### 1.4.9 Pedigree field format

It is possible to record relationships between genomes using the following syntax:

```
##PEDIGREE=<ID=TumourSample,Original=GermlineID>
##PEDIGREE=<ID=SomaticNonTumour,Original=GermlineID>
##PEDIGREE=<ID=ChildID,Father=FatherID,Mother=MotherID>
##PEDIGREE=<ID=SampleID,Name_1=Ancestor_1,...,Name_N=Ancestor_N>
```

or a link to a database:

```
##pedigreeDB=URL
```

See 5.4.11 for details.

### 1.4.10 Sequence collection format

The reference sequences can be recorded as a sequence collection identifier[‡]

```
##seqcol=<ID=ga4gh:SC.hsgd70lp87g2u8sgsnd0g2fxxc,url=ga4gh_resolver_url>
```

The ID section should contain the sequence collection identifier of the reference sequence that was used to generate the VCF prefixed with ga4gh:SC. Only one refget-seqcol entry can be added per VCF.

## 1.5 Header line syntax

The mandatory header line names the 8 fixed, mandatory columns. These columns are as follows:

#CHROM POS ID REF ALT QUAL FILTER INFO

If genotype data is present in the file, these are followed by a FORMAT column header, then an arbitrary number of sample IDs. Duplicate sample IDs are not allowed. The header line is tab-delimited and there must be no tab characters at the end of the line.

## 1.6 Data lines

All data lines are tab-delimited with no tab character at the end of the line. The last data line must end with a line separator. In all cases, missing values are specified with a dot ('.').

### 1.6.1 Fixed fields

There are 8 fixed fields per record. Fixed fields are:

1. CHROM — chromosome: An identifier from the reference genome or an angle-bracketed ID String ("<ID>") pointing to a contig in the assembly file (cf. the ##assembly line in the header). All entries for a specific CHROM must form a contiguous block within the VCF file. (String, no whitespace permitted, Required).

2. POS — position: The reference position, with the 1st base having position 1. Positions are sorted numerically, in increasing order, within each reference sequence CHROM. It is permitted to have multiple records with the same POS. Telomeres are indicated by using positions 0 or N+1, where N is the length of the corresponding chromosome or contig. (Integer, Required)

3. ID — identifier: Semicolon-separated list of unique identifiers where available. If this is a dbSNP variant the rs number(s) should be used. No identifier should be present in more than one data record. If there is no identifier available, then the MISSING value should be used. (String, no whitespace or semicolons permitted, duplicate values not allowed.)

4. REF — reference base(s): Each base must be one of A,C,G,T,N (case insensitive). Multiple bases are permitted. The value in the POS field refers to the position of the first base in the String. For simple insertions and deletions in which either the REF or one of the ALT alleles would otherwise be null/empty, the REF and ALT Strings must include the base before the variant (which must be reflected in the POS field), unless the variant occurs

[‡] https://ga4gh.github.io/refget/seqcols/