

---

# Protocol Documentation

## Table of Contents

vmc.proto .....	1
Allele .....	1
Genotype .....	1
Haplotype .....	1
Interval .....	2
IntervalEdit .....	2
SequenceReference .....	2
Scalar Value Types .....	2

## vmc.proto

### Allele

Allele represents a single contiguous change on a specific reference sequence

**Table 1. Allele Fields**

Field	Type	Label	Description
seqref	SequenceReference	optional	sequence reference (namespace and accession)
interval	Interval	optional	location of sequence change
replacement	string	optional	replacement sequence
id	string	optional	Allele identifier

### Genotype

Genotype represents multiple changes at a single location

**Table 2. Genotype Fields**

Field	Type	Label	Description
allele_ids	string	optional	list of haplotypes by id
id	string	optional	Genotype identifier

### Haplotype

Haplotype represents a collection of phased changes on a single reference.

**Table 3. Haplotype Fields**

Field	Type	Label	Description
allele_ids	string	optional	list of haplotypes by id

Field	Type	Label	Description
id	string	optional	Haplotype identifier

## Interval

Represents the definite (i.e., not fuzzy) location of a sequence feature using an interval of interbase coordinates.

Interbase coordinates refer to the points *between* residues. For a sequence of length  $n$ ,  $0 \leq \text{start} \leq \text{end} \leq n$ , where 0 refers to the point before the start of the sequence,  $n$  refers to the point at the end of the sequence. An interval in which  $\text{start} == \text{end}$  is a zero width point between two nucleotides. See [http://gmod.org/wiki/Introduction\\_to\\_Chado#Interbase\\_Coordinates](http://gmod.org/wiki/Introduction_to_Chado#Interbase_Coordinates) for more information.

**Table 4. Interval Fields**

Field	Type	Label	Description
start	uint64	optional	start position
end	uint64	optional	end position

## IntervalEdit

IntervalEdit represents a located sequence change.

Consider renaming fields to match message name. One possibility is `location`  $\Rightarrow$  `interval` and `replacement`  $\Rightarrow$  `edit`, thus matching the message name `IntervalEdit`.

**Table 5. IntervalEdit Fields**

Field	Type	Label	Description
location	Interval	optional	location of sequence change
replacement	string	optional	replacement sequence; empty for deletion

## SequenceReference

SequenceReference represents a named reference to a sequence in a database. For the purposes of VMC, it is essential that the mapping from SequenceReference to sequence is many-to-one and immutable.

**Table 6. SequenceReference Fields**

Field	Type	Label	Description
namespace	string	optional	name of recognized sequence reference
accession	string	optional	replacement sequence; empty for deletion

## Scalar Value Types

.proto Type	Notes	C++ Type	Java Type	Python Type
double		double	double	float

<b>.proto Type</b>	<b>Notes</b>	<b>C++ Type</b>	<b>Java Type</b>	<b>Python Type</b>
float		float	float	float
int32	Uses variable-length encoding. Inefficient for encoding negative numbers – if your field is likely to have negative values, use sint32 instead.	int32	int	int
int64	Uses variable-length encoding. Inefficient for encoding negative numbers – if your field is likely to have negative values, use sint64 instead.	int64	long	int/long
uint32	Uses variable-length encoding.	uint32	int	int/long
uint64	Uses variable-length encoding.	uint64	long	int/long
sint32	Uses variable-length encoding. Signed int value. These more efficiently encode negative numbers than regular int32s.	int32	int	int
sint64	Uses variable-length encoding. Signed int value. These more efficiently encode negative numbers than regular int64s.	int64	long	int/long
fixed32	Always four bytes. More efficient than uint32 if values are often greater than 2 <sup>28</sup> .	uint32	int	int
fixed64	Always eight bytes. More efficient than uint64 if values are often greater than 2 <sup>56</sup> .	uint64	long	int/long
sfixed32	Always four bytes.	int32	int	int
sfixed64	Always eight bytes.	int64	long	int/long
bool		bool	boolean	boolean
string	A string must always contain UTF-8 encoded or 7-bit ASCII text.	string	String	str/ unicode
bytes	May contain any arbitrary sequence of bytes.	string	ByteString	str