

DPCC Data Standard Reference for Sequence Metadata v2.3						
Input Type	Project_Identifier	Contributing_Institution	Sample_Identifier	BioProject_Accession_Number	Embargo_End_Date	Provisional_Authors
Text Field	Text Field	Text Field	Text Field	Text Field	Date Field	Text Field
<b>Definition</b>	A unique Project Identifier generated by the DPCC by combining the Center-generated Project Code and a random 4-digit number	The institution code for the group that PERFORMED THE ANALYSIS. Must be one of the CEIRS institution codes assigned by NIAID	Identifier initially assigned to each sample collected. If multiple samples are taken from the same host, each sample should have its own identifier.	The BioProject unique accession number(s) associated with the sequence submission	In case an embargo of the information is needed, the date that the information should be released to the public databases by the DPCC	The list of authors associated with the sequence submission. Default entry in absence of a Publication_PMID will be the authors listed by the user. If a Publication_PMID is provided, publication authors will be listed.
<b>Format</b>	Project_Code_XXXX Maximum length: 21 characters	Center three-letter code followed by three digits. Maximum length: 6 characters	Center-specific Allowed characters include alphanumeric, hyphen, and underscore: a-z, A-Z, 0-9, -, _ Maximum length: 50 characters	BioProject ID Maximum length: 100 characters	DD-Mon-YYYY DD-Mon-YY NA Maximum length: 11 characters	FirstName LastName FirstName M. LastName Maximum length: 500 characters
<b>Value List</b>	None	None	None	None	Date NA	Text NA
<b>Duration</b>	The entry must be a Project Identifier value registered with the DPCC.	The entry must be an Institution Code value registered with the DPCC. Center 3-letter codes are case-sensitive and must be entered in all-caps.	The Sample_Identifier initially assigned to the surveillance sample must be provided.	The entry must be a valid BioProject accession number previously registered with the DPCC.	1. Leading 0 in DD is optional. 2. Month must match the first three letters of the month. Month is NOT case-sensitive. 3. Years may have two or four digits. 4. Date must conform to NIAID data release policies.	If a Publication_PMID is entered, use NA. The list of authors from the publication will be used.
<b>Examples</b>	SJCProj02_4001	SJC101	22258468	PRJNA37813	3-Mar-2011, 03-Mar-2011, 03-MAR-2011, 3-MAR-11, or NA	Indiana Jones, John M. Henry, or NA
<b>Notes</b>	None	None	Submissions without a corresponding surveillance submission will be listed as orphaned records.  For non-surveillance submissions, enter a unique Sample_Identifier.	The sequence must be linked to a BioProject.  If no relevant BioProject exists, one must be created.	An exact date must be provided.  If Embargo_End_Date is NA, information will be released without delay.  Embargo_End_Date cannot be more than 12 months after submission.	The DPCC will update the GenBank entry records if and when the authors provide a publication referenced after submission.  For listing of multiple names, comma-separate the names maintaining the order of FirstName and LastName or FirstName M. LastName
<b>Dependent Fields</b>						Publication_PMID
<b>Validation</b>	Project_Identifier should be a valid project identifier.	Must be an Institution Code value registered with the DPCC	Validate field length	Must be a valid BioProject accession number.	Embargo_End_Date must be exact date and cannot be a past date.	Validate field length
<b>Message Code</b>	Error_9_PROJECT_NOT_FOUND	Error_1_INVALID_VALUE	Error_70_INVALID_FIELD_LENGTH	Error_104_INVALID_BIOPROJECT	Error_7_INVALID_EMBARGO_DATE	Error_70_INVALID_FIELD_LENGTH
<b>Validation</b>	Project_Identifier exists but user does not have permission to access or edit the project.		If Surveillance_Sample is N, Sample_Identifier should be unique across all DPCC data.	Validate field length	Embargo_End_Date cannot be more than 12 months after submission.	Multiple names must be entered as comma-separated values in the order of FirstName LastName or FirstName M. LastName.
<b>Message Code</b>	Error_4_DENIED_USER_ACCESS		Error_66_NON_UNIQUE_SAMPLE_ID	Error_70_INVALID_FIELD_LENGTH	Error_100_EMBARGO_DATE_12M	Error_76_INVALID_NAME_LIST
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						

	Submission_Title	Publication_PMID	Molecule_Type	Sequencing_Technology	Forward_Primer	Reverse_Primer
Input Type	Text Field	Text Field	Text Field	Text Field	Text Field	Text Field
<b>Definition</b>	The descriptive title assigned to the sequence submission. Default entry in absence of a Publication_PMID will be "Direct Submission (CEIRS DPCC)". If a Publication_PMID is provided, Submission_Title should be the title of the referenced publication.	The PubMed Unique Identifier (PMID) for the publication in which the sequence was published.	The type of organic molecule that was sequenced.	The name of the sequencing technology used to obtain the submitted sequences.	The forward PCR primer that was used to amplify the nucleic acid that was sequenced.	The reverse PCR primer that was used to amplify the nucleic acid that was sequenced.
<b>Format</b>	Text Maximum length: 200 characters	PMID Maximum length: 50 characters	Text Maximum length: 15 characters	Text Maximum length: 250 characters	Name:Sequence Maximum length: 500 characters	Name:Sequence Maximum length: 500 characters
<b>Value List</b>	Text NA	Text NA	genomic DNA genomic RNA mRNA other RNA other DNA rRNA transcribed RNA iRNA unassigned DNA unassigned RNA viral cRNA	454 Helicos Illumina Ion Torrent Oxford Nanopore Pacific Biosciences Sanger SOLID OTH-	Text U	Text U
<b>Duration</b>	If no Publication_PMID is provided, use NA. If a Publication_PMID is entered, use the title of the publication exactly.	The entry must be a valid PMID number: 7 or 8-digit number with no leading zeros. <a href="http://www.nlm.nih.gov/bsd/mms/mmedlineelements.html#pmid">http://www.nlm.nih.gov/bsd/mms/mmedlineelements.html#pmid</a>	The entry must be one and only one member of the Value List.	The entry must be one or more comma-separated members of the Value List. If OTH- is selected, then sequencing technology should be entered as free text.	The entry must include the forward primer name and nucleotide sequence separated by a colon.	The entry must include the reverse primer name and nucleotide sequence separated by a colon.
<b>Examples</b>	Analysis of pandemic influenza H1N1 sequences from New York in 2010, or NA	19465683, or NA	viral cRNA	IonTorrent	fwd_seq:catgttgcacaagaaggaaa, or U	rev_seq:atgtggatgcagatgttggaa, or U
<b>Notes</b>	The DPCC will update the GenBank entry records if and when the authors provide a publication reference after submission.	If the sequence was referenced in multiple publications, comma-separate the individual PMIDs. Enter NA if the sequence was not referenced in a publication.	Enter viral cRNA for influenza virus sequences. Enter genomic RNA for SARS-CoV-2 sequences. Enter genomic DNA for plasmid sequences or sequences of ribosomal RNA genes.	If more than one sequencing technology is used, comma-separate individual technologies.	If multiple forward primers were used, comma-separate individual forward primers. Enter U if the forward primer is unknown.	If multiple reverse primers were used, comma-separate individual reverse primers. Enter U if the reverse primer is unknown.
<b>Dependent Fields</b>	Publication_PMID	Provisional_Authors, Submission_Title				
<b>Validation</b>	Validate field length	Validate field length	Field value should be one of valid values as in list.	Field value should be one of valid values as in list. NOTE: User can enter other value by prefixing 'OTH-'	Validate primer name and sequence format of primename:primer sequence. Multiple primers will be separated by comma.	Validate primer name and sequence format of primename:primer sequence. Multiple primers will be separated by comma.
<b>Message Code</b>	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH	Error_1_INVALID_VALUE	Error_1_INVALID_VALUE	Error_83_INVALID_PRIMER_NAME	Error_83_INVALID_PRIMER_NAME
<b>Validation</b>		Must be a valid PMID		Field length including values from the Value List and free text following 'OTH-' must be less than 250 characters.	Validate field length	Validate field length
<b>Message Code</b>		Error_96_INVALID_PMID		Error_75_INVALID_FIELD_LENGTH_OT	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						

	Assembly_Method	Assembler_Version	Coverage	Genbank_Accession_Numbers
<b>Input Type</b>	Text Field	Text Field	Text Field	Text Field
<b>Definition</b>	The name of the program used to assemble next-generation or sanger sequencing reads	The version of the assembly program used or, if not available, the date the assemblies were made	The average number of reads representing a given nucleotide in the sequence	The GenBank accession number(s) for the sequence submission, if a GenBank number has been assigned
<b>Format</b>	Text Maximum length: 150 characters	Text DD-Mon-YYYY DD-Mon-YY D-Mon-YYYY D-Mon-YY Maximum length: 50 characters	Text Maximum length: 50 characters	Text Maximum length: 1000 characters
<b>Value List</b>	None	Text Date	Number U	Text NA
<b>Curation</b>	The entry must be the name of a valid sequence assembly program.	The entry must be the version number of the assembly program used in format v.x.x or the date the assemblies were created.	The entry must be a number or enter U if unknown.	The entry must be a valid GenBank accession number or NA if none is available.
<b>Examples</b>	IonTorrent Assembler, or BioEdit	v3.2, 03-Mar-2011, 3-Mar-2011, or 3-Mar-11	25.47	HM006759, or NA
<b>Notes</b>	Sequences must be pre-assembled. Raw sequence reads from next generation sequencing technologies should not be submitted to GenBank.  If more than one assembly method is used, comma-separate individual methods.	If more than one assembly method is used, comma-separate individual versions.	If more than one coverage is used, comma-separate individual coverages.	If a GenBank_Accession_Number is provided, a FASTA file is not required.
<b>Dependent Fields</b>				
<b>Validation</b>	Validate field length	Validate field length	Validate field length	Validate field length
<b>Message Code</b>	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH
<b>Validation</b>		Validate assembler version format or date format.	Coverage has an invalid value. Must be a number or U	Sequence in fasta file or Genbank_Accession_Numbers is required
<b>Message Code</b>		Error_64_VALIDATION_ASSEMBLER_VERSION	Error_65_VALIDATION_COVERAGE	Error_60_SEQUENCE_NOFILE_NOACCESSION
<b>Validation</b>				
<b>Message Code</b>				
<b>Validation</b>				
<b>Message Code</b>				
<b>Validation</b>				
<b>Message Code</b>				
<b>Validation</b>				
<b>Message Code</b>				
<b>Validation</b>				
<b>Message Code</b>				
<b>Validation</b>				
<b>Message Code</b>				
<b>Validation</b>				
<b>Message Code</b>				
<b>Validation</b>				
<b>Message Code</b>				
<b>Validation</b>				
<b>Message Code</b>				
<b>Validation</b>				
<b>Message Code</b>				

Input Type	Strain Name
Definition	The WHO or ICTV strain name of the virus that was sequenced
Format	<p><b>Influenza A virus:</b> Antigenic Type/Host of Origin/Geographical Origin/Strain Number/Year of Isolation (Subtype)</p> <p><b>Influenza B, C, or D virus:</b> Antigenic Type/Host of Origin/Geographical Origin/Strain Number/Year of Isolation</p> <p><b>SARS-CoV-2 and other viruses:</b> Virus Name/Host of Origin/Geographical Origin/Strain Number/Year of Isolation</p> <p>Maximum length: 150 characters</p>
Value List	Text U
Curation	<p><b>For influenza viruses,</b> strain names most follow WHO naming convention: Fields must be ordered as follows and separated with the "/" character:</p> <ol style="list-style-type: none"> <li>The antigenic type (e.g., A, B, C, D)</li> <li>The host of origin (e.g., swine, equine, chicken. For human-origin viruses, no host of origin designation is given.)</li> <li>Geographical origin (e.g., Denver, Taiwan)</li> <li>Strain number (e.g., 15, 7)</li> <li>Year of isolation (e.g., 2009, 1934)</li> <li>For influenza A viruses, the hemagglutinin and neuraminidase antigen description in parentheses (e.g., (H1N1), (H3N2))</li> </ol> <p><b>For SARS-CoV-2 viruses,</b> strain names most follow ICTV naming convention: Fields must be ordered as follows and separated with the "/" character:</p> <ol style="list-style-type: none"> <li>Virus name (e.g., SARS-CoV-2)</li> <li>The host of origin (e.g., human. Human-origin viruses must include the origin designation.)</li> <li>Country of geographical origin as a three-letter code from the DPCC's Country Codes list (e.g., USA, MEX, CAN)</li> <li>Strain number (e.g., 15, 7)</li> <li>Year of isolation (e.g., 2019, 2020)</li> </ol> <p><b>For all other viruses,</b> please use the following convention: Fields must be ordered as follows and separated with the "/" character:</p> <ol style="list-style-type: none"> <li>Virus name (e.g., MERS-CoV, Bat-CoV, etc.)</li> <li>The host of origin (e.g., human, bat, camel. Human-origin viruses must include the origin designation.)</li> <li>Geographical origin, either regional locality or country (e.g., Denver, Taiwan)</li> <li>Strain number (e.g., 15, 7)</li> <li>Year of isolation (e.g., 2019, 2020)</li> </ol>
Examples	<p><b>For influenza:</b> A/Hong Kong/1/1968 (H3N2), A/chicken/Fujian/4/2002 (H3N6), A/chicken/Fujian/4/2002 (H3Nx), A/chicken/Fujian/4/2002 (mixed), A/swine/Iowa/233-56/2011 (H3N2), A/duck/Alberta/35/1976 (H1N1), B/Hong Kong/432/2014, C/Texas/1987/2011, or D/swine/Oklahoma/1334/2011</p> <p><b>For SARS-CoV-2:</b> SARS-CoV-2/human/USA/NY-PV08486/2020</p> <p><b>For other viruses:</b> PHEV-CoV/swine/USA/15TOSU25049/2015</p>
Notes	<p>WHO Reference for influenza viruses: <a href="http://www.cdc.gov/flu/about/viruses/types.htm">http://www.cdc.gov/flu/about/viruses/types.htm</a> ICTV Reference for SARS-CoV-2 viruses: <a href="https://pubmed.ncbi.nlm.nih.gov/32123347">https://pubmed.ncbi.nlm.nih.gov/32123347</a></p> <p>(H3Nx) can be used in cases where a partial subtype has been determined (e.g., H5Nx, H3N2).</p> <p>If there are mixed subtypes contained within a sample use A/chicken/Fujian/4/2002 (mixed) for Strain Name and enter additional subtype information in the Comments field.</p>
Dependent Fields	
Validation	Validate field length
Message Code	Error_70_INVALID_FIELD_LENGTH
Validation	Geographical origin element of strain name must be alphanumeric, underscore, dash, period, or single quote: a-Z, A-Z, 0-9, '-', '.', '_'
Message Code	Error_138_INVALID_LOCATION_STRAIN_TEXT
Validation	Strain number element of strain name must be alphanumeric, underscore, dash, or period: a-z, A-Z, 0-9, '-', '.', '_'
Message Code	Error_139_INVALID_STRAIN_NUMBER_STRAIN_TEXT
Validation	Year of isolation element of strain name must be 4-digit year.
Message Code	Error_110_INVALID_STRAIN_YEAR
Validation	If strain is influenza A, subtype must be present as text between parentheses.
Message Code	Error_112_INVALID_STRAIN_SUBTYPE
Validation	Brackets cannot be present if strain does not have rg_ prefix.
Message Code	Error_119_INVALID_STRAIN_BRACKETS
Validation	If included, the strain must have a matched pair of opening and closing brackets or parentheses.
Message Code	Error_158_INVALID_STRAIN_OPEN_BRACKETS
Validation	Brackets cannot be present if strain does not have rg_ prefix.
Message Code	Error_159_INVALID_NUMBER_OF_STRAIN_ELEMENTS
Validation	If included, the strain must have a matched pair of opening and closing brackets or parentheses.
Message Code	Error_158_INVALID_STRAIN_OPEN_BRACKETS
Validation	Brackets cannot be present if strain does not have rg_ prefix.
Message Code	Error_159_INVALID_NUMBER_OF_STRAIN_ELEMENTS

Input Type	Sample_Material	Surveillance_Sample	Host_Species	Host_Common_Name	Collection_Date	Collection_Country
	Text Field	Text Field	Text Field	Text Field	Date Field	Text Field
<b>Definition</b>	Material on which the testing was performed. If multiple samples are taken from the same host, they must be entered as separate records.	Is the sequence derived from a surveillance sample?	Full scientific name of host genus and species, without abbreviations, as listed in the DPCC Species Dictionary	The English common name given to a particular species as listed in the DPCC Species Dictionary	Date on which the sample was collected.	Country in which the original sample was collected, as listed in the DPCC Country Codes list
<b>Format</b>	Text Maximum length: 30 characters	Text Maximum length: 1 character	Text Maximum length: 100 characters	Text Maximum length: 100 characters	DD-Mon-YYYY DD-Mon-YY Mon-YYYY Mon-YY YYYY YY Maximum length: 11 characters	Text Maximum length: 60 characters
<b>Value List</b>	AIR BAL BLO CCO CCT CLO FEC LLF LUN NAL NAS NPH NTS ORP OTH- OTT PLS RCS SER SLU SOI SPU TFB TFT TRS WAT U	Y N	DPCC Species Dictionary ENV/ U NA	DPCC Species Dictionary U NA	Date U	ISO 3166 Standard Country Code U
<b>Duration</b>	The entry must be one and only one member of the Value List.  Values are case-sensitive and must be entered in all-caps.	The entry must be one and only one member of the Value List.	If the entry is not ENV or U or NA, the host species name is validated against the DPCC Species Dictionary.  For reverse genetic viruses or non surveillance-derived viruses, enter NA.	If the entry is not U or NA, the host common name is validated against the DPCC Species Dictionary.  For reverse genetic viruses or non surveillance-derived viruses, enter NA.	1. Leading 0 in DD is optional. 2. Month must match the first three letters of the month. Month is NOT case-sensitive. 3. Years may have two or four digits. 4. Use U (Unknown) if date is not known.	The entry must be one and only one member of the Value List.  Values are case-sensitive and must be entered in all-caps.
<b>Examples</b>	BLO	Y	Swiss scaup, or Anser albifrons	wild boar, or greater white-fronted goose	3-Mar-2011, 03-Mar-2011, 03-MAR-2011, Mar-2011, 3-Mar-11, 2011, 11, or U	VNM, or USA
<b>Notes</b>	AIR = Air BAL = Bronchoalveolar lavage BLO = Blood CCF = Combined cloacal and fecal CCO = Combined cloacal and oral-pharyngeal CCT = Combined cloacal and tracheal CLO = Cloacal FEC = Feces LLF = Lung lavage fluid LUN = Lungs NAL = Nasal lavage NAS = Nasal swab NPH = Nasopharyngeal swab NTS = Combined nasal throat swab ORP = Oral-pharyngeal OTH = Other; append free text to describe OTT = Other tissue PLS = Plasma RCS = Rectal swab SER = Serum SLU = Slurry SOI = Soil SPU = Sputum TFB = Tissue from brain TFT = Tissue from trachea TRS = Tracheal swab WAT = Water U = Unknown	Y = Yes N = No  Enter Y if the sequence was derived from a surveillance sample, even if the primary sample was not collected through CEIRS.  Enter N if the sequence is a lab strain or reverse genetic virus.	Please reference the DPCC Species Dictionary for allowed values.  Use ENV(ronment) for samples taken from the environment (e.g., a water sample or feces picked up from the beach) when the source species is not known.  If genus is known but species is unknown, then use the genus followed by 'sp.' (e.g., domestic duck would be reported as 'Anas sp.').  If both the genus and species are unknown, then use the scientific family name, without abbreviation (e.g., Charadriidae).  Enter U for all other cases.	Please reference the DPCC Species Dictionary for allowed values.  Use U if host common name is unknown or if the value entered under Host_Species is ENV or U.	At least one date must be provided for Sample_Receipt_Date or Collection_Date.	The ISO 3166 Standard for country codes may be found on the DPCC Portal.
<b>Dependent Fields</b>			Host_Common_Name	Host_Species		
<b>Validation</b>	Field value should be one of valid values as in list. NOTE: User can enter other value by prefixing 'OTH.'	Field value should be one of valid values as in list.	Validate field length	Validate field length	Validate date format	Field value should be a valid ISO three-letter country code or U
<b>Message Code</b>	Error_1_INVALID_VALUE	Error_1_INVALID_VALUE	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH	Error_67_INVALID_DATE_WITH_U	Error_1_INVALID_VALUE
<b>Validation</b>	Validate field length		Validate against DPCC Species Dictionary	Validate against DPCC Species Dictionary	Date cannot be future date	
<b>Message Code</b>	Error_75_INVALID_FIELD_LENGTH_OTH		Error_1_INVALID_VALUE	Error_1_INVALID_VALUE	Error_68_INVALID_FUTURE_DATE	
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						
<b>Validation</b>						
<b>Message Code</b>						

	Lab_Host	Parent_Strain_Name
<b>Input Type</b>	Text Field	Text Field
<b>Definition</b>	Description of the live animal host used for passaging the virus without additional information.	Name of the parental virus strain.
<b>Format</b>	Text Maximum length: 50 characters	<b>Influenza A virus:</b> Antigenic Type/Host of Origin/Geographical Origin/Strain Number/Year of Isolation (Subtype), or <b>Influenza B, C, or D virus:</b> Antigenic Type/Host of Origin/Geographical Origin/Strain Number/Year of Isolation, or <b>SARS-CoV-2 and other viruses:</b> Virus Name/Host of Origin/Geographical Origin/Strain Number/Year of Isolation, or Maximum length: 800 characters
<b>Value List</b>	Text NA	Text U
<b>Duration</b>	None	For <b>influenza viruses</b> , strain names most follow WHO naming convention: Fields must be ordered as follows and separated with the ^ character: 1. The antigenic type (e.g., A, B, C, D) 2. The host of origin (e.g., swine, equine, chicken. For human-origin viruses, no host of origin designation is given.) 3. Geographical origin (e.g., Denver, Taiwan) 4. Strain number (e.g., 15, 7) 5. Year of isolation (e.g., 2009, 1934) 6. For influenza A viruses, the hemagglutinin and neuraminidase antigen description in parentheses (e.g., (H1N1), (H3N2))  For <b>SARS-CoV-2 viruses</b> , strain names most follow ICTV naming convention: Fields must be ordered as follows and separated with the ^ character: 1. Virus name (e.g., SARS-CoV-2) 2. The host of origin (e.g., human. Human-origin viruses must include the origin designation.) 3. Country of geographical origin as a three-letter code from the DPCC's Country Codes list (e.g., USA, MEX, CAN) 4. Strain number (e.g., 15, 7) 5. Year of isolation (e.g., 2019, 2020)  For <b>all other viruses</b> , please use the following convention: Fields must be ordered as follows and separated with the ^ character: 1. Virus name (e.g., MERS-CoV, Bat-CoV, etc.) 2. The host of origin (e.g., human, bat, camel. Human-origin viruses must include the origin designation.) 3. Geographical origin, either regional locality or country (e.g., Denver, Taiwan) 4. Strain number (e.g., 15, 7) 5. Year of isolation (e.g., 2019, 2020)
<b>Examples</b>	ferret, or mouse	For <b>influenza</b> A/Hong Kong/1/1968 (H3N2), A/chicken/Fujian/4/2002 (H3N6), A/chicken/Fujian/4/2002 (H5N6), A/chicken/Fujian/4/2002 (mixed), A/swine/Iowa/233-56/2011 (H3N2), A/duck/Alberta/35/1976 (H1N1), B/Hong Kong/432/2014, C/Texas/19876/2011, D/swine/Oklahoma/1334/2011.  For <b>SARS-CoV-2</b> SARS-CoV-2/human/USANY-PV08486/2020  For <b>other viruses</b> PHEV-CoV/swine/USA/15TOSU25049/2015
<b>Notes</b>	This field is strictly reserved for the common or scientific name of live animals associated with generation of the sequence.  Enter passage history or experimental information in the Passage_History, Treatment, Transmission_Method, Severity, or Phenotype fields.	WHO Reference for influenza viruses: <a href="http://www.cdc.gov/flu/about/viruses/types.htm">http://www.cdc.gov/flu/about/viruses/types.htm</a> ICTV Reference for SARS-CoV-2 viruses: <a href="https://pubmed.ncbi.nlm.nih.gov/32123347">https://pubmed.ncbi.nlm.nih.gov/32123347</a>  (HxNx) can be used in cases where a partial subtype has been determined (e.g., H5Nx, HxN2).  If there are mixed subtypes contained within a sample use A/chicken/Fujian/4/2002 (mixed) for Strain_Name and enter additional subtype information in the Comments field.
<b>Dependent Fields</b>		
<b>Validation</b>	Validate field length	Validate field length
<b>Message Code</b>	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH
<b>Validation</b>		Influenza virus strain names must begin with the single letter A, B, C, or D, or be prefixed with rg-. Other strain names are not validated.
<b>Message Code</b>		QC_136_STRAIN_NAME_NOT_VALIDATED
<b>Validation</b>		Host of origin element of strain name must be alphanumeric, underscore, dash, period, or single quote: a-z, A-Z, 0-9, ., -, _
<b>Message Code</b>		Error_137_INVALID_HOST_SPECIES_STRAIN_TEXT
<b>Validation</b>		Geographical origin element of strain name must be alphanumeric, underscore, dash, period, or single quote: a-z, A-Z, 0-9, ., -, _
<b>Message Code</b>		Error_138_INVALID_LOCATION_STRAIN_TEXT
<b>Validation</b>		Strain number element of strain name must be alphanumeric, underscore, dash, or period: a-z, A-Z, 0-9, ., -, _
<b>Message Code</b>		Error_139_INVALID_STRAIN_NUMBER_STRAIN_TEXT
<b>Validation</b>		Year of isolation element of strain name must be 4-digit year.
<b>Message Code</b>		Error_110_INVALID_STRAIN_YEAR
<b>Validation</b>		If strain is influenza A subtype must be present as text between parentheses.
<b>Message Code</b>		Error_112_INVALID_STRAIN_SUBTYPE
<b>Validation</b>		Brackets cannot be present if strain does not have rg- prefix.
<b>Message Code</b>		Error_119_INVALID_STRAIN_BRACKETS
<b>Validation</b>		If included, the strain must have a matched pair of opening and closing brackets or parentheses.
<b>Message Code</b>		Error_158_INVALID_STRAIN_OPEN_BRACKETS
<b>Validation</b>		Brackets cannot be present if strain does not have rg- prefix.
<b>Message Code</b>		Error_159_INVALID_NUMBER_OF_STRAIN_ELEMENTS

	Passage_History	Antigenic_Characterization	Treatment	Transmission_Method
Input Type	Text Field	Text Field	Text Field	Text Field
Definition	Description of the passage history of the virus strain	Any information about antigenic characterization	Description of any experimental treatments	Description of the experimental method for virus transmission
Format	Text Maximum length: 50 characters	Text Maximum length: 50 characters	Text Maximum length: 200 characters	Text Maximum length: 50 characters
Value List	Text ORI X	Text NA	Text NA	Text NA
Duration	To indicate a change in passage history at one institution, combine passage host and number: ESC2  To indicate a change in passage history for samples between users or institutions, separate passage host and number with a slash '?': ESC/2	None	None	None
Examples	E2	NA	NA	Aerosol contact
Notes	Indicate lab host and number of passages: E = Embryonated chicken eggs C = MDCK cells S = MDCK-SIAT cells M = Monkey kidney cells  For other hosts, write out common name and include passage number e.g., Mouse2, E(duck)1, 293T1.  Use X if the passage number is unknown, e.g., EX, CXE2.  Enter X if passage history information is unknown.  Enter ORI if virus was not passaged and the sequence was derived directly from the original sample material.  Additional information relating to viral passaging can be entered in the Comments field.			
Dependent Fields				
Validation	Validate field length	Validate field length	Validate field length	Validate field length
Message Code	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH
Validation				
Message Code				
Validation				
Message Code				
Validation				
Message Code				
Validation				
Message Code				
Validation				
Message Code				
Validation				
Message Code				
Validation				
Message Code				
Validation				
Message Code				

Input Type	Severity		Phenotype		Comments		File Name		Internal Field Only		Internal Field Only	
	Text Field	Text Field	Text Field	Text Field	Text Field	Text Field	Text Field	Text Field	Text Field	Text Field	Text Field	Text Field
Definition	Description of the severity of infection	Description of the viral phenotype	Text describing anything else of interest related to the submission	The name of the sequence data file.	<b>Internal Field Only</b> Filled with the BioProject organism after successful registration.	<b>Internal Field Only</b> Filled with the BioProject locus tag ID after successful registration.						
Format	Text Maximum length: 50 characters	Text Maximum length: 50 characters	Text Maximum length: 2000 characters	Text Maximum length: 2000 characters	Text Maximum length: 100 characters	Text Maximum length: 100 characters						
Value List	Text NA	Text NA	Text NA	Text NA								
Duration	None	None	None	The entry must be the full file name, with extension.								
Examples	NA	LPAI	NA	22258468_1.fasta								
Notes			If there are no comments, enter NA.	There must be one sequence data file per sample. All sequence data files and the Sequence Metadata file must be submitted in one zip file.								
Dependent Fields												
Validation	Validate field length	Validate field length	Validate field length	Validate field length								
Message Code	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH								
Validation					File not found in ZIP file							
Message Code					Error_87_DATA_FILE_MISSING_FROM_ZIP							
Validation												
Message Code												
Validation												
Message Code												
Validation												
Message Code												
Validation												
Message Code												
Validation												
Message Code												
Validation												
Message Code												
Validation												
Message Code												
Validation												
Message Code												