GSA Data Items

Version 1.1   March 2016

Version	Authors	Date	Comments
1.1	Tingting Chen	March 2016	Item description added/corrected
1.0	Yanqing Wang, Junwei Zhu	February 2016	Several new items added
beta	Yanqing Wang, Junwei Zhu	October 2015	Data items summarized

Content

1. BioProject

（1） Submitter

*First name

First name of submitter.

Middle name

Middle name of submitter.

*Last name

Last name of submitter.

*Email (primary)

Primary Email address of submitter.

Email (secondary)

Secondary Email address of submitter.

*Submitting organization

Full name of organization.

Submitting organization URL

The URL of submitter's organization.

*Department

The department of submitter.

Phone

The phone number of submitter.

Fax

The Fax of submitter.

*Street

The Street.

*City

The City.

State/Province

The State/Province.

*Postal code

The Postal code.

*Country/Region

The Country/Region of submitter.

（2） General Info

* Release date

Select "Release immediately following curation" or "Release on specified date". If select "Release on specified date", provide release date in format (YYYY-MM-DD).

Umbrella project

1. eGPS: evolutionary Genotype-Phenotype Systems biology
2. MMDB: Molecule Model Design & Breeding

If your project belongs to a specific umbrella project, please choose it.

*Project title

Provide a brief title, as a phrase or short sentence for public display. Examples: 1) Chromosome Y sequencing; 2) Opportunistic pathogen that causes important food-born disease; 3) Global studies of microbial diversity on human skin.

*Relevance

Select the primary general relevance of the project.

*Public description

Provide a description (a paragraph) of the study purpose and relevance.

External Links

Link description

Display name of web site that is related to this project.

URL

URL of web site that is related to this project.

Related projects

Accession

Related project accession in GSA (e.g. PRJCA000001).

Description

Short description of relationship of the two projects.

Accessions in other database

Accession

Accession of the project in another database.

Database name

The name of the database where the project is also available, such as, NCBI, DDBJ, etc.

Grants

Funding information for a project.

Agency

The name of funding source. For example: 1) National Natural Science Foundation of China; 2) National Program on Key Basic Research Project (973 Program); 3) National High-Tech R&D Program of China (863 Program)

Agency abbr

Abbreviation of funding source. For example: 1) NSFC; 2) 973 Program; 3) 863 Program.

Grant ID

Grant number information is collected in support of searches (e.g., Grant numbers often cited in publications). For example: NSFC Grant Number.

Grant title:

The title of funded grant.

Consortium

Consortium name

If study is carried out as part of a consortium, please provide the consortium name.

Consortium URL

If the consortium has a web site, please provide the URL.

Data provider

Indicate the data provider (data submitter) if it is someone other than the submitting organization or consortium. For example, a sequencing center or a DACC.

Data provider URL

If you would like to present a link to the data provider, please provide the URL.

Biomaterial provider

Name and address of the Lab or PI, or a culture collection identifier.

（3） Project Type

*Project data type

A general label, indicating the primary study goal. Select an appropriate data type.

*Sample scope

The scope and purity of the biological sample used for the study.

*Material

The type of material that is isolated from the sample for use in the experimental study.

*Capture

The type, or scale, of information that the study is designed to generate from the sample material.

*Methodology

The general approach that is used to obtain data.

*Objective

Project goals with respect to the type of data that will be generated. Select all relevant menu options.

（4） Target

*Organism Name (You must select from the hint list which will be displayed after at least 3 letters are provided**)**

Name of the target organism.

Strain, breed, cultivar

Eukaryotic breed, or microbial strain name or cultivar name. Please provide this or "Isolate name or label".

Isolate name or label

Provide a label for an isolated sample, or name of an individual animal (e.g. Clint). Please provide this or "Strain, breed, cultivar".

Public description

Provide a brief description, to elaborate upon the brief label.

Biological Properties

Name and address of the Lab or PI, or a culture collection identifier.

I.   General Properties

General properties of target organism.

Cellularity

Select a cellularity.

Reproduction

Select a Reproduction.

Haploid genome size

Haploid genome size in Kb, Mb or cM.

Ploidy

Select a Ploidy.

II.   Organism Replicon

Describe how many replicons this organism has, how they are named (e.g., 1, 2, 3 vs. I, II, III), the replicon type (chromosome etc.), and the subcellular structure that the replicon is located in.

Name

The preferred standard for the replicon name.

Type

Select a type.

Location

The replicon subcellular location. For instance, the nucleus, or a differentiated organella. Please select "Nuclear or Prokaryote" for the chromosomes of eukaryotes, bacteria or archaea.

Size

The size and unit of measurement for the estimated genome size.

Description

A description of any unusual features of the replicon.

III.   Phenotype

Phenotype of target organism.

Disease

Provide a disease name.

Biotic relationship

Select a BioticRelationship.

Trophic level

Select a TrophicLevel.

IV.   Prokaryote Morphology

Describe the general morphology, when the target organism is prokaryote.

Shape

Select all relevant menu options.

Gram

Choose gram positive or negative.

Motility

Choose motility yes or not.

Enveloped

Choose enveloped or not.

Endospores

Choose target bacteria forms endospores or not.

V.   Ecological Environment

The general habitat for any organism. Please indicate additional extremophile parameters if known.

Habitat

Choose a Habitat.

Salinity

Choose a Salinity.

Oxygen requirement

Choose an Oxygen requirement.

Temperature range

Choose a temperature range.

Optimum temperature

Optimum temperature in Celsius or Farenheit.

（5） Publications

PubMed ID

Provide a PubMed ID.

DOI

Provide a DOI if a PubMed ID is not available.

2. BioSample

（1） Submitter

*First name

First name of submitter.

Middle name

Middle name of submitter.

*Last name

Last name of submitter.

*Email (primary)

Primary Email address of submitter.

Email (secondary)

Secondary Email address of submitter.

*Submitting organization

Full name of organization.

Submitting organization URL

The URL of submitter's organization.

*Department

The department of submitter.

Phone

The phone number of submitter.

Fax

The fax of submitter.

*Street

The street name.

*City

The city name.

State/Province

The state/province of the submitter.

*Postal code

The Postal code of the submitter.

*Country/Region

The Country/Region of the submitter.

（2） General Info

*Release date

Select "Release immediately following curation" or "Release on specified date". If select "Release on specified date", please provide release data in format (YYYY-MM-DD).

*Project accession

Select the project ID.

* Sample title

Provide a brief title, as a phrase or short sentence for public display.

Public description

Provide a description (a paragraph) of the study purpose and relevance.

（3） Sample Type

Select the package that best describes your samples

I. Clinical or Host-associated Pathogen

* Sample Name

A name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. Every Sample Name from a single Submitter must be unique.

* Organism (You must select from the hint list)

The most descriptive organism name for this sample (to the species, if relevant).

Isolate

Identification or description of the specific individual from which this sample was obtained.

Strain

Microbial or eukaryotic strain name.

* collected by

Name of persons or institute who collected the sample.

* collection date Example: 2015-01-01

Date of sampling, in "YYYY-MM-DD".

*geographic location

Geographical origin of the sample; use the appropriate name from this list http://www.insdc.org/documents/country-qualifier-vocabulary. Use a colon to separate the country or ocean from more detailed information about the location, e.g. "China: Hunan" or "China: Beijing, Daxing".

* host (You must select from the hint list)

The natural (as opposed to laboratory) host to the organism from which the sample was obtained.

*host disease (You must select from the hint list)

Name of relevant disease, e.g., Salmonella gastroenteritis.

* isolation source

Describes the physical, environmental and/or local geographical source of the biological sample from which the sample was derived.

* latitude and longitude

The geographical coordinates of the location where the sample was collected. Specify as degrees latitude and longitude in format "d[d.dddd] N|S d[dd.dddd] W|E", e.g., 38.98 N 77.11 W.

culture collection

Name of source institute and unique culture identifier. See information for the proper format and list of allowed institutes at http://www.insdc.org/controlled-vocabulary-culturecollection-qualifier.

genotype

Observed genotype.

host age

Age of host at the time of sampling.

host description

Additional information not included in other defined vocabulary fields.

host disease outcome

Final outcome of disease, e.g., death, chronic disease, recovery.

host disease stage

Stage of disease at the time of sampling.

host health state

Information regarding health state of the individual sampled at the time of sampling.

host sex

Gender or physical sex of the host.

host subject id

A unique identifier by which each subject can be referred to, de-identified, e.g. #131.

host tissue sampled(You must select from the hint list)

Type of tissue the initial sample was taken from.

passage history

Number of passages and passage method.

pathotype

Some bacterial specific pathotypes (example Eschericia coli - STEC, UPEC).

serotype

Taxonomy below subspecies; a variety (in bacteria, fungi or virus) usually based on its antigenic properties. Same as serovar and serogroup. E.g. serotype="H1N1" in Influenza A virus CY098518.

serovar

Taxonomy below subspecies; a variety (in bacteria, fungi or virus) usually based on its antigenic properties. Same as serovar and serotype. Sometimes used as species identifier in bacteria with shaky taxonomy, e.g., Leptospira, serovar saopaolo S76607 (65357 in Entrez).

specimen voucher

Identifier for the physical specimen.

subgroup

Taxonomy below subspecies; sometimes used in viruses to denote subgroups taken from a single isolate.

subtype

Used as classifier in viruses (e.g., HIV type 1, Group M, Subtype A).

II. Environmental, Food or Other Pathogen

* Sample Name

A name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent and as informative as possible. Every sample name from a single Submitter must be unique.

* Organism (You must select from the hint list)

The most descriptive organism name for this sample (to the species, if relevant).

Isolate

Organism group

Identification or description of the specific individual from which this sample was obtained.

Strain

Microbial or eukaryotic strain name.

* collected by

Name of persons or institute who collected the sample.

* collection date Example: 2015-01-01

Date of sampling, in "YYYY-MM-DD".

* geographic location

Geographical origin of the sample; use the appropriate name from this list http://www.insdc.org/documents/country-qualifier-vocabulary. Use a colon to separate the country or ocean from more detailed information about the location, e.g. "China: Hunan" or "China: Beijing, Daxing".

* isolation source

Describes the physical, environmental and/or local geographical source of the biological sample from which the sample was derived.

* latitude and longitude

The geographical coordinates of the location where the sample was collected. Specify as degrees latitude and longitude in format "d[d.dddd] N|S d[dd.dddd] W|E", e.g. 38.98 N 77.11 W.

culture collection

Name of source institute and unique culture identifier. See information for the proper format and list of allowed institutes at http://www.insdc.org/controlled-vocabulary-culturecollection-qualifier.

genotype

Observed genotype.

passage history

Number of passages and passage method.

pathotype

Some bacterial specific pathotypes (example Eschericia coli - STEC, UPEC).

serotype

Taxonomy below subspecies; a variety (in bacteria, fungi or virus) usually based on its antigenic properties. Same as serovar and serogroup, e.g., serotype="H1N1" in Influenza A virus CY098518.

serovar

Taxonomy below subspecies; a variety (in bacteria, fungi or virus) usually based on its antigenic properties. Same as serovar and serotype. Sometimes used as species identifier in bacteria with shaky taxonomy, e.g., Leptospira, serovar saopaolo S76607 (65357 in Entrez).

specimen voucher

Identifier for the physical specimen.

subgroup

Taxonomy below subspecies; sometimes used in viruses to denote subgroups taken from a single isolate.

subtype

Used as classifier in viruses (e.g. HIV type 1, Group M, Subtype A).

III. Microbe

* Sample Name

A name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. Every Sample Name from a single Submitter must be unique.

* Organism (You must select from the hint list)

The most descriptive organism name for this sample (to the species, if relevant).

* strain

microbial or eukaryotic strain name.

* isolation source

Describes the physical, environmental and/or local geographical source of the biological sample from which the sample was derived.

* collection date Example: 2015-01-01

Date of sampling, in "YYYY-MM-DD ".

* geographic location

Geographical origin of the sample; use the appropriate name from this list http://www.insdc.org/documents/country-qualifier-vocabulary. Use a colon to separate the

country or ocean from more detailed information about the location, e.g. "China: Hunan" or "China: Beijing, Daxing".

altitude

The altitude of the sample is the vertical distance between Earth's surface above Sea Level and the sampled position in the air.

biomaterial provider

Name and address of the lab or PI, or a culture collection identifier.

collected by

Name of persons or institute who collected the sample.

culture collection

Name of source institute and unique culture identifier. See information for the proper format and list of allowed institutes at http://www.insdc.org/controlled-vocabulary-culturecollection-qualifier.

depth

Depth is defined as the vertical distance below surface, e.g. for sediment or soil samples depth is measured from sediment or soil surface, respectively. Depth can be reported as an interval for subsurface samples.

environmental biome

Descriptor of the broad ecological context of a sample. Examples include: desert, taiga, deciduous woodland, or coral reef. EnvO (v 2013-06-14) terms can be found via the link: www.environmentontology.org/Browse-EnvO.

genotype

Observed genotype.

host (You must select from the hint list)

The natural (as opposed to laboratory) host to the organism from which the sample was obtained.

host tissue sampled (You must select from the hint list)

Type of tissue the initial sample was taken from.

identified by

Name of the taxonomist who identified the specimen.

lab host

Scientific name and description of the laboratory host used to propagate the source organism or material from which the sample was obtained, e.g., Escherichia coli DH5a, or Homo sapiens HeLa cells.

latitude and longitude

The geographical coordinates of the location where the sample was collected. Specify as degrees latitude and longitude in format "d[d.dddd] N|S d[dd.dddd] W|E", e.g. 38.98 N 77.11 W.

mating type

molecular mechanisms that regulate compatibility in sexually reproducing eukaryotes.

passage history

Number of passages and passage method.

sample size

Amount or size of sample (volume, mass or area) that was collected.

serotype

Taxonomy below subspecies; a variety (in bacteria, fungi or virus) usually based on its antigenic properties. Same as serovar and serogroup. E.g. serotype="H1N1" in Influenza A virus CY098518.

serovar

Taxonomy below subspecies; a variety (in bacteria, fungi or virus) usually based on its antigenic properties. Same as serovar and serotype. Sometimes used as species identifier in bacteria with shaky taxonomy, e.g. Leptospira, serovar saopaolo S76607 (65357 in Entrez).

specimen voucher

Identifier for the physical specimen.

temperature

Temperature of the sample at time of sampling.

IV. Virus Sample

* Sample Name

A name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent and as informative as possible. Every sample name from a single Submitter must be unique.

* Organism (You must select from the hint list)

The most descriptive organism name for this sample (to the species, if relevant).

* isolate

Identification or description of the specific individual from which this sample was obtained.

strain

Microbial or eukaryotic strain name.

host (You must select from the hint list)

The natural (as opposed to laboratory) host to the organism from which the sample was obtained.

lab host

Scientific name and description of the laboratory host used to propagate the source organism or material from which the sample was obtained, e.g., Escherichia coli DH5a, or Homo sapiens HeLa cells.

* collection date Example: 2015-01-01

Date of sampling, in "YYYY-MM-DD".

* geographic location

Geographical origin of the sample; use the appropriate name from this list at http://www.insdc.org/documents/country-qualifier-vocabulary. Use a colon to separate the country or ocean from more detailed information about the location, e.g., "China: Hunan" or "China: Beijing, Daxing".

* isolation source

Describes the physical, environmental and/or local geographical source of the biological sample from which the sample was derived.

altitude

The altitude of the sample is the vertical distance between Earth's surface above Sea Level and the sampled position in the air.

biomaterial provider

Name and address of the Lab or PI, or a culture collection identifier.

collected by

Name of persons or institute who collected the sample.

culture collection

Name of source institute and unique culture identifier. See information for the proper format and list of allowed institutes at http://www.insdc.org/controlled-vocabulary-culturecollection-qualifier.

depth

Depth is defined as the vertical distance below surface, e.g., for sediment or soil samples depth is measured from sediment or soil surface, respectively. Depth can be reported as an interval for subsurface samples.

disease (You must select from the hint list)

List of diseases diagnosed which can include multiple diagnoses. The value of the field depends on host; for humans the terms should be chosen from DO (Disease Ontology), free text for non-human.

environment biome

Descriptor of the broad ecological context of a sample. Examples include: desert, taiga, deciduous woodland, or coral reef. EnvO (v 2013-06-14) terms can be found via the link: www.environmentontology.org/Browse-EnvO.

genotype

Observed genotype.

host tissue sampled (You must select from the hint list)

Type of tissue the initial sample was taken from.

identified by

Name of the taxonomist who identified the specimen.

latitude and longitude

The geographical coordinates of the location where the sample was collected. Specify as degrees latitude and longitude in format "d[d.dddd] N|S d[dd.dddd] W|E", e.g. 38.98 N 77.11 W.

passage history

Number of passages and passage method.

sample size

Amount or size of sample (volume, mass or area) that was collected.

serotype

Taxonomy below subspecies; a variety (in bacteria, fungi or virus) usually based on its antigenic properties. Same as serovar and serogroup. E.g. serotype="H1N1" in Influenza A virus CY098518.

specimen vouch

Identifier for the physical specimen.

temperature

Temperature of the sample at time of sampling.

V. Human Sample

* Sample Name

A name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. Every Sample Name from a single Submitter must be unique.

* Organism (You must select from the hint list)

The most descriptive organism name for this sample (to the species, if relevant).

* isolate

Identification or description of the specific individual from which this sample was obtained.

age

Age at the time of sampling; relevant scale depends on species and study, e.g., could be seconds for amoebae or centuries for trees.

* biomaterial provider

Name and address of the Lab or PI, or a culture collection identifier.

* sex

Physical sex of sampled organism.

* tissue (You must select from the hint list)

Type of tissue the sample was taken from.

disease (You must select from the hint list)

List of diseases diagnosed which can include multiple diagnoses. the value of the field depends on host; for humans the terms should be chosen from DO (Disease Ontology), free text for non-human.

cell line

Name of the cell line.

cell subtype

The subtype of cell.

cell type (You must select from the hint list)

Type of cell of the sample or from which the sample was obtained.

culture collection

Name of source institute and unique culture identifier. See information for the proper format and list of allowed institutes at http://www.insdc.org/controlled-vocabulary-culturecollection-qualifier.

development stage

Developmental stage at the time of sampling.

disease stage

Stage of disease at the time of sampling.

ethnicity

Ethnicity of the subject.

health state

Health or disease status of sample at time of collection.

karyotype

Karyotype of sampled organism.

phenotype

Phenotype of sampled organism. For Phenotypic quality Ontology (PATO) (v1.269) terms, please see at http://bioportal.bioontology.org/visualize/44601.

population

Filial generation, number of progeny, genetic structure (For human and plants).

race

Race of sample.

type

Sample type, such as cell culture, mixed culture, tissue sample, whole organism, single cell, metagenomic assembly.

treatment

Treatment of sample.

VI. Model Organism or Animal Sample

* Sample Name

A name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. Every Sample Name from a single Submitter must be unique.

* Organism (You must select from the hint list)

The most descriptive organism name for this sample (to the species, if relevant).

Breed

Breed name - chiefly used in domesticated animals or plants.

Strain

Microbial or eukaryotic strain name.

Age

Age at the time of sampling; relevant scale depends on species and study, e.g., could be seconds for amoebae or centuries for trees.

* biomaterial provider

Name and address of the Lab or PI, or a culture collection identifier.

* sex

Physical sex of sampled organism.

* tissue (You must select from the hint list)

Type of tissue the sample was taken from.

birth date Example: 2015-01-01

The data of birth.

birth location

The location of birth.

breeding history

The history of breeding.

breeding method

The method of breeding.

cell line

Name of the cell line.

cell subtype

The subtype of cell.

cell type

Type of cell of the sample or from which the sample was obtained.

collected by

Name of persons or institute who collected the sample.

culture collection

Name of source institute and unique culture identifier. See information for the proper format and list of allowed institutes at http://www.insdc.org/controlled-vocabulary-culturecollection-qualifier.

death date Example: 2015-01-01

The data of death.

development stage

Age/stage group

Developmental stage at the time of sampling.

disease (You must select from the hint list)

List of diseases diagnosed which can include multiple diagnoses. The value of the field depends on host; for humans the terms should be chosen from DO (Disease Ontology),

free text for non-human.

disease stage

Stage of disease at the time of sampling.

genotype

Observed genotype.

geographic location

Geographical origin of the sample; use the appropriate name from this list http://www.insdc.org/documents/country-qualifier-vocabulary. Use a colon to separate the country or ocean from more detailed information about the location, e.g. "China: Hunan" or "China: Beijing, Daxing".

growth protocol

The protocol of growth.

health state

Health or disease status of sample at time of collection.

isolation source

Describes the physical, environmental and/or local geographical source of the biological sample from which the sample was derived.

latitude and longitude

The geographical coordinates of the location where the sample was collected. Specify as degrees latitude and longitude in format "d[d.dddd] N|S d[dd.dddd] W|E", e.g. 38.98 N 77.11 W

Phenotype

Phenotype of sampled organism. For Phenotypic quality Ontology (PATO) (v1.269) terms, please see http://bioportal.bioontology.org/visualize/44601.

specimen voucher

Identifier for the physical specimen.

storage conditions

Explain how and for how long the soil sample was stored before DNA extraction.

study book number

Study book number of sample.

treatment

Treatment of sample.

VII. Plant Sample

Use for any plant sample or cell line.

* Sample Name

Sample Name is a name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. Every Sample Name from a single Submitter must be unique.

* Organism (You must select from the hint list)

The most descriptive organism name for this sample (to the species, if relevant).

* cultivar

Organism group

Cultivar name - cultivated variety of plant.

* biomaterial provider

Name and address of the Lab or PI, or a culture collection identifier.

* tissue (You must select from the hint list)

Type of tissue the sample was taken from.

age

Age/stage group

Age at the time of sampling; relevant scale depends on species and study, e.g. could be seconds for amoebae or centuries for trees.

cell line

Name of the cell line.

cell type

Type of cell of the sample or from which the sample was obtained.

collected by

Name of persons or institute who collected the sample.

collection date Example: 2015-01-01

Date of sampling, in "YYYY-MM-DD".

culture collection

Name of source institute and unique culture identifier. See information for the proper format and list of allowed institutes at http://www.insdc.org/controlled-vocabulary-culturecollection-qualifier.

development stage

Age/stage group

Developmental stage at the time of sampling.

disease (You must select from the hint list)

List of diseases diagnosed which can include multiple diagnoses. The value of the field depends on host; for humans the terms should be chosen from DO (Disease Ontology), free text for non-human.

disease stage

Stage of disease at the time of sampling.

genotype

Observed genotype.

growth protocol

The protocol of growth.

height or length

Measurement of height or length.

isolation source

Describes the physical, environmental and/or local geographical source of the biological sample from which the sample was derived.

latitude and longitude

The geographical coordinates of the location where the sample was collected. Specify as degrees latitude and longitude in format "d[d.dddd] N|S d[dd.dddd] W|E", e.g. 38.98 N 77.11 W.

phenotype

Phenotype of sampled organism. For Phenotypic quality Ontology (PATO) (v1.269) terms, please see http://bioportal.bioontology.org/visualize/44601.

population

Filial generation, number of progeny, genetic structure (For human and plants).

type

Sample type, such as cell culture, mixed culture, tissue sample, whole organism, single cell, metagenomic assembly.

sex

Physical sex of sampled organism.

specimen vouch

Identifier for the physical specimen.

temperature

Temperature of the sample at time of sampling.

treatment

Treatment of sample.

3. Experiment

（1） Meta Information

*Platform

The sequencing platform and instrument model.

*Alias

Submission name of the experiment. This field is used when the record does not yet have an accession and needs to be referenced by other objects.

*Title

Short text that can be used to call out experiment records in searches or in displays. This element is technically optional but should be used for all new records.

*Project accession

Link data to BioProject that describes the research.

* Sample accession

Enter a BioSample or SRA Sample Accession. BioSample accessions have 'SAMN' prefix. SRA Sample Accessions have 'SRS' prefix. A BioSample describes the biological source material for your sequence library preparation.

Library Construction/Experiment design

Additional details about your experimental design and molecular strategies, including hybrid selection and affinity capture reagents; any detail that distinguishes your experiment from other similar experiments.

（2） Library

The library descriptor specifies the origin of the material being sequenced and any treatments that the material might have undergone that affect the sequencing result. This specification is needed even if the platform does not require a library construction step per se.

Library name

The submitter's name for this library.

*Strategy

Sequencing technique intended for this library.

*Source

The library source specifies the type of source material that is being sequenced.

*Selection

Whether any method was used to select and/or enrich the material being sequenced.

*Layout

Library Layout specifies whether to expect single, Pair-end, or other configuration of reads. In the case of paired reads, information about the relative distance and orientation is

specified.

Nominal size (bp)

Size of the insert for Paired reads.

Nominal standard deviation (bp)

Standard deviation of insert size (typically ~10% of Nominal Size)

Processing

This section varies with the sequencer selected. Please pay close attention to the answers provided in this section, as they may affect proper loading of data.

Release date

Giving the release data.

4. Run

（1） New Run

General info

* Alias

Submitter assigned name or id for the GSA submission object.

* Run data file type

The storage format of the sequence data being submitted.

Data blocks

File name

Name of the file transferred to GSA which we recommend to be compressed format (e.g. *.zip or *.tar.gz)

MD5 checksum

32-character alphanumeric string that can be computed with native command line tools 'md5' (Mac OS X) or 'md5sum' (Linux). Windows users will need to download a 3rd-party utility.

Release date

Submitter gave the release data.