Πληροφοριακά συστήματα Γονιδιώματος

GenBank

- Η GenBank είναι μία βάση δεδομένων με πληροφορίες DNA. Την βάση την διαχειρίζεται το National Center for Biotechnology Information (NCBI).). Η GenBank χωρίζεται σε 17 διακριτές κατηγορίες (divisions) ανάλογα με το είδος της αλληλουχίας DNA που περιέχουν:

Division	sequence subset	Division	sequence subset
PRI	primate	PHG	Bacteriophage
ROD	rodent	SYN	synthetic
MAM	other mammalian	UNA	unannotated
VRT	other vertebrate	EST	expressed sequence tags
INV	invertebrate	PAT	patent
PLN	plant, fungal, algal	STS	sequence tagged sites
BCT	bacterial	GSS	genome survey sequences
RNA	structural RNA	HTG	high throughput genomic seq.
VRL	viral		

- Πληροφορίες μπορεί κανείς να ανακτήσει από τη GenBank, χρησιμοποιώντας το ολοκληρωμένο σύστημα ανάκτησης πληροφοριών Entrez.

Δομή των καταχωρήσεων της GenBank

- Μια καταχώρηση στη GenBank περιέχει το αρχείο αλληλουχίας, το οποίο περιέχει εκτός από την αλληλουχία καθ'αυτή και διάφορες περιγραφικές πληροφορίες που σχετίζονται με αυτήν.

- Κάθε καταχώρηση (entry) επίσης περιλαμβάνει και διάφορες λέξεις-κλειδιά, σχετιζόμενα υπό-κλειδιά και ένα προαιρετικό αρχείο αξιοσημείωτων χαρακτηριστικών(feature table).

- Μια GenBank γραμμογράφηση μιας καταχώρηση για την ανθρώπινη cyclooxygenase (http://en.wikipedia.org/wiki/Cyclooxygenase)είναι η ακόλουθη:

- LOCUS HUMCYCLOX 3387 bp mRNA linear PRI 31-DEC-1994
- DEFINITION Homo sapiens cyclooxygenase-2 (Cox-2) mRNA, complete cds.
- ACCESSION M90100
- VERSION M90100.1 GI:181253
- KEYWORDS cyclooxygenase-2; prostaglandin synthase.
- SOURCE Homo sapiens (human)
- ORGANISM Homo sapiens
- Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
- Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
- REFERENCE 1 (bases 1 to 3387)
- AUTHORS Hla,T. and Neilson,K.
- TITLE Human cyclooxygenase-2 cDNA
- JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (16), 7384-7388 (1992)
- MEDLINE 92366465
- PUBMED 1380156
- COMMENT Original source text: Homo sapiens umbilical vein cDNA to mRNA.

- FEATURES Location/Qualifiers
- source 1..3387
- /organism="Homo sapiens"
- /mol_type="mRNA"
- /db_xref="taxon:9606"
- /cell_type="endothelial"
- /tissue_type="umbilical vein"
- gene 1..3387
- /gene="Cox-2"
- 5'UTR 1..97
- /gene="Cox-2"
- CDS 98..1912
- /gene="Cox-2"
- /EC_number="1.14.99.1"
- /codon_start=1
- /product="cyclooxygenase-2"
- /protein_id="AAA58433.1"
- /db_xref="GI:181254"
- /translation="MLARALLLCAVLALSHTANPCCSHPCQNRGVCMSVGFDQYKCDC
- TRTGFYGENCSTPEFLTRIKLFLKPTPNTVHYILTHFKGFWNVVNNIPFLRNAIMSYV
- …
- VEVGAPFSLKGLMGNVICSPAYWKPSTFGGEVGFQIINTASIQSLICNNVKGCPFTSF
- SVPDPELIKTVTINASSSRSGLDDINPTVLLKERSTEL"
- sig_peptide 98..148
- /gene="Cox-2"
- mat_peptide 149..1909
- /gene="Cox-2"
- /product="cyclooxygenase-2"
- /EC_number="1.14.99.1"
- 3'UTR 1913..3387
- /gene="Cox-2"
- polyA_signal 3369..3374
- /gene="Cox-2"

- ORIGIN
- 1 gtccaggaac tcctcagcag cgcctccttc agctccacag ccagacgccc tcagacagca
- 61 aagcctaccc ccgcgccgcg ccctgcccgc cgctgcgatg ctcgcccgcg ccctgctgct
- 121 gtgcgcggtc ctggcgctca gccatacagc aaatccttgc tgttcccacc catgtcaaaa
- 181 ccgaggtgta tgtatgagtg tgggatttga ccagtataag tgcgattgta cccggacagg
- 241 attctatgga gaaaactgct caacaccgga atttttgaca agaataaaat tatttctgaa
- 301 acccactcca aacacagtgc actacatact tacccacttc aagggatttt ggaacgttgt
- 361 gaataacatt cccttccttc gaaatgcaat tatgagttat gtgttgacat ccagatcaca
- …
- 3301 tacctgaact tttgcaagtt ttcaggtaaa cctcagctca ggactgctat ttagctcctc
- 3361 ttaagaagat taaaaaaaaa aaaaaag
- //

- Η Γραμμή LOCUS είναι το χαρακτηριστικό που υποδουλώνει αλληλουχία λειτουργικότητας (HUMCYCLOX υποδουλώνει ανθρώπινη cyclooxygenase), τον αριθμό βάσεων, την προέλευση της πληροφορίας αλληλουχίας (mRNA), το τμήμα της ΒΠ (PRI) και την ημερομηνία υποβολής των δεδομένων στη ΒΠ.

- Η γραμμή DEFINITION περιέχει την περιγραφή της αλληλουχίας.

- Η γραμμή ACCESSION είναι ένας μοναδικός κωδικός ο οποίος δίνεται σε κάθε μια καταχώρηση.

- Η γραμμή KEYWORDS περιέχει φράσεις που περιγράφουν τα γονίδια και άλλες σχετικές πληροφορίες.

- Η γραμμή SOURCE περιέχει πληροφορίες για τον ιστό από τον οποίον τα δεδομένα εξήχθησαν.

- Η γραμμή ORGANISM περιγράφει την βιολογική ταξινόμηση του οργανισμού προέλευσης.

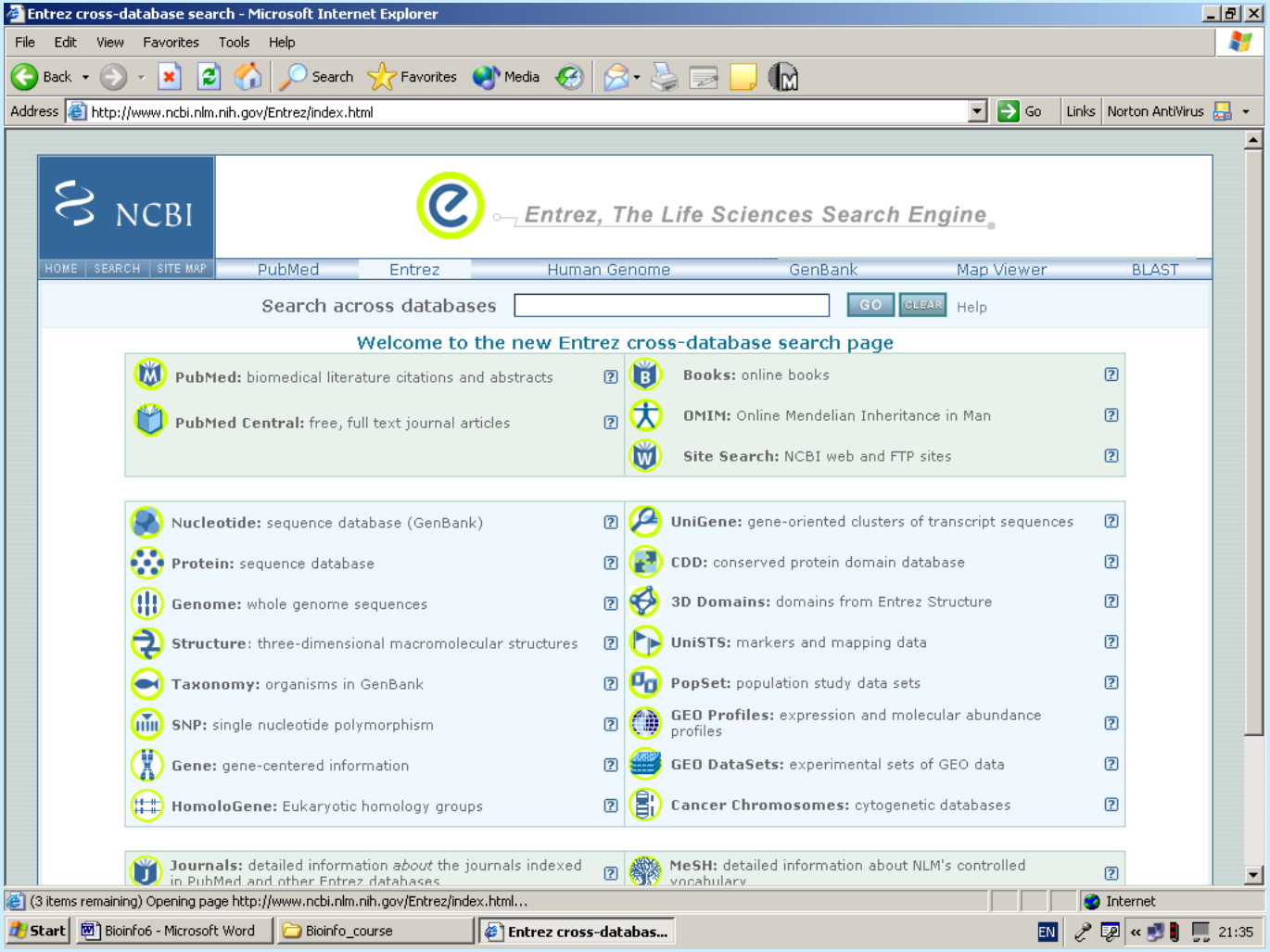
- Η γραμμή REFERENCE παρέχει παραπομπές στη βιβλιογραφία για την συγκεκριμένη αλληλουχία.

- Η γραμμή FEATURES παρέχει ακριβείς πληροφορίες για τα χαρακτηριστικά (Feature Table). Συντεταγμένες παρέχονται για το 5’ μη-μεταφρασμένο μέρος (1-97), για την αλληλουχία κωδικοποίησης (98-1912), για το 3’΄μη-μεταγραφσμένο μέρος (1913-3387), για τη polyadenylation ααλληλουχία (3369-3374) κ.τ.λ. Επίσης παρέχονται πληροφορίες για τη μετάφραση της πρωτεΐνης και τις θέσεις διαφόρων πεπτιδίων.

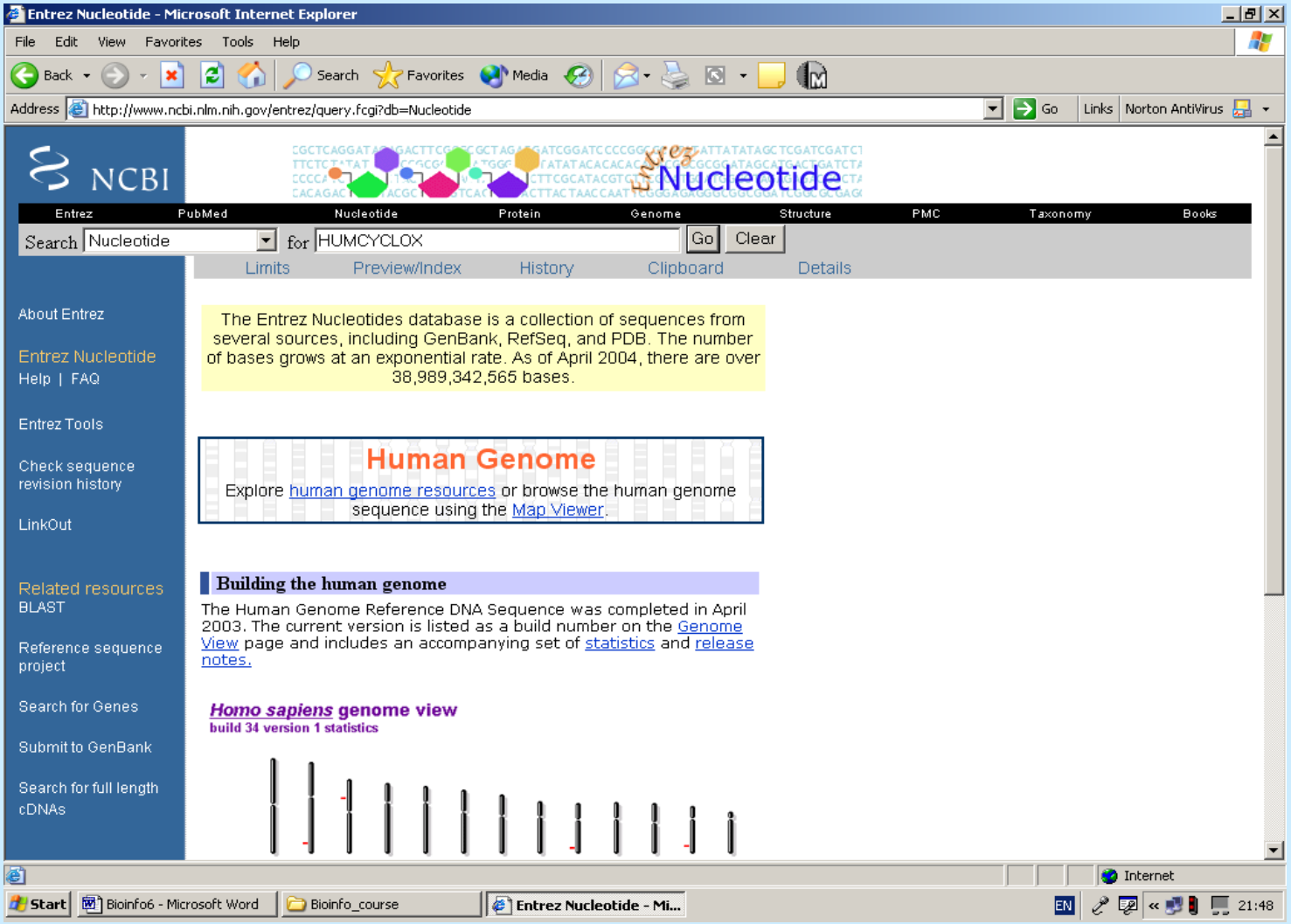
- Η γραμμή BASE COUNT δίνει πληροφορίες για την συχνότητα εμφάνισης των διαφορετικών τύπων βάσεων στην αλληλουχία (i.e. 1010 A, 712 C, 633 G, 1032 T).

- Η γραμμή ORIGIN σημειώνει την πρώτη βάση της αλληλουχίας στο γονιδίωμα.

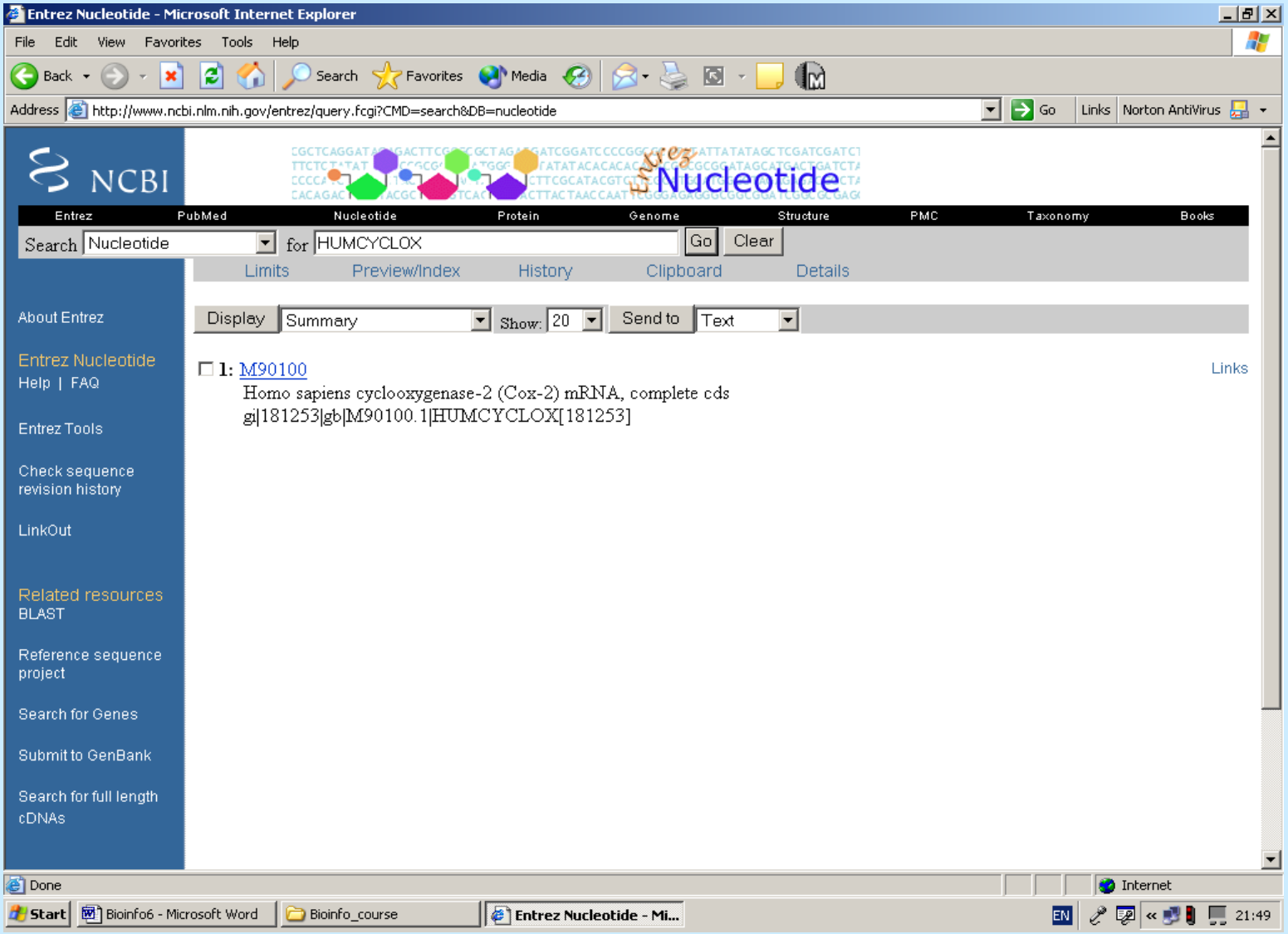
- **ΠΡΑΚΤΙΚΗ ΕΦΑΡΜΟΓΗ**
- ***GenBank***
- Για να προσπελάσετε τη ΒΠ GenBank, στον Internet Explorer πληκτρολογήστε την ακόλουθη διεύθυνση : www.ncbi.nlm.nih.gov/Entrez/index.html. Στη συνέχεια επιλέξετε “Nucleotide: sequence database (GenBank)”.



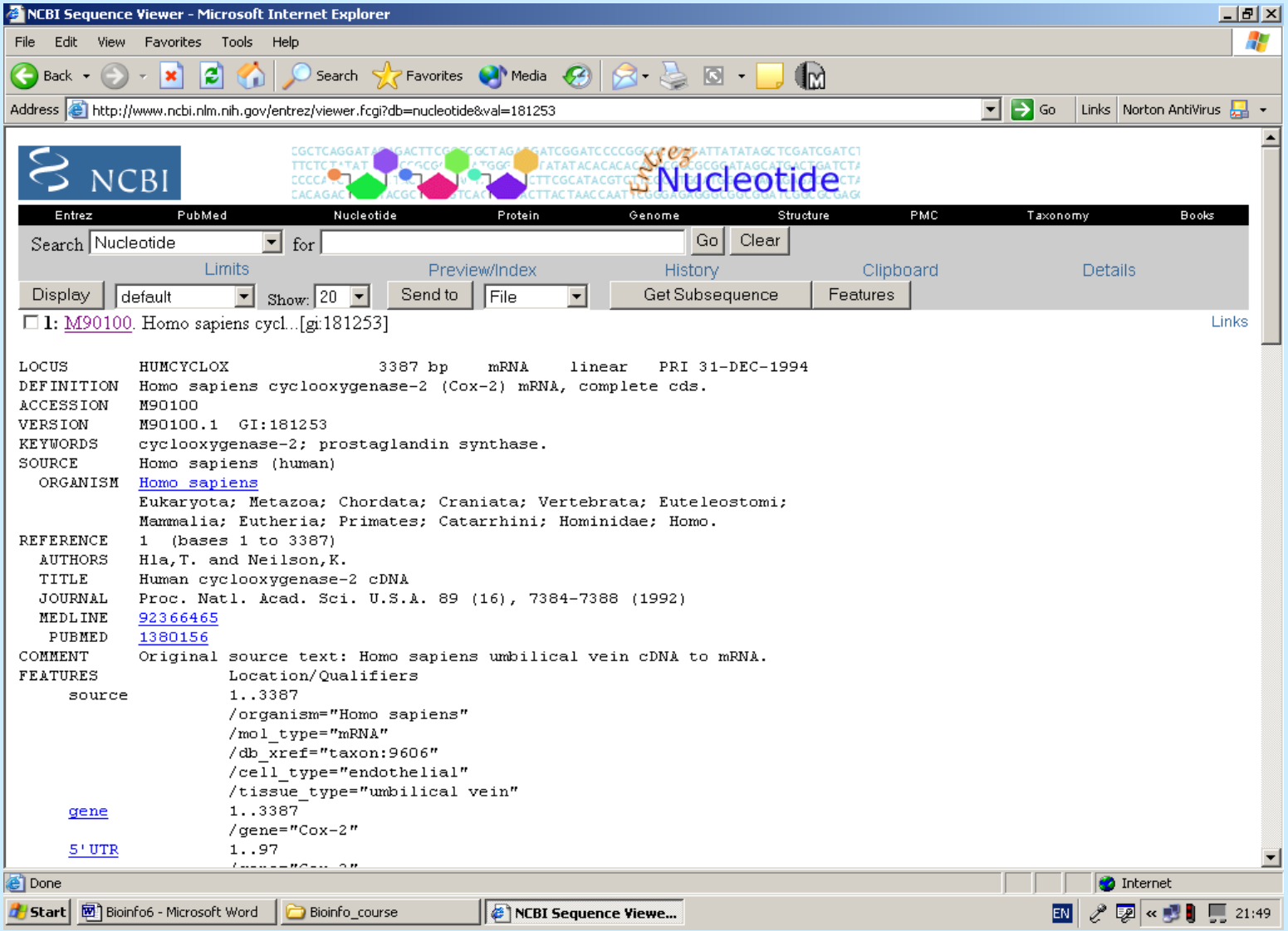
- Για να ανακτήσετε πληροφορίες για το HUMCYCLOX, πληκτρολογήστε τον Accession Code στο πλαίσιο παρακάτω και επιλέξετε “GO”.

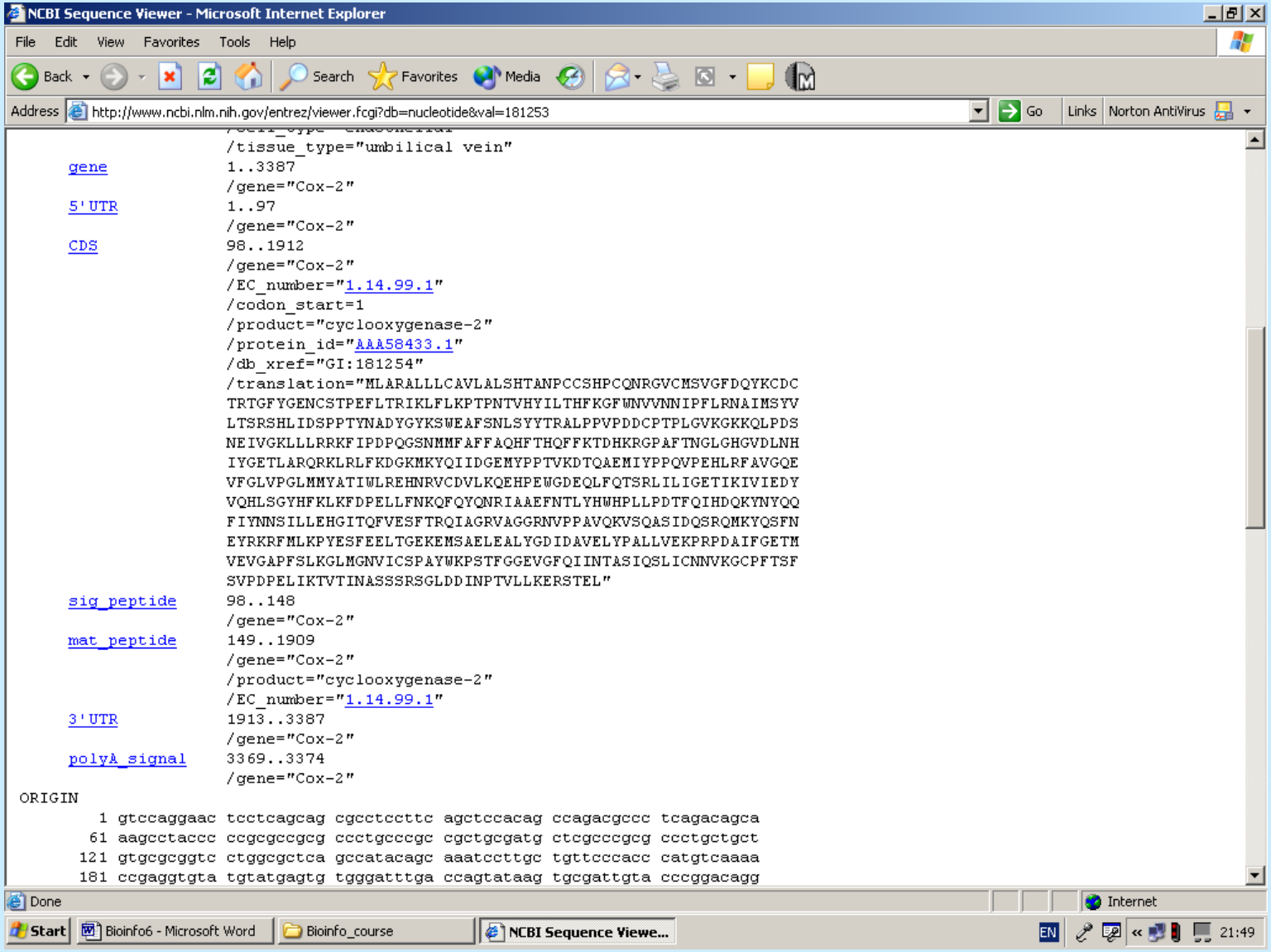


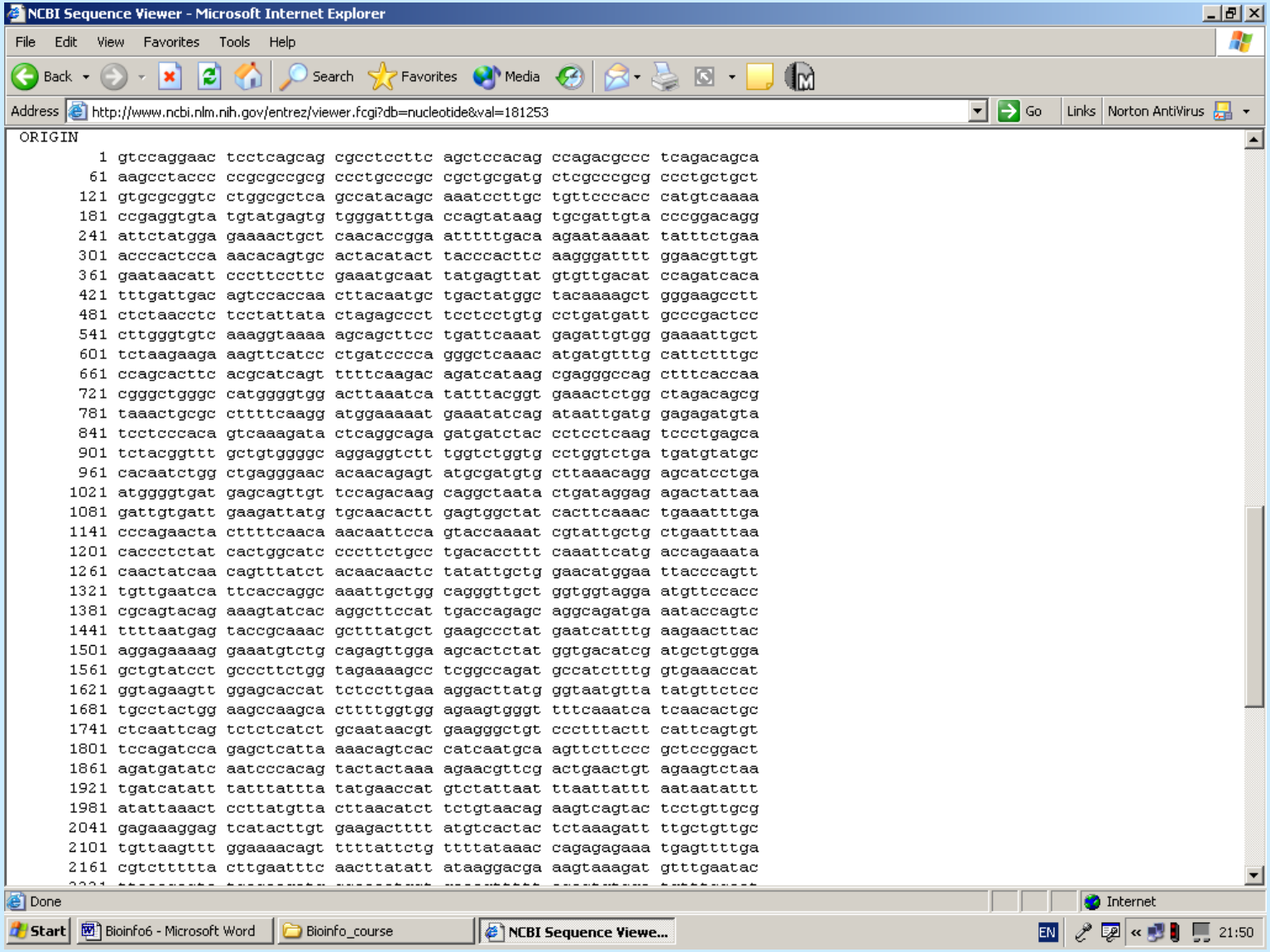
Στη συνέχεια επιλέξετε “M90100”, τον accession code για Homo sapiens cyclooxygenase-2 (Cox-2) mRNA.

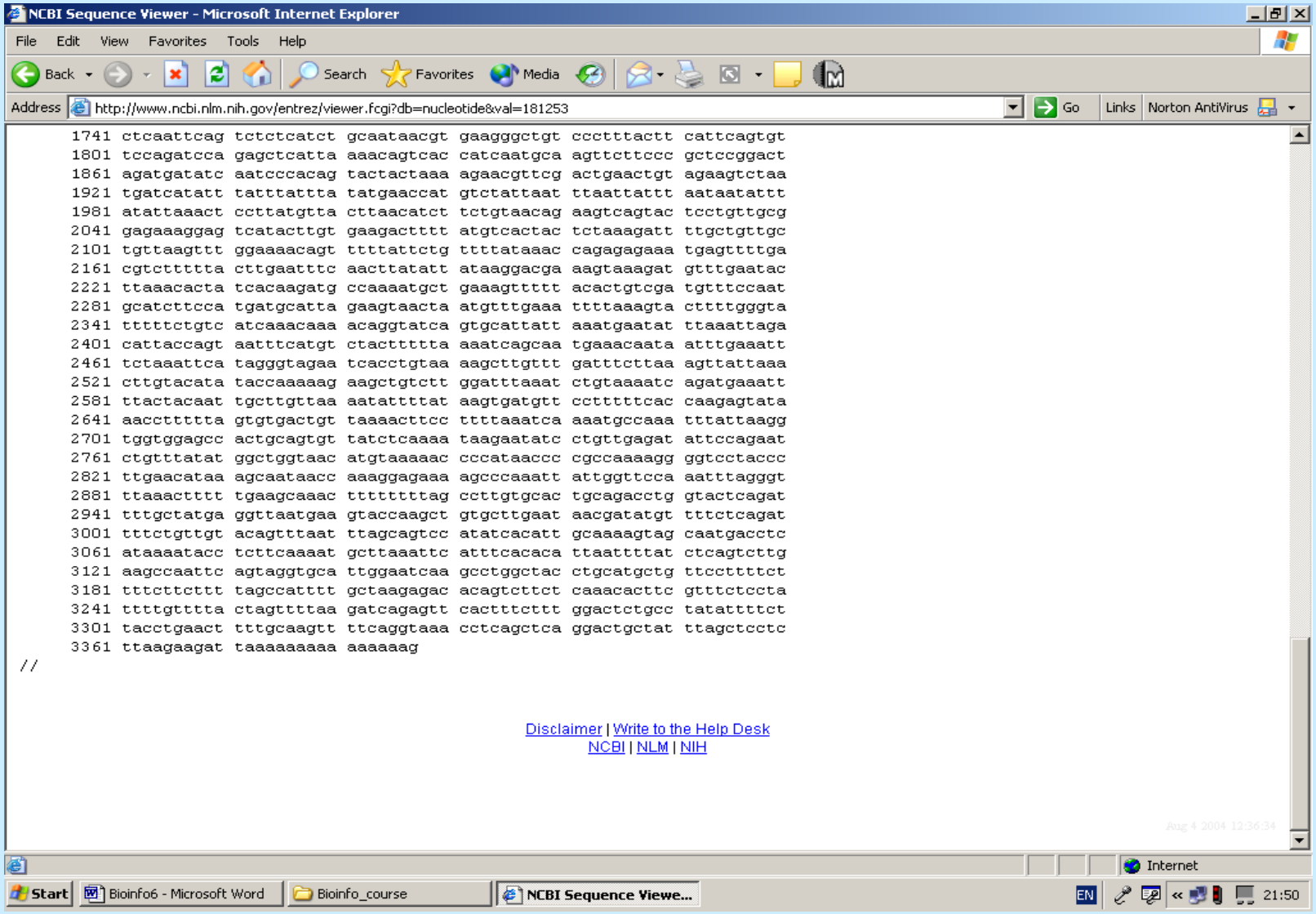


- Το σύστημα θα παρουσιάσει την καταχώρηση. Μπορείτε να κυλήσετε την οθόνη προς τα κάτω για να δείτε διάφορες πληροφορίες καθώς και την αλληλουχία.

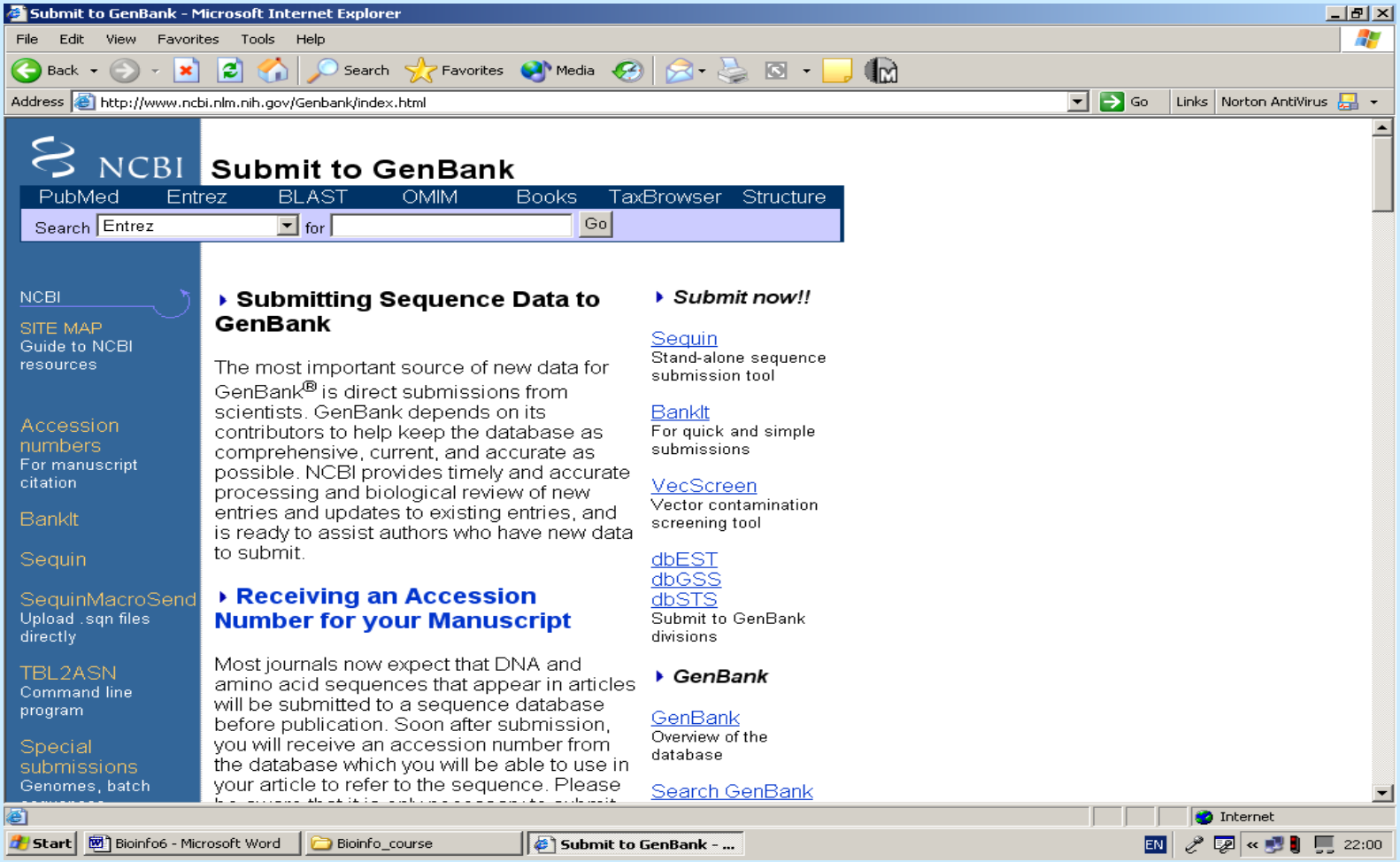



NCBI Sequence Viewer - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=181253
Go Links Norton AntiVirus
/tissue_type="umbilical vein"
gene 1..3387
/gene="Cox-2"
5'UTR 1..97
/gene="Cox-2"
CDS 98..1912
/gene="Cox-2"
/EC_number="1.14.99.1"
/codon_start=1
/product="cyclooxygenase-2"
/protein_id="AAA58433.1"
/db_xref="GI:181254"
/translation="MLARALLLCAVLALSHTANPCCSHPCQNRGVCMSVGFDQYKCDC
TRTGFYGENCSTPEFLTRIKLFLKPTPNTVHYILTHFKGFWNVVNNIPFLRNAIMSYV
LTSRSHLIDSPPTYNADYGYKSWEAFSNLSYYTRALPPVPDDCPTPLGVKGKKQLPDS
NEIVGKLLLRRKFIPDPQGSNMMFAFFAQHFTHQFFKTDHKRGPAFTNGLGHGVDLNH
IYGETLARQRKLRLFKDGKMKYQIIDGEMYPPTVKDTQAEMIYPPQVPEHLRFAVGQE
VFGLVPGLMMYATIWLREHNRVCDVLKQEHPEWGDEQLFQTSRLILIGETIKIVIEDY
VQHLSGYHFKLKFDPELLFNKQFQYQNRIAAEFNTLYHWHPLLPDTFQIHDQKYNYQQ
FIYNNSILLEHGITQFVESFTRQIAGRVAGGRNVPPAVQKVSQASIDQSRQMKYQSFN
EYRKRFMLKPYESFEELTGEKEMSAELEALYGDIDAVELYPALLVEKPRPDAIFGETM
VEVGAPFSLKGLMGNVICSPAYWKPSTFGGEVGFQIINTASIQSLICNNVKGCPFTSF
SVPDPELIKTVTINASSSRSGLDDINPTVLLKERSTEL"
sig_peptide 98..148
/gene="Cox-2"
mat_peptide 149..1909
/gene="Cox-2"
/product="cyclooxygenase-2"
/EC_number="1.14.99.1"
3'UTR 1913..3387
/gene="Cox-2"
polyA_signal 3369..3374
/gene="Cox-2"
ORIGIN
1 gtccaggaac tcctcagcag cgcctccttc agctccacag ccagacgccc tcagacagca
61 aagcctaccc ccgcgccgcg ccctgcccgc cgctgcgatg ctcgcccgcg ccctgctgct
121 gtgcgcggtc ctggcgctca gccatacagc aaatccttgc tgttcccacc catgtcaaaa
181 ccgaggtgta tgtatgagtg tgggatttga ccagtataag tgcgattgta cccggacagg
Done
Internet
Start
Bioinfo6 - Microsoft Word
Bioinfo_course
NCBI Sequence Viewe...
EN
21:49

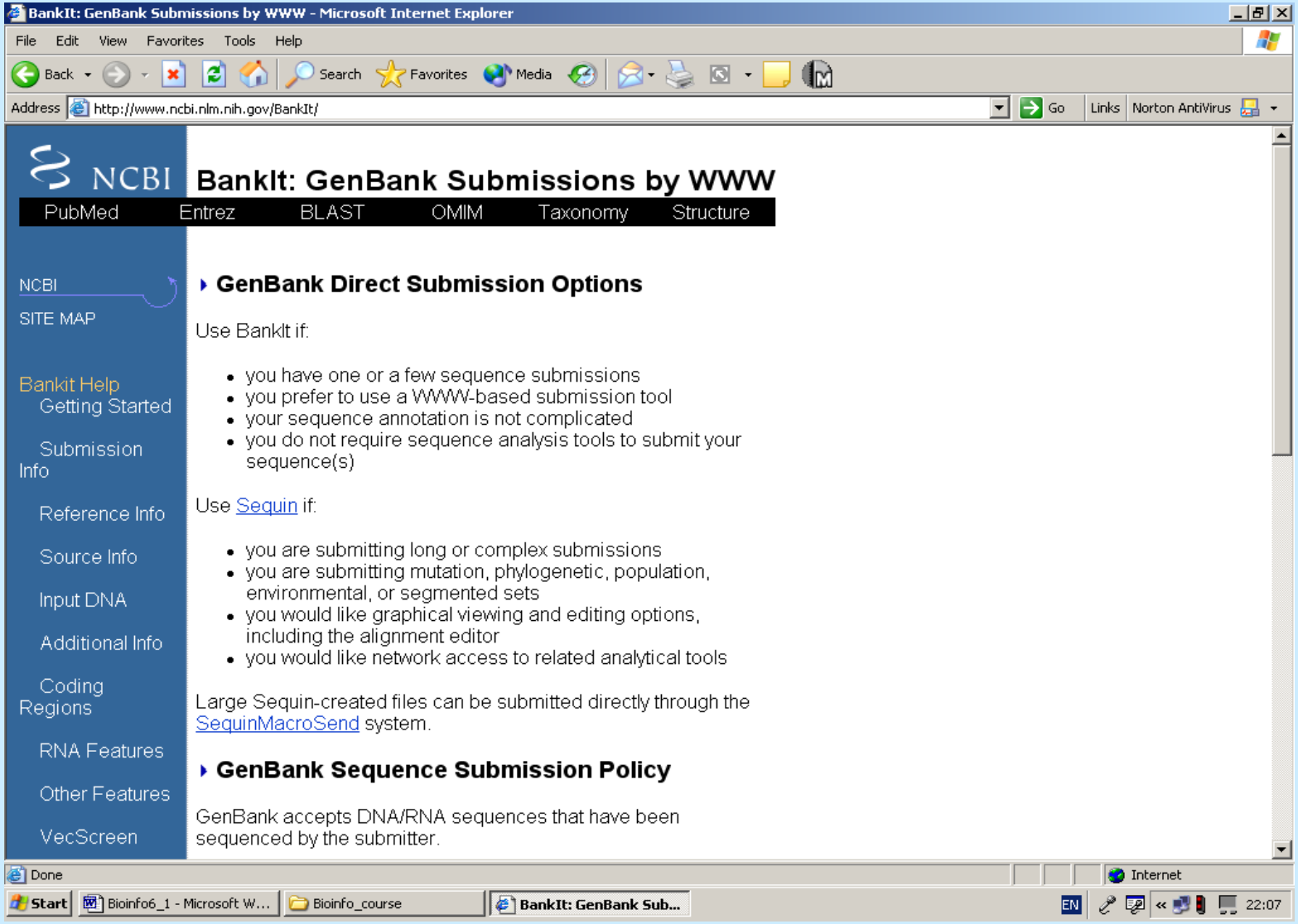

NCBI Sequence Viewer - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=181253
Go Links Norton AntiVirus
ORIGIN
1 gtccaggaac tcctcagcag cgcctccttc agctccacag ccagacgccc tcagacagca
61 aagcctaccc ccgcgccgcg ccctgcccgc cgctgcgatg ctcgcccgcg ccctgctgct
121 gtgcgcggtc ctggcgctca gccatacagc aaatccttgc tgttcccacc catgtcaaaa
181 ccgaggtgta tgtatgagtg tgggatttga ccagtataag tgcgattgta cccggacagg
241 attctatgga gaaaactgct caacaccgga atttttgaca agaataaaat tatttctgaa
301 acccactcca aacacagtgc actacatact tacccacttc aagggatttt ggaacgttgt
361 gaataacatt ccctttcctc gaaatgcaat tatgagttat gtgttgacat ccagatcaca
421 tttgattgac agtccaccaa cttacaatgc tgactatggc tacaaaagct gggaagcctt
481 ctctaacctc tcctattata ctagagccct tcctcctgtg cctgatgatt gcccgactcc
541 cttgggtgtc aaaggtaaaa agcagcttcc tgattcaaat gagattgtgg gaaaattgct
601 tctaagaaga aagttcatcc ctgatcccca gggctcaaac atgatgtttg cattctttgc
661 ccagcacttc acgcatcagt ttttcaagac agatcataag cgagggccag ctttcaccaa
721 cgggctgggc catggggtgg acttaaatca tatttacggt gaaactctgg ctagacagcg
781 taaactgcgc cttttcaagg atggaaaaat gaaatatcag ataattgatg gagagatgta
841 tcctcccaca gtcaaagata ctcaggcaga gatgatctac cctcctcaag tccctgagca
901 tctacggttt gctgtggggc aggaggtctt tggtctggtg cctggtctga tgatgtatgc
961 cacaatctgg ctgagggaac acaacagagt atgcgatgtg cttaaacagg agcatcctga
1021 atggggtgat gagcagttgt tccagacaag caggctaata ctgataggag agactattaa
1081 gattgtgatt gaagattatg tgcaacactt gagtggctat cacttcaaac tgaaatttga
1141 cccagaacta cttttcaaca aacaattcca gtaccaaaat cgtattgctg ctgaatttaa
1201 caccctctat cactggcatc cccttctgcc tgacaccttt caaattcatg accagaaata
1261 caactatcaa cagtttatct acaacaactc tatattgctg gaacatggaa ttacccagtt
1321 tgttgaatca ttcaccaggc aaattgctgg cagggttgct ggtggtagga atgttccacc
1381 cgcagtacag aaagtatcac aggcttccat tgaccagagc aggcagatga aataccagtc
1441 ttttaatgag taccgcaaac gctttatgct gaagccctat gaatcatttg aagaacttac
1501 aggagaaaag gaaatgtctg cagagttgga agcactctat ggtgacatcg atgctgtgga
1561 gctgtatcct gccccttctgg tagaaaagcc tcggccagat gccatctttg gtgaaaccat
1621 ggtagaagtt ggagcaccat tctccttgaa aggacttatg ggtaatgtta tatgttctcc
1681 tgcctactgg aagccaagca cttttggtgg agaagtgggt tttcaaatca tcaacactgc
1741 ctcaattcag tctctcatct gcaataacgt gaagggctgt ccctttactt cattcagtgt
1801 tccagatcca gagctcatta aaacagtcac catcaatgca agttcttccc gctccggact
1861 agatgatatc aatcccacag tactactaaa agaacgttcg actgaactgt agaagtctaa
1921 tgatcatatt tatttattta tatgaaccat gtctattaat ttaattattt aataatattt
1981 atattaaact ccttatgtta cttaacatct tctgtaacag aagtcagtac tcctgttgcg
2041 gagaaaggag tcatacttgt gaagactttt atgtcactac tctaaagatt ttgctgttgc
2101 tgttaagttt ggaaaacagt ttttattctg ttttataaac cagagagaaa tgagttttga
2161 cgtctttttа cttgaatttc aacttatatt ataaggacga aagtaaagat gtttgaatac
Done
Internet
Start Bioinfo6 - Microsoft Word Bioinfo_course NCBI Sequence Viewe...
EN 21:50


NCBI Sequence Viewer - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=181253
Go Links Norton AntiVirus
1741 ctcaattcag tctctcatct gcaataacgt gaagggctgt ccctttactt cattcagtgt
1801 tccagatcca gagctcatta aaacagtcac catcaatgca agttcttccc gctccggact
1861 agatgatatc aatcccacag tactactaaa agaacgttcg actgaactgt agaagtctaa
1921 tgatcatatt tatttattta tatgaaccat gtctattaat ttaattattt aataatattt
1981 atattaaact ccttatgtta cttaacatct tctgtaacag aagtcagtac tcctgttgcg
2041 gagaaaggag tcatacttgt gaagactttt atgtcactac tctaaagatt ttgctgttgc
2101 tgttaagttt ggaaaacagt ttttattctg ttttataaac cagagagaaa tgagttttga
2161 cgtcttttta cttgaatttc aacttatatt ataaggacga aagtaaagat gtttgaatac
2221 ttaaacacta tcacaagatg ccaaaatgct gaaagttttt acactgtcga tgtttccaat
2281 gcatcttcca tgatgcatta gaagtaacta atgtttgaaa ttttaaagta cttttgggta
2341 tttttctgtc atcaaacaaa acaggtatca gtgcattatt aaatgaatat ttaaattaga
2401 cattaccagt aatttcatgt ctacttttta aaatcagcaa tgaaacaata atttgaaatt
2461 tctaaattca tagggtagaa tcacctgtaa aagcttgttt gatttcttaa agttattaaa
2521 cttgtacata taccaaaaag aagctgtctt ggatttaaat ctgtaaaatc agatgaaatt
2581 ttactacaat tgcttgttaa aatattttat aagtgatgtt ccttttcac caagagtata
2641 aacctttta gtgtgactgt taaaacttcc ttttaaatca aaatgccaaa tttattaagg
2701 tggtggagcc actgcagtgt tatctcaaaa taagaatatc ctgttgagat attccagaat
2761 ctgtttatat ggctggtaac atgtaaaaac cccataaccc cgccaaaagg ggtcctaccc
2821 ttgaacataa agcaataacc aaaggagaaa agcccaaatt attggttcca aatttagggt
2881 ttaaacttt tgaagcaaac tttttttag ccttgtgcac tgcagacctg gtactcagat
2941 tttgctatga ggttaatgaa gtaccaagct gtgcttgaat aacgatatgt tttctcagat
3001 tttctgttgt acagtttaat ttagcagtcc atatcacatt gcaaaagtag caatgacctc
3061 ataaaatacc tcttcaaaat gcttaaattc atttcacaca ttaattttat ctcagtcttg
3121 aagccaattc agtaggtgca ttggaatcaa gcctggctac ctgcatgctg ttccttttct
3181 tttcttcttt tagccatttt gctaagagac acagtcttct caaacacttc gtttctccta
3241 ttttgtttta ctagttttaa gatcagagtt cactttcttt ggactctgcc tatattttct
3301 tacctgaact tttgcaagtt ttcaggtaaa cctcagctca ggactgctat ttagctcctc
3361 ttaagaagat taaaaaaaaa aaaaaag
//
Disclaimer | Write to the Help Desk
NCBI | NLM | NIH
Internet
Start Bioinfo6 - Microsoft Word Bioinfo_course NCBI Sequence Viewe...
EN 21:50

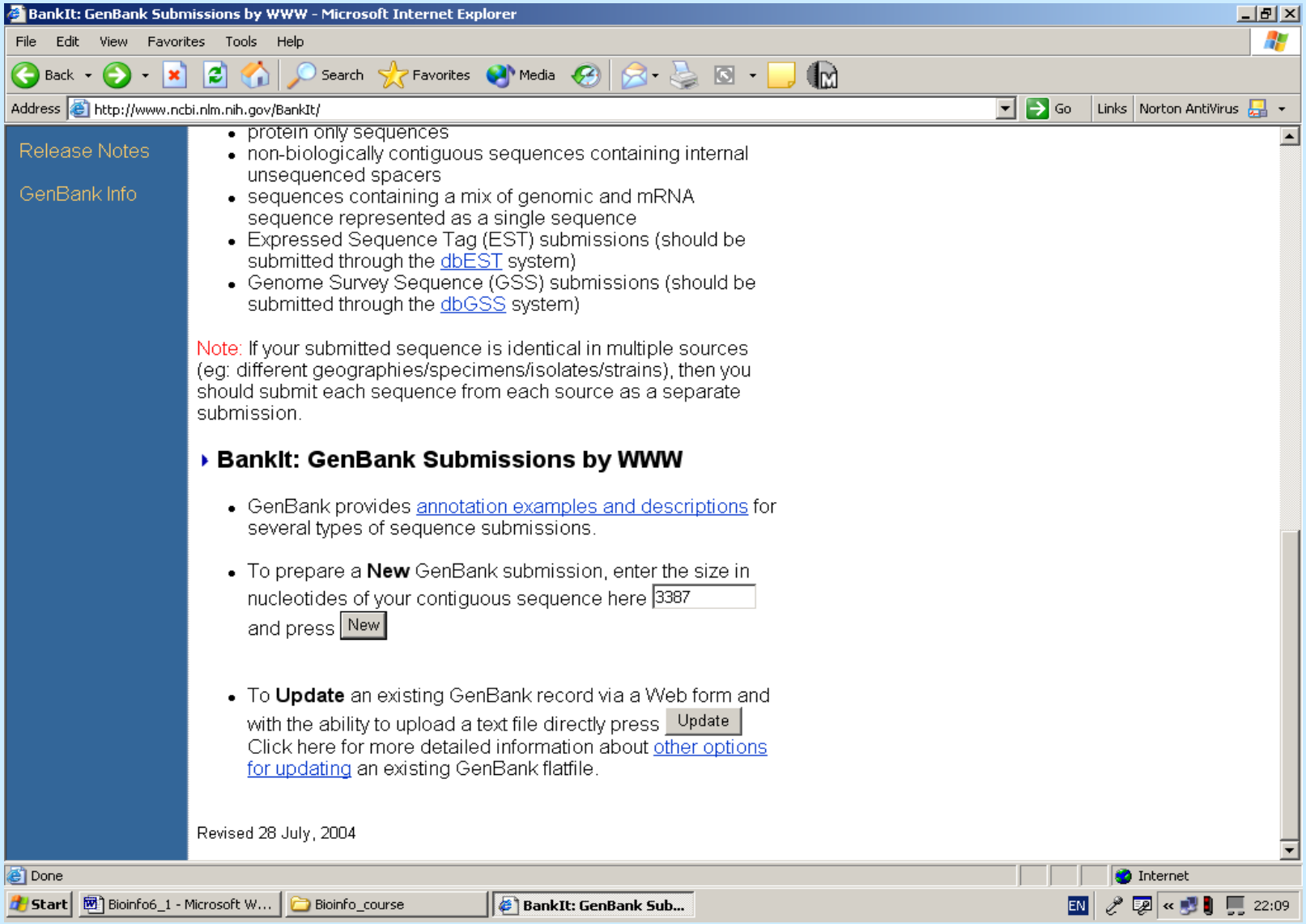
- Για να καταχωρήσετε μια νέα ακολουθία ή να ενημερώσετε μια ήδη υπάρχουσα, επιλέξετε “Submit to GenBank” στην ιστοσελίδα παρακάτω. Στην συνέχεια ακολουθήσετε τις οδηγίες που εμφανίζονται. Το “BankIt” είναι ένα εργαλείο που μπορείτε να χρησιμοποιήσετε για να καταχωρήσετε εύκολα και απλά μια νέα αλληλουχία. Η δομή της καταχώρησης θα πρέπει να μοιάζει με αυτή της HUMCYCLOX.



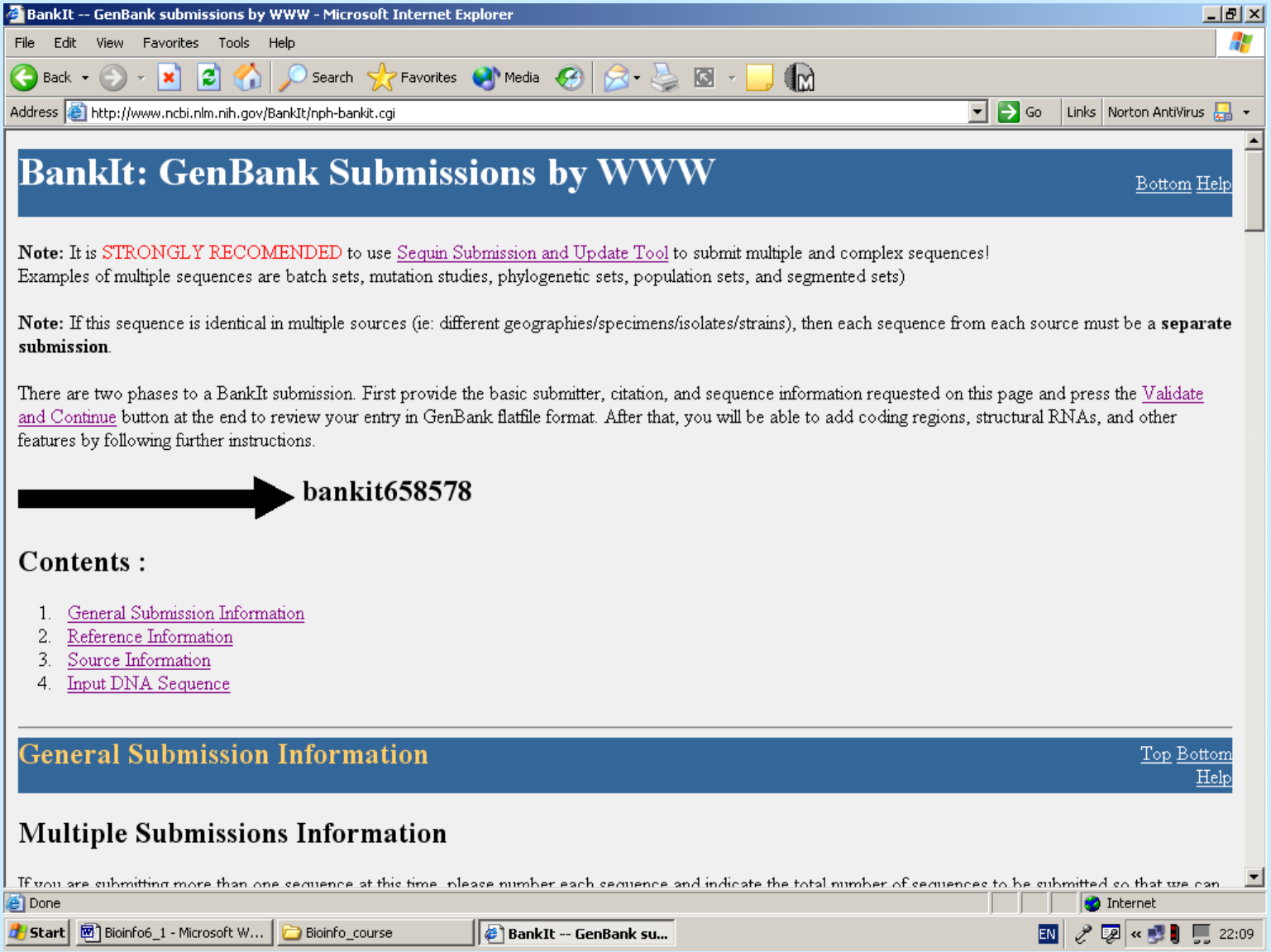
- Επιλέξτε “BankIt”.



- Καταχωρίστε το μέγεθος της αλληλουχίας, π.χ. 3387, και επιλέξτε “New”. Για ενημέρωση επιλέξτε “Update”.



- Κυλήστε την οθόνη προς τα κάτω για να δείτε όλες τις πληροφορίες που χρειάζονται για να καταχωρήσετε την αλληλουχία.



- Αφού έχετε συμπληρώσει όλα τα απαιτούμενα πεδία στην ιστοσελίδα, πληκτρολογήστε στο πεδίο παρακάτω την αλληλουχία, και επιλέξτε “Validate and Continue”.