Fastq file format

$$\underbrace{@\textit{SFHMR}:01926:02303}_{\text{Read ID}}$$

Fastq file format

@*SFHMR* : 01926 : 02303

Read ID

GAGACCCGCATCGCCATCTTGCGCAAGAAAGCACAGATG

DNA sequence

Fastq file format

$\underbrace{@SFHMR : 01926 : 02303}_{\text{Read ID}}$

$\underbrace{GAGACCCGCATCGCCATCTTGCGCAAGAAAGCACAGATG}_{\text{DNA sequence}}$

$\underbrace{+}_{\text{A plus}}$

Fastq file format

$\underbrace{@SFHMR : 01926 : 02303}_{\text{Read ID}}$

$\underbrace{GAGACCCGCATCGCCATCTTGCGCAAGAAAGCACAGATG}_{\text{DNA sequence}}$

$\underbrace{+}_{\text{A plus}}$

$\underbrace{?? >> 3D <; < @@BB9??@; 6 > 999?9CC : DCB >>?BBBB <}_{\text{Nucleotide qualities}}$

Dna sequence G A G A C C C G C A T C
Quality ? ? > > 3 D < ; < @ @ B

Nucleotide G has a quality of ?

Symbol $\rightarrow$ Encoded value $\rightarrow$ Quality score $\rightarrow$ Probability of error

http://www.ascii-code.com/

ASCII printable characters (character code 32-127)

Codes 32-127 are common for all the different variations of the ASCII table, they are called printable characters, represent letters, digits, punctuation marks, and a few miscellaneous symbols. You will find almost every character on your keyboard. Character 127 represents the command DEL.

DEC	OCT	HEX	BIN	Symbol	HTML Number	HTML Name	Description
32	040	20	00100000		 		Space
33	041	21	00100001	!	!		Exclamation mark
34	042	22	00100010	"	"	"	Double quotes (or speech marks)
35	043	23	00100011	#	#		Number
62	076	3E	00111110	>	>	>	Greater than (or close angled bracket)
63	077	3F	00111111	?	?		Question mark
64	100	40	01000000	@	@		At symbol

- Symbol ? is encoded as a decimal value 63
- Symbols used for quality score encoding (Sanger) start with symbol ! (value 33)
- Quality score = ? - ! = 63 - 33 = 30

$p = 10^{-\frac{Q}{10}}$

Our Q = 30, therefore

$p = 10^{-\frac{30}{10}} = 10^{-3} = 0.001$

Probability that nucleotide G is an error $= 0.001$

@2'4EH>>:A;?>>@8??CCBCCLC=BCB<@>>::+:::>>B<>>66*:66;:>D?7</<<::4<:99525.(/(66/)445;5*99:>==7.:,:::AACGCB<BB=F?B=FB@FMD>
@HLUIT:03001:01228
AACACATTCCTGACCCGAAATGCTGTTATTAAATTGGTCATAAAAATTCATGATAAATGCCAACAAAAATAAGTAAGGCATGTTGTTATGGATTTGACCAACCGTATTCATCAATATACTCGTGGGG
+
255---'-'569:/::@5@CCDD>8163-8+>@4<:94388*88884::8=863)449<59584*9999:;A?@;C=CBBBB<B?9??B<??9BB@A<A<A;AAAA;AABBB<:;:BAA<??@3???
@HLUIT:03099:01750
ATAAAAATTCATGATAAATGCCAACAAAAATAAGTAAGGCATGTTGTTATGGATTTGACCAACCGTATTCATCAATATACTCGTGGGGCTTGGACATTCGATGTCATTGATTCAGGCCCTATAGATGGCCCTCCTCGTGGTTTTATTGCATTGGCTTCCAGA
AACTGCCCATAGTTGGAGCAGACTTCGGAATTATTACATCAGCATCAATTTAGAACATTTGCGATCCATCAACGTGGTTATAGTCTAAATGCTAGACC
+
BB4FEFM?DCC@BBA9AB@?9C=DC4ICCCA:@?:2:3FCIDB>B?8?>:2::5@:999B<@;A@@>7?@@AA<EE::.A@>./.@@A>9?:===A>DCDBCCCF?:>>A<BCB=B9B@??>>7652.(...(//<8;66(5555/55:5'-'1-'-'8--(
66663(337:880.(//;<>>?:2999?:?:??<B@@AAAAFCCBC?:3:::@;@AA:AAA@@CE>BCCC>CCFB@B<BBBA?@?F;F@>>DDCCD>E
@HLUIT:03375:01864
GTAATTTACCTGCGGTAACAGGAGCTTCAGATTTTAGATCTTAGGCACGATCACGGCGGTATAAAGCTACTTATTCGATTTAATTGGACCGATTCGACCTCATAACACATTTCCTGACCCGAAATGCTGTTATTAAATTGGTCATAAAAATTCATGATAAAT
GCCAACAAAAATAAGTAAGGCATGTTGTTATGGATTTGACCAACCGTATTCATCAATATACTCGTGGGGCTTGGACATTCGATGTCATTGATTCAGGCCCTATAGATGGCCCTCTCGTGGTTTTAT
+
558>5?@@:?@>>8>9475.(./4-/44><5(88854767/886???>>==:;;::@@????5???@?=:4;:7?=>?6HA:A:@:;6A@@9>;9.'..99=8?;::7(88/?BCG9EEC9CCA@@@>6==:>4>>6>8A@@AA3AA@@:C?@@AB@9AAA
B>E;A:*::::::;@AB=C>NEFMC<C@>CC;7;:<BCAB<B;B<CB@9699CCC=BBCFFFFFBD9FFAB<C>?>AC>BBCCFDABB<CDB;BBC;@6@==?BBB@A;?9???ABA@A;A5AABBB
@HLUIT:01436:01892
AAATGCCAACAAAAATAAGTAAGGCATGTTGTTATGGATTTGACCAACCGTATTCATCAATATACTCGTGGGGCTTGGACATTCGATGTCATTGATTCAGGCCCTATAGATGGCCCTCTCGTGGTTTTATTGCATGGCTTTCCAGAAACTGCCCATAGTTGG
GAGCAGACTTCGGAATTATTACATCAGCATCAATTTAGAACATTTGCGATCCATCAACGTGGTTATAGTCTAAATGCTAGACCACAAAGCCGCTTTCATTACGCTTTA
+
AA9:;;9@::::::*:@:?@@:?9?FFDK?FC@BCC=BCCABBB>D@D=BBBC>CCEDD=BBBCCEBBAAAA5AA<C>CBBC=BDCCC@A?@:>=<6<99/48068<566?E<ED4@@?<?97-666'9B9@?B=82788-:3599:1@A@@@3>?<77.44
'11115557244-1-13895888@A=;<:???:@@7A??:ABAA;BA=??@394;<:>:==:>5=8;888688-44414111.3456-56145;>29--'--46:@5C
@HLUIT:03184:02349
GCCAACAAAAATAAGTAAGGCATGTTGTTATGGATTTGACCAACCGTATTCATCAATATACTCGTGGGGCTTGGACATTCGATGTCATTGATTCAGGCCCTATAGATGGCCCTCTCGTGGTTTTATTGCATGGCTTTCCAGAAACTGCCCATAGTTGGGAGC
AGACTTCGGAATTATTACATCAGCATCAATTTAGAACATTTGCGATCCATCAACGTGGTTATAGTCTAAATGCTAGACCACAAAGCCGCTTTCATTACGCTTTA
+
?@<B;>@@@?2@D>BBB<B<BCDDC=CC=AAA9??B;CDE=D=B=C?>><CBBAE<CBB>><;94667'7=6B=CEJG=CCCCDHBCB=BCF@F?:<@C:CDDCCCBB=CF=A@@>ECC;AAA6AC=CCECC=BCF@H=??@B;BCBCA9CCCAA;AE;CBC
BBBBB;AB;C;B;AA;ACCC@@@CBDCC=CH<C@?;@FFF;AACCCC?CCCB<AAAA<A;??:::CCCC8@@BDDECD;CAAA5?@::;@F:AAA;@@@?@@8?
@HLUIT:02849:02599
AACAAAAATAAGTAAGGCATGTTGTTATGGATTTGACCAACCGTATTCATCAATATACTCGTGGGGCTTGGACATTCGATGTCATTGATTCAGGCCCTATAGATGGCCCTCTCGTGGTTTTATTGCATGGCTTTCCAGAAACTGCCCATAGTTGGAGCAGAC
TTCGGAATTATTACATCAGCATCAATTTAGAACATTTGCGATCCATCAACGTGGTTATAGTCTAAATGCTAGACCACAAAGCCGCTTTCATTACGCTTTAA
+
6>:*::::::;E;:3A;ACBBC>FF=CC@9??9?>::2:7A;CA>>:@@??;:;>>;??BB@3DDDA;B<BB@A9@ABBBBECC=DF?9?>>7:,:::ACBEB=68-4-.::::>@<?/>=::8;>>97.>5(54-888/987;:5>><<..'.'..9<A@?>
:A@:@;@;AA<BBCCBABBDCCB<C<CCCD?EDB<CCCBBAE>DDCB=CCCB;?9??A?ACBB;DD@@>@AAB=BBB;BBC=CC@9???@:@>>C-1881D
@HLUIT:03305:01951
AACAAAAATAAGTAAGGCATGTTGTTATGGATTTGACCAACCGTATTCATCAATATACTCGTGGGGCTTGGACATTCGATGTCATTGATTCAGGCCCTATAGATGGCCCTCTCGTGGTTTTATTGCATGGCTTTCCAGAAACTGCCCATAGTTGGAGCAGAC
TTCGGAATTATTACATCAGCATCAATTTAGAACATTTGCGATCCATCAACGTGGTTATAGTCTAAATGCTAGACCACAAAGCCGCTTTCATTACGCTTTAA
+
9C:*:::::8C::4B<CBBAA=B@:AA:3::9AAA@9@9A;@@?@:@@>>>9CBBBBBC@@?3CCD@<A9??A?:??CFBBCEA;@AB<BBB<:3;::?@>>><69)978>>A>>8:+99857<99<4-43(77-::3(33393(779===592:9DA?A99
8CC=B=@<AA=BBCBBCDCCCCC=B:BBCA:;::+999A???<BBBA<A>@@<C=A>>>AA@@6AAA@@BDCB<BBA:AAA8@BB<ECBE@BBCCC<CC=H