



Influência do índice de confiança, *pipeline* de clusterização e classificadores em análises de *metabarcoding*

LAURA BEZERRA COUTINHO¹, AFONSO LUIS BARTH^{2,3}

¹Graduação em Biotecnologia/Bioinformática, UFRGS, Porto Alegre

²Laboratório de Pesquisa em Resistência Bacteriana, Centro de Pesquisa Experimental, Hospital de Clínicas de Porto Alegre, Porto Alegre

³Departamento de Análises, Faculdade de Farmácia, UFRGS, Porto Alegre

Introdução

A compreensão da composição e funcionamento da microbiota presente no corpo humano é, hoje, alvo de grande interesse científico, tendo em vista a gama de interações e funções nas quais ela está envolvida. Nesse cenário, o *metabarcoding* a partir do 16S rDNA é um método amplamente aplicado, e as análises bioinformáticas dos dados gerados enfrentam o grande desafio da garantia de qualidade e reprodutibilidade dos resultados. *Pipelines* de análise padronizadas são uma alternativa para essa questão, no entanto, é importante utilizar parâmetros adequados para garantir a qualidade e reprodutibilidade dos resultados obtidos.

Objetivos

O principal objetivo deste trabalho foi, portanto, elucidar a influência de parâmetros de análise - índices de confiança, classificadores e *pipelines* de clusterização de sequências - nos resultados de análises de dados de 16S rDNA referentes à microbiota intestinal humana, utilizando como referência o *pipeline* do BMP - *Brazilian Microbiome Project*.

Métodos

Foram utilizados os dados de sequenciamento da plataforma Ion Torrent PGM disponíveis no NCBI SRA Database sob o número de acesso SRX1044553. O diagrama a seguir representa todas as combinações de parâmetros testadas.

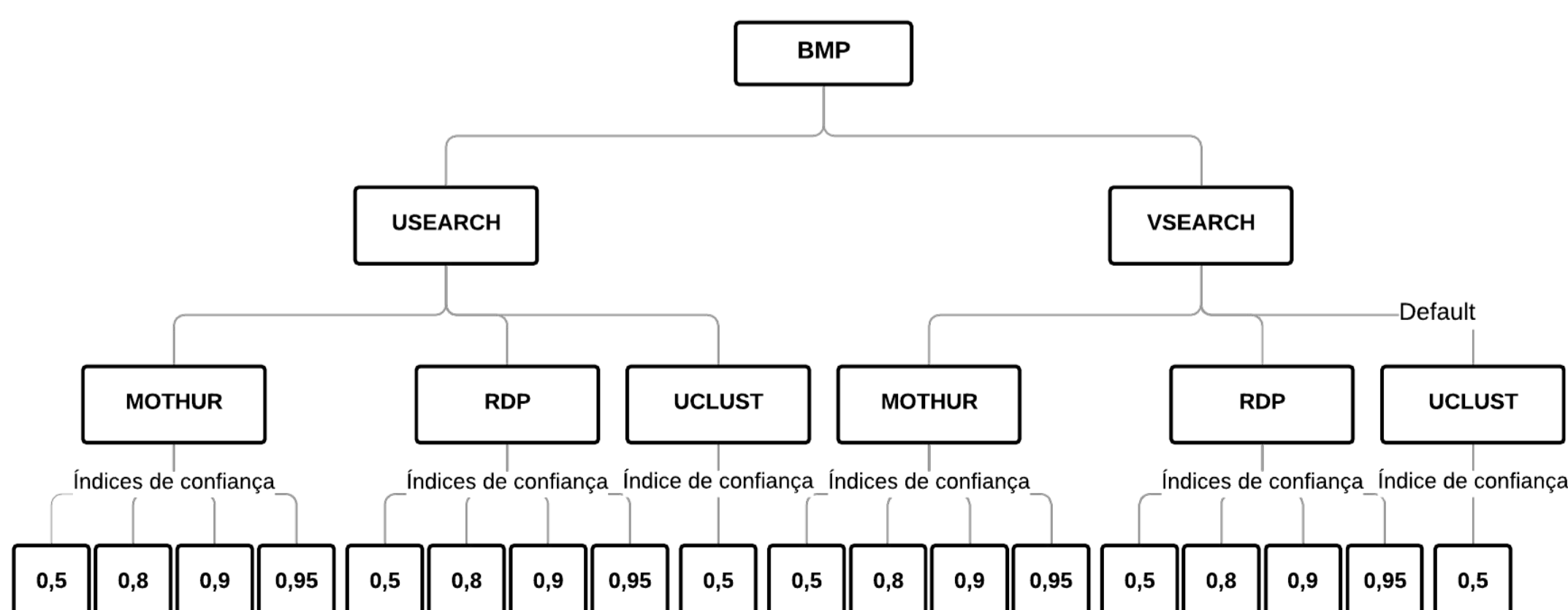


Figura 1 - Combinações de índices de confiança, classificadores e *pipelines* de clusterização de sequências testados.

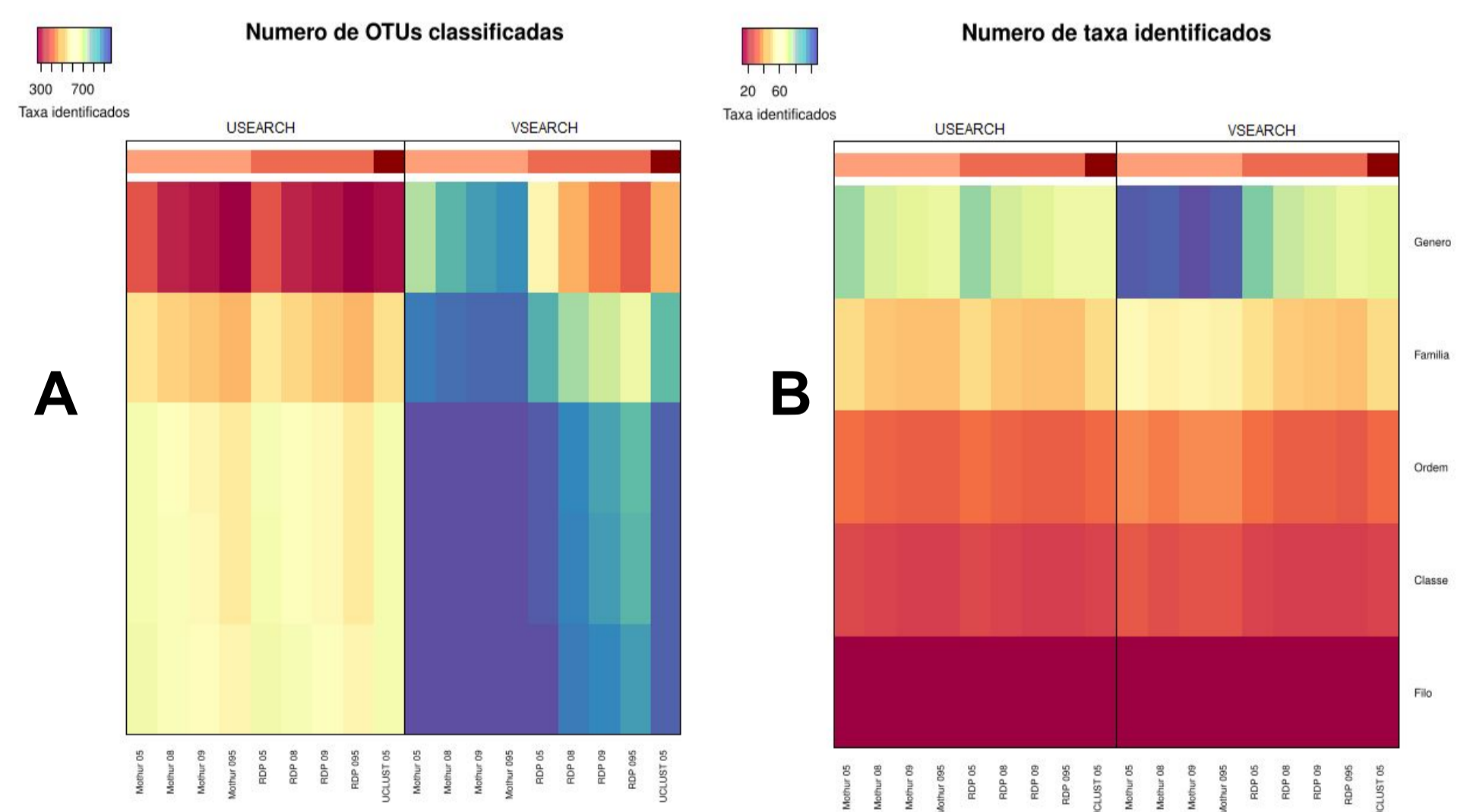
Apoio Financeiro

CAPES/HCPA e CNPq

Resultados

Nossos resultados indicam que há diferenças na aplicação de variados parâmetros ao *pipeline*, gerando diferentes efeitos na quantidade de taxa identificados, de OTUs classificadas e na precisão da classificação.

Resultados de combinações de classificadores e índices de confiança apresentam variações entre os dois *pipelines* de clusterização, no USEARCH havendo pouca variação com a alteração dos classificadores e no VSEARCH apresentando maiores disparidades - com destaque para o Mothur, cujos resultados de número de taxa identificados e OTUs classificadas foram acima dos demais, não respondendo inclusive ao aumento do índice de confiança. Destaca-se ainda o papel da remoção de sequências quiméricas nos resultados, o que é observável pelo número de OTUs detectadas em comparação entre USEARCH - 653 - e VSEARCH - 968.



Figuras 2A e 2B - Heatmaps representando, respectivamente, o número de OTUs classificadas e número de taxa identificados por cada combinação de parâmetros em cada nível taxonômico.

Conclusões

Os resultados apresentados salientam a importância da inclusão em estudos de microbiota de detalhes dos parâmetros e métodos aplicados, garantindo a validação, qualidade e reprodutibilidade do resultado. Compreender os *pipelines* aplicados e os efeitos de seus parâmetros, testar diferentes *pipelines* e variações dos parâmetros são práticas essenciais e recomendáveis.