

CAPÍTULO 1

Patogenômica: Explorando os Genomas de Patógenos

Lucas Gabriel Rodrigues Gomes, Eduarda Guimarães Sousa, Fernanda Diniz Prates, Talita Pereira Gomes, Gabriel Camargos Gomes, Janaíne Aparecida de Paula, Pedro Henrique Marques, Ana Lua de Oliveira Vinhal, Bernardo Buhr Alves Mendonça, Aline Ferreira Maciel de Oliveira, Mariana Letícia Costa Pedrosa, Luiza Pereira Reis, Marcus Vinicius Canário Viana, Siomar de Castro Soares, Sandeep Tiwari, Arun Kumar Jaiswal, Vasco Azevedo

<https://doi.org/10.69570/mp.978-65-84548-32-9.c1>

Resumo

Com o advento de desenvolvimentos na área da computação, e com a redução de custo de técnicas para o sequenciamento de genomas, a era do *Big Data* resultou na imensa expansão do nosso repertório de genomas de interesse. Hoje, bancos de dados como o GenBank contém grandes volumes de genomas de patógenos, e mais são depositados a cada dia. A disponibilidade desses dados permite o desenvolvimento de técnicas e ferramentas que servem tanto para estudar os genes que conhecemos quanto para usá-los como referência no estudo de novas descobertas. Através da patogenômica, é possível estudar os genomas de patógenos, identificando genes relacionados com a sua patogênese e entendendo a relação patógeno-hospedeiro. Desse modo, o campo busca guiar novos métodos diagnósticos, profiláticos, e de tratamento, usando o genoma como fundamentação. Neste capítulo, propomos explorar as metodologias e tecnologias usadas em estudos de patogenômica, como: sequenciamento de alto rendimento; genômica comparativa; e ferramentas e bancos de dados bioinformática. Estas ferramentas expandiram a capacidade de pesquisadores para caracterizar, compreender e rastrear espécies patogênicas. Também propomos explicar de modo amplo as aplicações práticas da patogenômica, incluindo: diagnóstico de doenças infecciosas; epidemiologia; e o desenvolvimento de novos métodos terapêuticos e vacinas. Também incluídos são exemplos de casos ilustrando como dados em patogenômica podem ser aplicados para desvendar o comportamento complexo de um patógeno, identificar fatores de virulência, e compreender mecanismos de resistência a antibióticos, bem como os desafios e direções futuras de exploração do campo.

Palavras-chave: Big Data, Patogenômica, Sequenciamento, Genomas, Bioinformática.

1. Introdução à Patogenômica

A genômica constitui um ramo especializado da genética dedicado à investigação do genoma, definido como a totalidade do material genético contido no DNA de um organismo. Esse campo de estudo, de caráter abrangente, inclui não apenas o sequenciamento e a análise dos genomas, mas também a elucidação das interações entre os genes e os fatores ambientais, proporcionando uma compreensão aprofundada dos mecanismos que regem a expressão gênica e a sua modulação em diferentes contextos biológicos (HIH, 2022).

O avanço nas técnicas de sequenciamento genético proporcionou à ciência um acesso sem precedentes a uma vasta gama de informações sobre diversas classes de organismos, incluindo microrganismos patogênicos de relevância médica, ambiental e veterinária. A integração da bioinformática nesse contexto permite que programas computacionais e bases de dados especializadas sejam empregados por pesquisadores, otimizando e ampliando o escopo das investigações científicas (PALLEN et al., 2007).

A patogenômica, campo de estudo dedicado à exploração de múltiplos aspectos da biologia dos patógenos, abrange desde a análise de seu conteúdo genético e suas relações evolutivas até a identificação de genes associados à resistência antimicrobiana e à virulência. Além disso, examina os mecanismos pelos quais esses microrganismos são capazes de infectar e se proliferar no hospedeiro. Entre os principais objetivos da patogenômica destacam-se a profilaxia, o diagnóstico e o monitoramento da evolução das doenças, bem como o desenvolvimento de tratamentos terapêuticos e vacinas com maior eficácia (DEMUTH et al., 2008).

A patogenômica está intrinsecamente vinculada a outras disciplinas, como a genômica funcional, estrutural e a metagenômica, que, quando integradas, fornecem uma visão completa da biologia dos patógenos e permitem a realização dos objetivos desta área de estudo. A genômica estrutural se destaca por sua missão de caracterizar tridimensionalmente as proteínas codificadas pelos genes, dada a estreita correlação entre estrutura e função (Goldsmith-Fischman e Honig, 2003). Essa abordagem não apenas facilita novas

descobertas sobre a dinâmica e a interação dessas proteínas com outras moléculas, mas também potencializa suas aplicações na descoberta e aprimoramento de fármacos. Por outro lado, a metagenômica, ao permitir o estudo das estruturas e funções de microrganismos a partir da análise de uma comunidade completa presente em uma amostra ambiental, oferece uma compreensão mais profunda de como esses genes interagem entre si e desempenham suas funções no contexto específico daquele ambiente (NRC, 2007).

Por fim, a genômica funcional oferece uma análise abrangente das funções de um gene, abrangendo seu papel integral em vias bioquímicas, celulares e fisiológicas (GARCIA, 2006). Ao empregar sequenciamentos genômicos completos e ferramentas bioinformáticas, é possível analisar com maior precisão os mecanismos que levam ao surgimento de fenótipos, permitindo a modificação direcionada das funções gênicas. Essa compreensão é importante para a patogenômica, na medida em que facilita a identificação de novos alvos terapêuticos, o desenvolvimento de novos fármacos, e até a personalização de tratamentos médicos. Dessa forma, é possível prever e otimizar a resposta individual de cada organismo a determinados fármacos, incrementando a segurança e a eficácia das intervenções terapêuticas (IOC, 2024).

Os genomas dos patógenos contêm o material genético que codifica os mecanismos necessários para estabelecer infecções e evadir o sistema imune do hospedeiro. Fatores genéticos chamados de fatores de patogenicidade são responsáveis pela capacidade de diferentes microrganismos a invadir o corpo do hospedeiro, a aderir às suas membranas, a evadir o sistema imune e a resistir a intervenções por antibióticos (IRAOLA e BOYCE, 2022). Sendo cruciais para a patogênese dos microrganismos, a compreensão desses fatores passa a ser essencial na compreensão e tratamento de doenças (IRAOLA e BOYCE, 2022).

Desses fatores de virulência, um conjunto é de particular preocupação global, os fatores de resistência. O desenvolvimento de resistência a antimicrobianos (AMR) em patógenos, frequentemente dado por fatores genéticos como mutações e transferência horizontal de genes, pode representar um risco expressivo do ponto de vista de saúde pública (WHO, 2023). Com isso,

um grande foco tem sido colocado na compreensão de mecanismos de resistência a antimicrobianos.

Os estudos em patogenômica feitos com bactérias do gênero *Leptospira* e com a espécie *Escherichia coli* podem exemplificar o quanto esse campo de estudo tem a contribuir com inovações no campo da saúde. As bactérias *E.coli*, por exemplo, presentes em dejetos, ocasionalmente podem colonizar o trato urinário e causar quadros de infecção urinária, doença que atinge muitos pacientes do sexo feminino, que possuem a uretra mais curta que o sexo masculino. Apesar de muito comum, é uma doença que, se não tratada corretamente, pode levar à morte por choque séptico. O crescente número de sequenciamento genético da *E. coli* tem proporcionado avanços no entendimento dos seus mecanismos de patogenicidade. Sendo um organismo que coloniza naturalmente o trato intestinal, a bactéria precisa de mecanismos adicionais para invadir e estabelecer-se na uretra, facilitando a transmissão entre hospedeiros. Esses mecanismos são denominados "fatores de virulência extraintestinais". A patogenômica tem revelado que essa bactéria pode colonizar o trato urinário através de duas principais vias: a via hematogênica, que envolve a infecção inicial do parênquima renal, e a via ascendente, que começa com a migração da bactéria das fezes para o trato urinário, atingindo eventualmente a bexiga (AGARWAL et al., 2012). O sequenciamento de genomas de linhagens multirresistentes como no caso de *Mycobacterium tuberculosis*, portanto, nos permite estudar a maneira em que esses mecanismos genéticos surgem, guiando o desenvolvimento de novos antibióticos e novos regimes de tratamento (EKTEFAIE et al., 2021).

Outro exemplo relevante são os estudos conduzidos com bactérias do gênero *Treponema*, responsáveis por uma variedade de treponematoses, dentre elas a sífilis, uma IST de importância global. Assim como no caso mencionado anteriormente, o aumento progressivo no sequenciamento de linhagens tem proporcionado descobertas significativas. Dentre essas análises, a realização de estudos de pangenoma focados na patogenômica desse gênero permitiu elucidar alguns dos processos biológicos cruciais dessas espiroquetas (KUMAR et al., 2017). É crucial ressaltar que a manipulação genética de bactérias do gênero *Treponema* enfrenta consideráveis desafios técnicos, devido às grandes

limitações para a manutenção de uma cultura ao longo prazo dos microrganismos (EDMONDSON et al., 2018). Devido a estas dificuldades, a aplicação da patogenômica tem possibilitado avanços no entendimento da patogênese e diversidade de *Treponema*, o que pode fomentar o desenvolvimento de meios de cultura aprimorados para o cultivo dessas bactérias, bem como novas abordagens terapêuticas. Adicionalmente, a patogenômica oferece perspectivas no desenvolvimento de uma vacina contra a sífilis, uma vez que, até o momento, não existe uma candidata viável para uso em humanos (GOMES et al., 2022).

Esses exemplos ilustram o valor substancial que os estudos em patogenômica podem agregar à saúde pública. A integração de dados de diferentes vertentes da genômica permite uma compreensão mais profunda de aspectos cruciais relacionados à biologia e transmissão de patógenos, incluindo suas relações evolutivas, metabolismo, fatores de virulência, e as interações patógeno-hospedeiro. Tais dados são fundamentais para o desenvolvimento de diagnósticos precoces, bem como para o aprimoramento de novas abordagens diagnósticas e terapêuticas. Nos Estados Unidos, os Centros de Controle e Prevenção de Doenças (CDC) já implementaram programas de investimento em sequenciamento *next generation* e patogenômica, com o objetivo de fortalecer o controle de doenças infecciosas. Um exemplo notável é o PulseNet, um banco de dados associado ao CDC, utilizado na detecção e controle de pequenos surtos de intoxicação alimentar, que anteriormente só eram identificáveis em larga escala (CDC, 2024). Outro exemplo está na formulação anual das vacinas contra o vírus da influenza. Duas vezes ao ano, a Organização Mundial da Saúde (OMS) organiza encontros com especialistas que, com base em informações sobre as cepas virais circulantes, indicam os melhores candidatos para a composição das vacinas sazonais, destacando o papel fundamental dos estudos genômicos na prevenção de epidemias globais (ARMSTRONG et al., 2019).

2. Histórico da Patogenômica

2.1. Origens e Desenvolvimento Inicial

O termo patogenômica surgiu por volta dos anos 2000 devido aos avanços do Projeto Genoma Humano. Patogenômica é a denominação utilizada

para a análise genômica de microrganismos patogênicos, ou seja, refere-se a análise do DNA, RNA e proteínas de diferentes patógenos, como bactérias, vírus ou fungos para caracterizar estruturas moleculares e funções, definir seus processos evolutivos e mapear os genomas destes microrganismos (PRADEL et al., 2006). Vale ressaltar que o primeiro genoma completamente sequenciado no mundo foi de um microorganismo, o bacteriófago ϕ x174, sequenciado em 1977 por Sanger et al., concretizando o marco inicial da era genômica (SANGER et al., 1977).

Mais tarde, em 1995, o primeiro patógeno tem seu genoma sequenciado por Fleischmann et al., o qual alcançou o genoma completo da bactéria *Haemophilus influenzae Rd* (Fleischmann et al., 1995). Um ano depois, em 1996, é sequenciado o primeiro genoma completo de um eucarioto que atua como patógeno de baixa virulência e que é de extremo interesse industrial, a levedura *Saccharomyces cerevisiae*, por obra de Goffeau et al (GOFFEAU et al., 1996). Em 2003, o primeiro genoma humano foi sequenciado pelo Projeto Genoma Humano iniciado em 1990 e esta prática recebeu ainda mais atenção (IHGSC, 2024).

Por volta de 1970 foi desenvolvido a primeira técnica de sequenciamento, Sanger, a qual permitiu o sequenciamento do primeiro genoma completo de vírus, o bacteriófago ϕ x174. (LAGE e SANTOS, 2022). Nos anos 2000, os sequenciadores de segunda geração (NGS) emergem revolucionando o cenário da genômica ao permitir o sequenciamento simultâneo de milhares de fragmentos de DNA com custo e tempo reduzido (SOUZA e BRUSAMARELLO, 2009).

O sequenciamento de terceira e quarta geração surge pouco depois e inclui novas técnicas de sequenciamento como Single Molecule Real-Time (SMRT) desenvolvida pela Pacific Biosystems e Nanopores proporcionando o sequenciamento de moléculas únicas capaz de realizar leituras mais longas, caracterizando um processo com melhor resolução e precisão (POLLARD et al., 2018). Essas tecnologias de sequenciamento, fornecem insights da expressão e regulação genética por análises de transcriptoma que permite a detecção de variantes splicing e outras modificações pós-traducionais essenciais para

desenvolver tratamentos personalizados a fim de melhorar sua eficácia, efetividade e segurança (STARK et al., 2019; SAMEEK e CHINNAIYAN, 2014).

O desenvolvimento do campo ao longo dos anos foi marcado por uma série de passos que nos permitiram expandir nossa compreensão e repertório de genomas e proteomas conhecidos. A **Figura 1** mostra uma linha do tempo destes avanços.



Figura 1. Linha do tempo com os principais pioneiros e que marcaram o início da patogenômica, marco da integração entre computação e biologia, principais descobertas que impulsionaram esse campo da bioinformática.

2.2. Marcos Históricos

Ao longo dos séculos, o *Mycobacterium tuberculosis* tem constituído uma ameaça significativa para a saúde mundial. Esta doença tem sido associada aos seres humanos há cerca de 70.000 anos (COMAS et al., 2013). A informação

anterior não surgiu ao acaso, e sim como resultado de uma abundância de estudos sobre genomas, bioinformática, filogenética e arqueologia. Nesta seção, o foco será o sequenciamento do genoma do *M. tuberculosis* e do *Plasmodium falciparum* e seu contexto evolutivo (BAÑUS et al., 2015; COLE et al., 1998).

A primeira estirpe de *Mycobacterium tuberculosis*, denominada H37Rv, foi obtida em 1998. Tal como descrito por S. T. Cole et al. sobre a investigação, uma grande parte do genoma foi obtida através dos cosmídeos, aproximadamente 3,2 Mb, enquanto a parte restante foi obtida utilizando 45.000 clones shotgun de genoma completo e clones BACs (Bacterial Artificial Chromosome). Os clones foram criados usando vectores M13, para os cosmídeos e BACs, e pUC18 para o DNA genómico, procurando uma melhor cobertura para a montagem, usando ambas as vertentes, “reverse” e “forward”, as regiões propensas à seleção e deleção foram obtidas por reação em cadeia da polimerase (PCR) e a partir de BACs. No final destes processos foram mostradas trinta diferenças entre as sequências shotgun e cosmídeos, resultantes de erros de sequenciação e mutações de cosmídeos. É também necessário contrastar os recursos computacionais utilizados ao longo do projeto, como o PHRAP, GAP4, um script perl, DIANA, TB-parse, EMBL, TREMBL, SwissProt, PROSITE, a sua base de dados, tRNAscan e tRNAscan-SE (COLE et al., 1998).

A malária é outra doença que tem causado muitos desafios aos seres humanos. Sendo um agente patogênico, o genoma do *Plasmodium falciparum* tem sido estudado há muitos anos (GARRIDO-CARDENAS et al., 2019). O primeiro genoma completo deste microrganismo foi sequenciado em 2002, e este é outro marco que será discutido nesta seção. A principal estratégia utilizada pelos investigadores, Malcolm J. Gardner et al, foi um shotgun de cromossomos inteiros, estes que foram separados em géis de campo pulsado, posteriormente o genoma cromossômico foi extraído para a construção de bibliotecas shotgun, um pormenor importante é o shotgun partilhado dos cromossomas 6, 7 e 8, devido resultados insatisfatórios em gel. Por fim foi necessário trabalhar com o fecho de lacunas, quase totalmente feito com Sequence tagged sites (STSs), marcadores microssatélites e mapeamento HAPPY, e alguns, entre 3 e 37, pequenos gaps ficaram nos limites do artigo (GARDNER et al., 2002).

O Projeto Genoma Humano, forneceu um avanço tecnológico e científico essencial para a era genômica a partir do desenvolvimento de tecnologias avançadas de sequenciamento de DNA, mais rápidas e acessíveis, as quais passaram a ser utilizadas para sequenciar o genoma de patógenos (WONG, 2018). Além disso, com o mapeamento de genes preciso e completo do DNA humano, foi possível identificar genes responsáveis por doenças através da comparação daqueles presentes DNA humano com os presentes em DNA de microrganismos (MSHG, 1988).

Vale salientar que compreender como o metabolismo e os mecanismos de defesa do organismo humano interagem com os agentes patogênicos é um forte contribuinte para a identificação de novos alvos terapêuticos e tratamentos mais efetivos e seguros (COVID-19 HOST GENETICS INITIATIVE, 2021). A descoberta de todos estes caminhos para a compreensão do combate a estes microrganismos permitiu maior rastreo e controle da disseminação de patógenos, bem como, de maiores precauções contra endemias e pandemias, tendo em vista o evento pandêmico mais recente, no início de 2020, a pandemia do SARSCoV-2 (KOMISSAROV et al., 2020; LI et al., 2021).

2.3. Evolução e Expansão do Campo

Por volta da década de 1960, a bioinformática emergiu como uma disciplina essencial para coordenar a enorme quantidade de dados biológicos gerados (PEREIRA et al., 2020). Essa integração entre biologia, informática, computação e estatística representa uma revolução na pesquisa biológica e médica, principalmente, com o surgimento de inúmeras ferramentas avançadas em análise genômica (SILVA e ALVES, 2023).

Dessa forma, a integração da bioinformática com a análise de dados apresenta uma importância técnico-científica e social significativa, pois como descrito, permite análises de grandes volumes de dados, facilita a identificação das variações bioquímicas e genéticas em organismos vivos promovendo um avanço na área das ômicas (COSTA, 2014).

Os diferentes marcos relativos às técnicas de sequenciamento deram origem à forma como diferentes metodologias são agrupadas, servindo como

uma associação autodescritiva, cronologicamente falando. Os métodos, atualmente, são divididos em 3 grandes gerações que revolucionaram o estudo de genomas, aprimorando e promovendo diferentes possibilidades de pesquisa (HU et al., 2021).

Inicialmente faz-se necessária a contextualização da era genômica que antecede a primeira geração de sequenciadores. Quando, na década de 1950, se tornou possível determinar os métodos de análise de proteínas emergentes, campo de estudo predecessor ao estudo de sequências genômicas. Inicialmente, isso foi feito de forma analógica e anos depois, com o trabalho de Margaret Dayhoff e Robert S. Ledley no COMPROTEIN, um programa foi escrito em FORTRAN para determinar sequências primárias de proteínas (PAREEK et al., 2011). Após alguns anos, sucederam grandes conquistas neste campo, e em 1977 foi divulgado o primeiro método de sequenciamento por Maxam e Walter Gilbert, com metodologia associada à degradação do DNA, uso de químicos perigosos e radiação, seguido pelo método de Frederick Sanger, associado à terminação de cadeia análogas à dideoxinucleotídeos, técnica utilizada para o sequenciamento do primeiro genoma, o bacteriófago Φ X174 (5386bp), coroando o surgimento da primeira geração de sequenciamento genômico. Até o início da segunda geração surgiram algumas outras metodologias de sequenciamento, bem como a primeira abordagem automatizada com fluorescência (GAUTHIER et al., 2018; PAREEK et al., 2011).

Marcadas pelo ano de 2004, tecnologias de segunda geração foram instauradas (HU et al., 2021), através do desenvolvimento do método Illumina pela Solexa, associadas à métodos de multiprocessamento de leituras, responsável pelo crescimento significativo da velocidade de sequenciamento, tecnologias baseadas em processamento de leituras curtas, a necessidade de criação de bibliotecas e nível de confiabilidade muito alto, sendo protagonizados pelos métodos Illumina, metodologia de fluorescência, e Ion Torrent, metodologia iônica (HU et al., 2021; RODRIGUEZ e KRISHNAN, 2023), que serão aprofundados posteriormente.

Em sequência, no ano de 2011 foi inaugurada a terceira e última geração de sequenciadores onde os pioneiros e protagonistas são o Pacific Biosciences (PacBio), *single-molecule real-time* (SMRT), e o Oxford Nanopore Technologies

(ONT), nanopores e corrente elétrica (XIAO e ZHOU, 2020). Esses métodos trazem como características comuns a utilização de leituras longas, maior eficiência ao trabalhar com genomas com sequências repetitivas, além da possibilidade de sequenciamento de materiais completos (SMS) e da redução da necessidade de multiplicação genômica (XIAO e ZHOU, 2020; ATHANASOPOULOU et al., 2021).

Por fim, vale ressaltar que nenhum desses métodos são uma abordagem perfeita. Isso se deve a complexidade de montagem que envolve sequenciamentos de segunda geração, ainda mais agravado quando o assunto são genomas mais complexos; a necessidade de uma biblioteca mais robusta, comparado com os métodos de terceira geração; e ao menor nível de confiabilidade nas leituras geradas por métodos de terceira geração (XIAO e ZHOU, 2020; ATHANASOPOULOU et al., 2021; WANG et al., 2021).

A análise de dados provenientes da genômica possui como principais objetivos a caracterização e estudo de determinado genoma para que suas características sejam melhor compreendidas, indicando potenciais favoráveis ou formas de tratamento de doenças causadas, no caso de patógenos (CHIVIAN et al., 2023). Nesse momento iremos abordar vias oportunas que envolvem as ciências de dados como instrumento para a análise de genomas.

O uso de análise computacional de dados, através de bancos de dados robustos, como GenBank, trouxe e trazem diversas transições positivas para a área, possibilitando a origem de algoritmos de análise comparativa, que serão melhor trabalhados posteriormente, análise de contaminação e a criação de tecnologias de Inteligência Artificial (IA) cada vez mais precisas e eficientes, decorrente da maior disponibilidade de dados e desenvolvimento do próprio campo de estudo. Dentre as técnicas envolvendo IA podemos destacar as técnicas de aprendizado profundo, que através de diferentes algoritmos de redes neurais artificiais, Redes Neurais Recorrentes (RNNs), Redes Neurais Convolucionais (CNNs) e algoritmo de atenção, permitem o avanço da área de pesquisa *in silico* nesse sentido, apesar da limitação muitas vezes imposta pela exigência de hardwares poderosos para trabalhar com tais algoritmos (SAYERS et al., 2023; CHOI e LEE, 2023).

Na saúde global, todas essas técnicas de sequenciamento e métodos de análises mais sofisticados originaram o conceito de vigilância genômica, definido como o monitoramento da evolução, disseminação, mecanismos de resistência e até mesmo a predição de cepas circulantes mais prevalentes ou incidentes em determinado período de tempo em um território ou em todo o mundo. Um exemplo positivo disso é o projeto *Global Influenza Surveillance and Response System* (GISRS) da OMS para rastreamento do vírus da influenza (HAY e MCCAULEY, 2018).

A relevância disso é evidente quando observamos os casos de fungos e protozoários que vem demonstrado uma crescente resistência medicamentosa nas últimas décadas (STRUELENS et al., 2024). O fungo *Candida auris*, por exemplo, vem provocando surtos em hospitais por sua resistência a múltiplos medicamentos (FORSBERG et al., 2019) e *Aspergillus* oferece grande risco de infecções graves em pacientes imunocomprometidos (CADENA et al., 2021). Protozoários como *Trichomonas vaginalis* ou *T. foetus* deixados em segundo plano, tanto que o mecanismo que desenvolveram de resistência a antimicrobianos ainda não é totalmente compreendido, além do pouco conhecimento que adquirimos quanto a seus processos metabólicos e algumas de suas estruturas singulares (KIRKCALDY et al., 2012).

Em complemento podemos descrever diversos outros momentos históricos em que a presença dos conhecimentos de bioinformática contribuíram para o entendimento de patógenos não bacterianos em momentos que vários destes configuraram entre novas doenças emergentes, sendo destaque no campo de identificação, Emerging Pathogen Identification (EPI), trazendo contribuições relativas à eficiência de tratamento, vacinas e até na identificação da origem do organismo, podem ser citados como exemplo o SARS-CoV-2, vírus que causou a última pandemia, a partir do ano de 2019, ou a epidemia de Ebola no oeste do continente africano (VASHISHT et al., 2023), através da prática de análise de material genético, a maior fonte de informações de organismos vivos, os pesquisadores identificam genes responsáveis por processos fundamentais do organismo do patógeno, incluindo a os genes de resistência e através destas técnicas computacionais mais aprimoradas e

robustas se tornou possível direcionar estes genes de maneira mais eficaz com novos fármacos (ENGELTHALER, 2024).

O estudo do genoma de patógenos, evidentemente, reinventou a patologia e a farmacologia, diante disto, é indiscutível que dispor da troca de conhecimentos, recursos e tecnologias de diferentes países é sem dúvidas o fator de impulsão do progresso científico e de inovação pela promoção de grandes volumes de informações na íntegra e de diversidade das mesmas (LINS et al, 2021). Um exemplo claro disso é a Iniciativa Global de Resistência Antimicrobiana (GARDP) divulgada em 2016, que envolve a participação da Organização Mundial da Saúde (OMS) e Iniciativa de Medicamentos para Doenças Negligenciadas (DNDi), como também, dos governos da África do Sul, Alemanha, Japão e Reino Unido.

GARDP analisa e classifica os patógenos, através da genômica, quanto a maior necessidade da criação de novas terapêuticas e menor probabilidade de serem alvos de pesquisa de terceiros. E para isso, esse plano de ação que exige parceria entre público e privado, a fim de incentivar o desenvolvimento de novos tratamentos e diagnósticos para combate a resistência antimicrobiana, um dos maiores problemas de saúde pública da atualidade (PIDDOCK, 2019).

Outros projetos de grande importância foram e são desenvolvidos, no entanto, a pandemia do SARS-CoV-2 se tornou o exemplo mais recente e que destacou ainda mais a importância da união de esforços, não somente no campo da patogenômica, mas de toda a ciência, devido magnitude alcançada por este sucesso. Este evento, se converteu em um marco histórico e em uma grande lição de como a conexão das informações impactou e ainda impacta imensamente, a velocidade de resposta do mundo moderno a episódios infortúnios, inesperados e que resultam em consequências destrutivas para a humanidade (SILVA e LIMA, 2021).

3. Áreas de Pesquisa em Patogenômica

3.1. Análise de Genomas de Patógenos

3.1.1. Sequenciamento e Montagem de Genomas

Ao longo do desenvolvimento do campo *in-silico* diversas tecnologias já foram utilizadas para conseguir determinar a sequência de certo fragmento de material genético. É possível dividir as tecnologias de sequenciamento em 2 grupos amplos: leituras pequenas, as quais surgiram na segunda geração, e leituras longas, que surgiram na terceira geração (SATAM et al., 2023). A seguir serão trabalhadas algumas nuances relativas aos métodos de maior destaque. Porém antes de explorar com mais detalhes os processos de sequenciamento é necessário ressaltar e destacar a importância do controle de qualidade na preparação de amostras e bibliotecas de sequenciamento (LARSON et al., 2023; HESS et al., 2020).

As tecnologias de segunda geração (SGS) trouxeram maior dinamicidade, logo maior capacidade de produção, comparado aos seus predecessores, tendo como características principais o uso de leituras pequenas, grande precisão de sequenciamento e sequenciamento paralelo, apesar da complexidade considerável no momento de montagem, devido às leituras, que possuem tamanho reduzido. Os nomes usados para ilustrar a segunda geração são Illumina e Ion Torrent (HU et al., 2021; SATAM et al., 2023).

Já os métodos de terceira geração (TGS), que trabalham com leituras longas, iremos destacar a tecnologia PacBio e Oxford Nanopore. Ambos os métodos oferecem como vantagem a capacidade de montagem mais simples, quando comparado aos SGS, dado o número menor de leituras a serem arranjadas, que, também, por serem mais longas preservam consideráveis sequências de nucleotídeos, apesar da precisão reduzida em comparação aos métodos das outras gerações. Além disso, os TGS possuem um ciclo de sequenciamento mais veloz (HU et al., 2021; SATAM et al., 2023).

Vale ressaltar, também, a possibilidade de abordagens híbridas de segunda e terceira gerações, que apesar de proporcionar um encarecimento do processo, combinam as características positivas de ambas, logo atenuando os pontos fracos da outra. Por exemplo podemos citar o eficiente papel que

sequenciamentos de terceira geração tem em contribuir no fechamento de gaps em sequências de segunda geração montadas, estas que, seguindo outra metodologia, ajudam a fornecer maior confiabilidade e corrigir potenciais erros de sequenciamentos de terceira geração (WANG et al., 2021).

A análise de dados gerados é feita com o uso de softwares apropriados para tratamento e curadoria. Na primeira etapa da análise de *short reads* é feito uma etapa digital de controle de qualidade e chamada de base, onde se infere a sequência de bases das leituras (CACHO et al., 2016), já na segunda é feito o alinhamento das sequências com o uso de um algoritmo de tabela *hash* e arquivo de formato BAM, para armazenamento dos resultados. Por fim, na terceira etapa é realizado um processo de chamada de variantes, buscando identificar mutações baseadas em algum genoma de referência (HU et al., 2021; LARSON et al, 2023; XU, 2018).

Após o processo bem sucedido de sequenciamento chegamos à etapa de montagem do genoma, que pode ser sintetizada como a tentativa de ligação das “peças”, as leituras, para encontrar a ordem correta delas e assim possuir os genomas completos (SEGERMAN, 2020). As metodologias para realização de montagem são *de novo*, baseado em sobreposição de leituras e sistemas de probabilidade, de referência, que utilizam um genoma semelhante como referência para o processo, e via transcriptoma, onde o RNAseq da espécie é a ferramenta de suporte para montagem (ESPINOSA et al., 2024; ZHANG et al., 2022). Alguns dos montadores de destaque ao longo do século são: Celera assembler(2000), *De novo*(Newbler)(2005), Velvet(2008), ALLPATHS(2008), SPAdes(2012), HGAP(2013), Skesa(2018) e Unicycler(2020) (SEGERMAN, 2020).

Quando tratamos da montagem de genomas mais complexos vemos que, por exemplo no caso de plantas, o tamanho do genoma, as diversas sequências com mesmas bases ao longo do material genético e genomas heterozigotos nas espécies são alguns fatores que crescem na complexidade da montagem (68). A grande demanda computacional do processo, além do fechamento de gaps e erros de montagem são desafios (ESPINOSA et al., 2024).

Os sequenciadores de terceira geração, ao passo que suas técnicas de *reads* longos trazem um atenuação para o problema das sequências curtas,

montagem dificultada pelo tamanho do genoma e áreas de sequências idênticas (67,69). Outra estratégia utilizada se faz a partir de uma progressão na tecnologia dos montadores para trabalhar diferentes tipos de genomas complexos e se complementarem no resultado final (ZHANG et al., 2022).

3.1.2. Análise Comparativa de Genomas

A genômica comparativa tem como finalidade mapear, comparar e analisar a estrutura e as regiões genéticas funcionais de genomas individuais de uma única unidade taxonômica, identificando assim as similaridades e diferenças entre elas (ABBY e DAUBIN, 2007). As análises destinadas à comparação de genomas têm evidenciado a existência de uma enorme diversidade genética mesmo entre linhagens de uma mesma espécie, além de uma grande complexidade na estrutura e organização desses genomas, o que tem levado a questionamentos importantes que contribuem para uma melhor compreensão acerca dos mecanismos de evolução (PRASAD et al., 2023).

Grande parte das principais ferramentas utilizadas na comparação de sequências de DNA, RNA ou proteína se baseiam em métodos clássicos de análise computacional, tais como, algoritmos de alinhamento local e global, e algoritmos de alinhamento simples e de múltiplas sequências.

No alinhamento local a comparação entre as sequências ocorre através de pequenas regiões (SMITH e WATERMAN, 1981). Os algoritmos de alinhamento local geralmente são muito utilizados para encontrar semelhanças entre regiões que podem estar relacionadas, mas que são invertidas ou reorganizadas umas em relação às outras (BRAY et al., 2003). Já o global ocorre quando os genomas são comparados levando em consideração toda a sua extensão, sendo recomendado para sequências similares e que apresentam tamanho aproximado (CHOWDHURY e GARAI, 2017).

Já os alinhamentos simples e de múltiplas sequências dizem respeito ao número de sequências que são comparadas durante o alinhamento. Enquanto o alinhamento simples compara apenas duas sequências entre si, o alinhamento múltiplo compara várias sequências ao mesmo tempo. A partir do resultado do alinhamento de múltiplas sequências, pode-se inferir a homologia e as relações

evolutivas entre as sequências analisadas (PAIS et al., 2014). Por outro lado, as ferramentas de alinhamentos simples são usadas para identificar regiões de similaridade, as quais podem indicar relações funcionais, estruturais e/ou evolutivas entre duas sequências biológicas (MULLAN, 2006).

Abaixo algumas ferramentas de bioinformática amplamente utilizadas para comparação genômica que se baseiam nos algoritmos citados acima:

- **BLAST (Basic Local Alignment Search Tool):** é uma das ferramentas de análise de sequências biológicas de domínio público mais utilizadas hoje (ALTSCHUL et al., 1990).
- **Clustal Ômega:** é a ferramenta mais recente dentro da família Clustal e foi desenvolvida com a finalidade de permitir o alinhamento de uma grande quantidade de sequências sem que houvesse um comprometimento nos resultados dos alinhamentos (SIEVERS et al., 2011).
- **MAFFT:** é uma ferramenta muito utilizada para realizar o alinhamento de múltiplas sequências, essencial em estudos de filogenia, estrutura de proteínas e análise funcional de genes, desempenhando um papel essencial na compreensão da biologia molecular (KATOH, 2002).

A análise comparativa de genomas tem sido muito utilizada na identificação de genes e regiões conservadas entre os genomas comparados visando compreender melhor a evolução e a relação filogenética entre os microrganismos, assim como no desenvolvimento de novas drogas e alvos terapêuticos (JUHAS et al. 2012). Os genes e regiões conservadas são aquelas que persistem no genoma de várias linhagens de um grupo e que apresentam mutações mínimas ao longo do tempo. Geralmente codificam proteínas que desempenham papéis significativos em vários processos celulares críticos e que são essenciais para a sobrevivência de um organismo, tais como transcrição e tradução do DNA, e proteínas envolvidas na respiração celular (ROSCONI et al., 2022).

Uma técnica amplamente utilizada e que se baseia na amplificação, sequenciamento e análise de genes conservados entre isolados de uma mesma

espécie é a tipagem molecular MLST, onde isolados são tipados a partir da sequência dos seus genes *housekeeping*. (Multilocus sequence typing) (MAIDEN et al. 1998). Essa metodologia é adequada para analisar a distribuição geográfica e temporal dos STs circulantes, clonalidade, assim como para obter um panorama sobre a epidemiologia e diversidade genética, auxiliando na investigação de casos esporádicos e surtos de várias doenças (IBARZ e MAIDEN., 2009).

A identificação de genes conservados entre as linhagens de uma mesma espécie também é uma das etapas para a análise do pangenoma, que é considerado atualmente o padrão ouro para a comparação de genomas e estudos de evolução e diversidade. O pangenoma corresponde ao repertório completo de genes observados nos genomas de organismos de uma determinada espécie ou grupo taxonômico (ROULI et al, 2015). Esse conjunto de genes é constituído pelo core genome, shared e singleton, onde o core é compartilhado por todos, o compartilhado é presente em múltiplas, mas não todas, linhagens, e o singleton é exclusivo a linhagens específicas (TETTELIN et al., 2005). Identificar quais genes são conservados e compartilhados por todas as linhagens é parte essencial para a escolha de genes adequados ou proteínas candidatas para o desenvolvimento de vacinas contra várias doenças infecciosas através da abordagem da vacinologia reversa (D'MELLO et al., 2019).

A plasticidade da molécula de DNA possibilita a variabilidade genética e a diversidade, que são fundamentais para o processo evolutivo das espécies e adaptação ao ambiente / hospedeiro (SELA et al., 2018). Uma vez que as características da molécula do DNA e seu processo de replicação contribuem para a continuidade da informação, outros mecanismos atuam para gerar variabilidade genética. O primeiro mecanismo engloba as estratégias que provocam alterações locais na sequência do DNA, tais como substituição, deleção ou inserção de alguns nucleotídeos. O segundo mecanismo envolve estratégias que promovem o rearranjo de sequências de DNA, enquanto o terceiro mecanismo baseia-se na abordagem da aquisição de DNA por transferência horizontal (ARBER, 2000). A **Tabela 1** tem exemplos de ferramentas capazes de identificar variantes nos genomas.

Tabela 1. Ferramentas para identificação de regiões conservadas e variantes genéticas

Ferramenta	Aplicação
VariantDetective (VD) (CHARRON e KANG et al., 2024)	Identificação de polimorfismos de nucleotídeo único (SNPs) e variantes estruturais (SVs)
HGTphyloDetect (YUAN et al., 2023)	Identificação e análises filogenéticas de genes obtidos por transferência horizontal.
progressiveMauve (DARLING et al., 2010)	Identificação e alinhamento de regiões conservadas do genoma na presença de eventos de rearranjos e transferência horizontal

3.1.3. Estudo de Variantes Genéticas

As variantes genéticas desempenham um papel fundamental na evolução, especialmente em microrganismos, onde mutações e recombinações promovem a variabilidade genética, mesmo com a reprodução assexuada. Essas variantes podem ser mutações pontuais, como polimorfismos de nucleotídeo único (SNPs), deleções, inserções de nucleotídeos, ou rearranjos de grandes segmentos gênicos e cromossômicos (ARENAS et al., 2018).

Uma mutação pontual refere-se a uma alteração permanente em uma única base do DNA de um gene, ou à deleção ou inserção de uma nova base. Quando uma mutação em um gene resulta em um códon que codifica o mesmo aminoácido, ela é chamada de mutação sinônima. Já quando o códon é alterado para codificar um aminoácido diferente, trata-se de uma mutação não sinônima, que pode modificar a estrutura ou função da proteína, influenciando diretamente a evolução do microrganismo (CHAHAL et al., 2019).

Por outro lado, a inserção ou deleção de bases (indels) pode alterar a sequência de códons e a fase de leitura do gene, o que pode modificar drasticamente a estrutura e função de uma proteína. Em alguns casos, isso pode gerar um códon de parada prematuro, resultando em uma proteína truncada, ou seja, interrompida precocemente durante sua síntese (CHEN et al., 2009).

As mutações geralmente surgem a partir de erros no processo de replicação do DNA, os quais podem ou não ser corrigidos pelos mecanismos de reparo da célula. Dessa forma, as taxas de mutação variam significativamente entre diferentes microrganismos (CHATTERJEE e WALKER, 2017). Embora mutações que afetam a sequência de aminoácidos muitas vezes possam ser prejudiciais, ocasionalmente elas proporcionam vantagens adaptativas, como maior capacidade reprodutiva ou maior sobrevivência em ambientes hostis (YOSHIDA et al., 1990).

Eventos de recombinação, como a transferência horizontal de genes em bactérias (por transdução, conjugação ou transformação), também impactam de forma significativa a variabilidade genética, ao gerar novas combinações ou reorganizar variações existentes que surgiram por mutações. Além disso, há as CNVs (variações no número de cópias), que são variantes estruturais desequilibradas caracterizadas pela perda ou ganho de fragmentos de DNA que variam de cerca de 50 pares de bases até cromossomos inteiros. (WU et al., 2020; TRALAMAZZA et al., 2024).

Um dos principais métodos para identificação de variantes é a chamada de variantes (do inglês, Variant Calling, VC). A chamada de variantes é o processo de comparar duas amostras de genoma para identificar suas diferenças. Geralmente, essas diferenças são restritas a polimorfismos de nucleotídeos únicos (SNPs) e pequenas inserções e deleções (indels). Alterações estruturais mais amplas, como inversões, duplicações e grandes deleções, normalmente não são incluídas na análise de chamada de variantes. Um dos métodos comumente usados é a ferramenta Snippy, disponível publicamente (<https://github.com/tseemann/snippy>), que compara dados de sequenciamento genômico contra um genoma de referência para identificação dos SNPs e Indels. Por esse motivo, a escolha de genomas de referência montados corretamente é essencial, pois todo o resultado da chamada de variantes será em relação ao genoma de referência.

Existem ainda ferramentas mais robustas, que implementam a chamada de variantes com base em diferentes métodos. As ferramentas atuais se dispõem em três categorias principais. A ferramenta bcftools (disponível em <https://github.com/samtools/bcftools>) por exemplo, realiza por meio de uma

estimativa de frequência de alelos com base em leituras mapeadas. A DeepVariant (POPLIN et al., 2018) usa um modelo de rede neural convolucional treinado em leituras mapeadas. E DiscoSNP (disponível em <https://github.com/GATB/DiscoSnp>) utiliza um método baseado em grafos de Bruijn de k-mers de leituras brutas. Demais ferramentas se baseiam em mudanças a partir desses métodos (SEAH et al., 2023). Esses métodos, por exemplo, são eficazes para chamadas de variantes em genomas bacterianos. Para identificar variações no número de cópias (CNV), há ainda para genomas bacterianos a ferramenta CNV-BAC (WU et al., 2020).

Em resposta à pandemia da COVID-19, muitas ferramentas surgiram para a chamada de variantes em genomas virais. A ferramenta ViReflow (MOSHIRI et al. 2022) por exemplo, é um pipeline de reconstrução e análise de sequência de consenso viral. O ViReflow foi criado em resposta à pandemia de COVID-19, mas pode ser aplicado a qualquer patógeno viral (LI, 2011; WILM et al., 2012).

Variantes genéticas, como SNPs, Indels e CNVs, presentes em regiões genômicas associadas à virulência ou resistência ao ambiente, estresse e antibióticos, no caso de bactérias, podem conferir vantagens ou desvantagens na adaptação e interação com o hospedeiro. Um exemplo recente vem da Índia (105), onde três linhagens de diferentes espécies de *Leptospira* spp. foram sequenciadas: *Leptospira interrogans* sorovar Pyrogenes cepa Salinem, *Leptospira santarosai* sorovar Huanuco cepa M-4, e *Leptospira kirschneri* sorovar Ratnapura cepa Wumalasena.

Além do sequenciamento, o estudo também incluiu a montagem de novo, a anotação estrutural e funcional, além da análise de variantes. Dentre as encontradas, foram relatadas variantes em proteínas que desempenham várias funções importantes em bactérias, como invasão celular, ancoragem à membrana e interação com o hospedeiro. A presença de um grande número de SNPs em genes associados à virulência indica que isso pode ser um dos fatores que contribuíram para o surgimento de linhagens altamente patogênicas. No entanto, são necessários estudos complementares para validar essa correlação (LATA et al., 2020).

Tratando-se de resistência a antibióticos, alguns estudos já aplicaram a chamada de variantes a genes de resistência com sucesso (NAIDENOV et al.,

2019). Por exemplo, a *Piscirickettsia salmonis*, patógeno intracelular Gram-negativo causador da Piscirickettsiose, é tratada frequentemente com altas doses de antibióticos. Análises genômicas *in silico* identificaram sete genes relacionados à resistência ao florfenicol e oxitetraciclina. Foram encontradas cepas com diferentes níveis de suscetibilidade aos antibióticos, e análises *in silico* revelaram que mudanças estruturais nas proteínas associadas à resistência, como OmpF, estavam relacionadas à presença de SNPs (FIGUEROA et al., 2019).

3.2. Análise de Interações Patógeno-Hospedeiro

O termo Interação Hospedeiro-Patógeno (IHPs) refere-se a abordagens diretamente relacionadas à forma como um determinado patógeno interage com seu organismo hospedeiro. Existem diferentes tipos de patógenos, como bactérias, fungos, vírus e outros tipos de organismos que, ao conseguirem se aderir ao sistema do hospedeiro, realizam diversas alternativas para a autopromoção da sobrevivência e da infecção, causando doenças em seu hospedeiro devido às suas características patogênicas. Nesta seção, analisaremos os principais aspectos comumente presentes nessas condições.

As IHPs podem ser descritas em níveis: populacional (como quando ocorreu a pandemia de Covid, em que o vírus Sars-Cov-2 causou infecção em uma população (WHO, 2024), organismal (diferentes patógenos infectando o mesmo hospedeiro) e/ou molecular (quando as proteínas do patógeno interagem com as proteínas do hospedeiro, com receptores presentes nas células humanas) (SEM et al., 2016). Além disso, os IHPs têm “estágios” que são respeitados entre os vários patógenos e hospedeiros possíveis, em ordem: invasão do patógeno no hospedeiro ao superar as barreiras primárias; evasão das defesas imunológicas do hospedeiro pelo patógeno; 3. replicação do patógeno e 4. capacidade imunológica do hospedeiro contra o patógeno (SEM et al., 2016). Os meios pelos quais os patógenos realizam esses feitos estão intimamente associados aos seus fatores de virulência.

3.2.1. Identificação de Fatores de Virulência

Fatores de virulência são moléculas específicas, na sua grande maioria representadas por proteínas produzidas e secretadas por bactérias, fungos, protozoários e vírus. Em patógenos bacterianos, os fatores de virulência são codificados por genes específicos que podem estar localizados no cromossomo ou em elementos genéticos móveis, tais como, plasmídeos e transposons (LEITÃO, 2020).

Os fatores de virulência desempenham diversos papéis importantes no estabelecimento de doenças, incluindo a colonização do hospedeiro, imunossupressão, aquisição de nutrientes da célula hospedeira e detecção de mudanças no ambiente (ALSPAUGH, 2015). Várias moléculas estão envolvidas neste cenário de interação patógeno-hospedeiro, englobando adesinas, toxinas, enzimas e cápsulas (CHEN, 2004).

Através da análise da sequência genômica de diversos tipos de patógenos, podemos obter uma melhor compreensão da interação patógeno-hospedeiro. Com isso, a identificação e caracterização precisa de genes e proteínas de virulência tem se tornado cada vez mais essencial na busca por novos alvos terapêuticos, para o desenvolvimento de testes mais específicos e direcionados, no desenvolvimento de vacinas ou para tratamentos que possam ser utilizados como alternativa aos antimicrobianos (OLI et al., 2020).

Com isso, várias ferramentas têm sido desenvolvidas para identificar com melhor precisão fatores de virulências que possam ser cruciais para a patogenicidade. Abaixo algumas dessas ferramentas:

- **Vfanalyzer:** O VFanalyzer constroi grupos ortólogos dentro do genoma de consulta e genomas de referência pré-analisados do banco de dados de fatores de virulência (VFDB) e, em seguida, conduz pesquisas de similaridade de sequência no VFDB para identificar com precisão potenciais fatores de virulência (CHEN, 2004).
- **VirulentPred 2.0:** O modelo de predição utilizado por este servidor web é baseado no treinamento de modelos de aprendizado de máquina com sequências de proteínas virulentas validadas experimentalmente (SHARMA et al., 2023).

Os fatores de virulência aumentam a habilidade de um patógeno em causar doença. De acordo com os mecanismos e suas funções, os fatores de virulência podem ser divididos nas seguintes categorias [Tabela 2]:

- (1) Adesinas: se ligam especificamente a receptores complementares encontrados na superfície de certos tipos celulares, promovendo assim a adesão da bactéria aos tecidos do hospedeiro (KLINE et al., 2009).
- (2) Evasinas: auxiliam as bactérias no processo de escape da fagocitose ou do sistema imune do hospedeiro (CHANDRA et al., 2022).
- (3) Invasinas: rompem as membranas do hospedeiro e estimulam a endocitose. (GARCIAL DEL PORTILLO e FINLAY, 1994).
- (4) Toxinas que provocam danos às células e tecidos do hospedeiro (DONG et al., 2019)
- (5) Polissacarídeos capsulares que protegem os patógenos das defesas do hospedeiro (PATON et al., 2019).
- (6) Sideróforos: secretados por algumas bactérias para a obtenção de ferro, necessário para o seu crescimento (ELLERMAN e ARTHUR, 2017).

Tabela 2. Exemplos de fatores de virulência em patógenos conhecidos

Espécie	Fator de virulência
<i>Escherichia coli</i> enteropatogênica	Proteínas intimina e seu receptor Tir: adesão à superfície do epitélio intestinal (PAKBIN et al., 2021)
<i>Staphylococcus aureus</i>	Toxinas: altamente inflamatórias e que podem causar morte de células leucocitárias por citólise e deleção clonal, respectivamente (TAM e TORRES, 2019).
<i>Neisseria gonorrhoeae</i>	Fimbrias: adesão à mucosa urogenital (TRAMONT, 2008). Quando as fimbrias estão ausentes, devido à mutação por exemplo, a colonização muitas vezes não ocorre, e a doença não evolui.
<i>Bacillus anthracis</i>	Cápsula de ácido D-glutâmico: inibição da fagocitose das células vegetativas pelos macrófagos durante a infecção por antraz (EZZELL e WELKOS, 1999).

Os fatores de virulência têm grande impacto no grau de patogenicidade das bactérias e até mesmo na progressão de algumas doenças [Figura 2]:

A infecção crônica pelo *Helicobacter pylori* é um fator essencial na patogênese do carcinoma gástrico, devido ao gene CagA fator de virulência relacionado a um risco aumentado para o desenvolvimento tumoral (WANG et al., 2014). Outro exemplo de patógeno com um fator de virulência bastante relevante é o *Corynebacterium diphtheriae*, principal agente etiológico da difteria, uma doença potencialmente fatal ao hospedeiro, principalmente devido à ação da toxina diftérica (TD), uma exotoxina produzida pelo microrganismo (SHARMA et al., 2019, SUGIMAN-MARANGOS et al., 2022).

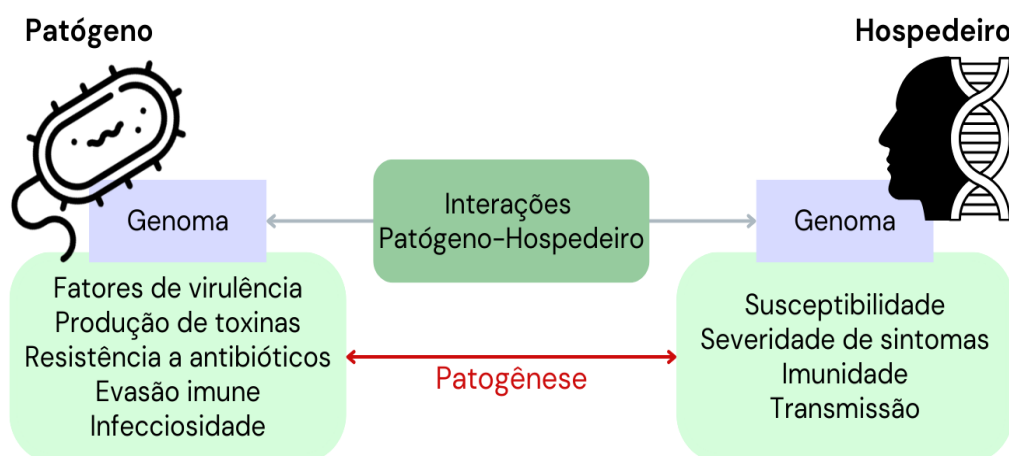


Figura 2. Interações genoma-genoma no contato entre patógeno e hospedeiro

3.2.2. Interações Moleculares

Técnicas para identificação de interações proteína-proteína (PPIs), como a co-imunoprecipitação, ensaios *pull-down*, EMSA (electrophoresis gel mobility shift analysis) e *DNase I footprinting* são bem desenvolvidas e estabelecidas em vários campos de estudo, mas técnicas *in vitro* tendem a ser processos muito caros e trabalhosos, pois exigem um grande número de anticorpos e equipamentos específicos, além da otimização das condições e de outros fatores. Além disso, o grande número de proteínas do hospedeiro e do patógeno representa um problema experimental em tempo real. Nesse sentido, a pesquisa

de bioinformática tende a ser bem-sucedida em vários tipos de análise. Assim, foram desenvolvidas abordagens para realizar a previsão *in silico* de PPIs com base no aprendizado de máquina, inicialmente por Bock et al. (2001) por meio do uso da Máquina de Vetor de Suporte (SVM) para treinar e prever interações com base na estrutura primária e nas propriedades físico-químicas relacionadas.

As técnicas de previsão de PPIs tendem a ser eficientes na triagem primária das interações realizadas entre o patógeno e o hospedeiro, fornecendo dados prováveis que podem ser verificados experimentalmente após o financiamento. Assim, estudos preditivos de PPIs putativos em HPIs são importantes para observar se determinadas vias tendem ou não a ser influenciadas por um determinado patógeno. Há vários tipos de previsão, a saber:

- **Aqueles baseados em homologia:** Quando uma determinada interação entre proteínas tende a ocorrer em sequências similares. (LEE et al., 2008; MATTHEWS et al., 2001).
- **Aquelas baseadas na estrutura:** Quando um determinado par de proteínas tem estruturas semelhantes a um par de proteínas que já interagem (DAVIS et al., 2007, DOOLITTLE e GOMEZ, 2010).
- **Aquelas baseadas na interação entre motivos conservados:** esses tipos de previsões integram PPIs intraespecíficas conhecidas com perfis de domínio/motivo de proteína que interagem (DYER et al., 2007, EVANSet al., 2009).
- **Aqueles baseados em aprendizado de máquina:** Eles extraem dados recuperados de interações entre hospedeiro e patógeno (THIEU et al., 2012, CHAUSSABEL et al., 2003).

3.2.3. Respostas Imunes e Evasão Imunológica

A resposta imunológica do hospedeiro contra um patógeno começa com o macrófago residente próximo ao sítio de entrada do patógeno no corpo humano e, ao observar isto, realiza fagocitose do patógeno visando controlar e “alertar” novas unidades de leucócitos sobre a infecção. Após a fagocitose, o macrófago

libera citocinas e quimiocinas que deixam o local da infecção e se ligam a proteoglicanos presentes na superfície do endotélio do vaso sanguíneo mais próximo que “espera” que um leucócito passe por esse vaso (ABBAS, 2018; KOLACZKOWSKA e KUBES, 2013; MOORE et al., 1995). Esse primeiro estágio é conhecido como o reconhecimento pelo sistema imunológico inato da existência do patógeno no corpo.

Posteriormente, a célula dendrítica (uma das APCs - Antigen Presenting Cells - Células Apresentadoras de Antígeno) leva o epítipo do patógeno, processado e expresso no MHC I ou II (dependendo do tipo de célula T que se deseja ativar), para os órgãos linfóides secundários, como os linfonodos, onde apresentará o epítipo aos linfócitos T ingênuos e desencadeará uma resposta imune celular, como a produção de anticorpos. Após a apresentação, os linfócitos sofrem expansão clonal, aumentando o número de linfócitos ativados específicos para esse agente infeccioso. Essa ação leva alguns dias. As células do sistema imunológico interagem entre si por meio de quimiocinas/citocinas, coordenando as atividades da resposta imune inata e adaptativa, uma vez que todas as células do sistema imunológico são capazes de secretar essas citocinas e receber sinais por meio de receptores para elas (ABBAS, 2018).

Assim, sabemos que, durante o curso de uma infecção, o que determinará o sucesso e a capacidade do patógeno de sobreviver e se replicar em um ambiente hostil, devido à ação do sistema imunológico do hospedeiro, depende de como será a interação entre os dois, principalmente um hospedeiro e um patógeno, que determina a capacidade do micróbio de sobreviver em um ambiente hostil e promove uma resposta imunológica do hospedeiro como medida de proteção. Por exemplo, no caso da *Burkholderia spp*, a ligação inicial aos receptores de superfície celular presentes no hospedeiro desencadeia uma série de liberação de mediadores pró-inflamatórios, iniciando importantes vias de sinalização, como NF κ B, Erk e Akt, bem como a indução de um grande número de citocinas, como as interleucinas 1 β , 6 e 8 (IL-1 β , 6 e 8), o fator de necrose tumoral alfa (TNF- α) e etc. das células epiteliais alveolares e bronquiolares (LU et al., 2012; GILLETTE et al., 2013; SIM et al., 2009). Todas essas vias estão relacionadas à resposta imune inicial contra o patógeno que acabou de entrar no corpo, com organização quimiotática destinada a controlar

e acabar com a infecção (DAVID et al., 2015), no entanto, ele tende a desenvolver mecanismos de evasão imune para sobreviver.

Uma das estratégias de imunização mais eficientes e mais comentadas, é baseada no desenvolvimento de vacinas. Vacinas são preparações biológicas projetadas para prevenir doenças infecciosas, através de treinamento do sistema imune a reconhecer e eliminar futuras infecções de maneira rápida e eficaz. A imunização através de vacinas é uma das ferramentas mais importantes da saúde pública, responsável por erradicar doenças como a varíola e controlar várias outras, como o sarampo e a poliomielite (DOWDLE, 1998; ROUSH, 2007; TALBIRD et al., 2022).

Há vários tipos de formulações vacinais possíveis, pois há inúmeras maneiras de fazer o sistema imunológico reconhecer partes do patógeno e induzir a sua eliminação. Um exemplo clássico e recente, foi quando, em relativamente um curto período de tempo, estavam disponíveis diversos tipos de vacinas contra o Sars-Cov-2, devido a pandemia que se iniciou em meados de 2019. A saber, foram desenvolvidas formulações do tipo “Vírus Inativado”, isto é, o antígeno como versão do vírus SARS-CoV-2 inativado; “Vivo atenuado”, com SARS-CoV-2 enfraquecido; de “Subunidade de proteína”, com a proteína spike do SARS-CoV-2 presente; “Partículas semelhantes a vírus (VLPs)” havendo múltiplas proteínas do SARS-CoV-2 na formulação vacinal; Vetor viral, DNA e RNA, contendo o ácido nucléico do vírus (RAHMAN et al., 2022).

Neste sentido, o desenvolvimento de uma vacina eficaz contra determinado patógeno deve levar em consideração características singulares do agente patológico, incluindo seus mecanismos de evasão imunológica. Ainda sobre vírus, sabe-se que o sistema de produção de interferon tipo I (IFNs- α/β) é ativado pelo reconhecimento de componentes virais por fatores da célula hospedeira (SAMUEL, 2001; SADLER, 2008). Contudo, um dos mecanismos de evasão do sistema imunológico pelos vírus é justamente o desenvolvimento de antagonismo de IFN viral por meio da codificação de proteínas virais capazes de antagonizar a via do IFN (SADLER, 2008; XU et al., 2021).

Ao considerar o desenvolvimento de vacinas para esses grupos de patógenos, é fundamental que elas estimulem tanto a resposta imune inata quanto a adaptativa. No campo da medicina veterinária, as vacinas vivas

atenuadas são geralmente preferidas por serem consideradas mais eficazes do que as inativadas. Essas vacinas vivas induzem tanto a imunidade humoral quanto a celular, enquanto as inativadas promovem principalmente uma resposta de anticorpos. A remoção do antagonismo de interferon (IFN) em vírus tem sido uma estratégia usada para aprimorar a imunogenicidade e aumentar a proteção em vacinas humanas e veterinárias (SU et al., 2023). Um exemplo é a proteína NS5 do vírus da dengue (DENV) no Zika virus (ZIKV), que interage com a proteína STAT2 humana e inibe a via do IFN, destacando-se como um importante fator de virulência e um possível alvo para o desenvolvimento de vacinas atenuadas contra flavivírus (GRANT et al., 2016; ASHOUR et al., 2010).

4. Aplicações da Patogenômica

4.1. Diagnóstico e identificação de patógenos

Organismos patogênicos, tais como bactérias e fungos, eram identificados pelas suas características fisiológicas e morfológicas, à partir da invenção da microscopia, e em métodos como a coloração de Gram, que resultavam em diagnósticos tardios, só possíveis após a manifestação dos primeiros sintomas, causando desconforto para o paciente, dificuldade para o tratamento (podendo levar à morte) e prevenção quanto à transmissão do patógeno. Os avanços na medicina tornaram possível a identificação das infecções de maneira mais precoce, através da identificação de biomarcadores que indicam uma infecção (DE NOOIJER et al., 2023), mas caracterizar e estudar qual patógeno está causando a infecção ainda é difícil, tendo em vista as características do meio ideais de cada microrganismo, como pH, temperatura, salinidade, oxigenação, entre outros (LIU et al., 2022). Os avanços no campo das ômicas (genômica, transcriptômica, proteômica, entre outros) estão tornando possível uma identificação mais precisa dos genes dos patógenos, com o desenvolvimento de programas e algoritmos mais eficientes, além da expansão dos bancos de dados. Pesquisadores estão sendo capazes de identificar marcadores e precursores de metabolismo, sendo possível o tratamento de infecções antes do surgimento de sintomas agudos ou marcadores toxicológicos que só aparecem após a doença, assim aumentando a eficiência dos tratamentos.

Os métodos convencionais envolvem técnicas associadas ou não a meios de cultura (NEHRA et al., 2022). Os mesmos podem ser baseados em imunologia, como por exemplo a detecção por ligação a anticorpos, que permite uma detecção não-invasiva (coleta de saliva), porém a confiabilidade só é alta após um certo tempo depois da infecção inicial, e os resultados perdem a exatidão dependendo da resposta imune dos hospedeiros, que pode não ser tão intensa (LEVA-BUENO et al., 2020). Os métodos baseados em ácidos nucleicos, tais como a reação em cadeia de polimerase (PCR), são atualmente os mais utilizados, especialmente após a pandemia do Sars-Cov (ZENG et al., 2024), e são fundamentais no trabalho com patógenos que não crescem *in vitro*, onde é possível amplificar e identificar genes específicos, porém também dependem do período de infecção para detectar a presença do microrganismo. Além disso, também são usadas técnicas relacionadas com espectrometria de massa, tais como a de MALDI-TOF, que apesar de rápida e relativamente barata, possui baixa discriminação entre espécies e sub-espécies de organismos, podendo levar a erros de identificação (ELBEHIRY et al., 2022). De forma geral, os métodos convencionais possuem uma série de dificuldades para serem executados, tais como o preço dos reagentes e equipamentos, o tempo para obter os resultados e a necessidade de pessoal qualificado, além da chance de resultar em falsos positivos/negativos por falha humana ou de produto.

Com os avanços na genômica, a partir do sequenciamento de nova geração (NGS), foi possível observar diferenças entre os genes de organismos, que, anteriormente eram classificados como taxonomicamente iguais, mas foram observadas grandes diferenças quanto a marcadores ribossomais e/ou metabólicos. Tendo bancos de dados com uma riqueza grande de genes confiáveis, veio à tona métodos de identificação taxonômica como a Identidade Média de Nucleotídeos (ANI), que realiza a identificação taxonômica a partir da semelhança dos genes ortólogos entre dois organismos (JAIN et al., 2018), e a Hibridização Digital de DNA-DNA (dDDH), que avalia a semelhança entre segmentos genômicos homologados para determinar a identidade nucleotídica média entre eles, e assim determinar a proximidade taxonômica com os genomas do banco de dados (MEIER-KOLTHOFF e GÖKER et al., 2019). Uma das aplicações da patogenômica é o monitoramento da presença e ausência de

genes de interesse, e no caso de patógenos, genes de virulência, resistência a antibióticos e elementos genéticos móveis (MGEs). Com essas informações é possível prever o comportamento de microrganismos patogênicos, e planejar o tratamento adequado para as doenças.

Em um estudo de *Daria Bottai et al* (BOTTAI et al., 2020), foi observado que diversas cepas de *Mycobacterium tuberculosis* presentes em epidemias globais de tuberculose teriam a deleção em comum de uma região gênica TbD1. Essa deleção, que aumenta a resistência a estresse oxidativo e hipóxia, é um fator evolutivo que pode ser um ponto de interesse no monitoramento desse patógeno, observando em cepas que ainda possuem o gene. Em outro caso, de *Migle Gabrielaite et al* (GABRIELAITE et al., 2021), foi observado que pacientes com fibrose cística estavam sendo infectados por *Achromobacter spp.*, gerando casos de resistência a antibióticos, por exemplo. Dessa forma foram analisados 101 genomas isolados de 51 pacientes, revelando padrões de mutação que indicam uma adaptação genética convergente, e esse padrão de comportamento pode ser fundamental para prever o resultado dessas mutações e como elas impactam na virulência do patógeno.

4.2. Desenvolvimento de terapias e vacinas

Com a melhoria nos bancos de dados, devido ao aumento do número de genomas e a qualidade dos mesmos, novas análises como a pangenômica estão se tornando viáveis para a identificação de novos meios de tratamento. Nas análises pangenômicas, a partir do estabelecimento de um grupo de organismos (de mesma espécie, por exemplo), se obtém um conjunto de genes essenciais (genes core) pertencentes a todos os indivíduos, fundamentais para a sobrevivência desses microrganismos. A implicação desse fato é, ao desenvolver um modelo vacinal, ou uma droga que atinge um gene essencial, em teoria todos os indivíduos do grupo são atingidos, assim criando um tratamento mais eficiente contra o patógeno. Os modelos de vacinologia estão se atualizando, tendo em vista que os modelos clássicos possuem um desenvolvimento muito custoso em termos de dinheiro e tempo. Com a patogenômica, é possível fazer a predição de quais são as proteínas de

membrana de um grupo de microrganismos, e assim desenvolver uma vacina que abrange todos.

Os microrganismos nosocomiais, ou seja, aqueles que estão relacionados à infecções em ambientes hospitalares, são um foco fundamental de monitoramento, pois estão constantemente expostos a novos hospedeiros, muitas vezes com outras infecções. Dessa forma, esses organismos recebem e transferem genes de forma horizontal, e com a adição do uso constante de antibióticos e drogas, o potencial de surgimento de patógenos com alta virulência e resistência a drogas é alto. Um estudo de *Cyril Lefebvre et al.* (LEFEBVRE et al., 2020) investigou fatores de patogenicidade de *Mycobacterium tuberculosis*, e foi observado que o gene *Rv0504c*, que produz a proteína *HadDMtb*, estaria relacionado com o metabolismo dessa bactéria, participando na formação de biofilmes e resistência a baixas temperaturas. Com essa evidência, a possibilidade de produção de um medicamento que pode inibir essa produção poderia reduzir a taxa de sobrevivência desse microrganismo. Certos tipos de microrganismos, como a espécie *Escherichia coli*, que possui uma grande variedade de cepas patogênicas, precisam ser monitoradas quanto ao seu potencial de se tornarem patógenos. J R Clark e A M Maresso (CLARK et al., 2021) construíram um patótipo a partir de 107 cepas reconhecidas como patógenos, e assim identificaram uma série de genes de interesse prevalentes nesses organismos (evasão do sistema imune, absorção de ferro, adesão e produção de toxinas). Com isso é possível monitorar se alguma cepa está se tornando patogênica, além do desenvolvimento de possíveis tratamentos para esses organismos.

4.3. Monitoramento e Controle de Epidemias

O monitoramento genômico tem transformado a forma pela qual monitoramos e controlamos epidemias infecciosas. Ao monitorar e analisar os genomas de patógenos, pesquisadores podem rastrear a origem de surtos, habilitando o controle de infecções (STRUELENS et al., 2024). Uma metodologia chave para esse monitoramento é o sequenciamento de genoma inteiro (WGS), que permite a pesquisadores o rastreamento de diferentes linhagens e variantes (NAFEA et al., 2024). Durante a pandemia do COVID-19, o WGS foi crucial para

rastrear variantes do vírus SARS-CoV-2, permitindo que cientistas e autoridades fizessem intervenções pontuais com agilidade (PAPA et al., 2023; YALÇIN et al., 2024). Portanto, os dados genômicos obtidos por WGS, em conjunto com modelos epidemiológicos, podem ser usados para prever e controlar epidemias, inclusive, no caso de modelos preditivos, de começar preparativos para controlá-las com antecedência (CARDENAS et al., 2022; WANG et al., 2022). Exemplos disso em tempos recentes foram os surtos de Ebola na África Ocidental e a pandemia da COVID-19, ambos casos em que a patogenômica teve função chave no controle de surtos, habilitando tomadas de decisão através de dados genômicos em tempo real (BARRETT et al., 2022, DI PAOLA et al., 2020).

No futuro, o conceito de patogenômica e seus insights poderá contribuir muito para o desenvolvimento de estratégias preventivas. O monitoramento do conteúdo genético de patógenos informando o desenvolvimento de vacinas e métodos diagnósticos, bem como a previsão de ameaças biológicas (FENG et al., 2024).

4.4. Pesquisa e Desenvolvimento de Novos Alvos Terapêuticos

Em tempos recentes, o desenvolvimento de novas plataformas de WGS permitiu a obtenção de conteúdo genômico de alta qualidade, o que viabiliza o uso da patogenômica para identificar novas proteínas ou genes como alvos de droga no genoma dos patógenos (ALTURKI et al., 2022; HYUN et al., 2023). Essas metodologias dependem do estudo cuidadoso do conteúdo genético dos patógenos em busca de genes chave que são essenciais para a sobrevivência do patógeno, fazendo deles excelentes alvos para intervenção terapêutica (LIANG et al., 2024; ALKATHERI et al., 2023) e também em busca de mecanismos de resistência a antibióticos no genoma (DJORDJEVIC et al., 2024). Algumas técnicas emergentes, como as técnicas de CRISPR-Cas baseadas no genoma e no RNA, são altamente precisas e ajudam na examinação dos genomas de patógenos, podendo indicar tratamentos ainda mais eficientes e precisos (LI et al., 2023; ZHU et al., 2022).

Além disso, com a ajuda do WGS e das descobertas em genômica, o campo de pesquisa farmacêutica visando a descoberta de novos medicamentos foi

revolucionado, levando ao desenvolvimento de alvos terapêuticos mais específicos. Por exemplo, a identificação *in silico* de novos alvos para o tratamento da sífilis por meio da análise das sequências do genoma de *Treponema pallidum*. (KUMAR et al., 2017).

5. Desafios e Perspectivas Futuras

A patogenômica, um campo emergente que combina dados genômicos com informação clínica e imunológica para avançar no diagnóstico e tratamento de doenças infecciosas, enfrenta desafios técnicos, metodológicos e éticos significativos, mas também oferece grandes promessas para a medicina de precisão. Este domínio tem evoluído ao longo dos anos através de constantes avanços tecnológicos, mas continuam a surgir novos desafios à medida que o volume de dados aumenta e cresce a necessidade da sua integração eficiente.

5.1. Desafios técnicos e metodológicos

Apesar dos avanços na tecnologia de sequenciação de nova geração (NGS), subsistem limitações consideráveis. A maioria dos métodos está limitada a dados unimodais, deixando a integração de modalidades, como os dados genômicos e imunológicos, relativamente pouco desenvolvida (FENG et al., 2024). Os desafios incluem a montagem exata de genomas complexos, como os de microrganismos com sequências repetitivas, e a dificuldade de integrar eficazmente informações fenotípicas e genotípicas (FENG et al., 2024). As ferramentas emergentes de inteligência artificial multimodal estão a começar a ultrapassar estas limitações, permitindo correlações entre características genômicas e patológicas, particularmente em oncologia.

Cornish e Guda (2015) sublinham a importância de pipelines bioinformáticas robustas para garantir a precisão na identificação de variantes genéticas, uma área crítica para o diagnóstico genómico preciso em agentes patogênicos (MILLER et al., 2010)

À medida que o volume de dados gerados pelas plataformas de sequenciação aumenta, são necessárias infraestruturas de computação potentes e algoritmos avançados para processar e interpretar estes dados em tempo útil. Um dos principais problemas enfrentados é a baixa proporção de DNA

patogénico nas amostras de infecção, o que exige tecnologia de sequenciação profunda e bioinformática avançada para reconstruir genomas patogénicos históricos ou incompletos (SATAN et al., 2023).

A montagem de genomas complexos, especialmente de agentes patogénicos, é outro grande desafio. As tecnologias atuais são frequentemente incapazes de lidar com genomas que têm sequências repetitivas ou altamente heterogêneas (LYNCH et al., 2016; SEREIKA et al., 2022; HAYASHI e OGURA, 2020). Destacam-se os desafios enfrentados pelos algoritmos de montagem, particularmente em genomas com sequências repetitivas, o que compromete a precisão dos dados de montagem de agentes patogénicos. Além disso, a identificação de variantes genéticas patogénicas é dificultada pela falta de genomas de referência bem anotados para algumas espécies (LYNCH et al., 2016).

O volume crescente de dados genômicos suscitou preocupações éticas sobre a privacidade, a segurança e a utilização desses dados (ODEH et al., 2021; WANGMO et al., 2018; MATHAIYAN et al., 2013). A proteção de dados sensíveis, como a informação genética relacionada com doenças infecciosas, exige o desenvolvimento de quadros éticos e jurídicos sólidos, especialmente no contexto da partilha global de dados durante emergências de saúde pública. A utilização da IA em grandes volumes de dados genômicos, como discutido por Vayena et al. (2018), levanta questões importantes sobre a privacidade e a segurança dos doentes.

Apesar dos avanços na bioinformática, continua a haver necessidade de ferramentas mais eficientes para analisar dados genômicos em grande escala e em tempo real. Embora os atuais pipelines bioinformáticos possam processar grandes volumes de dados, são limitados na sua capacidade de integrar de forma eficiente e normalizada diferentes tipos de dados multimodais (como dados clínicos, imagiológicos e genômicos) (HAYASHI e OGURA, 2020; MANIRUZZAMAN e NOKHODCHI, 2017, HE et al., 2023). A integração de dados de RNA-Seq com outras abordagens genômicas, tal como proposto por Wang, Gerstein e Snyder (2009), é uma estratégia importante para ultrapassar estas limitações e melhorar a análise de dados multimodais. Direcções futuras: Integração multimodal avançada e medicina de precisão

O futuro da patogenômica reside na integração total dos dados multimodais, combinando a genômica, a informação clínica e os dados imagiológicos com a ajuda da inteligência artificial e da aprendizagem automática (FENG et al., 2024; HE et al., 2023). Tal permitirá o desenvolvimento de biomarcadores específicos que combinam métricas genotípicas e fenotípicas, conduzindo a uma medicina de precisão mais eficaz. Além disso, serão essenciais melhores orientações éticas e tecnológicas para garantir a privacidade e a segurança dos dados genômicos a nível mundial, enquanto as ferramentas de análise em tempo real serão fundamentais para lidar com surtos infecciosos em rápida evolução (LEFTEROVA et al., 2015). A utilização da IA, como discutido por Topol (2019), pode acelerar a identificação de agentes patogênicos e a integração de dados multimodais para melhorar o diagnóstico e o tratamento em tempo real.

5.2. Tendências Futuras

O sequenciamento de nova geração (NGS) representa um avanço significativo na área da genômica, permitindo a análise de grandes volumes de dados genéticos de forma rápida e eficiente. Essa tecnologia, que surgiu em 2005, possibilita a leitura simultânea de milhões de bases nitrogenadas, desvendando os complexos mecanismos moleculares que subjazem à vida. As aplicações do NGS são vastas, abrangendo desde a medicina personalizada, com o desenvolvimento de tratamentos direcionados a mutações genéticas específicas, até a agricultura, com a criação de cultivos mais resistentes a pragas e doenças. (CARVALHO e SILVA, 2010)

A genômica e a proteômica revolucionaram a biologia molecular, alterando profundamente nossa compreensão da química dos ácidos nucleicos e das proteínas, e abrindo novas possibilidades para a manipulação de sistemas biológicos. O desenvolvimento de protocolos robustos para o uso de ferramentas moleculares impulsionou avanços em áreas como cultura de células, engenharia de tecidos e reprodução assistida. No Brasil, embora o Projeto Genoma Humano tenha tido menor impacto, os investimentos em genômica de microrganismos e parasitas, e mais recentemente em proteômica, têm sido cruciais. O conhecimento gerado por essas pesquisas têm o potencial de transformar a

prevenção e o tratamento de doenças. A diabetes tipo 1, por exemplo, tem sido alvo de estudos intensivos, e a descoberta do papel do gene Sumo-4 na regulação do sistema imunológico e a capacidade de manipular células-tronco para produzir insulina representam avanços promissores para o tratamento dessa doença (GARCIA, 2006).

O Sequenciamento de Nova Geração (NGS) oferece leituras mais longas e uma capacidade significativamente maior de gerar dados em comparação ao método tradicional de Sanger. Isso possibilita trabalhar com coberturas genômicas mais amplas, facilitando a montagem de genomas desconhecidos. Essa eficiência no sequenciamento é resultado da utilização de clonagem *in vitro* e sistemas de suporte sólido, que eliminam a necessidade de produzir clones bacterianos, montar placas de sequenciamento e separar fragmentos em géis. Essa técnica é aplicada, por exemplo, no desenvolvimento de marcadores moleculares, especialmente marcadores microssatélites de DNA, que ainda não existem para essa espécie (LEÃO, 2007).

Antes de 2000, as tecnologias de sequenciamento, como o método de Sanger, eram caras e demoradas, restringindo seu uso na patogenômica a estudos de pequena escala. Para enfrentar esses desafios, as técnicas de sequenciamento foram automatizadas, o que aprimorou a capacidade de processamento e diminuiu a margem de erro humano. Na década de 2000, com a introdução do Sequenciamento de Nova Geração (NGS), tornou-se possível realizar sequenciamentos em larga escala. No entanto, essa nova tecnologia ainda enfrentava problemas de capacidade e precisão, especialmente em relação a leituras curtas e erros em regiões genômicas complexas. O desenvolvimento de plataformas de NGS, como a Illumina, ajudou a reduzir custos e a facilitar estudos genômicos amplos, embora os desafios de montagem e precisão persistem. Entre 2010 e 2015, o aumento no número de projetos de sequenciamento resultou em grandes volumes de dados, sobrecarregando as ferramentas e a infraestrutura computacional disponíveis. Para resolver essa questão, surgiram pipelines de bioinformática, que incluíam ferramentas específicas para processar e analisar esses dados. No entanto, a padronização dos fluxos de trabalho e a integração de dados multimodais ainda eram limitadas. Entre 2015 e 2020, a integração de dados multimodais, que abrange informações

genômicas, clínicas e de imagem, continuou a ser um desafio. Apesar dos avanços no sequenciamento, a dependência de dados unimodais persistia. A solução surgiu com a introdução de ferramentas de inteligência artificial multimodal, que começaram a integrar esses diferentes tipos de dados. Desde 2020, o aumento do uso de dados patogenômicos levantou questões éticas, especialmente relacionadas à privacidade e consentimento, particularmente em situações de emergência de saúde pública. Para abordar essas preocupações, desenvolveu-se frameworks éticos e ferramentas de bioinformática que permitem análises em tempo real, com o objetivo de proteger a privacidade e promover o compartilhamento responsável de dados (LARSON et al., 2023; LYNCH et al., 2023; ARELLANO et al., 2018).

A combinação de dados genômicos, transcriptômicos e proteômicos é fundamental para uma compreensão mais completa das doenças. Pesquisas futuras devem se concentrar em desenvolver plataformas que integrem esses dados para fornecer insights mais profundos.

A fusão de dados multimodais tem como ideia extrair e integrar informações contextuais complementares de diferentes fontes, proporcionando uma base mais sólida para a tomada de decisões. Essa abordagem é especialmente importante na medicina, onde resultados semelhantes em uma única modalidade podem ser interpretados de maneiras distintas quando considerados em conjunto com outras modalidades (LEÃO, 2007).

6. Conclusão

A patogenômica, como abordado ao longo deste capítulo, representa um avanço essencial no estudo de doenças infecciosas, proporcionando uma compreensão profunda dos mecanismos genéticos que permitem aos patógenos causar infecções e resistir a tratamentos. A disponibilidade de grandes volumes de dados genômicos, combinada com o uso de ferramentas de bioinformática, tem revolucionado o diagnóstico e o tratamento de doenças, permitindo não apenas a identificação precisa de patógenos, mas também o desenvolvimento de estratégias terapêuticas personalizadas e vacinas mais eficazes.

O impacto desse campo para a ciência e a saúde pública é inegável, especialmente no monitoramento e controle de doenças emergentes e patógenos multirresistentes. Ao sequenciar e analisar genomas de patógenos, é possível identificar fatores de virulência e resistência, prever a evolução de surtos e orientar políticas de saúde pública, como ilustrado pela resposta à pandemia de COVID-19. Ademais, as metodologias de sequenciamento permitiram uma abordagem mais abrangente e precisa no diagnóstico molecular, melhorando significativamente as estratégias de detecção precoce.

Além disso, a patogenômica contribui para o desenvolvimento de novos fármacos e antibióticos, auxiliando na identificação de novos alvos terapêuticos e na compreensão das interações patógeno-hospedeiro, que são cruciais para a inovação no combate a patógenos de alta resistência. Em suma, este campo emergente continua a transformar o modo como a ciência compreende e responde às doenças infecciosas, sendo uma ferramenta indispensável na promoção de diagnósticos rápidos, tratamentos eficazes e controle eficiente de surtos.

Diante disso, é evidente a revolução que a patogenômica trouxe para o contexto de terapêutica e prevenção. E à medida que a compreensão da interação patógeno-hospedeiro avançou e a bioinformática foi inserida no desenvolvimento farmacêutico, também foi possível identificar seus mecanismos de resistência a medicamentos antimicrobianos, um dos maiores desafios da era moderna. Dessa forma, a aplicação da genômica na patologia e farmacologia torna a identificação de genes e proteínas essenciais para a sobrevivência destes microrganismos patogênicos uma arma em potencial, reduzindo bastante as chances destes seres contornarem efeitos terapêuticos e conferindo mais eficácia e segurança a nova geração de fármacos.

Neste capítulo, reunimos evidências de como o conhecimento sobre o genoma de patógenos tem revolucionado a medicina nas últimas décadas. O avanço tecnológico proporcionado pela computação e pela internet foi significativo, e a era do Big Data trouxe consigo um volume de dados complexo e desafiador para análise. Isso ressalta a necessidade não apenas de novas ferramentas analíticas, mas também de profissionais qualificados com uma visão

multidisciplinar e interdisciplinar, capazes de integrar diferentes campos do conhecimento.

Espera-se que a acessibilidade às ferramentas de sequenciamento e análise de genomas de patógenos, de forma mais rápida, permita o desenvolvimento de pesquisas e aplicações na profilaxia, tratamento e controle de patógenos na saúde pública global. Na profilaxia, será possível identificar genes candidatos para vacinas, por exemplo. No tratamento, será viável identificar genes candidatos como alvos de fármacos. Quanto ao controle, a vigilância genômica de patógenos permitirá o monitoramento de novas variantes circulantes, identificando aquelas com alta morbidade, o que pode levar a uma pandemia, além de possibilitar o acompanhamento da circulação de genes de resistência a antimicrobianos.

Além disso, espera-se que a patogenômica integre-se a outras “ômicas”, como a transcriptômica e a metabolômica, para melhorar o entendimento sobre a biologia dos patógenos e a interação patógeno-hospedeiro. Combinados esses esforços, espera-se que as 'ômicas' continuem revolucionando a biomedicina e que os profissionais atuantes nesta área estejam atentos aos desafios e preparados para enfrentá-los, promovendo o constante desenvolvimento do conhecimento científico e tecnológico.

7. Referências

- ABBAS, A. K. *Imunologia celular e molecular*. Elsevier, 2018.
- ABBY, S.; DAUBIN, V. Comparative genomics and the evolution of prokaryotes. **Trends Microbiol.**, v. 15, n. 3, p. 135–41
- AGARWAL, J.; SRIVASTAVA, S.; SINGH, M. Pathogenomics of uropathogenic **Escherichia coli**. **Indian J Med Microbiol.**, v. 30, n. 2, p. 141–149, abr. 2012.
- ALKATHERI, A. H. et al. Microbial Genomics: Innovative Targets and Mechanisms. *Antibiotics*, v. 12, n. 2, p. 190, 17 jan. 2023.
- ALSPAUGH, J. A. Virulence mechanisms and *Cryptococcus neoformans* pathogenesis. *Fungal Genet Biol.*, v. 78, p. 55–58, maio 2015.
- ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local alignment search tool. *J Mol Biol.*, v. 215, n. 3, p. 403–410, out. 1990.

- ALTURKI, N. A.; MASHRAQI, M. M.; JALAL, K.; KHAN, K.; BASHARAT, Z.; ALZAMAMI, A. Therapeutic Target Identification and Inhibitor Screening against Riboflavin Synthase of Colorectal Cancer Associated **Fusobacterium nucleatum**. **Cancers**, v. 14, n. 24, p. 6260, 19 dez. 2022.
- ARBER, W. Genetic variation: molecular mechanisms and impact on microbial evolution. **FEMS Microbiol Rev.**, v. 24, n. 1, p. 1–7, jan. 2000.
- ARELLANO, A. M.; DAI, W.; WANG, S.; JIANG, X.; OHNO-MACHADO, L. Privacy Policy and Technology in Biomedical Data Science. *Annu Rev Biomed Data Sci.*, v. 1, n. 1, p. 115–129, 20 jul. 2018.
- ARENAS, M.; ARAUJO, N. M.; BRANCO, C.; CASTELHANO, N.; CASTRO-NALLAR, E.; PÉREZ-LOSADA, M. Mutation and recombination in pathogen evolution: Relevance, methods and controversies. **Infect Genet Evol.**, v. 63, p. 295–306, set. 2018.
- ARMSTRONG, G. L. et al. Pathogen Genomics in Public Health. **N Engl J Med.**, v. 381, n. 26, p. 2569–2580, 26 dez. 2019.
- ASHOUR, J. et al. Mouse STAT2 Restricts Early Dengue Virus Replication. **Cell Host Microbe**, v. 8, n. 5, p. 410–421, nov. 2010.
- ATHANASOPOULOU, K.; BOTI, M. A.; ADAMOPOULOS, P. G.; SKOUROU, P. C.; SCORILAS, A. Third-Generation Sequencing: The Spearhead towards the Radical Transformation of Modern Genomics. **Life**, v. 12, n. 1, p. 30, 26 dez. 2021.
- BAÑULS, A. L.; SANOU, A.; VAN ANH, N. T.; GODREUIL, S. Mycobacterium tuberculosis: ecology and evolution of a human bacterium. *J Med Microbiol.*, v. 64, n. 11, p. 1261–1269, 1 nov. 2015.
- BARRETT, A. D. T. et al. The rapid progress in COVID vaccine development and implementation. **Npj Vaccines**, v. 7, n. 1, p. 20, s41541-022-00442–8, 10 fev. 2022.
- BOCK, J. R.; GOUGH, D. A. Predicting protein–protein interactions from primary structure. **Bioinformatics**, v. 17, n. 5, p. 455–460, 1 maio 2001.
- BOTTAI, D. et al. TbD1 deletion as a driver of the evolutionary success of modern epidemic **Mycobacterium tuberculosis** lineages. **Nat Commun.**, v. 11, n. 1, p. 684, 4 fev. 2020.
- BRAY, N.; DUBCHAK, I.; PACHTER, L. AVID: A Global Alignment Program. *Genome Res.*, v. 13, n. 1, p. 97–102, 1 jan. 2003.
- CACHO, A.; SMIRNOVA, E.; HUZURBAZAR, S.; CUI, X. A Comparison of Base-calling Algorithms for Illumina Sequencing Technology. **Brief Bioinform.**, v. 17, n. 5, p. 786–795, set. 2016.
- CADENA, J.; THOMPSON, G. R.; PATTERSON, T. F. Aspergillosis. **Infect Dis*

- Clin North Am.*, v. 35, n. 2, p. 415–434, jun. 2021.
- CÁRDENAS, P.; CORREDOR, V.; SANTOS-VEGA, M. Genomic epidemiological models describe pathogen evolution across fitness valleys. *Sci Adv.*, v. 8, n. 28, p. eabo0173, 15 jul. 2022.
- CARVALHO, M. C. D. C. G. D.; SILVA, D. C. G. D. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. Ciênc Rural, v. 40, n. 3, p. 735–744, mar. 2010.
- CDC. Outbreak Detection [Internet]. CDC PulseNet, 2024 [citado em 4 out. 2024]. Disponível em: <https://www.cdc.gov/pulsenet/hcp/about/outbreak-detection.html>
- CHAHAL, G.; TYAGI, S.; RAMIALISON, M. Navigating the non-coding genome in heart development and Congenital Heart Disease. *Differentiation*, v. 107, p. 11–23, maio 2019.
- CHANDRA, P.; GRIGSBY, S. J.; PHILIPS, J. A. Immune evasion and provocation by *Mycobacterium tuberculosis*. *Nat Rev Microbiol.*, v. 20, n. 12, p. 750–766, dez. 2022.
- CHARRON, P.; KANG, M. VariantDetective: an accurate all-in-one pipeline for detecting consensus bacterial SNPs and SVs. ALKAN, C. (Ed.). *Bioinformatics*, v. 40, n. 2, p. btae066, 1 fev. 2024.
- CHATTERJEE, N.; WALKER, G. C. Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen.*, v. 58, n. 5, p. 235–263, jun. 2017.
- CHAUSSABEL, D. et al. Unique gene expression profiles of human macrophages and dendritic cells to phylogenetically distinct parasites. Blood. 2003 Jul 15;102(2):672–81.
- CHEN, J. Q.; WU, Y.; YANG, H.; BERGELSON, J.; KREITMAN, M.; TIAN, D. Variation in the Ratio of Nucleotide Substitution and Indel Rates across Genomes in Mammals and Bacteria. *Mol Biol Evol.*, v. 26, n. 7, p. 1523–1531, 1 jul. 2009.
- CHEN, L. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.*, v. 33, n. Database issue, p. D325–D328, 17 dez. 2004.
- CHIVIAN, D. et al. Metagenome-assembled genome extraction and analysis from microbiomes using KBase. *Nat Protoc.*, v. 18, n. 1, p. 208–238, jan. 2023.
- CHOI, S. R.; LEE, M. Transformer Architecture and Attention Mechanisms in Genome Data Analysis: A Comprehensive Review. *Biology*, v. 12, n. 7, p. 1033, 22 jul. 2023.
- CHOWDHURY, B.; GARAI, G. A review on multiple sequence alignment from the perspective of genetic algorithm. Genomics, v. 109, n. 5–6, p. 419–431, out. 2017.

- CLARK, J. R.; MARESSO, A. M. Comparative Pathogenomics of *Escherichia coli*: Polyvalent Vaccine Target Identification through Virulome Analysis. MONACK, D. (Ed.). *Infect Immun.* v. 89, n. 8, p. e00115-21, 15 jul. 2021.
- COLE, S. T. et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, v. 393, n. 6685, p. 537–544, jun. 1998.
- COMAS, I. et al. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nat Genet.*, v. 45, n. 10, p. 1176–1182, out. 2013.
- COSTA, F. F. Big data in biomedicine. *Drug Discov Today*, v. 19, n. 4, p. 433–440, abr. 2014.
- COVID-19 Host Genetics Initiative, COVID-19 Host Genetics Initiative Leadership, Niemi, M. E. K., Karjalainen, J., Liao, R. G., Neale, B. M. et al. Mapping the human genetic architecture of COVID-19. *Nature*, v. 600, n. 7889, p. 472–477, 16 dez. 2021.
- D'MELLO, A.; AHEARN, C. P.; MURPHY, T. F.; TETTELIN, H. ReVac: a reverse vaccinology computational pipeline for prioritization of prokaryotic protein vaccine candidates. *BMC Genomics*, v. 20, n. 1, p. 981, dez. 2019.
- DARLING, A. E.; MAU, B.; PERNA, N. T. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. STAJICH, J. E. (Ed.). *PLoS ONE*, v. 5, n. 6, p. e11147, 25 jun. 2010.
- DAVID, J.; BELL, R. E.; CLARK, G. C. Mechanisms of Disease: Host-Pathogen Interactions between Burkholderia Species and Lung Epithelial Cells. *Front Cell Infect Microbiol* [Internet], v. 5, 18 nov. 2015 [citado em 16 out. 2024]. Disponível em: <http://journal.frontiersin.org/Article/10.3389/fcimb.2015.00080/abstract>
- DAVIS, F. P.; BARKAN, D. T.; ESWAR, N.; MCKERROW, J. H.; SALI, A. Host-pathogen protein interactions predicted by comparative modeling. *Protein Sci.* v. 16, n. 12, p. 2585–2596, dez. 2007.
- DE NOOIJER, A. H.; PICKKERS, P.; NETEA, M. G.; KOX, M. Inflammatory biomarkers to predict the prognosis of acute bacterial and viral infections. *J Crit Care*, v. 78, p. 154360, dez. 2023.
- DEMUTH, A. et al. Pathogenomics: An updated European Research Agenda. *Infect Genet Evol.* v. 8, n. 3, p. 386–393, maio 2008.
- DI PAOLA, N.; SANCHEZ-LOCKHART, M.; ZENG, X.; KUHN, J. H.; PALACIOS, G. Viral genomics in Ebola virus research. *Nat Rev Microbiol.* v. 18, n. 7, p. 365–378, jul. 2020.
- DJORDJEVIC, S. P. et al. Genomic surveillance for antimicrobial resistance — a One Health perspective. *Nat Rev Genet.*, v. 25, n. 2, p. 142–157, fev.

2024.

- DONG, M.; MASUYER, G.; STENMARK, P. Botulinum and Tetanus Neurotoxins. **Annu Rev Biochem.**, v. 88, n. 1, p. 811–837, 20 jun. 2019.
- DOOLITTLE, J. M.; GOMEZ, S. M. Structural similarity-based predictions of protein interactions between HIV-1 and **Homo sapiens**. **Virol J.**, v. 7, n. 1, p. 82, dez. 2010.
- DOWDLE, W. R. The principles of disease elimination and eradication. *Bull World Health Organ.*, v. 76, n. Suppl 2, p. 22–25, 1998.
- DYER, M. D.; MURALI, T. M.; SOBRAL, B. W. Computational prediction of host-pathogen protein–protein interactions. **Bioinformatics**, v. 23, n. 13, p. i159–i166, 1 jul. 2007.
- EDMONDSON, D. G.; HU, B.; NORRIS, S. J. Long-Term In Vitro Culture of the Syphilis Spirochete **Treponema pallidum** subsp. **pallidum**. ROSA, P. A. (Ed.). **mBio**, v. 9, n. 3, p. e01153-18, 5 jul. 2018.
- EKTEFAIE, Y. et al. Globally diverse **Mycobacterium tuberculosis** resistance acquisition: a retrospective geographical and temporal analysis of whole genome sequences. **Lancet Microbe**, v. 2, n. 3, p. e96–e104, mar. 2021.
- ELBEHIRY, A. et al. How MALDI-TOF Mass Spectrometry Technology Contributes to Microbial Infection Control in Healthcare Settings. **Vaccines**, v. 10, n. 11, p. 1881, 8 nov. 2022.
- ELLERMANN, M.; ARTHUR, J. C. Siderophore-mediated iron acquisition and modulation of host-bacterial interactions. **Free Radic Biol Med.**, v. 105, p. 68–78, abr. 2017.
- ENGELTHALER, D. M. Genomic surveillance and pathogen intelligence. **Front Sci.**, v. 2, p. 1397048, 25 abr. 2024.
- ESPINOSA, E.; BAUTISTA, R.; LARROSA, R.; PLATA, O. Advancements in long-read genome sequencing technologies and algorithms. **Genomics**, v. 116, n. 3, p. 110842, maio 2024.
- EVANS, P.; DAMPIER, W.; UNGAR, L.; TOZEREN, A. Prediction of HIV-1 virus-host protein interactions using virus and host sequence motifs. **BMC Med Genomics.**, v. 2, n. 1, p. 27, dez. 2009.
- EZZELL, J. W.; WELKOS, S. L. The capsule of **Bacillus anthracis**, a review. **J Appl Microbiol.**, v. 87, n. 2, p. 250–250, ago. 1999.
- FENG, X. et al. Pathogenomics for accurate diagnosis, treatment, prognosis of oncology: a cutting edge overview. **J Transl Med.**, v. 22, n. 1, p. 131, 3 fev. 2024.
- FIGUEROA, J. et al. Analysis of single nucleotide polymorphisms (SNPs) associated with antibiotic resistance genes in Chilean *Piscirickettsia*

- salmonis strains. *J Fish Dis.*, v. 42, n. 12, p. 1645–1655, dez. 2019.
- FLEISCHMANN, R. D. et al. Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd. *Science*, v. 269, n. 5223, p. 496–512, 28 jul. 1995.
- FORSBERG, K. et al. *Candida auris*: The recent emergence of a multidrug-resistant fungal pathogen. *Med Mycol.*, v. 57, n. 1, p. 1–12, 1 jan. 2019.
- GABRIELAITE, M.; NIELSEN, F. C.; JOHANSEN, H. K.; MARVIG, R. L. *Achromobacter* spp. genetic adaptation in cystic fibrosis. *Microb Genomics* [Internet], v. 7, n. 7, 28 jul. 2021 [citado em 16 out. 2024]. Disponível em: <https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000582>
- GARCIA, E. S. *Genes: fatos e fantasias* [Internet]. Editora FIOCRUZ, 2006 [citado em 16 out. 2024]. Disponível em: <http://books.scielo.org/id/m7tn7>
- GARCIA-DEL PORTILLO, F.; FINLAY, B. B. *Salmonella* invasion of nonphagocytic cells induces formation of macropinosomes in the host cell. *Infect Immun.*, v. 62, n. 10, p. 4641–4645, out. 1994.
- GARDNER, M. J. et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, v. 419, n. 6906, p. 498–511, out. 2002.
- GARRIDO-CARDENAS, J. A.; GONZÁLEZ-CERÓN, L.; MANZANO-AGUGLIARO, F.; MESA-VALLE, C. *Plasmodium* genomics: an approach for learning about and ending human malaria. *Parasitol Res.*, v. 118, n. 1, p. 1–27, jan. 2019.
- GAUTHIER, J.; VINCENT, A. T.; CHARETTE, S. J.; DEROME, N. A brief history of bioinformatics. *Brief Bioinform.*, v. 20, n. 6, p. 1981–1996, 27 nov. 2019.
- GILLETTE, D. D. et al. Analysis of Human Bronchial Epithelial Cell Proinflammatory Response to *Burkholderia cenocepacia* Infection. *J Biol Chem.*, v. 288, n. 6, p. 3691–3695, fev. 2013.
- GOFFEAU, A. et al. Life with 6000 Genes. *Science*, v. 274, n. 5287, p. 546–567, 25 out. 1996.
- GOLDSMITH-FISCHMAN, S.; HONIG, B. Structural genomics: Computational methods for structure analysis. *Protein Sci.*, v. 12, n. 9, p. 1813–1821, set. 2003.
- GOMES, L. G. R. et al. In Silico Designed Multi-Epitope Immunogen “Tpme-VAC/LGCM-2022” May Induce Both Cellular and Humoral Immunity against *Treponema pallidum* Infection. *Vaccines*, v. 10, n. 7, p. 1019, 25 jun. 2022.
- GRANT, A. et al. Zika Virus Targets Human STAT2 to Inhibit Type I Interferon

- Signaling. **Cell Host Microbe**, v. 19, n. 6, p. 882–890, jun. 2016.
- HAY, A. J.; MCCAULEY, J. W. The WHO global influenza surveillance and response system (GISRS)—A future perspective. **Influenza Other Respir Viruses**, v. 12, n. 5, p. 551–557, set. 2018.
- HAYASHI, S.; OGURA, Y. ERK signaling dynamics in the morphogenesis and homeostasis of *Drosophila*. *Curr Opin Genet Dev.*, v. 63, p. 9–15, ago. 2020.
- HE, Q. et al. A graph-based genome and pan-genome variation of the model plant *Setaria*. *Nat Genet.*, v. 55, n. 7, p. 1232–1242, jul. 2023.
- HESS, J. F. et al. Library preparation for next generation sequencing: A review of automation strategies. **Biotechnol Adv.**, v. 41, p. 107537, jul. 2020.
- HU, T.; CHITNIS, N.; MONOS, D.; DINH, A. Next-generation sequencing technologies: An overview. *Hum Immunol.*, v. 82, n. 11, p. 801–811, nov. 2021.
- HYUN, J. C.; MONK, J. M.; SZUBIN, R.; HEFNER, Y.; PALSSON, B. O. Global pathogenomic analysis identifies known and candidate genetic antimicrobial resistance determinants in twelve species. **Nat Commun.**, v. 14, n. 1, p. 7690, 24 nov. 2023.
- IBARZ PAVÓN, A. B.; MAIDEN, M. C. J. Multilocus Sequence Typing. In: CAUGANT, D. A. (Ed.). **Molecular Epidemiology of Microorganisms** [Internet]. Totowa, NJ: Humana Press, 2009 [citado em 16 out. 2024]. p. 129–140. (Methods in Molecular Biology; v. 551). Disponível em: http://link.springer.com/10.1007/978-1-60327-999-4_11
- INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM. Finishing the euchromatic sequence of the human genome. **Nature**, v. 431, n. 7011, p. 931–945, out. 2004.
- IOC J. Genômica funcional: desafios [Internet]. IOC, 2010 [citado em 4 out. 2024]. Disponível em: <https://www.ioc.fiocruz.br/noticias/genomica-funcional-desafios>
- IRAOLA, G.; BOYCE, J. D. Understanding Pathogenesis Through Pathogenomics and Bioinformatics. In: PRESCOTT, J. F.; RYCROFT, A. N.; BOYCE, J. D.; MACINNES, J. I.; VAN IMMERSEEL, F.; VÁZQUEZ-BOLAND, J. A. (Eds.). **Pathogenesis of Bacterial Infections in Animals** [Internet]. 1. ed. Wiley, 2022 [citado em 16 out. 2024]. p. 32–56. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/9781119754862.ch3>
- JAIN, C.; RODRIGUEZ-R, L. M.; PHILLIPPY, A. M.; KONSTANTINIDIS, K. T.; ALURU, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. **Nat Commun.**, v. 9, n. 1, p. 5114, 30 nov. 2018.
- JUHAS, M.; EBERL, L.; CHURCH, G. M. Essential genes as antimicrobial

- targets and cornerstones of synthetic biology. **Trends Biotechnol.**, v. 30, n. 11, p. 601–607, nov. 2012.
- KATOH, K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.*, v. 30, n. 14, p. 3059–3066, 15 jul. 2002.
- KIRKCALDY, R. D. et al. **Trichomonas vaginalis** Antimicrobial Drug Resistance in 6 US Cities, STD Surveillance Network, 2009–2010. **Emerg Infect Dis.**, v. 18, n. 6, p. 939–943, jun. 2012.
- KLINE, K. A.; FÄLKER, S.; DAHLBERG, S.; NORMARK, S.; HENRIQUES-NORMARK, B. Bacterial Adhesins in Host-Microbe Interactions. **Cell Host Microbe**, v. 5, n. 6, p. 580–592, jun. 2009.
- KOLACZKOWSKA, E.; KUBES, P. Neutrophil recruitment and function in health and inflammation. *Nat Rev Immunol.*, v. 13, n. 3, p. 159–175, mar. 2013.
- KOMISSAROV, A. B. et al. Genomic epidemiology of the early stages of SARS-CoV-2 outbreak in Russia [Internet]. 2020 [citado em 16 out. 2024]. Disponível em: <http://medrxiv.org/lookup/doi/10.1101/2020.07.14.20150979>
- KUMAR JAISWAL, A.; TIWARI, S.; JAMAL, S.; BARH, D.; AZEVEDO, V.; SOARES, S. An In Silico Identification of Common Putative Vaccine Candidates against **Treponema pallidum**: A Reverse Vaccinology and Subtractive Genomics Based Approach. **Int J Mol Sci.**, v. 18, n. 2, p. 402, 14 fev. 2017.
- LAGE, F. S. D.; SANTOS, F. B. Sequenciamento de primeira geração: método de Sanger. In: MARIANO, D.; DEZORDI, F. Z.; MARTINS, P.; XAVIER, J.; SOUSA, T. D. J.; LIMA, L. et al. (Eds.). **BIOINFO - Revista Brasileira de Bioinformática e Biologia Computacional** [Internet]. 1. ed. Alfahelix, 2021 [citado em 16 out. 2024]. Disponível em: <https://bioinfo.com.br/sequenciamento-de-primeira-geracao-metodo-de-sanger/>
- LARSON, N. B.; OBERG, A. L.; ADJEI, A. A.; WANG, L. A Clinician’s Guide to Bioinformatics for Next-Generation Sequencing. **J Thorac Oncol.**, v. 18, n. 2, p. 143–157, fev. 2023.
- LATA, K. S. et al. Whole genome sequencing and de novo assembly of three virulent Indian isolates of **Leptospira**. **Infect Genet Evol.**, v. 85, p. 104579, nov. 2020.
- LEÃO, U. S. Sequenciamento de Nova Geração (NGS) de *Urochloa mosambicensis*: uma forrageira promissora para o semiárido. *RENORBIO 2017.*, 2017 [citado em 4 out. 2024]. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/handle/doc/1080722>
- LEE, S. A. et al. Ortholog-based protein-protein interaction prediction and its application to inter-species interactions. **BMC Bioinformatics**, v. 9, n. S12,

p. S11, dez. 2008.

- LEFEBVRE, C. et al. Discovery of a novel dehydratase of the fatty acid synthase type II critical for ketomycolic acid biosynthesis and virulence of *Mycobacterium tuberculosis*. **Sci Rep.**, v. 10, n. 1, p. 2112, 7 fev. 2020.
- LEFTEROVA, M. I.; SUAREZ, C. J.; BANAEI, N.; PINSKY, B. A. Next-Generation Sequencing for Infectious Disease Diagnosis and Management. *J Mol Diagn.*, v. 17, n. 6, p. 623–634, nov. 2015.
- LEITÃO, J. H. Microbial Virulence Factors. *Int J Mol Sci.*, v. 21, n. 15, p. 5320, 27 jul. 2020.
- LEVA-BUENO, J.; PEYMAN, S. A.; MILLNER, P. A. A review on impedimetric immunosensors for pathogen and biomarker detection. **Med Microbiol Immunol (Berl)**, v. 209, n. 3, p. 343–362, jun. 2020.
- LI, B. sheng et al. Genomic Evolution and Variation of SARS-CoV-2 in the Early Phase of COVID-19 Pandemic in Guangdong Province, China. *Curr Med Sci.*, v. 41, n. 2, p. 228–235, abr. 2021.
- LI, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. **Bioinformatics**, v. 27, n. 21, p. 2987–2993, 1 nov. 2011.
- LI, T. et al. CRISPR/Cas9 therapeutics: progress and prospects. *Signal Transduct Target Ther.*, v. 8, n. 1, p. 36, 16 jan. 2023.
- LI, W. et al. Plant pan-genomics: recent advances, new challenges, and roads ahead. **J Genet Genomics.**, v. 49, n. 9, p. 833–846, set. 2022.
- LIANG, Y.; LUO, H.; LIN, Y.; GAO, F. Recent advances in the characterization of essential genes and development of a database of essential genes. **iMeta**, v. 3, n. 1, p. e157, fev. 2024.
- LINS, R. A.; ASSEF NETO, R. C. D. S.; SILVA, C. H. D.; GUIMARÃES, M. C. S. Coronavírus: um panorama sobre a colaboração científica internacional do Brasil ao longo da história. **Rev Eletrônica Comun Informação E Inov Em Saúde** [Internet], v. 15, n. 4, 10 nov. 2021 [citado em 16 out. 2024]. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/2398>
- LIU, S.; MOON, C. D.; ZHENG, N.; HUWS, S.; ZHAO, S.; WANG, J. Opportunities and challenges of using metagenomic data to bring uncultured microbes into cultivation. **Microbiome**, v. 10, n. 1, p. 76, 12 maio 2022.
- LU, R.; POPOV, V.; PATEL, J.; EAVES-PYLES, T. *Burkholderia mallei* and *Burkholderia pseudomallei* stimulate differential inflammatory responses from human alveolar type II cells (ATII) and macrophages. *Front Cell Infect Microbiol* [Internet], v. 2, 2012 [citado em 16 out. 2024]. Disponível em: <http://journal.frontiersin.org/article/10.3389/fcimb.2012.00165/abstract>

- LYNCH, F. et al. Australian public perspectives on genomic data storage and sharing: Benefits, concerns and access preferences. *Eur J Med Genet.*, v. 66, n. 1, p. 104676, jan. 2023.
- LYNCH, T. et al. A Primer on Infectious Disease Bacterial Genomics. *Clin Microbiol Rev.*, v. 29, n. 4, p. 881–913, out. 2016.
- MAIDEN, M. C. J. et al. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. **Proc Natl Acad Sci.**, v. 95, n. 6, p. 3140–3145, 17 mar. 1998.
- MANIRUZZAMAN, M.; NOKHODCHI, A. Continuous manufacturing via hot-melt extrusion and scale up: regulatory matters. *Drug Discov Today*, v. 22, n. 2, p. 340–351, fev. 2017.
- Mapping and Sequencing the Human Genome [Internet]. Washington, D.C.: National Academies Press, 1988 [citado em 16 out. 2024]. Disponível em: <http://www.nap.edu/catalog/1097>
- MATHAIYAN, J.; CHANDRASEKARAN, A.; DAVIS, S. Ethics of genomic research. *Perspect Clin Res.*, v. 4, n. 1, p. 100, 2013.
- MATTHEWS, L. R. et al. Identification of Potential Interaction Networks Using Sequence-Based Searches for Conserved Protein-Protein Interactions or “Interologs.” **Genome Res.**, v. 11, n. 12, p. 2120–2126, 1 dez. 2001.
- MEIER-KOLTHOFF, J. P.; GÖKER, M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. **Nat Commun.**, v. 10, n. 1, p. 2182, 16 maio 2019.
- MILLER, J. R.; KOREN, S.; SUTTON, G. Assembly algorithms for next-generation sequencing data. *Genomics*, v. 95, n. 6, p. 315–327, jun. 2010.
- MOORE, K. L. et al. P-selectin glycoprotein ligand-1 mediates rolling of human neutrophils on P-selectin. *J Cell Biol.*, v. 128, n. 4, p. 661–671, 15 fev. 1995.
- MOSHIRI, N. et al. The ViReflow pipeline enables user friendly large scale viral consensus genome reconstruction. **Sci Rep.**, v. 12, n. 1, p. 5077, 24 mar. 2022.
- MULLAN, L. Pairwise sequence alignment—it’s all about us! *Brief Bioinform.*, v. 7, n. 1, p. 113–115, mar. 2006.
- NAFEA, A. M. et al. Application of next-generation sequencing to identify different pathogens. **Front Microbiol.**, v. 14, p. 1329330, 29 jan. 2024.
- NAIDENOV, B. et al. Pan-Genomic and Polymorphic Driven Prediction of Antibiotic Resistance in **Elizabethkingia**. **Front Microbiol.**, v. 10, p. 1446, 4 jul. 2019.
- NATIONAL RESEARCH COUNCIL (U.S.). **New science of metagenomics:*

- revealing the secrets of our microbial planet*. Washington, DC: National Academies Press, 2007. 1 p.
- NEEDLEMAN, S. B.; WUNSCH, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol.*, v. 48, n. 3, p. 443–453, mar. 1970.
- NEHRA, M.; KUMAR, V.; KUMAR, R.; DILBAGHI, N.; KUMAR, S. Current Scenario of Pathogen Detection Techniques in Agro-Food Sector. **Biosensors**, v. 12, n. 7, p. 489, 4 jul. 2022.
- NIH. A Brief Guide to Genomics [Internet]. genome.gov. 2022 [citado em 4 out. 2024]. Disponível em: <https://www.genome.gov/about-genomics/fact-sheets/A-Brief-Guide-to-Genomics#:~:text=Genomics%20is%20the%20study%20of,and%20with%20the%20person's%20environment>
- ODEH, M. et al. iOntoBioethics: A Framework for the Agile Development of Bioethics Ontologies in Pandemics, Applied to COVID-19. *Front Med.*, v. 8, p. 619978, 21 maio 2021.
- OLI, A. N. et al. Immunoinformatics and Vaccine Development: An Overview. **ImmunoTargets Ther.**, v. 9, p. 13–30, fev. 2020.
- PAIS, F. S. M.; RUY, P. D. C.; OLIVEIRA, G.; COIMBRA, R. S. Assessing the efficiency of multiple sequence alignment programs. *Algorithms Mol Biol.*, v. 9, n. 1, p. 4, dez. 2014.
- PAKBIN, B.; BRÜCK, W. M.; ROSSEN, J. W. A. Virulence Factors of Enteric Pathogenic *Escherichia coli*: A Review. **Int J Mol Sci.**, v. 22, n. 18, p. 9922, 14 set. 2021.
- PALLEN, M. J.; WREN, B. W. Bacterial pathogenomics. **Nature**, v. 449, n. 7164, p. 835–842, out. 2007.
- PAPA MZE, N. et al. High Throughput SARS-CoV-2 Genome Sequencing from 384 Respiratory Samples Using the Illumina COVIDSeq Protocol. **Genes**, v. 14, n. 3, p. 681, 9 mar. 2023.
- PAREEK, C. S.; SMOCZYNSKI, R.; TRETYN, A. Sequencing technologies and genome sequencing. **J Appl Genet.**, v. 52, n. 4, p. 413–435, nov. 2011.
- PATON, J. C.; TRAPPETTI, C. **Streptococcus pneumoniae** Capsular Polysaccharide. FISCHETTI, V. A.; NOVICK, R. P.; FERRETTI, J. J.; PORTNOY, D. A.; BRAUNSTEIN, M.; ROOD, J. I. (Eds.). **Microbiol Spectr.**, v. 7, n. 2, p. 7.2.33, 12 abr. 2019.
- PEREIRA, R.; OLIVEIRA, J.; SOUSA, M. Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. *J Clin Med.*, v. 9, n. 1, p. 132, 3 jan. 2020.
- PIDDOCK, L. J. V. The global antibiotic research and development partnership

- (GARDP): researching and developing new antibiotics to meet global public health needs. **MedChemComm.**, v. 10, n. 8, p. 1227–1230, 2019.
- POLLARD, M. O.; GURDASANI, D.; MENTZER, A. J.; PORTER, T.; SANDHU, M. S. Long reads: their purpose and place. **Hum Mol Genet.**, v. 27, n. R2, p. R234–R241, 1 ago. 2018.
- POPLIN, R. et al. A universal SNP and small-indel variant caller using deep neural networks. **Nat Biotechnol.**, v. 36, n. 10, p. 983–987, nov. 2018.
- PRADEL, G.; TEMPLETON, T. J. Genomics of Pathogenic Parasites. In: HACKER, J.; DOBRINDT, U. (Eds.). **Pathogenomics** [Internet]. 1. ed. Wiley, 2006 [citado em 16 out. 2024]. p. 417–444. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/352760801X.ch19>
- PRASAD, A.; ENE, A.; JABLONSKA, S.; DU, J.; WOLFE, A. J.; PUTONTI, C. Comparative Genomic Study of *Streptococcus anginosus* Reveals Distinct Group of Urinary Strains. RAO, K. (Ed.). *mSphere*, v. 8, n. 2, 20 abr. 2023.
- RAHMAN, M. d. M.; MASUM, M. d. H. U.; WAJED, S.; TALUKDER, A. A comprehensive review on COVID-19 vaccines: development, effectiveness, adverse effects, distribution and challenges. *VirusDisease*, v. 33, n. 1, p. 1–22, mar. 2022.
- RODRIGUEZ, R.; KRISHNAN, Y. The chemistry of next-generation sequencing. **Nat Biotechnol.**, v. 41, n. 12, p. 1709–1715, dez. 2023.
- ROSCONI, F. et al. A bacterial pan-genome makes gene essentiality strain-dependent and evolvable. **Nat Microbiol.**, v. 7, n. 10, p. 1580–1592, 12 set. 2022.
- ROULI, L.; MERHEJ, V.; FOURNIER, P. E.; RAOULT, D. The bacterial pangenome as a new tool for analysing pathogenic bacteria. **New Microbes New Infect.**, v. 7, p. 72–85, set. 2015.
- ROUSH, S. W. Historical Comparisons of Morbidity and Mortality for Vaccine-Preventable Diseases in the United States. *JAMA*, v. 298, n. 18, p. 2155, 14 nov. 2007.
- SADLER, A. J.; WILLIAMS, B. R. G. Interferon-inducible antiviral effectors. *Nat Rev Immunol.*, v. 8, n. 7, p. 559–568, jul. 2008.
- SAMEEK, R.; CHINNAIYAN, A. M. Translating Genomics for Precision Cancer Medicine. *Annu Rev Genomics Hum Genet.*, v. 15, n. 1, p. 395–415, 31 ago. 2014.
- SAMUEL, C. E. Antiviral Actions of Interferons. *Clin Microbiol Rev.*, v. 14, n. 4, p. 778–809, out. 2001.
- SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain-terminating inhibitors. **Proc Natl Acad Sci.**, v. 74, n. 12, p. 5463–5467, dez. 1977.

- SATAM, H. et al. Next-Generation Sequencing Technology: Current Trends and Advancements. **Biology**, v. 12, n. 7, p. 997, 13 jul. 2023.
- SAYERS, E. W. et al. Database resources of the National Center for Biotechnology Information in 2023. **Nucleic Acids Res.**, v. 51, n. D1, p. D29–D38, 6 jan. 2023.
- SEAH, Y. M. et al. In Silico Evaluation of Variant Calling Methods for Bacterial Whole-Genome Sequencing Assays. DEKKER, J. P. (Ed.). **J Clin Microbiol.**, 10 jul. 2023.
- SEGERMAN, B. The Most Frequently Used Sequencing Technologies and Assembly Methods in Different Time Segments of the Bacterial Surveillance and RefSeq Genome Databases. **Front Cell Infect Microbiol.**, v. 10, p. 527102, 19 out. 2020.
- SELA, I.; WOLF, Y. I.; KOONIN, E. V. Genome plasticity, a key factor of evolution in prokaryotes [Internet]. 2018 [citado em 16 out. 2024]. Disponível em: <http://biorxiv.org/lookup/doi/10.1101/357400>
- SEN, R.; NAYAK, L.; DE, R. K. A review on host–pathogen interactions: classification and prediction. *Eur J Clin Microbiol Infect Dis.*, v. 35, n. 10, p. 1581–1599, out. 2016.
- SEREIKA, M. et al. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat Methods*, v. 19, n. 7, p. 823–826, jul. 2022.
- SHARMA, A.; GARG, A.; RAMANA, J.; GUPTA, D. VirulentPred 2.0: An improved method for prediction of virulent proteins in bacterial pathogens. **Protein Sci.**, v. 32, n. 12, p. e4808, dez. 2023.
- SHARMA, N. C.; EFSTRATIOU, A.; MOKROUSOV, I.; MUTREJA, A.; DAS, B.; RAMAMURTHY, T. Diphtheria. **Nat Rev Dis Primer**, v. 5, n. 1, p. 81, 5 dez. 2019.
- SIEVERS, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.*, v. 7, n. 1, p. 539, jan. 2011.
- SILVA, H. P. D.; LIMA, L. D. D. Política, economia e saúde: lições da COVID-19. **Cad Saúde Pública**, v. 37, n. 9, p. e00200221, 2021.
- SILVA, R. C. C. D.; ALVES, M. C. S. O uso de ferramentas de bioinformática para análise de dados genéticos: uma revisão. *Sci Electron Arch [Internet]*, v. 17, n. 1, 11 dez. 2023 [citado em 16 out. 2024]. Disponível em: <https://sea.ufr.edu.br/SEA/article/view/1872>
- SIM, S. H. et al. Innate Immune Responses of Pulmonary Epithelial Cells to Burkholderia pseudomallei Infection. NG, L. F. P. (Ed.). *PLoS ONE*, v. 4, n. 10, p. e7308, 6 out. 2009.

- SMITH, T. F.; WATERMAN, M. S. Identification of common molecular subsequences. *J Mol Biol.*, v. 147, n. 1, p. 195–197, mar. 1981.
- SOUZA, A. L. F. D.; BRUSAMARELLO, L. C. C. Sequenciamento de DNA: decifrando o manual de instruções dos seres vivos. **Genética Na Esc.**, v. 4, n. 1, p. 45–52, 20 maio 2009.
- STARK, R.; GRZELAK, M.; HADFIELD, J. RNA sequencing: the teenage years. **Nat Rev Genet.**, v. 20, n. 11, p. 631–65
- STRUELENS, M. J. et al. Real-time genomic surveillance for enhanced control of infectious diseases and antimicrobial resistance. **Front Sci.**, v. 2, p. 1298248, 25 abr. 2024.
- SU, C. M.; DU, Y.; ROWLAND, R. R. R.; WANG, Q.; YOO, D. Reprogramming viral immune evasion for a rational design of next-generation vaccines for RNA viruses. *Front Immunol.*, v. 14, p. 1172000, 17 abr. 2023.
- SUGIMAN-MARANGOS, S. N.; GILL, S. K.; MANSFIELD, M. J.; ORRELL, K. E.; DOXEY, A. C.; MELNYK, R. A. Structures of distant diphtheria toxin homologs reveal functional determinants of an evolutionarily conserved toxin scaffold. **Commun Biol.**, v. 5, n. 1, p. 375, 19 abr. 2022.
- TALBIRD, S. E. et al. Impact of Routine Childhood Immunization in Reducing Vaccine-Preventable Diseases in the United States. *Pediatrics*, v. 150, n. 3, p. e2021056013, 1 set. 2022.
- TAM, K.; TORRES, V. J. **Staphylococcus aureus* Secreted Toxins and Extracellular Enzymes*. FISCHETTI, V. A.; NOVICK, R. P.; FERRETTI, J. J.; PORTNOY, D. A.; BRAUNSTEIN, M.; ROOD, J. I. (Eds.). **Microbiol Spectr.**, v. 7, n. 2, p. 7.2.16, 12 abr. 2019.
- TETTELIN, H. et al. Genome analysis of multiple pathogenic isolates of **Streptococcus agalactiae**: Implications for the microbial “pan-genome.” **Proc Natl Acad Sci.**, v. 102, n. 39, p. 13950–13955, 27 set. 2005.
- THIEU, T.; JOSHI, S.; WARREN, S.; KORKIN, D. Literature mining of host–pathogen interactions: comparing feature-based supervised learning and language-based approaches. **Bioinformatics**, v. 28, n. 6, p. 867–875, 15 mar. 2012.
- TRALAMAZZA, S. M.; GLUCK-THALER, E.; FEURTEY, A.; CROLL, D. Copy number variation introduced by a massive mobile element facilitates global thermal adaptation in a fungal wheat pathogen. **Nat Commun.**, v. 15, n. 1, p. 5728, 8 jul. 2024.
- TRAMONT, E. C. Adhesion of **Neisseria Gonorrhoeae** and Disease. In: ELLIOTT, K.; O’CONNOR, M.; WHELAN, J. (Eds.). **Novartis Foundation Symposia** [Internet]. Chichester, UK: John Wiley & Sons, Ltd., 2008 [citado em 16 out. 2024]. p. 188–201. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/9780470720639.ch12>

- VASHISHT, V. et al. Genomics for Emerging Pathogen Identification and Monitoring: Prospects and Obstacles. **BioMedInformatics**, v. 3, n. 4, p. 1145–1177, 7 dez. 2023.
- VESTERGAARD, L. K.; OLIVEIRA, D. N. P.; HØGDALL, C. K.; HØGDALL, E. V. Next Generation Sequencing Technology in the Clinic and Its Challenges. **Cancers**, v. 13, n. 8, p. 1751, 7 abr. 2021.
- WANG, F.; MENG, W.; WANG, B.; QIAO, L. **Helicobacter pylori**-induced gastric inflammation and gastric cancer. **Cancer Lett.**, v. 345, n. 2, p. 196–202, abr. 2014.
- WANG, Y. et al. Linking genomic and epidemiologic information to advance the study of COVID-19. **Sci Data.**, v. 9, n. 1, p. 121, 30 mar. 2022.
- WANG, Y.; ZHAO, Y.; BOLLAS, A.; WANG, Y.; AU, K. F. Nanopore sequencing technology, bioinformatics and applications. **Nat Biotechnol.**, v. 39, n. 11, p. 1348–1365, nov. 2021.
- WANGMO, T. et al. An update on the “empirical turn” in bioethics: analysis of empirical research in nine bioethics journals. *BMC Med Ethics*, v. 19, n. 1, p. 6, dez. 2018.
- WHO. Antimicrobial Resistance [Internet]. WHO Fact-sheets, 2023 [citado em 4 out. 2024]. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance>
- WHO. Coronavirus disease (COVID-19) pandemic [Internet]. WHO Emergencies, 2024 [citado em 2024]. Disponível em: <https://www.who.int/europe/emergencies/situations/covid-19>
- WILM, A. et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. **Nucleic Acids Res.**, v. 40, n. 22, p. 11189–11201, 1 dez. 2012.
- WONG, D. W. S. Whole Genome and Next Generation Sequencing. In: *The ABCs of Gene Cloning* [Internet]. Cham: Springer International Publishing, 2018 [citado em 16 out. 2024]. p. 219–230. Disponível em: http://link.springer.com/10.1007/978-3-319-77982-9_24
- WU, L.; WANG, H.; XIA, Y.; XI, R. CNV-BAC: Copy number Variation Detection in Bacterial Circular Genome. *BIROL*, I. (Ed.). **Bioinformatics**, v. 36, n. 12, p. 3890–3891, 1 jun. 2020.
- XIAO, T.; ZHOU, W. The third generation sequencing: the advanced approach to genetic diseases. **Transl Pediatr.**, v. 9, n. 2, p. 163–173, abr. 2020.
- XU, C. A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. **Comput Struct Biotechnol J.**, v. 16, p. 15–24, 2018.

- XU, Q.; TANG, Y.; HUANG, G. Innate immune responses in RNA viral infection. *Front Med.*, v. 15, n. 3, p. 333–346, jun. 2021.
- YALÇIN, S. et al. Genomic surveillance during the first two years of the COVID-19 pandemic – country experience and lessons learned from Türkiye. **Front Public Health.**, v. 12, p. 1332109, 24 maio 2024.
- YOSHIDA, H.; BOGAKI, M.; NAKAMURA, M.; NAKAMURA, S. Quinolone resistance-determining region in the DNA gyrase *gyrA* gene of **Escherichia coli**. **Antimicrob Agents Chemother.**, v. 34, n. 6, p. 1271–1272, jun. 1990.
- YUAN, L.; LU, H.; LI, F.; NIELSEN, J.; KERKHOVEN, E. J. HGTphyloDetect: facilitating the identification and phylogenetic analysis of horizontal gene transfer. **Brief Bioinform.**, v. 24, n. 2, p. bbad035, 19 mar. 2023.
- ZENG, D.; JIAO, J.; MO, T. Combination of nucleic acid amplification and CRISPR/Cas technology in pathogen detection. **Front Microbiol.**, v. 15, p. 1355234, 6 fev. 2024.
- ZHANG, T. et al. Complex genome assembly based on long-read sequencing. **Brief Bioinform.**, v. 23, n. 5, p. bbac305, 20 set. 2022.
- ZHU, Y. et al. RNA-based therapeutics: an overview and prospectus. *Cell Death Dis.*, v. 13, n. 7, p. 644, 23 jul. 2022.

Autor

Lucas Gabriel Rodrigues Gomes¹, Eduarda Guimarães Sousa¹, Fernanda Diniz Prates^{1,2}, Talita Pereira Gomes^{1,3}, Gabriel Camargos Gomes¹, Janaíne Aparecida de Paula^{1,4}, Pedro Henrique Marques⁴, Ana Lua de Oliveira Vinhal¹, Bernardo Buhr Alves Mendonça¹, Aline Ferreira Maciel de Oliveira¹, Mariana Letícia Costa Pedrosa¹, Luiza Pereira Reis¹, Marcus Vinicius Canário Viana¹, Siomar de Castro Soares⁴, Sandeep Tiwari⁵, Arun Kumar Jaiswal¹, Vasco Azevedo¹

1. Laboratório de Genética Celular e Molecular, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, 31270-901, Minas Gerais, Brasil.

2. Núcleo Técnico Operacional, Microbiologia, Instituto Hermes Pardini, Vespasiano, Minas Gerais, Brasil.
3. Instituto Nacional de Ciência e Tecnologia em Estudos Interdisciplinares e Transdisciplinares em Ecologia e Evolução (INCT IN-TREE), Instituto de Biologia, Universidade Federal da Bahia, Salvador, Bahia, Brasil.
4. Laboratório de Bioinformática, Departamento de Microbiologia, Imunologia e Parasitologia, Instituto de Ciências Biológicas e Naturais, Universidade Federal do Triângulo Mineiro, Uberaba, Minas Gerais, Brasil.
5. Instituto de Biologia, Universidade Federal da Bahia, Salvador, Bahia, Brasil.