



Decompressing Data Obtained from dbSNP's FTP Site

Created: July 7, 2005; Updated: February 18, 2014.

I am using gunzip -c (on UNIX) to decompress files retrieved from the dbSNP FTP site, but when the decompressed file in the buffer reaches 2 GB, the process dies. How do I get around this problem?

We have no problems running gzip 1.2.4 on UNIX. More than likely, your file system doesn't allow files larger than 2 Gb. For now, you can use a Batch query to request a range of SNPs with a total size not exceeding 2 Gb, because the average SNP size in the XML format is about 11 K. You could also try querying [Entrez SNP](#) to get a subset, and then retrieve the XML report using the batch query service or directly from Entrez. Once you get the query result, select Display and then select dbSNP Batch Report; then select Display and XML. There are also [Entrez Programming Utilities](#) that provide access to Entrez data and report retrieval in batches. We have plans for re-engineering the dbSNP system to allow for smaller data chunks to be downloaded. This feature should be available in the beginning of 2005.

I have tried expanding a file that I downloaded from the dbSNP FTP site using both PKZIP and GZIP, but both applications report an unexpected EOF condition. Why?

Please make sure that you are using PKZIP or GZIP capable of handling files over 2 Gb. The uncompressed size of the file in question is about 6.5 Gb. I had no problems uncompressing it with 64bit GZIP on a Solaris SPARC.