

# Formatos archivos de secuencias

[http://www.ebi.ac.uk/help/formats\\_frame.html](http://www.ebi.ac.uk/help/formats_frame.html)

[http://www.genomatix.de/online\\_help/help/sequence\\_formats.html](http://www.genomatix.de/online_help/help/sequence_formats.html)

**FASTA (\*.fas):** es uno de los formatos más empleados en la actualidad. Puede incluir una o más secuencias lineales de nucleótidos o aminoácidos no alineadas.

```
>SASP of: sap3 check: -1 from: 1 to: 279 July 29, 1996 13:56
MTDNLTTTHGGSTTLELLAQYNDHRSKKDKSIEHIEKGTCSGKERNPSYDEIFTENIKLK
LQVQEYETEIESLEKVIDMLQKNREASLEVVLEQVQNSRDSYVNDQSFVLPPrSAERKA
HIKSLNLP IPTLS PPLQOGSDVALET SVTPTVPQIGVTSNTSISRKHLQNMILNDEIEAN
SSFSSPKIINRSVSSPTKIHSEQLASPAASVTYTTSRITIKSPNKGSKSPLQERLRSPQN
PNRMTAVINNHLHSPLKASTSNNLDELTESKSQQLTNDAIQKNDRVYSSITSSAYTTGTP
TSAAKSPSSLLEVKEGENKALGFSPASKEKLDDFTQLLDSSFGEEDLVNTDSKDPLSIKS
```

**Alineamiento (\*.aln):** es uno de los muchos formatos para archivos de secuencias alineadas más utilizados por los programas de análisis de secuencias (**MUSCLE** o **CLUSTAL**).

CLUSTAL W(1.60) multiple sequence alignment

```

YVKE1      SQLSWKRLLMKGYIPPYKPAVS-----NSMDTSNFDEEFTR-----EKPIDSVVDEYLSSES
YVKE2      KDISWKKLLLKGYIPPYKPIVK-----SEIDTANFDQEFTK-----EKPIDSVVDEYLSAS
YVKE_H     RRIDWEKLENREIQPPFKPKVC-----GKGAENFDKFFTR-----GQPVLTPPDQLVIAN
YVKE_M     RSIDWDLLEKKQALPPFQPQIT-----DDYGLDNFDTQFTS-----EPVQLTPDDEDAIKR
          * .

YVKE1      -----QKQF
YVKE2      -----QKQF
YVKE_H     D-----QSDF
YVKE_M     D-----QSEF
          *  *
    
```

Sus principales características son: tanto nucleótidos como aminoácidos son representados por el código de una letra; los *gaps* son indicados con un guión "-" y al final de las secuencias alineadas, puede aparecer una especie de secuencia consenso con signo como: "\*"

**Formato PHYLIP (\*.phy):** es un formato de secuencias de nucleótidos y de residuos de aminoácidos que se utiliza frecuentemente como formato de entrada de diferentes programas o paquetes de análisis tales como: [PHYLIP](#) (el cual le dio el nombre), [LARD](#), [PLATO](#), [SPOT](#); entre otros.

*Interleaved*

6 39

Archaeopt	CGATGCTTAC	CGCCGATGCT
Hesperorn	CGTTACTCGT	TGTCTCTGCC
Baluchith	TAATGTTAAT	TGTCGATGCT
B. virgin	TAATGTTTCGT	TGTCGAAGCA
Brontosau	CAAAACCCAT	CATCGTTGCT
B. subtili	GGCAGCCAAT	CACCGAAGCT

TACCGCCGAT	GCTTACCGC
CGTTGTCGTT	ACTCGTTGT
AATTGTTAAT	GTTAATTGT
CGTTGTTAAT	GTTTCGTTGT
CATCATCAA	ACCCATCAT
AATCACGGCA	GCCAATCAA

## Formato PHYLIP (\*.phy): Secuencial

6 39

```
Archaeopt  CGATGCTTAC  CGCCGATGCT  TACCGCCGAT  GCTTACCGC
Hesperorn  CGTTACTCGT  TGTCTCTGCC  CGTTGTCGTT  ACTCGTTGT
Baluchith  TAATGTTAAT  TGTCGATGCT  AATTGTTAAT  GTTAATTGT
B. virgin  TAATGTTTCG  TGTCGAAGCA  CAAAACCCAT  CATCGTTGC
Brontosau  CGATGCTTAC  CGCCGATGCT  TACCGCCGAT  GCTTACCGC
B. subtili TAATGTTTCG  TGTCGAAGCA  CAAAACCCAT  CATCGTTGC
```

**GenBank:** este formato es el utilizado para los reportes de secuencias tanto de nucleótidos como de aminoácidos (genPept) en el [GenBankS](#). Se caracteriza por proporcionar al usuario una información bastante completa sobre la secuencia.

```
LOCUS          CTRL          1184 bp      mRNA          PRI          19-MAR-1999
DEFINITION     Homo sapiens chymotrypsin-like (CTRL) mRNA.
ACCESSION      NM_001907
NID            g4503136
VERSION        NM_001907.1  GI:4503136
KEYWORDS       .
SOURCE         human.
  ORGANISM     Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 1184)
  AUTHORS      Larsen,F., Soliheim,J., reseland,J., Thorsen,L., Eriksen,J.A. and
                Prydz,H.
  TITLE        Molecular cloning and immunological detection of a novel
                Chymotrypsin-like pancreatic protease
  JOURNAL      Unpublished
FEATURES       Location/Qualifiers
  source       1..1184
                /organism="Homo sapiens"
                /db_xref="taxon:9606"

                /map="16q22.1"
                /clone="C1 to C25"
```

```
gene      /tissue_type="pancreas"  
          /clone_lib="Clontech cDNA lambd_gt11 HL11630"  
          1..1184  
          /gene="CTRL"  
          /db_xref="MIM:118888"  
          /db_xref="LocusID:1506"  
sig_peptide 11..64  
          /product="chymotrypsin-like"  
CDS       11..805  
          /gene="CTRL"  
          /EC_number="3.4.21.-"  
          /codon_start=1  
          /db_xref="MIM:118888"  
          /translation="MLLLSLTSLVLLGSSWGCIPAIKPALSFSQRIVNGENAVLGS  
          WPWQVSLQDSSGFHFCGSLISQSWVVTAAHCNVSPGRHFVVLGEYDRSSNAEPLQVL  
          SVSRAITHPSWNSTTMNNDVTLLKLASPAQYTTTRISPVCLASSNEALTEGLTCVTTGW"  
misc_feature 65..109  
          /note="activation peptide"  
mat_peptide 110..802  
          /product="chymotrypsin-like"  
polyA_signal 836..841  
          /note="A, primary"
```

BASE COUNT      258 a      361 c      307 g      258 t

ORIGIN

```
1 atctgccacg atgttgctgc tcagcctgac cctaagcctg gttctcctcg gctcctcctg  
61 gggctgcggc attcctgcca tcaaaccggc actgagcttc agccagagga ttgtcaacgg  
121 ggagaatgca gtggtgggct cctggccctg gcagggtgcc ctgcaggaca gcagcggctt  
181 ccacttctgc ggtggttctc tcatcagcca gtctctgggtg gtcactgctg cccactgcaa
```

**EMBL:** los archivos en este formato son reportes generados por la base de datos de secuencias de nucleótidos y de proteínas respaldada por el **EMBL** ([European Molecular Biology Laboratory](#))

```
ID ECGOR          standard; DNA; PRO; 1500 BP.
AC M13141;
SV M13141.1
NI g146247
DT 19-SEP-1987 (Rel. 13, Created)
DT 19-APR-1990 (Rel. 23, Last updated, Version 1)
DE E.coli gor gene encoding glutathione reductase, complete cds.
KW flavoprotein disulfide oxidoreductase; glutathione reductase; gor gene.
OS Escherichia coli
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP 1-1500
RX MEDLINE; 86243410.
RA Greer S., Perham R.N.;
RT "Glutathione reductase from Escherichia coli: ...
RL Biochemistry 25:2736-2742(1986).
DR SWISS-PROT; P06715; GSHR_ECOLI.
CC Potential -35 and -10 regions are located at positions 21-26 and
CC 43-48 respectively, and a potential stem-loop terminator is found
CC at 1438-1468.
XX
FH Key           Location/Qualifiers
```



```
FT source 1..1500
FT /organism="Escherichia coli"
FT /db_xref="taxon:562"
FT CDS 79..1431
FT /db_xref="PID:g146248"
FT /db_xref="SWISS-PROT:P06715"
FT /note="glutathione reductase (EC 1.6.4.2)"
FT /transl_table=11
FT /protein_id="AAA23926.1"
FT /translation="MTKHYDYIAIGGGSGGIASINRAAMYQKCALIEAKELGGTCVNV
FT GCVPPKVMWHAAQIREAIHMYGPDYGFDTTINKFNWETLIASRTAYIDRIHTSYENVLG
FT EKIVGIHGIGFGMDEMLQGFVAVALKMGATKKDFDNTVAIHPTAAEEFVTMR"
XX
SQ Sequence 1500 BP; 369 A; 375 C; 420 G; 336 T; 0 other;
GGAGTAATTG CAGCCATTGC TGGCACCTAT TACGTCTCGC GCTACAATCG CGGTAATCAA 60
CGATAAGGAC ACTTTGTCAT GACTAAACAC TATGATTACA TCGCCATCGG CGGCGGCAGC 120
GGCGGTATCG CCTCCATCAA CCGCGCGGCT ATGTACGGCC AGAAATGTGC GCTGATTGAA 180
GCCAAAGAGC TGGGCGGCAC CTGCGTAAAT GTTGGCTGTG TGCCGAAAAA AGTGATGTGG 240
CACGCGGCGC AAATCCGTGA AGCGATCCAT ATGTACGGCC CGGATTATGG TTTTGATACC 300
ACTATCAATA AATTCAACTG GGAAACGTTG ATCGCCAGCC GTACCGCCTA TATCGACCGT 360
```

//

**Formato Nexus:** Es el formato de archivo utilizado por PAUP, PAML, entre otros para análisis evolutivos y filogenético. Se caracteriza por iniciar con el Numeral seguido por la palabra NEXUS. En muchos casos es necesario editarlo “manualmente” según el tipo de análisis que se quiera usar.

```
#NEXUS
```

```
[TITLE: Written by EMBOSS 22/01/02]
```

```
begin data;
```

```
dimensions ntax=2 nchar=2016;
```

```
format interleave datatype=DNA missing=N gap=-;
```

```
matrix
```

```
HSFAU          ttctcttttctcgactccatcttcgcggtagctgggaccgccgttcagtc
```

```
HSFAU1        ctaccattttccctctcgattctatatgtacactcgggacaagttctcct
```

```
HSFAU          gccaatatgcagctctttgtccgcccaggagctacacaccttcgaggt
```

```
HSFAU1        gatcgaaaacggcaaaactaaggccccaagtaggaatgccttagttttcg
```

```
HSFAU          gaccggccaggaaacggctcgcccagatcaaggctcatgtagcctcactgg
```

```
HSFAU1        gggttaacaatgattaacactgagcctcacaccacgcgatgccctcagc
```

```
HSFAU          agggcattgccccggaagatcaagtcgtgctcctggcagggcgccccctg
```

```
HSFAU1        tcctcgctcagcgcctctcaccaacagccgtagccccgcagccccgctggac
```

## Formato MEGA (\*meg):

#Mega

!Title cytb.todos.aln;

#Branchiostomafloridae

```
-----ATGTC---TGGTCCTTTGCGGAAACAT-CATCC
GTTACTAAAAGTTGTGAATCATTCTGTGATTGACTTGCCCTGTCCCTAGCAATATTTCTGT
TATGTGGAATTTTGGATCCTTATTGGGTCTTTGCTTAGTGAGTCAAATTTTACTGGGCT
CTTTTGGCAATACACTATACAGCTGATGTAAATTTGGCTTTTTCATCTGTAGCTCATAT
CTGTCGAGATGTTAACTATGGGTGGTTATTGCGTAATCTTCATGCTAATGGTGCATCTTT
CATATTTATTTGTTTGTATATACATATCGGGCGAGGGCTATACTACGGGTCCTACTTTTA
TCGAGAGACATGAAACATTGGTGTAAATGTTGCTAGTTCTAACTATAGCAACCGCCTTTT
AGGTTATGTTCTTCCCTGGGGGCAGATGTCTTTTGGGGTGCTACTGTTATTACTAATTT
ATTTTCAGCAATCCCTTATCTCGGGCCAGACTTGGTTCAGTGGCTGTGAGGTGGATTCTC
TGGGGTATGTAGAGAATAAGCTTTTGTTTAGTTAG-----
-----
```

#Sibynophis

```
-----ATGCC-----CAAC-----CATTC
GCTTATA-----TTATTCAACCTCTTACCAGTCGGATCAAATATCTCCAC
GTGATGAACTTCGGCTCAATACTACTATCCTGTTCAATCCTACAAACATTGACCGGCTT
TTTTCTAGCAATTCACTATACAGCCAACATCAACCTAGCATTCTCATCTATCATTCATAT
CACACGAGACGTACCATACGGCTGAATTATACAAAACCTACATGCCATCGGAGCATCTAT
ATTCTTCATCTGTATCTATGTACACATCGCACGAGGGCTGTACTACGGATCCTACCTAAA
TAAAGAAGTCTGACTATCAGGAACCATACTATTAGTTATTCTAATAGCAACAGCCTTCTT
TGGCTATGTATTACCATGAGGACAAATATCATTCTGAGCAGCAACAGTAATTACCAACTT
GCTCACAGCTGTACCATATTTAGGTAATACCCTTACAACCTGACTTTGAGGCGGGTTCTC
TGGAACCTAATCAGACATCGACAAAATCCCATTCCACCCTACCCTCTCATAAAGACGC
```