

## DPCC Data Standard Reference for Sequence FASTA v2.1

Sequence_Identifier	
<b>Input Type</b>	Text Field
<b>Definition</b>	A unique identifier for each sequence in a submission
<b>Format</b>	Text  Allowed characters include alphanumeric, hyphen, and underscore: a-z, A-Z, 0-9, -, _  Maximum length: 90 characters
<b>Value List</b>	None
<b>Curation</b>	The entry must be unique for each sequence in the FASTA file and must contain a Sample_Identifier, Segment, and optional Clone_Identifier.  The Sample_Identifier initially assigned to the surveillance sample must be provided. For non-surveillance submissions, enter a unique Sample_Identifier.  Allowed values for Segment are: Influenza A virus: PB2, PB1, PA, HA, NP, NA, MP, NS Influenza B virus: PB1, PB2, PA, HA, NP, NA, MP, NS Influenza C virus: PB2, PB1, P3, HE, NP, MP, NS Influenza D virus: PB2, PB1, P3, HE, NP, P42, NS Other virus types: Enter gene name
<b>Examples</b>	22258468.PB2, 512245.RdRp, 452134.HA.1, 452134.HA.2, 452134.NA.Major_variant, or 452134.NA.Minor_variant
<b>Notes</b>	Submissions without a corresponding surveillance submission will be listed as orphaned records.  Clone_Identifier is required only when submitting multiple sequences for the same Segment and Sample_Identifier, i.e., from mixed or co-infected samples.
<b>Dependent Fields</b>	
<b>Validation</b>	Validate field length
<b>Message Code</b>	Error_70_INVALID_FIELD_LENGTH
<b>Validation</b>	Validate segment name based on influenza type ( A, B, C or D). Influenza type will be determined based on Strain_Name's first character. 'rg-' prefixes are escaped from this validation. Non-influenza sequences also escaped from this validation
<b>Message Code</b>	Error_85_INVALID_SEGMENT_FOR_TYPE
<b>Validation</b>	If Clone_Identifier is present then it should be unique for a sample_identifier and a segment
<b>Message Code</b>	Error_105_INVALID_CLONE_ID
<b>Validation</b>	Must match Sample_Identifier in metadata template.
<b>Message Code</b>	Error_60_SEQUENCE_NOFILE_NOACCESSION
<b>Validation</b>	Sequence Identifier should be unique
<b>Message Code</b>	Error_57_SEQUENCE_DUPLICATE_SEQUENCE_IDENTIFIER