

Bioinformatics : Notes on Pileup format

This page last changed on Aug 23, 2010 by [gramasamy](#).

samtools Pileup, when invoked with -c produces a tab delimited file with 10 columns.

col #	Title	e.g	Explanation
1	Name of the chr/contig	LmjF_chr01	.
2	Reference position	1234	pos in the chr/contig where bases are piled (1 based coordinate)
3	Reference base	ATGCN	ref. base in that position of chr/contig
4	Consensus base	ATGCN*....	can be IUPAC ambiguity codes/
5	Consensus qlty	105	
6	SNP quality	132	
7	Max Mapping qlty	31	
8	Read Depth	20	
9	pileup of reads	.,+1A-1G*\$^	
10	Base quality	ASCII char	

Some confusion & clarification on field 9:

Title	Format	e.g	Remarks
Start of reads	^<map_quality><base>	^I.	
End of reads	<base>\$.\$	
Insertion in Reads	+<#of insertions><inserted bases>	+2AG	this read suggests, there is a two bp insert(AG) compared to ref.seq. The insertion happens at the position next to current position (col#2 +1). Status of the current pos(col#2) is actually presented before '+' symbol.
Deletion in Reads (in following positions)	-<#of deletions><deleted bases>	-3ATG	This suggests, there is a three base pair deletion in this read(when compared with reference).The deletion happens at the position next to current position (col#2

			+ 1). Status of the current pos(col#2) is actually presented before '-' symbol.
Deletion in Reads (current position)	*	*	This indicates, there is a read, aligned overlapping this reference position(col#2), has a deletion in it. Deletion happens at current position (col#2). This is a necessary redundant information. The above line also point to the same information but in a different format.

And there is an extra pileup line for each position with a indel

col #	Title	e.g	Explanation
1	Name of the chr/contig	LmjF_chr01	.
2	Reference position	1234	.
3	Reference base	*	As there can be no reference base.
4	Consensus base	*/*	
.	.	*/T	Not sure about the difference between these two different values. Does not look like a consensus of the in/del position
5	Consensus qlty	105	
6	SNP quality	0	
7	Max Mapping qlty	31	
8	Read Depth	20	
9	pileup of reads	*	
10	Deleted/inserted string with +/- symobl	-TT	
11	#of reads with No del/insertion	26	includes the count of reads with mismatch
12	#of reads with Insertion/Deletion at following pos	1	
13	#of reads with Insertion/Deletion at following pos	2	Its not clear how the order of using 12, 13 for in or del

14	Unknown to me	0
15	Unknown to me	0