

Module 1: Artemis

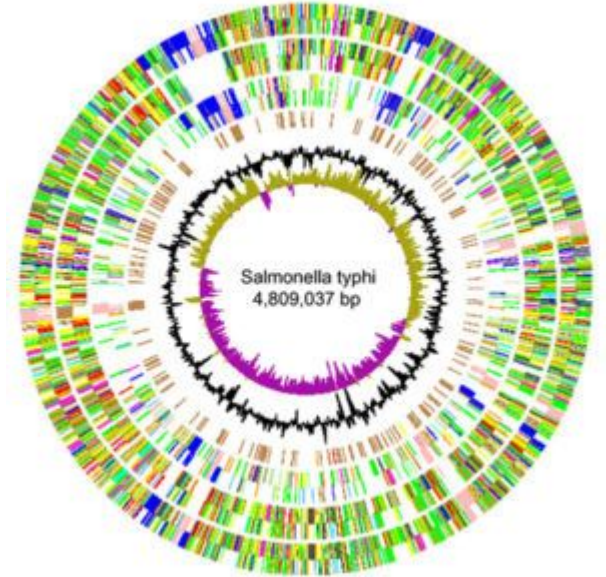
Lionel Uran Landaburu

Instituto de Investigaciones Biotecnológicas
lionel.u.l@iib.unsam.edu.ar

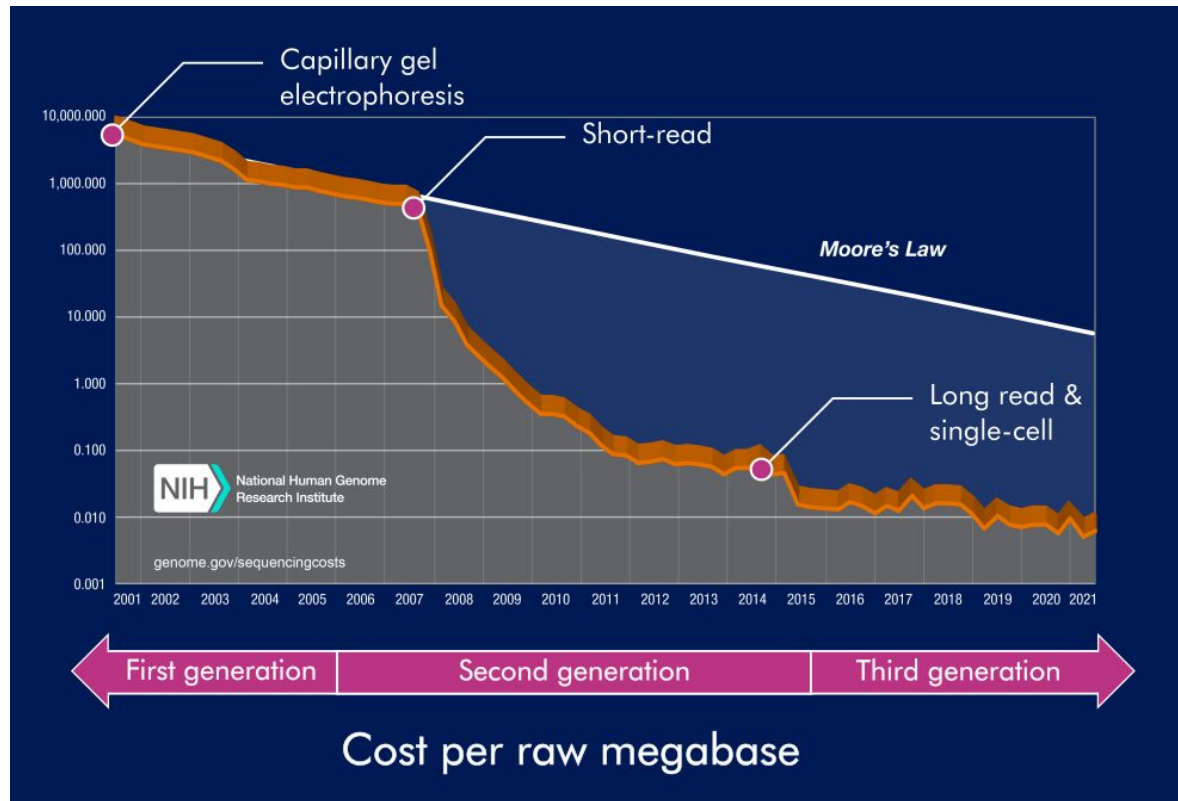
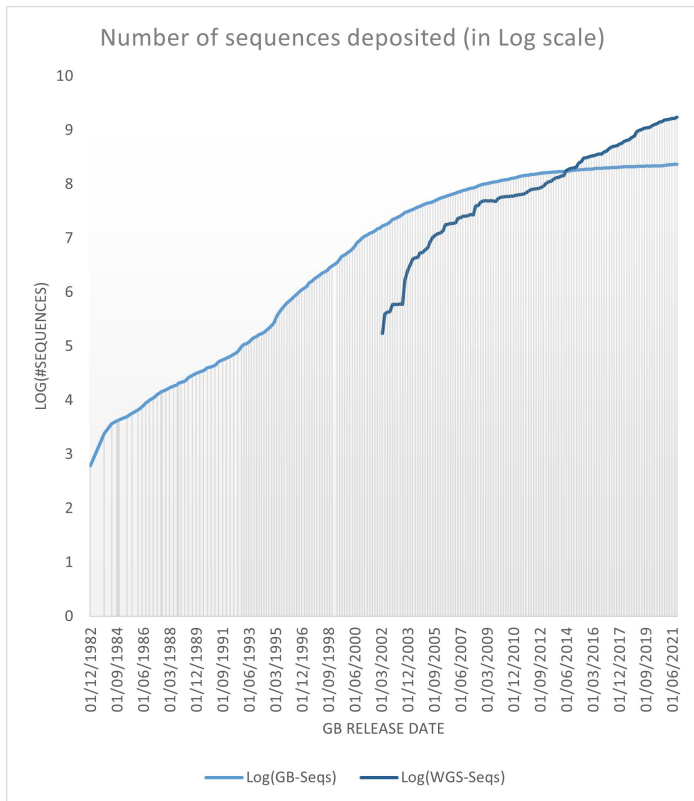
Rodrigo Quiroga

Instituto De Investigaciones En Fisico-Quimica De Cordoba
rodrigoquiroga77.yahoo@gmail.com

Working with Pathogen Genomes 7-11th, February 2022.



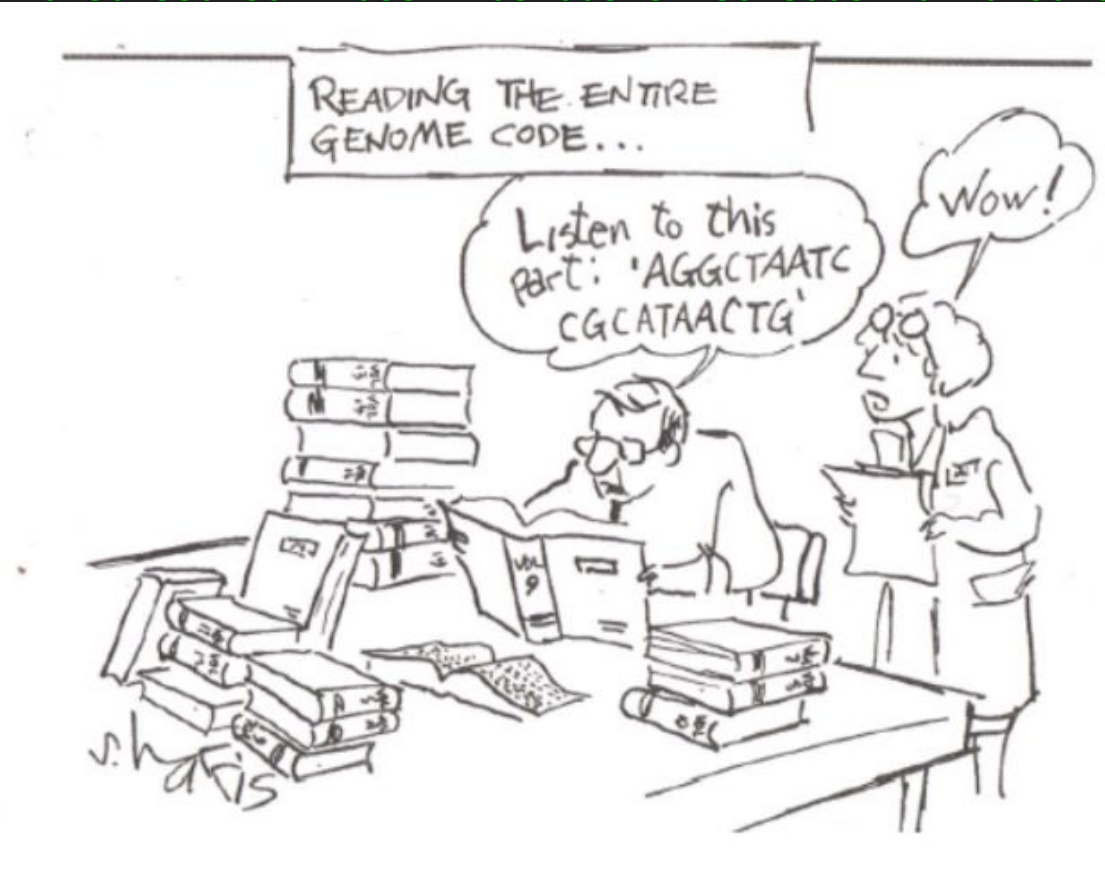
Next generation sequencing, a paradigm shift



Source: GenBank statistics
 (<https://www.ncbi.nlm.nih.gov/genbank/statistics/>)

Adapted from genome.gov/sequencingcosts & Athina Gkazi (2021). An Overview of Next-Generation Sequencing. Technology Networks: Genomics Research.

We have sequences... now what?



ATTAAGGTTTATACCTTCCAGGTAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTCTCTAAACGAACTTTAAAA
TCTGTGTGGCTGTCACCTCGGCTGCACTTTCGCACTTCCGCGTAAATTTTAAATTACTGTCGTTGACAGGACAC
GAGTAACTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGCTCCGTGTTGCAGCCGATCATCAGCACATCTAGGTTCGTCCG
GGTGTGACCGAAAGGTA
CAGGTTTCGCGACGTGC
GCACTTGTGGCTTAGTA
TCGAACTGCACCTCAT
CTTGGTGCCTTGTCC
CTGGTGGCCATAGTTA
AGAAACTGGAACACTA
GTCGATAACAACCTTCT
GCACTTTGTCCGAACA
GTACACGGAACGTTCT
GGGAATGTCCAATT
GCTTTATGGGTAGAAT
GTGTGATCATTGTGGT
ACTAAAGAAGGTGCCA
AAGTAGGACCTGAGCA
TGCCTTTGGAGGCTGT
GGTTGTAACCATACAG
TCAACATCAATATTGT
TTTTGTGGAACTGTG
GGAAAAGCTAAAAAAG

CAGTTTGCCTGTTTTA
CAACATCTTAAAGATG
TCAAACGTTCCGATGC
TCGTAGTGGTGAGACA
AACGGTAATAAAGGAG
CTTATGAAGATTTTCA
GGCATACTCGCTAT
GCTGGTAAAGCTTCAT
AGCATGAAATTGCTTG
ATTTGACACCTTCAAT
AAGAAAAGCTTGATG
TTTCAACTCTCATGAA
TGGCACTGAGAATTTG
GCATGTCACAATTCAG
AGGGTGGTTCGACTAT
TGCTAGCGCTAACATA
CTCCAAAAGAGAAAG
CTGCTTCCACAAGTGC
TTTTAAAGTTACAAA
TTTGCATCAGAGGCTG

CTCGTGTGTGACGATCAATTTTCTCCGGCACTCTTGAACCTGCTCAAATTTCTGTGCGTGTTTTACAGAAGGCCGCTATAAC

See anything yet?

Entry: sequence.fasta

Nothing selected



I K G L Y L P R # Q T N Q L S I S C R S V L # T N F K I C V A V T R L H A + C T H A V # L I T N Y C R * Q D T S N S S I
L K V Y T F P G N K P T N F R S L V D L F S K R T L K S V W L S L G C M L S A L T Q Y N # # L I T V V D R T R V T R L S
. # R F I P S Q V T N Q P T F D L L + I C S L N E L # N L C G C H S A A C L V H S R S I I N N # L L S L T G H E # L V Y L
ATTAAAGGTTTATAOCTTCCCAGGTAACAAACCAACCACTTTTCGATCTCTTGTAGATCTGTTCTCTAAACGAACCTTAAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACCTCACGCAGTATAAATTAACTAACTACTGTGCGTTGACAGGACACGAGTAACTCGTCTATCT
TAATTTCCAAATATGGAAGGGTCCATTGTTTGGTTGGTTGAAAGCTAGAGAACATCTAGACAAGAGATTTGCTTGAAATTTAGACACACCGACAGTGAGCCGACGTACGAATCACGTGAGTGGGTGATTAATTATTGATTAAATGACAGCAACTGTCCCTGTGCTCATTGAGCAGATAGA
N F T # V K G P L L G V L K R D R T S R N E L R V K F D T H S D S P Q M S L A S V C Y L # Y S I V T T S L V R T V R R D
L N I G E W T V F W G V K S R K Y I Q E R F S S # F R H P Q * E A A H K T C E R L I I L L + N S D N V P C S Y S T + R
. L P K Y R G L Y C V L W S E I E Q L P D T R + V F K L I Q T A T V R S C A # H V * A T Y N I V L # Q R Q C S V L L E D I H

How about now? Does it ring a bell?

Entry: sequence.gb

Selected feature: bases 3822 amino acids 1273 S (/gene="S" /locus tag="GU280 gp02" /gene synonym="spike glycoprotein" /codon start=1 /product="sur")

ORF1ab S ORF E ORF6'a ORF10 ORF7b ORF7a

5'UTR ib ab ORF1ab ORF1 ORF ORF ORF ORF1ab ORF1ab ORF1ab ORF1 ORF1ab ORF E 4 OF ORF7b ORF10 loop

2200 4400 6600 8800 11000 13200 15400 17600 19800 22000 24200 26400 28600

```
I K G L Y L P R # Q T N Q L S I S C R S V L # T N F K I C V A V T R L H A + C T H A V # L I T N Y C R * Q D T S N S S I F C R L L
L K V Y T F P G N K P T N F R S L V D L F S K R T L K S V W L S L G C M L S A L T Q Y N # # L I T V V D R T R V T R L S S A G C
. # R F I P S Q V T N Q P T F D L L + I C S L N E L # N L C G C H S A A C L V H S R S I I N N # L L S L T G H E # L V Y L L Q A A
ATTAAAGGTTTATACCTTCCCAGGTAAACAAACCAACCAACTTTCGATCTCTGTAGATCTGTCTCTAAACGAACCTTAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTAAATTAATAACTAATTACTGTCGTTGACAGGACACGAGTAACCTCGTCTATCTTCTGACGGCTGCT
20 40 60 80 100 120 140 160 180
TAATTTCCAAATATGGAAGGGTCCATTGTTTGGTTGGTTGAAAGCTAGAGAACATCTAGACAAGAGATTGCTTGAATTTAGACACACCGACAGTGAGCGAGCTACGAATCACTGAGTGAGTGGTGCATATTAAATTTGATTAAATGACAGCAACTGTCCTGTGCTCATTGAGCAGATAGAAGACGTCGACGA
NFT#VKGPLLGV LKRDRTSRNELRVKFDTHSDSPQMSLASV CYL#YSIVTTS LVRRTVRRDEAPQK
#LNI GEWT VFWGVKSRKYIQERFSS#FRHPQ*EAAAHKTCERLLIILL+NSDNVPCSYST+RRC AA
LPKYRGLYCVLWSEIEQLDTR+VFKLIQQTATVRS CA#H V*ATYNIIVL#QRQCSVLLLEDIKQLSS
```

CDS	21563	25384	structural protein; spike protein
gene	25393	26220	
CDS	25393	26220	
gene	26245	26472	
CDS	26245	26472	ORF4; structural protein; E protein

Genome browser and annotation tool



Artemis

- Visualization of sequence
 - DNA
 - six frame translation
 - Panoramic and sequence view
- Annotation
 - Features
 - Mapped and listed
 - Editable
 - In layers (entry)
- Perform and view analysis
 - basic analysis
 - Basic stats & index can be plotted
 - import and view the results of other searches/analysis
 - Different lines of evidence can be seen together

The screenshot shows the Artemis genome browser interface. The title bar reads "Artemis Entry Edit: S_typhi.dna". The menu bar includes "File", "Entries", "Select", "View", "Goto", "Edit", "Create", "Write", "Run", "Graph", "Display", and "(Noddy)". The main window displays a DNA sequence with several features highlighted in different colors (yellow, green, orange, red). The sequence is shown in a panoramic view with a scale bar at the bottom. The legend at the bottom of the window lists various features and their corresponding colors:

CDS	190	255	Orthologue of E. coli thrL (LPT_ECOLI); Pasta hit to LPT_ECOLI
CDS	337	2799	Orthologue of E. coli thrA (AKIH_ECOLI); Pasta hit to AKIH_ECOLI
misc_feature	343	369	PS00324 Aspartokinase signature
misc_feature	2314	2382	PS01042 Homoserine dehydrogenase signature
CDS	2801	3730	Orthologue of E. coli thrB (KHSE_ECOLI); Pasta hit to KHSE_ECOLI
misc_feature	3066	3103	PS00627 GHMP kinases putative ATP-binding domain
CDS	3734	5020	Orthologue of E. coli thrC (THRC_ECOLI); Pasta hit to THRC_ECOLI
misc_feature	4022	4066	PS00165 Serine/threonine dehydratases pyridoxal-phosphate at
CDS	5114	5887	Orthologue of E. coli yaaA (YAAA_ECOLI); Pasta hit to YAAA_ECOLI
CDS	5966	7396	c Similar to Bacillus subtilis amino acid carrier protein alst
misc_feature	7091	7138	c PS00873 Sodium:alanine symporter family signature
CDS	7665	8618	Pasta hit to TALA_ECOLI (316 aa), 65% identity in 311 aa over
misc_feature	7755	7781	PS01054 Transaldolase signature 1

Artemis



Overview



Download



Learn



License



Contact



Related



Publications

Overview

Artemis is a free genome browser and annotation tool that allows visualisation of sequence features, next generation data and the results of analyses within the context of the sequence and also its six-frame translation.

Download and Installation

Please see our [GitHub page](#) for download and installation instructions.

Learn and Support

For information and advice on using this software please see our [GitHub page](#).

In addition, an email discussion list called [artemis-users](#) is available and posts to the list since September 2001 are archived at [mail-archive.com](#)

Chado

Artemis and ACT can be used to connect to [Chado](#) databases. They are being developed to read from and write to the database and perform the same functions as the standard Artemis and ACT.

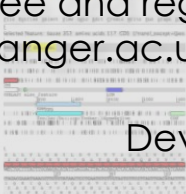
- [Chado overview](#)
- [Chado admin](#)
- [Chado practical guide](#)
- [Data storage](#)

Genome browser and annotation tool (Rutherford et al, 2000)

Tool Type

- [Annotation](#)

Screenshots



Free and regularly updated software
 • (<https://www.sanger.ac.uk/science/tools/artemis>)

Developed in JAVA in 1998

Available for UNIX, macOS and Windows

Related Tools

- [Alien Interpreter](#)
- [Annotation Tool \(ACT\)](#)
- [BamView](#)
- [DNAPlotter](#)
- [GeneDB](#)

Developed for annotation and analysis of prokaryotic and small eukaryotic genomes

Can be used simply as a sequence viewer allowing the visualization of sequence and annotation taken directly from EMBL or GeneBank.

Files in Artemis



Sequence	Annotation
.fasta	.tab
.seq	
.dna	
	.embl

It can read several file formats (FASTA, EMBL, GENE BANK, GFF).

EMBL

Two-character line code indicates the type of information contained in the line

```

ID ECRSMA      standard; DNA; PRO; 500 BP.
XX
AC L40173;
XX
XX L40173.1
SV
XX
DT 10-AUG-1995 (Rel. 44, Created)
DT 04-MAR-2000 (Rel. 63, Last updated, Version 4)
XX
DE Erwinia carotovora repressor (rsmA) gene, complete cds.
XX
KW repressor; rsmA gene.
XX
OS Pectobacterium carotovorum
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriaceae;
OC Pectobacterium.
XX
RN [1]
RP 1-500
RA Cui Y., Chatterjee A., Liu Y., Dumenyo C.K., Chatterjee A.K.;
RT "Identification of a global repressor gene, rsmA, of Erwinia carotovora
subsp. carotovora that controls extracellular enzymes,
N-(3-oxohexanoyl)-L-homoserine lactone, and pathogenicity in soft-rotting
RT Erwinia spp";
RL J. Bacteriol. 177(17):0-0(1995).
XX
DB GOA; Q47620; Q47620.
DE SWISS-PROT; Q47620; CSRA_ERWCA.
XX

```

Feature Key

```

FH Key          Location/Qualifiers
FH
FT source       1..500
FT              /db_xref="taxon:554"
FT              /organism="Pectobacterium carotovorum"
FT              /strain="71"
FT              /sub_species="carotovora"
FT -10_signal   107..112
FT              /gene="rsmA"
FT RBS          235..239
FT              /gene="rsmA"
FT CDS          246..431
FT              /codon_start=1
FT              EIQYRIQAEKSQPTSY"
XX

```

Qualifier

```

SQ Sequence 500 BP; 140 A; 101 C; 120 G; 139 T; 0 other;
ggatccggca agcaggatag aaagtgtgtt accttcagat attctgaagc ttatcatgct 60
cagttctgtt gttgtgataa caaaagcaca agctactgat atcgactaaa ctaacaagta 120
gtgacaaacc ggagtgatgat ggtgtggtta taccatcgct taggtttacg ttttccacagc 180
acatgatgga taatggcggg gagacagaga gacccgactc tttataatct ttcaaggagc 240
aaagaatgct tattttgact cgtcagagtg gcgaaacctc catcatcgcg gatgaggtaa 300
cgyttaccgt attagagtg aaaggcaacc aggtgcgat tggtgttaac gcacctaaag 360
aggtttctgt ccaccgtgaa gagatctatc agcgtatcca ggcgaaaaaa ttccaaccaa 420
cgtcaatatt attgacaatt cgtctcgtgt tcgcgggagc caattgttat ttccggtttt 480
tcccccaaac attatcogat

```

Genbank

```

LOCUS      ERWRMSMA              500 bp    DNA     linear   BCT 19-AUG-1995
DEFINITION Erwinia carotovora repressor (rsmA) gene, complete cds.
ACCESSION  L40173
VERSION   L40173.1  GI:927031
KEYWORDS  repressor; rsmA gene.
SOURCE    Pectobacterium carotovorum
ORGANISM  Pectobacterium carotovorum
           Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriaceae;
           Pectobacterium.
REFERENCE 1 (bases 1 to 500)
AUTHORS   Cui,Y., Chatterjee,A., Liu,Y., Dumenyo,C.K. and Chatterjee,A.K.
TITLE     Identification of a global repressor gene, rsmA, of Erwinia
carotovora subsp. carotovora that controls extracellular enzymes,
N-(3-oxohexanoyl)-L-homoserine lactone, and pathogenicity in
soft-rotting Erwinia spp
JOURNAL   J. Bacteriol. 177(17) (1995) In press
COMMENT   Original source text: Erwinia carotovora (strain 71, sub_species
carotovora) DNA.

```

Header

```

FEATURES             Location/Qualifiers
     source            1..500
                     /organism="Pectobacterium carotovorum"
                     /strain="71"
                     /sub_species="carotovora"
                     /db_xref="taxon:554"
     gene              107..431
                     /gene="rsmA"
     -10_signal        107..112
                     /gene="rsmA"
     RBS                235..239
                     /gene="rsmA"
     CDS                246..431
                     /gene="rsmA"
                     /function="global repressor"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA74502.1"
                     /db_xref="GI:927032"
                     /translation="MLILTRRVGETLIIGDEVTVLTVLGVKNQVRIGVNPAPKEVSVHR
EETIYORIQAEEKSQPTSY"

```

Annotation

```

BASE COUNT      140 a      101 c      120 g      139 t
ORIGIN
1  ggatccggca agcaggatag aaagtgtgtt accttcagat attctgaagc tttac
61  cagttctgtt gttgtgataa caaaagcaca agctactgat atcgactaaa ctaacaagta
121 gtgacaaacc ggagtgatgat ggtgtggtta taccatcgct taggtttacg ttttccacagc
181 acatgatgga taatggcggg gagacagaga gacccgactc tttataatct ttcaaggagc
241 aaagaatgct tattttgact cgtcagagtg gcgaaacctc catcatcgcg gatgaggtaa
301 cgyttaccgt attagagtg aaaggcaacc aggtgcgat tggtgttaac gcacctaaag
361 aggtttctgt ccaccgtgaa gagatctatc agcgtatcca ggcgaaaaaa ttccaaccaa
421 cgtcaatatt attgacaatt cgtctcgtgt tcgcgggagc caattgttat ttccggtttt
481 tcccccaaac attatcogat

```

Sequence

Artemis panels & navigation

Drop Down Menus

Entry Button Line

Main Sequence View Panel

Magnified Sequence View Panel

Feature Menu

The screenshot displays the Artemis genome browser interface for the *S. typhi* DNA sequence. The window title is "Artemis Entry Edit: S_typhi.dna". The menu bar includes "File", "Entries", "Select", "View", "Goto", "Edit", "Create", "Write", "Run", "Graph", "Display", and "(Noddy)".

The main sequence view shows a genomic map with features represented by colored bars. Features include STY0002, STY0003, STY0004, STY0005, and STY0006. Below the map, the amino acid sequence is displayed, with a magnified view of a specific region highlighted in yellow: "M V K V V A P A S S A N M S V G".

At the bottom, a feature menu lists various features with their coordinates and descriptions:

CDS	190	255	Orthologue of E. coli thrL (LPT_ECOLI); Fasta hit to LPT_EC
CDS	337	2799	Orthologue of E. coli thrA (AK1H_ECOLI); Fasta hit to AK1H_E
misc_feature	343	369	PS00324 Aspartokinase signature
misc_feature	2314	2382	PS01042 Homoserine dehydrogenase signature
CDS	2801	3730	Orthologue of E. coli thrB (KHSE_ECOLI); Fasta hit to KHSE E
misc_feature	3068	3103	PS00627 GHMP kinases putative ATP-binding domain
CDS	3734	5020	Orthologue of E. coli thrC (THRC_ECOLI); Fasta hit to THRC E
misc_feature	4022	4066	PS00165 Serine/threonine dehydratases pyridoxal-phosphate at
CDS	5114	5887	Orthologue of E. coli yaaA (YAAA_ECOLI); Fasta hit to YAAA E
misc_feature	5966	7396	c Similar to Bacillus subtilis amino acid carrier protein alst
misc_feature	7091	7138	c PS00873 Sodium:alanine symporter family signature
CDS	7665	8618	Fasta hit to TALA_ECOLI (316 aa), 65% identity in 311 aa ove
misc_feature	7755	7781	PS01054 Transaldolase signature

Sliders

A word on genome browsers

	Artemis	GIVE	IGB	IGV	Jbrowse	Tablet	UCSC
LOCAL INSTALLATION							
Native app	●		●	●	●	●	
Web app		●●		●●	●●		●
Creation date	1999	2017	2001	2008	2009	2007	2014
Development status (2019)	stalled	early	stalled	mature/ early for web	active	mature	active
Software License	GPL3	Apache 2.0	Common Public License 1.0	MIT License	GNU LGPL v2.1	BSD-2 Clause	Copyright © 2001 UC Regents
PUBLIC WEB INSTANCE							
Creation date		2017		2018			2000
Development status (2019)		early		early			active

Features of the Artemis module

Exercise 1: Familiarize with Artemis

- Loading sequences and annotation files
- Changing the view
- Searching and getting around

Exercise 2: Find insight within a genome

- Graphs and plots

Exercise 3: Analyze a genome

- Basic analysis
- Generating features

Exercise 4: Feature editing

- Adding and Modifying annotations
- Finding evidence : Database searches



**HANDS-ON
LEARNING**