

## Day 2: Data structures

Let's start by importing the **data** module

```
In [13]: from data import DNA_Sequence, human_sequence, Species_list, sequences
```

### Problem 1

The `*DNA_Sequence*` variable imported from the **data** module contains DNA sequence

a) Print the **DNA\_Sequence** variable

```
In [4]:
```

b) Print the length of the **DNA\_Sequence** variable

```
In [5]:
```

c) Count (with a method) the number of "A", "C", "T", and "G" nucleotides in the **DNA\_Sequence** variable and assign each one to a different variable (e.g, assign the number of "A"s to a variable named `a_count`, etc) and then print each variable

```
In [10]:
```

d) Transcribe the **DNA\_Sequence** and assign it to a new variable. Print the new variable.

(Tip: replace the "T" nucleotides for "U" nucleotides)

```
In [15]:
```

e) Split the **DNA\_Sequence** at each "GAT" motif and store the resulting list in a new variable. Determine the number of the resulting fragments. Print the result.

(Tip: Determine the length of the resulting list of fragments to get the number of fragments)

```
In [21]:
```

f) Merge the first and last fragments of the list resulted from e) and store it in a new variable. Print the new variable.

```
In [25]:
```

### Problem 2

The **human\_sequence** variable imported from the **data** module contains a human DNA sequence

a) Print the **human\_sequence** variable

In [26]:

b) Notice that both ends of the sequence contain gaps "-". Eliminate the gaps from boths ends of the sequence, and assign the resulting sequence to a new variable. Print the result.

In [28]:

c) Change the capitalization of the **human\_sequence** variable and print

In [29]:

## Problem 3

The **Species\_list** variable imported from the **data** module contains a list with species names.

a) Determine the number of species in **Species\_list** and print it.

In [30]:

b) Sort the **Species\_list** variable by alphabetical order and print.

In [41]:

c) Change the "C\_albicans" entry by a new species: "D\_melanogaster". Print Species\_list.

In [42]:

d) Store the first 3 species of the **Species\_list** list in a new variable: "First\_species" and print.

In [44]:

e) Create a new empty list and add the following species as entries: "C\_kahawae","Q\_suber","L\_lepida"

In [46]:

## Problem 4

a) Create a new string variable composed of 100 "N" characters

In [48]:

b) Create two new number variables with the numbers 23 and 323. Determine their sum, difference, division and multiplication

In [49]:

|

c) Notice that the division of 23 by 323 results in "0". Convert both numbers into floating point variables and repeat the division

In [50]:

|

## Problem 5

The **sequences** variable contains a dictionary with taxon name as keys, and their DNA sequence of the Cytb gene as values.

a) Determine the number of taxa contained in the dictionary

In [53]:

|

b) Print both the taxon name and sequence of the 3<sup>o</sup>, 5<sup>o</sup> and 7<sup>o</sup> dictionary item.

In [61]:

|

In [ ]:

|