

DPCC FASTA Reference for Sequence Submission v2.0

	Sequence_Identifier	Sample_Identifier	Segment	Clone_Identifier
Input Type	Text Field	Text Field	Text Field	Text Field
Definition	A unique identifier for each sequence in a submission	The identifier assigned to the virus strain	The viral gene name of the sequence submission	The identifier assigned to the individual clone of the influenza virus genome segment
Format	Text Maximum length: 90 characters	Text Maximum length: 50 characters	Text Maximum length: 25 characters	Text Maximum length: 15 characters
Value List	None	None	Influenza A virus: PB2, PB1, PA, HA, NP, NA, MP, or NS Influenza B virus: PB1, PB2, PA, HA, NP, NA, MP, or NS Influenza C virus: PB2, PB1, P3, HE, NP, MP, or NS Influenza D virus: PB2, PB1, P3, HE, NP, P42, or NS Other virus types: Enter gene name	None
Curation	The entry must be unique for each sequence in the FASTA file. Data elements must match those listed in the Sample_Identifier, Segment, and optional Clone_Identifier fields.	The Sample_Identifier initially assigned to the surveillance sample must be provided.	No spaces are allowed. Segment must match the influenza virus type (A, B, C, or D) listed in the Strain_Name field on the Sequence Metadata template. For other virus types, if the sequence covers multiple genes, comma-separate individual gene names.	No spaces are allowed. Optional field
Examples	22258468.PB2 or 452134.HA.1 or 452134.HA.2 or 452134.NA.Major_variant or 452134.NA.Minor_variant	22258468 or 452134	PB2 or HA or RdRp or Capsid,RdRp	1 or 2 or Major_variant or Minor_variant
Notes		Submissions without a corresponding surveillance submission will be listed as orphaned records. For non-surveillance submissions, enter a unique Sample_Identifier.		This field is not required when submitting a single sequence for an individual genome Segment. This field is required when submitting multiple sequences for the same Segment and Sample_Identifier i.e., from mixed or co-infected samples.