

DPCC Data Standard Reference for Sequence FASTA Submission v2.1

	Sequence_Identifier
Input Type	Text Field
Definition	A unique identifier for each sequence in a submission
Format	Text Allowed characters include alphanumeric, hyphen, and underscore: a-z, A-Z, 0-9, -, _ Maximum length: 90 characters
Value List	None
Curation	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: <u>Influenza A virus:</u> PB2, PB1, PA, HA, NP, NA, MP, NS <u>Influenza B virus:</u> PB1, PB2, PA, HA, NP, NA, MP, NS <u>Influenza C virus:</u> PB2, PB1, P3, HE, NP, MP, NS <u>Influenza D virus:</u> PB2, PB1, P3, HE, NP, P42, NS <u>Other virus types:</u> For full genomes or sequences from only one virus, please enter the virus name (e.g., SARS-CoV-2). For multiple sequences or for gene segments, enter the gene name (e.g., Capsid).
Examples	22258468.PB2, 512245.RdRp, PV10234.SARS-CoV-2, 452134.HA.1, 452134.HA.2, 452134.NA.Major_variant, or 452134.NA.Minor_variant
Notes	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.
Dependent Fields	
Validation	Validate field length
Message Code	Error_70_INVALID_FIELD_LENGTH
Validation	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
Message Code	Error_85_INVALID_SEGMENT_FOR_TYPE
Validation	If Clone_Identifier is present then it should be unique for a sample_identifier and a segment
Message Code	Error_105_INVALID_CLONE_ID
Validation	Must match Sample_Identifier in metadata template.
Message Code	Error_60_SEQUENCE_NOFILE_NOACCESSIOIN
Validation	Sequence Identifier should be unique
Message Code	Error_57_SEQUENCE_DUPLICATE_SEQUENCE_IDENTIFIER