, encoding a CadC-like transcriptional regulator. This overlapping gene shows evidence of transcription and translation in Luria-Bertani (LB) and brain-heart infusion (BHI) medium based on RNA sequencing (RNAseq) and ribosomal-footprint sequencing (RIBOseq). The transcriptional start site is 289 base pairs (bp) upstream of the start codon and transcription termination is 155 bp downstream of the stop codon. Overexpression of LaoB fused to an enhanced green fluorescent protein (EGFP) reporter was possible. The sequence upstream of the transcriptional start site displayed strong promoter activity under different conditions, whereas promoter activity was significantly decreased in the presence of L-arginine. A strand-specific translationally arrested mutant of laoB provided a significant growth advantage in competitive growth experiments in the presence of L-arginine compared to the wild type, which returned to wild type level after complementation of laoB in trans. A phylostratigraphic analysis indicated that the novel gene is restricted to the Escherichia/Shigella clade and might have originated recently by overprinting leading to the expression of part of the antisense strand of ECs5115.

CONCLUSIONS:
Here, we present evidence of a novel small protein-coding gene laoB encoded in the antisense frame -2 of the annotated gene ECs5115. Clearly, laoB is evolutionarily young and it originated in the Escherichia/Shigella clade by overprinting, a process which may cause the de novo evolution of bacterial genes like laoB.

KEYWORDS: De novo gene; EHEC; Overlapping gene; Overprinting; Small protein

PMID: 29433444
DOI: 10.1186/s12862-018-1134-0