

Appendix B: Format of the FASTA file

Sample of Fasta file containing Pyrosequenced Reads¹⁶⁸:

```
>FYRUZ7J01BX62B rank=0000730 x=682.0 y=1713.0 length=48
TTCATAATCAAGGTAATTAGAAGTGGTCAACAGGAGATGGCAAGAGTG
>FYRUZ7J01A1I23 rank=0001179 x=310.0 y=1357.0 length=40
ATGCTGCTGCTGCTGCTGCTGCTAAGTCGCTTCAGTCATG
>FYRUZ7J01DYUWR rank=0001219 x=1510.0 y=361.0 length=55
TTAGCATGGATGCATCTAGATCAGTGCTTCTCAATCTTGACGTCAGATTATGG
AC
>FYRUZ7J01CBMWB rank=0001288 x=835.0 y=2073.0 length=54
TTCATGCTGTGCATCAGTTGCAGGTCGTCGCGACAGCGAACGCGACGCTG
AA
>FYRUZ7J01BHYB5 rank=0001511 x=497.0 y=1667.5 length=50
TTCCTCTTAGCTCAGTTGGTAAGAATCTACCTGCAATGACAGAGACTTGG
>FYRUZ7J01BWLTF rank=0001639 x=664.0 y=1249.0 length=50
CACAGTATATGAGAGTTGACACATATGGAGTAAGACACAGAGTAGTGTTG
>FYRUZ7J01B3YZA rank=0001782 x=748.0 y=836.5 length=105
GTACTATGTATGTGGCGTGTCTGTAGCGTGTGGCTTGGCTGTAGTATGTATGT
GGCAGTGTCTGTAGTGTGTGGCGTGTCTGTGCGCATGTAGCCGTGTCTGTCAT
>FYRUZ7J01DWXCE rank=0002575 x=1488.0 y=316.0 length=47
ATGGCTGCAGTCACCATCTGCAGTGATTGTTGGAGTCTACAACACTC
```

¹⁶⁸ Courtesy

(n.d.).

Steen, A. S. (2003). *Distributed Systems : Principles and Paradigms*. New Jersey: Prentice-Hall.

(Site) (LectureNotes) (<http://www.genome.gov/10001772>) (NCBI) (<http://www.genome.gov/10001772>) (<http://www.genome.gov/10001772>) (Site) : Anand Agricultural University

>FYRUZ7J01B3MEI rank=0002790 x=744.0 y=920.5 length=43
ATCTATATACTGATGAGATGAGACTACACACACGTCTCGGTAC

>FYRUZ7J01CK34I rank=0002918 x=943.0 y=1936.0 length=44
ACAGTCACAGAGTCAGACATGATGTAGTGACTGACACACACATT

>FYRUZ7J01DYUUN rank=0002991 x=1510.0 y=285.0 length=58
TTGGTGCAATGGAAGTGGCCGTCATCTCGAGTTGAGGCAGGAACCTCAGGGT
TCTGCT

>FYRUZ7J01DUZSB rank=0003179 x=1466.0 y=281.0 length=49
GTAATGTAGACTTGAATCTTACAACTAGTCACACTTGCAGATAGGAGTA

>FYRUZ7J01DKJ7E rank=0003577 x=1347.0 y=952.0 length=43
ATGCTATATAGATCAGATCAGTCACTGATACAGATTCACTTNT

>FYRUZ7J01B2ZU8 rank=0003621 x=737.0 y=386.0 length=46
CATATACACACATGGACGTATACTACAGCGTACGAGTAGATATAGA

>FYRUZ7J01BSZ6H rank=0003929 x=623.0 y=1175.0 length=65
TTACTATGTTACAATCTAACAACAGCAACAAGATAGACGGAGCGCGAATAAG
GCAATGGGAGTAC

>FYRUZ7J01A1PTO rank=0003965 x=312.0 y=1898.5 length=181
CATACACACACATATACACACACACACACACACACACACACACTACTACAC
ACA
CAC
ACACACACACACACACTACTAC

>FYRUZ7J01B11G2 rank=0004085 x=726.0 y=868.0 length=78
TTTAATGCAGATACGCAACTCCATTCTACTAAGATTAGCTGGCTCAACGGTAAG
AAGCTGTGTCTTCCTATTTCTGACC

>FYRUZ7J01D60MQ rank=0004286 x=1603.0 y=96.0 length=38
CTCATCGCGCATGCGCTGACGCACGCACGACGCGCGCG

>FYRUZ7J01DX5GF rank=0004377 x=1502.0 y=141.0 length=42

TTCATCTATCGTGATGAGGTGTATCACGAGAACAGCGATCTG
>FYRUZ7J01B3S66 rank=0004397 x=746.0 y=1536.0 length=45
TTTGCTGATGCGCTCCGCCAGCTGTTCCAACGCCTGCGCAATGCG
>FYRUZ7J01CRVAZ rank=0004779 x=1020.0 y=1705.0 length=119
CACTCACACACACATACACACTCACACACACACGGACAGACAGACACACACAT
ATAGAGACACACAATACACACTCACACACACGGACAGACAGACACACACAGA
GACACACAATACCA
>FYRUZ7J01C3CH5 rank=0004899 x=1151.0 y=635.5 length=84
GTTTCGACATCGGCACAGAGGCGATAGCCATAGCGCAACTTGTCCAGCGGTTC
GATAGTGATCAACCGGCATTTCGCGCCGCGACG

**Sample of Fasta file containing Contigs Assembled using Pyrosequence
Reads:**

>contig00001 length=156 numreads=30
GCCACCCgaGAAAGACCTCGAGAGTCCTTCTTCAACGCGTCTTGAGGCCCTAG
TCaCTTCCCATGACTCGAGAGTGATGACCAGCTCCCCcTCgCCATGCGCATGGA
GACATGAgtTCCCTGGCaCCGCACGAGAGGCTCCCTGAGGTCCTCGTCG
>contig00002 length=392 numreads=4
GAGTCCTTCTTCAACGCGTCTcGaGgCCCTAGTCACCTCCcGTGACTCgAgAGCC
AGgACGcGCTCCCCcTcGCCATgCGCATGGaGACCTGACTtCCCTGgCGCCGCac
GAGAGGCTCCcTGAGATCCTAGtCgtGCCTCGGgaGAAAACCCACGGGCGTC
ACAGCTCGAAGATACCCCGGAGACGCCcACGTCCTCGGGAGCTGAGGGCCTT
CTTtCCTGCATGGCCTGGAGAGCAATCCCGGGTCcTCTCTCCAAaCGGAGGAG
GAGGCTGGACTCCCTGgAGGCCGCTCAGGGGGGctCCAAGAGACCCGCGTCGC
GACTCGAGAGGAGAGCGGAGtCCTTGGCTTCCCcTCGAGACAAGGCCTGACT
CCTCAGGTGAGCCTGGAATG

>contig00003 length=104 numreads=361

AGTGTTCcAAACTGCCAAGGGCTACGGCAATGCGAGCTCTGTCAGAGGTAA
GAAAAGAGCAATTGGTTGGATTCTTCAGCTGCTGCGTGGCTGCACAGCTTCC

>contig00004 length=441 numreads=294

TTGGAGCcAGaaTATtCCTGAaaGACTTAGGCTCACCAaaGCTGTCCTCCAGAG
CAAaGACATTGGATGTGGTTTTtGCCTTCCCTTGTCAGTATGGGTCCCCTAGCTT
GGCCCCAGCTCCGGGGATACAGCCTCTGAAAAAATCTTCCAAGAAGCGATTTT
CCTCCAGGCTGTTAAGCGGAATTCGCAGTTTGCAAGTTAAGGAAGCCTGCTTC
GTGGAAGAGCAAGCAGCCAAGTGGCCCTGTGTTAGGAAGCCTGTATTTGAAA
GGAATGGGTAGCTTCTACAGCGGGAGAACCTCAGATCTGCTCAGGGGgCTG
AAATACTAGAGTCAGGGTCGAGGTTCTCCCTTGTCCTTTGCGCACCTTGCAAA
GACATTCCTAGCTTCTAGCGTTAGTGTCCCTGTAGTAGCAGGCTGCAGGAGCG
TGTGATAAATGACTGCC

>contig00005 length=245 numreads=38

AaaCcACGTGCTTCCCTCGTCATCGCAAGATGAGGCCCTTTGTCCGCTACAGCA
TCTCAAGAGAAGTCCCACGTTCCATCCTGAAGTGCGAAACGGTACTTGGCACC
CTTGATGCGACCCcAAAagTTCCGCGACATACCAGACTCACTCAAGGGGAACA
CCGAGGTTCCCGGCACCACTTCATCTGAgCCCcTTCTCCTCTCCTGATTGCGAC
AGGAGGGTCGACTCCCCTTCTCTGTCTGCAA

>contig00006 length=414 numreads=338

TCATTTGAGCCCCTTCTCCCCTCCTGATTGCAACAGGAGTGTCgaTTtCTCCTGC
TTTGTgTGGAAGGGGTACCGATCTTCCCAGCGCACaTCAGGATGAGGCcGT
CTCACGAGGAAATTGGAGATGTAGCCTCGTGGGTGGTGCCACATGCGGAAA
GACCCCATTTCCcGGTCaGCTATTAATAAGAACCCGATGCCCCGTACACTTCTTG
gAATGCATCCCGTaGGATGAAGGCACAACAAGGAGGGGaCCCTGAaCACCCGT
GCATCTTTTGAAAAATTCTGCAGGTTCCACATACAGCTCGACAAGTGACCTGT

TACCTCGTGAACAACCTCGAGAGGCAAGCTAAGTTTCATTTCTCCACAGAATAC
GAGGCCTGACTCTCCTGTCCCAACTCTGCAGGAATCCTGAGAT

>contig00007 length=118 numreads=18

GGCcTCAGCTCcGCTATTGCAtGTTcATCTGAACTTGAACCCAGGCCTACTTGG
AGAAGTGTTtGTGTGCCGATTTCCAGAACTGCGAAGGGCTACAGCAATGAT
AACTCTCTCAGT

>contig00008 length=156 numreads=27

GAGTTATCATTGCTGTAGCCCTTCGCAGTTTCTGGAAATCGGCACACAAACAC
TTCTCCAAGTAGGCCTGGGTTCAAGTTCaGATGAACATGCAATAGCGGAGCTG
AGGCCAGTAAGGCACCAAGTCAGGTGTCACACACTCgggCAGcctGCTCCT

>contig00009 length=855 numreads=130

TTGTCAAACATCCTTGGTACGCTAGGGATAGCTTTCTCTGTGGAAGGTGGAAG
CCTCCCTTCCATCTCTCGGTGCAGACAGGCAAACCGACCACAACTCAAAGCAA
CGGTGTTTGTGCCACGTGGGGCATACTTCCACGGTCTCTACTGATGGACACA
CCACCCACTTCACCGAGCACATCCGAGGAGCCTTCACCCAAAGCTCTTGCAAA
ATTCTTCAGAACTCTCAGTCTTCCAGTCACCTGCCACCATCAGGCGAGGTGTA
GAGAACCGACGGAGAACTGGAGAAATGACTTCTTCCGTGGGGTTGGATGA
ACTGAGGAACAGGATGCTGAAGTGGATGCCTGACGTCTTGAGGAAAAGAGG
ATGGGGCGATGGGGGgTGGGGCGGGGgAGGAAGGGCAGGACGGTCCACTT
AGGAGGGCTGGTTGAACATGGACACATAAACTGTACATCCAAAGAATCCACA
AGCACCTACTCTATACTATGCCAGCACGGGAAAATAGACGTATGACGTTGCAT
GAACCTCTAAGGGAAAAGACTCTGAACACCTGAAACGAACACGGCTTTGGAA
ACCCACCCTAGACTTCATTTCTGCCGTGAGGGTGCTGTCGTGTGCATAgCTGA
GGGGATTGATATGTCTCCCGGCCATCTGGATTCCAGCTTGTGCTTCATCCAGC
CCAGCGCTTCTCATGATGTGTACTCAGCGTGTACGTTCAATAAGCAGGGTGAC
GATGGACAGCCTTGATGGACTCCTTTCCCTTATTTGGAACCAGCCTGTTGGGT
TCCATGTCTAGTTCTAACTGTTGCTTCCTGACCTGCAGAAaTaCAGATTCC

Sample of Fasta file containing Chromosome

(The following is the sample of the fasta file storing Chromosome X of Bos Taurus (Cow). The content displayed here is Less than 0% of the actual file):

```
>gi|194719540|ref|NC_007331.3|NC_007331 Bos taurus chromosome
X, reference assembly (based on Btau_4.0), whole genome shotgun
sequence
ATTCTCCAGGCCAGAATACTGGGATGGGTAGCCTTTCCTCCTCCAGGTGATC
TTCCCAACCCAGGAATCGAACCCAGGTCTCCTGCATTGCAGGTAGATTCGTTA
CCAGCTGAGACACAAGGGAAGCCCAAGAGTATTGGAGTGGGTAGCCTATCCC
TTCTCCAGCGGATTTTCCCAATCCAGGAATCAAACCTGGGGTCTCCTGCATTGC
AGGTGGATTCTTTACCAACTGAGCCACAAGGGAAGCCCCTCACAAATATTGGT
CCTATGAAAATCAGCTCCCTTGCA GTTACAGAAAGGAGCAAATTGTATATAAA
TTTCTCAAAAATTTCCCATTCCTGGAGCATTCTGTTTCTAGAGCTTTGACACT
GTTTGACCTGTTTCATAGCTCCCTAGAAAAATCCCACTCCAAACCTGTTATTAT
TAGAATCTGACTCAGCACTTGCTTTGCAAAACAGCCCTGTCCCCAGGTCATTTG
TCAAAAACAATCAGTGGCAATTGTTTAACACCTTAGTTGCTTGAGGCAGCAAT
AACAGTTGGAGCCAAAAAGAGGCTAACCAAGAAGCTAAAAAAAAACAATGA
TGTGGGGGGGAAAAAAGACAAATATAGGGGAATTTGAAAAGCTCTGAAATGT
TCCCAGGCAATCTAAAAGGCCCTGTGCATGTGCAGAGTTGTGTGTATGCTCAG
GAAATACCTGAGAGAAAGCCAGTCCTTCATCTCTGGCTGGCCTTGAGGCTTTA
TTCAAGCAGGCTATGCAGATGTAAACTACCTGGCTGAGTGTTGCAGACATGCT
CCAACCTGCACAGAACGTTCCCCTTGGTAAAGACTGGGAGATACACTGGATCC
AAGTTTTTAAGGAAATCTTAGCTCAATCATTAGCTTACCAGTAACCTGTGCAAT
GTGCTCAGTTTCTCAGTCATATCTGACTCTTTGCATCCCCACGGACTATAATCC
ACTAGGCTCCACTGTCCATGGAATTTTCCAAGCAAGAATACTGGAGTGGGTTG
```

CCATTTCTACTTTAGGGGATCTTCCTGAACCAGGGATCAAACCCTCATCTCCT
GAGTCTCCTGCATTGGCCGACAACTTCTTTACGTATGTGCCACCTGGGAAGCC
CCGACCAATAACCTAGCAAGGCAAAATTCACATTGGCCGCACACGCACATTCA
GTAAGAAATACAGACTTTACAGAATTAGCCCAGAAAAGTCATTAAACAAGGA
GAAGCAAAAGTAACAAAAAAGAATAGTTAATAACAAATCATAAGAAAGTGA
GATCTGATTTCAAGAGTTGCCACATTATCTAAAAGGCCTGCTGCTGCTGCTGC
TGCTAAGTCCCTTCAGTCATGTCCGACTCTGTGCAACCCCATAGACGGCAGCC
CACCAGGCTTCCCTGTCCCTGGGATTCTCCAGGCAAGAACTGGAGTGGTTT
GCCATTTCTTCTCCAATGCATGAAAGTGAAAAGTGAAAGTGAAGCCACTCAG
TCGTGTCCGACTCCCAGCGACCCCATGGACGGCAGGCTACCAGGCTCCTCCAT
CCACGGAATTTTCCAGGCAAGAGTACTGGAGTGGGGTGCCATTGCCTTCTCCA
TCTAAAAGGCCTAGTTTTCAACAAAGAATTATTAGATACACATAGAACTGGA
AGATAGGTTTATACACAGGAAAAAAAAAAGCAGTTAGTAGAAATTATTCCT
GAGGGGGCCCAGATGTTGACCTTTCTATACAAAGACTTTAAATCATTTACTAT
GAATATATTCAAAGAACTAAAATAAATCATGTCTAAATAATTAAACCAAGTAA
GATAATGATGTTTCAACAAAGGGATAATATTAGTAGAAATTATTTTAAAAAGA
AACTAAATAGAAATTCTAGAGTTGAAAAGTGAAATATTTGAAATAAGAAATG
TACAAGAGAGATTACACTCTACCAACACTTATCTTTTGTCATCTTGATAACAG
TCATCCTAACAGGTATAAGGTAACATTGCATTGATGTTTAAATTTGCATTTCCC
TAGGGAAATGCAGTGAGCACCTTTTCATATACCCATTGGCCATGTGTATATAT
TCTTTGGAAAGATGTCTATTCAACTTTTTGTACTTAGATCCTGAATGAATTTCTT
ATCTATTTTGTATATTAACCCCTTATGATATTTGTGAGTTGCCCTGTACCAAATC
CAAATAAAAACATCACAAGAAAAATATAGCCCAATATTTCTGTGAATATAGT
TGTA AAAACCCTCAACAAAATACCAACAACTAAATTCAGCAACATATTAAGT
ATATTATACATCATGAGAAAGTGAGATTTATCCCAGGATCCAAGGGTGGCTCA
ATATACAAAAATAAAATAATATGCTGCATTAACAAAATGAAGGAAACAAATCA
CATGATCATCTCAAAATTGATACAGAAAAGGCATCTGAGAAATCCAACATTTT

ATTATAAAAATCAAATTAGGAATTGAAAGGAACTTCCTTAGTATAATAAAGGA
TAACATATGAAAAACCCATGGCTGATACTGTACTGAATGGTGAAAACCTAAAA
ATTTTCTTCTAAGATCAGGAAGAAGAGAATCAACTCCTTTCCTACTTCTATT
CAACATTGTACTGTAACTAGAGCAATTAGGCAATAAAAAACCTAAAGGTATCCA
AATTGGAAAGGAAGAATTAAAATTATCTCTATTTACAGATGGTATGATCTCAT
ATATAGCAATACAAAAAATCTGCACACATCAAAGAAACACTTATGGCTAATA
AATGAACACAGCATATTTGCAGAATGTAAGATCAACATGGAGCTTCCCTGGCA
GCTCATGTGGTAAAGAATCTGCCTGCATTATAAGAGACCCTGGCTTGATTCTT
GGTCAGGAAGATCCCCTGGAGAAGGGATAGGCTACCCACTCCAGTATTCTTG
GGCTTCTCTGGTAGCTCAGTCAGTAAAGAATCCACCTGCAATGAGGGAGACC
TGGGTTCAATCCCTGGGTTGGGAAGATCCCCTGGAGGGGGCAAGCCACTCCAG
TATTCTTGACTGGATAATCCCGATGGACAGAGGAGCCTGGTGGGCGCAGTC
CATGGGGTCACAAAGAGTCAGACATGACTGAGCAACTAAGCACACACACACA
ACATCAACGTGCACAAAAAATAGCTTTCTTTCTATACACTAGCAACAAACATTC
CAGACAGGAAATTGAAAAGGAACTTCATTAAATTAGCATCAAAAGAATAAAA
ATACTGGGGATAAATTGAACCAGGGA.....