



Evento	Salão UFRGS 2015: SIC - XXVII SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2015
Local	Porto Alegre - RS
Título	Análise in silico da composição de nucleotídeos de sequências promotoras de genes de reposta ao choque térmico em Escherichia coli
Autor	GABRIEL DALL'ALBA
Orientador	SCHEILA DE AVILA E SILVA
Instituição	Universidade de Caxias do Sul

Análise *in silico* da composição de nucleotídeos de sequências promotoras de genes de
reposta ao choque térmico em *Escherichia coli*

Nome do Autor: Gabriel Dall'Alba

Nome do Orientador: Scheila de Avila e Silva

Instituição: Universidade de Caxias do Sul

O estresse desencadeado pelo choque térmico em uma bactéria provoca rápidas mudanças em alguns processos biológicos, por exemplo, o envelopamento e a fixação de proteínas. A resposta biológica a este estímulo é regulada pela ligação dos fatores sigma (σ) 24 e 32 na enzima RNA polimerase (RNAP) e o posterior reconhecimento da região promotora. A identificação dos pares de base constituintes do promotor auxiliam o estudo da regulação gênica. Deste modo, amplia-se a compreensão da capacidade e dos mecanismos de sobrevivência utilizados pelos organismos. Com base nestes aspectos, o objetivo deste trabalho foi analisar a composição de nucleotídeos das regiões promotoras reconhecidas pelos fatores σ 24 e σ 32 de *Escherichia coli*. Foram coletadas do banco de dados RegulonDB dois conjuntos de dados: (i) 521 sequências reconhecidas pelo σ 24; (ii) 324 sequências reconhecidas pelo σ 32. Posteriormente, os promotores foram codificadas em valores numéricos conforme metodologia descrita na literatura, a fim de deixá-las adequadas ao uso da técnica computacional de clusterização. Esta técnica utiliza o algoritmo K-Means, o qual requer um valor numérico K (atribuído pelo usuário da ferramenta) equivalente ao número de agrupamentos desejados. Como a metodologia formal para a obtenção do número ótimo de k é empírica, foram realizadas simulações com diferentes valores de k: 3, 6 e 8. Os agrupamentos foram analisados em relação à pureza do *cluster*, ou seja, quantas sequências de um mesmo σ foram agrupados em um mesmo *cluster*. Com a análise dos resultados obtidos, verificou-se que a simulação com k=6 apresentou agrupamentos com uma pureza média de 93%. A quantidade de sequências em cada conjunto, bem como o fator σ associado, foram, respectivamente: (i) *Cluster* 1: 125 sequências σ 32 e 3 sequências σ 24 (97% exatidão); (ii) *Cluster* 2: 138 sequências σ 24 e 1 σ 32 (99% exatidão); (iii) *Cluster* 3: 102 sequências σ 24 e 3 σ 32 (97% exatidão); (iv) *Cluster* 4: 148 sequências σ 24 (100% exatidão); (v) *Cluster* 5: 131 sequências σ 24 (100% exatidão); (vi) *Cluster* 6: 184 sequências σ 32 (67% exatidão). Após esta análise, a composição consensual dos nucleotídeos foi obtida com a ferramenta WebLogo. Assim, foi possível observar que a região canônica denominada -10 apresentou conteúdo similar ao consenso biológico previamente descrito em todos os *clusters*, com exceção de um agrupamento (*cluster* 5). Já para a região -35, não foi possível estabelecer relação com o consenso estabelecido na literatura. Estas observações são fundamentadas pela presença da região -10 estendida para os promotores do σ 32, ou seja, a informação região -35 foi realocada nesta região, a fim de proporcionar o funcionamento da RNAP. Outra característica evidente foi a prevalência dos nucleotídeos A e G na região +1 (início da transcrição). Conforme relatos prévios, a presença do nucleotídeo G na posição +1 é responsável pela pausa na transcrição, a qual trata-se de um importante mecanismo para reparo de erros de mutação ao longo da sequência da região codificante. Deste modo, a compreensão dos mecanismos de regulação gênica relacionado ao choque térmico é importante uma vez que este fenômeno altera a forma e composição do genoma. Percebe-se que um certo grau de degeneração das regiões consenso é tolerado no processo de transcrição gênica, porém evidenciam o desafio computacional. Assim, estabelecer os distintos perfis encontrados auxilia na redução do número de falsos positivos em ferramentas *in silico* relacionados à predição de promotores.