

Filtrado de las bases de datos

Las bases de datos fueron descargadas mediante BASH de la página web de: el nombre de la página. Los filtros se hicieron de la siguiente manera:

1. Se buscó en cada columna de cada dato la palabra "Illumina" para quedarnos solo con los datos que fueron estudiados o secuenciados con tecnología Illumina.
2. Dependiendo del mapa que se quería crear, se hicieron 2 filtros diferentes:
 - a) "wgs" para quedarnos con los estudios de metagenomas.
 - b) ".amplicorn" para quedarnos con los estudios de amplicornes.

En este paso se crearon dos sub bases de datos, "wgs" ".amplicorn". Los siguientes filtrados se hicieron idénticos en ambos casos, así que solo llamaré sub base, que puede ser ya sea "wgs" ".amplicorn".

3. Aquí, dependiendo de lo que queríamos, se hicieron tres búsquedas diferentes
 - "water"
 - rhizosphere"
 - "sediment"

para crear las sub bases respectivas:

- "sub base_water"
- "sub base_rhizosphere"
- "sub base_sediment"

Y a partir de cada base se crearon los mapas.

Creación de los mapas

Todos los mapas se crearon de la misma manera. La base de datos con la que trabajamos tenía la ubicación en el formato "45.241 N, 51.8426 W" como ejemplo, pero la paquetería "ggplot" de R no podía entender la ubicación. Por lo tanto, se creó una función para transformar las coordenadas al siguiente formato: "(15.241,-51.826)", que es:

```
parse_coordinates <- function(coord_string) {  
  
  parts <- strsplit(coord_string, " ")  
  lat <- as.numeric(parts[[1]][1])  
  lon <- as.numeric(parts[[1]][3])  
  
  if (grepl("S", coord_string)) {  
    lat <- -lat  
  }  
  if (grepl("W", coord_string)) {  
    lon <- -lon  
  }  
  
  return(c(lat, lon))  
}
```

Entonces la función hace lo siguiente

$$\text{parse_coordinates}(45,241N, 51,846W) = (45,241, -51,846)$$

y ya nada mas con "ggplot" se hicieron las graficas.