|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Identity** | **Name** | **Tipo e dominio** | **Multivalue** | **Zero** | **Keys** |
| Gene-“GenBank” | **ID\_genbank** | int | N | N | S |
| Description | Text | N | N | N |
| Organism | Text | N | N | N |
| Last\_actualization | Data | N | N | N |
| Len(sequence) | int | N | N | N |
| Sequence | Text | N | N | N |
| A | int | N | S | N |
| C | int | N | S | N |
| G | int | N | S | N |
| T | int | N | S | N |
| PubMed | **PubMed\_ID** | int | N | N | S |
| Publishing data | Text | N | N | N |
| Title | Text | N | N | N |
| Doi number | Int | N | N | S |
| Authors | Text | N | N | N |
| Affiliation | Text | N | N | N |
| CDS | **CDS\_ID/protein\_ID** | int | N | N | S |
| Product | text | N | N | N |
| Location | int | N | N | N |
| Translation | text | N | N | N |
| Uniprot | **ID** | int | N | N | S |
| Function | text | N | N | N |
| Subcellular Location | text | N | N | N |
| Protein sequence | text | N | N | N |
| Length (a.a) | int | N | N | N |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Identity** | **#** | **Relationship** | **Identity** | **#** |
| Genebank | 1…N | Reference | PubMed | 1…N |
| 1…1 | TYPE=CDS | CDS | 1…N |
| CDS | 1…N | Research query | Uniprot | 1…1 |

With the table presented we want to explore certain problems such as:

1) number of CDS;

2) percentage GC;

3) top 3 most common amino acids.

4)etc.

Guilherme Sousa, Karyna Lysenko, Rodrigo Esperança.