BRAZILIAN REPOSITORIUM FOR PROTEIN-PROTEIN INTERACTIONS (BRF- PPI)

Iandê de Freitas Richalski[1*], Bruno Rafael Florentino[2], Robson Parmezan Bonidia[1]
[1]Universidade Tecnológica Federal do Paraná, Campus Cornélio Procópio
[2]Universidade de São Paulo, Campus São Carlos
[*]e-mail: iande@alunos.utfpr.edu.br

Área Temática: 05 – Ciência/Engenharia da Computação/Software

Palavras-chave: interações proteína-proteína; interface web; bioinformática; visualização de redes; inteligência artificial

RESUMO

Contexto: Dados de interações proteína-proteína (PPIs, do inglês *Protein-protein interactions*) em plantas têm crescido com o desenvolvimento de ferramentas computacionais de predição. Os países do Mercosul têm alta capacidade produtiva, mas a região ainda enfrenta insegurança alimentar, e as mudanças climáticas tendem a agravar esse quadro[1]. Uma das estratégias para enfrentá-lo é inserir em cultivos proteínas de espécies que resistem naturalmente à seca e ao calor. Para que a nova característica funcione, essas proteínas precisam interagir com as proteínas da planta que as recebe. Mapear tais interações em laboratório, porém, leva tempo, custa caro e tem escala limitada[2]. Técnicas de Inteligência Artificial (IA) vêm sendo desenvolvidas para predizê-las, o que torna a busca por novas interações mais rápida e barata. O projeto BRF-PPI reúne duas frentes: um modelo preditivo, aplicado ao banco STRING[4] para reavaliar a confiabilidade das interações registradas em espécies do clado Viridiplantae[7], e a camada de aplicação que disponibiliza esses resultados.
Lacuna/Problema: O acesso organizado a essas informações ainda é um desafio para pesquisadores e profissionais da área. Parte das plataformas disponíveis exige familiaridade com ferramentas computacionais, o que limita a aplicação prática dos dados[2]. Bancos de dados de PPIs também apresentam dificuldades de integração, decorrentes do uso de identificadores distintos e de padrões de anotação inconsistentes[3]. Há ainda um segundo problema: conjuntos de dados preditos costumam ser distribuídos como pares de identificadores e pontuações, sem informação funcional associada. Nesse formato, avaliar se uma interação faz sentido exige consultar cada proteína em outras bases.
Objetivo: Este trabalho desenvolve a camada de aplicação do projeto BRF-PPI. Ela disponibiliza as predições em uma interface de consulta e acrescenta dois recursos que apoiam a interpretação dos resultados: a exploração em rede e a anotação funcional.
Método: A interface foi construída em Python com o framework Streamlit. Os dados ficam em arquivos Parquet, um por espécie, identificadas pelo taxon ID do banco STRING. A leitura usa a biblioteca Polars em modo sob demanda, de forma que apenas as porções relevantes de cada arquivo chegam à memória. A navegação está organizada em duas abas. A primeira permite consultar os dados em formato de tabela, com filtros por espécie, identificador e faixa do Score STRING. A segunda exibe a rede de uma proteína: o usuário informa um identificador e recebe o grafo de suas interações diretas, montado com NetworkX e desenhado com Plotly. A cor das arestas indica a categoria de confiança e a cor dos nós, o compartimento subcelular. A anotação funcional é obtida do banco UniProt[5], espécie por espécie, e reúne termos de Gene Ontology[6] e localização subcelular com vocabulário de onze compartimentos, para que ajuste em uma escala de cores legível. As duas abas permitem exportar os resultados em CSV, e a busca oferece ainda o download em ZIP dos conjuntos completos das espécies selecionadas. Na rede, cada proteína leva a sua entrada no UniProt por hiperlink. A plataforma está disponível em português, inglês e espanhol.
Resultados: A plataforma está operacional e reúne 24,6 milhões de registros distribuídos entre 124 espécies do clado Viridiplantae. Entre esses registros estão as mais de 3 milhões de novas interações de altíssima confiança preditas pelo projeto, que não constavam nessa categoria no banco original[7]; as demais foram reavaliadas nas categorias de alta e média confiança. Os identificadores do repositório correspondem aos

códigos de proteína do banco UniProt, o que permitiu integrar as duas bases sem etapa intermediária de mapeamento. Obteve-se assim anotação funcional para ao menos metade das proteínas em 103 dessas espécies. Na aba de rede, o pesquisador consulta uma proteína e vê, para cada parceiro, a confiança atribuída e o compartimento em que atua. A leitura sob demanda permitiu manter todas essas espécies disponíveis sem carregar os arquivos completos em memória.
Conclusão: O BRF-PPI reúne em uma única plataforma as interações do clado Viridiplantae, organizadas por categoria de confiança. Com isso, facilita o acesso a esses dados e os aproxima das pesquisas de melhoramento genético voltadas a cultivos capazes de suportar condições climáticas adversas. Com a visualização em rede e a anotação funcional, o pesquisador não vê apenas quais proteínas interagem, mas também onde e com que função atuam, o que ajuda a avaliar se a interação faz sentido. Os próximos passos são publicar a plataforma em servidor aberto, disponibilizar os dados para download e testá-la com pesquisadores da área. Espera-se assim acelerar o estudo dos mecanismos de interação proteica em espécies ainda pouco caracterizadas.

AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Universidade Tecnológica Federal do Paraná (UTFPR) pelo apoio à iniciação científica. Ao Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC/USP) pela colaboração no desenvolvimento do projeto. Este trabalho também foi financiado pelo *Canadian International Development Research Centre* (IDRC) e pelo *UK government's Foreign, Commonwealth and Development Office*.

REFERÊNCIAS

[1] ZOLOTNYTSKA, Y.; KRZYZANOWSKI, J.; WIGIER, M.; KRUPIN, V.; WOJCIECHOWSKA, A., Food security strategy for Mercosur countries in response to climate and socio-economic challenges, **Sustainability**, v.17 (2025);
[2] BIAN, Q. et al., PPI-CoAttNet: a web server for protein–protein interaction tasks using a coattention model, **Journal of Chemical Information and Modeling**, v.65, p. 461–471 (2025);
[3] LEHNE, B.; SCHLITT, T., Protein–protein interaction databases: keeping up with growing interactomes, **Human Genomics**, v.3, p. 291–297 (2009);
[4] SZKLARCZYK, D. et al., The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest, **Nucleic Acids Research**, v.51, p. D638–D646 (2023);
[5] THE UNIPROT CONSORTIUM, UniProt: a worldwide hub of protein knowledge, **Nucleic Acids Research**, v.47, p. D506–D515 (2019);
[6] THE GENE ONTOLOGY CONSORTIUM, The Gene Ontology knowledgebase in 2023, **Genetics**, v.224 (2023);
[7] FLORENTINO, B. R. et al.. Predicting protein-protein interactions to accelerate genetic improvement aimed at understanding the plant proteome. **11th Workshop on Probabilistic and Statistical Methods** (2026);