

Example Records

Source Record

<i>Observed Sequence</i>	
id	: O100
version	: 23
type	: mRNA
db	: Genbank
ref	: GI:BB234567
position	: 50-250
cnf_score	: 1
sequence	: acgu...
md5_sum	: e8f612a77ac1
...	

<i>Reference Genome</i>	
id	: R100
version	: 83
assembly	: GRCh38
patch	: p7
gff_seqid	: CHR_HSCHR14_3_CTG1
gff_start	: 106739363
gff_end	: 106739862
gff_strand	: +
cnf_score	: 1
sequence	: agct...
md5_sum	: e8f612a77ac1
...	

<i>Publication</i>	
id	: P100
version	: 93
PubMed id	: 25501908
doi	: 10.1038/nm.3743
Citation	: In-depth determination...
cnf_score	: 1
...	

Gene Description

id	: D100
version	: 9
authors	:
species	: Mus musculus
subspecific_name	: domesticus
additional_info	: BALB/C
canonical_name	: IGHV2-6-7*01
Alt-name-1	: musIGHV134
Alt-name-2	:
...	
inference_class	: 2
cnf_score	: 2
paralog_of	: D101
...	
delineation:	
delin_name	: IMGT
FR1	: 1-24
CDR1	: 25-30
...	
delin_name	: Chothia
FR1	: 1-30
CDR1:	: 31-36
...	
observed_sequences	: O500,O503,...
reference_genome_links	: R456, R837,...
publications	: P398, P8902, ...

Germline Set

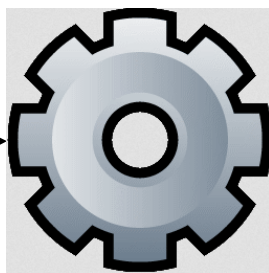
id	: S100
version	: 48
title	: "High confidence B-cell repertoire of the BALB/C mouse"
authors	:
reported_in	: P302, P109,...
species	: Mus musculus
subspecific_name	: domesticus
additional_info	: BALB/C
...	
genes	: D100, D101,...
...	

Only illustrative fields are shown. For the full list of fields, consult the schema definition.

OGRDB In Use



Germline
Set

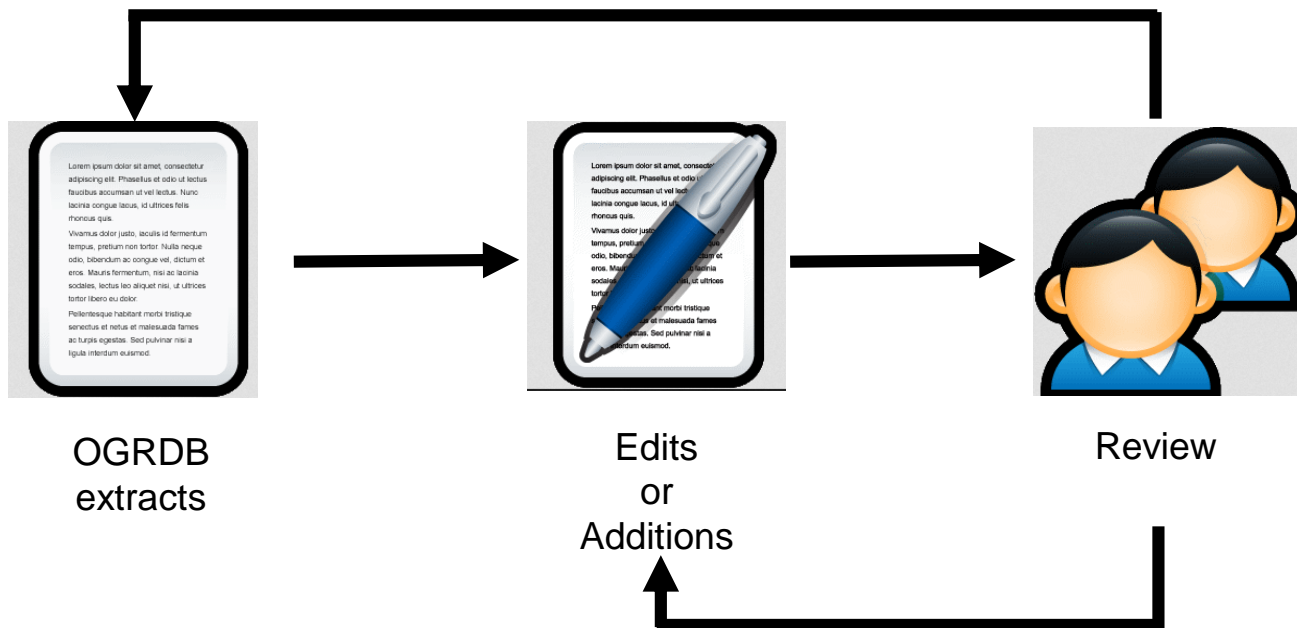


Analysis
Tool

OGRDB will generate germline sets for use by repertoire parsers and other tools.

The format of the germline set will be convenient for tools to import.

Some users may wish to modify the germline sets, or produce their own sets in the same format: we hope to establish a standard for the interchange of germline sets.



Curators will download records from OGRDB for review and modification.

Downloaded records will be provided in CSV format for convenience.

OGRDB will support a review and acceptance process.

Proposed modifications will be checked by OGRDB.

Some fields will be auto-populated (e.g, if a PubMed ID is provided, OGRDB will populate the citation and doi fields).

All records have a version number which is generated by OGRDB. This will be used to resolve conflicting updates to the same record by different editors.

OGRDB File Formats

Source Record

All source records of a particular type, or an extract

id	version
O001	23	
O002	54	
...		

Germline Set Definition

The contents of a Germline Set Record

id : S100

version : 48

title : "High confidence B-cell repertoire of the BALB/C mouse"

authors :

reported_in : P302, P109,...

species : Mus musculus

subspecific_name : domesticus

additional_info : BALB/C

genes : D100, D101,...

...

Gene Description

A Gene Description Record, plus all associated Source Records in full

id : D100

version : 9

authors :

species : Mus musculus

subspecific_name : domesticus

additional_info : BALB/C

canonical_name : IGHV2-6-7*01

Alt-name-1 : musIGHV134

Alt-name-2 :

...

inference_class : 2

cnf_score : 2

paralog_of : D101

...

delineation:

delin_name : IMGT

FR1 : 1-24

CDR1 : 25-30

...

Observed Sequences

id	version
O045	12	
...		

Reference Genome Sequences

id	version
R045	12	
...		

Publications

id	version
P045	12	
...		

Primarily for curators

←

Primarily for tool users

→

Germline Set

The contents of a Germline Set Record, plus extracts from all associated Gene Descriptions

id : S100

version : 48

title : "High confidence B-cell repertoire of the BALB/C mouse"

authors :

reported_in : P302, P109,...

species : Mus musculus

subspecific_name : domesticus

additional_info : BALB/C

id	version
D100	12	
D101	54	
...		

The Gene Description and Germline Set formats are non-canonical files meant to ease curation or use in specific circumstances