Sequence File Formats

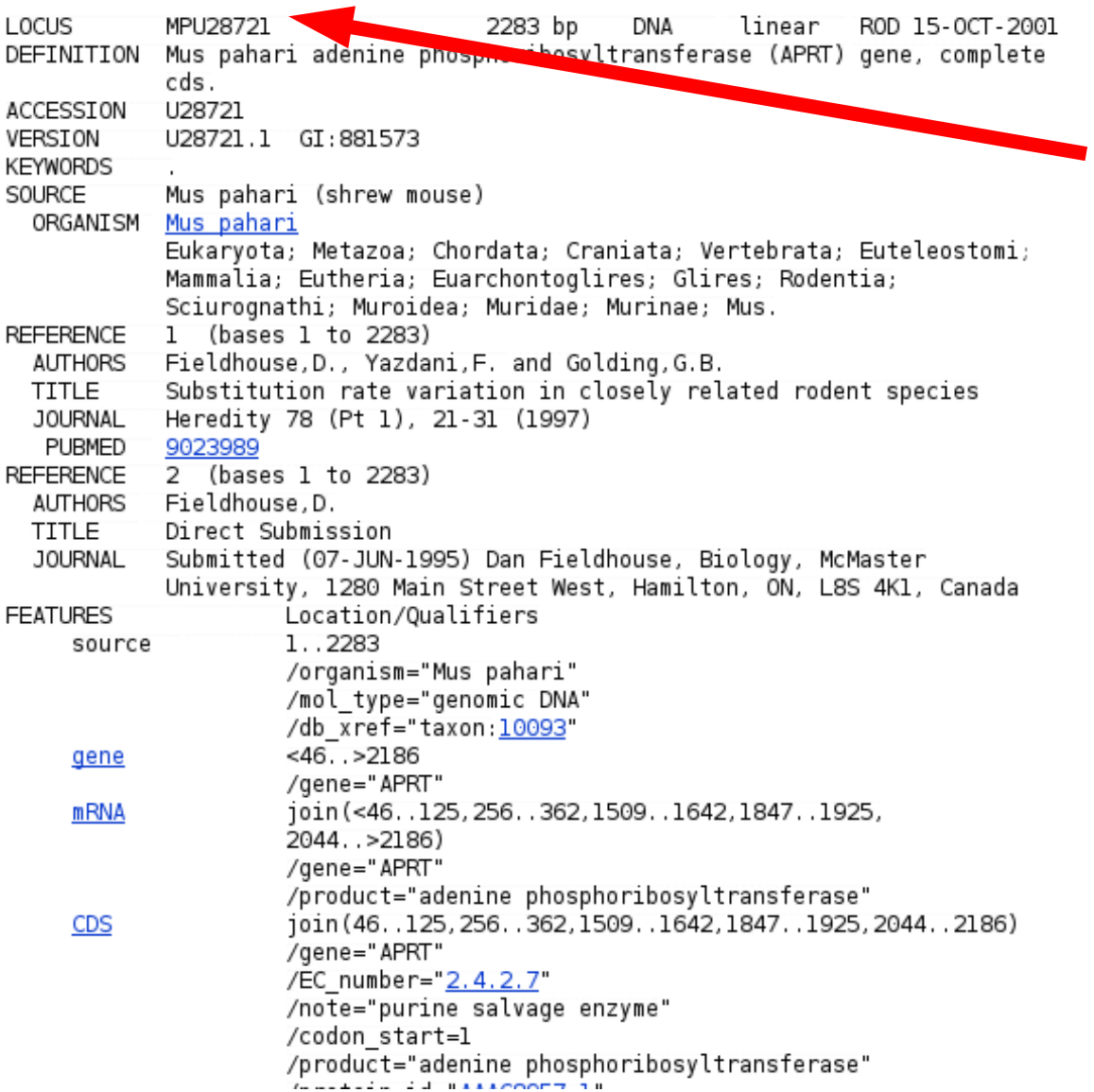
Major File Formats

- GenBank
- EMBL
- FASTA / Pearson
- FASTQ
- Phylip
- GDE
- Nexus
- ASN.1
- VCF
- PDB
- SAM/BAM
- GFF

GenBank

- Developed for a single nucleic sequence
- Does not display homology information
- Can display more than one gene, but only one primary sequence
- Large amount of details
  - organism
  - references
  - features

```
LOCUS       MPU28721                2283 bp    DNA     linear   ROD 15-OCT-2001
DEFINITION  Mus pahari adenine phosphoribosyltransferase (APRT) gene, complete
            cds.
ACCESSION   U28721
VERSION     U28721.1  GI:881573
KEYWORDS    .
SOURCE      Mus pahari (shrew mouse)
  ORGANISM  Mus pahari
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D., Yazdani,F. and Golding,G.B.
  TITLE     Substitution rate variation in closely related rodent species
  JOURNAL   Heredity 78 (Pt 1), 21-31 (1997)
   PUBMED   9023989
REFERENCE   2  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-1995) Dan Fieldhouse, Biology, McMaster
            University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada
FEATURES             Location/Qualifiers
     source          1..2283
                     /organism="Mus pahari"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10093"
     gene            <46..>2186
                     /gene="APRT"
     mRNA            join(<46..125,256..362,1509..1642,1847..1925,
                     2044..>2186)
                     /gene="APRT"
                     /product="adenine phosphoribosyltransferase"
     CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
                     /gene="APRT"
                     /EC_number="2.4.2.7"
                     /note="purine salvage enzyme"
                     /codon_start=1
                     /product="adenine phosphoribosyltransferase"
                     /protein_id="AAA68957.1"
```



```
LOCUS       MPU28721                2283 bp    DNA     linear   ROD 15-OCT-2001
DEFINITION  Mus pahari adenine phosph[illegible]ibosyltransferase (APRT) gene, complete
            cds.
ACCESSION   U28721
VERSION     U28721.1  GI:881573
KEYWORDS    .
SOURCE      Mus pahari (shrew mouse)
  ORGANISM  Mus pahari
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D., Yazdani,F. and Golding,G.B.
  TITLE     Substitution rate variation in closely related rodent species
  JOURNAL   Heredity 78 (Pt 1), 21-31 (1997)
   PUBMED   9023989
REFERENCE   2  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-1995) Dan Fieldhouse, Biology, McMaster
            University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada
FEATURES             Location/Qualifiers
     source          1..2283
                     /organism="Mus pahari"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10093"
     gene            <46..>2186
                     /gene="APRT"
     mRNA            join(<46..125,256..362,1509..1642,1847..1925,
                     2044..>2186)
                     /gene="APRT"
                     /product="adenine phosphoribosyltransferase"
     CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
                     /gene="APRT"
                     /EC_number="2.4.2.7"
                     /note="purine salvage enzyme"
                     /codon_start=1
                     /product="adenine phosphoribosyltransferase"
                     /protein_id="[illegible]"
```

Locus name
must be unique
16 characters or less:
- first two/three – designate organism (first letter of the genus (M), first letter of the species (P))
- 4th and 5th other group information

Sequence length
Type
Topology
Division
Creation date
the position of each field (column) is fixed

Position	Content
01-05	'LOCUS'
06-12	spaces
13-28	Locus name
29-29	space
30-40	Length of sequence, right-justified
41-41	space
42-43	bp
44-44	space
45-47	spaces, ss- (single-stranded), ds- (double-stranded), or ms- (mixed-stranded)
48-53	NA, DNA, RNA, tRNA (transfer RNA), rRNA (ribosomal RNA), mRNA (messenger RNA), uRNA (small nuclear RNA), snRNA, snoRNA. Left justified.
54-55	space
56-63	'linear' followed by two spaces, or 'circular'
64-64	space
65-67	The division code (see Section 3.3)
68-68	space
69-79	Date, in the form dd-MMM-yyyy (e.g., 15-MAR-1991)

```
LOCUS       MPU28721                2283 bp    DNA     linear   ROD 15-OCT-2001
DEFINITION  Mus pahari adenine phosphoribosyltransferase (APRT) gene, complete
            cds.
ACCESSION   U28721
VERSION     U28721.1  GI:881573
KEYWORDS    .
SOURCE      Mus pahari (shrew mouse)
  ORGANISM  Mus pahari
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D., Yazdani,F. and Golding,G.B.
  TITLE     Substitution rate variation in closely related rodent species
  JOURNAL   Heredity 78 (Pt 1), 21-31 (1997)
   PUBMED   9023989
REFERENCE   2  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-1995) Dan Fieldhouse, Biology, McMaster
            University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada
FEATURES             Location/Qualifiers
     source          1..2283
                     /organism="Mus pahari"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10093"
     gene            <46..>2186
                     /gene="APRT"
     mRNA            join(<46..125,256..362,1509..1642,1847..1925,
                     2044..>2186)
                     /gene="APRT"
                     /product="adenine phosphoribosyltransferase"
     CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
                     /gene="APRT"
                     /EC_number="2.4.2.7"
                     /note="purine salvage enzyme"
                     /codon_start=1
                     /product="adenine phosphoribosyltransferase"
                     /protein_id="AAA68957.1"
```

DEFINITION [TITL]
Brief description of the sequence proceeds from general to specific should include:

- organism
- gene/protein name
- description of sequence function (non-coding)
- presence of a CDS (complete / partial)

no limit in number of lines but the last line must ends by a period.

DETERMINED BY THE AUTHORS, MAY VARY

```
LOCUS       MPU28721                2283 bp    DNA     linear   ROD 15-OCT-2001
DEFINITION  Mus pahari adenine phosphoribosyltransferase (APRT) gene, complete
            cds.
ACCESSION   U28721
VERSION     U28721.1  GI:881573
KEYWORDS    .
SOURCE      Mus pahari (shrew mouse)
  ORGANISM  Mus pahari
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D., Yazdani,F. and Golding,G.B.
  TITLE     Substitution rate variation in closely related rodent species
  JOURNAL   Heredity 78 (Pt 1), 21-31 (1997)
   PUBMED   9023989
REFERENCE   2  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-1995) Dan Fieldhouse, Biology, McMaster
            University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada
FEATURES             Location/Qualifiers
     source          1..2283
                     /organism="Mus pahari"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10093"
     gene            <46..>2186
                     /gene="APRT"
     mRNA            join(<46..125,256..362,1509..1642,1847..1925,
                     2044..>2186)
                     /gene="APRT"
                     /product="adenine phosphoribosyltransferase"
     CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
                     /gene="APRT"
                     /EC_number="2.4.2.7"
                     /note="purine salvage enzyme"
                     /codon_start=1
                     /product="adenine phosphoribosyltransferase"
                     /protein_id="AAA68957.1"
```

ACCESSION [ACCN]
Unique identifier
1 letter followed by 5 digits
U28721
2 letters followed by 6 digits
AF123456
(3 letters would indicate a protein)

ACCESSION NUMBERS DO NOT CHANGE, ENTRIES ARE JUST UPDATED

for the RefSeq database:
2 letters, underscore and then 6 or more digits:
NT_ genomic contigs
NM_ mRNAs
NP_ proteins
NC_ chromosomes

```
LOCUS       MPU28721                2283 bp    DNA     linear   ROD 15-OCT-2001
DEFINITION  Mus pahari adenine phosphoribosyltransferase (APRT) gene, complete
            cds.
ACCESSION   U28721
VERSION     U28721.1  GI:881573
KEYWORDS    .
SOURCE      Mus pahari (shrew mouse)
  ORGANISM  Mus pahari
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D., Yazdani,F. and Golding,G.B.
  TITLE     Substitution rate variation in closely related rodent species
  JOURNAL   Heredity 78 (Pt 1), 21-31 (1997)
   PUBMED   9023989
REFERENCE   2  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-1995) Dan Fieldhouse, Biology, McMaster
            University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada
FEATURES             Location/Qualifiers
     source          1..2283
                     /organism="Mus pahari"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10093"
     gene            <46..>2186
                     /gene="APRT"
     mRNA            join(<46..125,256..362,1509..1642,1847..1925,
                     2044..>2186)
                     /gene="APRT"
                     /product="adenine phosphoribosyltransferase"
     CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
                     /gene="APRT"
                     /EC_number="2.4.2.7"
                     /note="purine salvage enzyme"
                     /codon_start=1
                     /product="adenine phosphoribosyltransferase"
```

VERSION

Identification number for a single nucleotide sequence

Format implemented by GenBank/EMBL/DDBJ

2 numbers:

compound number – a stable part (Accession), and a sequentially increasing number

in case of change: U28721.1→U28721.2

A new GI (GenInfo Identifier) number is assigned for every change

The GI number is a unique number associated to every sequence NCBI has processed

```
LOCUS       MPU28721                2283 bp    DNA     linear   ROD 15-OCT-2001
DEFINITION  Mus pahari adenine phosphoribosyltransferase (APRT) gene, complete
            cds.
ACCESSION   U28721
VERSION     U28721.1  GI:881573
KEYWORDS    .
SOURCE      Mus pahari (shrew mouse)
  ORGANISM  Mus pahari
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D., Yazdani,F. and Golding,G.B.
  TITLE     Substitution rate variation in closely related rodent species
  JOURNAL   Heredity 78 (Pt 1), 21-31 (1997)
   PUBMED   9023989
REFERENCE   2  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-1995) Dan Fieldhouse, Biology, McMaster
            University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada
FEATURES             Location/Qualifiers
     source          1..2283
                     /organism="Mus pahari"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10093"
     gene            <46..>2186
                     /gene="APRT"
     mRNA            join(<46..125,256..362,1509..1642,1847..1925,
                     2044..>2186)
                     /gene="APRT"
                     /product="adenine phosphoribosyltransferase"
     CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
                     /gene="APRT"
                     /EC_number="2.4.2.7"
                     /note="purine salvage enzyme"
                     /codon_start=1
                     /product="adenine phosphoribosyltransferase"
                     /protein_id="AAA68957.1"
```

KEYWORDS [KYWD]

Words or sentence describing the sequence

Each keyword is separated by a semicolon

The last line should end by a period

NOT PRESENT IN EVERY RECORD

```
LOCUS       MPU28721                2283 bp    DNA     linear   ROD 15-OCT-2001
DEFINITION  Mus pahari adenine phosphoribosyltransferase (APRT) gene, complete
            cds.
ACCESSION   U28721
VERSION     U28721.1  GI:881573
KEYWORDS    .
SOURCE      Mus pahari (shrew mouse)
  ORGANISM  Mus pahari
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D., Yazdani,F. and Golding,G.B.
  TITLE     Substitution rate variation in closely related rodent species
  JOURNAL   Heredity 78 (Pt 1), 21-31 (1997)
   PUBMED   9023989
REFERENCE   2  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-1995) Dan Fieldhouse, Biology, McMaster
            University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada
FEATURES             Location/Qualifiers
     source          1..2283
                     /organism="Mus pahari"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10093"
     gene            <46..>2186
                     /gene="APRT"
     mRNA            join(<46..125,256..362,1509..1642,1847..1925,
                     2044..>2186)
                     /gene="APRT"
                     /product="adenine phosphoribosyltransferase"
     CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
                     /gene="APRT"
                     /EC_number="2.4.2.7"
                     /note="purine salvage enzyme"
                     /codon_start=1
                     /product="adenine phosphoribosyltransferase"
```

SOURCE
Free format information

```
LOCUS       MPU28721                2283 bp    DNA     linear   ROD 15-OCT-2001
DEFINITION  Mus pahari adenine phosphoribosyltransferase (APRT) gene, complete
            cds.
ACCESSION   U28721
VERSION     U28721.1  GI:881573
KEYWORDS    .
SOURCE      Mus pahari (shrew mouse)
  ORGANISM  Mus pahari
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D., Yazdani,F. and Golding,G.B.
  TITLE     Substitution rate variation in closely related rodent species
  JOURNAL   Heredity 78 (Pt 1), 21-31 (1997)
   PUBMED   9023989
REFERENCE   2  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-1995) Dan Fieldhouse, Biology, McMaster
            University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada
FEATURES             Location/Qualifiers
     source          1..2283
                     /organism="Mus pahari"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10093"
     gene            <46..>2186
                     /gene="APRT"
     mRNA            join(<46..125,256..362,1509..1642,1847..1925,
                     2044..>2186)
                     /gene="APRT"
                     /product="adenine phosphoribosyltransferase"
     CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
                     /gene="APRT"
                     /EC_number="2.4.2.7"
                     /note="purine salvage enzyme"
                     /codon_start=1
                     /product="adenine phosphoribosyltransferase"
                     /protein_id="AAA68957.1"
```

ORGANISM [ORGN]
Scientific name of the organism following the NCBI Taxonomy Database
- first line:
Genus species
- second line:
Full formal classification

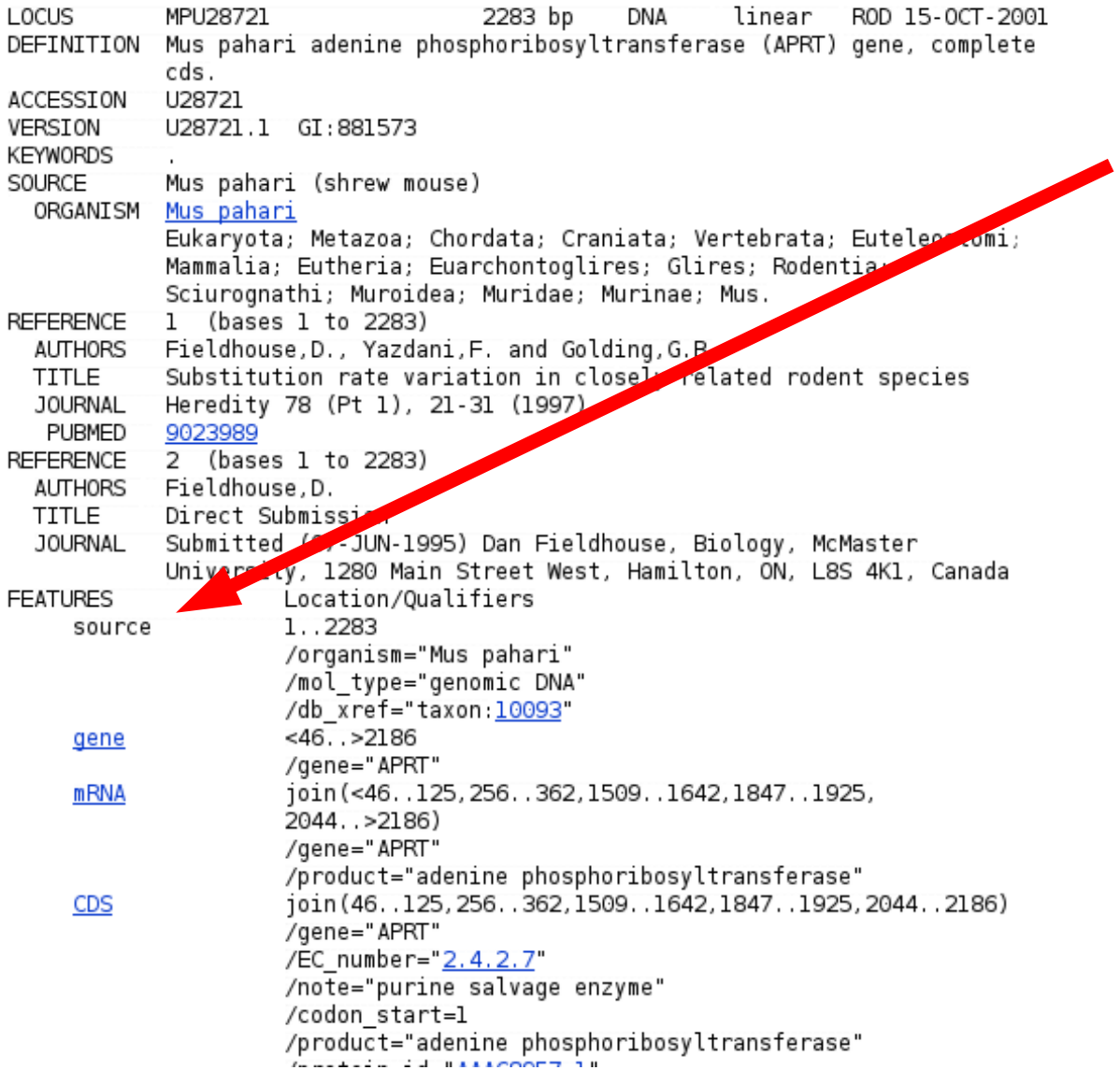
```
LOCUS       MPU28721                2283 bp    DNA     linear   ROD 15-OCT-2001
DEFINITION  Mus pahari adenine phosphoribosyltransferase (APRT) gene, complete
            cds.
ACCESSION   U28721
VERSION     U28721.1  GI:881573
KEYWORDS    .
SOURCE      Mus pahari (shrew mouse)
  ORGANISM  Mus pahari
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D., Yazdani,F. and Golding,G.B.
  TITLE     Substitution rate variation in closely related rodent species
  JOURNAL   Heredity 78 (Pt 1), 21-31 (1997)
   PUBMED   9023989
REFERENCE   2  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-1995) Dan Fieldhouse, Biology, McMaster
            University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada
FEATURES             Location/Qualifiers
     source          1..2283
                     /organism="Mus pahari"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10093"
     gene            <46..>2186
                     /gene="APRT"
     mRNA            join(<46..125,256..362,1509..1642,1847..1925,
                     2044..>2186)
                     /gene="APRT"
                     /product="adenine phosphoribosyltransferase"
     CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
                     /gene="APRT"
                     /EC_number="2.4.2.7"
                     /note="purine salvage enzyme"
                     /codon_start=1
                     /product="adenine phosphoribosyltransferase"
                     /protein_id="AAA68957.1"
```

REFERENCE
Information about the publication of the sequence, divided into optional parts:

- Author(s) name(s) [AUTH]
- Title [TITL]
- Journal of submission [JOUR]
- PUBMED reference
- REMARK

The reference field contains the number and the range of bases used in the citation

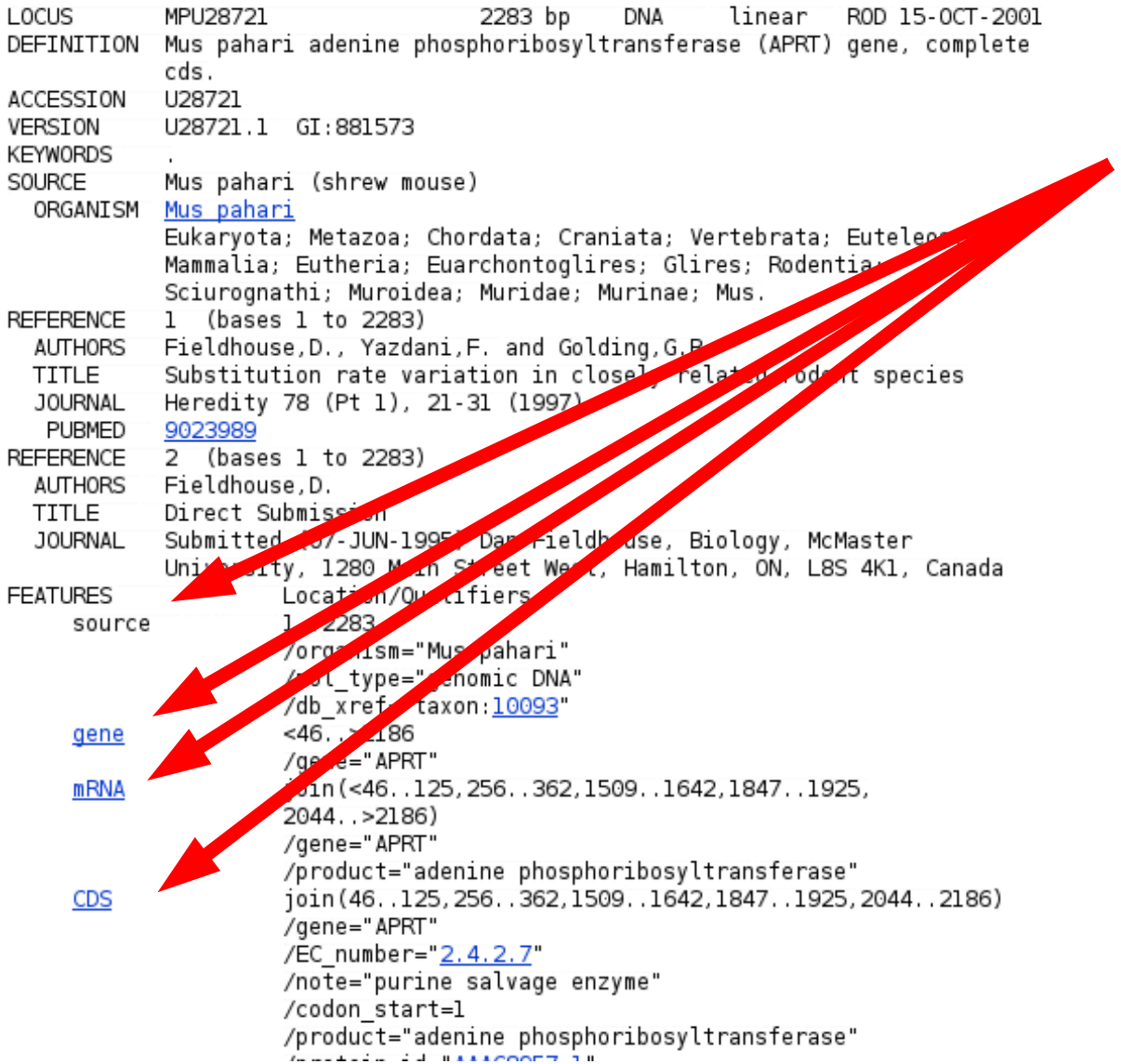
The last publication field may contain information about the submitter.



```
LOCUS       MPU28721                2283 bp    DNA     linear   ROD 15-OCT-2001
DEFINITION  Mus pahari adenine phosphoribosyltransferase (APRT) gene, complete
            cds.
ACCESSION   U28721
VERSION     U28721.1  GI:881573
KEYWORDS    .
SOURCE      Mus pahari (shrew mouse)
  ORGANISM  Mus pahari
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D., Yazdani,F. and Golding,G.B.
  TITLE     Substitution rate variation in closely related rodent species
  JOURNAL   Heredity 78 (Pt 1), 21-31 (1997)
   PUBMED   9023989
REFERENCE   2  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-1995) Dan Fieldhouse, Biology, McMaster
            University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada
FEATURES             Location/Qualifiers
     source          1..2283
                     /organism="Mus pahari"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10093"
     gene            <46..>2186
                     /gene="APRT"
     mRNA            join(<46..125,256..362,1509..1642,1847..1925,
                     2044..>2186)
                     /gene="APRT"
                     /product="adenine phosphoribosyltransferase"
     CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
                     /gene="APRT"
                     /EC_number="2.4.2.7"
                     /note="purine salvage enzyme"
                     /codon_start=1
                     /product="adenine phosphoribosyltransferase"
```

FEATURES

Information about genes and genes products
Regions of biological significance



Keys

Gives the biological nature of the annotated feature

source
promoter
mRNA
CDS

```
LOCUS       MPU28721                2283 bp    DNA     linear   ROD 15-OCT-2001
DEFINITION  Mus pahari adenine phosphoribosyltransferase (APRT) gene, complete
            cds.
ACCESSION   U28721
VERSION     U28721.1  GI:881573
KEYWORDS    .
SOURCE      Mus pahari (shrew mouse)
  ORGANISM  Mus pahari
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D., Yazdani,F. and Golding,G.B.
  TITLE     Substitution rate variation in closely related rodent species
  JOURNAL   Heredity 78 (Pt 1), 21-31 (1997)
   PUBMED   9023989
REFERENCE   2  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-1995) Dan Fieldhouse, Biology, McMaster
            University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada
FEATURES             Location/Qualifiers
     source          1..2283
                     /organism="Mus pahari"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10093"
     gene            <46..>2186
                     /gene="APRT"
     mRNA            join(<46..125,256..362,1509..1642,1847..1925,
                     2044..>2186)
                     /gene="APRT"
                     /product="adenine phosphoribosyltransferase"
     CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
                     /gene="APRT"
                     /EC_number="2.4.2.7"
                     /note="purine salvage enzyme"
                     /codon_start=1
                     /product="adenine phosphoribosyltransferase"
                     /protein_id="AAA68057.1"
```

```
LOCUS       MPU28721                2283 bp    DNA     linear   ROD 15-OCT-2001
DEFINITION  Mus pahari adenine phosphoribosyltransferase (APRT) gene, complete
            cds.
ACCESSION   U28721
VERSION     U28721.1  GI:881573
KEYWORDS    .
SOURCE      Mus pahari (shrew mouse)
  ORGANISM  Mus pahari
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D., Yazdani,F. and Golding,G.B.
  TITLE     Substitution rate variation in closely related rodent species
  JOURNAL   Heredity 78 (Pt 1), 21-31 (1997)
   PUBMED   9023989
REFERENCE   2  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-1995) Dan Fieldhouse, Biology, McMaster
            University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada
FEATURES             Location/Qualifiers
     source          1..2283
                     /organism="Mus pahari"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10093"
     gene            <46..>2186
                     /gene="APRT"
     mRNA            join(<46..125,256..362,1509..1642,1847..1925,
                     2044..>2186)
                     /gene="APRT"
                     /product="adenine phosphoribosyltransferase"
     CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
                     /gene="APRT"
                     /EC_number="2.4.2.7"
                     /note="purine salvage enzyme"
                     /codon_start=1
                     /product="adenine phosphoribosyltransferase"
                     /protein_id="AAA98857.1"
```

Locations
Region of the sequence where the feature is located

“<” indicates that the start is unknown
“>” indicates that the end is unknown

Different operators
- complement
- join
- order

```
LOCUS       MPU28721                2283 bp    DNA     linear   ROD 15-OCT-2001
DEFINITION  Mus pahari adenine phosphoribosyltransferase (APRT) gene, complete
            cds.
ACCESSION   U28721
VERSION     U28721.1  GI:881573
KEYWORDS    .
SOURCE      Mus pahari (shrew mouse)
  ORGANISM  Mus pahari
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D., Yazdani,F. and Golding,G.B.
  TITLE     Substitution rate variation in closely related rodent species
  JOURNAL   Heredity 78 (Pt 1), 21-31 (1997)
   PUBMED   9023989
REFERENCE   2  (bases 1 to 2283)
  AUTHORS   Fieldhouse,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-1995) Dan Fieldhouse, Biology, McMaster
            University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada
FEATURES             Location/Qualifiers
     source          1..2283
                     /organism="Mus pahari"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10093"
     gene            <46..>2186
                     /gene="APRT"
     mRNA            join(<46..125,256..362,1509..1642,1847..1925,
                     2044..>2186)
                     /gene="APRT"
                     /product="adenine phosphoribosyltransferase"
     CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
                     /gene="APRT"
                     /EC_number="2.4.2.7"
                     /note="purine salvage enzyme"
                     /codon_start=1
                     /product="adenine phosphoribosyltransferase"
                     /protein_id="AAA68057.1"
```

Qualifiers

Always in the form:
/[qualifer]=” “

/product="adenine phosphoribosyltransferase"
/protein_id="AAA68957.1"
/db_xref="GI:881574"
/translation="MSESELKLVARRIRSFPDFPIPGVLFRDISPLLKDPDSFRASIR
LLASHLKSTHSGKIDYIAGLDSRGFLFGPSLAQELGVGCVLIRKQGKLPGPTISASYA
LEYGKAELEIQKDALEPGQRVVIVDDLLATGGTMFAACDLLHQLRAEVVECVSLVELT
SLKGRERLGPIPFFSLLQYD"

ORIGIN

ORIGIN

```
        1 cctgcggata ctcacctcct ccttgtctcc tacaagcacg cggccatgtc cgagtctgag
       61 ttgaaactgg tggcgcggcg catccgcagc ttccccgact tccccatccc gggcgtgctg
      121 ttcaggtgcg gtcacgagcc ggcgaggcgt tggcgccgta ctctcatccc ccggcgcagg
      181 cgcgtgggca gccttgggga tcttgcgggg cctctgcccg gccacacgcg gtcactctcc
      241 tgtccttgtt cccagggata tctcgccccc cttgaaagat ccggactcct tccgagcttc
      301 catccgcctc ctggccagtc acctgaagtc cacgcacagc ggcaagatcg actatatcgc
      361 agggcaaggt ggccttgcta ggccgtactc atcccccacg gtcctatccc ctatcccctt
      421 tcccctcgtg tcacccacag tctaccccac acccatccat tctttcttta acctctgact
      481 cttcctcctt ggtttctcac tgccttggac gcttgttcac cccggatgaa ctccgtaggc
      541 gtctcccttc cctgcttggt accctaaggt gccctcggtg cttgttcgta gagacgaact
      601 ctgctctgtc cttgtgtcca gaaccaagcc ttcctctttt agggcacaag gctggccagc
      661 atcctgacag caggctggga gacctggatc ctccagatga cggacatcct tacttagggg
      721 tagcctctgg gatgaactag atattaaaaa ttaggtaacc ttggggcgtg cctgggcaga
      781 cctcaagtct ggtagcttca ggggctgctt ctccccagga ctacaccggg gcactcttct
      841 cttgtcctcc ccacccccca caagcttgtg ctaaacaact gctgtatacc aggctccatg
      901 cttgagcttc agaaacaccc tagggcagct gaatgtccac caggagtgtc cagagggagg
      961 gtgagcaccc caagagaaca gagtggccct agtaaatgct cagggaccac agaacttttg
     1021 cccactccac ttcctattgg taccccggc catgccccag aaatcagggc atgtttgtac
     1081 cctccccacg acagctcggg ccgcctggaa ctgacctgta gacagtgctc ctgggtagat
     1141 gctgcatttg aaaggtggca agagggctgg tgagatggct cagcggttag gagcactgac
     1201 tgctcttcca aaggtcctga gttcaaatcc cagcaaccac atggtggctc acaaccacct
     1261 acagctacag tgtacacaca tataataaaa taaataaaca aatcttaaaa aaaaaaaaaa
     1321 gaaagaaagg tggcaagagc caccatagtg gagaaggcag gtaggatccc caaggctaag
     1381 atgctaccga gtaaccatca gtgttcttct agccatagtg ggcaagacct agtgttccta
     1441 gtcaatgttg acctctccat acttgcctct cggctccatc ccacaccctt ccctccttac
     1501 cctaacaggt ctagactcca ggggcttcct gtttggccct tccctagctc aggagctggg
     1561 cgtgggctgc gtgctcatcc ggaagcaggg gaagctgccg ggccccacta tatcagcctc
     1621 ctatgctctg gagtatggga aggtaaggga gctgtgggta gaggaagggc agggtcttat
     1681 taccacggct accagtgcct aggagtaaat gtgggtgctc agagaggttg agacattggg
     1741 gtgaggttta caactcctga aatgctcagc ctcagaaatg ctccaggcta gggaggtggc
     1801 cacttgttag catctagact ctcttaacgc tacttcctgt ctgcaggctg agctggaaat
     1861 ccagaaagat gccttagaac ccgggcagag agtggtcatt gtggatgacc tcctggccac
     1921 tggaggtaaa gaaccagccc aagacaaaca ggcttcaaag ggccaggccc tgtctggggt
     1981 gctgactaaa caaagcgctt gaataccttc tctttctctg tcccttcccc ccccccccc
     2041 caggaaccat gtttgcagcc tgtgatctgc tgcaccagct acgggctgag gtggtggagt
     2101 gtgtgagcct ggtggagctg acctcgctga agggcaggga gaggctagga cctataccat
     2161 tcttctctct cctccagtat gactgagctg gctagatggt cacacccctg ctcacagcag
     2221 cagtaactgc gcggtggctc agccctgggc gcctaagtga cctttgtgag ctacctgctg
     2281 ccc
//
```

/product="adenine phosphoribosyltransferase"
/protein_id="AAA68957.1"
/db_xref="GI:881574"
/translation="MSESELKLVARRIRSFPDFPIPGVLFRDISPLLKDPDSFRASIR
LLASHLKSTHSGKIDYIAGLDSRGFLFGPSLAQELGVGCVLIRKQGKLPGPTISASYA
LEYGKAELEIQKDALEPGQRVVIVDDLLATGGTMFAACDLLHQLRAEVVECVSLVELT
SLKGRERLGPIPFFSLLQYD"

ORIGIN

```
       1 cctgcggata ctcacctcct ccttgtctcc tacaagcacg cggccatgtc cgagtctgag
      61 ttgaaactgg tggcgcggcg catccgcagc ttccccgact tccccatccc gggcgtgctg
     121 ttcaggtgcg gtcacgagcc ggcgaggcgt tggcgccgta ctctcatccc ccggcgcagg
     181 cgcgtgggca gccttgggga tcttgcgggg cctctgcccg gccacacgcg gtcactctcc
     241 tgtccttgtt cccagggata tctcgccccc cttgaaagat ccggactcct tccgagcttc
     301 catccgcctc ctggccagtc acctgaagtc cacgcacagc ggcaagatcg actatatcgc
     361 agggcaaggt ggccttgcta ggccgtactc atcccccacg gtcctatccc ctatcccctt
     421 tcccctcgtg tcacccacag tctaccccac acccatccat tctttcttta acctctgact
     481 cttcctcctt ggtttctcac tgccttggac gcttgttcac cccggatgaa ctccgtaggc
     541 gtctcccttc cctgcttggt accctaaggt gccctcggtg cttgttcgta gagacgaatt
     601 ctgctctgtc cttgtgtcca gaaccaagcc ttcctctttt agggcacaag gctgcctagc
     661 atcctgacag caggctggga gacctggatc ctccagatga cggacatcct tacctagggg
     721 tagcctctgg gatgaactag atattaaaaa ttaggtaacc ttggggcgtg gctgggcaga
     781 cctcaagtct ggtagcttca ggggctgctt ctccccagga ctacaccgga gcactcttct
     841 cttgtcctcc ccacccccca caagcttgtg ctaaacaact gctgtatacc aggctccatg
     901 cttgagcttc agaaacaccc tagggcagct gaatgtccac caggagtgtc cagagggagg
     961 gtgagcaccc caagagaaca gagtggccct agtaaatgct caaggaccac agaacttttg
    1021 cccactccac ttcctattgg taccccggc catgccccag gaatcagggc atgtttgtac
    1081 cctccccacg acagctcggg ccgcctggaa ctgacctgtc gacagtgctc ctgggtagat
    1141 gctgcatttg aaaggtggca agagggctgg tgagatggct cagcggttag gagcactgac
    1201 tgctcttcca aaggtcctga gttcaaatcc caggaaccac atggtggctc acaaccacct
    1261 acagctacag tgtacacaca tataataaaa tgataaaaca aatcttaaaa aaaaaaaaaa
    1321 gaaagaaagg tggcaagagc caccatagtg gagaaggcag gtaggatccc caaggctaag
    1381 atgctaccga gtaaccatca gtgttctttc agccatagtg ggcaagacct agtgttccta
    1441 gtcaatgttg acctctccat acttgattct cggctccatc ccacaccctt ccctccttac
    1501 cctaacaggt ctagactcca gggcattcct gtttggccct tccctagctc aggagctggg
    1561 cgtgggctgc gtgctcatcc cgaagcaggg gaagctgccg ggccccacta tatcagcctc
    1621 ctatgctctg gagtatgggc aggtaaggga gctgtgggta gaggaagggc agggtcttat
    1681 taccacggct accagtgact aggagtaaat gtgggtgctc agagaggttg agacattggg
    1741 gtgaggttta caactcctga aatgctcagc ctcagaaatg ctccaggcta gggaggtggc
    1801 cacttgttag catctagact ctcttaacgc tacttcctgt ctgcaggctg agctggaaat
    1861 ccagaaagat cctcttagaac ccgggcagag agtggtcatt gtggatgacc tcctggccac
    1921 tggaggtatc gaaccagccc aagacaaaca ggcttcaaag ggccaggccc tgtctggggt
    1981 gctgacaaaa caaagcgctt gaataccttc tctttctctg tcccttcccc ccccccccc
    2041 caggtaccat gtttgcagcc tgtgatctgc tgcaccagct acgggctgag gtggtggagt
    2101 gtggtgagcct ggtggagctg acctcgctga agggcaggga gaggctagga cctataccat
    2161 ccttctctct cctccagtat gactgagctg gctagatggt cacacccctg ctcacagcag
    2221 cagtaactgc gcggtggctc agccctgggc gcctaagtga cctttgtgag ctacctgctg
    2281 ccc
```

//

End of the entry, always //

```
                     /db_xref="taxon:7227"
                     /chromosome="2L"
     gene            7529..9491
                     /gene="CG11023"
                     /locus_tag="Dmel_CG11023"
                     /old_locus_tag="CG11023"
                     /note="CG11023"
                     /map="21A5-21A5"
                     /db_xref="FLYBASE:FBgn0031208"
                     /db_xref="GeneID:33155"
     mRNA            join(7529..8116,8229..8589,8668..9491)
                     /gene="CG11023"
                     /locus_tag="Dmel_CG11023"
                     /old_locus_tag="CG11023"
                     /product="CG11023-RA"
                     /transcript_id="NM_175941.1"
                     /db_xref="GI:28573981"
                     /db_xref="FLYBASE:FBgn0031208"
                     /db_xref="GeneID:33155"
     CDS             join(7680..8116,8229..8589,8668..9276)
                     /gene="CG11023"
                     /locus_tag="Dmel_CG11023"
                     /old_locus_tag="CG11023"
                     /codon_start=1
                     /protein_id="NP_787955.1"
                     /db_xref="GI:28573982"
                     /db_xref="FLYBASE:FBgn0031208"
                     /db_xref="GeneID:33155"
     gene            complement(9836..21372)
                     /gene="l(2)gl"
                     /locus_tag="Dmel_CG2671"
                     /old_locus_tag="CG2671"
                     /note="lethal (2) giant larvae; synonyms: CG2671, D-LGL,
                     L(2)GL, LGL, dlgl, gl, l(2), l(2) giant larva, l(2)giant
                     larvae, l-gl, l[[2]]gl, p127, p127l(2)gl"
                     /map="21A5-21A5"
                     /db_xref="FLYBASE:FBgn0002121"
                     /db_xref="GeneID:33156"
     mRNA            complement(join(9836..11344,11410..11518,11779..12221,
                     12286..12928,13520..13625,13683..14874,14933..15711,
                     19880..20020,21066..21200,21349..21372))
                     /gene="l(2)gl"
                     /locus_tag="Dmel_CG2671"
                     /old_locus_tag="CG2671"
                     /product="lethal (2) giant larvae CG2671-RF, transcript
                     variant F"
```

GenBank format can display more than one gene in a single sequence

RefSeq Genbank Format

- A more formalized recording for “reference” sequences. For example the accession prefixes ...

Accession prefix	Molecule type	Comment
AC_	Genomic	Complete genomic molecule, usually alternate assembly
NC_	Genomic	Complete genomic molecule, usually reference assembly
NG_	Genomic	Incomplete genomic region
NT_	Genomic	Contig or scaffold, clone-based or WGS
NW_	Genomic	Contig or scaffold, primarily WGS
NS_	Genomic	Environmental sequence
NZ_	Genomic	Unfinished WGS
NM_	mRNA	
NR_	RNA	
XM_	mRNA	Predicted model
XR_	RNA	Predicted model
AP_	Protein	Annotated on AC_ alternate assembly
NP_	Protein	Associate with an N_ or NC_ accession
YP_	Protein	
XP_	Protein	Predicted model, associated with an XM_ accession
ZP_	Protein	Predicted model, annotated on NZ_ genomic records

EMBL Format

- Developed for a single nucleic sequence
- Can display more than one gene
- Large amount of detail
- Same as GenBank format but with 2 letters codes

```
ID   U28721; SV 1; linear; genomic DNA; STD; ROD; 2283 BP.
XX
AC   U28721;
XX
DT   04-JUL-1995 (Rel. 44, Created)
DT   17-APR-2005 (Rel. 83, Last updated, Version 5)
XX
DE   Mus pahari adenine phosphoribosyltransferase (APRT) gene, complete cds.
XX
KW   .
XX
OS   Mus pahari (shrew mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea;
OC   Muridae; Murinae; Mus.
XX
RN   [1]
RP   1-2283
RX   DOI; 10.1038/sj.hdy.6881150
RX   PUBMED; 16397634.
RA   Fieldhouse D., Yazdani F., Golding G.B.;
RT   "Substitution rate variation in closely related rodent species";
RL   Heredity 78(1):21-31(1997).
XX
RN   [2]
RP   1-2283
RA   Fieldhouse D.;
RT   ;
RL   Submitted (07-JUN-1995) to the EMBL/GenBank/DDBJ databases.
RL   Dan Fieldhouse, Biology, McMaster University, 1280 Main Street West,
RL   Hamilton, ON, L8S 4K1, Canada
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..2283
FT                   /organism="Mus pahari"
FT                   /mol_type="genomic DNA"
FT                   /db_xref="taxon:10093"
FT   mRNA            join(<46..125,256..362,1509..1642,1847..1925,2044..>2186)
FT                   /gene="APRT"
FT                   /product="adenine phosphoribosyltransferase"
FT   CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
```

Line codes

ID - identification
AC - accession number
PR - project identifier
DT - date
DE - description
KW - keyword
OS - organism species
OC - organism classification
OG - organelle
RN - reference number
RC - reference comment
RP - reference positions
RX - reference cross-reference
RG - reference group
RA - reference author(s)
RT - reference title
RL - reference location
DR - database cross-reference
CC - comments or notes
AH - assembly header
AS - assembly information
FH - feature table header
FT - feature table data
XX - spacer line
SQ - sequence header
CO - contig/construct line
bb - (blanks) sequence data
// - termination line

```
ID   U28721; SV 1; linear; genomic DNA; STD; ROD; 2283 BP.
XX
AC   U28721;
XX
DT   04-JUL-1995 (Rel. 44, Created)
DT   17-APR-2005 (Rel. 83, Last updated, Version 5)
XX
DE   Mus pahari adenine phosphoribosyltransferase (APRT) gene, complete cds.
XX
KW   .
XX
OS   Mus pahari (shrew mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea;
OC   Muridae; Murinae; Mus.
XX
RN   [1]
RP   1-2283
RX   DOI; 10.1038/sj.hdy.6881150
RX   PUBMED; 16397634.
RA   Fieldhouse D., Yazdani F., Golding G.B.;
RT   "Substitution rate variation in closely related rodent species";
RL   Heredity 78(1):21-31(1997).
XX
RN   [2]
RP   1-2283
RA   Fieldhouse D.;
RT   ;
RL   Submitted (07-JUN-1995) to the EMBL/GenBank/DDBJ databases.
RL   Dan Fieldhouse, Biology, McMaster University, 1280 Main Street West,
RL   Hamilton, ON, L8S 4K1, Canada
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..2283
FT                   /organism="Mus pahari"
FT                   /mol_type="genomic DNA"
FT                   /db_xref="taxon:10093"
FT   mRNA            join(<46..125,256..362,1509..1642,1847..1925,2044..>2186)
FT                   /gene="APRT"
FT                   /product="adenine phosphoribosyltransferase"
FT   CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
```

DATA CLASS

How the data was generated

CON	fragments of entered sequences
ANN	fragment of entries sequences with its own annotation
PAT	Patent
EST	Expressed Sequence Tag
GSS	Genome Survey Sequence
HTC	High throughput cDNA
HTG	High throughput genomic
MGA	Mass Genome Annotation
WGS	Whole Genome Shotgun
TPA	Third Party Annotation
STS	Sequence Tag Site
STD	Standard

```
ID   U28721; SV 1; linear; genomic DNA; STD; ROD; 2283 BP.
XX
AC   U28721;
XX
DT   04-JUL-1995 (Rel. 44, Created)
DT   17-APR-2005 (Rel. 83, Last updated, Version 5)
XX
DE   Mus pahari adenine phosphoribosyltransferase (APRT) gene, complete cds.
XX
KW   .
XX
OS   Mus pahari (shrew mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea;
OC   Muridae; Murinae; Mus.
XX
RN   [1]
RP   1-2283
RX   DOI; 10.1038/sj.hdy.6881150
RX   PUBMED; 16397634.
RA   Fieldhouse D., Yazdani F., Golding G.B.;
RT   "Substitution rate variation in closely related rodent species";
RL   Heredity 78(1):21-31(1997).
XX
RN   [2]
RP   1-2283
RA   Fieldhouse D.;
RT   ;
RL   Submitted (07-JUN-1995) to the EMBL/GenBank/DDBJ databases.
RL   Dan Fieldhouse, Biology, McMaster University, 1280 Main Street West,
RL   Hamilton, ON, L8S 4K1, Canada
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..2283
FT                   /organism="Mus pahari"
FT                   /mol_type="genomic DNA"
FT                   /db_xref="taxon:10093"
FT   mRNA            join(<46..125,256..362,1509..1642,1847..1925,2044..>2186)
FT                   /gene="APRT"
FT                   /product="adenine phosphoribosyltransferase"
FT   CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
```

TAXONOMIC DIVISION

PHG	Bacteriophage
ENV	Environmental
FUN	Fungal
HUM	*Homo sapiens*
INV	Invertebrates
MAM	Other mammals
VRT	Vertebrate
MUS	*Mus musculus*
PLN	Plant
PRO	Prokaryote
ROD	Other rodents
SYN	Synthetic
TGN	Transgenic
UNC	Unclassified
VRL	Viral

```
ID   U28721; SV 1; linear; genomic DNA; STD; ROD; 2283 BP.
XX
AC   U28721;
XX
DT   04-JUL-1995 (Rel. 44, Created)
DT   17-APR-2005 (Rel. 83, Last updated, Version 5)
XX
DE   Mus pahari adenine phosphoribosyltransferase (APRT) gene, complete cds.
XX
KW   .
XX
OS   Mus pahari (shrew mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea;
OC   Muridae; Murinae; Mus.
XX
RN   [1]
RP   1-2283
RX   DOI; 10.1038/sj.hdy.6881150
RX   PUBMED; 16397634.
RA   Fieldhouse D., Yazdani F., Golding G.B.;
RT   "Substitution rate variation in closely related rodent species";
RL   Heredity 78(1):21-31(1997).
XX
RN   [2]
RP   1-2283
RA   Fieldhouse D.;
RT   ;
RL   Submitted (07-JUN-1995) to the EMBL/GenBank/DDBJ databases.
RL   Dan Fieldhouse, Biology, McMaster University, 1280 Main Street West,
RL   Hamilton, ON, L8S 4K1, Canada
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..2283
FT                   /organism="Mus pahari"
FT                   /mol_type="genomic DNA"
FT                   /db_xref="taxon:10093"
FT   mRNA            join(<46..125,256..362,1509..1642,1847..1925,2044..>2186)
FT                   /gene="APRT"
FT                   /product="adenine phosphoribosyltransferase"
FT   CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
```

FEATURE

Qualifiers:
SAME FORMAT AS FOR GenBank

```
FT                   /translation="MSESELKLVARRIRSFPDFPIPGVLFRDISPLLKDPDSFRASIRL
FT                   LASHLKSTHSGKIDYIAGLDSRGFLFGPSLAQELGVGCVLIRKQGKLPGPTISASYALE
FT                   YGKAELEIQKDALEPGQRVVIVDDLLATGGTMFAACDLLHQLRAEVVECVSLVELTSLK
FT                   GRERLGPIPFFSLLQYD"
XX
SQ   Sequence 2283 BP; 485 A; 696 C; 590 G; 512 T; 0 other;
     cctgcggata ctcacctcct ccttgtctcc tacaagcacg cggccatgtc cgagcctgag        60
     ttgaaactgg tggcgcggcg catccgcagc ttccccgact tccccatccc gggcgtgctg       120
     ttcaggtgcg gtcacgagcc ggcgaggcgt tggcgccgta ctctcatccc ccggcgcagg       180
     cgcgtgggca gccttgggga tcttgcgggg cctctgcccg gccacacgcg gtcactctcc       240
     tgtccttgtt cccagggata tctcgcccct cttgaaagat ccggactcct tccgagcttc       300
     catccgcctc ctggccagtc acctgaagtc cacgcacagc ggcaagatcg actatatcgc       360
     agggcaaggt ggccttgcta ggccgtactc atcccccacg gtcctatccc ctatcccctt       420
     tcccctcgtg tcacccacag tctaccccac acccatccat tctttcttta acctctgact       480
     cttcctcctt ggtttctcac tgccttggac gcttgttcac cccggatgaa ctccgtaggc       540
     gtctcccttc cctgcttggt accctaaggt gccctcggtg cttgttcgta gagacgaact       600
     ctgctctgtc cttgtgtcca gaaccaagcc ttcctctttt agggcacaag gctggccagc       660
     atcctgacag caggctggga gacctggatc ctccagatga cggacatcct tacttagggg       720
     tagcctctgg gatgaactag atattaaaaa ttaggtaacc ttggggcgtg cctgggcaga       780
     cctcaagtct ggtagcttca ggggctgctt ctccccagga ctacaccggg gcactcttct       840
     cttgtcctcc ccacccccca caagcttgtg ctaaacaact gctgtatacc aggctccatg       900
     cttgagcttc agaaacaccc tagggcagct gaatgtccac caggagtgtc cagagggagg       960
     gtgagcaccc caagagaaca gagtggccct agtaaatgct cagggaccac agaacttttg      1020
     cccactccac ttcctattgg taccccggc catgccccag aaatcagggc atgtttgtac      1080
     cctccccacg acagctcggg ccgcctggaa ctgacctgta gacagtgctc ctgggtagat      1140
     gctgcatttg aaaggtggca agagggctgg tgagatggct cagcggttag gagcactgac      1200
     tgctcttcca aaggtcctga gttcaaatcc cagcaaccac atggtggctc acaaccacct      1260
     acagctacag tgtacacaca tataataaaa taaataaaca aatcttaaaa aaaaaaaaaa      1320
     gaaagaaagg tggcaagagc caccatagtg gagaaggcag gtaggatccc caaggctaag      1380
     atgctaccga gtaaccatca gtgttcttct agccatagtg ggcaagacct agtgttccta      1440
     gtcaatgttg acctctccat acttgcctct cggctccatc ccacaccctt ccctccttac      1500
     cctaacaggt ctagactcca ggggcttcct gtttggccct tccctagctc aggagctggg      1560
     cgtgggctgc gtgctcatcc ggaagcaggg gaagctgccg ggccccacta tatcagcctc      1620
     ctatgctctg gagtatggga aggtaaggga gctgtgggta gaggaagggc agggtcttat      1680
     taccacggct accagtgcct aggagtaaat gtgggtgctc agagaggttg agacattggg      1740
     gtgaggttta caactcctga aatgctcagc ctcagaaatg ctccaggcta gggaggtggc      1800
     cacttgttag catctagact ctcttaacgc tacttcctgt ctgcaggctg agctggaaat      1860
     ccagaaagat gccttagaac ccgggcagag agtggtcatt gtggatgacc tcctggccac      1920
     tggaggtaaa gaaccagccc aagacaaaca ggcttcaaag ggccaggccc tgtctggggt      1980
     gctgactaaa caaagcgctt gaataccttc tctttctctg tcccttcccc ccccccccc       2040
     caggaaccat gtttgcagcc tgtgatctgc tgcaccagct acgggctgag gtggtggagt      2100
     gtgtgagcct ggtggagctg acctcgctga agggcaggga gaggctagga cctataccat      2160
     tcttctctct cctccagtat gactgagctg gctagatggt cacacccctg ctcacagcag      2220
     cagtaactgc gcggtggctc agccctgggc gcctaagtga cctttgtgag ctacctgctg      2280
     ccc                                                                   2283
//
```

Sequence

```
FT                   /translation="MSESELKLVARRIRSFPDFPIPGVLFRDISPLLKDPDSFRASIRL
FT                   LASHLKSTHSGKIDYIAGLDSRGFLFGPSLAQELGVGCVLIRKQGKLPGPTISASYALE
FT                   YGKAELEIQKDALEPGQRVVIVDDLLATGGTMFAACDLLHQLRAEVVECVSLVELTSLK
FT                   GRERLGPIPFFSLLQYD"
XX
SQ   Sequence 2283 BP; 485 A; 696 C; 590 G; 512 T; 0 other;
     cctgcggata ctcacctcct ccttgtctcc tacaagcacg cggccatgtc cgagtctgag        60
     ttgaaactgg tggcgcggcg catccgcagc ttccccgact tccccatccc gggcgtgctg       120
     ttcaggtgcg gtcacgagcc ggcgaggcgt tggcgccgta ctctcatccc ccggcgcagg       180
     cgcgtgggca gccttgggga tcttgcgggg cctctgcccg gccacacgcg gtcactctcc       240
     tgtccttgtt cccagggata tctcgcccct cttgaaagat ccggactcct tccgagcttc       300
     catccgcctc ctggccagtc acctgaagtc cacgcacagc ggcaagatcg actatatcgc       360
     agggcaaggt ggccttgcta ggccgtactc atcccccacg gtcctatccc ctatcccctt       420
     tcccctcgtg tcacccacag tctaccccac acccatccat tctttcttta acctctgact       480
     cttcctcctt ggtttctcac tgccttggac gcttgttcac cccggatgaa ctccgtaggc       540
     gtctcccttc cctgcttggt accctaaggt gccctcggtg cttgttcgta gagacgaact       600
     ctgctctgtc cttgtgtcca gaaccaagcc ttcctctttt agggcacaag gctggccagc       660
     atcctgacag caggctggga gacctggatc ctccagatga cggacatcct tacttagggg       720
     tagcctctgg gatgaactag atattaaaaa ttaggtaacc ttggggcgtg cctgggcaga       780
     cctcaagtct ggtagcttca ggggctgctt ctccccagga ctacaccggg gcactcttct       840
     cttgtcctcc ccacccccca caagcttgtg ctaaacaact gctgtatacc aggctccatg       900
     cttgagcttc agaaacaccc tagggcagct gaatgtccac caggagtccc cagagggagg       960
     gtgagcaccc caagagaaca gagtggccct agtaaatgct cagggtccac agaacttttg      1020
     cccactccac ttcctattgg taccccggc catgccccag aaatcagggc atgtttgtac      1080
     cctccccacg acagctcggg ccgcctggaa ctgacctgta cacagtgctc ctgggtagat      1140
     gctgcatttg aaaggtggca agagggctgg tgagatggct cagcggttag gagcactgac      1200
     tgctcttcca aaggtcctga gttcaaatcc cagcaacaac atggtggctc acaaccacct      1260
     acagctacag tgtacacaca tataataaaa taaatgaaca aatcttaaaa aaaaaaaaaa      1320
     gaaagaaagg tggcaagagc caccatagtg gaggaggcag gtaggatccc caaggctaag      1380
     atgctaccga gtaaccatca gtgttcttct agccatagtg ggcaagacct agtgttccta      1440
     gtcaatgttg acctctccat acttgcctct cggctccatc ccacaccctt ccctccttac      1500
     cctaacaggt ctagactcca ggggcttgct gtttggccct tccctagctc aggagctggg      1560
     cgtgggctgc gtgctcatcc ggaaccaggg gaagctgccg ggccccacta tatcagcctc      1620
     ctatgctctg gagtatggga aggtaaggga gctgtgggta gaggaagggc agggtcttat      1680
     taccacggct accagtgcct tggagtaaat gtgggtgctc agagaggttg agacattggg      1740
     gtgaggttta caactcctgg aatgctcagc ctcagaaatg ctccaggcta gggaggtggc      1800
     cacttgttag catctaggct ctcttaacgc tacttcctgt ctgcaggctg agctggaaat      1860
     ccagaaagat gccttagaac ccgggcagag agtggtcatt gtggatgacc tcctggccac      1920
     tggaggtaaa gaccagccc aagacaaaca ggcttcaaag ggccaggccc tgtctggggt      1980
     gctgactaaa caaagcgctt gaataccttc tctttctctg tcccttcccc ccccccccc      2040
     caggaaccat gtttgcagcc tgtgatctgc tgcaccagct acgggctgag gtggtggagt      2100
     gtgtgaggct ggtggagctg acctcgctga agggcaggga gaggctagga cctataccat      2160
     tcttctctct cctccagtat gactgagctg gctagatggt cacacccctg ctcacagcag      2220
     cagtaactgc gcggtggctc agccctgggc gcctaagtga cctttgtgag ctacctgctg      2280
     cc                                                                    2283
//
```

End of the entry, always //

FASTA / Pearson

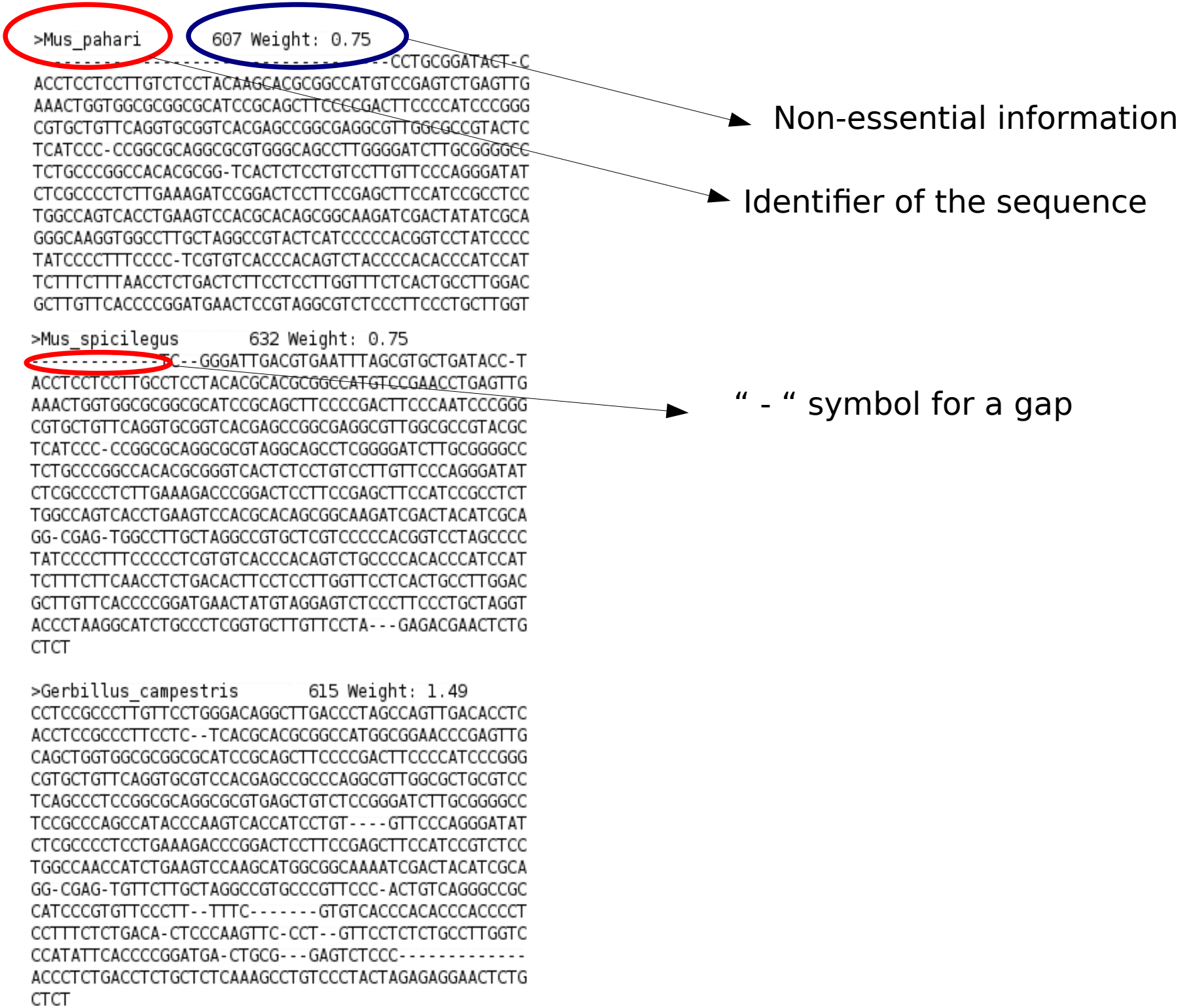
- Simplest format used
- No feature information is stored in this format
- Each sequence start with “>” and a sequence title
- Can be used for single sequences or multiple sequences (aligned or not)
- Information for homologous sites can be available (“-”).
- used by many software codes

Mus pahari, Mus spicilegus and Gerbillus campestris partial APRT gene sequences

```
>Mus_pahari       607 Weight: 0.75
-----------------------------------CCTGCGGATACT-C
ACCTCCTCCTTGTCTCCTACAAGCACGCGGCCATGTCCGAGTCTGAGTTG
AAACTGGTGGCGCGGCGCATCCGCAGCTTCCCCGACTTCCCCATCCCGGG
CGTGCTGTTCAGGTGCGGTCACGAGCCGGCGAGGCGTTGGCGCCGTACTC
TCATCCC-CCGGCGCAGGCGCGTGGGCAGCCTTGGGGATCTTGCGGGGCC
TCTGCCCGGCCACACGCGG-TCACTCTCCTGTCCTTGTTCCCAGGGATAT
CTCGCCCCTCTTGAAAGATCCGGACTCCTTCCGAGCTTCCATCCGCCTCC
TGGCCAGTCACCTGAAGTCCACGCACAGCGGCAAGATCGACTATATCGCA
GGGCAAGGTGGCCTTGCTAGGCCGTACTCATCCCCCACGGTCCTATCCCC
TATCCCCTTTCCCC-TCGTGTCACCCACAGTCTACCCCACACCCATCCAT
TCTTTCTTTAACCTCTGACTCTTCCTCCTTGGTTTCTCACTGCCTTGGAC
GCTTGTTCACCCCGGATGAACTCCGTAGGCGTCTCCCTTCCCTGCTTGGT

>Mus_spicilegus       632 Weight: 0.75
-------------TC--GGGATTGACGTGAATTTAGCGTGCTGATACC-T
ACCTCCTCCTTGCCTCCTACACGCACGCGGCCATGTCCGAACCTGAGTTG
AAACTGGTGGCGCGGCGCATCCGCAGCTTCCCCGACTTCCCAATCCCGGG
CGTGCTGTTCAGGTGCGGTCACGAGCCGGCGAGGCGTTGGCGCCGTACGC
TCATCCC-CCGGCGCAGGCGCGTAGGCAGCCTCGGGGATCTTGCGGGGCC
TCTGCCCGGCCACACGCGGGTCACTCTCCTGTCCTTGTTCCCAGGGATAT
CTCGCCCCTCTTGAAAGACCCGGACTCCTTCCGAGCTTCCATCCGCCTCT
TGGCCAGTCACCTGAAGTCCACGCACAGCGGCAAGATCGACTACATCGCA
GG-CGAG-TGGCCTTGCTAGGCCGTGCTCGTCCCCCACGGTCCTAGCCCC
TATCCCCTTTCCCCCTCGTGTCACCCACAGTCTGCCCCACACCCATCCAT
TCTTTCTTCAACCTCTGACACTTCCTCCTTGGTTCCTCACTGCCTTGGAC
GCTTGTTCACCCCGGATGAACTATGTAGGAGTCTCCCTTCCCTGCTAGGT
ACCCTAAGGCATCTGCCCTCGGTGCTTGTTCCTA---GAGACGAACTCTG
CTCT

>Gerbillus_campestris       615 Weight: 1.49
CCTCCGCCCTTGTTCCTGGGACAGGCTTGACCCTAGCCAGTTGACACCTC
ACCTCCGCCCTTCCTC--TCACGCACGCGGCCATGGCGGAACCCGAGTTG
CAGCTGGTGGCGCGGCGCATCCGCAGCTTCCCCGACTTCCCCATCCCGGG
CGTGCTGTTCAGGTGCGTCCACGAGCCGCCCAGGCGTTGGCGCTGCGTCC
TCAGCCCTCCGGCGCAGGCGCGTGAGCTGTCTCCGGGATCTTGCGGGGCC
TCCGCCCAGCCATACCCAAGTCACCATCCTGT----GTTCCCAGGGATAT
CTCGCCCCTCCTGAAAGACCCGGACTCCTTCCGAGCTTCCATCCGTCTCC
TGGCCAACCATCTGAAGTCCAAGCATGGCGGCAAAATCGACTACATCGCA
GG-CGAG-TGTTCTTGCTAGGCCGTGCCCGTTCCC-ACTGTCAGGGCCGC
CATCCCGTGTTCCCTT--TTTC-------GTGTCACCCACACCCACCCCT
CCTTTCTCTGACA-CTCCCAAGTTC-CCT--GTTCCTCTCTGCCTTGGTC
CCATATTCACCCCGGATGA-CTGCG---GAGTCTCCC-------------
ACCCTCTGACCTCTGCTCTCAAAGCCTGTCCCTACTAGAGAGGAACTCTG
CTCT
```

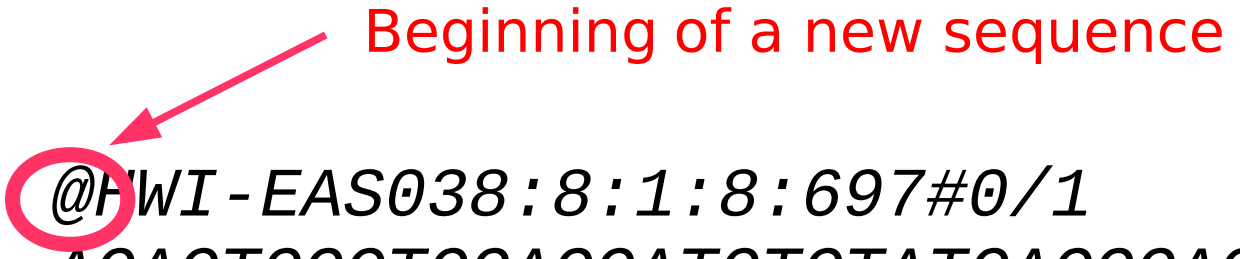

>Mus_pahari 607 Weight: 0.75
----------CCTGCGGATACT-C
ACCTCCTCCTTGTCTCCTACAAGCACGCGGCCATGTCCGAGTCTGAGTTG
AAACTGGTGGCGCGGCGCATCCGCAGCTTCCCCGACTTCCCCATCCCGGG
CGTGCTGTTCAGGTGCGGTCACGAGCCGGCGAGGCGTTGGCGCCGTACTC
TCATCCC-CCGGCGCAGGCGCGTGGGCAGCCTTGGGGATCTTGCGGGGCC
TCTGCCCGGCCACACGCGG-TCACTCTCCTGTCCTTGTTCCCAGGGATAT
CTCGCCCCTCTTGAAAGATCCGGACTCCTTCCGAGCTTCCATCCGCCTCC
TGGCCAGTCACCTGAAGTCCACGCACAGCGGCAAGATCGACTATATCGCA
GGGCAAGGTGGCCTTGCTAGGCCGTACTCATCCCCCACGGTCCTATCCCC
TATCCCCTTTCCCC-TCGTGTCACCCACAGTCTACCCCACACCCATCCAT
TCTTTCTTTAACCTCTGACTCTTCCTCCTTGGTTTCTCACTGCCTTGGAC
GCTTGTTCACCCCGGATGAACTCCGTAGGCGTCTCCCTTCCCTGCTTGGT
>Mus_spicilegus 632 Weight: 0.75
----------C--GGGATTGACGTGAATTTAGCGTGCTGATACC-T
ACCTCCTCCTTGCCTCCTACACGCACGCGGCCATGTCCGAACCTGAGTTG
AAACTGGTGGCGCGGCGCATCCGCAGCTTCCCCGACTTCCCAATCCCGGG
CGTGCTGTTCAGGTGCGGTCACGAGCCGGCGAGGCGTTGGCGCCGTACGC
TCATCCC-CCGGCGCAGGCGCGTAGGCAGCCTCGGGGATCTTGCGGGGCC
TCTGCCCGGCCACACGCGGGTCACTCTCCTGTCCTTGTTCCCAGGGATAT
CTCGCCCCTCTTGAAAGACCCGGACTCCTTCCGAGCTTCCATCCGCCTCT
TGGCCAGTCACCTGAAGTCCACGCACAGCGGCAAGATCGACTACATCGCA
GG-CGAG-TGGCCTTGCTAGGCCGTGCTCGTCCCCCACGGTCCTAGCCCC
TATCCCCTTTCCCCCTCGTGTCACCCACAGTCTGCCCCACACCCATCCAT
TCTTTCTTCAACCTCTGACACTTCCTCCTTGGTTCCTCACTGCCTTGGAC
GCTTGTTCACCCCGGATGAACTATGTAGGAGTCTCCCTTCCCTGCTAGGT
ACCCTAAGGCATCTGCCCTCGGTGCTTGTTCCTA---GAGACGAACTCTG
CTCT
>Gerbillus_campestris 615 Weight: 1.49
CCTCCGCCCTTGTTCCTGGGACAGGCTTGACCCTAGCCAGTTGACACCTC
ACCTCCGCCCTTCCTC--TCACGCACGCGGCCATGGCGGAACCCGAGTTG
CAGCTGGTGGCGCGGCGCATCCGCAGCTTCCCCGACTTCCCCATCCCGGG
CGTGCTGTTCAGGTGCGTCCACGAGCCGCCCAGGCGTTGGCGCTGCGTCC
TCAGCCCTCCGGCGCAGGCGCGTGAGCTGTCTCCGGGATCTTGCGGGGCC
TCCGCCCAGCCATACCCAAGTCACCATCCTGT----GTTCCCAGGGATAT
CTCGCCCCTCCTGAAAGACCCGGACTCCTTCCGAGCTTCCATCCGTCTCC
TGGCCAACCATCTGAAGTCCAAGCATGGCGGCAAAATCGACTACATCGCA
GG-CGAG-TGTTCTTGCTAGGCCGTGCCCGTTCCC-ACTGTCAGGGCCGC
CATCCCGTGTTCCCTT--TTTC-------GTGTCACCCACACCCACCCCT
CCTTTCTCTGACA-CTCCCAAGTTC-CCT--GTTCCTCTCTGCCTTGGTC
CCATATTCACCCCGGATGA-CTGCG---GAGTCTCCC------------
ACCCTCTGACCTCTGCTCTCAAAGCCTGTCCCTACTAGAGAGGAACTCTG
CTCT
Non-essential information
Identifier of the sequence
“ - “ symbol for a gap

FASTQ

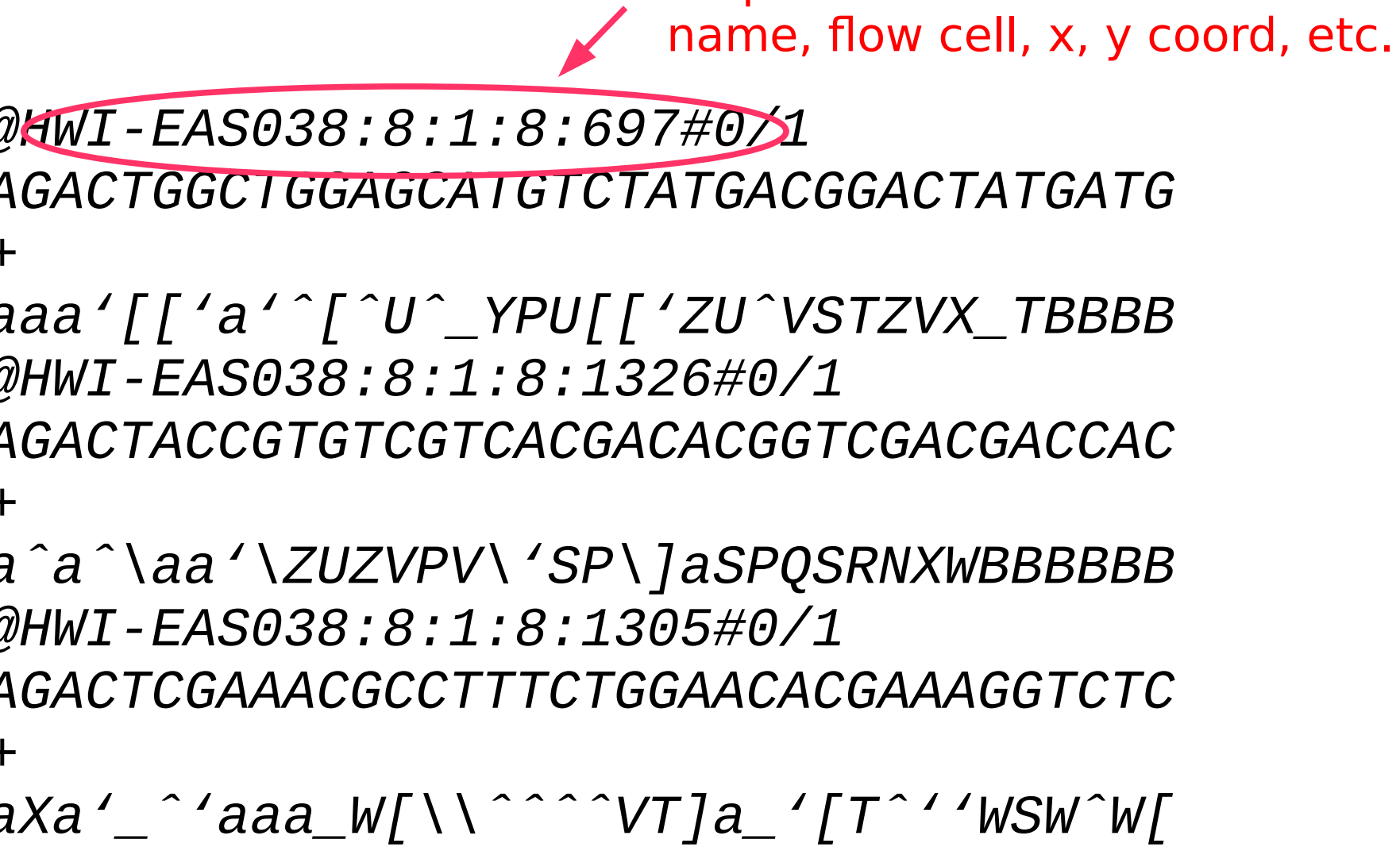
- Format adapted for high-throughput short reads
- Only sequence and quality information is stored in this format
- Each sequence start with “@” and a sequence title (usually machine generated)
- Sequences are on one line
- Third line starts with “+”
- Fourth line stores base quality score

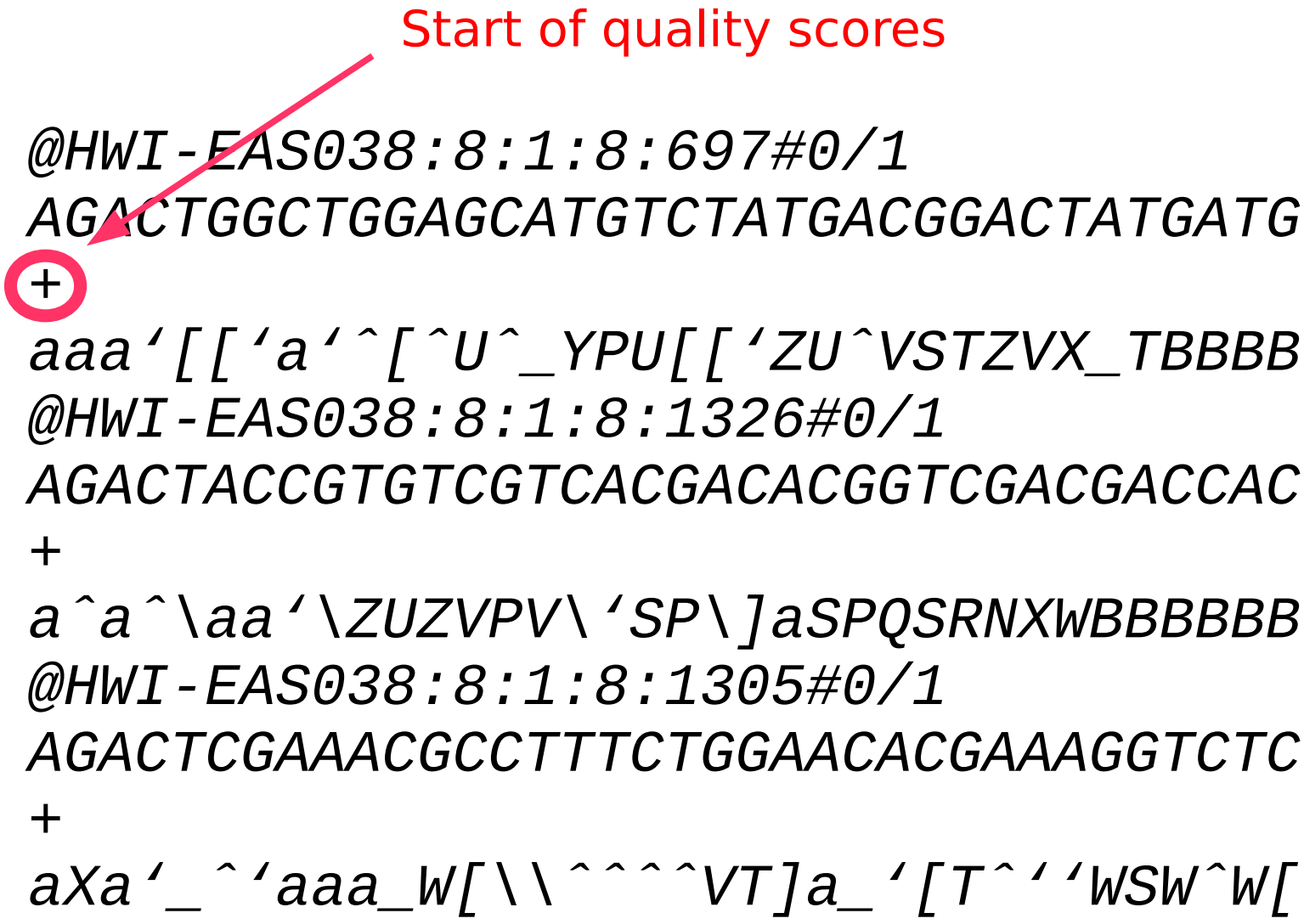
(Phred = ord(Q) - 33)

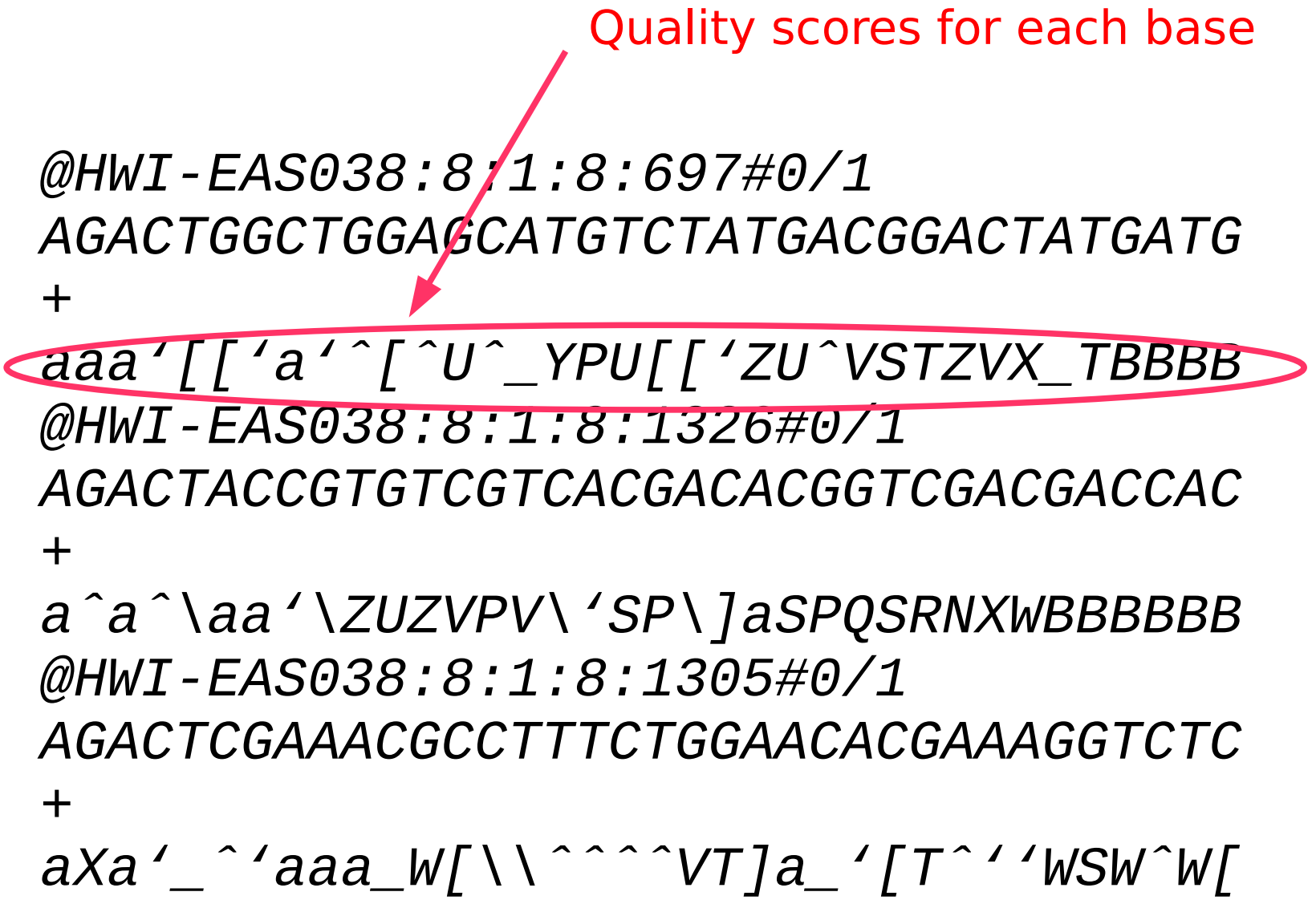
```
@HWI-EAS038:8:1:8:697#0/1
AGACTGGCTGGAGCATGTCTATGACGGACTATGATG
+
aaa‘[[‘a‘^[^U^_YPU[[‘ZU^VSTZVX_TBBBB
@HWI-EAS038:8:1:8:1326#0/1
AGACTACCGTGTCGTCACGACACGGTCGACGACCAC
+
a^a^\aa‘\ZUZVPV\‘SP\]aSPQSRNXWBBBBBB
@HWI-EAS038:8:1:8:1305#0/1
AGACTCGAAACGCCTTTCTGGAACACGAAAGGTCTC
+
aXa‘_^‘aaa_W[\\^^^VT]a_‘[T^‘‘WSW^W[
```




```
@HWI-EAS038:8:1:8:697#0/1
AGACTGGCTGGAGCATGTCTATGACGGACTATGATG
+
aaa‘[[‘a‘^[^U^_YPU[[‘ZU^VSTZVX_TBBBB
@HWI-EAS038:8:1:8:1326#0/1
AGACTACCGTGTCGTCACGACACGGTCGACGACCAC
+
a^a^\aa‘\ZUZVPV\‘SP\]aSPQSRNXWBBBBBB
@HWI-EAS038:8:1:8:1305#0/1
AGACTCGAAACGCCTTTCTGGAACACGAAAGGTCTC
+
aXa‘_^‘aaa_W[\\^^^VT]a_‘[T^‘‘WSW^W[
```


Sequence title: often machine name, flow cell, x, y coord, etc.
@HWI-EAS038:8:1:8:697#0/1
AGACTGGCTGGAGCATGTCTATGACGGACTATGATG
+
aaa‘[[‘a‘^[^U^_YPU[[‘ZU^VSTZVX_TBBBB
@HWI-EAS038:8:1:8:1326#0/1
AGACTACCGTGTCGTCACGACACGGTCGACGACCAC
+
a^a^\aa‘\ZUZVPV\‘SP\]aSPQSRNXWBBBBBB
@HWI-EAS038:8:1:8:1305#0/1
AGACTCGAAACGCCTTTCTGGAACACGAAAGGTCTC
+
aXa‘_^‘aaa_W[\\^^^VT]a_‘[T^‘‘WSW^W[


Start of quality scores
@HWI-EAS038:8:1:8:697#0/1
AGACTGGCTGGAGCATGTCTATGACGGACTATGATG
+
aaa‘[[‘a‘^[^U^_YPU[[‘ZU^VSTZVX_TBBBB
@HWI-EAS038:8:1:8:1326#0/1
AGACTACCGTGTCGTCACGACACGGTCGACGACCAC
+
a^a^\aa‘\ZUZVPV\‘SP\]aSPQSRNXWBBBBBB
@HWI-EAS038:8:1:8:1305#0/1
AGACTCGAAACGCCTTTCTGGAACACGAAAGGTCTC
+
aXa‘_^‘aaa_W[\\^^^VT]a_‘[T^‘‘WSW^W[


Quality scores for each base
@HWI-EAS038:8:1:8:697#0/1
AGACTGGCTGGAGCATGTCTATGACGGACTATGATG
+
aaa‘[[‘a‘^[^U^_YPU[[‘ZU^VSTZVX_TBBBB
@HWI-EAS038:8:1:8:1326#0/1
AGACTACCGTGTCGTCACGACACGGTCGACGACCAC
+
a^a^\aa‘\ZUZVPV\‘SP\]aSPQSRNXWBBBBBB
@HWI-EAS038:8:1:8:1305#0/1
AGACTCGAAACGCCTTTCTGGAACACGAAAGGTCTC
+
aXa‘_^‘aaa_W[\\^^^VT]a_‘[T^‘‘WSW^W[

```
@HWI-EAS038:8:1:8:697#0/1
AGACTGGCTGGAGCATGTCTATGACGGACTATGATG
+
aaa‘[[‘a‘^[^U^_YPU[[‘ZU^VSTZVX_TBBBB
@HWI-EAS038:8:1:8:1326#0/1
AGACTACCGTGTCGTCACGACACGGTCGACGACCAC
+
a^a^\aa‘\ZUZVPV\‘SP\]aSPQSRNXWBBBBBB
@HWI-EAS038:8:1:8:1305#0/1
AGACTCGAAACGCCTTTCTGGAACACGAAAGGTCTC
+
aXa‘_^‘aaa_W[\\^^^VT]a_‘[T^‘‘WSW^W[
```

ASCII Table

	30	40	50	60	70	80	90	100	110	120
0		(	2	<	F	P	Z	d	n	x
1		)	3	=	G	Q	[	e	o	y
2		*	4	>	H	R	\	f	p	z
3	!	+	5	?	I	S	]	g	q	{
4	″	,	6	@	J	T	^	h	r	|
5	#	-	7	A	K	U	_	i	s	}
6	$	.	8	B	L	V	`	j	t	~
7	%	/	9	C	M	W	a	k	u	DEL
8	&	0	:	D	N	X	b	l	v	
9	′	1	;	E	O	Y	c	m	w	

American Standard Code for Information Interchange

Fastq-sanger $Q = ord(q) - 33$

Fastq-solexa $Q = 10*\log_{10}(1 + 10^{((ord(q)-64)/10)})$

Fastq-illumina $Q = ord(q) - 64$

As of CASAVA 1.8, the Illumina FASTQ variant use 33-offset quality encoding (ASCII ‘!’ = 0) and have a stylized format:

@<instrument>:<run number>:<flowcell ID>:<lane>:<tile>:<x-pos>:<y-pos>
<read>:<is filtered>:<control number>:<index sequence>

Specific example:

```
@EAS139:136:FC706VJ:2:5:1000:12850 1:Y:18:ATCACG
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
+
BBBBCCCC?<A?BC?7@@???????DBBA@@@@A@@
```

Example: Casava 1.8

If q = `?` then ord(q) = 63

Fastq-sanger = 63 – 33 = 30

Fastq-solexa = $10*\log_{10}(1+10^{-0.1}) = 2$

Fastq-illumina = 63 - 64 = “undefined”

Example: older Illumina

If q = `a` then ord(q) = 97

Fastq-sanger = 97 – 33 = “undefined”

Fastq-solexa = $10*\log_{10}(1+10^{3.3}) = 33$

Fastq-illumina = 97 - 64 = 33

GFF file format; General feature format

Example:

```
##gff-version 3
#!gff-spec-version 1.20
#!processor NCBI annotwriter
##sequence-region NC_004354.3 1 22422827
##species http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=7227
NC_004354.3     RefSeq  region  1       22422827        .       +       .
ID=id0;Dbxref=taxon:7227;chromosome=X;gbkey=Src;genome=chromosome;genotype=y%5B1%5D%3B cn%5B1%5D bw
%5B1%5D sp%5B1%5D%3B Rh6%5B1%5D;mol_type=genomic DNA;old-lineage=Eukaryota%3B Metazoa%3B Arthropoda
%3B Hexapoda%3B Insecta%3B Pterygota%3B Neoptera%3B Endopterygota%3B Diptera%3B Brachycera%3B
Muscomorpha%3B Ephydroidea%3B Drosophilidae%3B Drosophila%3B Sophophora

NC_004354.3     RefSeq  region  1       4229    .       +       .   \
ID=id1;Dbxref=FLYBASE:FBti0102096;gbkey=mobile_element
NC_004354.3     RefSeq  sequence_variant        148     148     .       +       .   \
ID=id2;Dbxref=dbSNP:207064200;gbkey=variation


NC_004354.3     RefSeq  region  10396   10433   .       +       .       \
ID=id888;Dbxref=FLYBASE:FBti0063564;gbkey=mobile_element
```

GFF file format; General feature format

Example:

```
NC_004354.3     RefSeq  region  10396   10433  \
 .       +       .       \
ID=id888;Dbxref=FLYBASE:FBti0063564;gbkey=mobile_element
```

A tab delimited file.

#1 “Name” sequence, #2 Source (program/database), #3 name of feature (gene/exon), #4 start, #5 end, #6 confidence value ('.' undefined), #7 strand, #8 frame/phase, #9 description

SAM/BAM format

```
@SQ SN:NC_009456 LN:1108250
@SQ SN:NC_009457 LN:3024069

HWUSI-EAS1786:60:FC62MTAAXX:1:1:5522:1000 0 NC_009456
267580 25 32M * 0 0
NGGACGGGTTGGATGCCGACGATCTCAATTTG
#111166446@@@@C@@CCCCC@@C@@@CCCC
XT:A:U NM:i:2 X0:i:1 X1:i:0 XM:i:2 XO:i:0 XG:i:0 MD:Z:0T0T30

HWUSI-EAS1786:60:FC62MTAAXX:1:1:1955:1000 0 NC_009456
875747 0 30M * 0 0
NCATCGCGGCTCAATGGGACTGGAAACGCC
#223054535@@@@@@@@@@@@@CCCC@8CC
XT:A:R NM:i:1 X0:i:2 X1:i:44 XM:i:1 XO:i:0 XG:i:0 MD:Z:0C29

HWUSI-EAS1786:60:FC62MTAAXX:1:1:9547:1001 16 NC_009457
2485619 0 48M * 0 0
ATCCAGCGCCAGGTTCCCCTAGCGCTACCTTGTTACGACTTCCCCCCN
@?@C@@C@@@@@CCC@C@@@@@@@@@@@@@@@@@@@@@@@44003//30#
 XT:A:R NM:i:2 X0:i:8 X1:i:0 XM:i:2 XO:i:0 XG:i:0 MD:Z:42A4A0
```

SAM/BAM format

```
@SQ SN:NC_009456 LN:1108250
@SQ SN:NC_009457 LN:3024069
HWUSI-EAS1786:60:FC62MTAAXX:1:1:5522:1000 0 NC_009456
267580 25 32M * 0 0
NGGACGGGTTGGATGCCGACGATCTCAATTTG
#111166446@@@@C@@CCCCC@@C@@@CCCC
XT:A:U NM:i:2 X0:i:1 X1:i:0 XM:i:2 XO:i:0 XG:i:0 MD:Z:0T0T30

HWUSI-EAS1786:60:FC62MTAAXX:1:1:1955:1000 0 NC_009456
875747 0 30M * 0 0
NCATCGCGGCTCAATGGGACTGGAAACGCC
#223054535@@@@@@@@@@@@@CCCC@8CC
XT:A:R NM:i:1 X0:i:2 X1:i:44 XM:i:1 XO:i:0 XG:i:0 MD:Z:0C29

HWUSI-EAS1786:60:FC62MTAAXX:1:1:9547:1001 16 NC_009457
2485619 0 48M * 0 0
ATCCAGCGCCAGGTTCCCCTAGCGCTACCTTGTTACGACTTCCCCCCN
@?@C@@C@@@@@CCC@C@@@@@@@@@@@@@@@@@@@@@@@@44003//30#
XT:A:R NM:i:2 X0:i:8 X1:i:0 XM:i:2 XO:i:0 XG:i:0 MD:Z:42A4A0
```

#1

#2

#3

#4-#9

#12

Tab-delimited

1. **QNAME Query template/pair NAME**

2. **FLAG bitwise FLAG**

3. **RNAME Reference sequence NAME**

4. **POS The left most coordinate of the read using the number of the sequence in the reference genome.**

5. **MAPQ MAPping Quality**
   The map quality is Phred-scaled. A value of 255 is used for an unknown map quality.

6. **CIAGR extended CIGAR string**
   This string describes features of the match between the read and the reference sequence. In the cases above it is '[0-9M' indicating a perfect match for the length of the read. The format is a number followed by a letter. The number indicates the number of bases and the letter designates a category; M for match, I for an insert in the read, D for a deletion in the read, N for a region skipped, etc.

7. **MRNM Mate Reference sequence NaMe**
   In the cases above it is '*' meaning that there is no mate; these were unpaired reads.

8. **MPOS Mate POSistion**
   The bp location in the reference genome where the leftmost bp of the mate read maps.

9. **TLEN inferred Template LENgth**
   The length of the insert between mate pairs.

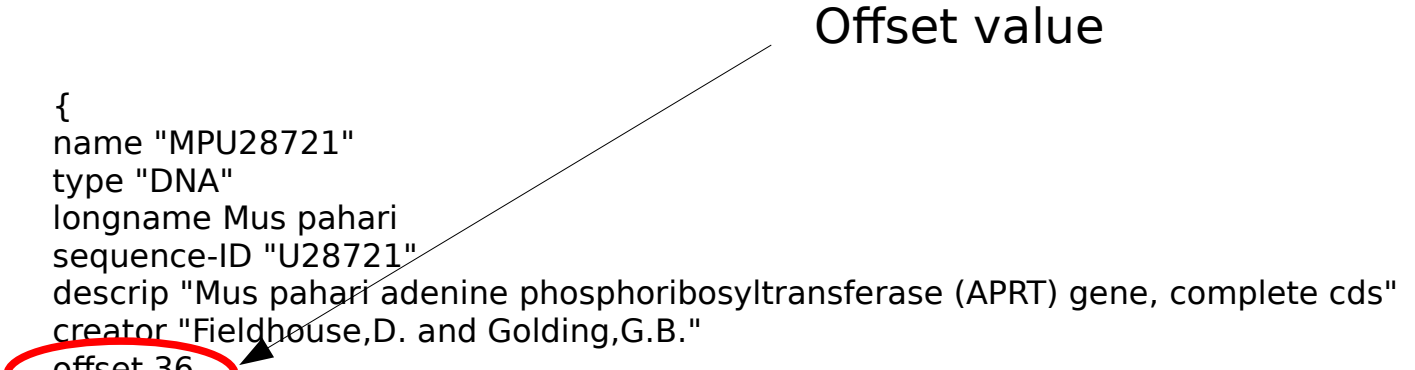
10. **SEQ query SEQuence**
    The sequence of the read.

11. **QUAL query QUALity**
    The quality is given is ord(ASCII)-33 (Sanger Phred scores).

12. **OPT variable OPTional fields**

GDE

- Tagged file format storing all the information about a sequence (similar to GenBank format)
- Can contain alignment information
- Text enclosed in “{}”
- All tagged values are enclosed in “”




```
{
name "MPU28721"
type "DNA"
longname Mus pahari
sequence-ID "U28721"
descrip "Mus pahari adenine phosphoribosyltransferase (APRT) gene, complete cds"
creator "Fieldhouse,D. and Golding,G.B."
offset 36
creation-date      1/31/98 14:18:24
direction 1
strandedness 1
comments "
NID   g881573
KEYWORDS.
SOURCE             shrew mouse.
                   Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
                   Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
                   Murinae; Mus.
REFERENCE          1      (bases 1 to 2283)
TITLE              Rates of substitution in closely related rodent species
JOURNAL            Unpublished
REFERENCE          2      (bases 1 to 2283)
TITLE Direct       Submission
JOURNAL            Submitted (07-JUN-1995) Dan Fieldhouse, Biology, McMaster
                   University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada
FEATURES                   Location/Qualifiers
      source               1..2283
                           /organism=?Mus pahari?
                           /db_xref=?taxon:10093?
      gene                 join(46..125,256..362,1509..1642,1847..1925,2044..2186)
                           /gene=?APRT?
```

```
     CDS             join(46..125,256..362,1509..1642,1847..1925,2044..2186)
                     /gene=?APRT?
                     /EC_number=?2.4.2.7?
                     /note=?purine salvage enzyme?
                     /codon_start=1
                     /product=?adenine phosphoribosyltransferase?
                     /db_xref=?PID:g881574?
                     /translation=?MSESELKLVARRIRSFPDFPIPGVLFRDISPLLKDPDSFRASIR
                     LLASHLKSTHSGKIDYIAGLDSRGFLFGPSLAQELGVGCVLIRKQGKLPGPTISASYA
                     LEYGKAELEIQKDALEPGQRVVIVDDLLATGGTMFAACDLLHQLRAEVVECVSLVELT
                     SLKGRERLGPIPFFSLLQYD?
BASE COUNT       485 a        696 c        590 g        512 t
"
sequence "CCTGCGGATACTCACCTCCTCCTT
GTCTCCTACAAGCACGCGGCCATGTCCGAGTCTGAGTTGAAACTGGTGGCGCGGCGCATC
CGCAGCTTCCCCGACTTCCCCATCCCGGGCGTGCTGTTCAGGTGCGGTCACGAGCCGGCG
AGGCGTTGGCGCCGTACTCTCATCCC-CCGGCGCAGGCGCGTGGGCAGCCTTGGGGATCT
TGCGGGGCCTCTGCCCGGCCACACGCGG-TCACTCTCCTGTCCTTGTTCCCAGGGATATC
TCGCCCCTCTTGAAAGATCCGGACTCCTTCCGAGCTTCCATCCGCCTCCTGGCCAGTCAC
CTGAAGTCCACGCACAGCGGCAAGATCGACTATATCGCAGGGCAAGGTGGCCTTGCTAGG
CCGTACTCATCCCCCACGGTCCTATCCCCTATCCCCTTTCCCC-TCGTGTCACCCACAGT
CTACCCCACACCCATCCATTCTTTCTTTAACCTCTGACTCTTCCTCCTTGGTTTCTCACT
GCCTTGGACGCTTGTTCACCCCGGATGAACTCCGTAGGCGTCTCCCTTCCCTGCTTGGTA
CCCTAAGG----TGCCCTCGGTGCTTGTTCGTAGAGACGAACTCTGCTCT"
}
```

Phylip

- Used by a large number of programs (PHYLIP)
- Two different formats
  - sequential format
  - interleaved : 2 numbers on the first line (number of sequences, sequence length)
- No more than 10 characters for the names of the sequences (warning)
- Can store alignment information

Interleaved format


```
 3   650
Mus_pahari ---------- ---------- ---------- ------CCTG CGGATACT-C ACCTCCTCCT
Mus_spicil ---------- ---TC--GGG ATTGACGTGA ATTTAGCGTG CTGATACC-T ACCTCCTCCT
Gerbillus_ CCTCCGCCCT TGTTCCTGGG ACAGGCTTGA CCCTAGCCAG TTGACACCTC ACCTCCGCCC

           TGTCTCCTAC AAGCACGCGG CCATGTCCGA GTCTGAGTTG AAACTGGTGG CGCGGCGCAT
           TGCCTCCTAC ACGCACGCGG CCATGTCCGA ACCTGAGTTG AAACTGGTGG CGCGGCGCAT
           TTCCTC--TC ACGCACGCGG CCATGGCGGA ACCCGAGTTG CAGCTGGTGG CGCGGCGCAT

           CCGCAGCTTC CCCGACTTCC CCATCCCGGG CGTGCTGTTC AGGTGCGGTC ACGAGCCGGC
           CCGCAGCTTC CCCGACTTCC CAATCCCGGG CGTGCTGTTC AGGTGCGGTC ACGAGCCGGC
           CCGCAGCTTC CCCGACTTCC CCATCCCGGG CGTGCTGTTC AGGTGCGTCC ACGAGCCGCC

           GAGGCGTTGG CGCCGTACTC TCATCCC-CC GGCGCAGGCG CGTGGGCAGC CTTGGGGATC
           GAGGCGTTGG CGCCGTACGC TCATCCC-CC GGCGCAGGCG CGTAGGCAGC CTCGGGGATC
           CAGGCGTTGG CGCTGCGTCC TCAGCCCTCC GGCGCAGGCG CGTGAGCTGT CTCCGGGATC

           TTGCGGGGCC TCTGCCCGGC CACACGCGG- TCACTCTCCT GTCCTTGTTC CCAGGGATAT
           TTGCGGGGCC TCTGCCCGGC CACACGCGGG TCACTCTCCT GTCCTTGTTC CCAGGGATAT
           TTGCGGGGCC TCCGCCCAGC CATACCCAAG TCACCATCCT GT----GTTC CCAGGGATAT

           CTCGCCCCTC TTGAAAGATC CGGACTCCTT CCGAGCTTCC ATCCGCCTCC TGGCCAGTCA
           CTCGCCCCTC TTGAAAGACC CGGACTCCTT CCGAGCTTCC ATCCGCCTCT TGGCCAGTCA
           CTCGCCCCTC CTGAAAGACC CGGACTCCTT CCGAGCTTCC ATCCGTCTCC TGGCCAACCA

           CCTGAAGTCC ACGCACAGCG GCAAGATCGA CTATATCGCA GGGCAAGGTG GCCTTGCTAG
           CCTGAAGTCC ACGCACAGCG GCAAGATCGA CTACATCGCA GG-CGAG-TG GCCTTGCTAG
           TCTGAAGTCC AAGCATGGCG GCAAAATCGA CTACATCGCA GG-CGAG-TG TTCTTGCTAG
```

Sequential format

```
3 650
Mus_pahari
-----------------------------------CCTGCGGATACT-CACCTCCTCCT
TGTCTCCTACAAGCACGCGGCCATGTCCGAGTCTGAGTTGAAACTGGTGGCGCGGCGCAT
CCGCAGCTTCCCCGACTTCCCCATCCCGGGCGTGCTGTTCAGGTGCGGTCACGAGCCGGC
GAGGCGTTGGCGCCGTACTCTCATCCC-CCGGCGCAGGCGCGTGGGCAGCCTTGGGGATC
TTGCGGGGCCTCTGCCCGGCCACACGCGG-TCACTCTCCTGTCCTTGTTCCCAGGGATAT
CTCGCCCCTCTTGAAAGATCCGGACTCCTTCCGAGCTTCCATCCGCCTCCTGGCCAGTCA
CCTGAAGTCCACGCACAGCGGCAAGATCGACTATATCGCAGGGCAAGGTGGCCTTGCTAG
GCCGTACTCATCCCCCACGGTCCTATCCCCTATCCCCTTTCCCC-TCGTGTCACCCACAG
TCTACCCCACACCCATCCATTCTTTCTTTAACCTCTGACTCTTCCTCCTTGGTTTCTCAC
TGCCTTGGACGCTTGTTCACCCCGGATGAACTCCGTAGGCGTCTCCCTTCCCTGCTTGGT
ACCCTAAGG----TGCCCTCGGTGCTTGTTCGTA---GAGACGAACTCTG
Mus_spicil
-------------TC--GGGATTGACGTGAATTTAGCGTGCTGATACC-TACCTCCTCCT
TGCCTCCTACACGCACGCGGCCATGTCCGAACCTGAGTTGAAACTGGTGGCGCGGCGCAT
CCGCAGCTTCCCCGACTTCCCAATCCCGGGCGTGCTGTTCAGGTGCGGTCACGAGCCGGC
GAGGCGTTGGCGCCGTACGCTCATCCC-CCGGCGCAGGCGCGTAGGCAGCCTCGGGGATC
TTGCGGGGCCTCTGCCCGGCCACACGCGGGTCACTCTCCTGTCCTTGTTCCCAGGGATAT
CTCGCCCCTCTTGAAAGACCCGGACTCCTTCCGAGCTTCCATCCGCCTCTTGGCCAGTCA
CCTGAAGTCCACGCACAGCGGCAAGATCGACTACATCGCAGG-CGAG-TGGCCTTGCTAG
GCCGTGCTCGTCCCCCACGGTCCTAGCCCCTATCCCCTTTCCCCCTCGTGTCACCCACAG
TCTGCCCCACACCCATCCATTCTTTCTTCAACCTCTGACACTTCCTCCTTGGTTCCTCAC
TGCCTTGGACGCTTGTTCACCCCGGATGAACTATGTAGGAGTCTCCCTTCCCTGCTAGGT
ACCCTAAGGCATCTGCCCTCGGTGCTTGTTCCTA---GAGACGAACTCTG
Gerbillus_
CCTCCGCCCTTGTTCCTGGGACAGGCTTGACCCTAGCCAGTTGACACCTCACCTCCGCCC
TTCCTC--TCACGCACGCGGCCATGGCGGAACCCGAGTTGCAGCTGGTGGCGCGGCGCAT
CCGCAGCTTCCCCGACTTCCCCATCCCGGGCGTGCTGTTCAGGTGCGTCCACGAGCCGCC
CAGGCGTTGGCGCTGCGTCCTCAGCCCTCCGGCGCAGGCGCGTGAGCTGTCTCCGGGATC
TTGCGGGGCCTCCGCCCAGCCATACCCAAGTCACCATCCTGT----GTTCCCAGGGATAT
CTCGCCCCTCCTGAAAGACCCGGACTCCTTCCGAGCTTCCATCCGTCTCCTGGCCAACCA
TCTGAAGTCCAAGCATGGCGGCAAAATCGACTACATCGCAGG-CGAG-TGTTCTTGCTAG
GCCGTGCCCGTTCCC-ACTGTCAGGGCCGCCATCCCGTGTTCCCTT--TTTC-------G
TGTCACCCACACCCACCCCTCCTTTCTCTGACA-CTCCCAAGTTC-CCT--GTTCCTCTC
TGCCTTGGTCCCATATTCACCCCGGATGA-CTGCG---GAGTCTCCC------------
ACCCTCTGACCTCTGCTCTCAAAGCCTGTCCCTACTAGAGAGGAACTCTG
```

NEXUS

- Madison et al. 1997. Syst. Biol. 46: 590-621
- Format used with PAUP, McClade and Mr. Bayes
- Composed of different modules
  - starting with “ BEGIN XXXXX:”
  - ending with “END;”
- Standard blocks: TAXA, CHARACTERS, TREE
- Comments are enclosed within “[]”

```
#NEXUS

[Name: Mus_pahar、       Len:   680  Check: 70E718C]
[Name: Mus_spici        Len:   680  Check: D5E622FB]
[Name: Gerbillus        Len:   680  Check: FBE40A58]


BEGIN TAXA;
 DIMENSIONS NTAX=3;
 TAXLABELS Mus_pahar Mus_spici Gerbillus;
END;

BEGIN CHARACTERS;
 DIMENSIONS NCHAR=680;
 FORMAT MISSING=? DATATYPE=DNA INTERLEAVE GAP=-;
  MATRIX

Mus_pahari            -------------------- ----------------CCTG CGGATACT-CACCTCCTCCT TGTCTCCTACAAGCACGCGG CCATGTCCGAGTCTGAGTTG
Mus_spicilegus        -------------TC--GGG ATTGACGTGAATTTAGCGTG CTGATACC-TACCTCCTCCT TGCCTCCTACACGCACGCGG CCATGTCCGAACCTGAGTTG
Gerbillus_campestris  CCTCCGCCCTTGTTCCTGGG ACAGGCTTGACCCTAGCCAG TTGACACCTCACCTCCGCCC TTCCTC--TCACGCACGCGG CCATGGCGGAACCCGAGTTG

Mus_pahari            AAACTGGTGGCGCGGCGCAT CCGCAGCTTCCCCGACTTCC CCATCCCGGGCGTGCTGTTC AGGTGCGGTCACGAGCCGGC GAGGCGTTGGCGCCGTACTC
Mus_spicilegus        AAACTGGTGGCGCGGCGCAT CCGCAGCTTCCCCGACTTCC CAATCCCGGGCGTGCTGTTC AGGTGCGGTCACGAGCCGGC GAGGCGTTGGCGCCGTACGC
Gerbillus_campestris  CAGCTGGTGGCGCGGCGCAT CCGCAGCTTCCCCGACTTCC CCATCCCGGGCGTGCTGTTC AGGTGCGGTCACGAGCCGGC GAGGCGTTGGCGCCGTACGC

Mus_pahari            TCATCCC-CCGGCGCAGGCG CGTGGGCAGCCTTGGGGATC TTGCGGGGCCTCTGCCCGGC CACACGCGG-TCACTCTCCT GTCCTTGTTCCCAGGGATAT
Mus_spicilegus        TCATCCC-CCGGCGCAGGCG CGTAGGCAGCCTCGGGGATC TTGCGGGGCCTCTGCCCGGC CACACGCGGGTCACTCTCCT GTCCTTGTTCCCAGGGATAT
Gerbillus_campestris  TCAGCCCTCCGGCGCAGGCG CGTGAGCTGTCTCCGGGATC TTGCGGGGCCTCCGCCCAGC CATACCCAAGTCACCATCCT GT----GTTCCCAGGGATAT

Mus_pahari            CTCGCCCCTCTTGAAAGATC CGGACTCCTTCCGAGCTTCC ATCCGCCTCCTGGCCAGTCA CCTGAAGTCCACGCACAGCG GCAAGATCGACTATATCGCA
Mus_spicilegus        CTCGCCCCTCTTGAAAGACC CGGACTCCTTCCGAGCTTCC ATCCGCCTCTTGGCCAGTCA CCTGAAGTCCACGCACAGCG GCAAGATCGACTACATCGCA
Gerbillus_campestris  CTCGCCCCTCCTGAAAGACC CGGACTCCTTCCGAGCTTCC ATCCGTCTCCTGGCCAACCA TCTGAAGTCCAAGCATGGCG GCAAAATCGACTACATCGCA

Mus_pahari            GGGCAAGGTGGCCTTGCTAG GCCGTACTCATCCCCCACGG TCCTATCCCCTATCCCCTTT CCCC-TCGTGTCACCCACAG TCTACCCCACACCCATCCAT
Mus_spicilegus        GG-CGAG-TGGCCTTGCTAG GCCGTGCTCGTCCCCCACGG TCCTAGCCCCTATCCCCTTT CCCCCTCGTGTCACCCACAG TCTGCCCCACACCCATCCAT
Gerbillus_campestris  GG-CGAG-TGTTCTTGCTAG GCCGTGCCCGTTCCC-ACTG TCAGGGCCGCCATCCCGTGT TCCCTT--TTTC-------G TGTCACCCACACCCACCCCT
```

File identifier

```
#NEXUS

[Name: Mus_pahar.       Len:   680  Check: 70E718C]
[Name: Mus_spici        Len:   680  Check: D5E622FB]
[Name: Gerbillus        Len:   680  Check: FBE40A58]
```

Sequences information

```
BEGIN TAXA;
 DIMENSIONS NTAX=3;
 TAXLABELS Mus_pahar Mus_spici Gerbillus;
END;
```

Block of information

```
BEGIN CHARACTERS;
 DIMENSIONS NCHAR=680;
 FORMAT MISSING=? DATATYPE=DNA INTERLEAVE GAP=-;
  MATRIX

Mus_pahari            ---------- ----------CCTG CGGATACT-CACCTCCTCCT TGTCTCCTACAAGCACGCGG CCATGTCCGAGTCTGAGTTG
Mus_spicilegus        ----------TC--GGG ATTGACGTGAATTTAGCGTG CTGATACC-TACCTCCTCCT TGCCTCCTACACGCACGCGG CCATGTCCGAACCTGAGTTG
Gerbillus_campestris  CCTCCGCCCTTGTTCCTGGG ACAGGCTTGACCCTAGCCAG TTGACACCTCACCTCCGCCC TTCCTC--TCACGCACGCGG CCATGGCGGAACCCGAGTTG

Mus_pahari            AAACTGGTGGCGCGGCGCAT CCGCAGCTTCCCCGACTTCC CCATCCCGGGCGTGCTGTTC AGGTGCGGTCACGAGCCGGC GAGGCGTTGGCGCCGTACTC
Mus_spicilegus        AAACTGGTGGCGCGGCGCAT CCGCAGCTTCCCCGACTTCC CAATCCCGGGCGTGCTGTTC AGGTGCGGTCACGAGCCGGC GAGGCGTTGGCGCCGTACGC
Gerbillus_campestris  CAGCTGGTGGCGCGGCGCAT CCGCAGCTTCCCCGACTTCC CCATCCCGGGCGTGCTGTTC AGGTGCGGTCACGAGCCGGC GAGGCGTTGGCGCCGTACGC

Mus_pahari            TCATCCC-CCGGCGCAGGCG CGTGGGCAGCCTTGGGGATC TTGCGGGGCCTCTGCCCGGC CACACGCGG-TCACTCTCCT GTCCTTGTTCCCAGGGATAT
Mus_spicilegus        TCATCCC-CCGGCGCAGGCG CGTAGGCAGCCTCGGGGATC TTGCGGGGCCTCTGCCCGGC CACACGCGGGTCACTCTCCT GTCCTTGTTCCCAGGGATAT
Gerbillus_campestris  TCAGCCCTCCGGCGCAGGCG CGTGAGCTGTCTCCGGGATC TTGCGGGGCCTCCGCCCAGC CATACCCAAGTCACCATCCT GT----GTTCCCAGGGATAT

Mus_pahari            CTCGCCCCTCTTGAAAGATC CGGACTCCTTCCGAGCTTCC ATCCGCCTCCTGGCCAGTCA CCTGAAGTCCACGCACAGCG GCAAGATCGACTATATCGCA
Mus_spicilegus        CTCGCCCCTCTTGAAAGACC CGGACTCCTTCCGAGCTTCC ATCCGCCTCTTGGCCAGTCA CCTGAAGTCCACGCACAGCG GCAAGATCGACTACATCGCA
Gerbillus_campestris  CTCGCCCCTCCTGAAAGACC CGGACTCCTTCCGAGCTTCC ATCCGTCTCCTGGCCAACCA TCTGAAGTCCAAGCATGGCG GCAAAATCGACTACATCGCA

Mus_pahari            GGGCAAGGTGGCCTTGCTAG GCCGTACTCATCCCCCACGG TCCTATCCCCTATCCCCTTT CCCC-TCGTGTCACCCACAG TCTACCCCACACCCATCCAT
Mus_spicilegus        GG-CGAG-TGGCCTTGCTAG GCCGTGCTCGTCCCCCACGG TCCTAGCCCCTATCCCCTTT CCCCCTCGTGTCACCCACAG TCTGCCCCACACCCATCCAT
Gerbillus_campestris  GG-CGAG-TGTTCTTGCTAG GCCGTGCCCGTTCCC-ACTG TCAGGGCCGCCATCCCGTGT TCCCTT--TTTC-------G TGTCACCCACACCCACCCCT
```

```
Mus_pahari            GGGCAAGGTGGCCTTGCTAG GCCGTACTCATCCCCCACGG TCCTATCCCCTATCCCCTTT CCCC-TCGTGTCACCCACAG TCTACCCCACACCCATCCAT
Mus_spicilegus        GG-CGAG-TGGCCTTGCTAG GCCGTGCTCGTCCCCCACGG TCCTAGCCCCTATCCCCTTT CCCCCTCGTGTCACCCACAG TCTGCCCCACACCCATCCAT
Gerbillus_campestris  GG-CGAG-TGTTCTTGCTAG GCCGTGCCCGTTCCC-ACTG TCAGGGCCGCCATCCCGTGT TCCCTT--TTTC-------G TGTCACCCACACCCACCCCT

Mus_pahari            TCTTTCTTTAACCTCTGACT CTTCCTCCTTGGTTTCTCAC TGCCTTGGACGCTTGTTCAC CCCGGATGAACTCCGTAGGC GTCTCCCTTCCCTGCTTGGT
Mus_spicilegus        TCTTTCTTCAACCTCTGACA CTTCCTCCTTGGTTCCTCAC TGCCTTGGACGCTTGTTCAC CCCGGATGAACTATGTAGGA GTCTCCCTTCCCTGCTAGGT
Gerbillus_campestris  CCTTTCTCTGACA-CTCCCA AGTTC-CCT--GTTCCTCTC TGCCTTGGTCCCATATTCAC CCCGGATGA-CTGCG---GA GTCTCCC------------

Mus_pahari            ACCCTAAGG----TGCCCTC GGTGCTTGTTCGTA---GAG ACGAACTCTGCTCTGTCCTT GTGTCCAGAACCAAGCCTTC
Mus_spicilegus        ACCCTAAGGCATCTGCCCTC GGTGCTTGTTCCTA---GAG ACGAACTCTGCTCTGTCCTT GTGTCCAGAACCAGGCCTCC
Gerbillus_campestris  ACCCTCTGACCTCTGCTCTC AAAGCCTGTCCCTACTAGAG AGGAACTCTGCTCTGTCCAT GTGTGCAGGGCCAGCTCTTC

;
END;
BEGIN TREES;
        TREE tree1 = (Mus_pahar, (Mus_spici,Gerbillus));
        TREE tree2 = (Mus_spici, (Mus_pahar,Gerbillus));
END;
BEGIN NOTES;
        PICTURE TAXON=3 FORMAT=GIF SOURCE=FILE
        PICTURE=a_rodent.gif
END;
```

Tree

ASN1

- Abstract Syntax Notation
- developed to aid computer access
- made to be read/written by computers

```
Seq-entry ::= set {
  level 1 ,
  class nuc-prot ,
  descr {
    title "Mus pahari adenine phosphoribosyltransferase (APRT) gene, and
 translated products" ,
    source {
      org {
        taxname "Mus pahari" ,
        common "shrew mouse" ,
        db {
          {
            db "taxon" ,
            tag
              id 10093 } } ,
        orgname {
          name
            binomial {
              genus "Mus" ,
              species "pahari" } ,
          lineage "Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
 Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muroidea; Muridae; Murinae; Mus" ,
          gcode 1 ,
          mgcode 2 ,
          div "ROD" } } } ,
    create-date
      std {
```

```
      year 1995 ,
      month 6 ,
      day 28 } ,
    pub {
      pub {
        article {
          title {
            name "Substitution rate variation in closely related rodent
species." } ,
          authors {
            names
              std {
                {
                  name
                    name {
                      last "Fieldhouse" ,
                      initials "D." } } ,
                {
                  name
                    name {
                      last "Yazdani" ,
                      initials "F." } } ,
                {
                  name
                    name {
                      last "Golding" ,
                      initials "G.B." } } } ,
```

```
          affil
            str "Department of Biology, McMaster University, Hamilton,
Ontario, Canada." } ,
        from
          journal {
            title {
              iso-jta "Heredity" ,
              ml-jta "Heredity" ,
              issn "0018-067X" ,
              name "Heredity." } ,
            imp {
              date
                std {
                  year 1997 ,
                  month 1 } ,
              volume "78" ,
              issue " Pt 1" ,
              pages "21-31" ,
              language "eng" } } ,
        ids {
          pubmed 9023989 ,
          medline 97176443 } } ,
      pmid 9023989 ,
      muid 97176443 ,
      gen {
        serial-number 1 } } } ,
    pub {
```

```
    pub {
      gen {
        serial-number 2 } ,
      sub {
        authors {
          names
            std {
              {
                name
                  name {
                    last "Fieldhouse" ,
                    initials "D." } } } ,
          affil
            str "Dan Fieldhouse, Biology, McMaster University, 1280 Main
Street West, Hamilton, ON, L8S 4K1, Canada" } ,
        medium other ,
        date
          std {
            year 1995 ,
            month 6 ,
            day 7 } } } } ,
  update-date
    std {
      year 2001 ,
      month 10 ,
      day 15 } } ,
 seq-set {
  seq {
```

```
    id {
      genbank {
        name "MPU28721" ,
        accession "U28721" ,
        version 1 } ,
      gi 881573 } ,
    descr {
      title "Mus pahari adenine phosphoribosyltransferase (APRT) gene,
complete cds." ,
      molinfo {
        biomol genomic } } ,
    inst {
      repr raw ,
      mol dna ,
      length 2283 ,
      seq-data
```

```
ncbi2na
'5E68C745D75FB75C4246694ED62DE2F807AE99A64D649F5587D54D5A9B9EF4AE6B4
625A629BE996C774D55A64A66EA497EA8DF9AA5DE5694466B4775ED7EF54A8CDD95
77E0235A1D7D627D4D65D7A52D1782D46449A42361CCD92A42BA5F9CA5B1D35546
B5CD57355FD576ED1512DC55115353DFDFC177877D75FAFDD1E5FA19FBD1568E075
B29B757D5E7EB15C2B95DAE7EF6C8860779DED7EED481425F5DFF2A4429E94935E1
24A7A885E8D75238684D7C7CAAC977A8E07233C00F2B05FAA6E5EA485D0B7AC9F4A
A79F755287116A91DF77ED7551554427EE701079ECC529D4E7E27D20115CA92783B5
14A2ED48A8AB8915422048BA572C0E74A851207FE54751F5CFAC55694E552034A93B
FB15D5461276A597A0785EC84B9D7AB239E4FE02BA422A7AE23A749AF2891E1E77D4
0AD78BD035490513AE9D1051712712EC444CC300C0C040DF00000008080AE908945
32E88292B28D5429C239C58B0534BBDF72532EA42172EF5CB43BE177531F97769D4D
5115F575F15C12B721D4AA7D7BFA57D5C9D289EA6EA79B9D35A092A82796A551CCD
25D739DE8B3A82B0A89EEAC8A0A92ADF3C51A714B9728B03BAB9D222BE213EAB8AF
C41D780E7497480E7529CA8AE947EF24DC8777C19C7D7B7929E27A035202397C815
A9222EB4FBA385D7A51E8AC08149508404A7D02A5295EDEAB9E1C04099F8317DDFD
DED5F55555554A053BF925EE379E45271A9E2BAE8BBB897AE89E176782A4A88A7285
CC53DF7775D4B38789E9C8EB4455E744924B0799AE9D257A9970B85FEE27179E54'H
}
    annot {
      {
        data
          ftable {
            {
              data
                rna {
                  type mRNA ,
                  ext
```

```
      name "adenine phosphoribosyltransferase" } ,
partial TRUE ,
location
  mix {
    int {
      from 45 ,
      to 124 ,
      id
        gi 881573 ,
      fuzz-from
        lim lt } ,
    int {
      from 255 ,
      to 361 ,
      id
        gi 881573 } ,
    int {
      from 1508 ,
      to 1641 ,
      id
        gi 881573 } ,
    int {
      from 1846 ,
      to 1924 ,
      id
        gi 881573 } ,
```

```
                int {
                  from 2043 ,
                  to 2185 ,
                  id
                    gi 881573 ,
                  fuzz-to
                    lim gt } } } ,
              {
                data
                  gene {
                    locus "APRT" } ,
                partial TRUE ,
                location
                  int {
                    from 45 ,
                    to 2185 ,
                    strand plus ,
                    id
                      gi 881573 ,
                    fuzz-from
                      lim lt ,
                    fuzz-to
                      lim gt } } } } } } ,
  seq {
    id {
      genbank {
        accession "AAA68957" ,
```

```
      version 1 } ,
    gi 881574 } ,
  descr {
    molinfo {
      biomol peptide ,
      tech concept-trans } ,
    title "adenine phosphoribosyltransferase [Mus pahari]" } ,
  inst {
    repr raw ,
    mol aa ,
    length 180 ,
    seq-data
      ncbieaa
"MSESELKLVARRIRSFPDFPIPGVLFRDISPLLKDPDSFRASIRLLASHLKSTHSGKID
YIAGLDSRGFLFGPSLAQELGVGCVLIRKQGKLPGPTISASYALEYGKAELEIQKDALEPGQRVVIVD
DLLATGGTMF
AACDLLHQLRAEVVECVSLVELTSLKGRERLGPIPFFSLLQYD" } ,
  annot {
    {
      data
        ftable {
          {
            data
              prot {
                name {
                  "adenine phosphoribosyltransferase" } ,
```

```
          ec {
            "2.4.2.7" } } ,
        location
          whole
            gi 881574 } } } ,
{
  db other ,
  name "Annot:CDD" ,
  desc {
    name "CDDSearch" ,
    create-date
      std {
        year 2007 ,
        month 6 ,
        day 18 ,
        hour 23 ,
        minute 46 ,
        second 57 } } ,
  data
    ftable {
      {
        data
          region "PRK02304" ,
        comment "adenine phosphoribosyltransferase" ,
        location
          int {
            from 5 ,
            to 178 ,
```

```
    id
      gi 881574 } ,
ext {
  type
    str "cddScoreData" ,
  data {
    {
      label
        str "definition" ,
      data
        str "PRK02304" } ,
    {
      label
        str "short_name" ,
      data
        str "PRK02304" } ,
    {
      label
        str "score" ,
      data
        int 575 } ,
    {
      label
        str "evalue" ,
      data
        real { 307765, 10, -64 } } ,
```

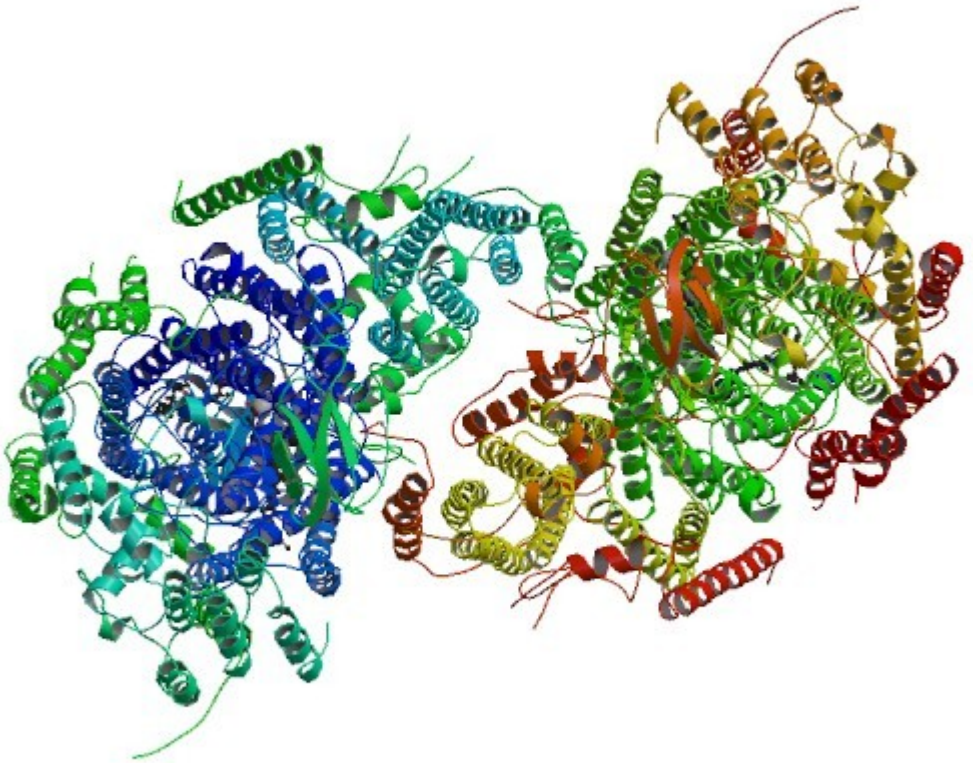
```
                {
                  label
                    str "bit_score" ,
                  data
                    real { 225315, 10, -3 } } } } ,
              dbxref {
                {
                  db "CDD" ,
                  tag
                    id 74170 } } } } } } } } ,
annot {
  {
    data
      ftable {
        {
          data
            cdregion {
              frame one ,
              code {
                id 1 } } ,
          comment "purine salvage enzyme" ,
          product
            whole
```

```
gi 881574 ,
  location
    mix {
      int {
        from 45 ,
        to 124 ,
        id
          gi 881573 } ,
      int {
        from 255 ,
        to 361 ,
        id
          gi 881573 } ,
      int {
        from 1508 ,
        to 1641 ,
        id
          gi 881573 } ,
      int {
        from 1846 ,
        to 1924 ,
        id
          gi 881573 } ,
      int {
        from 2043 ,
        to 2185 ,
        id
          gi 881573 } } } } } } }
```

PDB

- Protein file
- Store the 3D position of each amino acid
- Primary and secondary structures
- Crystallographic experiments and parameters

CYTOCHROME C OXIDASE



```
HEADER    OXIDOREDUCTASE                          26-MAY-98   2OCC
TITLE     BOVINE HEART CYTOCHROME C OXIDASE AT THE FULLY OXIDIZED
TITLE    2 STATE
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: CYTOCHROME C OXIDASE;
COMPND   3 CHAIN: A, N;
COMPND   4 SYNONYM: FERROCYTOCHROME C\:OXYGEN OXIDOREDUCTASE;
COMPND   5 EC: 1.9.3.1;
COMPND   6 OTHER_DETAILS: THIS ENZYME IS A HYBRID PROTEIN COMPLEX AND
COMPND   7 IS A HOMODIMER. ONE MONOMER IS COMPOSED OF 13 DIFFERENT
COMPND   8 SUBUNITS AND SEVEN METAL CENTERS, HEME A, HEME A3, CUA,
COMPND   9 CUB, MG, NA AND ZN.;

     ------------------------------------------------------------

SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: BOS TAURUS;
SOURCE   3 ORGANISM_COMMON: BOVINE;
SOURCE   4 ORGAN: HEART;
SOURCE   5 TISSUE: HEART MUSCLE;
SOURCE   6 ORGANELLE: MITOCHONDRION;

     ------------------------------------------------------------

KEYWDS    OXIDOREDUCTASE, CYTOCHROME(C)-OXYGEN, CYTOCHROME C OXIDASE
EXPDTA    X-RAY DIFFRACTION
AUTHOR    T.TSUKIHARA,M.YAO
REVDAT   1   13-JAN-99 2OCC    0
JRNL        AUTH   S.YOSHIKAWA,K.SHINZAWA-ITOH,R.NAKASHIMA,R.YAONO,
JRNL        AUTH 2 E.YAMASHITA,N.INOUE,M.YAO,M.J.FEI,C.P.LIBEU,
JRNL        AUTH 3 T.MIZUSHIMA,H.YAMAGUCHI,T.TOMIZAKI,T.TSUKIHARA
JRNL        TITL   REDOX-COUPLED CRYSTAL STRUCTURAL CHANGES IN BOVINE
JRNL        TITL 2 HEART CYTOCHROME C OXIDASE.
JRNL        REF    SCIENCE                       V. 280  1723 1998
JRNL        REFN   ASTM SCIEAS  US ISSN 0036-8075
```

```
REMARK   1
REMARK   1 REFERENCE 1
REMARK   1  AUTH   T.TSUKIHARA,H.AOYAMA,E.YAMASHITA,T.TOMIZAKI,
REMARK   1  AUTH 2 H.YAMAGUCHI,K.SHINZAWA-ITOH,R.NAKASHIMA,R.YAONO,
REMARK   1  AUTH 3 S.YOSHIKAWA
REMARK   1  TITL   THE WHOLE STRUCTURE OF THE 13-SUBUNIT OXIDIZED
REMARK   1  TITL 2 CYTOCHROME C OXIDASE AT 2.8 A
REMARK   1  REF    SCIENCE                       V. 272  1136 1996
REMARK   1  REFN   ASTM SCIEAS  US ISSN 0036-8075
REMARK   1 REFERENCE 2
```

```
REMARK   2 RESOLUTION. 2.30 ANGSTROMS.
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : X-PLOR 3.84
REMARK   3   AUTHORS     : BRUNGER
REMARK   3
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) : 2.30
REMARK   3   RESOLUTION RANGE LOW  (ANGSTROMS) : 15.00
REMARK   3   DATA CUTOFF            (SIGMA(F)) : 2.000
REMARK   3   DATA CUTOFF HIGH         (ABS(F)) : 100000.000
REMARK   3   DATA CUTOFF LOW          (ABS(F)) : 0.1000
REMARK   3   COMPLETENESS (WORKING+TEST)   (%) : 88.9
REMARK   3   NUMBER OF REFLECTIONS             : 278049
REMARK   3
REMARK   3  FIT TO DATA USED IN REFINEMENT.
REMARK   3   CROSS-VALIDATION METHOD          : THROUGHOUT
REMARK   3   FREE R VALUE TEST SET SELECTION  : RANDOM
REMARK   3   R VALUE            (WORKING SET) : 0.209
```

```
REMARK   4
REMARK   4 2OCC COMPLIES WITH FORMAT V. 3.0, 1-DEC-2006
REMARK   4
REMARK   4 THIS IS THE REMEDIATED VERSION OF THIS PDB ENTRY.
REMARK   4 REMEDIATED DATA FILE REVISION 3.101 (2007-05-01)
REMARK 200
REMARK 200 EXPERIMENTAL DETAILS
REMARK 200  EXPERIMENT TYPE                : X-RAY DIFFRACTION
REMARK 200  DATE OF DATA COLLECTION        : MAY-1996
REMARK 200  TEMPERATURE           (KELVIN) : 283.0
REMARK 200  PH                             : 6.80
REMARK 200  NUMBER OF CRYSTALS USED        : 32
REMARK 200
REMARK 200  SYNCHROTRON              (Y/N) : Y
REMARK 200  RADIATION SOURCE               : PHOTON FACTORY
REMARK 200  BEAMLINE                       : BL-6B
REMARK 200  X-RAY GENERATOR MODEL          : NULL
REMARK 200  MONOCHROMATIC OR LAUE    (M/L) : M
```

```
DBREF  2OCC A    1   514  UNP    P00396   COX1_BOVIN       1    514
DBREF  2OCC B    1   227  UNP    P00404   COX2_BOVIN       1    227
DBREF  2OCC C    1   261  UNP    P00415   COX3_BOVIN       1    261
DBREF  2OCC D    1   147  UNP    P00423   COX4_BOVIN      23    169
DBREF  2OCC E    1   109  UNP    P00426   COXA_BOVIN       1    109
DBREF  2OCC F    1    98  UNP    P00428   COXB_BOVIN       1     98
DBREF  2OCC G    1    84  UNP    P07471   COXD_BOVIN      13     96
DBREF  2OCC H    1    85  UNP    P00429   COXG_BOVIN       1     85
DBREF  2OCC I    1    73  UNP    P04038   COXH_BOVIN       1     73
DBREF  2OCC J    1    59  UNP    P07470   COXK_BOVIN      22     80
DBREF  2OCC K    1    56  UNP    P13183   COXM_BOVIN      33     88
DBREF  2OCC L    1    47  UNP    P00430   COXO_BOVIN      17     63
```

```
DBREF  2OCC M    1    46  UNP    P10175   COXQ_BOVIN      25     70
DBREF  2OCC N    1   514  UNP    P00396   COX1_BOVIN       1    514
DBREF  2OCC O    1   227  UNP    P00404   COX2_BOVIN       1    227
DBREF  2OCC P    1   261  UNP    P00415   COX3_BOVIN       1    261
DBREF  2OCC Q    1   147  UNP    P00423   COX4_BOVIN      23    169
DBREF  2OCC R    1   109  UNP    P00426   COXA_BOVIN       1    109
DBREF  2OCC S    1    98  UNP    P00428   COXB_BOVIN       1     98
DBREF  2OCC T    1    84  UNP    P07471   COXD_BOVIN      13     96
DBREF  2OCC U    1    85  UNP    P00429   COXG_BOVIN       1     85
DBREF  2OCC V    1    73  UNP    P04038   COXH_BOVIN       1     73
DBREF  2OCC W    1    59  UNP    P07470   COXK_BOVIN      22     80
DBREF  2OCC X    1    56  UNP    P13183   COXM_BOVIN      33     88
DBREF  2OCC Y    1    47  UNP    P00430   COXO_BOVIN      17     63
DBREF  2OCC Z    1    46  UNP    P10175   COXQ_BOVIN      25     70
SEQRES   1 A  514  MET PHE ILE ASN ARG TRP LEU PHE SER THR ASN HIS LYS
SEQRES   2 A  514  ASP ILE GLY THR LEU TYR LEU LEU PHE GLY ALA TRP ALA
SEQRES   3 A  514  GLY MET VAL GLY THR ALA LEU SER LEU LEU ILE ARG ALA
SEQRES   4 A  514  GLU LEU GLY GLN PRO GLY THR LEU LEU GLY ASP ASP GLN
SEQRES   5 A  514  ILE TYR ASN VAL VAL VAL THR ALA HIS ALA PHE VAL MET
SEQRES   6 A  514  ILE PHE PHE MET VAL MET PRO ILE MET ILE GLY GLY PHE
SEQRES   7 A  514  GLY ASN TRP LEU VAL PRO LEU MET ILE GLY ALA PRO ASP
SEQRES   8 A  514  MET ALA PHE PRO ARG MET ASN ASN MET SER PHE TRP LEU

    ------------------------------------------------------------

SEQRES   1 B  227  MET ALA TYR PRO MET GLN LEU GLY PHE GLN ASP ALA THR
SEQRES   2 B  227  SER PRO ILE MET GLU GLU LEU LEU HIS PHE HIS ASP HIS
SEQRES   3 B  227  THR LEU MET ILE VAL PHE LEU ILE SER SER LEU VAL LEU
SEQRES   4 B  227  TYR ILE ILE SER LEU MET LEU THR THR LYS LEU THR HIS
SEQRES   5 B  227  THR SER THR MET ASP ALA GLN GLU VAL GLU THR ILE TRP
SEQRES   6 B  227  THR ILE LEU PRO ALA ILE ILE LEU ILE LEU ILE ALA LEU
SEQRES   7 B  227  PRO SER LEU ARG ILE LEU TYR MET MET ASP GLU ILE ASN

    ------------------------------------------------------------
```

```
HET    HEA  A 515      60
HET    HEA  A 516      60
HET    PER  A 520       2
HET    HEA  N 515      60
HET    HEA  N 516      60
HET    PER  N 520       2
HETNAM      CU COPPER (II) ION
HETNAM      MG MAGNESIUM ION
HETNAM      NA SODIUM ION
HETNAM      ZN ZINC ION
HETNAM     HEA HEME-A
HETNAM     PER PEROXIDE ION
FORMUL  27   CU    6(CU 2+)
FORMUL  28   MG    2(MG 2+)
FORMUL  29   NA    2(NA 1+)
FORMUL  32   ZN    2(ZN 2+)
FORMUL  39  HEA    4(C49 H56 FE N4 O6)
FORMUL  41  PER    2(O2 2-)
HELIX    1   1 PHE A    2  TRP A    6  1                                   5
HELIX    2   2 HIS A   12  LEU A   41  1                                  30
HELIX    3   3 ASP A   51  ILE A   87  1                                  37
HELIX    4   4 PRO A   95  MET A  117  1                                  23
HELIX    5   5 ALA A  141  ASN A  170  1                                  30
HELIX    6   6 LEU A  183  ASN A  214  1                                  32

     ------------------------------------------------------------

SHEET    1   A 5 LEU B 116  SER B 120  0
SHEET    2   A 5 TYR B 105  TYR B 110 -1  N  TYR B 110   O  LEU B 116
SHEET    3   A 5 LEU B  95  HIS B 102 -1  N  HIS B 102   O  TYR B 105
SHEET    4   A 5 ILE B 150  SER B 156  1  N  ARG B 151   O  LEU B  95
SHEET    5   A 5 ASN B 180  LEU B 184 -1  N  LEU B 184   O  ILE B 150
SHEET    1   B 3 VAL B 142  PRO B 145  0
SHEET    2   B 3 ILE B 209  VAL B 214  1  N  GLU B 212   O  VAL B 142
SHEET    3   B 3 GLY B 190  GLY B 194 -1  N  GLY B 194   O  ILE B 209
SHEET    1   C 2 HIS B 161  VAL B 165  0
SHEET    2   C 2 LEU B 170  ALA B 174 -1  N  ALA B 174   O  HIS B 161
SHEET    1   D 3 ASN F  47  SER F  51  0
```

```
SHEET    2   D 3 GLY F  86  PRO F  93  1  N  LYS F  90   O  ASN F  47
SHEET    3   D 3 GLN F  80  CYS F  82 -1  N  CYS F  82   O  GLY F  86
SHEET    1   E 2 LYS F  55  CYS F  60  0
SHEET    2   E 2 ILE F  70  HIS F  75 -1  N  LEU F  74   O  ARG F  56
SHEET    1   F 5 LEU O 116  SER O 120  0
SHEET    2   F 5 TYR O 105  TYR O 110 -1  N  TYR O 110   O  LEU O 116
SHEET    3   F 5 LEU O  95  HIS O 102 -1  N  HIS O 102   O  TYR O 105
SHEET    4   F 5 ILE O 150  SER O 156  1  N  ARG O 151   O  LEU O  95
SHEET    5   F 5 ASN O 180  LEU O 184 -1  N  LEU O 184   O  ILE O 150
SHEET    1   G 3 VAL O 142  PRO O 145  0
SHEET    2   G 3 ILE O 209  VAL O 214  1  N  GLU O 212   O  VAL O 142
SHEET    3   G 3 GLY O 190  GLY O 194 -1  N  GLY O 194   O  ILE O 209
SHEET    1   H 2 HIS O 161  VAL O 165  0
SHEET    2   H 2 LEU O 170  ALA O 174 -1  N  ALA O 174   O  HIS O 161
SHEET    1   I 3 ASN S  47  SER S  51  0
SHEET    2   I 3 GLY S  86  PRO S  93  1  N  LYS S  90   O  ASN S  47
SHEET    3   I 3 GLN S  80  CYS S  82 -1  N  CYS S  82   O  GLY S  86
SHEET    1   J 2 LYS S  55  CYS S  60  0
SHEET    2   J 2 ILE S  70  HIS S  75 -1  N  LEU S  74   O  ARG S  56
SSBOND   1 CYS H   29    CYS H   64
SSBOND   2 CYS H   39    CYS H   53
SSBOND   3 CYS U   29    CYS U   64
SSBOND   4 CYS U   39    CYS U   53
LINK         FE   HEA A 515                NE2 HIS A  61
LINK         FE   HEA A 515                NE2 HIS A 378
LINK         FE   HEA A 516                NE2 HIS A 376
LINK         FE   HEA A 516                O1  PER A 520
LINK         CU    CU A 517                ND1 HIS A 240
LINK         CU    CU A 517                NE2 HIS A 290
```

```
ATOM      1  N   MET A   1      55.242 340.693 224.088  1.00 68.90           N
ATOM      2  CA  MET A   1      54.908 339.282 224.487  1.00 71.09           C
ATOM      3  C   MET A   1      54.673 338.307 223.329  1.00 66.66           C
ATOM      4  O   MET A   1      55.350 337.285 223.238  1.00 67.66           O
```

```
ATOM      5  CB  MET A   1      53.723 339.248 225.450  1.00 79.30           C
ATOM      6  CG  MET A   1      54.110 339.452 226.915  1.00 87.90           C
ATOM      7  SD  MET A   1      55.300 338.229 227.515  1.00 97.07           S
ATOM      8  CE  MET A   1      54.166 336.799 228.014  1.00 96.59           C
ATOM      9  N   PHE A   2      53.673 338.579 222.494  1.00 61.89           N
ATOM     10  CA  PHE A   2      53.412 337.739 221.322  1.00 56.50           C
ATOM     11  C   PHE A   2      54.569 337.917 220.303  1.00 53.31           C
ATOM     12  O   PHE A   2      55.076 336.947 219.739  1.00 53.84           O
ATOM     13  CB  PHE A   2      52.077 338.127 220.683  1.00 55.21           C
ATOM     14  CG  PHE A   2      51.737 337.334 219.459  1.00 54.54           C
ATOM     15  CD1 PHE A   2      51.050 336.138 219.565  1.00 55.24           C
ATOM     16  CD2 PHE A   2      52.126 337.775 218.200  1.00 55.62           C
ATOM     17  CE1 PHE A   2      50.756 335.388 218.432  1.00 58.99           C
ATOM     18  CE2 PHE A   2      51.839 337.035 217.059  1.00 57.84           C
ATOM     19  CZ  PHE A   2      51.155 335.840 217.171  1.00 58.36           C
ATOM     20  N   ILE A   3      55.010 339.158 220.116  1.00 47.37           N

     ------------------------------------------------------------

HETATM28635 CU    CU A 517      67.173 310.978 190.358  1.00 16.27          CU
HETATM28636 MG    MG A 518      62.605 315.176 179.115  1.00 19.26          MG
HETATM28637 NA    NA A 519      42.250 318.661 179.405  1.00 26.18          NA
HETATM28638 CU    CU B 228      57.527 320.742 171.423  1.00 21.99          CU
HETATM28639 CU    CU B 229      56.638 319.970 173.568  1.00 24.27          CU
HETATM28640 ZN    ZN F  99      71.521 300.480 232.843  1.00 33.09          ZN
```

```
    ------------------------------------------------------------

CONECT  35128637
CONECT  47428647
CONECT 183628635
CONECT 223928635
CONECT 224928635
CONECT 283428636
CONECT 284228636
CONECT 290228707
CONECT 292328647
CONECT 343128637

    ------------------------------------------------------------

CONECT2888828880
CONECT288892882928890
CONECT288902864128889
MASTER      425    0   18   98   30    0    2    928864   26  308  292
END
```