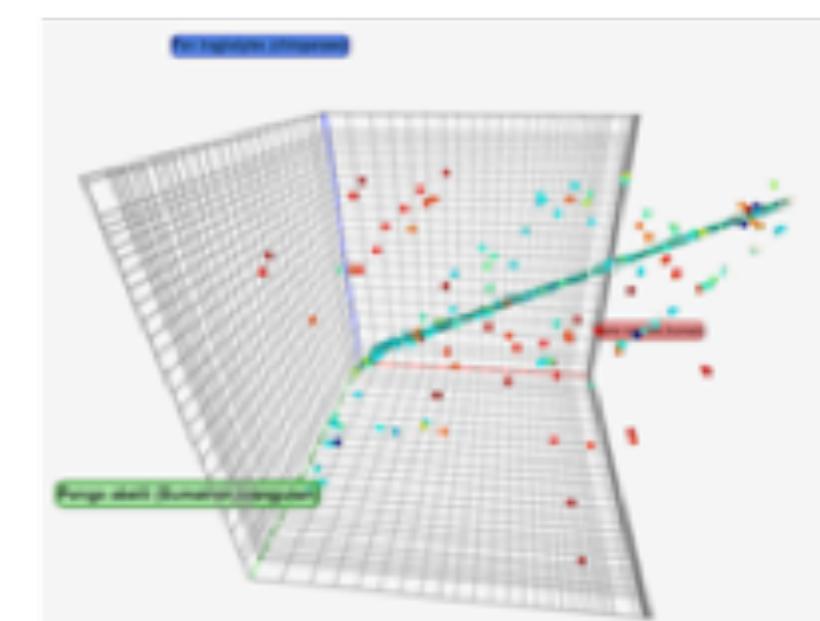
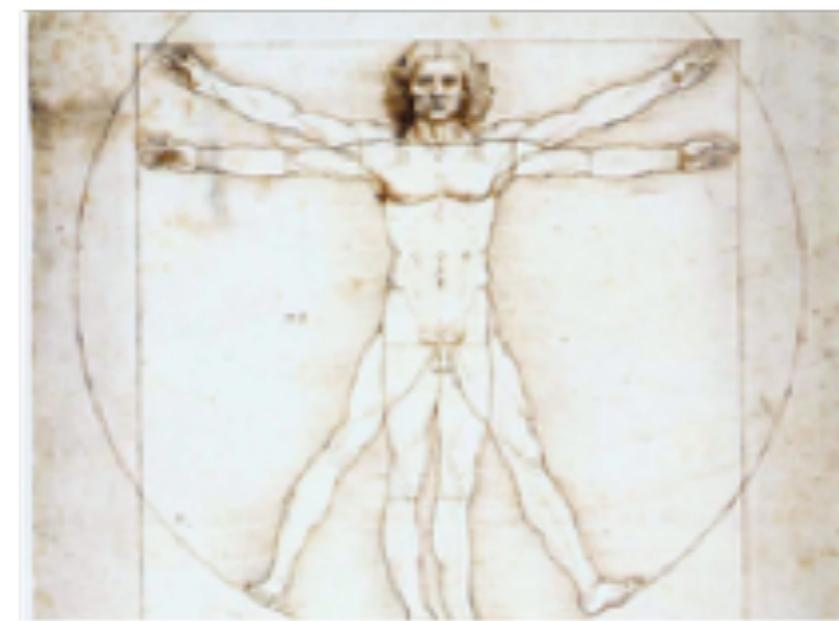


# Computational Genomics

## Genome Files

### FASTA Format



# Fasta Format

## Organizing Data Into Tables

### About Data and Tables

**Column 01**

**Column 02**

**Row 01**

**Row 02**

Variable-01	Variable-02
Observation-01	Observation-01
Observation-02	Observation-02
Observation-03	Observation-03
Observation-04	Observation-04
Observation-05	Observation-05
Observation-06	Observation-06
Observation-07	Observation-07

# Fasta Format

## Organizing Data Into Tables

### About Data and Tables

#### The Atomic Principle of Data Structure

- **Columns for variables and rows for observations**
- **Never enter more than one piece of information in a cell**
- **Always use descriptive fields names and never include spaces, numbers, or special characters of any kind. Use underscores ( \_ ), or dashes ( - ), instead.**
- **Example: Dates as data:**

November 15

November 15, 2023

15-11-2023

2023-11-15

2023-11-15:15:25:25

# Fasta Format

## Organizing Data Into Tables

### About Data and Tables

#### The Atomic Principle of Data Structure

- Example: Dates as data:

**YYYY:** the full year, i.e. 2023

**MM:** the month, i.e. 11

**DD:** the day of month, i.e. 15

**hh:** hour of day, i.e. 15

**mm:** minutes, i.e. 25

**ss:** seconds, i.e. 25

Year	Month	Day	Time
2023	11	15	15:25:25

Year	Month	Day	Hr	Min	Sec
2023	11	15	15	25	25

# Fasta Format

## Organizing Data Into Tables

Table containing  
12 fields (columns) and 15 records (rows or lines)

Field 01  
Record 01

Field 01  
Record 08

Fields Separators

Records Separators

1	2	3	4	5	6	7	8	9	10	11	12
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											

# Fasta Format

## Organizing Data Into Tables

### About Invisible Characters

```
>ENSPF0T00000024091.1 cds scaffold:PoeFor_5.1.2:KI520192.1:72966:81842:-1
gene:ENSPF0G0000000548.2 gene_biotype:protein_coding
transcript_biotype:protein_coding description:transmembrane protein 88
[Source:NCBI gene;Acc:103129828]
ATGTGTGGTGTGGATGTGGACCTGGAGGACGGCGGCTCAGCGGACGAGGAGAAGGAGGAG
GAGTTCTGGGTGGGGACGCCGTAAAGATGCTGCCCCCCCAGTGGCGCACAGTGACGGC
AGTCGTGGGCGGCCGCAGGGAGCGTGCCTCGGTGGCTGCAGGGCGTTCTGGTC
CTGTGGAACCTGTGCGTGGTGTCCGCCGGCGCCCTGCTGCTGGCGCTGGTCTTCTGTG
GTGCTTCTGCCTGCGCGCTGCTGTACGCCGGCTTCCTCTGCCACTCCAGAGTCCTC
GACGCCCGGCAGCTATGCCGCTACCTCGACGACAACAGCTGCTCCGCCCTCATCATC
CTGGGCTTCTGTAATGATGTCGCCCTCGTAGTTTGGCGGGCGGTTTCTGCGGGCTG
CTGAGGAAGTTCCGGCTCCTCTTCCAGCCGCTGTCCAGGGCGCGGTACCGCGGG
AGGCTGCTGGACTGGGTGGGAAGAGTCCAGGCTTGGGTGTGA
```

# Fasta Format

## Organizing Data Into Tables

### About Invisible Characters

```
>ENSPF0T00000024091.1 · cds · scaffold:PoeFor_5.1.2:KI520192.1:72966:81842:-1  
gene:ENSPF0G0000000548.2 · gene_biotype:protein_coding · ←  
transcript_biotype:protein_coding · description:transmembrane · protein · 88 ·  
[Source:NCBI · gene;Acc:103129828] ←  
ATGTGTGGTGTGGATGTGGACCTGGAGGACGGCGGCTCAGCGGACGAGGAGAAGGAGGAG ←  
GAGTTCTGGGTGGGGACGCCGTAAAGATGCTGCCCCCCCAGTGGCGCACAGTGACGGC ←  
AGTGCCTGGGGCGGCCGCAGGGGAGCGTGCCTCGGTGGCCTGCAGGGCGGTTCTGGTC ←  
CTGTGGAACCTGTGCGTGGTGTCTGGCCGGCGCCCTGCTGGCGCTGGTCTTCTGTGTG ←  
GTGCTTCTGCCTGCGCGCTGCTGTACGCCGGCTTCCTCTGCCACTCCAGAGTCCTC ←  
GACGCCCGGCAGCTATGCCGCTACCTCGACGACAACAGCTGCTCCGCCATCATC ←  
CTGGGCTTCTGTAATGATGTCGCCCTCGTAGTTTGGCGGGCGGTTTCTGCGGGCTG ←  
CTGAGGAAGTTCCGGCTCCTCTTCCAGCCGCTGTCCAGGGCGCGGTACCGCGGG ←  
AGGCTGCTGGACTGGGTGGGAAGAGTCCAGGCTTGGGTGTGA ←
```

# Fasta Format

## Organizing Data Into Tables

Field 01    Field 02

Record 01 →  
Record 02 →

ID_Seq_01	Seq
2	
3	
4	
5	
6	
7	
8	

# Fasta Format

A sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line (defline) is distinguished from the sequence data by a greater-than (">") symbol at the beginning. It is recommended that all lines of text be shorter than 80 characters in length. An example sequence in FASTA format is:

Seq ID: { >ENSPF0T00000000487.2 cds scaffold:PoeFor\_5.1.2:KI520192.1:729  
66:83042:-1 gene:ENSPF0G00000000548.2 gene\_biotype:protein\_coding transcript\_biotype:protein\_coding description:transmembrane protein 88 [Source:NCBI gene;Acc:103129828]  
ATGTGTGGTGTGGATGTGGACCTGGAGGACGGCGGCTCAGCGGACGAGGAGAAGGAGGAG  
GAGTTCTGGGTGGGGACGCCGTAAAGATGCTGCCCCCCCAGTGGCGCACAGTGACGGC  
AGTGCCTGGGGCGGCCGCAGGGGAGCGTGCCTGGCTGGCCTGCCGGGGCGGTTCTGGTC  
CTGTGGAACCTGTGCGTGGTGTCCGCCGGCCCTGCTGCTGGCGCTGGTCTTCTGTGTG  
GTGCTTCTGCCTGCCCGCTGCTGTACGCCGGCTTCCTCTGCCACTCCAGAGTCCTC  
GACGCCCGGCAGCTATCTGCCGCTACCTCGACGACAACAGCTGCTCCGCCCTCATCATC  
CTGGGCTTCGTAATGATGTCGCCCTCGTAGTTTGGCGGCGGGTTTCTGCGGGCTG  
CTGAGGAAGTTCCGGCTCCTCTTTCCAGCCGCTGTCCAGGGCGCGTACCGCGGG  
AGGCTGCTGGACTGGTGGGAAGAGTCCAGGCTTGGGTGTGA

Seq: {

Record 01

|-----Sequence-----|

# Fasta Format

**Seq ID:**

```
>ENSPF0T0000000487.2 cds scaffold:PoeFor_5.1.2:KI520192.1:729  
66:83042:-1 gene:ENSPFOG0000000548.2 gene_biotype:protein_coding transcript_biotype:protein_coding description:transmembrane protein 88 [Source:NCBI gene;Acc:103129828]
```

**Seq:**

```
ATGTGTGGTGTGGATGTGGACCTGGAGGACGGCGGCTCAGCGGACGGAGAAGGAGGAG  
GAGTTCTGGGTGGGGACGCCGTGAAGATGCTGCCCCCCCAGTGGCGCACAGTGACGGC  
AGTCGTGGGGCGGCCGCAGGGGAGCGTGCAGCTCGGTGGCCTGCGGGCGGTTCTGGTC  
CTGTGGAACCTGTGCGTGGTGTGGCCGGCGCCCTGCTGCTGGCGCTGGTCTTGTGTG  
GTGCTTCTGCCTGCAGCGCTGCTGTACGCCGGCTTCTCTGCCACTCCAGAGTCCTC  
GACGCCCGCAGCTATCTGCCGTACCTCGACGACAACAGCTGCTCCGCCATCATC  
CTGGGCTTCGTAATGATGTCGCCCTCGTAGTTTGGCGGCCGGTTCTGCGGGCTG  
CTGAGGAAGTTCCGGCTCCTCTTCCAGCCGCTGTCCAGGGCGCGGTACCGCGGG  
AGGCTGCTGGACTGGGTGGGAAGAGTCAGGCTGGGTGTGA
```

**Seq ID:**

```
>ENSPF0T00000024091.1 cds scaffold:PoeFor_5.1.2:KI520192.1:729  
66:81842:-1 gene:ENSPFOG0000000548.2 gene_biotype:protein_coding transcript_biotype:protein_coding description:transmembrane protein 88 [Source:NCBI gene;Acc:103129828]
```

**Seq:**

```
ATGTGTGGTGTGGATGTGGACCTGGAGGACGGCGGCTCAGCGGACGGAGAAGGAGGAG  
GAGTTCTGGGTGGGGACGCCGTGAAGATGCTGCCCCCCCAGTGGCGCACAGTGACGGC  
AGTCGTGGGGCGGCCGCAGGGGAGCGTGCAGCTCGGTGGCCTGCGGGCGGTTCTGGTC  
CTGTGGAACCTGTGCGTGGTGTGGCCGGCGCCCTGCTGCTGGCGCTGGTCTTGTGTG  
GTGCTTCTGCCTGCAGCGCTGCTGTACGCCGGCTTCTCTGCCACTCCAGAGTCCTC  
GACGCCCGCAGCTATCTGCCGTACCTCGACGACAACAGCTGCTCCGCCATCATC  
CTGGGCTTCGTAATGATGTCGCCCTCGTAGTTTGGCGGCCGGTTCTGCGGGCTG  
CTGAGGAAGTTCCGGCTCCTCTTCCAGCCGCTGTCCAGGGCGCGGTACCGCGGG  
AGGCTGCTGGACTGGGTGGGAAGAGTCAGGCTGGGTGTGA
```

**Seq ID:**

```
>ENSPF0T0000000511.2 cds scaffold:PoeFor_5.1.2:KI520192.1:873  
71:97453:-1 gene:ENSPFOG0000000555.2 gene_biotype:protein_coding transcript_biotype:protein_coding description:neuroblastoma suppressor of tumorigenicity 1 [Source:NCBI gene;Acc:103129830]
```

**Seq:**

```
ATGTGGCAGAGGATTCAAATTGCTGTGCGCTGTTGCGCTGCATTAGCGGCAGCGCCT  
GCACACATCAACCGCCTGGCGCTGTTCCCTGACAAGAGCGCCTGGTGCAGCCAAGAAC  
ATCACTCAGATTGTCGGACACACAGGATGTCAGCCGCGCTCTATCCAAAACAGAGCCTGT  
CTGGGCCAGTGCTTAGCTACAGCGTCCAAACACGTTCCCACAGTCCACCGAGTCCCTG  
GCACACTGTGACTCCTGCACGCCCTGCCAGACACAGTGGGAAGTGGTGAECTCTGGATGC  
CCCGCAGCGAGGATTCTCCACGCGTGGACAAGTTGGTGGAGCGGATCCTGCACTGCAGC  
TGCCAGTCCTGCAGTAAGGAAAGTGTCCAGGAGGGGGCGGTGATGCAGCTATATCAGCA  
GACAACACCCCTGGATTCAAGCTCCGTCTGTCCGACCTCTGATCGGAGACCCGGCCAC  
TCTCTGCCACGCAAGACAGCCTGTCTGACAAGCAGGCCACCCCTACGCGCACCACT  
CCACGCTGCCACGCGTCAGACGGAGGCTAG
```

**Seq ID:**

```
>ENSPF0T00000027073.1 cds scaffold:PoeFor_5.1.2:KI520192.1:103  
580:109381:-1 gene:ENSPFOG00000022757.1 gene_biotype:protein_coding transcript_biotype:protein_coding description:MICOS complex subunit Mic10-like [Source:NCBI gene;Acc:103129831]
```

**Seq:**

```
ATGGCAGACGAGCTGGGCAGAAATGGGACCGCTGTCTGCTGACACAGCCTAAAAACA  
GCGACTGGCGTCGGTATCGGCATCGTGTCTCCGTCTTCTTAAACGTCGAACGTGG  
CCTGTGTCGTTGGATCGGGCCTGGGGTTGGGAATGGGATACAGCAACTGCCAGCATGAC  
TTCAGGTCAACCATATCTGATTCAAGGACCGAGTGA
```



# Fasta Format

Seq ID:

>ENSPF0T0000000487.2 cds scaffold:PoeFor\_5.1.2:KI520192.1:729  
66:83042:-1 gene:ENSPFOG0000000548.2 gene\_biotype:protein\_coding transcript\_biotype:protein\_coding description:transmembrane protein 88 [Source:NCBI gene;Acc:103129828]

ATGTGTGGTGTGGATGTGGACCTGGAGGACGGCGCTCAGCGGACGAGGAGAAGGAGGAG  
GAGTTCTGGTGGGGACGCCGTAAAGATGCTGCCCCCCCAGTGGCGCACAGTGACGGC  
AGTCGTGGGGCGGCCAGGGAGCGTGCAGGGCTCGGTGGCCTGCCACTCCAGAGTC  
CTGTGGAACCTGTGCGTGGTGCAGGCCCTGCTGGCGCTGGTCTTGTGTG  
GTGCTTCTGCCTGCCGCTGCTGACGCCCTCTGCCACTCCAGAGTC  
GACGCCCGGAGCTATCTGCCCTACCTGACGACAACAGCTGCTCCGCCCTCATCATC  
CTGGGCTTCGTAATGATGTCGCCCTCGTAGTTTGGCGGCGGGTTCTGCGGGCTG  
CTGAGGAAGTCCGGCTCCTCTTCCAGCCGTGTCAGGGCGCGGTACCGCGGG  
AGGCTGCTGGACTGGTGGGAAGAGTCCAGGCTGGGTGTGA

>ENSPF0T0000024091.1 cds scaffold:PoeFor\_5.1.2:KI520192.1:729  
66:81842:-1 gene:ENSPFOG0000000548.2 gene\_biotype:protein\_coding transcript\_biotype:protein\_coding description:transmembrane protein 88 [Source:NCBI gene;Acc:103129828]

ATGTGTGGTGTGGATGTGGACCTGGAGGACGGCGCTCAGCGGACGAGGAGAAGGAGGAG  
GAGTTCTGGTGGGGACGCCGTAAAGATGCTGCCCCCCCAGTGGCGCACAGTGACGGC  
AGTCGTGGGGCGGCCAGGGAGCGTGCAGGGCTCGGTGGCCTGCCACTCCAGAGTC  
CTGTGGAACCTGTGCGTGGTGCAGGCCCTGCTGGCGCTGGTCTTGTGTG  
GTGCTTCTGCCTGCCGCTGCTGACGCCCTCTGCCACTCCAGAGTC  
GACGCCCGGAGCTATCTGCCCTACCTGACGACAACAGCTGCTCCGCCCTCATCATC  
CTGGGCTTCGTAATGATGTCGCCCTCGTAGTTTGGCGGCGGGTTCTGCGGGCTG  
CTGAGGAAGTCCGGCTCCTCTTCCAGCCGTGTCAGGGCGCGGTACCGCGGG  
AGGCTGCTGGACTGGTGGGAAGAGTCCAGGCTGGGTGTGA

>ENSPF0T0000000511.2 cds scaffold:PoeFor\_5.1.2:KI520192.1:873  
71:97453:-1 gene:ENSPFOG0000000555.2 gene\_biotype:protein\_coding transcript\_biotype:protein\_coding description:neuroblastoma suppressor of tumorigenicity 1 [Source:NCBI gene;Acc:103129830]

ATGTGGCAGAGGATTCAAATTGCTGCGCTTGGCTGCATTAGCGGCAGGCC  
GCACACATCAACGCCCTGGCGCTGTCCTGACAAGAGCGCCTGGTGCAGCCAAGAAC  
ATCACTCAGATTGTCGGACACACAGGATGTCAGCCGCTCTACAAAACAGAGCCTGT  
CTGGGCCAGTCTTACAGCGTCCAAACACGTTCCACAGTCCACCGAGTCCTG  
GCACACTGTGACTCCTGCACGCCAGACACAGTGGGAAGTGGTACTCTGAATGC  
CCCGGCAGCGAGGATTCTCACGCGTGGACAAGTGGTGGAGCGGATCTGCACTGCAGC  
TGCCAGTCTGCAGTAAGGAAAGTGTCCAGGAGGGGGCGGTGATGCAAGCTATATCCAGCA  
GACAACACCCTGGATTCAAGCTCCGTCTGTCCGACCCCTGATGGAGACCCGGCCAC  
TCTCTGCCCAACGAGACAGCCTGTGACAAGCAGGCCACCCACGCGCACCA  
CCCACGCTGCCCAACGCGTCAGACGGAGGCTAG

>ENSPF0T0000027073.1 cds scaffold:PoeFor\_5.1.2:KI520192.1:103  
580:109381:-1 gene:ENSPFOG00000022757.1 gene\_biotype:protein\_coding transcript\_biotype:protein\_coding description:MICOS complex subunit Mic10-like [Source:NCBI gene;Acc:103129831]

ATGGCAGACGAGCTGGCGAAAATGGGACCGCTGTCTGCTGACACAGCCTAAAAACA  
GCGACTGGCGTCGGTATCGGCATCGTGTCTCCGTCTTAAACGTCGAACGTGG  
CCTGTGTCGTTGGATCGGGCTGGGGTGGGAATGGGATACAGCAACTGCCAGCATGAC  
TTCAGGTACCATATCTGATTGATGCCGCTGGTTAAGGACCAAGTGA

>ENSPFOP0000024938.1 pep scaffold:PoeFor\_5.1.2:KI520192.1:179  
93:27836:-1 gene:ENSPFOG0000000511.2 transcript:ENSPF0T000000  
4762.1 gene\_biotype:protein\_coding transcript\_biotype:protein\_coding

MDDQRRLDFSLIPQIILHRIDSLQCNHQQERRSTLDQEELEPLKGKQDQEETEDHQIK  
EEKEDLKHQQIKEEQEDLKHQQIKEEQEDLKHQQIKEEVTEVYCSQGE  
EWIELKQETDCMVFPADEQTYQTESEQSGNKVIFQEPAEAENQNQERRKPFSCVICKKA  
FSGKYALNNHIRTHTGEKPFSCGNCGECFSHQQLLTQHMRIRGKPFSCVNGKSFQSQK  
RHLTQHLVIHTGEKPFSCGNCGKRFSSQPKNLTRHMMIHTGEKPFSCVNGKSFQSQKQHIT  
QHMKIHTCVKPFSCVNGCESFSHKQYLTHQMMIHTGEKLFCLCVNGKRFSSQPKPLSQHMR  
IHTGEKPFSCVNGKRFSSQPKNLTRHMMIHTGEKPFSCVNGKSFQSQPKNLTRHMMIHTG  
EKPFSCVNGKSFQSQPKNLTRHMMIHTGEKPFSCVNGKSFQSQPKNLTRHMMIHTG

>ENSPFOP0000025428.1 pep scaffold:PoeFor\_5.1.2:AYCK01029644.1  
:207:5683:-1 gene:ENSPFOG00000024132.1 transcript:ENSPF0T0000  
028006.1 gene\_biotype:protein\_coding transcript\_biotype:protein  
\_coding description:trafficking protein particle complex subu  
nit 8-like [Source:NCBI gene;Acc:103133310]

QVFFLQFPTALLCGEIRKAYVEFCNVSAVALCGLRVASTHPDFFTGSQSTAPTPPSPD  
SAENCSAYVTPSPLQPGSMSSETLVAPACSQPSRVTEIPIHGGTLQPGQSVRLPLWLR  
GPDQEGVHEINFLYYENTEKGLKISHRVLRHTVFICASRSLSVQASACRSSASPHHSDL  
DKGSGGTLVFDVENVNASESGVREFHIVQVSSSSQHWRLLHYINPCNNYDCRLSSRERA  
KLCFKAMRCKPHEGSDGVKEYTFADLNLngERIMSSSTPCGDFFFSCRQSSVLEQVNSAP  
RNSRSQSGSGHSDNTASDITNIAKKCNELDLNIIIVIWKAYVVEDNQLILEGQLHVALKTF  
GKEASSLIPKEETQEMMLLRFKSELPPPVILPYPELSQLIRTNLCYPETFTHF

>ENSPFOP0000000483.2 pep scaffold:PoeFor\_5.1.2:KI520192.1:385  
00:47984:-1 gene:ENSPFOG0000000529.2 transcript:ENSPF0T00000  
00484.2 gene\_biotype:protein\_coding transcript\_biotype:protein  
\_coding description:von Willebrand factor A domain-containing  
protein 1-like [Source:NCBI gene;Acc:103129824]

MKSFLRCIFVFAMILQQSIRQTAASPDTELNCCEGDLILLLDSSGSVTSEFSSFLSFAA  
SLLYPFTLGRGHVRVALLLVGTTPHLEFSLDVHSNQESLLRALQSVNQQQGDNTKAAIE  
VAQRLLSEADGDVPKVLLWLTDGVQTGDVEKAMSELKARGVYVLIVYTIHGNHRVLQRVA  
TPPLESHLYSVDMESIDIITDDLREAIKIIIRAERLSVVRLTSHAVLQWRPVLAADSGH  
YELSYKAVDDQHSGKTTVLPSSSSQTELIQLQPDTTYTASLRPESNQRLHSTLSVQFTTL  
PDVLSPAEVTLSDPGPRQVRVSWGPLQPDRVQRTLEYGAIPSGRVHTVDVSNQKSSVFL  
RNLEPGTRYLITVSALHRDGKERAMSVRACTQEEARPALTLRLTLTDRQEGNEVQASWE  
TNTEGLKGWVSWERKGSQNSQSERSPSTAYLPPSPSSVRLAHLAPNSRVCVSPVYSSGR  
GDGICCTAKNSR

>ENSPFOP0000000486.1 pep scaffold:PoeFor\_5.1.2:KI520192.1:729  
66:83042:-1 gene:ENSPFOG0000000548.2 transcript:ENSPF0T00000  
00487.2 gene\_biotype:protein\_coding transcript\_biotype:protein  
\_coding description:transmembrane protein 88 [Source:NCBI gene  
;Acc:103129828]

MCGVDVDLEDGGSADEEKEEFWVGDAVKMLPPPVAHSDGSAWGGRRGACGSVACGAVLV  
LWNLCVVSAGALLALVFCVLLPAALLLYAGFLCHSRVLDAPAICRYL  
LGFVMMSPVVLAAGVFCGLLRKFRLLLFLQPLSRARYRGRLLDWGRVQ



# Fasta Format

HPRC Kaiser Galaxy

Workflow Visualize Shared Data Admin Help User

Using 0%

Tools

search tools

Upload Data

This dataset is large and only the first megabyte is shown below.  
Show all | Save

ENSPFOP00000000459.1 pep scaffold:PoeFor\_5.1.2:K1520192.1:5697:27330:1 gene:ENSPFOG00000000511.2 transcript:ENSPFOT00000000460.1 gene\_biotype:protein\_coding  
SSPDSLQCYPEAEAENQNQERRKPFSCVICKRGFTVKYALNNHIRTHTGEKPFSCVNCGK  
RFSQKRKLTLQHMIVHSSDKPFSCVNCGKKFIHKRNLTQHMRHSQEKFSCVNCGKRSQ  
KRNLTQHMMIHSSEKFSCVNCGKSFIFIHKRNLTQHMRHTGEKPFSCVNCGKTFQRKRN  
TQHMIHTSEKPFSCVNCGKCFQKQHLTRHLMHTGEKPFSCVNCGKSFGKQHLTQQM  
RIHSGEKPFSFSCGNCGKSFQSQKHVLTQHMMIHTGEKPFSCVNCGKSFSRKQILTQHMT  
IHTGEKPFSCGNCGKSFQSQKVSLTQHMTIHTGEKPFSCGNCGKSFQSQKVSLTQHMTIHTG  
EKPFSCGNCGKSFQSQKLNLTLHMKIHTGEKPFSCVNCGKSFSRKYLTVLQHMSHTGEKPF  
SCGNCGKGFSQKLNLTRHMMIHTGEKPFSCGNCGKSFQSQKHALTQHMRHTGEKPFSCGN  
CGKGFSQKLNLTRHMMIHTGEKPFSCGNCGKSFQSQKHALTQHMRHTGEKPFSCGNCGK  
FSHVNLTLQHMRHTGEKPFSCGNCGKSFNLKPTLTRHMRHTGEKPFSCGNCGKSFQSQ  
KQHVTQHMTIHTGEKPFSCGNCGKSFQHKHLTQHVIHNGEKFSCVNCGKSFQSQKLSL  
TRHMMIHTGEKPFSCVNCGKSFNQKKNLTQHMIHTGEKPFSCVNCGKSFSHKQNFTQHM  
RIHSDEKPFSCVNCGKSFSRKYALNIHRTHTGEKPFSCGNCGKSFQGKQLTRHMMIHT  
GEKPFSCVNCGKSFSSKHILTQHMMIHTGEKPFSCVNCGKSFSHKRLTQHMRHTGEKPF  
FSCVNCGKSFQSQKVSVTRHMMIHTGEKPLCGNCRGSFSYKQHFTQHMRHTGEKPFSCG  
NCGKSFSHKCDLNHVRMIRHGKKPFSCVNCGKSFQSQKRHLTQHVLVIHTGEKPFSCGNCG  
RFSQKPNLTRHMMIHTGEKPFSCVNCGKSFQSQKHITQHMKIHTCVKPFSCVNCGKRSQ  
KPLLSQHMRHTGEKPFSCVNCGKRFQSKPNLTRHMMIHTGEKPFSCVNCGKSFQSQPNL  
TRHMMIHTGEKPFSCVNCGKSFQSQPKPNLTRHMMIHT  
ENSPFOP00000024938.1 pep scaffold:PoeFor\_5.1.2:K1520192.1:17993:27836:1 gene:ENSPFOG00000000511.2 transcript:ENSPFOT0000024762.1 gene\_biotype:protein\_coding  
MDDQRLLDFSLIPQIILHRIDLSQLCNIHHQERRSTLDQEELEPLKGKQDQEETEDHQIK  
EEKEDLKHHQQIKEEQEDLKHHQQIKEEQDKLKHQQIKEEQEDLKHHQQIKEVEETEVYCSQE  
EWIELKQETDTCMVFPADAEQTYQTESEQSGNKVIFQEPAEAENQNQERRKPFSCVICKKA  
FSGKYALNNHIRTHTGEKPFSCGNCGECFSHQQLLTQHMRIRHGKKPFSCVNCGKSFQSQ  
RHLTQHVLVIHTGEKPFSCGNCGKRFQSKPNLTRHMMIHTGEKPFSCVNCGKSFQSQKHT  
QHMKIHTCVKPFSCVNCGESFSHKQYLTQHMMIHTGEKLFLCVNCGKRFQSKPLLSQHMR  
IHTGEKPFSCVNCGKRFQSKPNLTRHMMIHTGEKPFSCVNCGKSFQSQPKNLTRHMMIHTG  
EKPFSCVNCGKSFQSQPKNLTRHMMIHTGEKPFSCVNCGKSFQSQPKNLTRHMMIHTGEKPF  
ENSPFOP00000025428.1 pep scaffold:PoeFor\_5.1.2:AYCK01029644.1:207:5683:-1 gene:ENSPFOG00000024132.1 transcript:ENSPFOT0000028006.1 gene\_biotype:protein\_coding  
QVFFLQFTALLCGERKAYVEFCNSAVALCGLRVASTHPDFFTGSQSTAPTPPPSD  
SAENCSAYTFPSPLQPGSMSSETLVSAPACSQPSRVEIPIHGGTLQPGQSVRLPLWLR  
GPDQEGVHEINFLFYENTEKGLKISHRVRRLHTVFICASRSLSVQASACRSSASPHSLD  
DKGSGGTIVFIDVENVNASESGVREFHVQVSSSSQHWRLHKYINPCNNYDCRLSSRERA  
KLCFKAMRCKPHEGSDGVKEYTFADLNGLNERIMSSSTPCGDFFFSCRQSSVLEQVNSAP  
RNSRSQSGHSNDNTASDITNIACKCNELDLNIIIVIKAYVVEDNQLILEGQLHVALKTF  
GKEASSLIPKEETQEMMLLRFKSELPPPVILPYPELSQLIRTNLCYPEETFTHHF  
ENSPFOP00000000483.2 pep scaffold:PoeFor\_5.1.2:K1520192.1:38500:47984:-1 gene:ENSPFOG00000000529.2 transcript:ENSPFOT00000000484.2 gene\_biotype:protein\_coding  
MKSFLLRCIFVFAMLQQSIRQTAASPDTTELNCCEGDILLLDSSGVTSEFSSFLSFAA  
SLLYPFTLGRGHVRVALLLVGTTPHLEFSLDVHSNQESLLRALQSVNQQQGDTNTKAIE  
VAQRLLSEADGDVPKVLLWLTDGVQTGDKVEKAMSELKARGVYVLIVYTIHGNHRLQRA  
TPPLESHLYSDMESIDIITDDLREAIKKIIRAERLSVRLTSQAVLQWRPVLAADSGH  
YELSYKAVDDQHSGKTTLPSSSSQTELIQLQDPDTTYSASLRPESNQRLHSTLSVQFTTL  
PDVLSPAETLSDPGPRQVRVSWGPLQPDFRVQRYTLEYGAIPSGRVHTDVSNQKSSVFL  
RNLEPGTRYLITVSALHRDGKERAMSVRACTQEEARPALTDRLTLTDRQEGNEVQASWE  
TNTEGLKGWVWSWERKGQSNSQSERPSTAYLPPSPSSVRLAH LAPNSRVCVSPVYSSGR  
GDGICCTAKTNSR  
ENSPFOP00000000486.1 pep scaffold:PoeFor\_5.1.2:K1520192.1:72966:83042:-1 gene:ENSPFOG00000000548.2 transcript:ENSPFOT00000000487.2 gene\_biotype:protein\_coding

History

search datasets

L04-A

1 shown  
24.82 MB

1: Poecilia\_formosa.Poe  
For\_5.1.2.pep.all.fa

# Fasta Format

## Compute Sequence Length

### Tools

search tools x

Upload Data

Barcode Splitter

Remove sequencing artifacts

FASTQ to FASTA converter from  
FASTX-toolkit

Filter by quality

Draw quality score boxplot

RNA/DNA converter

FASTA Width formatter

Length Distribution chart

Pear Paired-End read merger

FLASH adjust length of short reads

FASTQ Summary Statistics by  
column

FASTQ de-interlacer on paired end  
reads

FASTQ interlacer on paired end  
reads

Combine FASTA and QUAL into  
FASTQ

Fasta Statistics display summary  
statistics for a FASTA file

Filter sequences by length

Compute sequence length

Cutadapt Remove adapter  
sequences from FASTQ/FASTA

Trim sequences

RNA-seq

CD-HIT

Datamash

EMBOSS

MUMmer4

Nanopore

Picard

Built-in Converters

Compute sequence length (Galaxy Version 1.0.3)

Sequences

From History

Compute length for these sequences

1: Poecilia\_formosa.PoeFor\_5.1.2.pep.all.fa

How many title characters to keep?

0  
'0' = keep the whole thing

Strip fasta description from header?

No  
Stripping the description will truncate the fasta header to just the sequence ID. Otherwise the header description will be kept. This step is done before the 'How many characters to keep' option.

Job Resource Parameters

Specify job resource parameters

Memory (GB)

7

Maximum Job Memory

Time (hours)

24

Maximum job time

Execute

### What it does

This tool counts the length of each fasta sequence in the file. The output file has two columns per line (separated by tab): fasta titles and lengths of the sequences. The option *How many characters to keep?* allows to select a specified number of letters from the beginning of each FASTA entry.

### Example

Suppose you have the following FASTA formatted sequences from a Roche (454) FLX sequencing run:

```
>EYKX4VC02EQL05 length=108 xy=1826_0455 region=2 run=R_2007_11_07_16_15_57_
TCGGCGCCGAGCATGCCATCTGGATTCCGGCGCGATGACCATGCCGCTCCACCACG
TTCGGCGGCCCTCTCGCGAGGAATGACACCAGCGCTTCGCCAACG
&gt;EYKX4VC02D4GS2 length=60 xy=1573_3972 region=2 run=R_2007_11_07_16_15_57_
ATAAAACTAAATCAGCAAAGACTGGCAAATACTCACAGGTTATAACAATACAATGTAAfa
```

Running this tool while setting *How many characters to keep?* to 14 will produce this:

```
EYKX4VC02EQL05 108
EYKX4VC02D4GS2 60
```

However, if your IDs are not all the same length, you may wish to just keep the fasta ID, and not the description:

```
>EYKX4VC02EQL05 length=108 xy=1826_0455 region=2 run=R_2007_11_07_16_15_57_
```

### History

search datasets x

L04-A

1 shown

24.82 MB

1: Poecilia\_formosa.PoeFor\_5.1.2.pep.all.fa x

# Fasta Format

## Compute Sequence Length

## Tools

search tools

Upload Data

## Barcode Splitter

## Remove sequencing artifacts

FASTQ to FASTA converter from  
FASTX-toolkit

## Filter by quality

## Draw quality score boxplot

## RNA/DNA converter

## FASTA Width formatter

## Length Distribution chart

## Pear Paired-End read merger

## FLASH adjust length of short reads

FASTQ Summary Statistics by  
columnFASTQ de-interlacer on paired end  
readsFASTQ interlacer on paired end  
readsCombine FASTA and QUAL into  
FASTQFasta Statistics display summary  
statistics for a FASTA file

## Filter sequences by length

## Compute sequence length

Cutadapt Remove adapter  
sequences from FASTQ/FASTA

## Trim sequences

## RNA-seq

## CD-HIT

## Datamash

## EMBOSS

## MUMmer4

## Nanopore

## Picard

## Built-in Converters

ng description:trafficking protein particle complex subunit 8-like [Source:NCBI gene;Acc:103133310]	1116
ng description:von Willebrand factor A domain-containing protein 1-like [Source:NCBI gene;Acc:103129824]	451
g description:transmembrane protein 88 [Source:NCBI gene;Acc:103129828]	414
description:transmembrane protein 88 [Source:NCBI gene;Acc:103129828]	493
description:neuroblastoma suppressor of tumorigenicity 1 [Source:NCBI gene;Acc:103129830]	173
g description:MICOS complex subunit Mic10-like [Source:NCBI gene;Acc:103129831]	173
description:forkhead box I3b [Source:ZFIN;Acc:ZDB-GENE-031126-4]	190
g description:capping actin protein of muscle Z-line subunit beta [Source:NCBI gene;Acc:103129832]	75
g description:lymphocyte cytosolic protein 2b [Source:ZFIN;Acc:ZDB-GENE-081107-21]	389
g description:lymphocyte cytosolic protein 2b [Source:ZFIN;Acc:ZDB-GENE-081107-21]	274
g description:dual specificity phosphatase 16 [Source:ZFIN;Acc:ZDB-GENE-040426-2360]	396
description:protein Wnt-8a-like [Source:NCBI gene;Acc:103132073]	454
g description:cAMP responsive element binding protein-like 2 [Source:NCBI gene;Acc:436975]	112
ption:GTPase IMAP family member 8-like [Source:NCBI gene;Acc:103132645]	635
escription:GTPase IMAP family member 8-like [Source:NCBI gene;Acc:103132645]	350
escription:GTPase IMAP family member 8-like [Source:NCBI gene;Acc:103132645]	358
description:wingless-type MMTV integration site family, member 8a [Source:NCBI gene;Acc:30122]	120
g description:cadherin, EGF LAG seven-pass G-type receptor 2 [Source:ZFIN;Acc:ZDB-GENE-070122-2]	450
g description:sterile alpha motif domain containing 12 [Source:HGNC Symbol;Acc:HGNC:31750]	518
scription:protein NLRC3-like [Source:NCBI gene;Acc:103132588]	280
escription:TANK binding kinase 1 [Source:NCBI gene;Acc:103129555]	222
g description:exostosin glycosyltransferase 1b [Source:NCBI gene;Acc:497280]	487
ing description:mediator complex subunit 30 [Source:NCBI gene;Acc:103155381]	778
escription:Na(+)/H(+) exchange regulatory cofactor NHE-RF1-like [Source:NCBI gene;Acc:103152893]	180
scription:Na(+)/H(+) exchange regulatory cofactor NHE-RF1-like [Source:NCBI gene;Acc:103152893]	361
g description:dual serine/threonine and tyrosine protein kinase [Source:NCBI gene;Acc:103129834]	362
ding description:solute carrier family 30 member 8 [Source:ZFIN;Acc:ZDB-GENE-060315-10]	724
g description:ras-related protein Rab-37-like [Source:NCBI gene;Acc:103152894]	373
ing	244
description:LDL receptor related protein 6 [Source:NCBI gene;Acc:103129556]	97
g description:39S ribosomal protein L27, mitochondrial-like [Source:NCBI gene;Acc:103152895]	1608
escription:crossover junction endonuclease EME1-like [Source:NCBI gene;Acc:103152896]	152
scription:golgi-associated, gamma adaptin ear containing, ARF binding protein 1 [Source:ZFIN;Acc:ZDB-GENE-040822-13]	519
description:PAX interacting (with transcription-activation domain) protein 1 [Source:NCBI gene;Acc:449658]	452
escription:PAX interacting (with transcription-activation domain) protein 1 [Source:NCBI gene;Acc:449658]	1055
description:MANSC domain containing 1 [Source:NCBI gene;Acc:103129557]	1047
description:MANSC domain containing 1 [Source:NCBI gene;Acc:103129557]	413
description:MANSC domain containing 1 [Source:NCBI gene;Acc:103129557]	228

History

search datasets

L04-A

2 shown

32.65 MB

2: Compute sequence length on data 1

30,898 lines

format: tabular, database: ?

1: Poecilia\_formosa.Poef or\_5.1.2.pep.all.fa

# Fasta Format

## Convert Fasta to Tabular

Tools

table x

Upload Data

Hide Sections

## Get Data

UCSC Archaea table browser

UCSC Main table browser

Upload File from your computer

## BED

**bedtools FlankBed** create new intervals from the flanks of existing intervals

## Picard

**CollectRnaSeqMetrics** collect metrics about the alignment of RNA to various functional classes of loci in the genome

**AddOrReplaceReadGroups** add or replaces read group information

**FastqToSam** convert Fastq data into unaligned BAM

## SAM/BAM

**Samtools idxstats** reports stats of the BAM index file

## Collection Operations

**Relabel identifiers**

## FASTQ Quality Control

**FastQC** Read Quality reports

## Assembly

**velveth** Prepare a dataset for the Velvet velvetg Assembler

## Convert Formats

**FASTA-to-Tabular** converter

## Nanopore

**Flye** de novo assembler for single molecule sequencing reads

## WORKFLOWS

All workflows

FASTA-to-Tabular converter (Galaxy Version 1.1.1)

Convert these sequences

1: Poecilia\_formosa.PoeFor\_5.1.2.pep.all.fa

How many columns to divide title string into?

1

Typically 2 to take the ID (first word) and description (rest) as two columns, or 1 to give a single column

How many title characters to keep?

0

Applies only to the first column taken from the title string ('0' = keep the whole thing), useful when your sequence identifiers are all the same length.

Job Resource Parameters

Specify job resource parameters

Memory (GB)

7

Maximum Job Memory

Time (hours)

24

Maximum job time

Execute

## What it does

This tool converts FASTA formatted sequences to TAB-delimited format.

Many tools consider the first word of the FASTA ">" title line to be an identifier, and any remaining text to be a free form description. It is therefore useful to split this text into two columns in Galaxy (identifier and any description) by setting **How many columns to divide title string into?** to 2. In some cases the description can be usefully broken up into more columns -- see the examples .

The option *How many characters to keep?* allows to select a specified number of letters from the beginning of each FASTA entry. With the introduction of the **How many columns to divide title string into?** option this setting is of limited use, but does still allow you to truncate the identifier.

## Example

Suppose you have the following FASTA formatted sequences from a Roche (454) FLX sequencing run:

```
>EYKX4VC02EQL05 length=108 xy=1826_0455 region=2 run=R_2007_11_07_16_15_57_
TCCGCGCCGAGCATGCCATCTGGATTCCGGCGCATGACCATGCCGCTCCACCACG
TTCGGCGGCCCTCTCGAGGAATGACACCAGCGCTTCGCCAACG
>EYKX4VC02D4GS2 length=60 xy=1573_3972 region=2 run=R_2007_11_07_16_15_57_
AATAAAACTAAATCAGCAAAGACTGGCAAATCTCACAGGTTATACAATACAATGTAA
```

Running this tool with the default settings will produce this (2 column output):

EYKX4VC02EQL05 length=108 xy=1826_0455 region=2 run=R_2007_11_07_16_15_57_	TCCGCGCCGAGCATGCCATCTGGATTCCGGC...ACG
EYKX4VC02D4GS2 length=60 xy=1573_3972 region=2 run=R_2007_11_07_16_15_57_	AATAAAACTAAATCAGCAAAGACTGGCAAATAC...TAA

Having the full title line (the FASTA ">" line text) as a column is not always ideal.

History

search datasets

L04-A  
2 shown  
32.65 MB

2: Compute sequence length on data 1

1: Poecilia\_formosa.PoeFor\_5.1.2.pep.all.fa

# Fasta Format

## Convert Fasta to Tabular

**HPRC Kaiser Galaxy**

Workflow Visualize Shared Data Admin Help User

Using 0%

Tools

- table x
- [Upload Data](#)
- [Hide Sections](#)

Get Data

- [UCSC Archaea table browser](#)
- [UCSC Main table browser](#)
- [Upload File from your computer](#)

BED

- [bedtools FlankBed](#) create new intervals from the flanks of existing intervals

Picard

- [CollectRnaSeqMetrics](#) collect metrics about the alignment of RNA to various functional classes of loci in the genome
- [AddOrReplaceReadGroups](#) add or replaces read group information
- [FastqToSam](#) convert Fastq data into unaligned BAM

SAM/BAM

- [Samtools idxstats](#) reports stats of the BAM index file

Collection Operations

- [Relabel identifiers](#)

FASTQ Quality Control

- [FastQC](#) Read Quality reports

Assembly

- [velveth](#) Prepare a dataset for the Velvet velvetg Assembler

Convert Formats

- [FASTA-to-Tabular converter](#)

Nanopore

- [Flye](#) de novo assembler for single molecule sequencing reads

WORKFLOWS

- All workflows

SSPDSLQCYEPAEAENQNQERRKPFSCVICKRGFTVKYALNNHI  
MDDQRRLDFSLIPQIILHRIDSLQCNIIHQERRSTLDQEELEPLK  
QVFFLQFPTALLCGEIRKAYVEFCNVSVALCGLRVASTHPDFFT  
MKSFLRCIFVFAMLQQSIRQTAASPDTELNCCEGDLILLDSGG  
MCGVDVLEDGGSADEEKEEEFWVGDAVKMLPPPVAHDGSA  
MCGVDVLEDGGSADEEKEEEFWVGDAVKMLPPPVAHDGSA  
MWQRIQICCALFALHSAAAPAHINRLALFPDKSAWCEAKNITQIV  
MADELGRKWDRCLADTALKATGVGIGIVFSVLFFKRRTWPVSF  
MSSFETQNQSPPRCPGPQFPLSGQEPEPELSMYGECYYPPPSLPS  
MNDQQLDCAALDMRRLLPPQQIEKNLSDLIDLVPSLCEDLLSSVD  
TDGPIKEDVMAWEPKKLADYMRQLKLFGCDNVVLKRGITGAQF  
MSLTDGPIKEDVMAWEPKKLADYMRQLKLFGCDNVVLKRGITG  
MLHGRIDAKPFVIERSIGSARYVPEPVQPHKDPLQLQLVSLQKASG  
ERSAMPRTVRPIGPEALVVLEGGGLDRVVLIDSRRPFDYNTSH  
MAHLIFWILFLHLKMDVGHAWVASNFLMTGPKAYLTYAKSVQVC  
MFECFSKHLYQRAVNPAEATKEFKEIIVKLEKEMDDRRRLDFSI  
MEDNKMVAGVKKPGKGRKPAKIDLKAKLERSRSQSARECRAR  
MVKSEEELDDQRGLLDFSRTPRIILHRIDSLQYNIKQERRSTLDG  
LVGQKCSKGSTTGNNILVKDQNFQNYQDHIETSTCVKHEGRF  
MAAMSNSAQSHPATELRIVLIGGGRNNLRSKSLAGNFIQGNV  
LRIVMIGGSCAGKSTTGNIILGKNFKVHEKTYQDQRETSTCVK  
MDDQRRLDFSRTPRIILHRIDRQQHNIYKEKLCQERRSTLDQE  
MAQLDVLSAMALSTLCLCQLASTWTNNFLMTGPKAFLTYAS  
VRWELRLMLWCYELVSLAQAYTVHISEDAGERTVLAGEQRGAW  
MVCLVESSSVSLKMSVHLKRVYQRAVNPAEETTEFKGIIVKSK  
MAVEAVHCNLSQNGIDHQLRAEDMTSQLEEDELVDAEAVLLD  
NVVVTQSAAKSVQLGQTVTIDCKASRALSTSDCSIGSQVCLSW  
MCEKKDLADILDDLSEKEFNAFKWHLKEEKWNGMEPIKWSKLE  
MQSTTNYLWLISDLLGQGATANVYRGRHKKTGDLYAVKFNNL  
MQAKKRYLILLSAGACALLFYLGGVELPAARRSRQDRSRSGYR  
MTTPPLVPPFVGQPPPPQQQQTQAARDVNTSSLCRIGQETVG  
MSHLRPRLCVLEKSPTGYGFHLHAEKGRTGQFIRLVEPDSPA  
LSNLANKMSHLPRLCVCLEKSPTGYGFHLHAEKGRTGQFIRLVE  
ETKVMVPCPSVQPIEEALEDCIRHVPVILYAINQDSLSEQVAE  
MFKSKDPERRHILAEHSKAYSLPGSVNHEELQKDGTFKHCHDN  
MERMESMEPISACYKTAYPEINPDGYGSGENTENQAKTPPAPS  
MERDSHSEDTLSTMVLENVKNKLIHAFRVTAESRAGQREAGGG  
MGMCVRILLLGSCWLVSGEPLLYANRRDLRLVDAARDKANAT  
MTQFKMAALASMLKSAGLLVSSESYLLDSVRFASKKAGGSS  
MCSDLSQLSEDLRSSSDSDEELPVFDLQPRPSTDVASQKAPKKI  
QVLETCKMKNCGNKFHGEVGKFRFLNELIKVVSPPKKIASQAGEP  
MSEEEIKVDPALFKDVKFYVGDIETKVVQLLKAGKGKEVSYNAL  
MSEEEIKVDPALFKDVKFYVGDIETKVVQLLKAGKGKEVSYNAL  
MTPPTDLRPLPRLFVMLLMSLPAAALEPETCFSRQHQGAAINV  
MTPPTDLRPLPRLFVMLLMSLPAAALEPETCFSRQHQGAAINV

History

search datasets

L04-A

3 shown

57.18 MB

3: FASTA-to-Tabular on data 1

30,898 lines

format: tabular, database: ?

1

ENSPF0P0000000459.1 pep scaffold:PoeFor\_

ENSPF0P0000024938.1 pep scaffold:PoeFor\_

ENSPF0P0000025428.1 pep scaffold:PoeFor\_

ENSPF0P0000000483.2 pep scaffold:PoeFor\_

ENSPF0P0000000486.1 pep scaffold:PoeFor\_

2: Compute sequence length on data 1

1: Poecilia\_formosa.PoeFor\_5.1.2.pep.all.fa

# Fasta Format

## Convert Tabular to Fasta

### Tools

tabular x

Upload Data

Hide Sections

query sequence(s)

**NCBI BLAST+ tblastx** Search  
translated nucleotide database with  
translated nucleotide query  
sequence(s)

**NCBI BLAST+ rpstblastn** Search  
protein domain database (PSSMs)  
with translated nucleotide query  
sequence(s)

**NCBI BLAST+ rpsblast** Search  
protein domain database (PSSMs)  
with protein query sequence(s)

### Convert Formats

**Tabular-to-Fasta** converts tabular  
file to FASTA format

**FASTA-to-Tabular** converter

### Built-in Converters

**csv-to-tabular** converter

**dbnsfp.tabular-to-snpsiftdbnsfp**  
converter

**fasta-to-tabular** converter

### Datamash

**Transpose** rows/columns in a  
tabular file

**Reverse** columns in a tabular file

**Datamash** (operations on tabular  
data)

### Text Manipulation

**Arithmetic Operations** on tables

**Change Case** of selected columns

**Change Case** of selected columns

### Collection Operations

**Apply rules**

### WORKFLOWS

All workflows

**Tabular-to-Fasta** converts tabular file to FASTA format (Galaxy Version 1.1.1)

**Tab-delimited file**

3: FASTA-to-Tabular on data 1

**Title column(s)**

Select/Unselect all

Column: 1

Multi-select list - hold the appropriate key while clicking to select multiple columns

**Sequence column**

Column: 2

**Job Resource Parameters**

Specify job resource parameters

**Memory (GB)**

7

Maximum Job Memory

**Time (hours)**

24

Maximum job time

Execute

**What it does**

Converts tab delimited data into FASTA formatted sequences.

**Example**

Suppose this is a sequence file produced by Illumina (Solexa) sequencer:

```
5 300 902 419 GACTCATGATTCTTACCTATTAGTGGTGAACATC
5 300 880 431 GTGATATGTATGTTGACGCCATAAGGCTGCTTCTT
```

Selecting c3 and c4 as the **Title column(s)** and c5 as the **Sequence column** will result in:

```
>902_419
GACTCATGATTCTTACCTATTAGTGGTGAACATC
>880_431
GTGATATGTATGTTGACGCCATAAGGCTGCTTCTT
```

**Requirements:** ?

- python (Version 3.7)

### History

search datasets

L04-A

3 shown

57.18 MB

3: FASTA-to-Tabular on d  
ata 1

2: Compute sequence len  
gth on data 1

1: Poecilia\_formosa.Poef  
or\_5.1.2.pep.all.fa

# Fasta Format

## Convert Tabular to Fasta

# Fasta Format

## Correct Fasta Width

HPRC Kaiser Galaxy      Using 0%

Tools  
width  
Upload Data  
Hide Sections

FASTA/FASTQ  
FASTA Width formatter  
Length Distribution chart

Collection Operations  
Filter empty datasets  
Unzip collection  
Filter failed datasets  
Flatten collection  
Zip collections  
Build list

BED  
**bedtools getfasta** use intervals to extract sequences from a FASTA file

Picard  
CollectInsertSizeMetrics plots distribution of insert sizes

WORKFLOWS  
All workflows

**FASTA Width formatter (Galaxy Version 1.0.1)**

**Input FASTA file**  
4: Tabular-to-FASTA on data 3

**New width for nucleotides strings**  
0  
Use 0 for single line out.

**Job Resource Parameters**  
Specify job resource parameters

**Memory (GB)**  
7  
Maximum Job Memory

**Time (hours)**  
24  
Maximum job time

**Execute**

**What it does**  
This tool re-formats a FASTA file, changing the width of the nucleotides lines.  
**TIP:** Outputting a single line (with **width = 0**) can be useful for scripting (with **grep**, **awk**, and **perl**). Every odd line is a sequence identifier, and every even line is a nucleotides line.

**Example**  
Input FASTA file (each nucleotides line is 50 characters long):  

```
>Scaffold3648
AGGAATGATGACTACAATGATCAACTAACCTATCTATTAAATTAGTTC
CTTAATGTCAGGGACCTACCTGTTTGTATGTTGGGTTTGTGTTG
TTGTTTTTTAATCTGAAGGTATTGTGCATTATATGACCTGTAATACACA
ATTAAGTCATTTAATGAACATGTAGAAAAACT
>Scaffold9299
CAGCATCTACATAATATGATCGCTATTAACCTAAATCTCCTGACGGAG
TCTTCGGTCATAACACAAACCCAGACCTACGTATATGACAAAGCTAATAG
aactggctttacctTTAAGTTG
```

  
Output FASTA file (with width=80):  

```
>Scaffold3648
AGGAATGATGACTACAATGATCAACTAACCTATCTATTAAATTAGTCCCTAATGTCAAGGGACCTACCTGTTTGT
ATGTTGGGTTTGTGTTGTTTTAATCTGAAGGTATTGTGCATTATATGACCTGTAATACACAATTAAAGTC
ATTTAATGAACATGTAGAAAAACT
>Scaffold9299
CAGCATCTACATAATATGATCGCTATTAACCTAAATCTCCTGACGGAGCTTCGGTCATAACACAAACCCAGACCTAC
GTATATGACAAAGCTAATAGaactggctttacctTTAAGTTG
```

History  
search datasets  
L04-A  
4 shown  
81.73 MB  
4: Tabular-to-FASTA on data 3  
3: FASTA-to-Tabular on data 1  
2: Compute sequence length on data 1  
1: Poecilia\_formosa.PoeFor\_5.1.2.pep.all.fa

# Fasta Format

## Correct Fasta Width

HPRC Kaiser Galaxy      Using 0%

Tools  
width  
Upload Data  
Hide Sections

FASTA/FASTQ  
FASTA Width formatter  
Length Distribution chart

Collection Operations  
Filter empty datasets  
Unzip collection  
Filter failed datasets  
Flatten collection  
Zip collections  
Build list

BED  
**bedtools getfasta** use intervals to extract sequences from a FASTA file

Picard  
CollectInsertSizeMetrics plots distribution of insert sizes

WORKFLOWS  
All workflows

**FASTA Width formatter (Galaxy Version 1.0.1)**

**Input FASTA file**  
4: Tabular-to-FASTA on data 3

**New width for nucleotides strings**  
100  
Use 0 for single line out.

**Job Resource Parameters**  
Specify job resource parameters

**Memory (GB)**  
7  
Maximum Job Memory

**Time (hours)**  
24  
Maximum job time

**Execute**

**What it does**  
This tool re-formats a FASTA file, changing the width of the nucleotides lines.  
**TIP:** Outputting a single line (with **width = 0**) can be useful for scripting (with **grep**, **awk**, and **perl**). Every odd line is a sequence identifier, and every even line is a nucleotides line.

**Example**  
Input FASTA file (each nucleotides line is 50 characters long):  

```
>Scaffold3648
AGGAATGATGACTACAATGATCAACTAACCTATCTATTAAATTAGTTC
CTTAATGTCAGGGACCTACCTGTTTGTATGTTGGGTTTGTGTTG
TTGTTTTTTAATCTGAAGGTATTGTGCATTATATGACCTGTAATACACA
ATTAAGTCATTTAATGAACATGTAGAAAAACT
>Scaffold9299
CAGCATCTACATAATATGATCGCTATTAACCTAAATCTCCTGACGGAG
TCTTCGGTCATAACACAAACCCAGACCTACGTATATGACAAAGCTAATAG
aactggctttacctTTAAGTTG
```

  
Output FASTA file (with width=80):  

```
>Scaffold3648
AGGAATGATGACTACAATGATCAACTAACCTATCTATTAAATTAGTCCCTAATGTCAAGGGACCTACCTGTTTGT
ATGTTGGGTTTGTGTTGTTTTAATCTGAAGGTATTGTGCATTATATGACCTGTAATACACAATTAAAGTC
ATTTAATGAACATGTAGAAAAACT
>Scaffold9299
CAGCATCTACATAATATGATCGCTATTAACCTAAATCTCCTGACGGAGCTTCGGTCATAACACAAACCCAGACCTAC
GTATATGACAAAGCTAATAGaactggctttacctTTAAGTTG
```

History  
search datasets  
L04-A  
4 shown  
81.73 MB  
4: Tabular-to-FASTA on data 3  
3: FASTA-to-Tabular on data 1  
2: Compute sequence length on data 1  
1: Poecilia\_formosa.PoeFor\_5.1.2.pep.all.fa

# Fasta Format

## Correct Fasta Width

HPRC Kaiser Galaxy

Workflow Visualize Shared Data Admin Help User

Using 0%

Tools width

Upload Data Hide Sections

FASTA/FASTQ

FASTA Width formatter Length Distribution chart

Collection Operations

Filter empty datasets Unzip collection Filter failed datasets Flatten collection Zip collections Build list

BED

bedtools getfasta use intervals to extract sequences from a FASTA file

Picard

CollectInsertSizeMetrics plots distribution of insert sizes

WORKFLOWS

All workflows

This dataset is large and only the first megabyte is shown below.

Show all | Save

```
>ENSPF0P0000000459.1 pep scaffold:PoeFor_5.1.2:KI520192.1:5697:27330:1 gene:ENSPFOG0000000511.2 transcript:ENSPF0T0000000460.1 gene_biotype:protein_coding
SSPDSLQCYEPAEAENQNQERRKPFSCVICKRGFTVKYALNNHIRTHTGEKPFCVNCGKRSQRKLQHMHIVHSSDKPFSCVNCGGKFIHKRNLTTHM
RIHSGEKPFCVNCGKRSQRNLQHMIHSSEKPFSCVNCGKSFHKRNLTQHMRINTGEKPFCVNCGKTFRQKRNLQHMIHSEKPFSCVNCGK
CFCQKQHLTRHLMINTGEKPFCVNCGKSFQHKQHLLQQMIRHSGEKPFFSCGNCGKFSQHKVLTQHMMIHTGEKPFCVNCGKSFQKQILTQHMT
IHTGEKPFCGNGKSFQKVLSTQHMTIHTGEKPFCGNGKSFQKVLSTQHMTIHTGEKPFCGNGKSFQKLNLTQHMKIHTGEKPFCVNCGK
FSRKYVLTQHMSHTGEKPFCGNGKGSQKLNLTQHMKIHTGEKPFCGNGKSFQKLNLTQHMMIHTGEKPFCGNGKGSQKLNLTQHMMIHTGEKPFC
EKPFCGNGKSFQKHALTQHMRINTGEKPFCGNGKSFQKLNLTQHMRINTGEKPFCGNGKSFQKLNLTQHMMIHTGEKPFCGNGKSFQ
KQHVTQHMTIHTGEKPFCGNGKSFQKHLTQHVIINHGEKPFCGNGKSFQKSLTRHMMIHTGEKPFCGNGKSFQKLNLTQHMIHTEKEP
FSCVNCGKSFQSHKQNFQHMRISDEKPFSCVNCGKSFQKRSRKYALNIHIRTHTGEKPFCGNGKSFQGKKHLTRHMMIHTGEKPFCVNCGKSFSSKHIL
TQHMMIHTGEKPFCVNCGKSFQSHKRILTQHMRINTGEKPFCVNCGKSFQKVSQVTRHMMIHTGEKPFCGNGCRGSFSYQHFTQHMRINTGEKPFCG
NCGKSFQSHKCDLNVMRIHRGKKPFSCVNCGKSFQKRLTQHVIHTGEKPFCGNGKRSQPKNLTRHMMIHTGEKPFCVNCGKSFQKQHITQH
KIHTCVKPFCVNCGKRSQPKLSSQHMRINTGEKPFCVNCGKRSQPKNLTRHMMIHTGEKPFCVNCGKSFQPKNLTRHMMIHTGEKPFCVNCGK
SFQPKNLTRHMMIHTGEKPFCVNCGKSFQPKNLTRHMMIHTGEKPFCVNCGKSFQPKNLTRHMMIHTGEKPFCVNCGK
>ENSPF0P00000024938.1 pep scaffold:PoeFor_5.1.2:KI520192.1:17993:27836:1 gene:ENSPFOG0000000511.2 transcript:ENSPF0T00000024762.1 gene_biotype:protein_coding
MDDQRRLDFSLIPQIILHRIDSLQCNIIHQERRSTLDQEELEPLKGKQDQEETEDHQIKEEKEDLKHHQIKEEQDLKHQHQIKEEQDLKHQHQIKEEQE
DLKHQHQIKEETEVYCSQGEWEIELKQETDTCMVFPDAEQTYQTESEQSGNKVIFQEPAEAEQNQERRKPFSCVICKKAFGSKYALNNHIRTHTGEKP
SCGNCECFSHQQLLTQHMRITHRGKKPFSCVNCGKSFQKRLTQHVIHTGEKPFCGNGKRSQPKNLTRHMMIHTGEKPFCVNCGKSFQKQHIT
QHMKIHTCVKPFCVNCGKSFQSHKQYLTQHMMIHTGEKFLFCVNCGKRSQPKLSSQHMRINTGEKPFCVNCGKRSQPKNLTRHMMIHTGEKPFCV
CGKSFQPKNLTRHMMIHTGEKPFCVNCGKSFQPKNLTRHMMIHTGEKPFCVNCGKSFQPKNLTRHMMIHTGEKPFCVNCGK
>ENSPF0P00000025428.1 pep scaffold:PoeFor_5.1.2:AYCK01029644.1:207:5683:-1 gene:ENSPFOG00000024132.1 transcript:ENSPF0T00000028006.1 gene_biotype:protein_coding
QVFFLQPTALLGEIRKAYVEFCNVSAVALCGLRASTHPDFTFGSQTAPTPPSDSAENCASAYTFPSPLQPGSMSSETLVASAPACSQPSRVTEI
PIHGGTLQPGQSVRLPLWLRGPQEGVHEINFLFYENTEKGLKISHRVLRHTFICASRSLSVQASACRSSASPHISLDDKGSGGTLVFIDVENVASE
SGVREFHVQVSSSSQHWRHLKYINPCNYDCLRSSRERAKLCFKAMRCKPHEGSDGVKEYTFADLNGLNERIMSSSTPCGDFFFSCRQSSVLEQVNAP
RNSRSQGSHSDNTASDITNIAKKCNELDLNIIVIWKAYVVEDNKQLILEGQLHVALKTFGEASSLIPKEETQEMMLLRFKSELPPPVILPYPELSQI
RTNLCPETFTHF
>ENSPF0P0000000483.2 pep scaffold:PoeFor_5.1.2:KI520192.1:38500:47984:-1 gene:ENSPFOG0000000529.2 transcript:ENSPF0T0000000484.2 gene_biotype:protein_coding
MKSFLRCIFVAMILQSQIRTAASPDTTELNCCEGDIILLLDSSGVTSEFSSFLFAASLLYPFTLGRGHVRALLVGTTPHLEFLDVHSNQESLL
RALQSVNQQGDTNTKAAIEVAQRLLSEADGDPVKLLWLTDGVQTVDEKAMSELKARGVVYLIVYTQHNRVLRQVATPPLESHLYSVDMESIDIIT
DDLREAIKIIRERAERLSQLVRLTSQAVLQWRPVLAQDSGHYELSYKAVDDQHSGKTTVLPSSSSQTELQLPQDTTTASLRPESNQLRHSTLQVFTTL
PDVLSPAEVTLSDPGPRQVRVSWGPLQPDFRVRQRTLEYGAIPSGRVHTVDVSNSQVFLRNLEPGTRYLITVSALHRDGKERAMSVRACTQEEARPALT
DLRLTLTDRQEGNEVQASWETNTEGLKGYWVSWERKGQSNSQSERSPSTAYLPPSPSSVRLAHAPNSRVCVSPVYSSGRGDGICCTAKTNSR
>ENSPF0P0000000486.1 pep scaffold:PoeFor_5.1.2:KI520192.1:72966:83042:-1 gene:ENSPFOG0000000548.2 transcript:ENSPF0T0000000487.2 gene_biotype:protein_coding
MCGVVDLEDGSADEEEFWVGDAVKMLPPPVAHDGSAWGRRGACGSVACGAVLVLWNLCVVSAGALLLALVFCVLLPAALLLYAGFLCHSRVL
DAPAAICRYLDDNSCSALIILGFVMMSPLVLAACVFCGLLRKFRLLLFLQPLSRARYRGRLLDWVGRVQAWV
>ENSPF0P00000029118.1 pep scaffold:PoeFor_5.1.2:KI520192.1:72966:81842:-1 gene:ENSPFOG0000000548.2 transcript:ENSPF0T00000024091.1 gene_biotype:protein_coding
MCGVVDLEDGSADEEEFWVGDAVKMLPPPVAHDGSAWGRRGACGSVACGAVLVLWNLCVVSAGALLLALVFCVLLPAALLLYAGFLCHSRVL
DAPAAICRYLDDNSCSALIILGFVMMSPLVLAACVFCGLLRKFRLLLFLQPLSRARYRGRLLDWVGRVQAWV
>ENSPF0P0000000510.2 pep scaffold:PoeFor_5.1.2:KI520192.1:87371:97453:-1 gene:ENSPFOG0000000555.2 transcript:ENSPF0T0000000511.2 gene_biotype:protein_coding
MWQRQICCALFALHSAAAPAHINRRLAFPDKSAWCEAKNITQIVGHTGCQPRSQIQRACLGQCFSYSPNTFPQSTESLAHCDSCPQATQWEVVTLEC
PGSEDSPRDKLVERILHCSCQCSKESVQEGAVMQLYPADNTLDSAPSLSDPPLIGDPAHSLPHADSLSQAHPHAHHTPLPHASDGG
>ENSPF0P00000022570.1 pep scaffold:PoeFor_5.1.2:KI520192.1:103580:109381:-1 gene:ENSPFOG00000022757.1 transcript:ENSPF0T00000027073.1 gene_biotype:protein_coding
MADELGRKWDRCLADTALKTATGVGIGIVFSVLFKRTWPVSGSGLGMGYSNCQHDFRSPYLIHGRVVKDQ
>ENSPF0P00000015794.2 pep scaffold:PoeFor_5.1.2:KI520523.1:19077:21348:-1 gene:ENSPFOG00000015741.2 transcript:ENSPF0T0000015816.2 gene_biotype:protein_coding
MSSFETQNQSPRCGPQFPLQGQEPPELSMYGECYYPPPSLSPQRTAPSSYELSDYTSSPNYLWFNNSGINSSSYLSNNGPPGPPFVPQHYGMQ
RPLYLGAGPGPQGGELSWFSLPSQEDLMKLRPQPSYNSALIAMAIHGDKRQLTSQIYQYVADNFPFYNSKAGWQNSIRHNLSLNDCFKVPRDEDDP
GKGNYWTLDNCEKMDNGNFRRKRRKRSDSLNSGDGASGAPESGDSERGSPKHPALMSPTPDRIPSLSGPTPCSSLSEMSTVSTGAANEVGGDG
LARPLQINLPLDGPHRPTQPGNFSNSYSPNSGVSEWPQVPAPSVLTSSPSHSSLGYSSPILSQYTGSQHFPYPSLSSTGIYHRDGTEV
>ENSPF0P0000000622.2 pep scaffold:PoeFor_5.1.2:KI520192.1:111691:128820:1 gene:ENSPFOG0000000600.2 transcript:ENSPF0T0000000623.2 gene_biotype:protein_coding
MNDQQLDCALDLMRRLPQQIEKNLSDLIDLVLPSLCEDLSSVQDPLKIAIRDVKVGDYLLCDYNRDGSYRSPWSNKYEPPIEDGAMPSARLRKLEVEA
NNAFDQYRDLYFEFGVSSVYLWDLDHGFAGVILIKKAGDGSKKIKGCGWDSIHSVVEQKSSGRTAHYLTSTVMLWLQTTKTGSGTMNLGGSLTRQMEKD
ETVGESSHIANIGRLVEDMENKIRSTLNEIYFGKTKDIVNGLRSIESLPDNQKYRQLQRELSQLTQRQIFID
>ENSPF0P00000024666.1 pep scaffold:PoeFor_5.1.2:KI520523.1:25902:30531:-1 gene:ENSPFOG00000022030.1 transcript:ENSPF0T0000028540.1 gene_biotype:protein_coding
TDGPIKEDVMWEPKKKLADYMRQLKLFCDNVVLKRGITGAQFLHMIQRLDQVFPTVPLAKIQRDVNQKEKLREIILQKVTKHPLNFANKINQPY
FLQFDSNDSESDEYEFNEEAQDNYIVPLTEPLTADNLHKAKAREDHLRHPVFSLPPDQVRVSKVHSLSPTGNEKLPCSAHSSNFQRAANKKPEKPPVQRA
FTRLGANNITPHVVPVTSVHLSNQPTRTGHVNKQQVQLVTQKSAFSESKQQMAMLYNTTAKTFSESDTELKNTFVLTQGLDPSWYKGRITRNQAEVSL
WEVNDGAFLVRDSSHGSTQHPYTLMLNQKGKIFNIKIQRQNDFYFLGNGTNNTKSFLRVKEMIHHMSTPLLKDATDQCSEEQPCQCCLLHPAGL
>ENSPF0P00000026960.1 pep scaffold:PoeFor_5.1.2:KI520523.1:25908:30540:-1 gene:ENSPFOG00000022030.1 transcript:ENSPF0T0000029530.1 gene_biotype:protein_coding
```

History search datasets

L04-A 5 shown 106.44 MB

5: FASTA Width on data 4 30,898 sequences format: fasta, database: ? display with IGV local

4: Tabular-to-Fasta on data 1

3: FASTA-to-Tabular on data 1

2: Compute sequence length on data 1

1: Poecilia\_formosa.PoeFor\_5.1.2.pep.all.fa