Genome 10K Data Submission Guidelines

Ngan Nguyen <nknguyen@soe.ucsc.edu>
Mark Diekhans <markd@soe.ucsc.edu>

Version 0.05: November 2, 2009

1 Introduction

This document assists the submission of data to Genome 10K. It describes required information and format of a correct submission.

2 Guidelines

These guidelines are to be followed for each submission:

- Submitters are to provide two documents: an `animal-sample` document and a `steward` document.

  `animal-sample` document (Table 1): a detailed list of the submitted samples, which contains all necessary information of the samples, such as the animal from which each sample is extracted, voucher number, species name, preservative type, storage location, steward, quantity, quality, and etc.

  `steward` document (Table 2): contains contact information of the individuals and institutions responsible for each animal and sample listed in the `animal-sample` document. Animals and samples may have different stewards.

- All data must follow the format specified in column "Accepted values" . If a set of values is provided in the column, then the input data for that field can be selected only from among those values. If submitters would like to add a new value, please email <submit-help@genome10k.org>

- Fields (columns) with an asterisk mark (*) are required. Other fields are optional and can be omitted or left blank.

- Samples with a submitted BOLD accession number can skip (**) marked columns. These columns will be imported automatically from the BOLD record.

- Submission must be either Excel spreadsheets or tab-separated files with column headers.

- The order of columns is not important.

3 Submission documents

3.1 `animal-sample` document (Table 1)

Field-name	Explanation	Accepted Values	Example
g10kAnimalId	id number assigned by G10K project - required for updated submissions and omitted for new submissions	numeric	123
genus*		alphabetical	Arremonops
species*		alphabetical	rufivirgatus
subspecies		alphabetical	
commonName		alphabetical, space, comma	olive sparrow
barcodeID	BOLD [4] barcode ID	alphanumeric, hyphen	KKBNA837-05
voucherCatalogNumber**		numeric	335431
voucherLocation**		alphabetical	MCZ
sex**		male, female, hermaphrodite, unknown, or other	unknown
animalSteward	steward for the animal or animal tissue	First Last	David M. Brown
stewardAnimalId	identifier for the animal used by the steward	free text	CF0001
birthDate		any date type	2009-07-28
birthLongtitude	longitude in decimal degrees	numeric	-12.345
birthLatitude	latitude in decimal degrees	numeric	12.345
birthDepth	depth or elevation in meters	numeric	100
deathDate		any date type	2009-07-28
collectionDate**		any date type	2009-07-28
collectionLongtitude**	longitude in decimal degrees	numeric	10.511
collectionLatitude**	latitude in decimal degrees	numeric	-85.363
collectionDepth**	depth or elevation in meters	numeric	320
collectors**	comma delimited list of collectors	First Last, First Last, . . .	Sue Smith, Sam Lee
hasCollectionPermit	whether or not the steward has permit for collection	yes, no, unknown, NA	yes

acucPermitsStatus	steward has permit of their institution's Animal Care and Use Committee	confirmed, pending, unknown, NA	NA
citesStatus	Convention on International Trade in Endangered Species of Wild Fauna and Flora for the species[2] appendix	I, II, III	I
citesPermitStatus	status of CITES permit for this animal	confirmed, pending, unknown, NA	pending
usesaStatus	U.S. Endangered Species Act [3] status for the species	LE, LT, PE, PT, C, SC, PDL, SAE/SAT, PSAE/PSAT, XE, XN, blank	C
usesaPermitStatus	status of USESA permit for this animal	confirmed, pending, unknown, NA	confirmed
animalNotes	notes on the animal of interest to G10K	free text	our finest specimen
g10kSampleId	id number assigned by G10K project - required for updated submissions and omitted for new submission	numeric	12345
sampleSteward**	steward for the sample, if different from animalSteward	First Last	David M. Brown
stewardSampleId	sample identifier used by the steward	free text	CFA0001-DN-01
tissueSampleIsTakenFrom		blood, brain, eggs, eye lense, fin, foot, general, gills, gut, heart, intestines, kidney, liver, muscle, ova, skin, spleen, stomach, tadpoles, tail, testes, toe, viscera, whole, wing, other	muscle
sampleType		DNA, blood, other tissue, cell line	DNA

sampleQuantity	quantity estimate	value(numeric) unit(g, mcg/ug/μg, ml) [1]	100 ug
sampleQuality		free text	
storageLocation		alphabetical	MCZ
preservativeType		EDTA, TE, Tris, Heparin, DMSO, ethanol, frozen (or specific temperature if available, e.g "-50C") , none, other	TE
sampleNotes	notes on the sample of interest to G10K	free text	a prefect sample

Table 1: `animal-sample`

3.2 `steward` document Table 2

The following fields are defined:

Field-name	Accepted Value
g10kStewardId (assigned by G10K project)	numeric
name	First Last
institution	free text
address	free text
email	free text
telephone	numeric and dashes, including country code
fax	numeric and dashes, including country code
notes	free text

Table 2: `steward`

4 Terminology

- An `animal` describes each individual animal tracked by the database. Biological samples from an animal are tracked in separate records, allowing multiple samples to be tracked for a given individual animal.
- A `sample` describes a biological sample from an animal. Multiple samples may be created from the same or different tissues from the same animal, including cell lines.

- Alphabetical: only the characters a-z, A-Z, and underscore (_) are allowed. Blank, space, and other characters are not accepted. Entries with "genus species", "name1/name2", "name1 x name2", "sp.", "cf.", "aff." are not accepted.
- Numeric: only the characters 0-9, floating point ".", and negative sign "-" are allowed.
- Date type: can be any format with date, month, and year information.
- * means required.
- ** means information can be taken automatically from the BOLD record if a BOLD accession number is provided.

References

[1] Abbreviations of Units of Measurement. Enlighsplus (http://englishplus.com/grammar/00000058.htm). Web, 2009.

[2] Appendices I, II and III. Convention on International Trade in Endangered Species of Wild Fauna and Flora (http://www.cites.org/). Web, May 2009.

[3] U.S. Endangered Species Act. NatureServe Explorer (http://www.natureserve.org/explorer/). Web, 2009.

[4] S. Ratnasingham and P. D. N. Hebert. BOLD: The barcode of life data system (http://www.barcodinglife.org). *Molecular Ecology Notes*, 7:355–364, 2007.