

2. Genes and proteins

- The sequence as a model of DNA
- Genes: from Mendel to molecular biology
- The genetic code
- A translation algorithm
- **Implementing the genetic code**
- Algorithms + data structures = programs
- The algorithm design trade-off
- DNA sequencing
- Whole genome sequencing
- How to find genes?

Implementing the genetic code

We still have to write the `lookupGeneticCode` function!

```
indexDNA, indexProt: integer
DNAsequence, ProteinSequence: character string [1:*]
indexProt, indexDNA ← 1
repeat
  ProtSequence [indexProt] ← lookupGeneticCode (DNAsequence [indexDNA],
  DNAsequence [indexDNA+1], DNAsequence [indexDNA+2])
  indexDNA ← indexDNA + 3
  indexProt ← indexProt + 1
until DNAsequence [indexDNA] = "*" or
DNAsequence [indexDNA+1] = "*" or
DNAsequence [indexDNA+2] = "*"
ProteinSequence [indexProt] ← "*"
```

lookupGeneticCode is a function
which is written separately to reduce the difficulty
of writing the entire algorithm at once

The genetic code as an array

- The genetic code can be represented as an array of 64 rows and 4 columns
- Here, the first 12 rows (over 64)

T	T	T	F
T	T	C	F
T	T	A	L
T	T	G	L
T	C	T	S
T	C	C	S
T	C	A	S
T	C	G	S
T	A	T	Y
T	A	C	Y
T	A	A	Stop
T	A	G	Stop
...

```

Function lookupGeneticCode (char1, char2, char3: character) returns character
/* GeneticCode: array [1:64, 1:4] of character
                                     is supposed to be known inside the function */
for i from 1 to 64 do
    if GeneticCode [i, 1] = char1 then
        for j from i to 64 do
            if GeneticCode [j, 2] = char2 then
                for k from j to 64
                    if GeneticCode [k, 3] = char3 then
                        return (GeneticCode [k, 4])
                    endfor
                endfor
            endfor
        endfor
    endfor
return("?")
end lookupGeneticCode

```

```

indexDNA, indexProtein: integer
DNAsequence, ProteinSequence: character string
[1:*]
indexProt, indexDNA ← 1
repeat
  ProtSequence [indexProt] ← lookupGeneticCode
  (DNAsequence [indexDNA], DNAsequence [indexDNA
  +1], DNAsequence [indexDNA+2])
  indexDNA←indexDNA + 3
  indexProt ← indexProt + 1
until DNAsequence [indexDNA] = "*" or
DNAsequence [indexDNA+1] = "*" or
DNAsequence [indexDNA+2] = "*"

ProteinSequence [indexProt] ← "*"

```

```

Function lookupGeneticCode (char1, char2, char3: character) returns
character
/* GeneticCode: array [1:64, 1:4] of character
           is supposed to be known inside the function */
for i from 1 to 64 do
  if GeneticCode [i, 1] = char1 then
    for j from i to 64 do
      if GeneticCode [j, 2] = char2 then
        for k from j to 64
          if GeneticCode [k, 3] = char3
            then return (GeneticCode [k, 4])
        endfor
      endif
    endif
  endif
endfor
return("?")
end lookupGeneticCode

```