

GenBank format

```
LOCUS       CMGLOAD                      1185 bp    DNA     linear   VRT 18-APR-2005
DEFINITION  Cairina moschata (duck) gene for alpha-D globin.
ACCESSION   X01831
VERSION     X01831.1  GI:62724
KEYWORDS    alpha-globin; globin.
SOURCE      Cairina moschata (Muscovy duck)
  ORGANISM  Cairina moschata
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Cairina.
REFERENCE   1  (bases 1 to 1185)
AUTHORS     Erbil,C. and Niessing,J.
TITLE       The primary structure of the duck alpha D-globin gene: an unusual
            5' splice junction sequence
JOURNAL     EMBO J. 2 (8), 1339-1343 (1983)
PUBMED      10872328
COMMENT     Data kindly reviewed (13-NOV-1985) by J. Niessing.
FEATURES             Location/Qualifiers
     source           1..1185
                    /organism="Cairina moschata"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:8855"
     CAAT_signal      20..24
     TATA_signal      69..73
     precursor_RNA    101..1114
                    /note="primary transcript"
     exon             101..234
                    /number=1
     CDS              join(143..234,387..591,939..1067)
                    /codon_start=1
                    /product="alpha D-globin"
                    /protein_id="CAA25966.2"
                    /db_xref="GI:4455876"
                    /db_xref="GOA:P02003"
                    /db_xref="InterPro:IPR000971"
                    /db_xref="InterPro:IPR002338"
                    /db_xref="InterPro:IPR002340"
                    /db_xref="InterPro:IPR009050"
                    /db_xref="UniProt/Swiss-Prot:P02003"
                    /translation="MLTAEDKKLIVQVWEKVAGHQEEFGSEALQRMFLAYPQTKTYFP
                    HFDLHPGSEQRVGHGKKVAALGNVAKSLDNLSQLSELSNLHAYNLRVDPVNFKLLA
                    QCFQVVLAAHLGKDYSPEMHAADFDFLSAVAVALAEKYR"
     repeat_region    227..246
                    /note="direct repeat 1"
     intron           235..386
                    /number=1
     repeat_region    289..309
                    /note="direct repeat 1"
     exon             387..591
                    /number=2
     intron           592..939
                    /number=2
     exon             940..1114
                    /number=3
     polyA_signal     1095..1100
     polyA_signal     1114

ORIGIN
1   ctgcgtggcc  toagcccctc  caccctccca  cgctgataag  ataaggccag  ggcgggagcg
61  cagggtgcta  taagagctcg  gccccgcggg  tgtctccacc  acagaaaccc  gtcagttgcc
121 agcctgccac  gccgctgccg  ccatgctgac  cgccagggac  aagaagctca  tcgtgcaggt
181 gtgggagaag  gtggctggcc  accaggagga  attcggaagt  gaagctctgc  agagggtgtg
241 gctgggccca  gggggcactc  acagggtggg  cagcaggggg  caggagccct  gcagcgggtg
301 tgggctggga  cccagagcgc  cacgggggtg  gggctgagat  gggcaaaagc  gcagggcacc
361 aaaactgact  ggccctgcct  cggcaggatg  ttctctgcct  acccccagac  caagacctac
421 ttccccactg  tcgacctgca  tcccggctct  gaacagggtc  gtggccatgg  caagaaagtg
481 cgggctgccc  tgggcaatgc  cgtgaagagc  ctggacaacc  tcagccaggc  cctgtctgag
541 ctcaagcaacc  tgcatgccta  caacctgcgt  gttgacctg  tcaacttcaa  ggcaagcggg
601 gactagggtc  cttgggtctg  ggggtctgag  ggtgtggggt  gcagggtctg  ggggtccagg
661 ggtctgagtt  tcctggggtc  tggcagtcct  gggggctgag  ggccagggtc  ctgtggtctt
721 gggtaaccag  gtctctgggg  ccagcagcca  gacagcaggg  gctgggattg  catctgggat
781 gtgggocaga  ggctgggatt  gtgtttgaa  tgggagctgg  gcaggggcta  gggccagggt
841 gggggactca  gggcctcagg  gggactcggg  gggggactga  gggagactca  gggccatctg
901 tccggagcag  gggctactaa  ccctggtttg  ccttgcaagt  gctggcacag  tgcttccagg
961 tgggtctggc  cgcacacctg  ggcaaaact  acagcccga  gatgcatgct  gcctttgaca
1021 agttctgttc  cgccgtggct  gccgtgctgg  ctgaaaagta  cagatgagcc  actgcctgca
1081 cccttgccac  ttcaataaag  acaccattac  cacagctctg  tgtctgtgtg  tgctgggact
1141 gggcatcggg  ggtcccaggg  agggctgggt  tgcttcaca  catcc
```

Header

Contains information about Organism, publications, Accession ID etc.

FEATURE table

Contains annotation of block of interest with in the sequence (e.g. Genes)..

CDS: **Coding Sequence.** Contains the coordinates of the protein coding part of a gene. May be split in multiple parts if the gene contains introns.

ORIGIN section

Contains the DNA sequence.

//

FASTA format

Header

Each sequence entry starts with a header line which must begin with the character ">". Typically this line is used to provide a name for the sequence. If the FASTA file is generated by automated extraction from a database, additional information about database ID's and sequence intervals will often be present.

Sequence

The FASTA format can be used to contain any kind of biological sequence data: DNA, RNA, Protein etc. The sequence block ends at the next ">" or by the end of the file, if it's the last entry.

>alpha-D

```
ATGCTGACCGACTCTGACAAGAAGCTGGTCTGCAGGTGTGGGAGAAGGTGATCCGCCAC
CCAGACTGTGGAGCCGAGGCCCTGGAGAGGTGCGGGCTGAGCTTGGGGAAACCATGGGCA
AGGGGGGCGACTGGGTGGGAGCCCTACAGGGCTGCTGGGGGTGTTCGGCTGGGGGTGAG
CACTGACCATCCCGCTCCCGCAGCTGTTACCACCTACCCCCAGACCAAGACCTACTTCC
CCCCTTCGACTTGCACCATGGCTCCGACCAGGTCCGCAACCACGGCAAGAAGGTGTTGG
CCGCCTTGGGCAACGCTGTCAAGAGCCTGGGCAACCTCAGCCAAGCCCTGTCTGACCTCA
GCGACCTGCATGCCTACAACCTGCGTGTGACCCCTGTCAACTTCAAGGCAGGCGGGGGAC
GGGGGTGAGGGGCCGGGGAGTTGGGGGCCAGGGACCTGGTTGGGGATCCGGGGCCATGCC
GGCGGTACTGAGCCCTGTTTTGCCTTGCAGCTGCTGGCGCAGTGCTTCCACGTGGTGCTG
GCCACACACCTGGGCAACGACTACACCCCGGAGGCACATGCTGCCTTCGACAAGTTCCTG
TCGGCTGTGTGCACCGTGCTGGCCGAGAAGTACAGATAA
```

>alpha-A

```
ATGGTGCTGTCTGCCAACGACAAGAGCAACGTGAAGGCCGTCTTCGGCAAAATCGGCGGC
CAGGCCGGTGACTTGGGTGGTGAAGCCCTGGAGAGGTATGTGGTCATCCGTCATTACCCC
ATCTCTTGTCTGTCTGTGACTCCATCCCATCTGCCCCATACTCTCCCCATCCATAACTG
TCCCTGTTCTATGTGGCCCTGGCTCTGTCTCATCTGTCCCCAACTGTCCCTGATTGCCTC
TGTCCCCCAGGTTGTTTCATCACCTACCCCCAGACCAAGACCTACTTCCCCACTTCGACC
TGTCACATGGCTCCGCTCAGATCAAGGGGCACGGCAAGAAGGTGGCGGAGGCACTGGTTG
AGGCTGCCAACACATCGATGACATCGCTGGTGGCCCTCTCCAAGCTGAGCGACCTCCACG
CCCAAAAGCTCCGTGTGGACCCCGTCAACTTCAAAGTGAGCATCTGGGAAGGGGTGACCA
GTCTGGCTCCCCCTCCTGCACACACCTCTGGCTACCCCCTCACCTCACCCCCTTGCTCACC
ATCTCCTTTTGCCTTTTTCAGCTGCTGGGTCACTGCTTCCCTGGTGGTCTGGCCGTCCACTT
CCCCCTCTCTCCTGACCCCGGAGGTCCATGCTTCCCTGGACAAGTTCGTGTGTGCCGTGGG
CACCGTCCTTACTGCCAAGTACCGTTAA
```