

**UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
DEPARTAMENTO DE BIOLOGIA CELULAR E MOLECULAR E BIOAGENTES
PATOGENICOS**

JOSHELIN HUANCA JUAREZ

**Identificação e caracterização funcional de genes de origem metagenômica associados a
resistência ao estresse em bactérias**

Ribeirão Preto
2022

JOSHELIN HUANCA JUAREZ

Identificação e caracterização funcional de genes de origem metagenômica associados a resistência ao estresse em bactérias

Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do título de Mestre em Ciências.

Área de concentração: Biologia Celular e Molecular.

Orientadora: Prof. Dra. María Eugenia Guazzaroni

Ribeirão Preto
2022

RESUMO

A biotecnologia industrial desenvolveu-se rapidamente como uma plataforma de produção de bens benéficos para a sociedade, atribuíveis decorrente ao uso de fábricas de células, tendo como uma de suas principais vantagens o uso de recursos renováveis com o mínimo de impacto ao meio ambiente.. Bioprodutos industriais são menos competitivos do que produtos produzidos quimicamente devido deficiências dos bioprocessos microbianos ou enzimáticos convencionais; portanto, tornam-se necessárias tecnologias aprimoradas para engenharia e otimização de deformação. O potencial genético de bactérias que vivem em ambientes extremos, como aqueles com alta salinidade, baixo pH e outros fatores, é explorado usando técnicas metagenômicas. O desenvolvimento de microrganismos robustos e eficientes pode beneficiar-se da compreensão dos mecanismos moleculares empregados para sobreviver em condições extremas. Neste trabalho, a fim de construir circuitos de genes sintéticos que possam aumentar a resistência bacteriana a várias situações de estresse, realizou-se a identificação de novos genes em bancos de dados metagenômicos por meio de abordagem in silico. Recuperamos informações de metagenomas de ambientes hostis para identificar genes que codificam chaperonas e outras proteínas como proteases, proteínas de ligação a ácidos nucleicos, entre outras, que fornecem resistência a condições de estresse. As sequências foram identificadas por meio de perfis de Hidden Markov Model (HMM), e posteriormente as mais relevantes foram selecionadas e categorizadas com base na identidade da sequência de proteínas e revisão da literatura. Dez sequências de proteínas, incluindo a proteína de ligação ao DNA HU, a protease dependente de ATP ClpP e a proteína chaperona DnaJ foram selecionadas e caracterizadas funcionalmente em *Escherichia coli* sob seis condições de estresse: alta temperatura, acidez, estresse oxidativo, estresse osmótico, radiação UV e ácido p-cumárico. Mediante esta abordagem, foram identificadas cinco proteínas que responderam ao menos a duas condições de estresse em comparação com culturas controle. Essas descobertas demonstram que as técnicas metagenômicas e as ferramentas de bioinformática podem atuar na prospecção de novos genes que apresentam potencial de aumentar a resiliência das bactérias a situações de estresse.

ABSTRACT

Industrial biotechnology has developed rapidly as a platform for producing goods beneficial to society attributable to the use of cell factories, with the use of renewable resources with minimum impact on the environment as one of its major advantages. Industrial bioproducts are less competitive than chemically produced goods due to the shortcomings of conventional microbial or enzymatic bioprocesses; therefore, improved technologies for strain engineering and optimization will be needed. The genetic potential of bacteria living in severe settings, such as those with high salinity, low pH, and other factors, is exploited using metagenomic techniques. Building strong, efficient microbes can benefit from understanding the molecular mechanisms they employ to survive under these conditions. In order to build synthetic gene circuits that can increase bacterial resistance to various stress situations, we addressed the identification of new genes in metagenomic databases using an *in silico* approach. We recovered information from harsh environment metagenomes to identify genes encoding chaperones and other proteins (such as proteases, nucleic-acid-binding proteins, etc.) that provide resistance to stress conditions. The most relevant sequences were chosen using Hidden Markov Model (HMM) profiles, which were then categorized based on their protein sequence identity and literature review. Ten protein sequences including the DNA-binding protein HU, the ATP-dependent protease ClpP, and the chaperone protein DnaJ were selected and functionally characterized in *Escherichia coli* under six stress conditions: high temperature, acidity, oxidative stress, osmotic stress, UV radiation, and p-coumaric acid. Further characterization allowed us to identify five proteins that responded under at least two stress conditions compared with control cultures. These findings demonstrated that metagenomic techniques and bioinformatic tools can prospect novel genes that could potentially increase bacteria's resilience to stressful situations.