**GSCID/BRC Project and Sample Application Standard**

**Sequencing Assay**

**v1.3**

**Finalized by the GSCID/BRC Metadata Working Group**

How to interpret the document:

**BOLD**: Field Name

*ITALICS*: Attributes of the field

1. **Sample ID - Sequencing Facility**

*Sequencing Assay Field ID*: SA1

*Description*: Unique identifier used by the relevant sequencing center to identify the sample submitted by the sample provider.

*Data Categories*: Sample Shipment

*Allowed Values*: Free text

*Syntax*: Alphanumeric

*Example Values*: V506

*Data Source*: GSCID

*Comments*: This is a candidate for the key ID that would be embedded in the GenBank record as a dbxref for linkage of BRC metadata records with GenBank sequences. Required field.

*OBO Foundry Synonym*: specimen identifier assigned by sequencing facility

*OBO Foundry Purl*: http://purl.obolibrary.org/obo/OBI\_0001901

*OBI Definition*: A specimen identifier which is assigned by a sequencing facility

*Other Synonyms*: GenBank Structured Comment: SRA\*

1. **Nucleic Acid Extraction Method**

*Sequencing Assay Field ID*: SA2

*Description*: Experimental procedure used to derive the nucleic acid fraction from the submitted sample used for the sequencing reaction.

*Data Categories*: Sequencing Sample Preparation

*Allowed Values*: OBI, http://bioportal.bioontology.org/ontologies/40832

*Syntax*: free text

*Example Values*: Illumina suggested standard method; CTAB/phenol/chloroform

*Data Source*: Sample Provider or GSCID

*Comments*: Required field with ‘not applicable’ as an allowed value

*OBO Foundry Synonym*: nucleic acid extraction

*OBO Foundry Purl*: http://purl.obolibrary.org/obo/OBI\_0666667

*OBI Definition*: a protocol is a plan specification that has sufficient level of detail and quantitative information to communicate it between domain experts, so that different domain experts will reliably be able to independently reproduce the process.

*Other Synonyms*: MIxS: sample material processing; GenBank structured comment: SRA\*; nucleic acid preparation; extraction method

1. **Nucleic Acid Preparation Method**

*Sequencing Assay Field ID*: SA3

*Description*: Details about the preparation of DNA samples for sequencing including if amplification was used (e.g., in the case of sequencing a single mosquito), and any other relevant molecular biology protocols done prior to sequencing.

*Data Categories*: Sequencing Sample Preparation

*Allowed Values*: Free text

*Syntax*: Free text

*Example Values*: Standard 454 LC

*Data Source*: GSCID

*Comments*: Required field with ‘not applicable’ as an allowed value

*OBO Foundry Synonym*: sample preparation for sequencing assay

*OBO Foundry Purl*: http://purl.obolibrary.org/obo/OBI\_0001902

*OBI Definition*: a protocol is a plan specification which has sufficient level of detail and quantitative information to communicate it between domain experts, so that different domain experts will reliably be able to independently reproduce the process.

*Other Synonyms*: MIxS:sample material processing; GenBank structured comment: SRA\*; Nucleic acid preparation by GSC; including amplification procedure

1. **Sequencing Technology**

*Sequencing Assay Field ID*: SA4

*Description*: The sequencing technology type(s) used to obtain the sequences

*Data Categories*: Sequencing Assay

*Allowed Values*: OBI, http://bioportal.bioontology.org/ontologies/40832

*Syntax*: Free text

*Example Values*: 454; Illumina

*Data Source*: GSCID

*Comments*: Required field

*OBO Foundry Synonym*: sequencing assay

*OBO Foundry Purl*: http://purl.obolibrary.org/obo/OBI\_0600047

*OBI Definition*: a protocol is a plan specification which has sufficient level of detail and quantitative information to communicate it between domain experts, so that different domain experts will reliably be able to independently reproduce the process.

*Other Synonyms*: MIxS: sequencing method

1. **Assembly Name**

*Sequencing Assay Field ID*: SA5

*Description*: a unique, short name suitable for display

*Data Categories*: Data Transformation

*Allowed Values*: Free text

*Syntax*: Free text

*Example Values*: Plas\_yoel\_17X\_V2 for Plasmodium yoellii 17X assembly version 2

*Data Source*: GSCID

*Comments*: Newly-proposed GenBank Structured Comment Field ID; Required field with ‘not applicable’ as an allowed value; ‘sequence assembly' has been submitted to OBI

*OBO Foundry Synonym*: sequence assembly name

*OBO Foundry Purl*: http://purl.obolibrary.org/obo/OBI\_0001948

*OBI Definition*: A textual entity that is used to denote a sequence assembly.

*Other Synonyms*:

1. **Assembly Method**

*Sequencing Assay Field ID*: SA6

*Description*: Computational algorithm used to assemble individual sequence reads into larger contigs. Assembly details including but not limited to assembler type (overlap-layout-consensus, deBrujn), assembler version, any relevant QC information such as % known genes/ESTs captured.

*Data Categories*: Data Transformation

*Allowed Values*: pick list

*Syntax*: free text

*Example Values*: Illumina GA pipeline ver1.3; Newbler MapAsmResearch-03/15/2010 -vs C\_elegans -e 45; No assembly. Reads mapped to reference genome withTopHat; Newbler de novo hibrid assembly/CLC reference mapping of 454 reads; AV454 v1.0

*Data Source*: GSCID

*Comments*: Required field with ‘not applicable’ as an allowed value

*OBO Foundry Synonym*: sequence assembly algorithm

*OBO Foundry Purl*: http://purl.obolibrary.org/obo/OBI\_0001522

*OBI Definition*: A plan specification that describes inputs, output of mathematical functions as well as workflow of execution for achieving an predefined objective. Algorithms are realized usually by means of implementation as computer programs for execution by automata.

*Other Synonyms*: MIxS: assembly

1. **Genome Coverage**

*Sequencing Assay Field ID*: SA7

*Description*: Depth of sequence coverage based both on external (e.g. Cot-based size estimates) and internal (average coverage in the assembly) measures of genome size.

*Data Categories*: Data Transformation

*Allowed Values*: Free text

*Syntax*: Free text

*Example Values*: 45X average

*Data Source*: GSCID

*Comments*: Required field that is generally not available until genome submitted. 'not applicable' is an allowed value.

*OBO Foundry Synonym*: average depth of sequence coverage

*OBO Foundry Purl*: http://purl.obolibrary.org/obo/OBI\_0001618

*OBI Definition*: An average value of the depth of sequence coverage based both on external (e.g. Cot-based size estimates) and internal (average coverage in the assembly) measures of genome size.

*Other Synonyms*: MIxS:finishing strategy

1. **Annotation Provider**

*Sequencing Assay Field ID*: SA8

*Description*: The name of the responsible person, group or institution providing the set of annotated features for a genome sequences that is submitted to a resource such as GenBank.

*Data Categories*: Data Transformation

*Allowed Values*: Free text

*Syntax*: Free text

*Example Values*: Broad Institute

*Data Source*: GSCID

*Comments*: Newly-proposed GenBank Structured Comment Field ID, sequence annotation' has been submitted to OBI

*OBO Foundry Synonym*: sequence annotation provider

*OBO Foundry purl*: http://purl.obolibrary.org/obo/OBI\_0001947

*OBI Definition (OBI)*: Person or organization report the feature annotation results of a macromolecular sequence.

*Other Synonyms*:

1. **Annotation Method**

*Sequencing Assay Field ID*: SA9

*Description*: The names and versions of the software and databases used in creating the set of annotated features that is submitted to a resource such as GenBank.

*Data Categories*: Data Transformation

*Allowed Values*: Free text

*Syntax*: Free text

*Example Values*:

*Data Source*: GSCID

*Comments*: Newly-proposed GenBank Structured Comment Field ID; sequence annotation' has been submitted to OBI

*OBO Foundry Synonym*: sequence annotation

*OBO Foundry Purl*: http://purl.obolibrary.org/obo/OBI\_0001944

*OBI Definition*: A planned process that identifies and reports sequence features (e.g. protein coding regions) in sequence data.

*Other Synonyms*:

1. **GenBank Record ID**

*Sequencing Assay Field ID*: SA10

*Description*: Unique identifier of the submitted GenBank sequence record(s).

*Data Categories*: Data Transformation

*Allowed Values*: Free text

*Syntax*: Free text

*Example Values*: KF973453

*Data Source*: GSCID

*Comments*: Required field but not available until genome submitted.

*OBO Foundry Synonym*: GenBank ID

*OBO Foundry ID*: http://purl.obolibrary.org/obo/OBI\_0001614

*OBI Definition*: An information content entity that consists of a CRID symbol and additional information about which CRID registry it belongs.

*Other Synonyms*: GenBank accession number