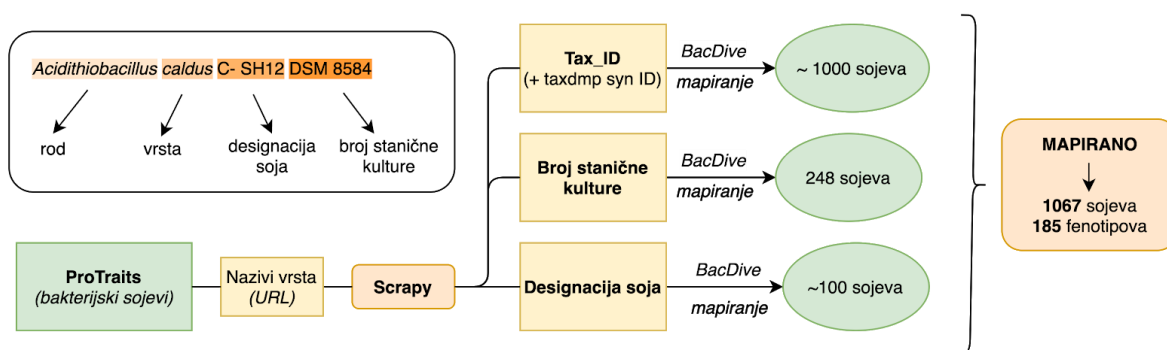


## MAPIRANJE BAKTERIJSKIH FENOTIPOVA S **BACDIVE** BAZE PODATKA



**BacDive** (*Bacterial Diversity Metadatabase*) baza podataka sadrži informacije o taksonomiji, morfologiji, fiziologiji, načinu izolacije i kultivacije te metabolizmu kod oko 64 000 bakterijskih sojeva (<https://bacdiv.dsmz.de/>).

### ALATI

- **Scrapy** - open source framework u Pythonu za ekstrakciju podataka s web stranica. Omogućava skidanje podataka po definiranim oznakama (*engl. selectors*), kao što su *tax\_ID*, *broj stanične kulture*, i sl.
- **grep()** i **grepl()** - osnovne R funkcije za rad s regularnim izrazima u svrhu pronalaženja preklapajućih uzoraka u tekstualnim podacima.

### CILJ

- Mapiranje što većeg broja dostupnih fenotipova s BacDive baze podataka na bakterijske sojeve u ProTraits-u.
- Priprema mapiranih podataka u ProTraits formatu.
- Validacija ProTraits fenotipskih predikcija za dane bakterijske sojeve s preciznošću  $\geq 0.8$ .

### REZULTAT (5 primjera)

Naziv soja	Tax_ID	Fenotip	ProTraits predikcija	BacDive provjera
<i>Borrelia garinii</i> PBi	290434	oxygenreq.strictaero <sup>1</sup>	+	wrong
<i>Rhodopirellula baltica</i> SWK14	993516	habitat.freeliving	+	correct
<i>Terriglobus saanensis</i> SP1PR4	401053	alkaline_phosphatase	+	correct
<i>Borrelia burgdorferi</i> B31	224326	ecosystem.insecta	+	wrong
<i>Sulfurimonas denitrificans</i> DSM1251	326298	oxygenreq.strictaero	-	correct

<sup>1</sup>oxygenreq.strictaero (oxygen requirement = strict aerobe)