[This is a quick reference for **BCF1**, as historically implemented in Samtools. The BCF1 format is obsolete; BCF2 is widely implemented and recommended for use instead of this legacy format.]

Field	Descrption	Type	Value
magic	Magic string	char[4]	BCF\4
l_seqnm	Length of concatenated sequence names	int32_t	
seqnm	Concatenated names, NULL padded	char[l_seqnm]	
l_smpl	Length of concatenated sample names	int32_t	
smpl	Concatenated sample names	char[l_smpl]	
l_meta	Length of the meta text (double-hash lines)	int32_t	
meta	Meta text, NULL terminated	char[l_meta]	
List of records until the end of the file			
seq_id	Reference sequence ID	int32_t	
pos	Position	int32_t	
qual	Variant quality	float	
l_str	Length of str	int32_t	
str	ID+REF+ALT+FILTER+INFO+FORMAT, NULL padded	char[l_str]	
Blocks of data; #blocks and formats defined by FORMAT (table below)			

Field	Type	Description
DP	uint16_t[n]	Read depth
GL	float[n*G]	Log10 likelihood of data; $G = \frac{A(A+1)}{2}$, $A = \#\{alleles\}$
GT	uint8_t[n]	missing<<7 | phased<<6 | allele1<<3 | allele2
_GT	uint8_t+uint8_t[n*P]	Generic GT; the first int equals the max ploidy P. If the highest bit is set, the allele is not present (e.g. due to different ploidy between samples).
GQ	uint8_t[n]	Genotype quality
HQ	uint8_t[n*2]	Haplotype quality
_HQ	uint8_t+uint8_t[n*P]	Generic HQ
IBD	uint32_t[n*2]	IBD
_IBD	uint8_t+uint32_t[n*P]	Generic IBD
PL	uint8_t[n*G]	Phred-scaled likelihood of data
PS	uint32_t[n]	Phase set
Integer	int32_t[n*X]	Fix-sized custom Integer; X defined in the header
Numeric	double[n*X]	Fix-sized custom Numeric
String	uint32_t+char*	NULL padded concat. strings (int equals to the length)

- A BCF file is in the BGZF format.
- All multi-byte numbers are little-endian.
- In a string, a missing value '.' is an empty C string "\0" (not ".\0")
- For GL and PL, likelihoods of genotypes appear in the order of alleles in REF and then ALT. For example, if REF=C, ALT=T,A, likelihoods appear in the order of CC,CT,TT,CA,TA,AA (NB: the ordering is different from the one in the original BCF proposal).
- Predefined FORMAT fields can be missing from VCF headers, but custom FORMAT fields are required to be explicitly defined in the headers.
- A FORMAT field with its name starting with '_' is specific to BCF only. It gives an alternative binary representation of the corresponding VCF field, in case the default representation is unable to keep the genotype information, for example, when the ploidy is not 2 or there are more than 8 alleles.