- PSL (List of Strings): The list of phase sets, one for each allele value specified in the `GT`. Unphased alleles (without a | separator before them) must have the value '.' in their corresponding position in the list. Unlike `PS` (which is defined per CHROM), records with different CHROM but the same phase-set name are considered part of the same phase set. If an implementation cannot guarantee uniqueness of phase-set names across the VCF (for example, phasing a streaming VCF or each CHROM is processed independently in parallel), new phase-set names should be of the format CHROM*POS*ALLELE-NUMBER of the "first" allele which is included in this set, with ALLELE-NUMBER being the one-based index of the allele in the `GT` field, since multiple distinct phase-sets could start at the same position. [§] A given sample-genotype must not have values for both PS and PSL. In addition, PS and PSL are not interoperable, in that a PS mentioned in one variant cannot be referenced in a PSL in another, since when used in PS it isn't connected to any specific haplotype (i.e. first or second), but PSL is.

  Example:

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	SAMPLE1
chr19	5	.	T	G	.	PASS	DP=100	GT:PSL	`|0/1:chr19*5*1,.`
chr20	10	.	A	T,G	.	PASS	DP=100	GT:PSL	`|1/2|3:chr20*10*1,.,chr19*5*1`
chr20	15	.	G	C	.	PASS	DP=100	GT:PSL	`1|2:.,chr20*10*1`

- PSO (List of integers): List of phase set ordinals. For each phase-set name, defines the order in which variants are encountered when traversing a derivative chromosome. The missing value '.' should be used when the corresponding PSO value is missing. For each phase-set name, PSO should be defined if any allele with that phase-set name on any record is symbolic structural variant or in breakpoint notation. Variants in breakpoint notation must have the same PSL and PSO on both records.

  Without explicitly specifying the derivative chromosome traversal order, multiple derivative chromosome reconstructions are possible. Take for example this tandem duplication in a triploid organism with SNVs (ID/QUAL/FILTER columns removed for clarity):

#CHROM	POS	REF	ALT	INFO	FORMAT	SAMPLE1
chr1	10	T	<DUP>	SVCLAIM=DJ	GT:PSL:PSO	`/0/0|1:.,.,chr1*10*3:.,.,3`
chr1	20	A	G	.	GT:PSL:PSO	`/0/0|0|1:.,.,chr1*10*1,chr1*10*3:.,.,4,`
chr1	30	G	T	.	GT:PSL:PSO	`/0/0|0|1:.,.,chr1*10*1,chr1*10*3:.,.,2,`

  Without defining PSO, it would be ambiguous as to which copy of the duplicated region the SNVs occur on. In this example, the presence of the PSO field clarifies that the SNVs are cis phased with the duplication, the first SNV occurs on the first copy of the duplicated region, and second SNV on the second copy.

- PSQ (List of integers): The list of PQs, one for each phase set in PSL (encoded like PQ). The missing value '.' should be used when the corresponding PSL value is missing, or when the phasing is of unknown quality.

2 Understanding the VCF format and the haplotype representation

VCF records use a single general system for representing genetic variation data composed of:

- Allele: representing single genetic haplotypes (A, T, ATC).
- Genotype: an assignment of alleles for each chromosome of a single named sample at a particular locus.
- VCF record: a record holding all segregating alleles at a locus (as well as genotypes, if appropriate, for multiple individuals containing alleles at that locus).

VCF records use a simple haplotype representation for REF and ALT alleles to describe variant haplotypes at a locus. ALT haplotypes are constructed from the REF haplotype by taking the REF allele bases at the POS in the reference genotype and replacing them with the ALT bases. In essence, the VCF record specifies a-REF-t and the alternative haplotypes are a-ALT-t for each alternative allele.

[§]The '*' character is used as a separator since ':' is not reserved in the CHROM column.