

# Abstrakt

Terpeny a terpenoidy představují největší a strukturně nejrozmanitější skupinu přírodních látek s využitím v mnoha oborech, včetně farmaceutického průmyslu. Tyto molekuly jsou v přírodě syntetizovány enzymy známými jako terpen syntázy. V této práci byla provedena bioinformatická analýza kurátorované databáze obsahující všech 1125 experimentálně charakterizovaných terpen syntáz se zaměřením na identifikaci vzorců v délkách sekvencí a doménových architekturách těchto enzymů napříč různými říšemi života.

Na základě poznatků této analýzy byl proveden sekvenčně založený mining s cílem identifikovat možné nové terpen syntázy. S využitím téměř 5,5 miliard proteinových sekvencí z různých rozsáhlých sekvenčních databází vedl mining k identifikaci více než 600 tisíc potenciálních terpen syntáz. Tyto potenciální terpen syntázy pocházejí převážně z bakterií a metagenomů, tedy ze zdrojů, které byly historicky méně zkoumány.

Výsledný dataset, doplněný fylogenetickým stromem, sítí sekvenční podobnosti a dvěma skóre prioritizace, nabízí cenný zdroj pro objevování nových terpenů.

**Klíčová slova:** terpen syntáza, TPS, mining, Pfam, SUPERFAMILY, doména, terpen