

UNIVERSIDADE FEDERAL DO PARANÁ

GIULIA LEONEL PASCHOAL
JÚLIA FERNANDES DA SILVA
LARA RIBEIRO RAMPIM
LUCAS MÜLLER
NICOLLI DE OLIVEIRA ROSA

MODELO OCULTO DE MARKOV PARA PREDIÇÃO DE TROPISMO EM HIV-1

CURITIBA

2024

Giulia Leonel Paschoal
Júlia Fernandes da Silva
Lara Ribeiro Rampim
Lucas Müller
Nicolli de Oliveira Rosa

MODELO OCULTO DE MARKOV PARA PREDIÇÃO DE TROPISMO EM HIV-1

Trabalho apresentado à disciplina de
Bioinformática, Setor de Ciências Exatas,
Universidade Federal do Paraná.

Orientador: Prof. Dr. Eduardo Jaques Spinosa

CURITIBA

2024

SUMÁRIO

1 INTRODUÇÃO	4
2 FUNDAMENTAÇÃO TEÓRICA	6
3 DESCRIÇÃO DA ABORDAGEM COMPUTACIONAL	11
4 DESCRIÇÃO DAS FERRAMENTAS	12
4.1 BASE DE DADOS DE LOS ALAMOS	12
4.1.1 dataset new_db	12
4.1.2 Sequências de entrada CXCR4 E R5X4	12
4.2 AWK	12
4.3 ClustalO	12
4.4 HMMER	13
4.4.1 hmmbuild	13
4.4.2 hmmsearch	13
4.5 OUTRAS CONSIDERAÇÕES	13
5 RESULTADOS	14
5.1 CO-RECEPTORES R5X4	14
5.2 CO-RECEPTORES CXCR4	15
6 CONCLUSÕES	17
REFERÊNCIAS BIBLIOGRÁFICAS	18
APÊNDICE A - GitHub	20
ANEXO A - Database	21

1 INTRODUÇÃO

O Vírus da Imunodeficiência Humana (HIV, do inglês *Human Immunodeficiency Virus*) é um lentivírus que afeta o sistema imunológico, levando a sua deterioração, e dessa forma, fragiliza o sistema imune do paciente, que se torna indefeso a outros tipos de doenças e infecções subsequentes (DOUEK; ROEDERER; KOUP, 2009). Ademais, o HIV é o responsável pela Síndrome da Imunodeficiência Adquirida (Aids), que inicialmente se apresenta como sintomas semelhantes ao da gripe, no entanto, conforme a doença avança e compromete cada vez mais o sistema imunológico, o indivíduo se torna cada vez mais suscetível a outras infecções e doenças, como câncer (WEISS, 1993).

São conhecidos dois tipos de HIV hoje: HIV-1 e HIV-2, sendo que possuem muitas similaridades quando se trata de replicação e transmissão. Essencialmente, sua principal diferença está na progressão da imunodeficiência, que acontece mais rapidamente no HIV-1 (NYAMWEYA *et al.*, 2013) e por isso, este será o foco da discussão. Entender as individualidades de cada subtipo permite a aplicação de ferramentas de forma mais específica, resultando em estudos proporcionais à patogênese e prevalência do HIV-1.

O HIV é transmitido por meio de fluídos corporais como saliva, sêmen, leite materno, lubrificação vaginal, fluido pré-ejaculatório. Dessa forma, as principais vias de transmissão são por meio de relações sexuais desprotegidas, compartilhamento de agulhas infectadas e a transmissão de mãe para filho durante a gravidez, parto ou amamentação (BRASIL. Ministério da Saúde)

A destruição do sistema imune pelo HIV se dá pela infecção de células imunes como linfócitos T CD4+, nos quais o vírus invade e altera o DNA dessa célula para reproduzir o material genético do vírus possibilitando a sua replicação e invasão de outras células. Além disso, os linfócitos T CD8+ são responsáveis por reconhecer e induzir os linfócitos T CD4+ infectados à morte (KUMAR; ABBAS; ASTER, 2012). Devido a isso, o número de linfócitos do indivíduo decresce e fica abaixo de valores aceitáveis, indicando a perda da imunidade inata e assim se torna cada vez mais imunocomprometido (KUMAR; ABBAS; ASTER, 2012).

A infecção por HIV em seres humanos tem se mostrado uma pandemia atualmente, visto que aproximadamente 37,6 milhões de pessoas estavam vivendo com HIV no mundo em 2020, das quais 690 mil de pessoas morreram de doenças relacionadas à AIDS em 2020 (UNAIDS, 2020). Dessa maneira, a infecção por HIV, mais especificamente do HIV-1, apresenta-se como uma problemática significativa na saúde pública global e pesquisas

envolvendo terapia antirretroviral permitiram o aumento da expectativa de vida de pacientes com HIV. No entanto, ainda há muito a ser descoberto pela pesquisa, como em relação ao tropismo das cepas desse vírus pelas células infectadas.

A determinação do tropismo do HIV é importante para orientar a escolha de medicamentos antirretrovirais para o paciente, visto que, ao saber qual receptor o vírus é mais provável de utilizar para infectar a célula, permite um tratamento mais direcionado para determinada cepa do HIV (LIN; KURITZKES, 2009). As diferentes cepas de HIV apresentam mudanças em suas sequências genéticas e isso interfere na interação do vírus com receptores celulares de células hospedeiras. Da mesma maneira, essas diferenças genéticas afetam o tropismo do HIV com os receptores de células hospedeiras, que pode ter afinidade com co-receptor CCR5 (tropismo CCR5), pelo co-receptor CXCR4 (tropismo CXCR4) ou por ambos. (ZAITSEVA; PEDEN; GOLDING, 2003). Nesse caso, uma grande maioria dos indivíduos positivos para HIV-1 são inicialmente infectados via CCR5, que resulta na alta patogenicidade conhecida.

Portanto, as diferentes cepas do HIV podem exibir diferentes padrões de tropismo, o que tem implicações na progressão da infecção, na resposta ao tratamento e na seleção de terapias antirretrovirais específicas. A análise do tropismo viral é fundamental para compreender a dinâmica da infecção pelo HIV e para personalizar o tratamento de acordo com o perfil genético do vírus presente em cada paciente.

Nesse sentido, para avaliar o tropismo entre sequências homólogas, ou seja, as cepas de HIV, pode ser usado o Perfil de Modelos de Markov Oculto (pHMM). O pHMM é criado a partir de alinhamentos de sequências homólogas e codifica informações específicas de posição sobre a frequência de aminoácidos em cada posição do alinhamento, bem como a frequência de inserções e deleções. Assim, permite resumir as sequências homólogas e suas semelhanças em um único modelo, permitindo testar outras sequências.(CRISTIANINI; HAHN, 2006).

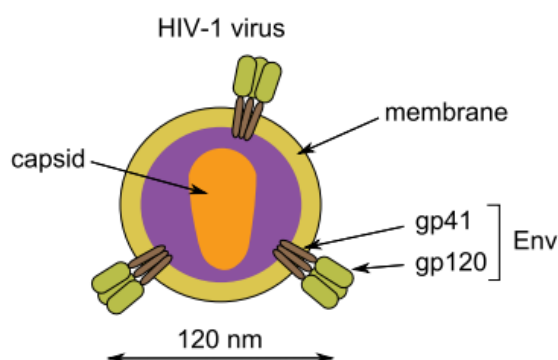
Dessa forma, o estudo foi baseado no experimento da pesquisa de Chen, Wang e Pan (2019), e o objetivo é testar a utilização do método HMM na predição de tropismo. Visto que o pHMM é fundamental para identificar recombinantes e novas cepas na sequência do HIV-1, que além de auxiliar na compreensão da diversidade viral, também ajuda a compreender qual vírus está acometendo o paciente, como será a progressão da doença e direciona o tratamento com os antirretrovirais mais adequados para cada caso.

2 FUNDAMENTAÇÃO TEÓRICA

O HIV é um dos vírus da família *retroviridae*, com uma membrana viral composta por uma bicamada lipídica, derivada da célula hospedeira e glicoproteínas do envelope codificadas pelo vírus (LLORENTE GARCÍA; MARSH, 2020) (ARENZANA-SEISDEDOS; PARMENTIER, 2006). A glicoproteína do envelope do HIV é sintetizada pelo retículo endoplasmático da célula hospedeira como a molécula precursora trimérica gp160, a qual, posteriormente, sofre a ação de proteases da família furina, originando duas subunidades proteicas ligadas não covalentemente (LLORENTE GARCÍA; MARSH, 2020) (ARENZANA-SEISDEDOS; PARMENTIER, 2006).

As subunidades são gp120, uma glicoproteína do envelope externo, e gp41, uma glicoproteína transmembrana e elas são responsáveis pela interação com os receptores da superfície celular, mediando a fusão de membranas (LLORENTE GARCÍA; MARSH, 2020) (Figura 1).

Figura 1. (LLORENTE GARCÍA; MARSH, 2020)



Estrutura do vírus HIV-1. Membrana plasmática derivada da célula hospedeira, envolvendo o capsídeo, que abriga o RNA viral e embebida pela proteína de envelope (Env) composta pelas glicoproteínas gp41 e gp120.

Normalmente, os receptores da superfície celular têm funções celulares essenciais, exercidas a partir da comunicação entre o ambiente extracelular e o interior, podendo desencadear em mudanças nos domínios citoplasmáticos do receptor, que por sua vez, podem afetar a interação com ligantes extracelulares (LLORENTE GARCÍA; MARSH, 2020). Os vírus podem explorar esse potencial de sinalização para sua entrada na célula e subsequente replicação (LLORENTE GARCÍA; MARSH, 2020).

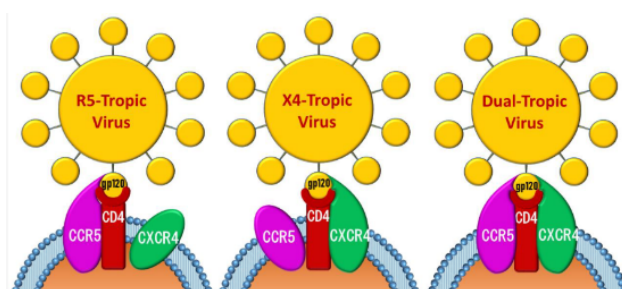
Frequentemente, é o uso desses receptores que determina o tropismo do vírus, que pode ser definido, portanto, como a suscetibilidade de determinados hospedeiros e tipos de células à infecção por um vírus específico (LLORENTE GARCÍA; MARSH, 2020).

Para o HIV-1, a interação entre a glicoproteína viral gp120 e o receptor celular CD4 é um passo inicial indispensável, responsável pela indução de mudanças conformacionais na proteína envelope, responsáveis pela exposição do sítio de ligação à co-receptores acoplados à proteína G, o qual envolve a base do loop V3 (LLORENTE GARCÍA; MARSH, 2020) (GRANDE; OCCHIUZZI; RIZZUTI; et al., 2019) (BRELOT; CHAKRABARTI, 2018) (Figura 2). Os receptores acoplados à proteína G, que funcionam como co-receptores do HIV, foram identificados como receptores de quimiocinas e apesar de vários identificados *in vitro*, os principais *in vivo* são CCR5 e CXCR4 (ZAITSEVA; PEDEN; GOLDING, 2003).

Desta forma, a glicoproteína viral consegue interagir com o co-receptor CCR5 ou com o CXCR4, que determinam o tropismo do HIV-1 e são responsáveis por modificações conformacionais tanto em gp120 quanto em gp41, levando à inserção dos peptídeos de fusão viral na membrana da célula-alvo, formação de um poro de fusão e permitindo a eventual entrada do vírus (LLORENTE GARCÍA; MARSH, 2020) (GRANDE; OCCHIUZZI; RIZZUTI; et al., 2019) (BRELOT; CHAKRABARTI, 2018).

Os vírus que utilizam preferencialmente o CCR5 são responsáveis pela infecção celular inicial, sendo esse co-receptor expresso principalmente em macrófagos, e denominados HIV R5 (GRANDE; OCCHIUZZI; RIZZUTI; et al., 2019). Em contraste, as cepas X4 do HIV, geralmente associadas a estágios tardios e a progressão da doença, usam o CXCR4 como co-receptor, predominante em células T (LLORENTE GARCÍA; MARSH, 2020) (GRANDE; OCCHIUZZI; RIZZUTI; et al., 2019) (CHEN; WANG; PAN, 2019). Os HIVs de tropismo duplo, também chamados de R5X4, podem infectar células CD4+ usando qualquer um dos dois co-receptores (LLORENTE GARCÍA; MARSH, 2020) (Figura 2).

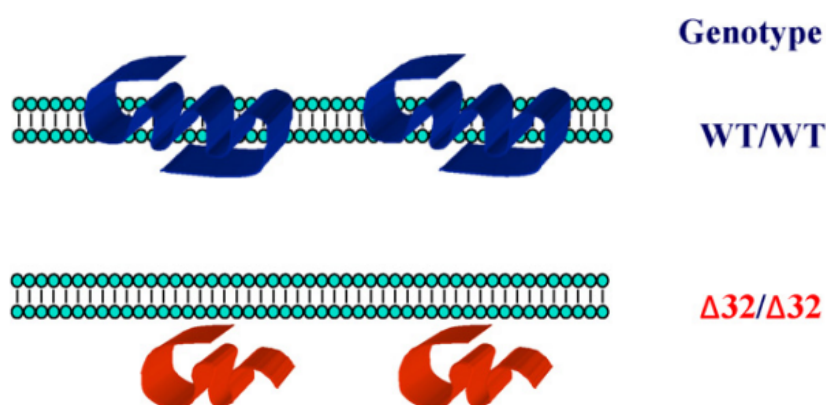
Figura 2. (GRANDE; OCCHIUZZI; RIZZUTI; et al., 2019)



Tropismo de HIV-1. Vírus R5 interagem preferencialmente com CCR5 (macrófagos), vírus X4 interagem com CXCR4 em estágios tardios (linfócitos T) e vírus de tropismo duplo conseguem interagir com ambos.

Uma evidência de que CCR5 é um receptor essencial para o estágio inicial da infecção é que variantes não funcionais do CCR5, haplótipos envolvendo o gene CCR5, o número de cópias dos genes que codificam ligantes do CCR5, estão estritamente associados à suscetibilidade individual de contrair infecções pelo HIV (ARENZANA-SEISDEDOS; PARMENTIER, 2006). O principal exemplo disso é a resistência ao HIV-1 em indivíduos onde a função desse co-receptor é desativada, quando o gene CCR4- Δ 32 aparece em homozigose (CHEN; WANG; PAN, 2019). Essa mutação de deleção resulta em um receptor não funcional que fica retido no retículo endoplasmático e não chega à superfície celular (ARENZANA-SEISDEDOS; PARMENTIER, 2006) (Figura 3).

Figura 3. (ARENZANA-SEISDEDOS; PARMENTIER, 2006)



Resistência ao HIV-1. Mutação de deleção de 32 pb Δ 32 de CCR5, em relação ao gene selvagem (WT: *Wild Type*), associada à resistência ao vírus devido a ausência do receptor na superfície celular.

Vários estudos indicaram a fusão do HIV, que ocorre na superfície das células-alvo, como o principal mecanismo de entrada para o vírus (LLORENTE GARCÍA; MARSH, 2020). É a partir dessa via que há liberação do capsídeo viral no citoplasma da célula-alvo, iniciando a replicação viral e as consequências patogênicas da infecção celular (LLORENTE GARCÍA; MARSH, 2020). Por isso, o CD4 e os co-receptores são alvos primários para várias classes de inibidores que estão atualmente em teste (ZAITSEVA; PEDEN; GOLDING, 2003). Principalmente CCR5, que tem sido foco de estudos intensivos, desde a descoberta de seu papel na infecção por HIV em 1996 (BRELOT; CHAKRABARTI, 2018). Esses estudos

levaram ao desenvolvimento de drogas que visam CCR5, como o Maraviroc (MVC), se tornando em 2007 o primeiro inibidor de receptor de quimiocina aprovado clinicamente pelo *Food and Drug Administration* (FDA) (BRELOT; CHAKRABARTI, 2018).

O tratamento mais bem-sucedido de pacientes infectados pelo HIV-1 até agora é a partir da administração de uma combinação de agentes antirretrovirais que visam proteínas específicas necessárias para a replicação viral, sendo mais comuns os agentes inibidores de entrada e fusão (GRANDE; OCCHIUZZI; RIZZUTI; et al., 2019). Com o advento dos antagonistas de receptores de quimiocinas, há um impulso adicional para decifrar os determinantes complexos do uso de co-receptores e os mecanismos que impulsionam a evolução dos vírus X4 (CLEMENTI; LAZZARIN, 2010).

Neste contexto, os ensaios de tropismo viral desempenham um papel crucial na identificação de pacientes candidatos ao tratamento com inibidores de entrada e no desenvolvimento contínuo de antagonistas de co-receptores (CLEMENTI; LAZZARIN, 2010). Mesmo que muitos agentes atuando seletivamente em CCR5 ou CXCR4 isoladamente tenham sido descritos, apenas alguns poucos antagonistas de ambos foram identificados e a escolha para o uso clínico de antagonistas CCR5 ou CXCR4 é feita com base na identificação do tropismo da cepa nos pacientes (GRANDE; OCCHIUZZI; RIZZUTI; et al., 2019). Além disso, dados epidemiológicos sobre o uso de co-receptores são cruciais para elucidar a relação entre o tropismo a cada co-receptor e a progressão da doença (CLEMENTI; LAZZARIN, 2010).

Nas últimas décadas, portanto, dois tipos de métodos de teste de tropismo têm sido desenvolvidos (CHEN; WANG; PAN, 2019). Os métodos fenotípicos costumam ser mais caros, longos, pouco acessíveis, devido à necessidade de centros especializados e possuem baixa sensibilidade, não promovendo resultados consistentes em amostras com carga viral abaixo de 1000 cópias/mL (CHEN; WANG; PAN, 2019). Os genotípicos, por outro lado, são preferidos devido ao baixo custo, menor tempo de resposta, grande acessibilidade e maior sensibilidade, fornecendo resultados confiáveis em amostras com carga viral abaixo de 1000 cópias/mL, baseados em estatística ou aprendizado de máquina (CHEN; WANG; PAN, 2019).

O estudo do HIV-1 apresenta uma dificuldade para detectar a recombinação de sequências de HIV-1 e seus subtipos, a solução desse problema é de extrema importância pois, devido à diversidade genética do vírus, classificar com precisão seus subtipos e suas formas recombinantes é essencial para entender a sua epidemiologia e orientar melhores decisões terapêuticas e preventivas. O HIV-1 é dividido em quatro principais grupos filogenéticos: M, N, O e P, sendo o grupo M o mais estudado, pois é o responsável pela pandemia de HIV. O

grupo M é dividido em 10 subtipos e dentre eles a recombinação é muito frequente. (BULLA et al., 2014; SCHULTZ et al., 2006).

Neste estudo, será analisado o tropismo do vírus HIV-1, que se altera de forma muito rápida por conta das recombinações de sequência do vírus. Prever o tropismo é crucial para determinar a eficácia de tratamentos, antecipar a evolução da doença e identificar variações do HIV-1 que podem desenvolver resistência a determinados medicamentos. O principal objetivo dessas análises é melhorar a precisão na predição do co-receptor que o HIV-1 utilizará para infectar as células (CHEN; WANG; PAN, 2019).

3 DESCRIÇÃO DA ABORDAGEM COMPUTACIONAL

São propostas por Chen, Wang e Pan (2019), duas soluções para a análise do tropismo do HIV-1, dentre elas o uso do Modelo Oculto de Markov (HMM), que é um modelo estatístico usado para representar sistemas que possuem estados ocultos, ou seja, estados que não são diretamente observáveis, mas que influenciam estados visíveis. O modelo de HMM é treinado por um conjunto de dados e, uma vez treinado, pode ser usado para inferir sequências de estados ocultos mais prováveis dada uma sequência de observação (CRISTIANINI; HAHN, 2006, p. 63).

O Modelo Oculto de Markov é um modelo finito aplicado em séries temporais e sequenciais lineares, utilizado para descrever famílias de proteínas. O modelo tem um bom desempenho em termos de alinhamento de sequências e reconhecimento de homologia, pois, por meio de cadeias de estado oculto, consegue lidar melhor com lacunas nas famílias de proteínas (CHEN; WANG; PAN, 2019).

Para a criação de um modelo HMM são necessários quatro componentes essenciais: (i) o alfabeto de símbolos, que são os diferentes símbolos que o modelo pode reconhecer (por exemplo, ACGT para DNA); (ii) o número de estados no modelo, representando as diferentes categorias ou situações possíveis; (iii) as probabilidades de emissão, que indicam a probabilidade de cada símbolo ser emitido de um determinado estado; e (iv) as probabilidades de transição, que mostram a probabilidade de mover de um estado para outro, incluindo permanecer no mesmo estado; as probabilidades de emissão e transmissão devem somar 1 para garantir um modelo válido. Qualquer modelo que tenha essas propriedades é um Modelo Oculto de Markov (EDDY, 2004).

4 DESCRIÇÃO DAS FERRAMENTAS

Neste estudo foi desenvolvido um script Bash que utiliza de várias ferramentas de bioinformática para manipulação e análise de sequências de DNA ou proteínas. As ferramentas utilizadas no script, bem como o seu propósito no contexto de processamento das sequências, são descritas a seguir conforme a ordem de uso.

4.1 BASE DE DADOS DE LOS ALAMOS

A base de dados de Los Alamos, indicada no anexo A, é um repositório abrangente de sequências de HIV, onde foi obtido o dataset utilizado neste estudo. As sequências de R5, X4 e duplo foram extraídas e filtradas para análise.

4.1.1 dataset new_db

O dataset new_db foi construído a partir de sequências do banco de dados. Este dataset contém 2335 sequências de tropismo R5, 245 sequências de tropismo X4 e 418 sequências de tropismo duplo, totalizando 2998 sequências. Sequências duplicadas foram removidas para garantir alta qualidade.

4.1.2 Sequências de entrada CXCR4 E R5X4

Foram extraídas sequências conhecidas dos co-receptores CXCR4 e R5X4 do loop V3 do banco de dados. Essas sequências serão utilizadas como entradas no experimento para demonstrar que os modelos HMM gerados para R5 e X4 são capazes de indicar o tropismo esperado para cada um dos co-receptores.

4.2 AWK

A ferramenta awk (AWK... 2023) foi utilizada para concatenar todas as sequências R5 em um único arquivo, e todas as sequências X4 em outro.

```
awk '{print}' newdb/ccr5.fasta newdb/r5x4.fasta > tmp/all_r5.fasta  
awk '{print}' newdb/cxcr4.fasta newdb/r5x4.fasta > tmp/all_x4.fasta
```

4.3 ClustalO

O Clustal Omega ou ClustalO (CLUSTAL... 2021), é uma ferramenta de alinhamento múltiplo de sequências, é usada no script para alinhar as sequências concatenadas dos arquivos FASTA.

```
clustalo -i tmp/all_r5.fasta -o tmp/aligned_r5.fasta --force  
clustalo -i tmp/all_x4.fasta -o tmp/aligned_x4.fasta --force
```

4.4 HMMER

O pacote HMMER (HMMER, 2024) é utilizado para construir e pesquisar modelos de perfil HMM. Duas ferramentas principais deste pacote são usadas no script, `hmmbuild` e `hmmsearch`.

4.4.1 `hmmbuild`

A ferramenta `hmmbuild` é usada para construir modelos de perfil HMM a partir dos alinhamentos de sequências gerados pelo ClustalO.

```
hmmbuild tmp/r5.hmm tmp/aligned_r5.fasta  
hmmbuild tmp/x4.hmm tmp/aligned_x4.fasta
```

4.4.2 `hmmsearch`

A ferramenta `hmmsearch` é utilizada para procurar sequências que correspondem aos perfis HMM criados. No script ela é utilizada para classificar sequências em um arquivo FASTA de entrada contra os perfis HMM criados para determinar a probabilidade de que a sequência pertença a um tropismo específico.

```
hmmsearch --tblout results_r5.txt tmp/r5.hmm "$1"  
hmmsearch --tblout results_x4.txt tmp/x4.hmm "$1"
```

4.5 OUTRAS CONSIDERAÇÕES

Todas as manipulações e análises são realizadas em arquivos no formato FASTA, que é um formato de texto utilizado para representar sequências de nucleotídeos ou proteínas.

Para a geração dos gráficos utilizados neste estudo, foi desenvolvido um script em Python que faz uso do pacote `matplotlib`.

5 RESULTADOS

Para a interpretação dos resultados, os campos de interesse são o E-value e o Score.

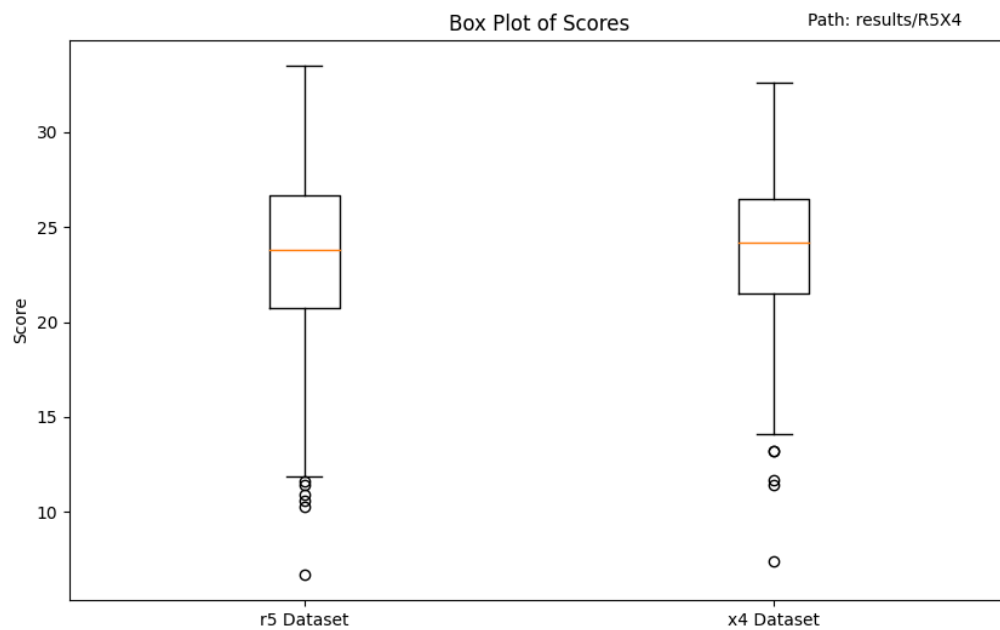
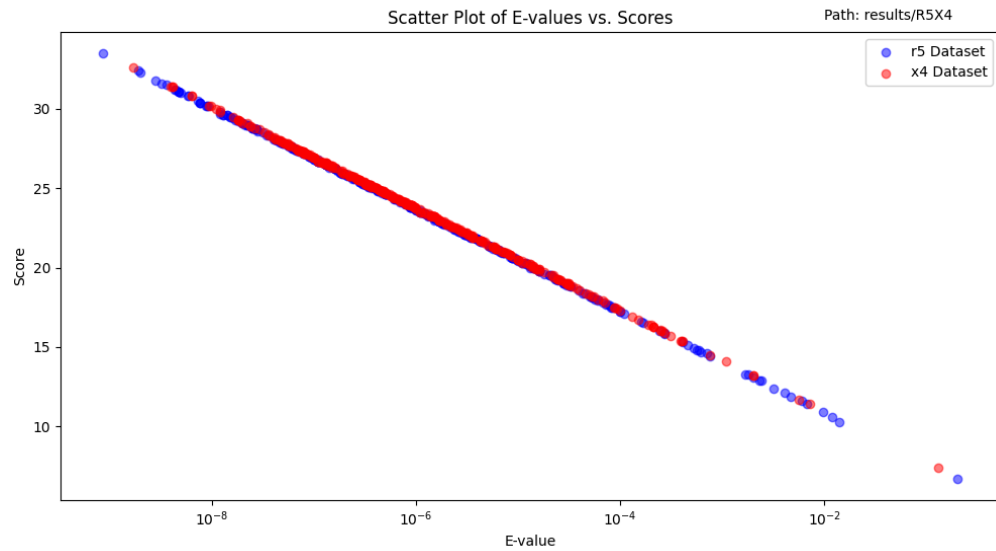
O E-value indica a significância de uma correspondência. Valores próximos de zero são mais significativos e indicam uma correspondência mais forte entre a sequência de consulta e o perfil HMM. Um E-value baixo sugere que a correspondência é improvável de ter ocorrido por acaso, o que aumenta a confiança na predição.

O Score representa a qualidade do alinhamento entre a sequência de consulta e o perfil HMM. Scores mais altos indicam melhores alinhamentos, sugerindo uma maior similaridade entre a sequência e o modelo HMM.

Utilizando esses dois valores, é possível identificar se um determinado loop V3 é associado aos co-receptores CCR5, CXCR4 ou ambos (R5X4). Scores elevados e E-values baixos são indicadores fortes de uma correspondência confiável, auxiliando na determinação do tropismo da sequência. As correspondências entre os co-receptores R5X4 e CXCR4 podem ser visualizadas analisando o alinhamento dessas sequências, extraídas da base de dados de Los Alamos, com os perfis HMM para R5 e X4 do dataset new_db. Os experimentos descritos podem ser reproduzidos a partir de nosso repositório no GitHub, conforme indicado no Apêndice A.

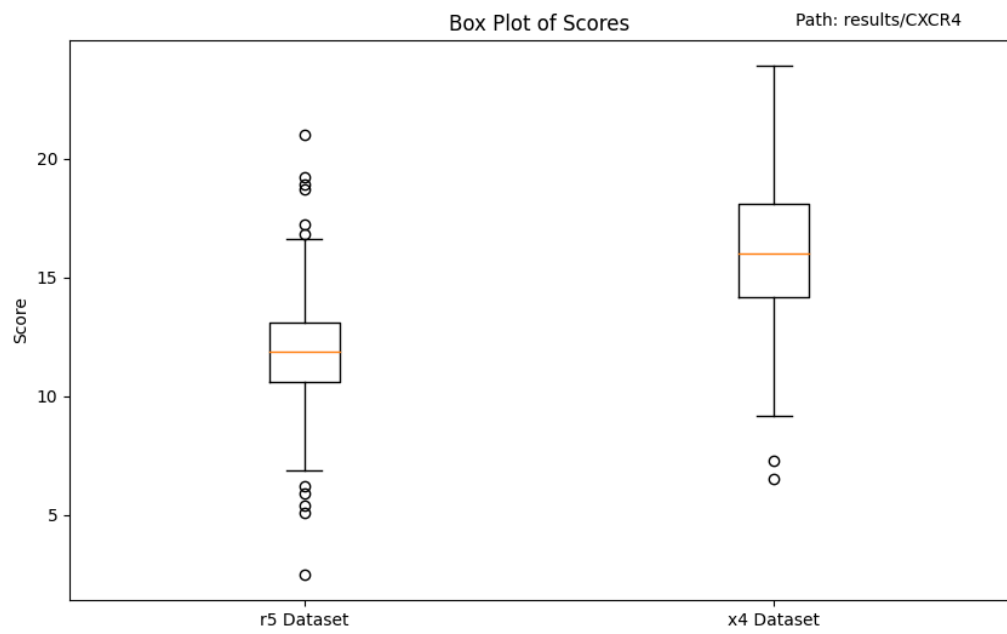
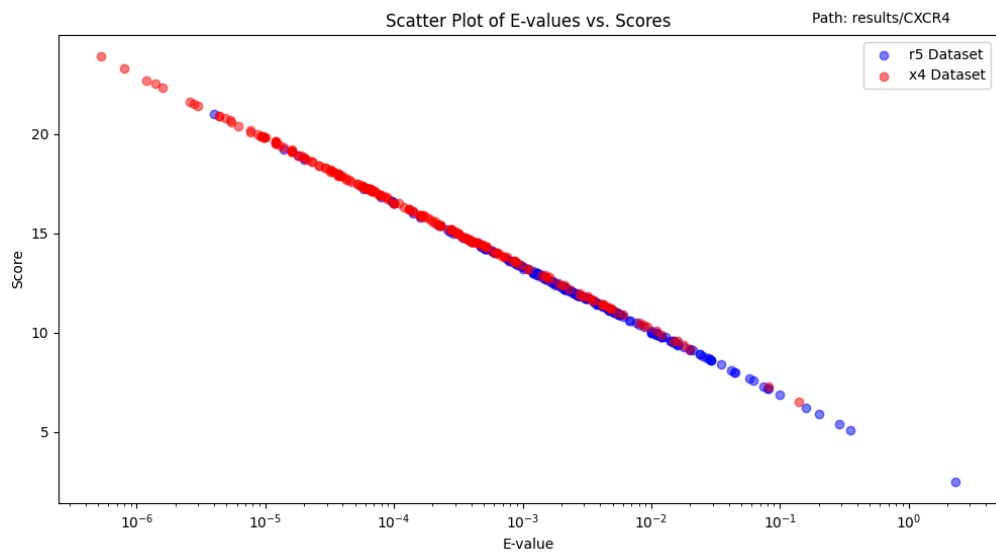
5.1 CO-RECEPTORES R5X4

Os resultados deste co-receptor mostram que as sequências testadas possuem fortes correspondências com ambos perfis HMM. Isso é um indicativo de que as sequências possuem características que são típicas dos tropismos R5 e X4, respectivamente. Como os resultados de R5 e X4 são comparáveis em termos de E-value e Score, pode-se concluir que a sequência testada apresenta tropismo duplo (R5X4).



5.2 CO-RECEPTORES CXCR4

Os resultados indicam que as sequências analisadas possuem uma correspondência muito mais forte com o perfil HMM para CXCR4 (X4) do que com o perfil para CCR5 (R5). Portanto, as sequências analisadas são classificadas como CXCR4, baseando-se nos resultados que mostram uma alta probabilidade de correspondência com o perfil HMM para tropismo X4.



6 CONCLUSÕES

A partir da necessidade de aprimoramento do tratamento dos pacientes com HIV, o presente trabalho objetivou uma análise detalhada do tropismo do HIV-1 utilizando o Modelo Oculto de Markov (HMM). Essa implementação permitiu a compreensão da interação viral com os receptores celulares CCR5 e CXCR4, e como consequência, a previsibilidade de prevalência de cepas do HIV-1.

A utilização das ferramentas do pacote HMMER permitiram identificar o tropismo das sequências virais com sucesso, ou seja, a implementação do Modelo Oculto de Markov conseguiu determinar quais co-receptores o vírus está mais propenso a utilizar e por isso, facilita a categorização dos subtipos virais e auxilia na adaptação das estratégias de personalização do tratamento antirretroviral de acordo com as características específicas de cada cepa.

A metodologia adotada neste estudo não apenas aprimora as técnicas de diagnóstico e monitoramento do HIV-1, mas também promove avanços significativos na compreensão da diversidade viral e na evolução do planejamento de saúde pública voltadas para o controle da pandemia de HIV discutida.

Dessa forma, essa análise realizada e descrita neste trabalho contribui para a seleção de terapias mais eficazes e para a previsão da progressão da doença em pacientes infectados pelo HIV-1.

REFERÊNCIAS BIBLIOGRÁFICAS

ARENZANA-SEISDEDOS, Fernando; PARMENTIER, Marc. **Genetics of resistance to HIV infection: Role of co-receptors and co-receptor ligands**. *Seminars in Immunology*, v. 18, n. 6, p. 387–403, dez. 2006.

AWK Programming: A User's Guide for GNU Awk. A User's Guide for GNU Awk. 2023. Disponível em: <https://www.gnu.org/software/gawk/manual/gawk.html>. Acesso em: 10 jul. 2024.

BRASIL. Ministério da Saúde. **Transmissão da AIDS/HIV**. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/a/aids-hiv/transmissao-da-aids-hiv>. Acesso em: 9 jul. 2024.

BRELOT, Anne; CHAKRABARTI, Lisa A. **CCR5 Revisited: How Mechanisms of HIV Entry Govern AIDS Pathogenesis**. *Journal of Molecular Biology*, v. 430, n. 17, p. 2557–2589, ago. 2018.

BULLA, Ingo; SCHULTZ, Anne-Kathrin; CHESNEAU, Christophe; MARK, Tanya; SEREA, Florin. **A model-based information sharing protocol for profile Hidden Markov Models used for HIV-1 recombination detection**. *BMC Bioinformatics*, v. 15, n. 1, p. 205, dez. 2014.

CHEN, Xiang; WANG, Zhi-Xin; PAN, Xian-Ming. **HIV-1 tropism prediction by the XGboost and HMM methods**. *Scientific Reports*, v. 9, n. 1, p. 9997, 10 jul. 2019.

CLEMENTI, M.; LAZZARIN, A. **Human immunodeficiency virus type 1 fitness and tropism: concept, quantification, and clinical relevance**. *Clinical Microbiology and Infection*, v. 16, n. 10, p. 1532–1538, out. 2010.

CLUSTAL Omega: The last alignment program you'll ever need. The last alignment program you'll ever need. 2021. Disponível em: <http://www.clustal.org/omega/#Documentation>. Acesso em: 10 jul. 2024.

CRISTIANINI, Nello; HAHN, Matthew W. **Introduction to Computational Genomics: A Case Studies Approach**. 1. ed. Cambridge: Cambridge University Press, 2006.

DOUEK, Daniel C.; ROEDERER, Mario; KOUP, Richard A.. **Emerging Concepts in the Immunopathogenesis of AIDS**. *Annual Review Of Medicine*, [S.L.], v. 60, n. 1, p. 471-484, 1 fev. 2009. Annual Reviews. <http://dx.doi.org/10.1146/annurev.med.60.041807.123549>.

EDDY, Sean R. **What is a hidden Markov model?** *Nature Biotechnology*, v. 22, n. 10, p. 1315–1316, out. 2004.

GRANDE, Fedora; OCCHIUZZI, Maria A.; RIZZUTI, Bruno; IOELE, Giuseppina; DE LUCA, Michele; TUCCI, Paola; SVICHER, Valentina; AQUARO, Stefano; GAROFALO, Antonio. **CCR5/CXCR4 Dual Antagonism for the Improvement of HIV Infection Therapy**. *Molecules*, v. 24, n. 3, p. 550, 2 fev. 2019.

HMMER. 2024. Disponível em: <http://hmmer.org/documentation.html>. Acesso em: 10 jul. 2024.

KUMAR, Vinay; ABBAS, Abdul K.; ASTER, Jon C.. **Robbins Basic Pathology**. 9. ed. Filadélfia: Elsevier, 2012.

LIN, Nina H; KURITZKES, Daniel R. Tropism testing in the clinical management of HIV-1 infection. **Current Opinion In HIV And AIDS**, [S.L.], v. 4, n. 6, p. 481-487, nov. 2009. Ovid Technologies (Wolters Kluwer Health). <http://dx.doi.org/10.1097/coh.0b013e328331b929>.

LLORENTE GARCÍA, Isabel; MARSH, Mark. **A biophysical perspective on receptor-mediated virus entry with a focus on HIV**. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, v. 1862, n. 6, p. 183158, jun. 2020.

NYAMWEYA, Samuel *et al.* **Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis**. 2013. Disponível em: https://www.researchgate.net/profile/Sarah-Rowland-Jones/publication/235741280_Comparing_HIV-1_and_HIV-2_infection_Lessons_for_viral_immunopathogenesis/links/5aeab832aca2725dabb6537e/Comparing-HIV-1-and-HIV-2-infection-Lessons-for-viral-immunopathogenesis.pdf. Acesso em: 09 jul. 2024.

SCHULTZ, Anne-Kathrin; ZHANG, Ming; LEITNER, Thomas; KUIKEN, Carla; KORBER, Bette; MORGENSTERN, Burkhard; STANKE, Mario. **A jumping profile Hidden Markov Model and applications to recombination sites in HIV and HCV genomes**. *BMC Bioinformatics*, v. 7, n. 1, p. 265, dez. 2006.

UNAIDS. **Folha informativa 2020**. Disponível em: https://unaids.org.br/wp-content/uploads/2021/06/2020_11_19_UNAIDS_FactSheet_PORT_Revisada-Final.pdf. Acesso em: 9 jul. 2024.

WEISS, Robin A.. **How Does HIV Cause AIDS?** *Science*, [S.L.], v. 260, n. 5112, p. 1273-1279, 28 maio de 1993. American Association for the Advancement of Science (AAAS). <http://dx.doi.org/10.1126/science.8493571>.

ZAITSOVA, Marina; PEDEN, Keith; GOLDING, Hana. **HIV coreceptors: role of structure, posttranslational modifications, and internalization in viral-cell fusion and as targets for entry inhibitors**. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, v. 1614, n. 1, p. 51–61, jul. 2003.

APÊNDICE A - GitHub

Para uma melhor organização do trabalho, foi criado um repositório no GitHub com informações adicionais do projeto, este repositório contém os arquivos e o script utilizados para o experimento. O link está disponível a seguir:

https://github.com/lcsmuller/CI1169_Bioinformatics

ANEXO A - Database

A new_db do artigo original foi extraída da Database Los Alamos, assim como as sequências utilizadas nos testes. O link está disponível a seguir:

<https://www.hiv.lanl.gov/content/index>