

47

Analýza bakteriálnych funkcií vo vzorke podľa 16s rRNA

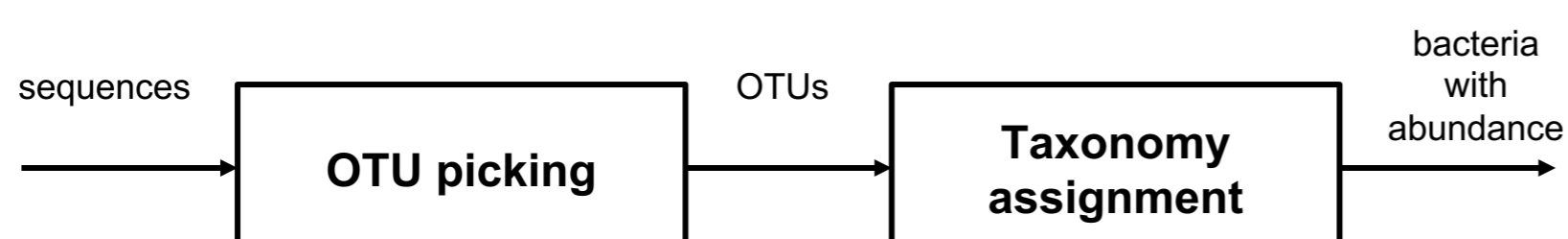
MOTIVÁCIA

Ľudské telo je prostredím pre život veľkého množstva mikróbov. Niektoré z nich môžu spôsobovať choroby a infekcie, ale iné, napríklad črevný mikrobióm, sú pre život a zdravie človeka nepostrádateľné. Paradoxne, črevný mikrobióm, ktorý ovplyvňuje zdravie každého človeka, zatiaľ nie je detailne preštudovaný, pretože obsahuje tisíce rôznych druhov baktérií, z ktorých väčšina sa nedá kultivovať v laboratórnych podmienkach. Riešením tohto problému je namiesto priameho pestovania baktérií v laboratóriách použiť na vzorku nové rýchle metódy sekvenovania a spracovávať získané sekvencie DNA. Tie sa potom analyzujú pomocou bioinformatických nástrojov, ktoré dokážu odpovedať na dve dôležité otázky - aké baktérie sú prítomné vo vzorke, a aké majú funkcie (trávenie cukrov, tukov, ...).

ANALÝZA VZORKY

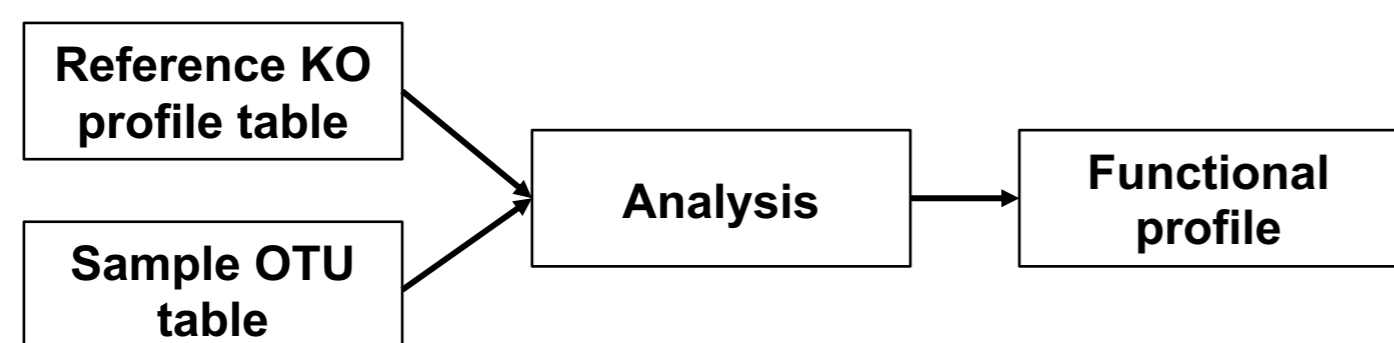
1. Taxonomická analýza - aké sú tu baktérie?

Sekvencie sa rozdelia do skupín nazývaných OTU (operational taxonomical unit) podľa vzájomnej podobnosti. V rovnakej skupine bývajú sekvencie, ktoré sú na 95 - 99 % identické. Skupiny sú reprezentované jednou sekvenciou, ktorá sa potom vyhľadáva v referenčnej databáze. Pre každé OTU sa potom určí príslušný druh baktérie podľa najlepšej zhody reprezentatívnej sekvencie so záznamom v referenčnej databáze.



2. Analýza funkčného profilu - čo dokážu tieto baktérie robiť?

Vstupom sú OTU nájdené vo vzorke a referenčná databáza obsahujúca známe funkčné profily. Pre OTU, ktoré majú zhodu v referenčnej databáze, sa funkčné profily vezmú priamo z tejto databázy. Pre OTU, ktoré dostatočnú zhodu nemajú, sa funkčný profil odvodzuje pomocou rôznych metód - môže to byť obyčajný priemer profilov najpodobnejších OTU v referenčnej databáze, alebo sa môže skúmať odhadovaný vývoj jednotlivých druhov baktérií a zodpovedajúcich funkcií pomocou fylogenetického (vývojového) stromu.

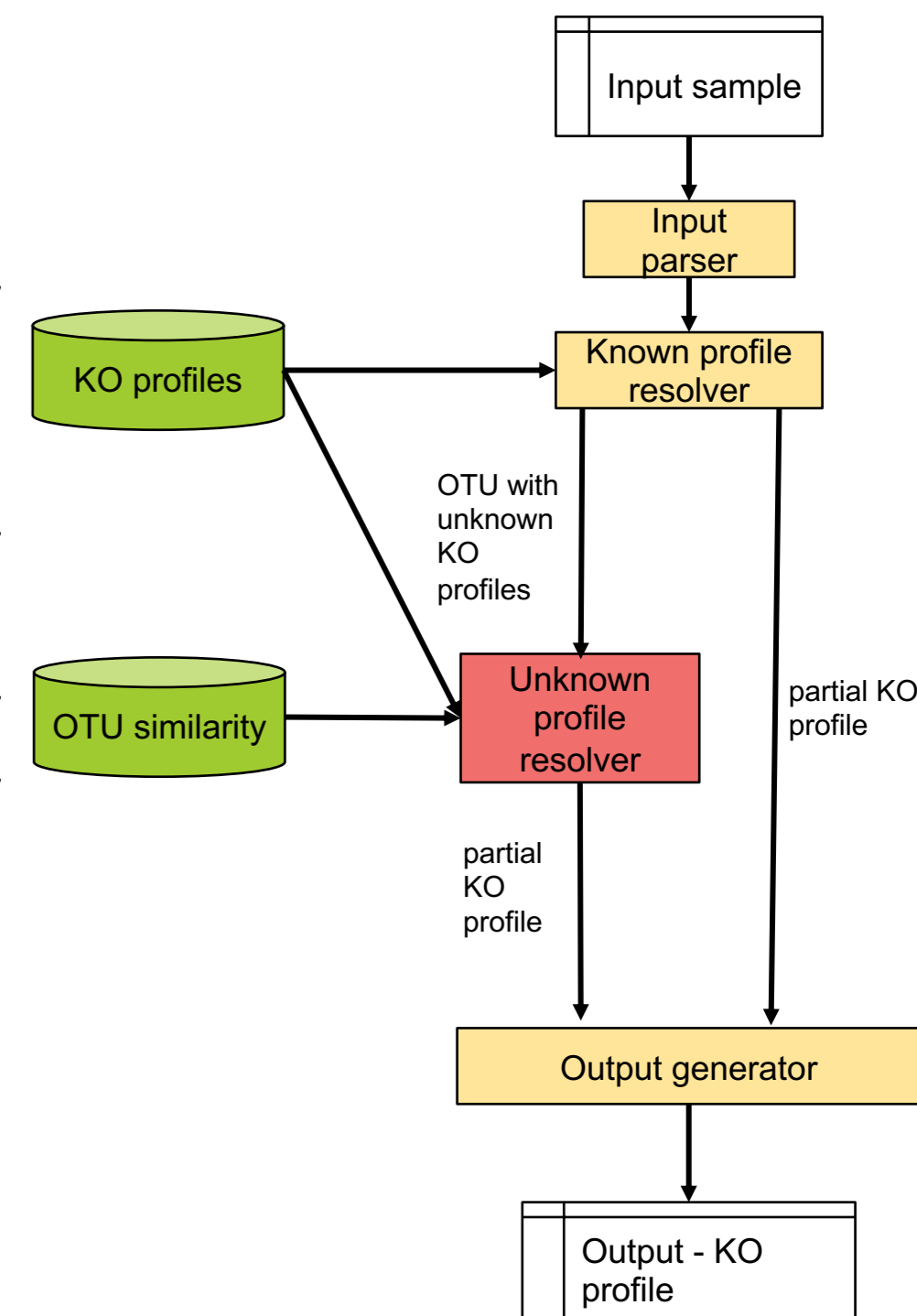


EXISTUJÚCE NÁSTROJE

- **Picrust** - rozšírený, používa zastaralú databázu, počíta funkčné profily pomocou fylogenetického stromu, Python
- **Picrust 2** - nová verzia Picrustu nezávislá na referenčnej databáze, aktuálne v beta verzii
- **Tax4Fun** - novšia databáza, vylepšená metóda fylogenetického stromu z Picrustu, R

VYTVORENÝ NÁSTROJ

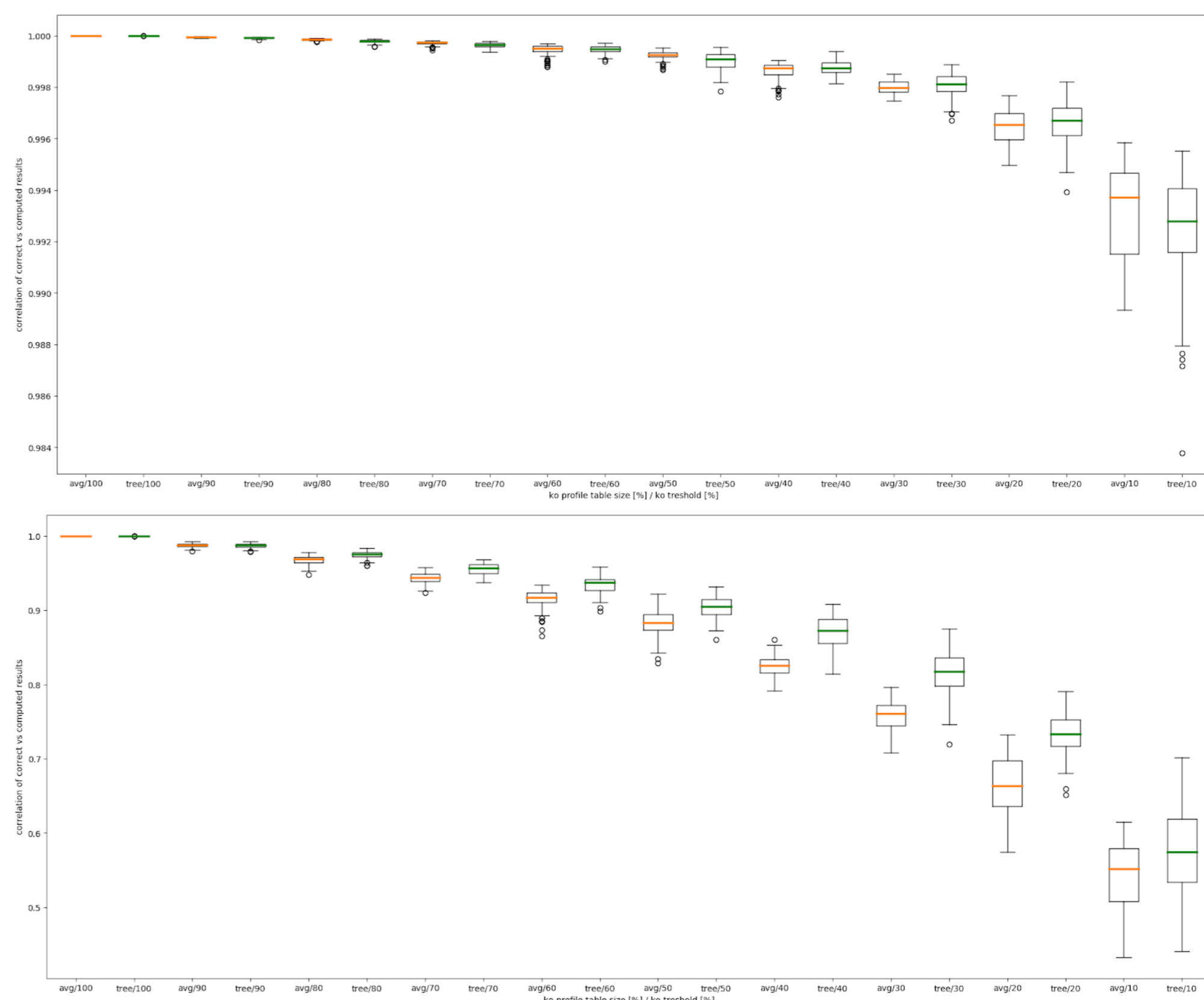
Cieľ knižnice bol otestovať a porovnať rôzne metódy na predikciu funkčného profilu. Predikciu neznámych profilov sa zaoberá modul "Unknown profile resolver". Sú v ňom implementované rôzne metódy pre odhad funkčných profilov, využívajúce priemer, vážený priemer, obmedzenie zhody prahom, fylogenetický strom, aj lineárnu regresiu.



Experimentovala som s rôznymi parametrami metód aj s rôznymi stratégiami na vyhodnotenie testov.

VÝSLEDKY EXPERIMENTU

Porovnanie metódy priemeru funkčných profilov a odvodenia z fylogenetického stromu. Výsledky zobrazujú koreláciu medzi funkčným profilom vypočítaným mojim nástrojom a očakávaným funkčným profilom. Horný graf zobrazuje všetky zdokumentované funkcie baktérií, dolný iba najšpecifickejšie.



ĎALŠIE CIELE

Najdôležitejšie je získať porovnanie mojej knižnice s Picrustom, ktorý je najpoužívanejším nástrojom v analýze funkčných profilov. Pri tomto porovnaní skúsím Picrust otestovať pomocou novej metódy evaluácie, ktorá berie do úvahy iba špecifické funkcie. Ak budú výsledky dosť zaujímavé, dali by sa neskôr publikovať.