1. Introduction
   1. Importance of mapping omics data

O mapeamento de dados ômicos visa entender a complexa rede de interações biológicas, assumindo um papel cada vez mais preponderante na ciência atual.

As comunidades microbianas são compostas por bactérias, archaea, fungos, leveduras, eucariotas e vírus, que frequentemente coabitam em conjunto num mesmo habitat. Com esta dimensão e complexidade das comunidades microbianas, os dados ómicos que delas resultam, tornam a análise desses mesmos dados numa tarefa difícil (Sequeira et al.), sendo necessária uma abordagem diferente para o tratamento desses dados, pois a diversidade total da biosfera não pode ser estudada se não for enquadrada num meta-contexto adequado. Deste modo, as tecnologias de mapeamento de dados ómicos assumem uma importância relevante na interpretação de resultados ómicos.

Os dados laboratoriais das ciências ómicas ganharam um interesse generalizado entre os investigadores nos últimos anos devido à sua complexidade, disponibilidade e potencial para gerar novos conhecimentos médicos (ou seja, compreender os processos biológicos subjacentes) (Toussaint et al.). Deste modo, o mapeamento de dados ómicos oferece oportunidades atrativas para compreender a complexidade, e obter informações sobre os processos biológicos subjacentes.

* 1. Software developed for omics and meta-omics analyses

As tecnologias ómicas e meta-ómicas são abordagens poderosas para explorar as funções dos microrganismos (Sequeira et al.), bem como os dados laboratoriais das ciências ómicas (Toussaint et al.).

O software desenvolvido para análises ómicas e meta-ómicas, juntamente com bases de conhecimento que incluem informação sobre genes, proteínas, anotação taxonómica e funcional, entre outros tipos de informação, são recursos valiosos para a análise de dados ómicos (Sequeira et al.). Estão disponíveis vários recursos bioinformáticos e interfaces Web para análises meta-ómicas.

Durante a última década, o mapeamento de dados ómicos surgiu como uma ajuda preciosa para compreender os dados gerados por várias tecnologias "ómicas". Como resultado, foram desenvolvidas várias ferramentas de software robustas, bem como interfaces Web para análises meta-ómicas, de modo a apoiar a análise de vias para estudos genómicos e proteómicos (Xia & Wishart).

* 1. Current challenges in mapping omics data

O software desenvolvido para análises ómicas e meta-ómicas, juntamente com bases de conhecimento que incluem informação sobre genes, proteínas, anotação taxonómica e funcional, entre outros tipos de informação (Sequeira et al.), tornam-se em recursos poderosos para análises ómicas. Deste modo, o novo panorama que se enfrenta exige o desenvolvimento de novas ferramentas de software (Filipo et al.), que convertem automaticamente os dados brutos em informações completas, através de mapas metabólicos.

Embora estejam disponíveis vários recursos bioinformáticos para análises meta-ómicas, muitos deles exigem conhecimentos computacionais significativos (Sequeira et al.) e uma grande capacidade computacional. As inter-faces Web são mais fáceis de utilizar, mas muitas vezes têm dificuldade em lidar com grandes ficheiros de dados (Sequeira et al.).

Uma das grandes vantagens de algumas ferramentas de software de mapeamento de dados ómicos, é a funcionalidade de identificação interativa, que permite aos utilizadores interagir com elementos específicos de uma via para uma exploração mais aprofundada ou para obter dados adicionais sobre genes, proteínas específicas ou até taxonomia de modo a identificar as relações evolutivas entre espécies. Esta caraterística é fundamental para os investigadores que pretendem compreender mais pormenorizadamente os seus dados experimentais.

* 1. Review of omics data analysis tools used in mapping pathways

Link Table

**Table 1.** The available analytical tools and web servers designed for mapping path- way data, including their functions, advantages, disadvantages, supported languages, packages, types of identifiers, interac-tivity, sample representation, taxonomy consid- erations, differential expression analysis capabili-ties, and input/output formats.

Com a disponibilidade de diagramas de vias de redes metabólicas e com um volume cada vez maior de dados ómicos a visualizar, foram desenvolvidas várias ferramentas para facilitar a interpretação dos resultados da anotação funcional e representar os genes ou proteínas identificadas nas vias metabólicas.

KEGGCharter (Sequeira et al.) é uma implementação de linha de comando do serviço de mapeamento do KEGG Pathway, obtendo também KOs adicionais e números EC, através dos métodos disponíveis no BioPython para aceder à API do KEGG. O KEGGCharter recebe como entrada uma tabela (TSV ou EXCEL), que contém os IDs KEGG, KOs ou números EC números, representa KOs identificados em mapas metabólicos e inclui informações sobre expressão genética diferencial. Quando são carregados dados de mais do que um organismo, o KEGGCharter liga a função à identificação taxonómica, que pode ser visualizada nos mapas. A expressão diferencial de genes/proteínas pode ser visualizada em mapas metabólicos, mostrando mini mapas de calor.

KEGG Mapper (Kanehisa et al.) é uma coleção de ferramentas de mapeamento de vias, BRITE e MÓDULOS. Fornece uma base de dados abrangente que integra informações funcionais sistêmicas com dados genômicos e químicos. KEGG Mapper (Kanehisa et al.) oferece funcionalidades interativas, possibilitando que os usuários explorem informações detalhadas sobre interações e funções moleculares. Suas ferramentas, como "Search&Color Pathway", permitem mapear dados de expressão gênica, indicando regulações positivas ou negativas (Kanehisa M. et al.).

MetPA (Xia & Wishart), é uma ferramenta desenhada para a visualização de dados metabolómicos. Utiliza pacotes como KEGGgraph, Graphviz e ImageMagick para renderizar redes metabólicas detalhadas, disponíveis numa interface web que suporta zoom sem perda de qualidade e manipulação dinâmica. A ferramenta cobre 11 modelos de organismos. As informações são visualizadas como uma rede numa interface web, proporcionando um ambiente de análise de vias metabólicas robusto e de fácil uso (Xia & Wishart).

DAVID (Dennis Jr. et al) é um software acessível pela web que integra anotações funcionais genômicas com resumos gráficos intuitivos, simplificando a anotação e sumarização rápida de dados conforme categorias compartilhadas, como Gene Ontology e domínios de proteínas. DAVID (Dennis Jr. et al) ajuda na interpretação de dados expressos diferencialmente.

CellDesigner 3.5 (Funahashi et al.), é uma ferramenta de modelagem para redes bioquímicas e gene-regulatórias, utilizando notação gráfica padronizada e SBML (Systems Biology Markup Language). Possui uma interface intuitiva e suporta vários identificadores biológicos. Não se aplica diretamente à representação de amostras de dados experimentais. Pode incluir expressões diferenciais de genes nos seus modelos. Exporta modelos em formatos gráficos como PNG e SVG (Funahashi et al.).

Da mesma forma, GenMAPP (Salomonis et al), é uma ferramenta que permite aos usuários visualizar e analisar dados de escala genômica, em vias biológicas. Implementada em Visual Basic 6.0, oferece suporte a múltiplas anotações de genes e espécies e permite a criação personalizada de bancos de dados para um número potencialmente ilimitado de espécies. Esta ferramenta integra processos de ETL para extração de dados de recursos públicos como Ensembl, Entrez Gene e Affymetrix, suportando uma ampla gama de identificadores, como IDs de genes Ensembl, IDs UniProt, IDs Entrez Gene e IDs de conjunto de sondas Affymetrix. Uma característica única é a coloração dinâmica dos genes nas vias com base em critérios definidos pelo usuário. GenMAPP (Salomonis et al) foi projetada para lidar com conjuntos de dados de escala de genoma e interpretar mudanças a nível de via numa ampla gama de organismos, e permite a visualização de dados de expressão gênica diferencial.

KGML-ED (Klukas & Schreiber), permite a visualização dinâmica, navegação interativa e edição de diagramas de vias do KEGG. Esta ferramenta suporta a exploração interativa e dinâmica das vias, permitindo a edição e criação de novas vias, facilitando a integração de dados específicos do usuário (Klukas & Schreiber). KGML-ED (Klukas & Schreiber) oferece técnicas de visualização semi-estáticas e dinâmicas para análises de vias melhoradas. Suporta identificadores KGML para genes, permitindo mapeamento e anotação detalhados dentro do framework KEGG. Possui capacidades interativas para navegar e modificar diagramas de vias, com opções de personalização conduzida pelo usuário como colorar componentes das vias (Klukas & Schreiber).

MetaCore™ (Ekins S. et al.) é uma plataforma web, que analisa os dados moleculares de alto rendimento e integra-os para visualização no contexto de mapas metabólicos. Com acesso pago e complexidade para novos usuários, usa Perl, HTML/JavaScript e Flash Player Plug-in. Aceita identificadores como LocusLink, SwissProt, RefSeq e Unigene. A interatividade é possível. Suporta análise de expressão gênica diferencial e aceita vários formatos de entrada e saída de dados, incluindo XML, JSON, CSV, TSV, PDF, PNG, JPEG, HTML, XLS e XLSX.

Pathway Tools (Karp. Et al., 2015) é projetado para simulação e visualização de coleções integradas de dados genômicos. Permite reconstruções metabólicas e previsões, além da visualização de interações regulatórias e análise de dados ômicos (Karp. Et al., 2015). Ao longo de suas versões, Pathway Tools aprimorou as suas capacidades e expandiu seu banco de dados, passando de 800 bases de dados de genoma/vias (PGDBs) (Karp. Et al., 2009) para mais de 20.000 (Karp. Et al., 2019). A aprendizagem da ferramenta pode ser íngreme para novos usuários. Pathway Tools (Karp. Et al., 2015) mantém a interatividade e a compatibilidade com diferentes formatos de entrada e saída de dados, como GenBank, SBML e BioPAX.

A ferramenta iPath (Yamada T. et al.) é uma ferramenta web para visualização e análise de vias celulares que fornece mapas interativos para o metabolismo central, biossíntese de metabólitos secundários e vias regulatórias (Yamada T. et al.). Oferece uma interface interativa e amigável para a exploração de uma ampla gama de dados de vias metabólicas, com a capacidade de mapear uma variedade de identificadores, como KEGG, e customizar mapas com dados como expressão diferencial. Ao longo das atualizações, iPath expandiu seu conjunto de dados e funcionalidades, incluindo novos módulos de vias KEGG, reações e espécies (Darzi Y. et al.). A versão mais recente suporta análise interativa (Darzi Y. et al.).

PathVisio (P van Iersel et al.) é uma ferramenta de software para visualizar, editar e analisar vias biológicas. Embora exija entrada manual de dados, o que pode ser demorado, PathVisio é reforçado pela sua capacidade de interagir com outras ferramentas científicas, como Cytoscape e Eu.Gene.

Pathview (Luo W. et al.) é um conjunto de ferramentas para integração e visualização de dados baseados em vias metabólicas, mapeando dados do usuário em gráficos de vias relevantes para facilitar a análise e interpretação. Como parte do projeto Bioconductor, ele integra-se com diversos pacotes R para análise de dados abrangente e suporta uma ampla gama de identificadores para genes/proteínas e compostos/metabólitos (Luo W. et al.). É interativo, lidando com conjuntos de dados de diferentes escalas e complexidade, e compatível com mais de 2000 espécies. Oferece suporte à visualização de expressão gênica diferencial.

Pathway Tools (Rahman S. et al.) é uma aplicação baseada em Java e oferece uma representação visual da rede bioquímica de um organismo, gerada automaticamente a partir de um Pathway/Genome Database (PGDB). A ferramenta é projetada para trabalhar com conjuntos de dados de organismos inteiros, integrando-se com o Omics Viewer para sobrepor dados de expressão gênica e outros dados quantitativos. Os aprimoramentos ao longo das versões incluem visualizações interativas na web e desktop, suporte a zoom semântico, geração de pôsteres e visualizações coloridas e animadas para dados de séries temporais (Paley S. & Karp P).

Reactome Knowledgebase (Fabregat et al.), é uma base de dados online que detalha processos celulares como transdução de sinal e replicação de DNA em uma rede de transformações moleculares ordenada. Permite visualização interativa de dados ômicos sobre diagramas de vias e oferece exportação de diagramas de vias para análise e personalização adicionais pelo usuário (Haw R. et al.).

As ferramentas CellDesigner e KGML-ED, são semelhantes mas não representam a taxonomia, o que as torna inválidas para quem quiserrelacionar dados metabolómicos com a taxonomia do organismo.

KEGG Mapper e Reactome oferecem funcionalidades profundas para exploração de dados genômicos e proteômicos, permitindo mapeamento detalhado de vias com ajustes visuais baseados em expressão gênica, diferindo do MetaCore™ (Ekins S. et al.) que, apesar de fornecer análises detalhadas, requer uma assinatura paga e tem uma curva de aprendizagem mais íngreme devido à sua complexidade. Já ferramentas como Pathway Tools (Rahman S. et al.) e MetPA, embora também focadas em vias metabólicas, divergem nas abordagens; Pathway Tools (Rahman S. et al.) permite simulações e predições detalhadas dentro de um ambiente integrado, enquanto MetPA especializa-se na visualização de dados metabolômicos com uma interface altamente interativa. Ferramentas como Pathview e PathVisio facilitam a integração e análise visual de dados sobre vias, sendo PathVisio reforçada pela sua capacidade de interação com outras ferramentas científicas, como Cytoscape. Em contraste, DAVID (Dennis Jr. et al) e KEGGCharter são mais focados na anotação funcional e sumarização de dados, com DAVID simplificando a visualização de categorias como Gene Ontology e KEGGCharter permitindo mapeamento direto através de linha de comando, o que torna esta ferramenta interessante do ponto de vista de personalização por parte do usuário. Essa diversidade de ferramentas oferece aos usuários uma grande gama de opções conforme suas necessidades específicas de análise e visualização.

Com exceção do KeggCharter, no global a maioria das ferramentas descritas apresentam interatividade, o que representa uma grande vantagem quando se fala sobre ferramentas de mapeamento de vias metabólicas, pois permite aos utilizadores a visualização dinâmica da informação nos mapas, dando assim a possibilidade de uma análise personalizada, com a integração de dados de taxonomia, expressão diferencial, entre outros, em simultâneo.

2. Motivation

2.1 KeggCharter

O KeggCharter é uma ferramenta que representa os resultados ómicos, incluindo [a expressão genética diferencial](https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/differential-gene-expression), nas vias metabólicas KEGG. Além disso, mostra a atribuição taxonómica das [enzimas](https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/enzyme) representadas, o que é particularmente útil em estudos metagenómicos em que estão presentes vários microrganismos. Contudo, os outputs gráficos que esta ferramenta produz são estáticos, fornecendo ao utilizador apenas a informação lá contida. Uma forma de contrariar essa desvantagem seria a integração de interatividade nos gráficos gerados pelo KeggCharter, de modo a dar a possibilidade ao leitor de ter uma análise personalizada da informação, com a integração de dados de taxonomia, expressão diferencial, entre outros, em simultâneo.

Uma maneira de introduzir a interatividade na ferramenta, seria implementar JavaScript para criar elementos interativos. Um exemplo prático disso será a KronaTools (<https://github.com/marbl/Krona/tree/master/KronaTools>), que gera gráficos Kronas, que oferecem uma maneira poderosa e intuitiva de visualizar e explorar dados hierárquicos, com interatividade que permite aos usuários analisar os dados em diferentes níveis de detalhe e personalizar a visualização conforme necessário

2.2 Objective of the study

Posto isto, o objetivo do projeto passa por alterar o KeggCharter de modo a fornecer outputs gráficos interativos, que entreguem ao utilizador a possibilidade de visualizar a informação em diferentes níveis, conforme a sua personalização. s gráficos do KEGGCharter serão expandidos para incluir a representação multi-nível da informação de expressão genética. O trabalho envolverá o desenvolvimento de novas representações interactivas sobre os gráficos atualmente produzidos, para permitir a inclusão de mais informação nos mapas metabólicos.

3. Methodology

Para atacar o problema da falta de interatividade nos gráficos gerados no output do KeggCharter, são propostos neste projeto dois diferentes workflows: workflow 1 vs workflow 2, descritos nas seguintes secções da metodologia.

3.1 Workflow 1

3.2 Workflow 2

4. Conclusion

4.1 Context and importance/relevance of the project

Na vanguarda da bioinformática, o projeto em análise insere-se no campo emergente das abordagens metagenómicas, uma área reconhecida pelo seu potencial em decifrar a complexidade e a dinâmica dos ecossistemas microbianos, bem como sua relevância no estudo da evolução de doenças. Neste contexto, a capacidade de mapear conjuntos de dados ómicos em vias metabólicas representa uma ferramenta indispensável na interpretação de interações complexas em redes genéticas e metabólicas, pois a análise de dados ómicos é bastante complexa, sendo bastante simplificada com a introdução de mapas metabólicos. A metagenómica abre portas para uma compreensão mais abrangente da ecologia microbiana, permitindo avanços significativos na ciência e na medicina, particularmente na identificação de novos biomarcadores e alvos terapêuticos.

Com o contributo proposto neste projeto, de forma a melhorar o KeggCharter introduzindo interatividade nos gráficos gerados pela ferramenta, está-se a dar a oportunidade aos usuários de integrarem as enumeras vantagens do KeggCharter num elemento interativo e ser mais fácil retirar conclusões dos seus dados ómicos presentes no mapa metabólico.