

## 2. ¿Cómo caracterizar una secuencia de DNA?

A la hora de identificar en una secuencia de nucleótidos regiones de interés: que se transcriban, que se encarguen de la regulación de otras que sí lo hagan, etc., puede ayudar conocer ciertas características de estas regiones. Por ejemplo, si se tiene el promotor de un gen al calcular el porcentaje de GC se podría determinar si forman parte de un gen constitutivo o uno que se expresa en tejidos específicos, según este porcentaje esté en torno al 0.61 o al 0.23, respectivamente.

Algunos de los criterios que van a permitir la caracterización de genes son: la cantidad y porcentaje de bases y de dinucleótidos, el peso molecular de la secuencia o el contenido en GC.

## Guía de usuario:

Sequence Manipulation Suite (<http://www.bioinformatics.org/sms2/>)

Dentro de esta página hay diferentes herramientas que se muestran en el menú de la izquierda de la página pero nos vamos a centrar en:

- DNA stats ([http://www.bioinformatics.org/sms2/dna\\_stats.html](http://www.bioinformatics.org/sms2/dna_stats.html)): va a determinar la cantidad y porcentaje de cada base y de dinucleótidos Basta con introducir la secuencia de nucleótidos en formato FASTA y clicar en Submit.

# Sequence Manipulation Suite:

## DNA Stats

DNA Stats returns the number of occurrences of each residue in the sequence you enter. Percentage totals are also given for each residue, and for certain groups of residues, allowing you to quickly compare the results obtained for different sequences.

Paste the raw sequence or one or more FASTA sequences into the text area below. Input limit is 500000 characters.

```
>Contig1_4711pb
TATTTTCAATTGTTTCTTCAACATTTACCGCCCGAGAACATGATAAGACATACTCAT
CCGGCGCTCTAGCTTTGCGTAGCGCTCGGTCAAGTTCAATGGTGTATGGAAAAATTC
CGCAACTAACAGTGGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
GAGCTGAAACGCTTGGTATACAAACAGCTTCGACAAAGCTCAGCATACTCAAGCTCA
```

Please check the browser compatibility page before using this program.

\*This page requires JavaScript. See [browser compatibility](#).

\*You can mirror this page or use it off-line.

El resultado que se mostrara sera una tabla similar a esta:

| DNA Stats results  |              |             |
|--|--------------|-------------|
| Results for 4711 residue sequence "Contig1" starting "TATTTCAAT" |              |             |
| Pattern:   | Times found: | Percentage: |
| g  | 1155         | 24.52       |
| a  | 1179         | 25.03       |
| t  | 1163         | 24.69       |
| c  | 1194         | 25.34       |
| n  | 20           | 0.42        |

|                       |      |       |
|-----------------------|------|-------|
| gg                    | 274  | 5.82  |
| ga                    | 276  | 5.86  |
| gt                    | 298  | 6.33  |
| gc                    | 307  | 6.52  |
| gn                    | 0    | 0.00  |
| ag                    | 241  | 5.12  |
| aa                    | 362  | 7.69  |
| at                    | 269  | 5.71  |
| ac                    | 307  | 6.52  |
| an                    | 0    | 0.00  |
| tg                    | 296  | 6.28  |
| ta                    | 224  | 4.76  |
| tt                    | 324  | 6.88  |
| tc                    | 319  | 6.77  |
| tn                    | 0    | 0.00  |
| cg                    | 344  | 7.30  |
| ca                    | 317  | 6.73  |
| ct                    | 271  | 5.75  |
| cc                    | 260  | 5.52  |
| cn                    | 1    | 0.02  |
| nc                    | 1    | 0.02  |
| nn                    | 19   | 0.40  |
| g,c                   | 2349 | 49.86 |
| a,t                   | 2342 | 49.71 |
| r,y,s,w,k             | 0    | 0.00  |
| b,h,d,v,n             | 20   | 0.42  |
| r,y,s,w,k,m,b,d,h,v,n | 20   | 0.42  |

- DNA Molecular weight ([http://www.bioinformatics.org/sms2/dna\\_mw.html](http://www.bioinformatics.org/sms2/dna_mw.html)): va a determinar el peso molecular de la secuencia  
Como en el caso anterior basta con introducir la secuencia de nucleótidos en formato FASTA y clicar en Submit.

**Sequence Manipulation Suite:**

**DNA Molecular Weight**

DNA Molecular Weight accepts one or more DNA sequences and calculates molecular weight. Sequences can be treated as double-stranded or single-stranded, and as linear or circular. Single-stranded sequences are assumed to have a 5' phosphate. Use DNA Molecular Weight when calculating molecule copy number.

Paste the raw sequence or one or more FASTA sequences into the text area below. Input limit is 200000 characters.

Contig 4\_5002ab  
GATTTCCTCGAAACCCGAGGCCCTACAGTGCTGGATTGTGAGCTTTACTTACCGCTG  
ACCTGAGCTTCAGGTGATGCTGAGCTTTCGCTGCTGTTAAATAACCAAGATCTCGCGGTAC  
CTTTATTACAGGCTCTGGCGACGGGATCGCAAGGGTCTCGAGCTATATCGTAGACGA  
GGCTAAAGTGGAGGGAGTTTCGAAATAAGGGGGTGTCAAGTAAGCTTTCATGGTGGAGAT  
ACCTTAAAGATAGATTTGAGCCCTGGTTTCCACCGGACACGGCTCTTAATGTGATTCCTTG

Please check the browser compatibility page before using this program.

Submit | Clear | Reset

• Treat sequences as  single  stranded  linear molecules.

\*This page requires JavaScript. See browser compatibility.  
\*You can mirror this page or use it off-line.

El resultado se mostrará en una ventana que indica un el peso molecular estimado teniendo en cuenta el número de bases de cada tipo que hay y el peso molecular de cada base:

**DNA Molecular Weight results**  
Results for 5002 residue sequence "Contig 4" starting "GATTTCCTCC"  
1544849.77 Da

En el caso de que parte de la secuencia haya sido enmascarada, y por tanto contenga "N" en lugar de una base en concreto, el programa dará como resultado un rango de pesos moleculares que tomaría la secuencia según por qué bases se sustituyan:

**DNA Molecular Weight results**  
Results for 4711 residue sequence "Contig1" starting "TATTTCAAT"  
1454378.27 to 1455178.87 Da

- CpG islands ([http://www.bioinformatics.org/sms2/cpg\\_islands.html](http://www.bioinformatics.org/sms2/cpg_islands.html)): va a determinar el contenido en islas CpG, el problema es que esta herramienta requiere de JAVA.

### **Endmemo** (<http://www.endmemo.com/bio/gc.php>)

Al igual que con la herramienta anterior se introduce la secuencia en formato FASTA y se clica en Calculate, de forma que calcula el contenido en GC y longitud de la secuencia.

**DNA/RNA GC Content Calculator**

Please Paste the DNA/RNA Sequence:

```
>Contig1_4711pb
TATTTCAATTGTTCTAACACTTTACCGCCAGAATCGATAAGACATACTCATCCGGCGCTTAGCTTG
CCGTACGCTCGCGTCAGTCAATGGTATTGAAAAATTCCGCAACTAAACGTGATCCTCTCTCTCTCTC
TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
TTTGAGCTGAAACGCTTGGGTATACAAACGCTTCGACAAGCCTCAGCCATACTCAGCTCACACAAAAC
CTCACC
GTTGTTGATGCAGATGCTTACTATGGCTGATAAAAATAGGTGACGATGTTACACATTTGTTGACCATGCC
ACT
TTGCGAAGGTGTCGTCACCGGCTCGTCTCGGCTCGTAGAAAAAGCTTCCCTACACACACAGCAGTGTGAG
TGCCTG
CACGAGTACACCGAAAGACCTTCGGTGGCATGGCGTAGCGCCGAGGGTGCACAGGCTTGGTGGCACAAG
TTT
TGCTTACTAACGTGATAAGCAAAACAAGCAAAGCGATCTGATAACGGTTGGCCATTCGGACGCATGCG
CAACCG
GGACTGGTGCCTGACTGAGCTTGGTGGCTATCACTCTGGTTGGTCGGGTTGGGAGAACGTTTATGTT
CTCTA
ATCTAACGTATGACAGAGAGTACAGTAGGGTACGAGAAAATGAAGAGAGAGACAGTGAGAGAGATTTG
AATGAAAT
TCACAGATTACTCAGGGTCTCGCGCTCGCTCCCTATGCTCGTATGATAAGACTACACGCCGCGCTGTG
CGAAAG
GGTTCCCATCCAAACCAATCCTAGTAGGCAACTAGTCGGTACTAACAAACGGCACCGGTTGGTCGGG
TGGGTGAC
TAGTGTCCCTGTCGATTGGTTCAAGTTTGTCACTCGAGTGATAACCTTCAAGCCGGTCAACCTAAC
CATCGCGCAATCATGGAGCACACACGGCTACTCTCAGCAGTGGTGCTTAGGCCACTGCTGCTGGGAA
GTGCTC
CACCATCGCAAGGGGCACGAATTCTCGGCATACTGCCGTCGGCGTTCGCACTACATCATCGGTGCC
GGGCT
GATGAAAGCACTGCTGGAGCGCTGGGCGACGAGGTGACCATCGTCAGTGGTCACTGGGATGAC
CGGGGGT
```

GC Content:  %

DNA Length:  bp

**Calculate** **Clear All**

Este último dato también lo proporciona la anterior herramienta, por lo que será muy útil para comparar ambos resultados.

El resultado se mostrará de la siguiente forma:

GC Content: **49.862025047761** %

DNA Length: **4711** bp

Para el contig empleado según Sequence Manipulate tiene una longitud de 4711 nucleótidos, en cambio según Endemo contiene 4718 residuos. Como la herramienta usada para localizar contigs, Cap3, coincide con Sequence Manipulate en que se trata de una secuencia de 4711 nucleótidos tomaremos este dato como más probable.