Database dùng để thử nghiệm có tên drosoph.nt là bộ gen loài ruồi giấm(Drosophila), loại gen là nucleic protein và theo định dạng FASTA (có cá định dạng khác FASTA, GenBank, EMBL, Swiss-Prot, GFF, SCF, ACE, SAM, PDB rồi mày dẫn link hoặc giải thích cho người khác biết FASTA là gì).

Database drosoph.nt có 1170 subject, và chứa tổng cộng 122,655,632 Nucleotide.

File database là file text có cấu trúc như sau

>gi|10726333|gb|AE003845.2|AE003845 Drosophila melanogaster genomic scaffold 142000013385266 section 1 of 3, complete sequence

CTGGAAGTTCCATGTAACTACAACCTTCGCCTGATATTCATGTGTATTAATCTCGTAGAAATATGGTGTTTTTTCGCAAA

CTACAATTTTATTTGACTCAACGTGAGCAGGGAGTAGCATCTGAGCAGCTTCAATATGTATTGTACATAAAAAACCAGCA

TGTGCGCTCTTGGGTTTTGGTAAATTTTCAATCTCTAGCCGAATTTCTATTGGCACTCTTACGGGTAGAAGTATTTCCGG

ACGATTTTTTTTTAAGTGGGGGCAGTGACCAACAGTACTTACAGCATTTTCTATGTTACGGCATTGCAAAGACTTGTGAA

CGCACTTATTATCAAAAATACACCAGTTGCAACCCCAGGAACTTTGTAGACACTCCTGGCAATTTCCATGATGTGAGCAG

TCAAAAAAAGCAAAATTCCTTGATACAAAATCCTTATTTGTCTCCGAACTTCTTACTGACAATGGCACCAAAATATGATC

Trong đó:

* Header của subject là 1 dòng bắt đầu bằng ký tự “>”. Sau ký tự “>” là các định danh của header kết thúc định danh đó là một khoảng trắng.
  + | dùng để phân tách các trường trong header.
  + Gi: xác định trường số maggie. Số maggie sau “gi” dùng nhận dạng chuỗi từ NCBI[tham khảo vào]. Trong ví dụ trên là số 10726333.
  + Gb: đánh dấu trường mã định danh trong GenBank. Mã đi sau gb là mã trong GenBank. Trong ví dụ trên là AE003845, .2 là version của mã trong genbank
  + sau khoảng trắng (space) chính là thông tin mô tả của chuỗi subject
* Sau header của subject là nội dung(payload) của subject. Nội dung được tách ra bằng dấu xuống dòng hoặc dấu cách. Nội dung của subject kết thúc khi bắt đầu subject mới hoặc hết file

<http://repository.lib.ncsu.edu/ir/bitstream/1840.16/6934/1/etd.pdf>

### [11\_blastplus.pdf (616 KB) - BRIC](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0ahUKEwjr4PiwrbHNAhVJMo8KHfmjA8EQFggoMAE&url=http%3A%2F%2Fwww.ibric.org%2Fmyboard%2Fview.php%3FBoard%3Dnews%26filename%3D11_blastplus.pdf%26id%3D256333%26fidx%3D1&usg=AFQjCNFngNb9RX6x5dt8c5ImikzuLWdSLg&sig2=8I_MLVuQ98bOmOkWzcnWyA)