

Introduction to Bioinformatics

Problem Set 6: Data Aggregates

A. Strings

A.1. From the example below:

```
(DEFINE test-string AS "ABCDEFGH")
test-string[2 -> 4]
test-string[2 5 1]
```

complete the following to extract the third codon from the gene sequence:

```
(DEFINE gene-seq AS (SEQUENCE-OF ssr1600)
  fill in
```

Confirm what you did by displaying the sequence and looking for the third codon.

A.2. Write code that will extract the numbers from a social security number of the form: NNN-NN-NNNN

A.3. The third position of codons is the position of greatest variability. It therefore may be that nucleotide frequencies vary more at that position than at position 1 and 2, and so it may be a better (or maybe a worse!) diagnostic for the genome identity of a sequence. Find the total nucleotide frequencies within the genes of *Prochlorococcus marinus* SS120 at each of the three positions. The following function might be of use to you:

```
(FROM 1 TO 30 BY 3)
--> (1 4 7 10 13 16 19 22 25 28)
```

B. Lists

B.1. From the example below:

```
(DEFINE test-list AS {1 2 3 4 5})
test-list[2 -> 4]
test-list[2 5 1]
```

Define the set of cyanobacteria that live in the ocean. Here are all available cyanobacteria:

```
*loaded-organisms
```

Of these, all *Prochlorococci*, *Synechococci*, *Crocophaera*, and *Trichodesmium* live in the ocean.

C. Tables

C.1. From the example below:

```
(DEFINE test-table[1 "A"] AS "1A")
(DEFINE test-table[2 "G"] AS "2G")
(DISPLAY-TABLE test-table)
```

Write a loop that creates and displays a multiplication table, showing the products of m times n for m and n going from 1 to 10.