Załącznik nr 1 do SWZ

znak sprawy: A-2402-9/2026

OPIS PRZEDMIOTU ZAMÓWIENIA

Projekt: ***FELU.01.01-IZ.00-0002/24 „Wzmocnienie potencjału laboratoriów Instytutu Agrofizyki PAN w celu prowadzenia badań w obszarze holobiontu roślinnego”***

CYKL SZKOLEŃ Z ANALIZY DANYCH BIOLOGICZNYCH – MIKROBIOM GLEBY, TRANSKRYPTOMIKA I SEKWENCJONOWANIE AMPLIKONOWE

1. *WPROWADZENIE DO R I RSTUDIO DLA POCZĄTKUJĄCYCH*

Kurs skierowany jest do osób analizujących i przetwarzających dowolne dane, w szczególności dane pochodzące z różnych rodzajów doświadczeń w języku programowania R. Jest on wykorzystywany do analizy danych, zarówno w publikacjach naukowych, jak i w laboratoriach. Przyjazne środowisko pracy, jakim jest RStudio, ułatwia rozwiązywanie problemów analitycznych. Kurs ugruntowuje znajomość języku R i otwiera jego nowe możliwości. Wprowadza podstawowe zagadnienia i typy językiem programowania w celu wyjaśnienia, przetworzenia, wizualizacji i wnioskowania na podstawie danych. Znajomość podstaw języka R nabyta na kursie pozwala na samodzielną pracę z tym narzędziem i pisanie własnego oprogramowania, co ma celu lepsze i dokładniejsze rozwiązania problemów w dowolnej dziedzinie.

Zagadnienia:

- Wprowadzenie do R i RStudio
- Podstawowe operatory logiczne, funkcje matematyczne, statystyczne, R jako kalkulator
- Podstawowe typy danych: wektory, macierze, ramki danych, listy
- Wprowadzenie do programowania: pętle, instrukcje warunkowe, funkcje
- Wprowadzenie do analizy danych: podstawy statystyki opisowej, testy statystyczne, test t
- Podstawy wizualizacji danych: pakiet ggplot2: wykresy słupkowe, histogram, boxploty
- ***Szkolenie obejmuje kompleksową analizę danych mikrobiomu gleby w środowisku R i RStudio, od importu i przygotowania tabel ASV/OTU, danych taksonomicznych oraz metadanych, przez kontrolę jakości, filtrowanie, normalizację i transformację danych, aż po ich analizę statystyczną i biologiczną interpretację. Uczestnicy poznają metody oceny różnorodności alfa i beta, analizy składu taksonomicznego, ordynacji PCA, PCoA, NMDS, RDA i CCA, testowania różnic między grupami za pomocą PERMANOVA oraz identyfikacji taksonów różnicujących warianty doświadczenia. Szkolenie obejmuje również analizę mikrobiomu rdzeniowego, zależności mikrobiom–środowisko, korelacji, sieci współwystępowania mikroorganizmów,***

potencjalnych biomarkerów oraz podstaw analizy funkcjonalnej i uczenia maszynowego. Szczególny nacisk zostanie położony na praktyczne wykorzystanie pakietów takich jak m.in. phyloseq, vegan, microbiome, microViz, ANCOMBC, DESeq2, MaAsLin2, NetCoMi, SpiecEasi, ComplexHeatmap i ggplot2 oraz na przygotowanie publikacyjnej jakości wizualizacji, w tym wykresów składu taksonomicznego, boxplotów, heatmap, wykresów ordynacyjnych, volcano plotów, bubble plotów, diagramów UpSet i sieci ekologicznych. Szkolenie ma charakter warsztatowy i przygotowuje uczestników do samodzielnego prowadzenia analiz mikrobiomu gleby oraz poprawnego raportowania i interpretowania uzyskanych wyników.

- ***Pakiety omawiane podczas szkolenia: m.in. phyloseq, microbiome, microViz, vegan, ape, qiime2R, TreeSummarizedExperiment, ANCOMBC, DESeq2, ALDEx2, MaAsLin2, ComplexHeatmap, NetCoMi, SpiecEasi, igraph, ggraph, mixOmics, randomForest, Boruta, caret, tidymodels, lme4, glmmTMB, emmeans, ggplot2, ggpubr, patchwork, ComplexUpset, ggalluvial, plotly***
- ***Typy wykresów wykonywane podczas szkolenia: m.in. wykresy liczby odczytów i krzywe rarefakcji; wykresy składu taksonomicznego; skumulowane wykresy słupkowe; boxploty, violin plots i raincloud plots; mapy cieplne; wykresy PCoA, PCA i NMDS; wykresy RDA, CCA i db-RDA; dendrogramy; wykresy korelacji; bubble plots; volcano plots; forest plots; wykresy prevalence–abundance; diagramy UpSet i Venna; wykresy Sankeya i alluvial; sieci współwystępowania; wykresy istotności zmiennych; krzywe ROC; ryciny wielopanelowe przeznaczone do publikacji***

Oprogramowanie: R, RStudio
Maksymalna liczba osób biorących udział w szkoleniu: 20
Minimalna ilość godzin w jednym bloku szkolenia: 5 godzin szkoleniowych (każda trwająca nie krócej niż 45 minut)
Czas trwania każdego z bloków: 2 dni szkoleniowe
Szkolenia 1, 2, 3 prowadzone w języku polskim
Szkolenie blok 4 – w języku angielskim

2. ***STATYSTYKA W R I RSTUDIO***

Warsztat dla osób, które mają doświadczenie w programowaniu w R. Innych językach lub ukończyły warsztat „Wprowadzenie do R i RStudio dla początkujących". Uczy jak używać wyżej zdefiniowanych narzędzi do analizy oraz interpretacji danych. Przekazuje podstawową wiedzę na temat analizy danych. Pokazuje i wizualizuje stosowane w publikacjach naukowych metody analizy statystycznej i wizualizacji.

Zagadnienia:

- Wprowadzenie do R i RStudio. Podstawowe statystyki opisowe: miary tendencji centralnej, miary rozrzutu
- Wizualizacja danych na potrzeby statystyki

- Podstawowe testy statystyczne: test t, ANOVA, porównania wielokrotne, testy sprawdzania równości wariancji, zgodności z rozkładem normalnym
- Podstawowe testy statystyczne nieparametryczne: Kruskala–Wallisa, test Friedmana, test Wilcoxona, porównania wielokrotne
- Testy zgodności, niezależności
- Regresja prosta, regresja wielokrotna
- Analiza korelacji
- Analiza składowych głównych

Oprogramowanie: R, RStudio
Maksymalna liczba osób biorących udział w szkoleniu: 20
Minimalna ilość godzin w jednym bloku szkolenia: 5 godzin szkoleniowych (każda trwająca nie krócej niż 45 minut)
Czas trwania każdego z bloków: 2 dni szkoleniowe
Szkolenia 1, 2, 3 prowadzone w języku polskim
Szkolenie blok 4 – w języku angielskim

3. ***ANALIZA DANYCH RNA-SEQ W R***

Kurs jest przeznaczony pracownikom naukowym, którzy mieli już do czynienia z programowaniem w R lub odbyli kurs z wprowadzenia do R. Warsztaty rozpoczynają się przedstawieniem specyfiki analiz danych wysokoprzepustowych, jakimi są dane pochodzące z eksperymentu RNA-seq. Zapoznaje z kompleksową analizą transkryptomu. Składają się na nią elementy takie jak analiza jakości, mapowanie, wstępna filtracja i normalizacja, analiza wyższego rzędu, czyli analiza różnicowa, oraz ontologia genomu. Przekazywana jest wiedza praktyczna poparta przykładami, ćwiczeniami oraz tzw. *case study* dotyczącymi nauk medycznych. Kurs ma na celu wprowadzić w podstawy analiz dużych zbiorów danych next generation sequencing (NGS).

Zagadnienia:

- Wprowadzenie do RNA-seq i Bioconductor
- Analiza jakości sekwencjonowania
- Mapowanie i zliczanie odczytów
- Wstępna obróbka danych countowych
- Podstawowe testy statystyczne dla danych NGS: edgeR, deseq, limma
- Analiza GO i KEGG
- Wizualizacja wyników
- Case study

Oprogramowanie: R, Rstudio
Maksymalna liczba osób biorących udział w szkoleniu: 20
Minimalna ilość godzin w jednym bloku szkolenia: 5 godzin szkoleniowych (każda trwająca nie krócej niż 45 minut)
Czas trwania każdego z bloków: 2 dni szkoleniowe
Szkolenia 1, 2, 3 prowadzone w języku polskim
Szkolenie blok 4 – w języku angielskim

4. ***AMPLICON SEQENCING ANALYSIS – KURS ANGLOJĘZYCZNY***

Szkolenie analizy danych sekwencjonowania amplikonu, z wykorzystaniem narzędzi takich jak Usearch, Vsearch i qiime2. Omawiane tematy obejmą kontrolę jakości, dopasowanie sekwencji, wstępne filtrowanie, normalizację i zaawansowane analizy, takie jak metryki różnorodności i przypisanie taksonomiczne. Szkolenie oparte o praktyczne przykłady, ćwiczenia i dedykowane studium przypadku. Podstawowym celem jest przekazanie praktycznej wiedzy niezbędnej do skutecznej analizy sekwencjonowania amplikonów.

Zagadnienia:

- Wprowadzenie do sekwencjonowania amplikonów i odpowiednich narzędzi (Usearch, Vsearch, qiime2)
- Ocena jakości danych sekwencjonowania amplikonu
- Mapowanie i techniki zliczania odczytu do sekwencjonowania amplikonu
- Wstępne przetwarzanie zliczeń amplikonów
- Kluczowe testy statystyczne dla danych sekwencjonowania amplikonu
- Badanie wskaźników różnorodności i przypisań taksonomicznych w sekwencjonowaniu amplikonów
- Wizualizacja wyników sekwencjonowania amplikonu
- Dogłębne studium przypadku dotyczące analizy sekwencjonowania amplikonu

Oprogramowanie: Usearch, Vsearch, qiime2
Maksymalna liczba osób biorących udział w szkoleniu: 20
Minimalna ilość godzin w jednym bloku szkolenia: 5 godzin szkoleniowych (każda trwająca nie krócej niż 45 minut)
Czas trwania każdego z bloków: 2 dni szkoleniowe
Szkolenia 1, 2, 3 prowadzone w języku polskim
Szkolenie blok 4 – w języku angielskim

WYMAGANE SĄ SZKOLENIA W FORMULE ON-LINE

Wykonawca zobowiązany jest do przygotowania i przekazania Zamawiającemu materiałów szkoleniowych do każdego z przeprowadzonych szkoleń, obejmujących zarówno prezentacje

wykorzystane podczas zajęć, jak i materiały opisowe rozwijające omawiane zagadnienia. Materiały opisowe powinny stanowić uzupełnienie prezentacji i umożliwiać uczestnikom samodzielne utrwalenie wiedzy oraz jej praktyczne wykorzystanie po zakończeniu szkolenia. Dopuszcza się przekazanie materiałów w wersji elektronicznej najpóźniej w dniu szkolenia.