

Local path /data/genome/human/illumina/ERR317482WGS/tmp/9827_2#49.sam
File format SAM
Organism Homo sapiens
Misc

Summary

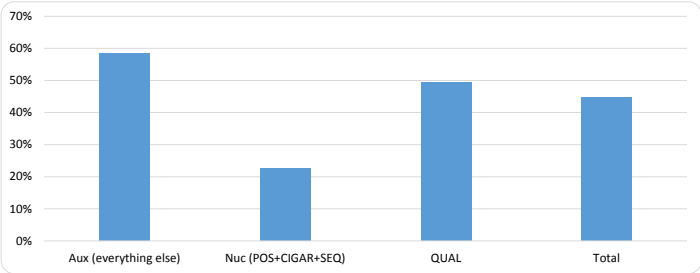
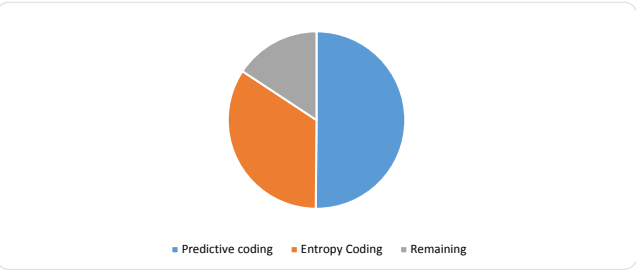
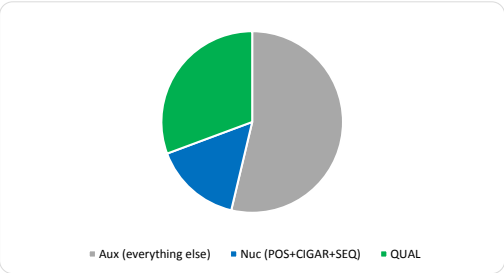
	File size (B)	File size (GiB)	Ratio	Factor	Time (us)	Time (h)	Parameters	Misc	Platform specification
Original File	360427371	0,34	100,00%	1,00					
Tsc	161678231	0,15	44,86%	2,23	741570270	0,21	FILECODEC_BLK_LC 10000 NUCCODEC_WINDOW_SZ 10		Intel® Xeon® CPU E5405 @ 2.00 GHz, 39GiB RAM, openSUSE 12.3 (Dartmouth) (x86_64)

Data distribution

	SAM size (B)	SAM size (GiB)	Ratio	Tsc size (B)	Tsc size (GiB)	Ratio	Compression ratio	Compression factor	Method
File format	0	0,00	0,00%	5628	0,00	0,00%			
SAM header	29279	0,00	0,01%	29279	0,00	0,02%	100,00%	1,00	
Aux (everything else)	148387525	0,14	41,17%	86824841	0,08	53,70%	58,51%	1,71	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Nuc (POS+CIGAR+SEQ)	112021467	0,10	31,08%	25260964	0,02	15,62%	22,55%	4,43	Predictive coding: local sequence alignment with short-time memory reference, entropy coding: 0-order arithmetic coder
QUAL	99989100	0,09	27,74%	49557519	0,05	30,65%	49,56%	2,02	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Total	360427371	0,34	100,00%	161678231	0,15	100,00%	44,86%	2,23	

Tsc timing statistics

	Time (us)	Time (h)	Ratio
Total	741570270	0,21	100,00%
Predictive coding	371774640	0,10	50,13%
Entropy Coding	253233893	0,07	34,15%
Remaining	116561737	0,03	15,72%
Misc	0	0,00	0,00%



Local path /data/genome/bacteria/DH10B/tmp/MiSeq_Ecoli_DH10B_110721_PF.sam
File format SAM
Organism Bacteria
Misc

Summary

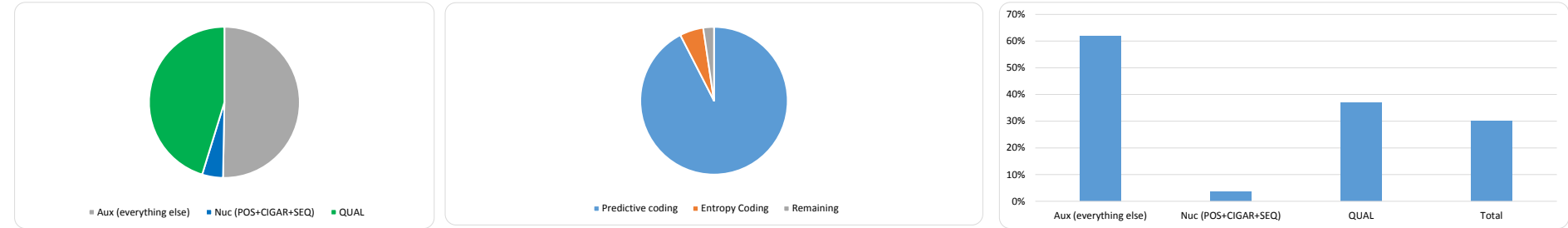
	File size (B)	File size (GiB)	Ratio	Factor	Time (us)	Time (h)	Parameters	Misc	Platform specification
Original File	409375634	0,38	100,00%	1,00					
Tsc	122698208	0,11	29,97%	3,34	1056081405	0,29	FILECODEC_BLK_LC 20000 NUCCODEC_WINDOW_SZ 10		Intel® Xeon® CPU E5405 @ 2.00 GHz, 39GiB RAM, openSUSE 12.3 (Dartmouth) (x86_64)

Data distribution

	SAM size (B)	SAM size (GiB)	Ratio	Tsc size (B)	Tsc size (GiB)	Ratio	Compression ratio	Compression factor	Method
File format	0	0,00	0,00%	5628	0,00	0,00%			
SAM header	197	0,00	0,00%	197	0,00	0,00%	100,00%	1,00	
Aux (everything else)	99699744	0,09	24,35%	61711510	0,06	50,30%	61,90%	1,62	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Nuc (POS+CIGAR+SEQ)	159676293	0,15	39,00%	5485868	0,01	4,47%	3,44%	29,11	Predictive coding: local sequence alignment with short-time memory reference, entropy coding: 0-order arithmetic coder
QUAL	149999400	0,14	36,64%	55495005	0,05	45,23%	37,00%	2,70	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Total	409375634	0,38	100,00%	122698208	0,11	100,00%	29,97%	3,34	

Tsc timing statistics

	Time (us)	Time (h)	Ratio
Total	1056081405	0,29	100,00%
Predictive coding	976340483	0,27	92,45%
Entropy Coding	54922435	0,02	5,20%
Remaining	24818487	0,01	2,35%
Misc	0	0,00	0,00%



Local path /data/genome/human/illumina/ERP001960/tmp/NA12878_S1.sam
File format SAM
Organism Homo sapiens
Misc

Summary

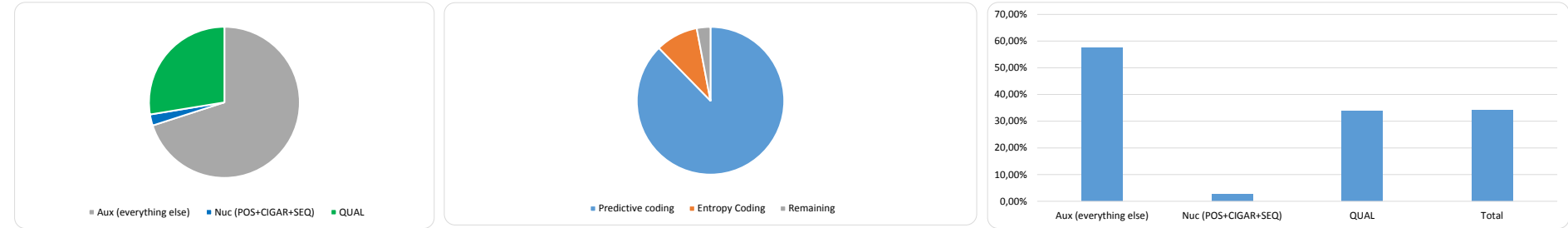
	File size (B)	File size (GiB)	Ratio	Factor	Time (us)	Time (h)	Parameters	Misc	Platform specification
Original File	360371641	0,34	100,00%	1,00					
Tsc	123395952	0,11	34,24%	2,92	66820930381	18,56	FILECODEC_BLK_LC 20000 NUCCODEC_WINDOW_SZ 10		Intel® Xeon® CPU E5620 @ 2.40 GHz, 94GiB RAM, openSUSE 13.1 (Bottle) (x86_64)

Data distribution

	SAM size (B)	SAM size (GiB)	Ratio	Tsc size (B)	Tsc size (GiB)	Ratio	Compression ratio	Compression factor	Method
File format	0	0,00	0,00%	5628	0,00	0,00%			
SAM header	696	0,00	0,00%	696	0,00	0,00%	100,00%	1,00	
Aux (everything else)	150486207	0,14	41,76%	86417617	0,08	70,03%	57,43%	1,74	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Nuc (POS+CIGAR+SEQ)	108887465	0,10	30,22%	2948748	0,00	2,39%	2,71%	36,93	Predictive coding: local sequence alignment with short-time memory reference, entropy coding: 0-order arithmetic coder
QUAL	100997273	0,09	28,03%	34023263	0,03	27,57%	33,69%	2,97	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Total	360371641	0,34	100,00%	123395952	0,11	100,00%	34,24%	2,92	

Tsc timing statistics

	Time (us)	Time (h)	Ratio
Total	66820930381	18,56	100,00%
Predictive coding	58572428778	16,27	87,66%
Entropy Coding	6243013640	1,73	9,34%
Remaining	2005487963	0,56	3,00%
Misc	0	0,00	0,00%



Local path

/data/genome/human/RNAseq/tmp/K562_cytosol_LID8465_TopHat_v2.sam.short.sam

File format

SAM

Organism

Homo sapiens

Misc

Short SAM file containing 100,000 lines; run with original file produced seg fault

Summary

	File size (B)	File size (GiB)	Ratio	Factor	Time (us)	Time (h)	Parameters	Misc	Platform specification
Original File	31167813	0,03	100,00%	1,00					
Tsc	13593736	0,01	43,61%	2,29	30943206	0,01	FILECODEC_BLK_LC 20000 NUCCODEC_WINDOW_SZ 10		Intel® Xeon® CPU E5-2690 v2 @ 3.00 GHz, 126GiB RAM, openSUSE 13.1 (Bottle) (x86_64)

Data distribution

	SAM size (B)	SAM size (GiB)	Ratio	Tsc size (B)	Tsc size (GiB)	Ratio	Compression ratio	Compression factor	Method
File format	0	0,00	0,00%	588	0,00	0,00%			
SAM header	3576	0,00	0,01%	3576	0,00	0,03%	100,00%	1,00	
Aux (everything else)	15057730	0,01	48,31%	9124121	0,01	67,12%	60,59%	1,65	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Nuc (POS+CIGAR+SEQ)	8513043	0,01	27,31%	1712317	0,00	12,60%	20,11%	4,97	Predictive coding: local sequence alignment with short-time memory reference, entropy coding: 0-order arithmetic coder
QUAL	7593464	0,01	24,36%	2753134	0,00	20,25%	36,26%	2,76	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Total	31167813	0,03	100,00%	13593736	0,01	100,00%	43,61%	2,29	

Timing statistics

	Time (us)	Time (h)	Ratio
Total	30943206	0,01	100,00%
Predictive coding	26625257	0,01	86,05%
Entropy Coding	2793129	0,00	9,03%
Remaining	1524820	0,00	4,93%
Misc	0	0,00	0,00%

