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	21058600749	19,61	100,00%	1,00					
Tsc	9769943603	9,10	46,39%	2,16	741570270	0,21	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%	135604	0,00	0,00%			
SAM header	29279	0,00	0,00%	29279	0,00	0,00%	100,00%	1,00	
Aux (everything else)	9765924270	9,10	46,37%	4927623278	4,59	50,44%	50,46%	1,98	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Nuc (POS+CIGAR+SEQ)	5646323600	5,26	26,81%	2023896048	1,88	20,72%	35,84%	2,79	Predictive coding: local sequence alignment with short-time memory reference, entropy coding: 0-order arithmetic coder
QUAL	5646323600	5,26	26,81%	2818259394	2,62	28,85%	49,91%	2,00	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Total	21058600749	19,61	100,00%	9769943603	9,10	100,00%	46,39%	2,16	

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%

Tsc data distribution

Tsc timing statistics

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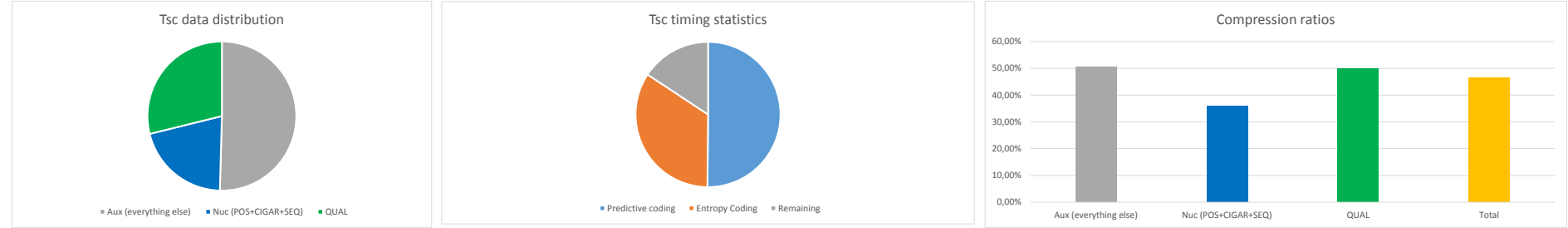
Local path	/data/genome/bacteria/DH10B/tmp/MiSeq_Ecoli_DH10B_110721_PF.sam
File format	SAM
Organism	Bacteria
Misc	

Summary

	File size (B)	File size (GiB)	Ratio	Factor	Time (us)	Time (h)	Parameters	Misc	Platform specification
Original File	5579036306	5,20	100,00%	1,00					
Tsc	2049681141	1,91	36,74%	2,72	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%	31684	0,00	0,00%			
SAM header	197	0,00	0,00%	197	0,00	0,00%	100,00%	1,00	
Aux (everything else)	1626332409	1,51	29,15%	819975129	0,76	40,01%	50,42%	1,98	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Nuc (POS+CIGAR+SEQ)	1976351850	1,84	35,42%	200821393	0,19	9,80%	10,16%	9,84	Predictive coding: local sequence alignment with short-time memory reference, entropy coding: 0-order arithmetic coder
QUAL	1976351850	1,84	35,42%	1028852738	0,96	50,20%	52,06%	1,92	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Total	5579036306	5,20	100,00%	2049681141	1,91	100,00%	36,74%	2,72	

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%

Tsc data distribution
- Aux (everything else)
- Nuc (POS+CIGAR+SEQ)
- QUAL

Tsc timing statistics
- Predictive coding
- Entropy Coding
- Remaining

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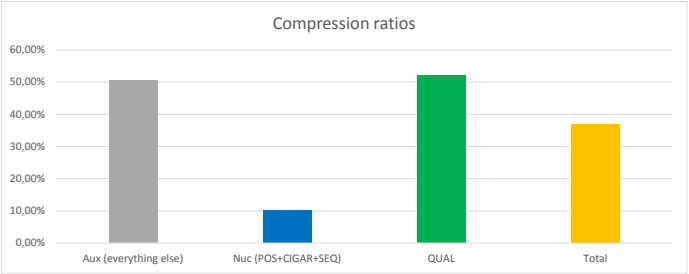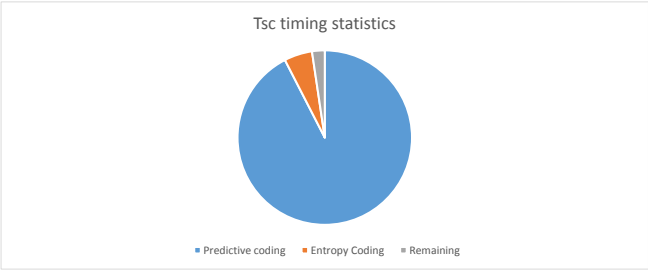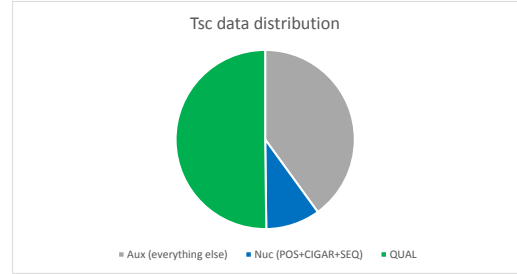
Local path	/data/genome/human/illumina/ERP001960/tmp/NA12878_S1.sam									
File format	SAM									
Organism	Homo sapiens									
Misc										

Summary

	File size (B)	File size (GiB)	Ratio	Factor	Time (us)	Time (h)	Parameters	Misc	Platform specification
Original File	5,89083E+11	548,63	100,00%	1,00					
Tsc	2,3531E+11	219,15	39,95%	2,50	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%	3798724	0,00	0,00%			
SAM header	696	0,00	0,00%	696	0,00	0,00%	100,00%	1,00	
Aux (everything else)	2,69363E+11	250,86	45,73%	1,32472E+11	123,37	56,30%	49,18%	2,03	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Nuc (POS+CIGAR+SEQ)	1,5986E+11	148,88	27,14%	28547578120	26,59	12,13%	17,86%	5,60	Predictive coding: local sequence alignment with short-time memory reference, entropy coding: 0-order arithmetic coder
QUAL	1,5986E+11	148,88	27,14%	74286605592	69,18	31,57%	46,47%	2,15	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Total	5,89083E+11	548,63	100,00%	2,3531E+11	219,15	100,00%	39,95%	2,50	

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%

Tsc data distribution — Aux (everything else), Nuc (POS+CIGAR+SEQ), QUAL

Tsc timing statistics — Predictive coding, Entropy Coding, Remaining

Compression ratios — Aux (everything else), Nuc (POS+CIGAR+SEQ), QUAL, Total
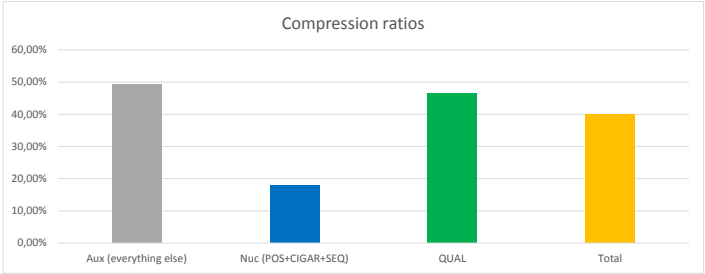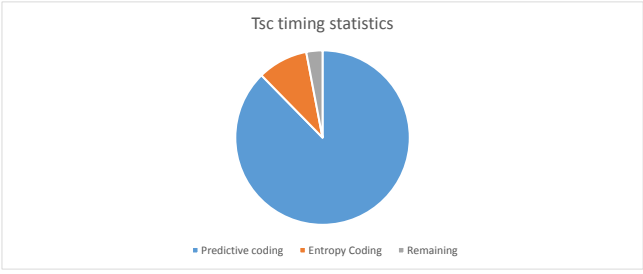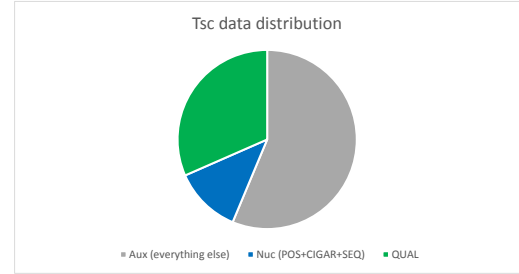
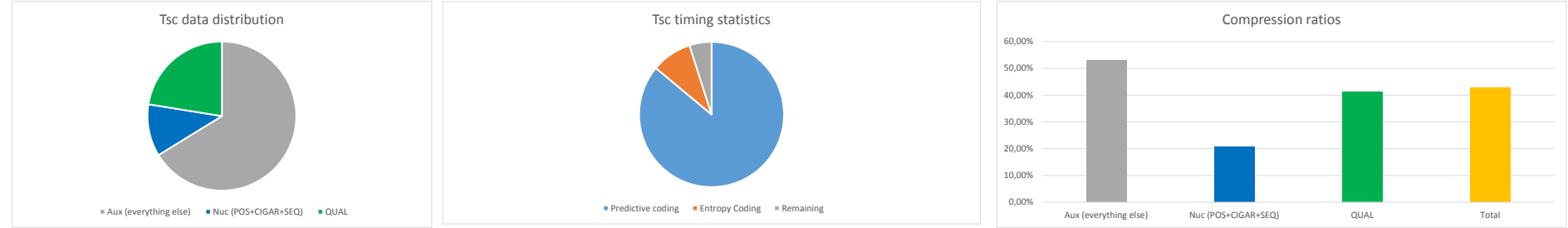|										
Local path	/data/genome/human/RNAseq/tmp/K562_cytosol_LID8465_TopHat_v2.sam.short.sam
File format	SAM
Organism	Homo sapiens
Misc	Short SAM file containing 100.000 lines,run with original file produced seg fault

Summary

	File size (B)	File size (GiB)	Ratio	Factor	Time (us)	Time (h)	Parameters	Misc	Platform specification
Original File	325843519	0,30	100,00%	1,00					
Tsc	139345899	0,13	42,76%	2,34	30943206	0,01	FILECODEC_BLK_LC 20000		Intel® Xeon® CPU E5-2690 v2 @ 3.00 GHz, 126GiB RAM, openSUSE 13.1 (Bottle) (x86_64)
							NUCCODEC_WINDOW_SZ 10		

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%	2452	0,00	0,00%			
SAM header	3576	0,00	0,00%	3576	0,00	0,00%	100,00%	1,00	
Aux (everything else)	173853015	0,16	53,35%	92298632	0,09	66,24%	53,09%	1,88	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Nuc (POS+CIGAR+SEQ)	75993464	0,07	23,32%	15736822	0,01	11,29%	20,71%	4,83	Predictive coding: local sequence alignment with short-time memory reference, entropy coding: 0-order arithmetic coder
QUAL	75993464	0,07	23,32%	31304417	0,03	22,47%	41,19%	2,43	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Total	325843519	0,30	100,00%	139345899	0,13	100,00%	42,76%	2,34	

Tsc 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%

Tsc data distribution

Aux (everything else) ■ Nuc (POS+CIGAR+SEQ) ■ QUAL

Tsc timing statistics

Predictive coding ■ Entropy Coding ■ Remaining

Compression ratios

Local path	/data/genome/cancercells/360b4736-6c5e-48df-af58-c1cf51609350/tmp/HCC1954.mix1.n80t20.sam.short.sam
File format	SAM
Organism	Cancer cells
Misc	Short SAM file containing 100.000 lines, run with original file was killed by the OS

Summary

	File size (B)	File size (GiB)	Ratio	Factor	Time (us)	Time (h)	Parameters	Misc	Platform specification
Original File	451835771	0,42	100,00%	1,00					
Tsc	228972746	0,21	50,68%	1,97	27927124	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%	2452	0,00	0,00%			
SAM header	23461	0,00	0,01%	23461	0,00	0,01%	100,00%	1,00	
Aux (everything else)	249834328	0,23	55,29%	147836783	0,14	64,57%	59,17%	1,69	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Nuc (POS+CIGAR+SEQ)	100988991	0,09	22,35%	35001148	0,03	15,29%	34,66%	2,89	Predictive coding: local sequence alignment with short-time memory reference, entropy coding: 0-order arithmetic coder
QUAL	100988991	0,09	22,35%	46108902	0,04	20,14%	45,66%	2,19	Predictive coding: nothing, entropy coding: 0-order arithmetic coder
Total	451835771	0,42	100,00%	228972746	0,21	100,00%	50,68%	1,97	

Tsc timing statistics

	Time (us)	Time (h)	Ratio
Total	27927124	0,01	100,00%
Predictive coding	14163171	0,00	50,71%
Entropy Coding	4323216	0,00	15,48%
Remaining	9440737	0,00	33,80%
Misc	0	0,00	0,00%

Tsc data distribution
- Aux (everything else)
- Nuc (POS+CIGAR+SEQ)
- QUAL

Tsc timing statistics
- Predictive coding
- Entropy Coding
- Remaining

Compression ratios