

ASSINATURA MOLECULAR NA DIFERENCIAÇÃO DA TONSILITE VIRAL E BACTERIANA

Introdução: A tonsilite aguda, de origem viral ou bacteriana, representa um desafio diagnóstico. A distinção inadequada entre as etiologias pode levar ao uso excessivo de antibióticos e à resistência antimicrobiana. Estratégias que integrem a caracterização microbiana e a resposta imunológica local podem aprimorar o diagnóstico diferencial.

Objetivo: Identificar um painel de biomarcadores *in situ* capazes de diferenciar tonsilites bacterianas e virais. **Metodologia:** Estudo transversal com 197 pacientes atendidos no Hospital Santa Izabel (Salvador, BA) com suspeita clínica de tonsilite aguda, além de 5 controles saudáveis. A inclusão nos grupos baseou-se no Escore de Centor modificado (bacteriana) e na presença de sintomas respiratórios altos com sinais tonsilares (viral). A classificação clínica foi realizada por emergencistas e especialistas. Excluíram-se pacientes com uso recente de antibióticos, anti-inflamatórios ou imunossupressores. Após a coleta das amostras de *swab* tonsilar, DNA e RNA foram convertidos em cDNA, preparados com painéis direcionados e sequenciados na plataforma Illumina MiSeq. As análises incluíram filtragem do hospedeiro, controle de qualidade, anotação taxonômica e normalização. A análise da metagenômica foi conduzida na plataforma CZID, com filtragem por abundância (≥ 100 rPM) e exclusão de espécies irrelevantes. Genes de resistência foram identificados com *ARIBA* e *ResFinder*. A expressão gênica foi avaliada pela tecnologia nCounter (NanoString). A análise transcriptômica incluiu expressão diferencial ($\log_2$), correção por FDR e enriquecimento funcional (GO, KEGG, Reactome) via *gProfiler*. Os dados foram filtrados por prevalência ($\geq 20\%$) e analisados por testes não paramétricos, *PLS-DA*, clusterização, correlações microrganismo-gene e curvas ROC/PR com validação cruzada 3-fold. Diversidades alfa e beta foram avaliadas com métricas ecológicas e PCoA. **Resultados:** Foram analisadas 46 amostras (31 virais, 10 bacterianas – emergencistas, 29 virais e 12 bacterianas – especialistas) e 5 controles. A análise metagenômica revelou uma comunidade microbiana complexa nos pacientes e controles saudáveis, incluindo patógenos e comensais. Não houve diferenças significativas nas diversidades alfa ou beta entre os grupos. Genes de resistência foram detectados independentemente da suspeita clínica. A análise da expressão gênica ($n=24$) distinguiu os grupos de tonsilite aguda. Nos casos bacterianos, houve regulação positiva de genes como IL21R, CD47, ITGAM e CD163, com enriquecimento da via MHC II. Já nos casos virais, genes como HLA-DRB5 e TUBA3E foram regulados positivamente, com enriquecimento da via de tráfego de conexons ($p < 0,01$). **Conclusão:** A expressão gênica *in situ* apresentou maior acurácia diagnóstica do que a simples detecção de patógenos, mostrando-se promissora na diferenciação etiológica. A metagenômica foi limitada pela complexidade inerente a essa análise e devido à diversidade da microbiota presente na cavidade oral.

Palavras-chave: 1. Tonsilite; 2. Diagnóstico diferencial; 3. Transcriptômica; 4. Metagenômica; 5. Biomarcadores; 6. Resistência a antimicrobianos.

MOLECULAR SIGNATURES IN THE DIFFERENTIATION OF VIRAL AND BACTERIAL TONSILLITIS

Introduction: Acute tonsillitis, of viral or bacterial origin, represents a diagnostic challenge. Inadequate distinction between etiologies may lead to the overuse of antibiotics and antimicrobial resistance. Strategies that integrate microbial characterization and the local immune response may improve differential diagnosis.

Objective: To identify an *in situ* biomarker panel capable of differentiating bacterial and viral tonsillitis. **Methods:** This was a cross-sectional study involving 197 patients seen at Hospital Santa Izabel (Salvador, BA) with clinical suspicion of acute tonsillitis, along with 5 healthy controls. Inclusion into groups was based on the modified Centor Score (bacterial) and the presence of upper respiratory symptoms with tonsillar signs (viral). Clinical classification was performed by emergency physicians and specialists. Patients with recent use of antibiotics, anti-inflammatory drugs, or immunosuppressants were excluded. Following collection of tonsillar swab samples, DNA and RNA were converted into cDNA, prepared using targeted panels, and sequenced on the Illumina MiSeq platform. Analyses included host filtering, quality control, taxonomic annotation, and normalization. Metagenomic analysis was conducted on the CZID platform, with abundance filtering (≥ 100 rPM) and exclusion of irrelevant species. Resistance genes were identified using ARIBA and ResFinder. Gene expression was evaluated using nCounter technology (NanoString). Transcriptomic analysis included differential expression ($\log_2$), FDR correction, and functional enrichment (GO, KEGG, Reactome) via gProfiler. Data was filtered by prevalence ($\geq 20\%$) and analyzed using non-parametric tests, PLS-DA, clustering, microorganism-gene correlations, and ROC/PR curves with 3-fold cross-validation. Alpha and beta diversities were evaluated using ecological metrics and PCoA. **Results:** A total of 46 samples were analyzed (31 viral, 10 bacterial – emergency physicians; 29 viral and 12 bacterial – specialists), along with 5 controls. Metagenomic analysis revealed a complex microbial community in both patients and healthy controls, including pathogens and commensals. There were no significant differences in alpha or beta diversity between groups. Resistance genes were detected regardless of clinical suspicion. Gene expression analysis ($n=24$) distinguished between acute tonsillitis groups. In bacterial cases, genes such as IL21R, CD47, ITGAM, and CD163 were upregulated, with enrichment of the MHC II pathway. In viral cases, genes such as HLA-DRB5 and TUBA3E were upregulated, with enrichment of the connexon trafficking pathway ($p < 0.01$). **Conclusion:** In situ gene expression demonstrated greater diagnostic accuracy than mere pathogen detection, showing promise in etiological differentiation. Metagenomics was limited by the inherent complexity of the analysis and the diversity of the oral cavity microbiota.

Keywords: 1. Tonsillitis; 2. Differential diagnosis; 3. Transcriptomics; 4. Metagenomics; 5. Biomarkers; 6. Antimicrobial resistance.

RESULTADOS

1. Seleção dos pacientes com suspeita de tonsilite infecciosa

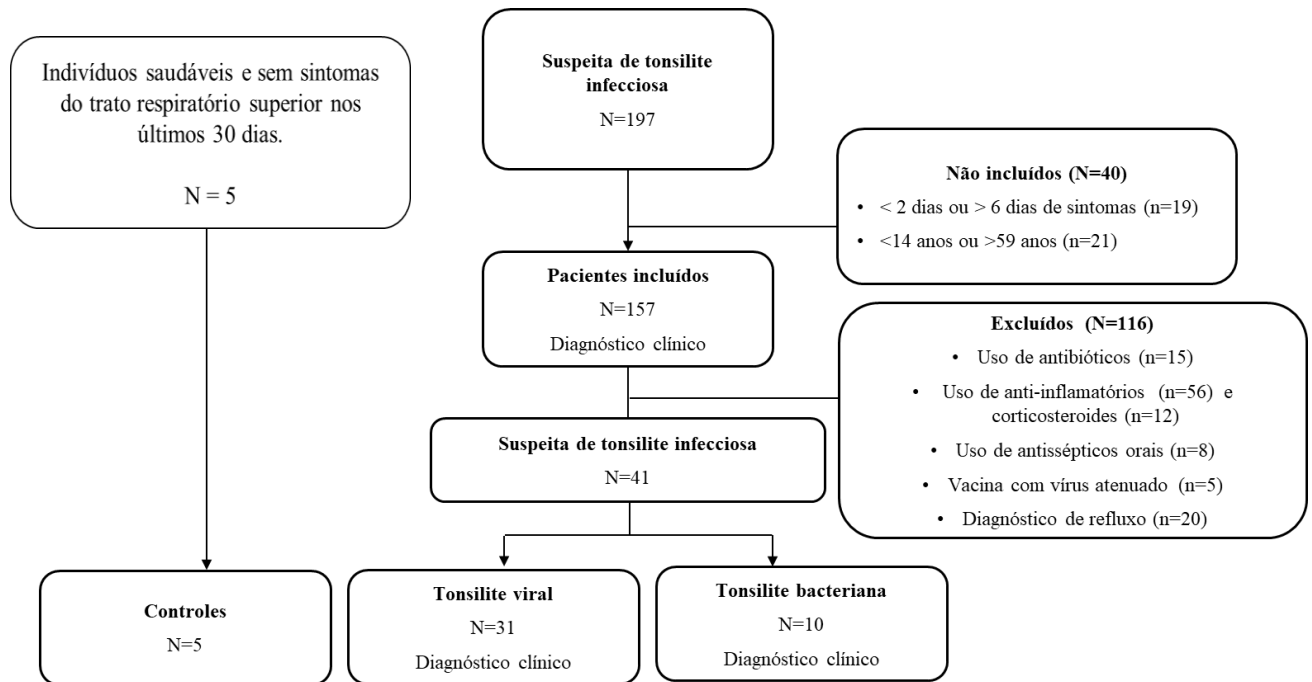


Figura 1. Fluxograma da seleção dos pacientes com suspeita clínica de tonsilite infecciosa. Pacientes foram submetidos a avaliação clínica com análise das tonsilas palatinas.

Tabela 1. Características clínicas e epidemiológicas dos 46 voluntários segundo emergencistas.

	SUSPEITA CLÍNICA		
	TONSILITE	TONSILITE	CONTROLES
	VIRAL n= 31	BACTERIANA n= 10	n=5
Sexo Feminino	24 (77)	8 (80)	4 (80)
Idade em anos - mediana (IIQ)	27 (21-35)	28 (24-42)	31 (22-45)
Sintomas – n (%)			
Febre	7 (23)	6 (60)	-
Tosse	18 (58)	3 (30)	-
Obstrução Nasal	26 (84)	0	-
Secreção Nasal			
Amarelada	10 (32)	1 (10)	-
Esbranquiçada	9 (29)	0	-
Sinais – n (%)			
Linfadenopatia cervical	11 (35)	7 (70)	-
Exsudato tonsilar			-
Amarelado	1 (17)	3 (60)	-
Esbranquiçado	5 (83)	2 (40)	-

IIQ = Intervalo Interquartil

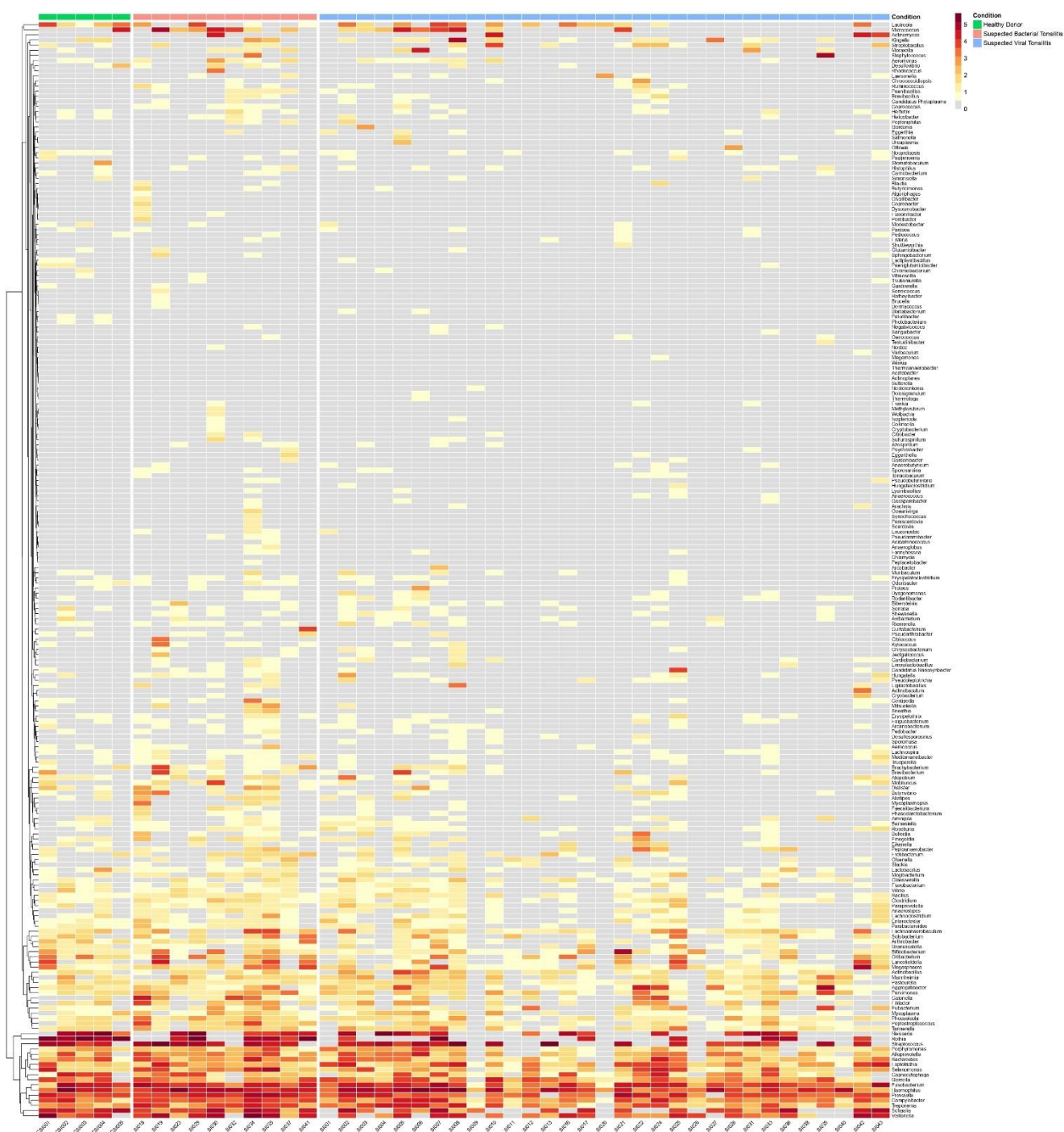


Figura 3. Heatmap das bactérias nas amostras de tonsilas de pacientes com suspeita de infecção viral, bacteriana e controles saudáveis, mediante análise por metagenômica, conforme classificação pelos emergencistas. A intensidade das cores indica a abundância relativa de cada vírus, com tons mais intensos indicando maior abundância. As linhas representam os diferentes organismos identificados, enquanto as colunas representam as amostras individuais. Os grupos de pacientes são segregados por categorias: suspeita de infecção viral, em azul; infecção bacteriana, em rosa; controles saudáveis, em verde; permitindo a comparação da composição microbiológica entre os grupos.

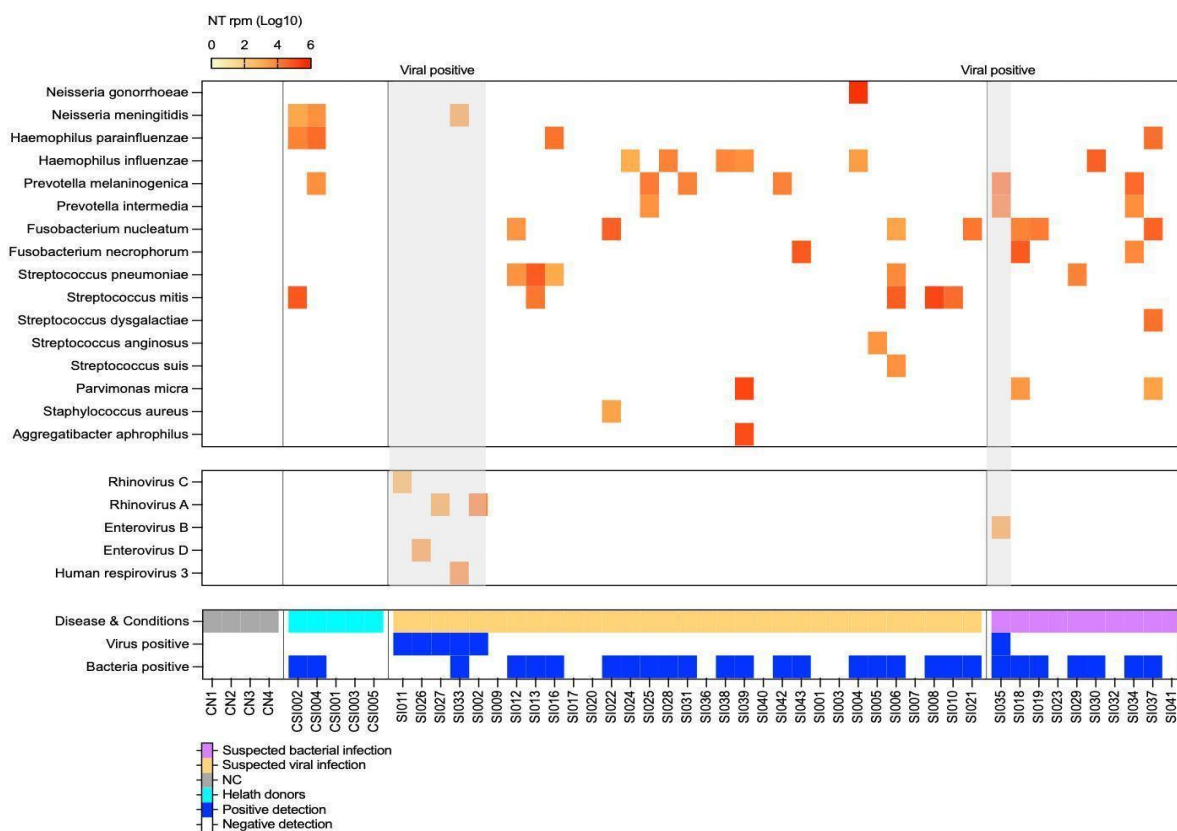


Figura 4. Heatmap do perfil de vírus e bactérias detectados nas amostras de tonsilas de pacientes com suspeita de infecção viral, bacteriana e controles saudáveis, conforme classificação pelos emergencistas. A intensidade das cores indica a abundância relativa de cada vírus ou bactéria, com tons mais intensos indicando maior abundância. As linhas representam os diferentes organismos identificados, enquanto as colunas representam as amostras individuais. Os grupos de pacientes são segregados por categorias: suspeita de infecção viral, de infecção bacteriana, controles saudáveis, além dos controles negativos, permitindo a comparação da composição microbiológica entre os grupos. "Disease and condition" = Suspeita clínica; "NC" = Controles negativos (Utilizado para verificar a ausência de contaminação e a especificidade do processo); "Healthy donors" = Controles do estudo.

Tabela 2. Sinais e sintomas clínicos da tonsilite e classificação clínica conforme especialistas e emergencistas

ID	Sinais e Sintomas clínicos	Classificação do especialista	Classificação do emergencista
SI001	Odinofagia, febre, secreção nasal esbranquiçada, obstrução nasal, tosse seca e rouquidão	Viral	Viral
SI002	Odinofagia, secreção nasal esbranquiçada, obstrução nasal, coceira, tosse com secreção esbranquiçada	Viral	Viral
SI003	Odinofagia, secreção nasal amarelada, obstrução nasal, tosse com secreção amarelada-esverdeada	Bacteriano	Viral
SI004	Odinofagia, obstrução nasal e tosse seca	Viral	Viral
SI005	Odinofagia, secreção nasal esbranquiçada, obstrução nasal, coceira, nódulo cervical elástico e doloroso, exsudato faringoamigdaliano confluyente esbranquiçado e unilateral	Viral	Viral
SI006	Odinofagia, secreção nasal esbranquiçada, obstrução nasal, rouquidão, nódulo cervical elástico indolor	Viral	Viral
SI007	Odinofagia, obstrução nasal, tosse, coceira, rouquidão, nódulo cervical não doloroso	Viral	Viral
SI008	Odinofagia, secreção nasal amarelada, obstrução nasal, coceira e tosse seca	Viral	Viral
SI009	Odinofagia e rouquidão	Viral	Viral
SI010	Odinofagia, secreção nasal amarelada, obstrução nasal e tosse com secreção amarelo-esverdeada, rouquidão, exsudato faringoamigdaliano	Bacteriano	Viral
SI011	Odinofagia, secreção nasal esbranquiçada, obstrução nasal e rouquidão	Viral	Viral
SI012	Odinofagia, coceira, tosse seca e rouquidão	Viral	Viral
SI013	Odinofagia, obstrução nasal, tosse seca, rouquidão, nódulo cervical elástico e doloroso	Viral	Viral
SI016	Odinofagia, febre, obstrução nasal, coceira, rouquidão	Viral	Viral
SI017	Odinofagia, secreção nasal amarelada, obstrução nasal, rouquidão, nódulo cervical elástico não doloroso	Viral	Viral
SI018	Odinofagia, febre, rouquidão, nódulo cervical elástico doloroso, exsudato faringoamigdaliano esbranquiçado	Viral	Bacteriano
SI019	Odinofagia, febre, secreção nasal amarelada, tosse com secreção amarelo-esverdeada, rouquidão, nódulo cervical elástico doloroso	Bacteriano	Bacteriano

SI020	Odinofagia, secreção nasal amarelada e rouquidão	Viral	Viral
SI021	Odinofagia, secreção nasal amarelada, obstrução nasal, coceira, tosse com secreção amarelo-esverdeada e rouquidão	Bacteriano	Viral
SI022	Odinofagia, secreção nasal esbranquiçada, obstrução nasal, tosse seca e rouquidão	Viral	Viral
SI023	Odinofagia, rouquidão, nódulo cervical elástico doloroso, exsudato faringoamigdaliano confluyente amarelado	Bacteriano	Bacteriano
SI024	Odinofagia, febre, secreção nasal amarelada, obstrução nasal, tosse com secreção amarelo-esverdeada, rouquidão, nódulo cervical elástico e não doloroso, exsudato amigdaliano confluyente e esbranquiçado	Bacteriano	Viral
SI025	Odinofagia, tosse com secreção amarelo-esverdeada, nódulo cervical elástico não doloroso	Bacteriano	Viral
SI026	Odinofagia, Obstrução nasal, coceira, tosse com secreção amarelo-esverdeada, rouquidão e exsudato faringoamigdaliano confluyente amarelado	Bacteriano	Viral
SI027	Odinofagia, secreção nasal esbranquiçada, obstrução nasal	Viral	Viral
SI028	Odinofagia, secreção nasal amarelada, obstrução nasal, tosse seca, rouquidão	Viral	Viral
SI029	Odinofagia	Viral	Bacteriano
SI030	Odinofagia, febre, tosse seca, nódulo cervical elástico doloroso, exsudato faringoamigdaliano esbranquiçado	Viral	Bacteriano
SI031	Odinofagia, Secreção amarelada, obstrução nasal, tosse seca, rouquidão	Viral	Viral
SI032	Odinofagia, febre, tosse amarelada esverdeada, rouquidão, nódulo cervical elástico doloroso	Bacteriano	Bacteriano
SI033	Odinofagia, secreção nasal esbranquiçada, obstrução nasal, coceira, tosse seca	Viral	Viral
SI034	Odinofagia, febre, exsudato faringoamigdaliano amarelado	Bacteriano	Bacteriano
SI035	Odinofagia, rouquidão	Viral	Bacteriano
SI036	Odinofagia, febre e obstrução nasal	Viral	Viral
SI037	Odinofagia, nódulo cervical elástico doloroso, exsudato faringoamigdaliano confluyente amarelado	Bacteriano	Bacteriano
SI038	Odinofagia, febre, obstrução nasal, tosse amarelada-esverdeada, nódulo cervical fibroelástico e doloroso.	Bacteriano	Viral
SI039	Odinofagia, Obstrução nasal, nódulo cervical elástico não doloroso, exsudato faringoamigdaliano esbranquiçado	Viral	Viral

SI040	Odinofagia, Secreção amarelada, obstrução nasal, nódulo cervical fibroelástico doloroso	Viral	Viral
SI041	Odinofagia, nódulo cervical fibroelástico doloroso	Viral	Bacteriano
SI042	Odinofagia, febre, secreção esbranquiçada, obstrução nasal, tosse seca, rouquidão, vesícula na mucosa jugal, exsudato faringoamigdaliano pontilhado esbranquiçado.	Viral	Viral
SI043	Odinofagia e febre	Viral	Viral

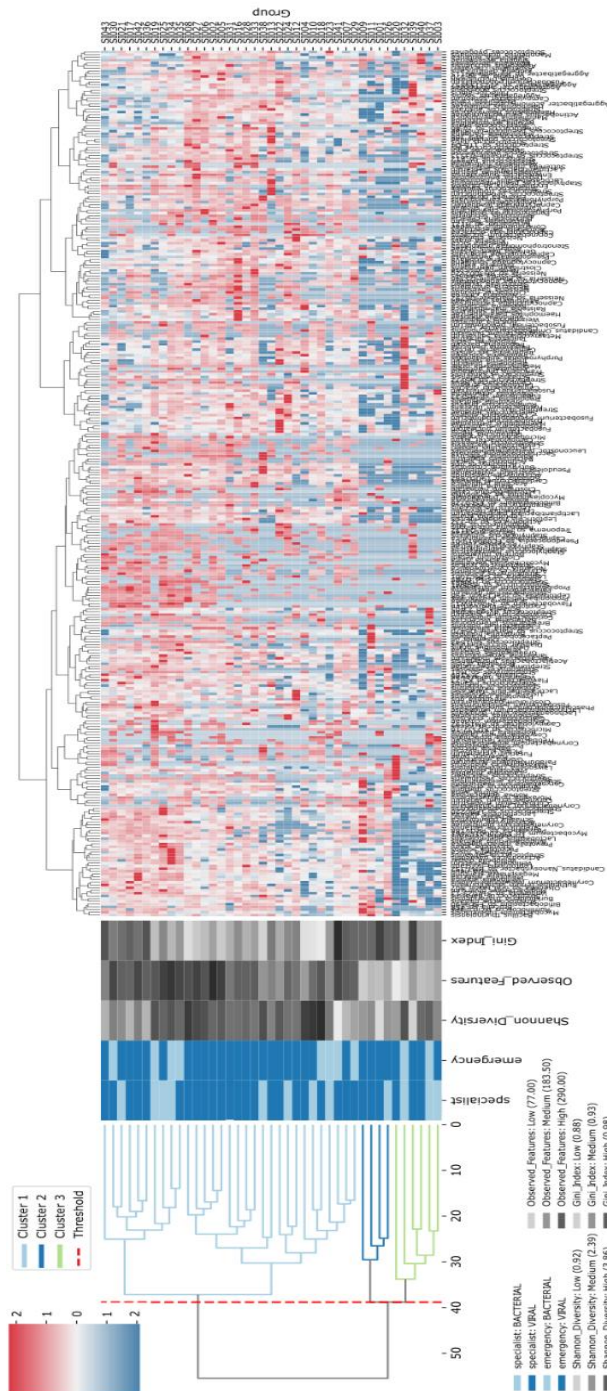


Figura 5. Heatmap da abundância relativa de espécies microbianas e variáveis de metadados associadas em diferentes amostras. O dendrograma à esquerda agrupa as amostras em três clusters distintos (Cluster 1, Cluster 2 e Cluster 3), com base na similaridade de seus perfis microbianos e características dos metadados. As barras coloridas adjacentes ao dendrograma detalham os valores de metadados para cada amostra: Specialist e emergency classificam as amostras em infecção bacteriana (azul claro) ou viral (azul escuro); Shannon_Diversity, representa a diversidade de espécies, categorizada como "Low" (0.92), "Medium" (2.39) ou "High" (3.86); Observed_Features, refere-se ao número de diferentes espécies observadas, divididas em "Low" (77.00), "Medium" (183.50) ou "High" (290.00) e Gini_Index avalia a heterogeneidade da amostra, com base nas espécies presentes, seguindo as categorias "Low" (0.88), "Medium" (0.93) ou "High" (0.98). O mapa de calor exibe a abundância relativa de diversas espécies microbianas em cada amostra. A escala de cores (vermelho para alta abundância, azul para baixa abundância) e o dendrograma superior agrupam as espécies com padrões de abundância semelhantes (o agrupamento foi definido com base no limiar marcado pela linha tracejada).

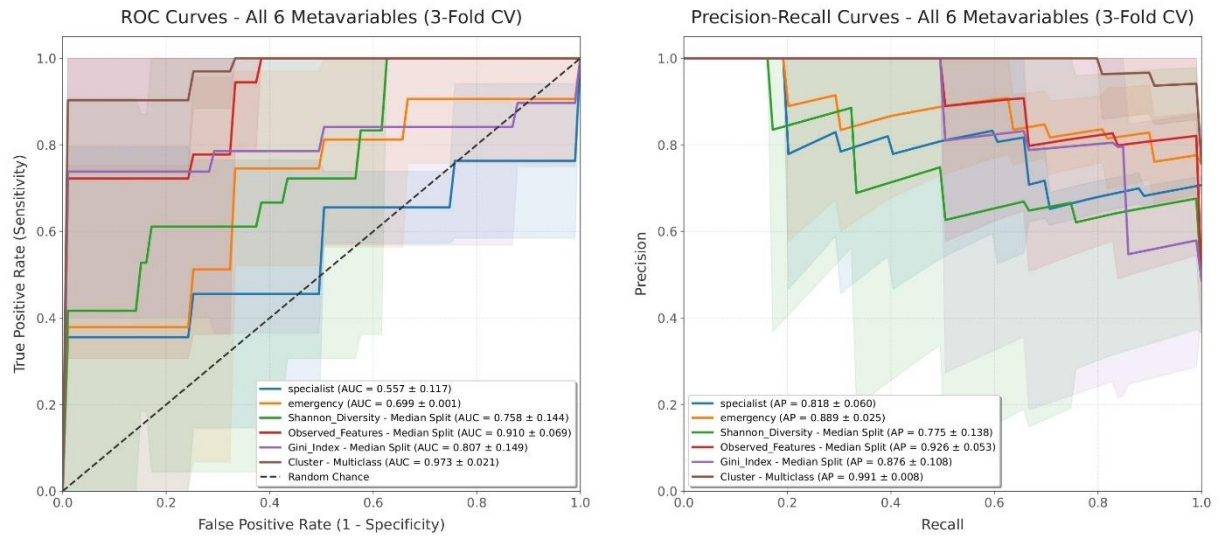


Figura 6. Curvas ROC e Precision-Recall para avaliação da capacidade preditiva das metavariáveis “specialist”, “emergency”, “Shannon diversity”, “Observed features”, “Gini index” e “Cluster-Multiclass”, utilizando validação cruzada de 3-Fold (3-Fold CV). A figura da esquerda apresenta as curvas ROC, com a Taxa de Verdadeiros Positivos (Sensibilidade) contra a Taxa de Falsos Positivos (1 - Especificidade). A linha tracejada representa a chance aleatória. A figura da direita exibe as curvas Precision-Recall, que plotam a Precisão contra o Recall. Para cada metavariável, os valores de Área Sob a Curva (AUC) e Precisão Média (AP) são apresentados na legenda como média $\pm$ desvio padrão.

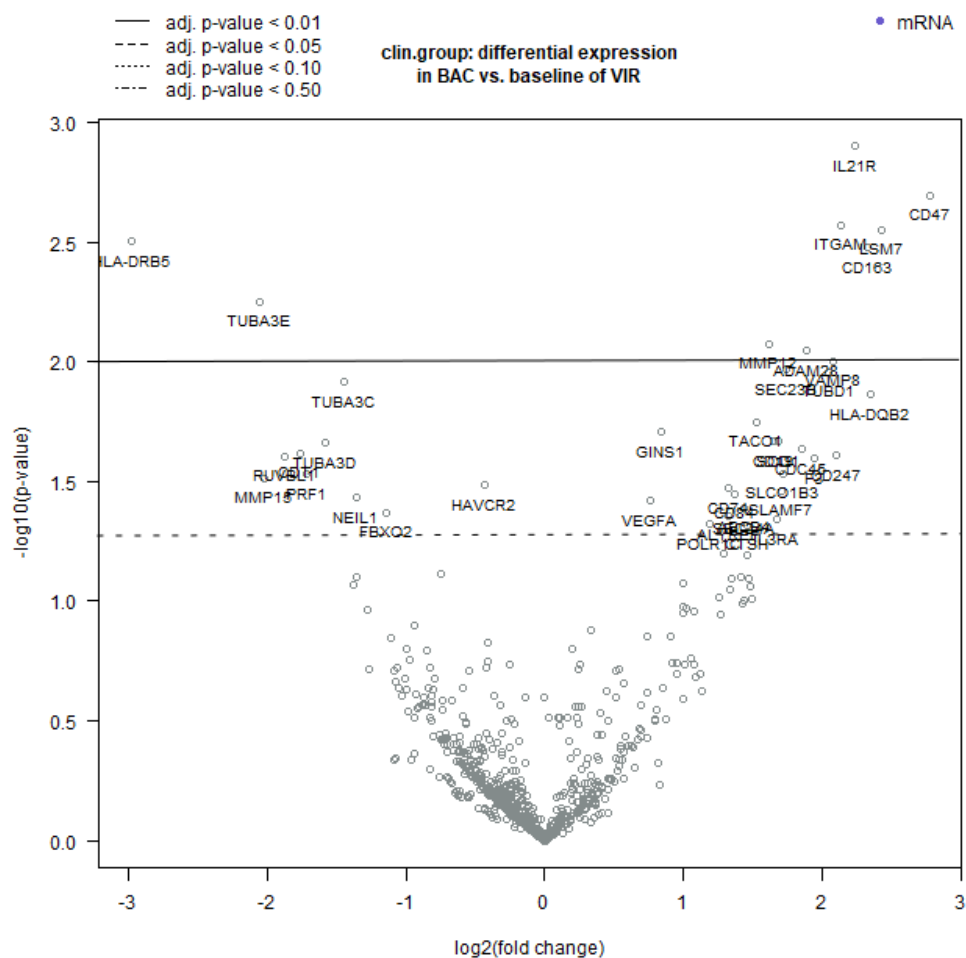


Figura 7. Volcano plot representando a expressão diferencial de genes em indivíduos com tonsilite bacteriana em comparação com os casos virais, quantificado por nCounter transcriptômica digital. Cada ponto no gráfico representa um gene, com o eixo x representando o $\log_2$ (fold change) na expressão do gene entre os grupos de infecção bacteriana e viral, e o eixo y representando o $-\log_{10}$ do valor de p associado a essa mudança na expressão. Um gene é considerado significativamente expresso, de forma diferencial, se apresentar valor $p < 0,01$. Os pontos à direita indicam os genes regulados positivamente e, à esquerda, os genes regulados negativamente na infecção bacteriana em comparação aos casos virais.

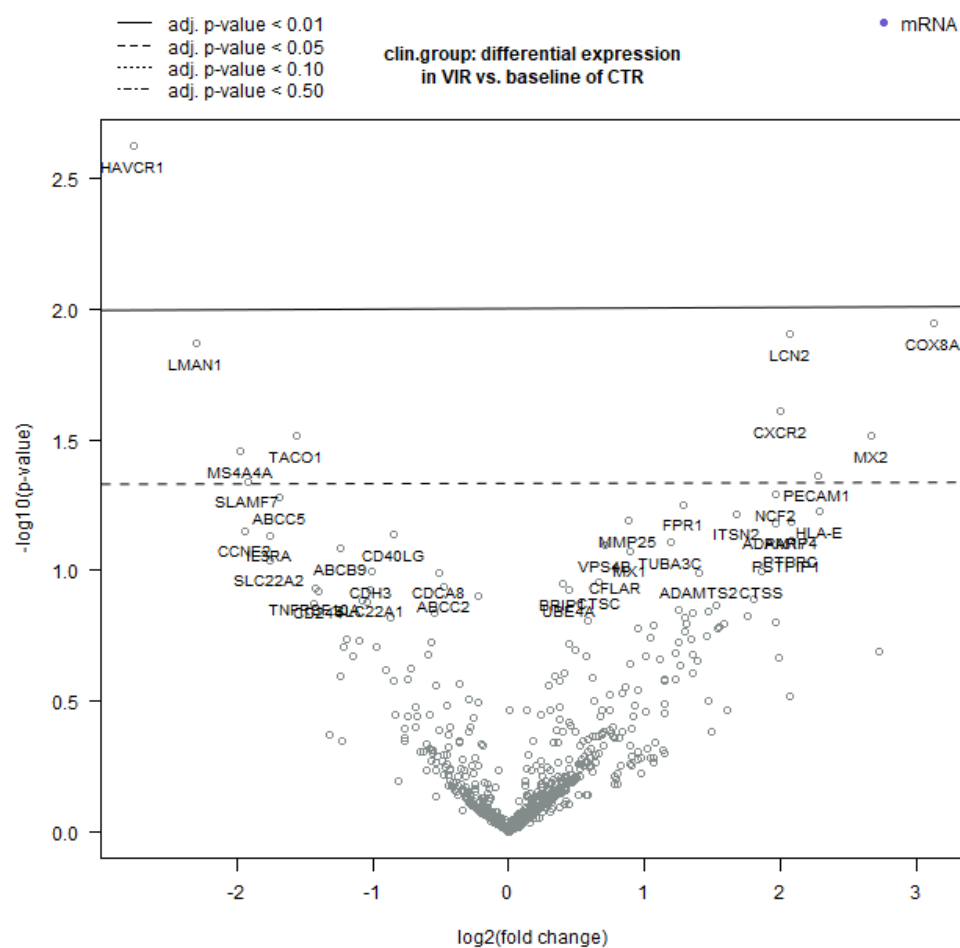


Figura 8. Volcano plot representando a expressão diferencial de genes em indivíduos com tonsilite viral em comparação com os controles, quantificado por nCounter transcriptômica digital. Cada ponto no gráfico representa um gene, com o eixo x representando o log₂ (fold change) na expressão do gene entre os grupos de infecção bacteriana e controle, e o eixo y representando o -log₁₀ do valor de p associado a essa mudança na expressão. Um gene é considerado significativamente expresso, de forma diferencial, se apresentar valor $p < 0,05$. Os pontos à direita indicam os genes regulados positivamente e à esquerda, os genes regulados negativamente na infecção viral em comparação com os controles.

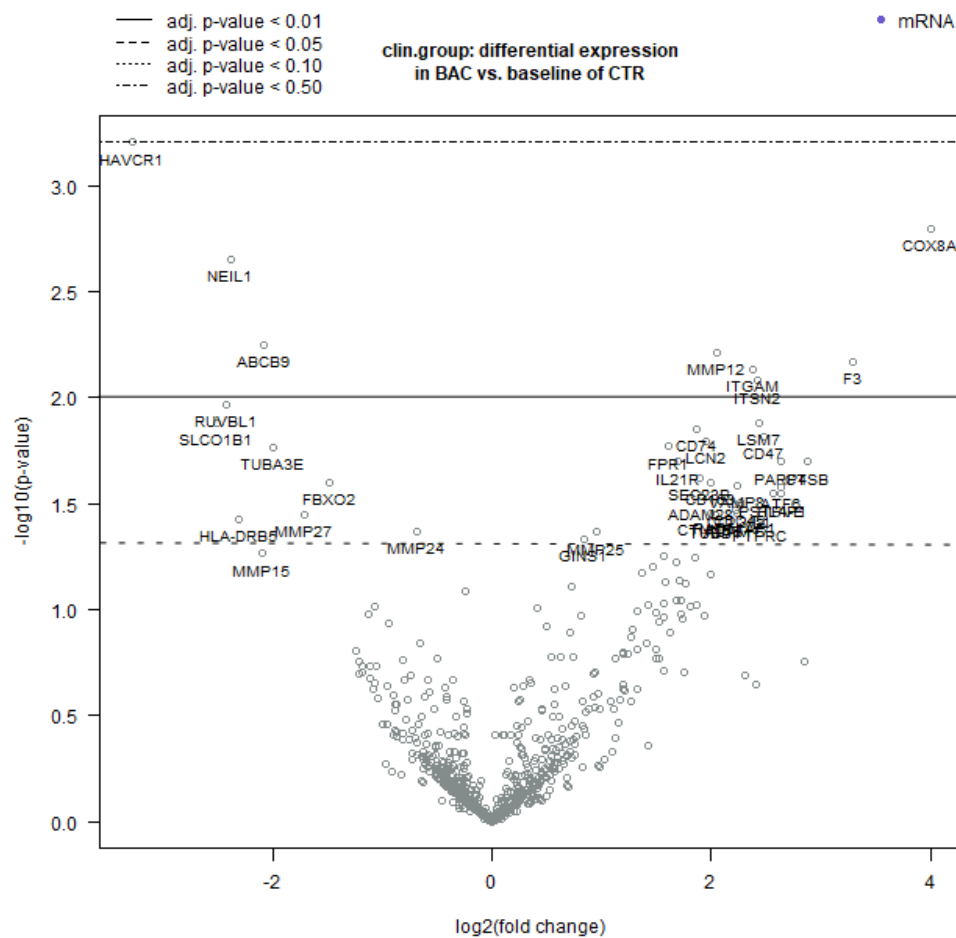


Figura 9. Volcano plot representando a expressão diferencial de genes em indivíduos com tonsilite bacteriana em comparação com os controles, quantificado por nCounter transcriptômica digital. Cada ponto no gráfico representa um gene, com o eixo x representando o $\log_2$ (fold change) na expressão do gene entre os grupos de infecção bacteriana e controle, e o eixo y representando o $-\log_{10}$ do valor de p associado a essa mudança na expressão. Um gene é considerado significativamente expresso, de forma diferencial, se apresentar valor $p < 0,01$. Os pontos à direita indicam os genes regulados positivamente e à esquerda, os genes regulados negativamente na infecção bacteriana em comparação com os controles

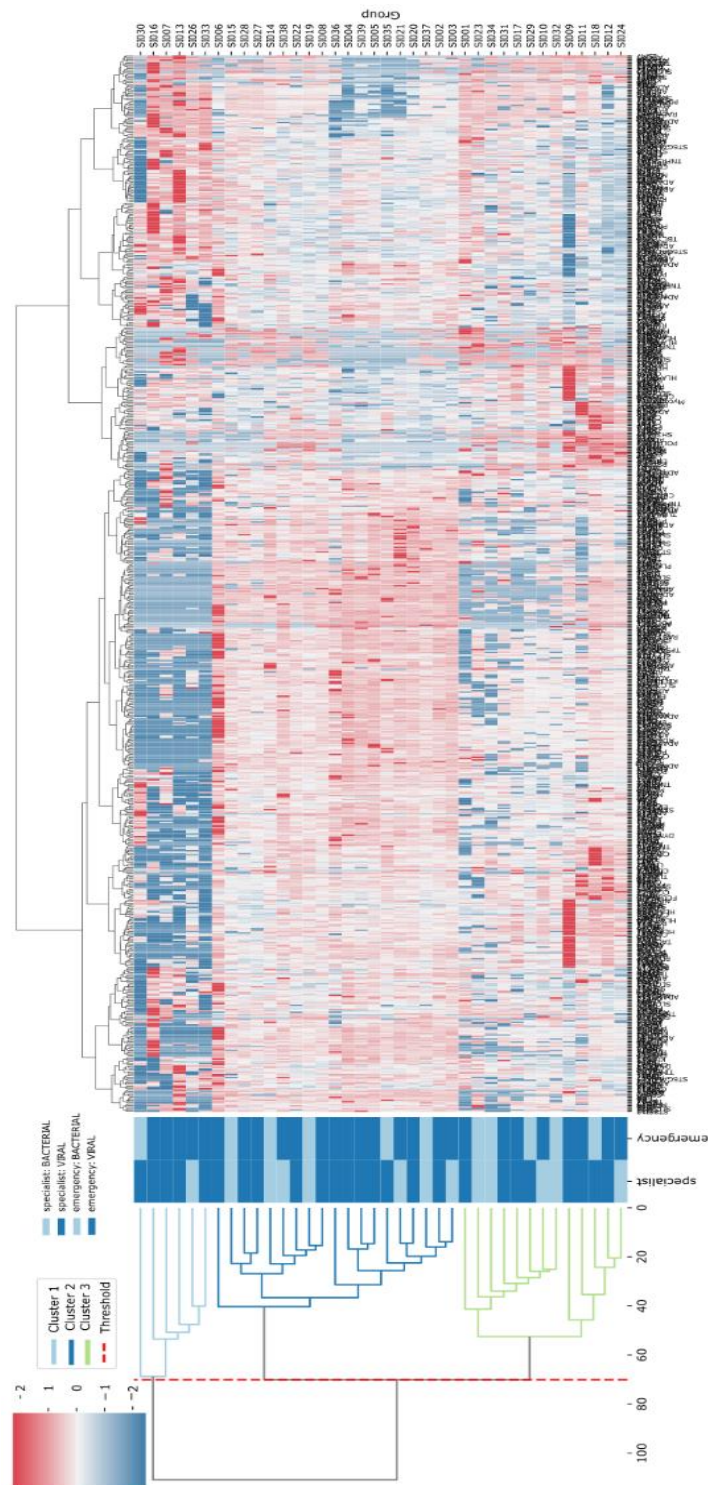


Figura 10. Clustermap da abundância relativa a 746 genes da resposta imune e variáveis de metadados associadas em diferentes amostras. O dendrograma à esquerda agrupa as amostras em três clusters distintos (Cluster 1, Cluster 2 e Cluster 3), com base na similaridade de seus perfis gênicos e características dos metadados. As barras coloridas adjacentes ao dendrograma detalham os valores de metadados para cada amostra: Specialist e emergency, os quais classificam as amostras em infecção bacteriana (azul claro) ou viral (azul escuro). O mapa de calor exibe a abundância relativa de diversas espécies microbianas em cada amostra. A escala de cores varia de azul (subexpressão) a vermelho (superexpressão), com valores z-score normalizados. A linha tracejada vermelha à esquerda (threshold) indica um limiar de corte utilizado na definição dos agrupamentos.

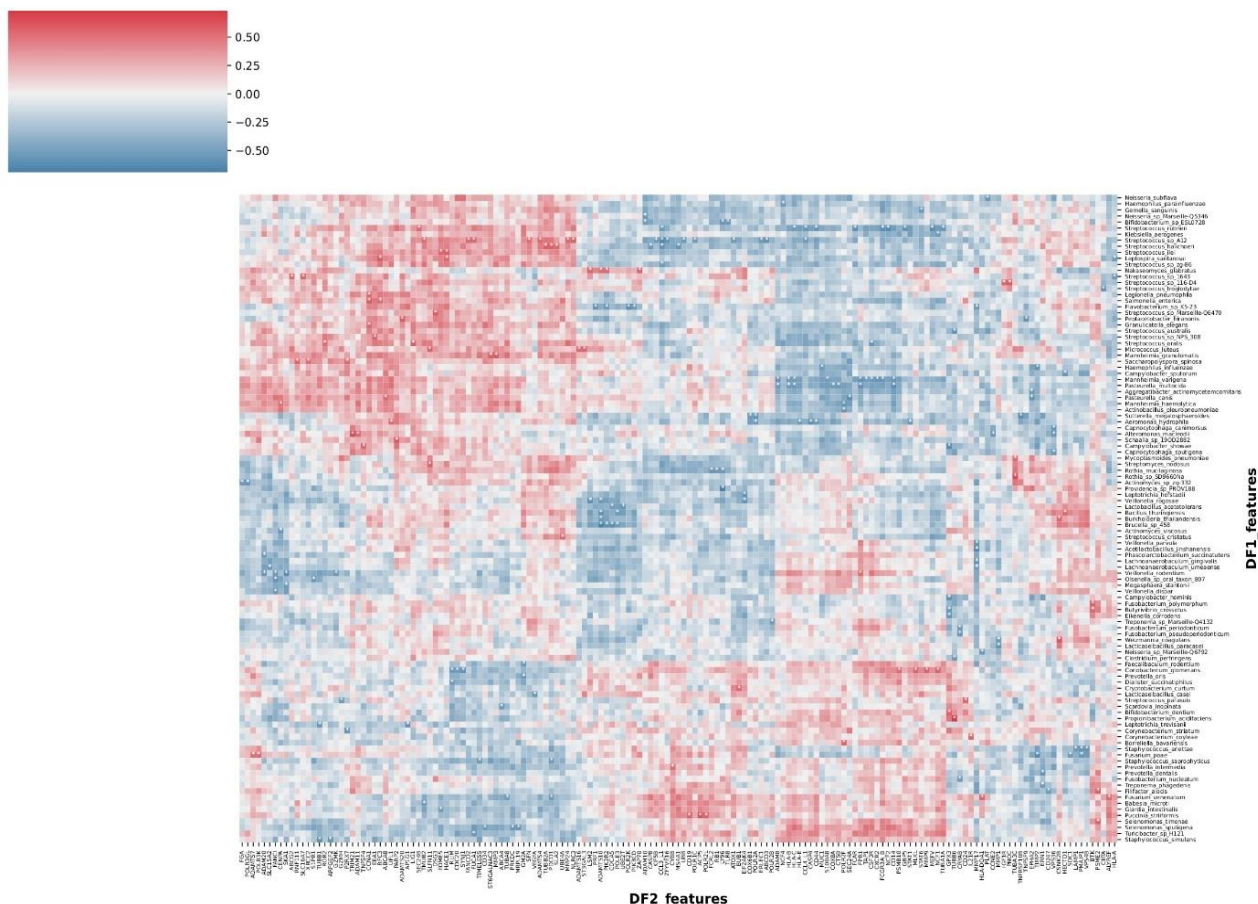


Figura 11. Heatmap de correlação entre os microrganismos e genes da resposta imune. Cada célula representa o coeficiente de correlação entre pares de variáveis dos dois conjuntos, com valores variando de $-0,5$ (correlação negativa, em azul) a $+0,5$ (correlação positiva, em vermelho). Células em branco indicam correlações próximas de zero. Asteriscos brancos sobre as células destacam correlações estatisticamente significativas ($p < 0,05$).

Enrichment Analysis ($-\log_{10}(p\text{-Val})$) with $p < 0.05$ threshold)

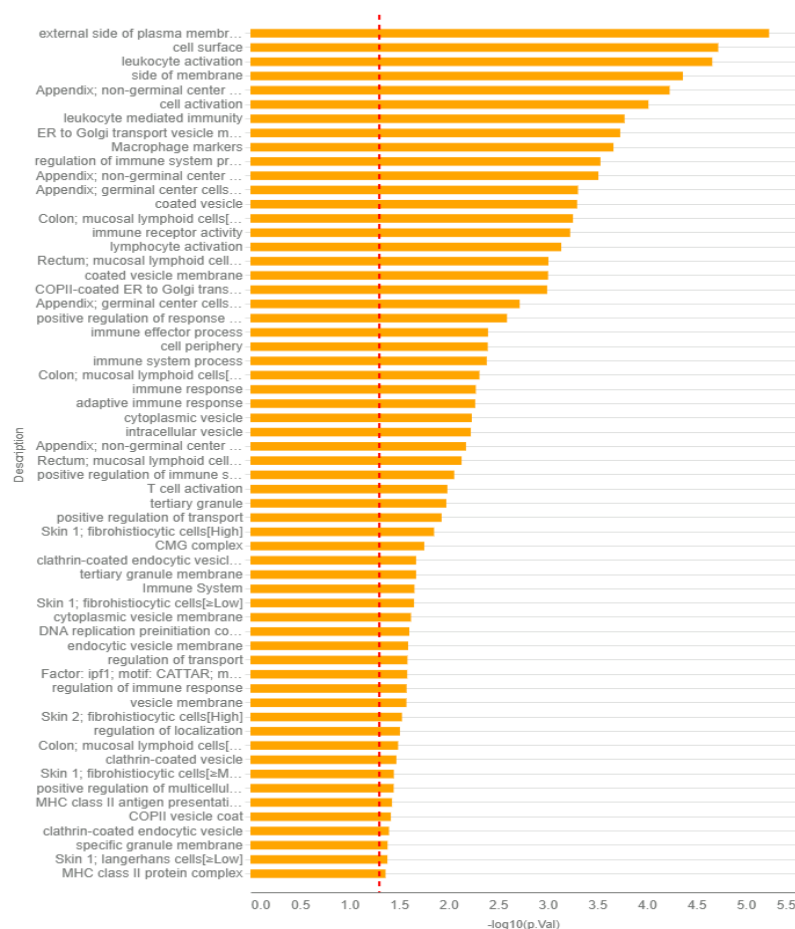


Figura 12. Análise de enriquecimento funcional de genes regulados positivamente nas suspeitas de infecção bacteriana, em comparação aos casos virais, utilizando o *gProfiler*. As barras em laranja representam os valores de $-\log_{10}(p\text{-valor})$ para cada termo biológico enriquecido, indicando a significância estatística do enriquecimento (quanto maior a barra, menor o p-valor e mais significativo o termo). A linha pontilhada vermelha serve como um limiar de significância, marcando o p-valor de 0.05. Todos os termos cujas barras se estendem além desta linha são considerados estatisticamente significativos.

Enrichment Analysis (-log₁₀(p.Val) with p < 0.05 threshold)

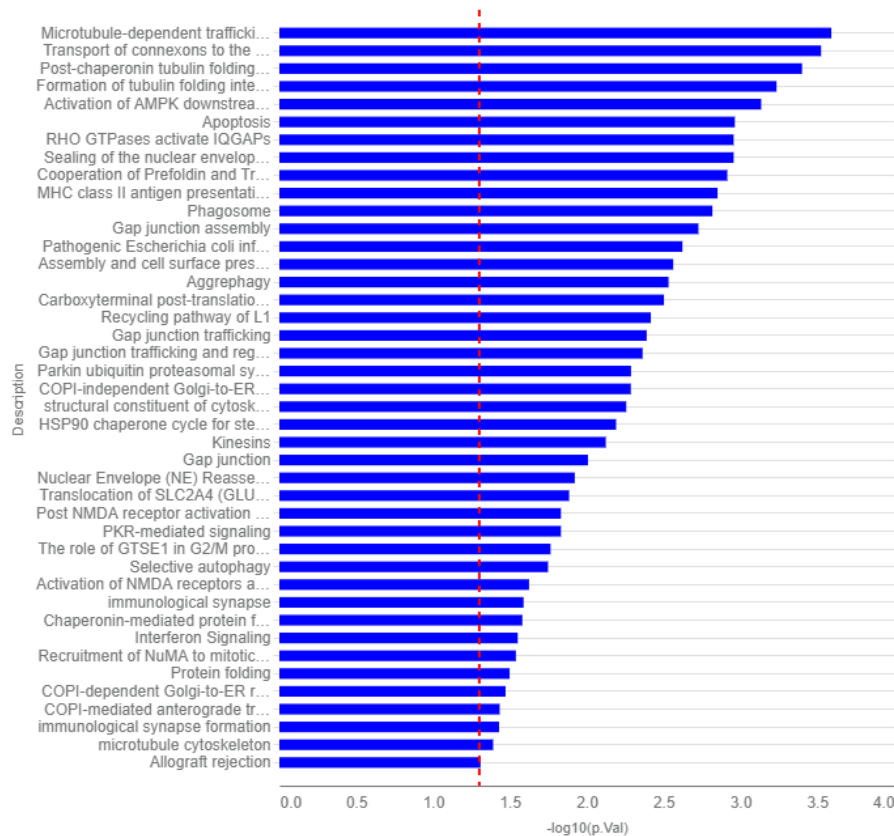


Figura 13. Análise de enriquecimento funcional de genes regulados positivamente nas suspeitas de infecção viral, em comparação aos casos bacterianos, utilizando o *gProfiler*. As barras em azul representam os valores de $-\log_{10}(\text{p-valor})$ para cada termo biológico enriquecido, indicando a significância estatística do enriquecimento (quanto maior a barra, menor o p-valor e mais significativo o termo). A linha vermelha serve como um limiar de significância, marcando o p-valor de 0.05. Todos os termos cujas barras se estendem além desta linha são considerados estatisticamente significativos.

Enrichment Analysis (-log₁₀(p.Val) with p < 0.05 threshold)

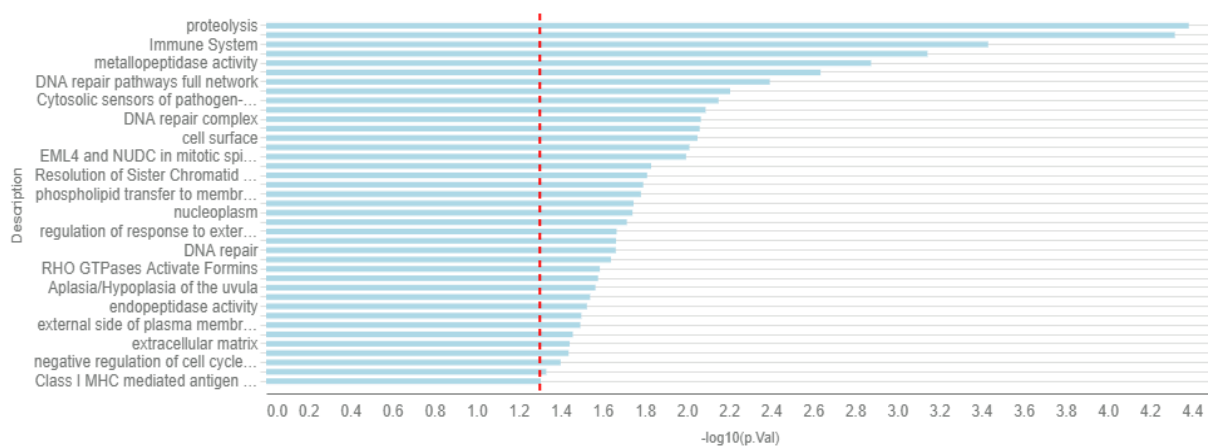


Figura 14. Análise de enriquecimento funcional dos genes diferencialmente expressos na matriz de correlação, em associação ao conjunto microbiano. As barras em azul representam os valores de $-\log_{10}(\text{p-valor})$ para cada termo biológico enriquecido, indicando a significância estatística do enriquecimento (quanto maior a barra, menor o p-valor e mais significativo o termo). A linha vermelha serve como um limiar de significância, marcando o p-valor de 0.05.

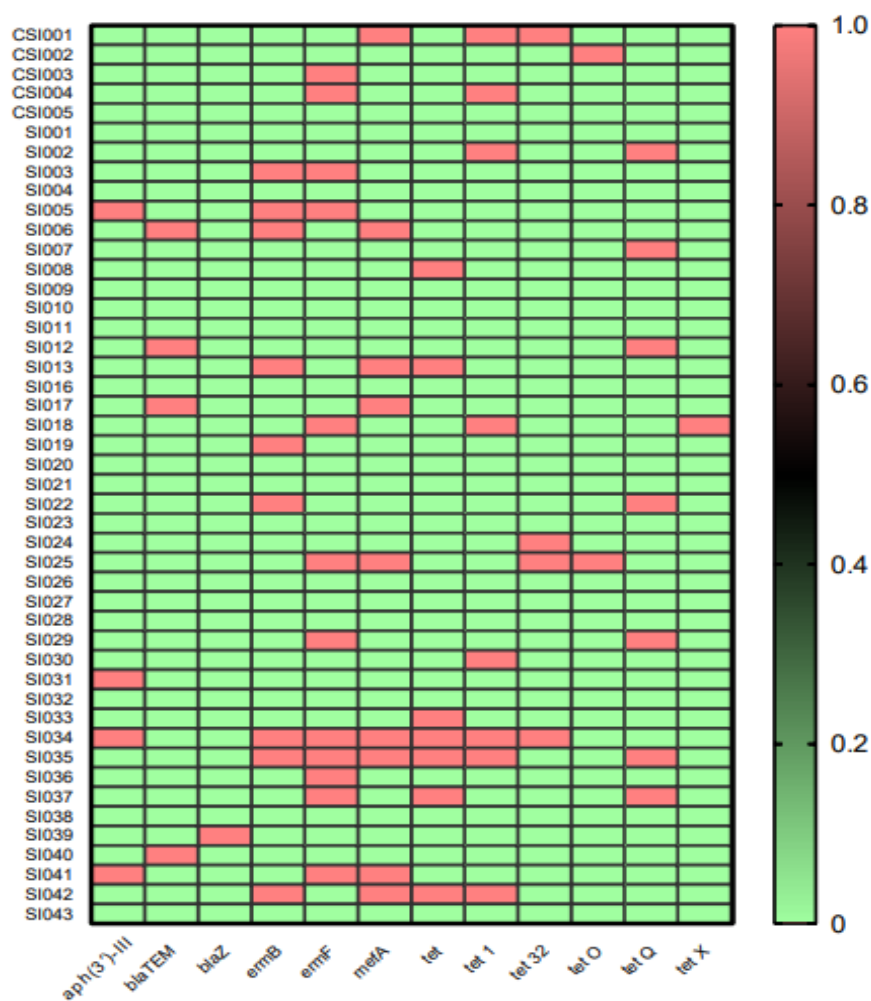


Figura 15. Heatmap da presença/ausência de genes de resistência antimicrobiana nas amostras dos pacientes com tonsilite. A cor vermelha representa uma correspondência perfeita e verde indica ausência de correspondência. As linhas representam as amostras individuais, enquanto as colunas representam os diferentes genes identificados.