Author

This exercise was authored and tested by Astrid Rasmussen.

Phred quality score exercise

@read1
ACGTGAGTCTAGCTAGCTAG
+
?<:>>C8<==0@B1?@B

@read2
CGTAGCTAGCTAGCTAGCTA
+
77<<7774:=4.4:=4.

1. Identify the Phred quality scores in read1 and read2 at the top of the page.

2. For each base, convert the ASCII character to a Phred quality score

3. Find the mean quality score of each read

   Note, Phred quality score are log probabilities so simply taking the average of these will give a wrong result. Instead you need to average the error probabilities and calculate the mean Q-score from this.

$$P = 10^{-Q/10}$$

$$Q = -10 \log_{10}(P)$$

ASCII_BASE=33 Illumina, Ion Torrent, PacBio and Sanger

Q	P_error	ASCII	Q	P_error	ASCII	Q	P_error	ASCII	Q	P_error	ASCII
0	1.00000	33 !	11	0.07943	44 ,	22	0.00631	55 7	33	0.00050	66 B
1	0.79433	34 "	12	0.06310	45 -	23	0.00501	56 8	34	0.00040	67 C
2	0.63096	35 #	13	0.05012	46 .	24	0.00398	57 9	35	0.00032	68 D
3	0.50119	36 $	14	0.03981	47 /	25	0.00316	58 :	36	0.00025	69 E
4	0.39811	37 %	15	0.03162	48 0	26	0.00251	59 ;	37	0.00020	70 F
5	0.31623	38 &	16	0.02512	49 1	27	0.00200	60 <	38	0.00016	71 G
6	0.25119	39 '	17	0.01995	50 2	28	0.00158	61 =	39	0.00013	72 H
7	0.19953	40 (	18	0.01585	51 3	29	0.00126	62 >	40	0.00010	73 I
8	0.15849	41)	19	0.01259	52 4	30	0.00100	63 ?	41	0.00008	74 J
9	0.12589	42 *	20	0.01000	53 5	31	0.00079	64 @	42	0.00006	75 K
10	0.10000	43 +	21	0.00794	54 6	32	0.00063	65 A			

(https://www.drive5.com/usearch/manual/quality_score.html)