

File	Description	Size (bytes)
linux	The Linux Kernel 2.6.31 tar file	365 711 360
HS-ch1	Homo-sapiens chromosome 1, from NCBI 36.49	251 370 600
ecoli	The file E.coli of the large Canterbury corpus	4 638 690
bible	The file bible.txt of the large Canterbury corpus	4 047 392
world	The file world192.txt of the large Canterbury corpus	2 473 400
a2M	The letter 'a' repeated two million times.	2 000 000

Canterbury corpus can be found at <http://corpus.canterbury.ac.nz/descriptions/#cantrbry>, while the FASTA files for the NCBI 36.49 DNA human genome are downloadable at ftp://ftp.ensembl.org/pub/release-49/fasta/homo_sapiens/dna/