KAIKObase manual

Version 4

October 2020

Outline

1 Introduction

2 Top page

- 2.1 Chromosome Overview
- 2.2 Keyword and Position Search
- 2.3 BLAST Search
- 2.4 Curated Genes

3 Description page

4 Genome browser

5 Contact

1 Introduction

KAIKObase is a silkworm genome database with a genetic map viewer and a genome viewer. The latest KAIKObase (version 4) was constructed based on the chromosomal-level genome assembly released in 2019 (10.1016/j.ibmb.2019.02.002) together with the genetic markers and old gene sets from the previous KAIKObase.

New updates in KAIKObase version 4 include that:

1. New genome assembly is used as the reference genome in the genome browser GBrowse.
2. 16,880 new gene models (protein-coding genes) and 51,926 reference transcripts are added and accessible in the genome browser. They are searchable in "BLAST search" and "Keyword search".
3. Functional annotations of each new gene model and its expression patterns in various tissues, as well as orthologs in closely related lepidopterans and model vertebrate and insect organisms, are available in "description page".

Old datasets kept in new KAIKObase include that:

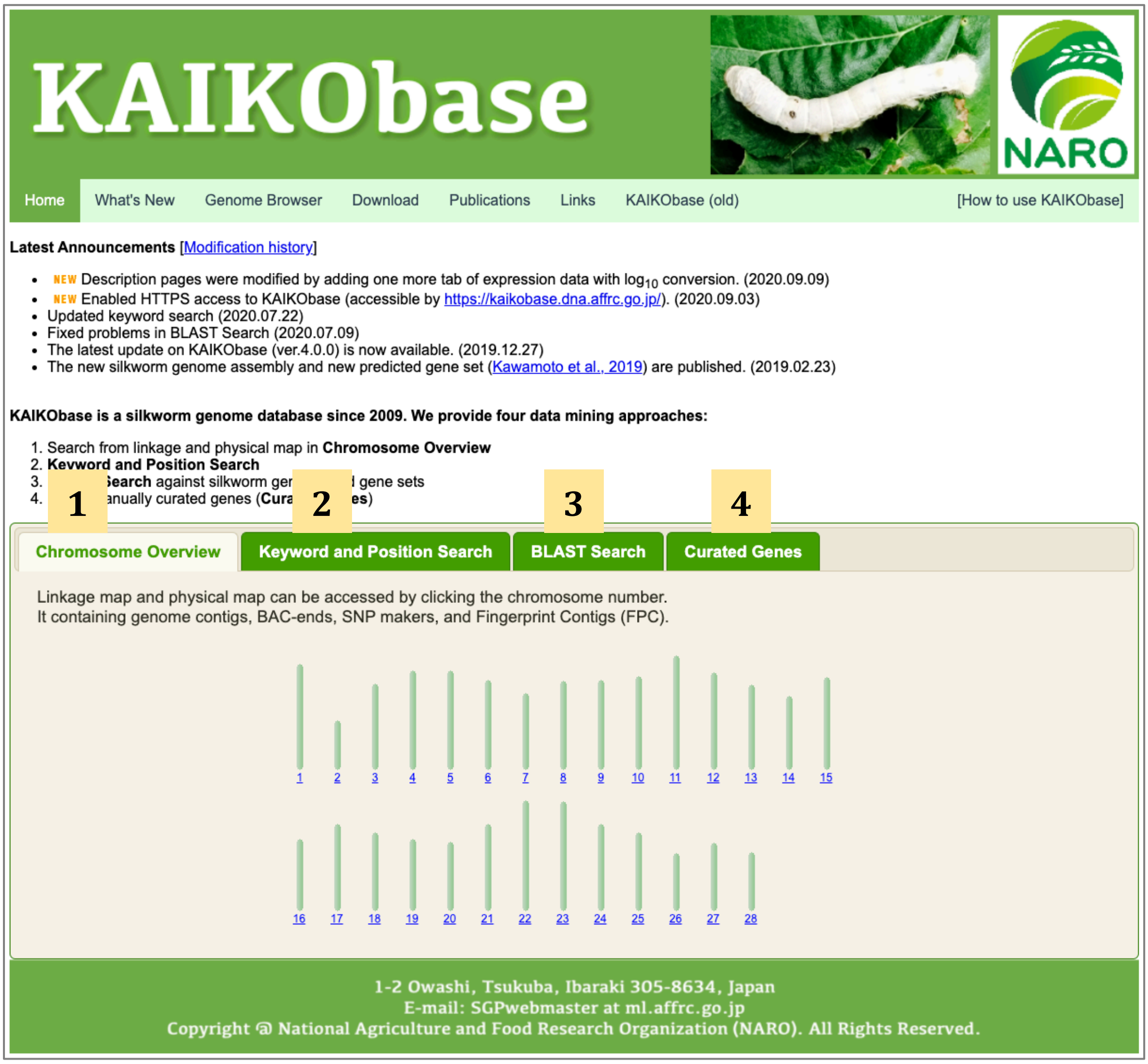
1. Genetic markers (SNP markers, BAC-end sequences, and fingerprint contigs) and scaffolds of old genome assembly are available on the base of new genome assembly.
2. Old protein-coding gene sets (Gene set A, China gene model, FLcDNA, and EST) were mapped onto new genome and are searchable in "BLAST search" and "Keyword search" and accessible in genome browser.

2 Top page

The top page of KAIKObase contains the announcement about database update and the data mining pages.

KAIKObase provides four data mining approaches as listed on the tabs:

1 Search from linkage and physical map in **Chromosome Overview**
2 Keyword and Position Search
3 BLAST Search against silkworm genome and gene sets
4 List of manually **curated genes**



Latest Announcements [Modification history]

- NEW Description pages were modified by adding one more tab of expression data with $\log_{10}$ conversion. (2020.09.09)
- NEW Enabled HTTPS access to KAIKObase (accessible by https://kaikobase.dna.affrc.go.jp/). (2020.09.03)
- Updated keyword search (2020.07.22)
- Fixed problems in BLAST Search (2020.07.09)
- The latest update on KAIKObase (ver.4.0.0) is now available. (2019.12.27)
- The new silkworm genome assembly and new predicted gene set (Kawamoto et al., 2019) are published. (2019.02.23)

KAIKObase is a silkworm genome database since 2009. We provide four data mining approaches:

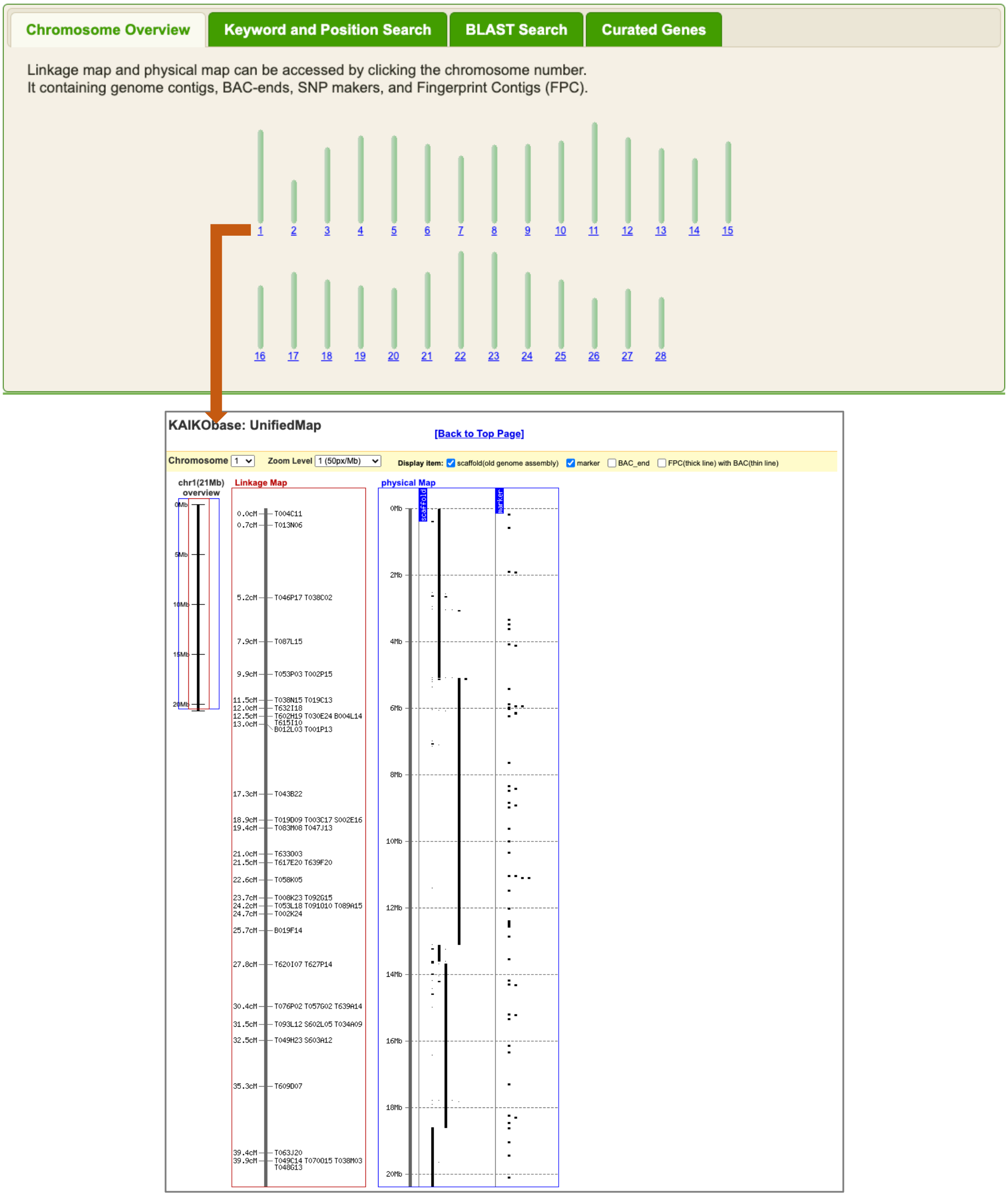
1. Search from linkage and physical map in **Chromosome Overview**
2. **Keyword and Position Search**
3. **earch** against silkworm ger d gene sets
4. anually curated genes (**Cura** **es**)

Linkage map and physical map can be accessed by clicking the chromosome number.
It containing genome contigs, BAC-ends, SNP makers, and Fingerprint Contigs (FPC).

1-2 Owashi, Tsukuba, Ibaraki 305-8634, Japan
E-mail: SGPwebmaster at ml.affrc.go.jp
Copyright @ National Agriculture and Food Research Organization (NARO). All Rights Reserved.

2.1 Chromosome Overview

Linkage and physical map of 28 chromosomes are provided. Click the number under the chromosome to see the genetic map. Detailed information of genetic map will be described in the next page.



Linkage map and physical map can be accessed by clicking the chromosome number.
It containing genome contigs, BAC-ends, SNP makers, and Fingerprint Contigs (FPC).

KAIKObase: UnifiedMap

Chromosome 1
Zoom Level 1 (50px/Mb)
Display item: ☑ scaffold(old genome assembly) ☑ marker ☐ BAC_end ☐ FPC(thick line) with BAC(thin line)

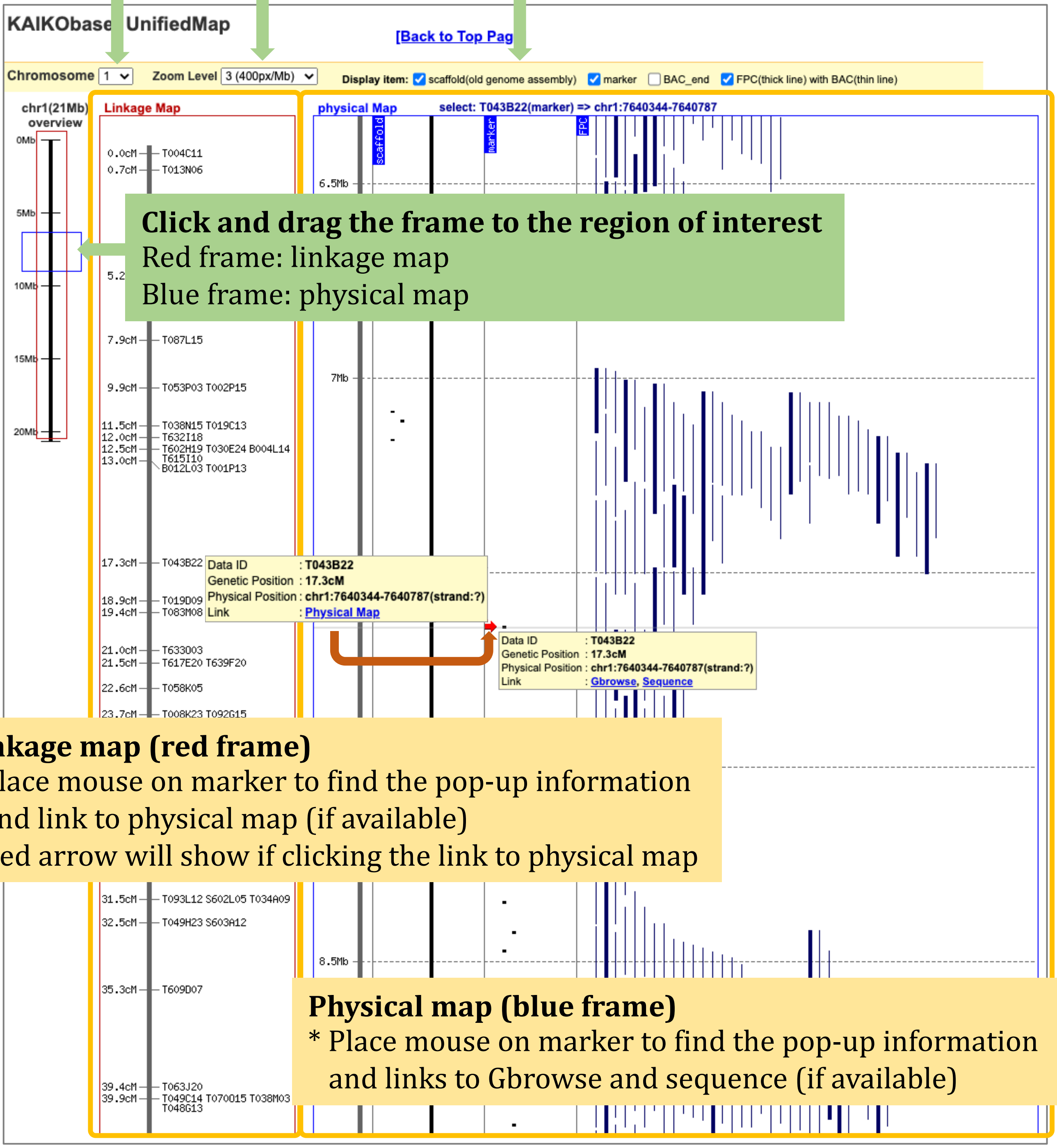
0.0cM	T004C11
0.7cM	T013N06
5.2cM	T046P17 T038C02
7.9cM	T087L15
9.9cM	T053P03 T002P15
11.5cM	T038N15 T019C13
12.0cM	T632I18
12.5cM	T602H19 T030E24 B004L14
13.0cM	T615I10 B012L03 T001P13
17.3cM	T043B22
18.9cM	T019D09 T003C17 S002E16
19.4cM	T083M08 T047J13
21.0cM	T633O03
21.5cM	T617E20 T639F20
22.6cM	T058K05
23.7cM	T008K23 T092G15
24.2cM	T053L18 T091O10 T089A15
24.7cM	T002K24
25.7cM	B019F14
27.8cM	T620I07 T627P14
30.4cM	T076P02 T057G02 T639A14
31.5cM	T093L12 S602L05 T034A09
32.5cM	T049H23 S603A12
35.3cM	T609D07
39.4cM	T063J20
39.9cM	T049C14 T070O15 T038M03 T048G13

Features of genetic map (UnifiedMap)

Specify zooming level

Select chromosome

Check items to display



KAIKObase UnifiedMap

[Back to Top Pag

Chromosome 1

Zoom Level 3 (400px/Mb)

Display item: ☑ scaffold(old genome assembly) ☑ marker ☐ BAC_end ☑ FPC(thick line) with BAC(thin line)

Data ID : T043B22
Genetic Position : 17.3cM
Physical Position : chr1:7640344-7640787(strand:?)
Link : Physical Map

Data ID : T043B22
Genetic Position : 17.3cM
Physical Position : chr1:7640344-7640787(strand:?)
Link : Gbrowse, Sequence

Click and drag the frame to the region of interest
Red frame: linkage map
Blue frame: physical map

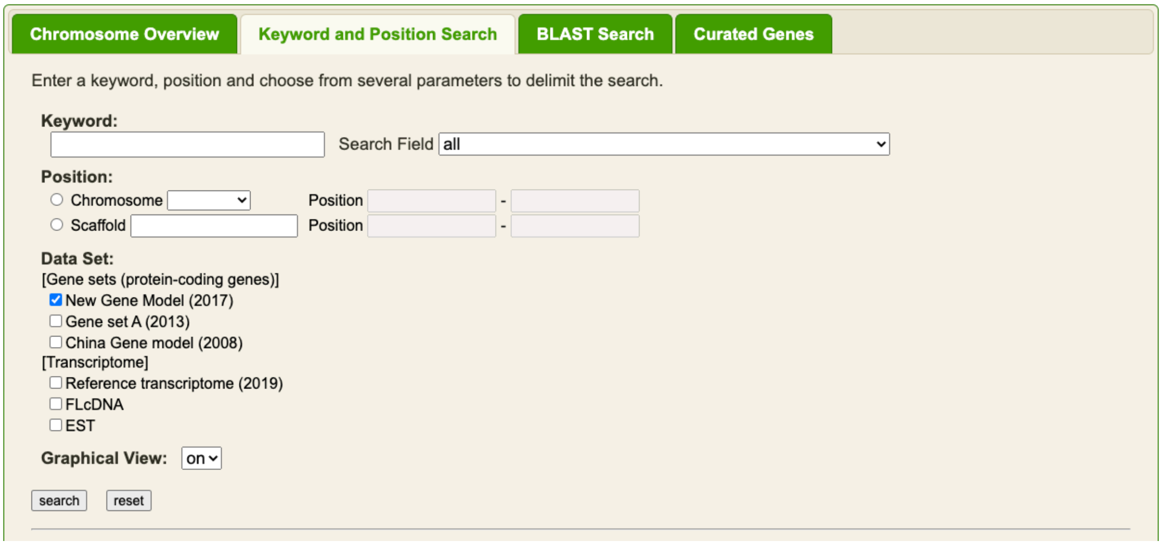
Linkage map (red frame)
- * Place mouse on marker to find the pop-up information and link to physical map (if available)
- * Red arrow will show if clicking the link to physical map

Physical map (blue frame)
- * Place mouse on marker to find the pop-up information and links to Gbrowse and sequence (if available)

2.2 Keyword and Position Search

Genes of interest can be searched by
1. Keyword (Gene ID, annotation, etc.)
2. Chromosomal position of a given region
3. A combination of keyword and position

The data set “New gene model” is selected for searching by default.



Enter a keyword, position and choose from several parameters to delimit the search.

Keyword:

Search Field all

Position:

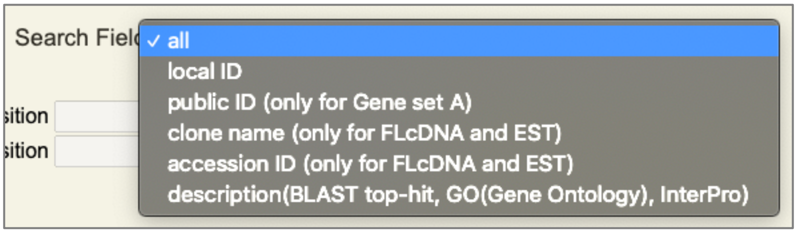
○ Chromosome Position -
○ Scaffold Position -

Data Set:

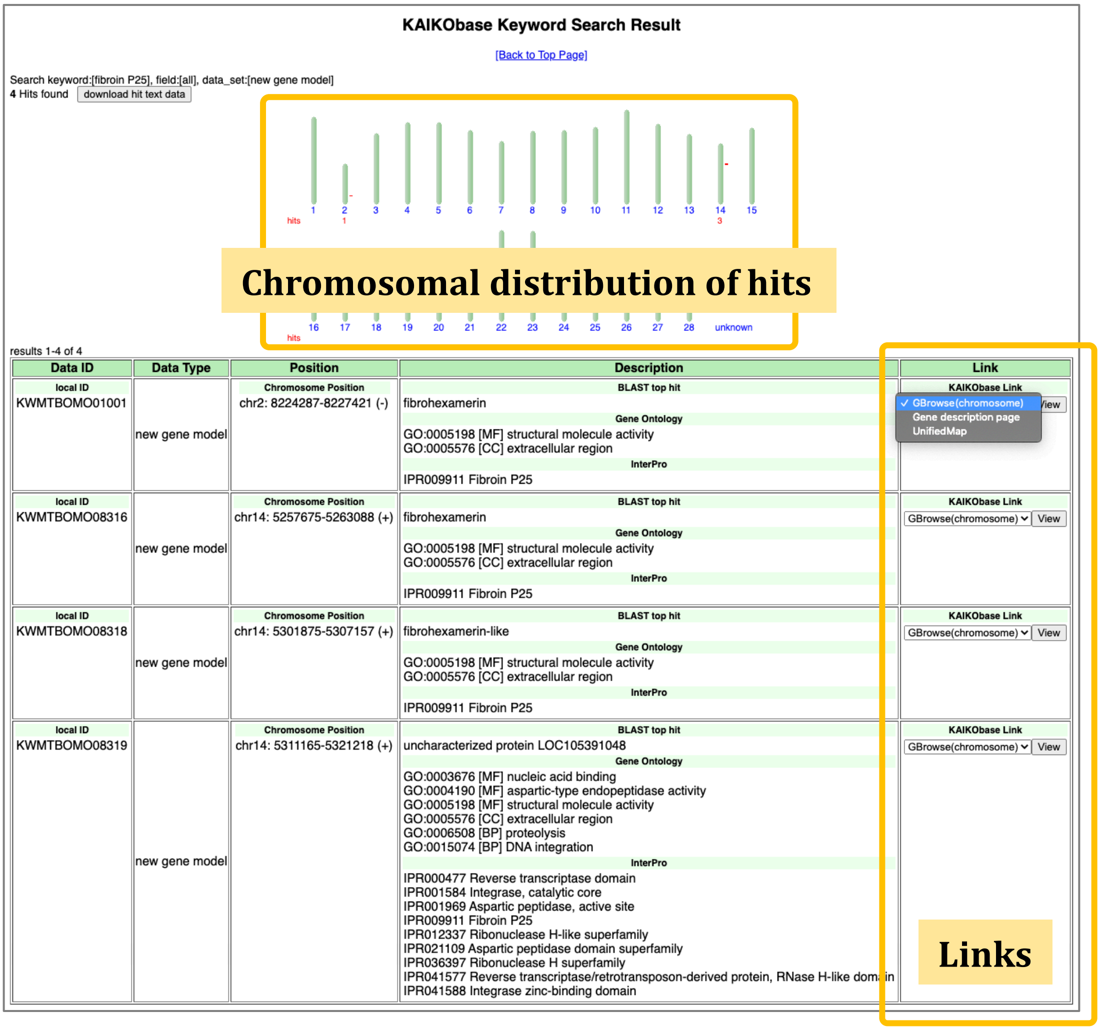
[Gene sets (protein-coding genes)]
☑ New Gene Model (2017)
☐ Gene set A (2013)
☐ China Gene model (2008)
[Transcriptome]
☐ Reference transcriptome (2019)
☐ FLcDNA
☐ EST

Graphical View: on

The entries of the pull-down menu in “Search Field” are:
- **all (default)** – applicable for all datasets
- **local ID** – applicable for all data sets
- **public ID** – used in Gene set A
- **clone name** – used in FLcDNA, EST
- **accession ID** – used in FLcDNA, EST
- **description** – applicable for all data sets but EST



“Search Results” show the hits on each chromosome and the list of hits. The pull-down menu provides links to genome browser and map viewer.



KAIKObase Keyword Search Result

[Back to Top Page]

Search keyword:[fibroin P25], field:[all], data_set:[new gene model]
4 Hits found

Chromosomal distribution of hits

results 1-4 of 4

Data ID	Data Type	Position	Description	Link
local ID KWMTBOMO01001	new gene model	**Chromosome Position** chr2: 8224287-8227421 (-)	**BLAST top hit** fibrohexamerin **Gene Ontology** GO:0005198 [MF] structural molecule activity GO:0005576 [CC] extracellular region **InterPro** IPR009911 Fibroin P25	**KAIKObase Link**
local ID KWMTBOMO08316	new gene model	**Chromosome Position** chr14: 5257675-5263088 (+)	**BLAST top hit** fibrohexamerin **Gene Ontology** GO:0005198 [MF] structural molecule activity GO:0005576 [CC] extracellular region **InterPro** IPR009911 Fibroin P25	**KAIKObase Link** GBrowse(chromosome) View
local ID KWMTBOMO08318	new gene model	**Chromosome Position** chr14: 5301875-5307157 (+)	**BLAST top hit** fibrohexamerin-like **Gene Ontology** GO:0005198 [MF] structural molecule activity GO:0005576 [CC] extracellular region **InterPro** IPR009911 Fibroin P25	**KAIKObase Link** GBrowse(chromosome) View
local ID KWMTBOMO08319	new gene model	**Chromosome Position** chr14: 5311165-5321218 (+)	**BLAST top hit** uncharacterized protein LOC105391048 **Gene Ontology** GO:0003676 [MF] nucleic acid binding GO:0004190 [MF] aspartic-type endopeptidase activity GO:0005198 [MF] structural molecule activity GO:0005576 [CC] extracellular region GO:0006508 [BP] proteolysis GO:0015074 [BP] DNA integration **InterPro** IPR000477 Reverse transcriptase domain IPR001584 Integrase, catalytic core IPR001969 Aspartic peptidase, active site IPR009911 Fibroin P25 IPR012337 Ribonuclease H-like superfamily IPR021109 Aspartic peptidase domain superfamily IPR036397 Ribonuclease H superfamily IPR041577 Reverse transcriptase/retrotransposon-derived protein, RNase H-like domain IPR041588 Integrase zinc-binding domain	**KAIKObase Link** GBrowse(chromosome) View

Links

Besides links to **genome browser** and **map viewer** (“UnifiedMap” in the menu), additional links are also provided as follows.

- **New gene model**: description page
- **Gene set A**: description page
- **China gene model**: gene viewer*, KAIKO2DDB*
- **FLcDNA**: FLcDNA DB*
- **EST**: EST DB*

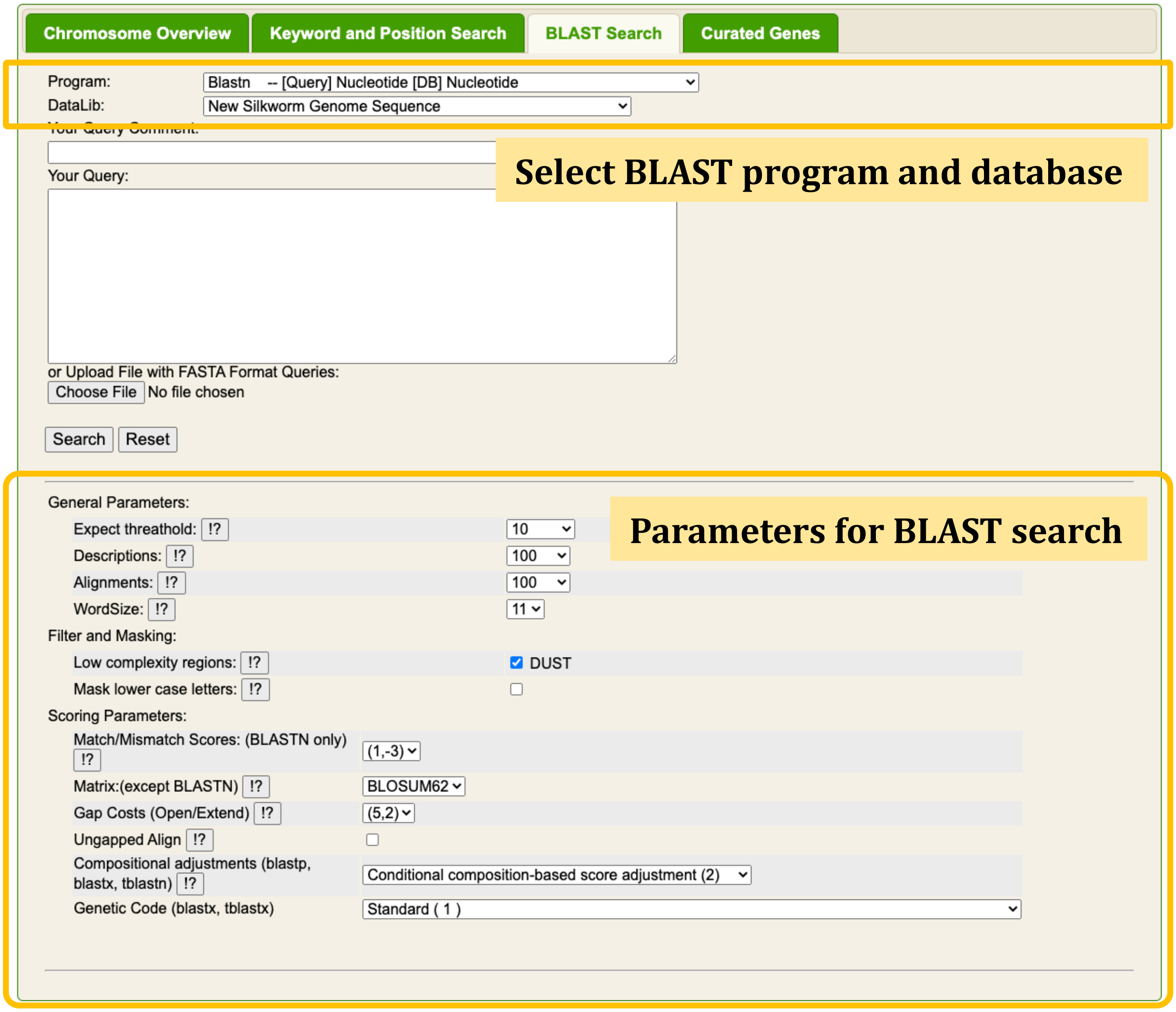
*: link to old KAIKObase

2.3 BLAST Search

“BLAST Search” provides five programs to search against five data sets.

Blastn, **tBlastn**, **tBlastx** can be used to search against
- Genome sequence
- New gene model (nucleotide)
- Reference transcript (nucleotide)

Blastp, **Blastx** can be used to search against
- New gene model (amino acid)
- Reference transcript (amino acid)



Program: Blastn -- [Query] Nucleotide [DB] Nucleotide
DataLib: New Silkworm Genome Sequence
No file chosen

General Parameters:
Expect threathold: 10
Descriptions: 100
Alignments: 100
WordSize: 11

Filter and Masking:
Low complexity regions: ☑ DUST
Mask lower case letters: ☐

Scoring Parameters:
Match/Mismatch Scores: (BLASTN only) (1,-3)
Matrix:(except BLASTN) BLOSUM62
Gap Costs (Open/Extend) (5,2)
Ungapped Align ☐
Compositional adjustments (blastp, blastx, tblastn) Conditional composition-based score adjustment (2)
Genetic Code (blastx, tblastx) Standard (1)

Select BLAST program and database

Parameters for BLAST search

The information page with an URL to the result appears after submitting sequence(s) to search.

Thanks!

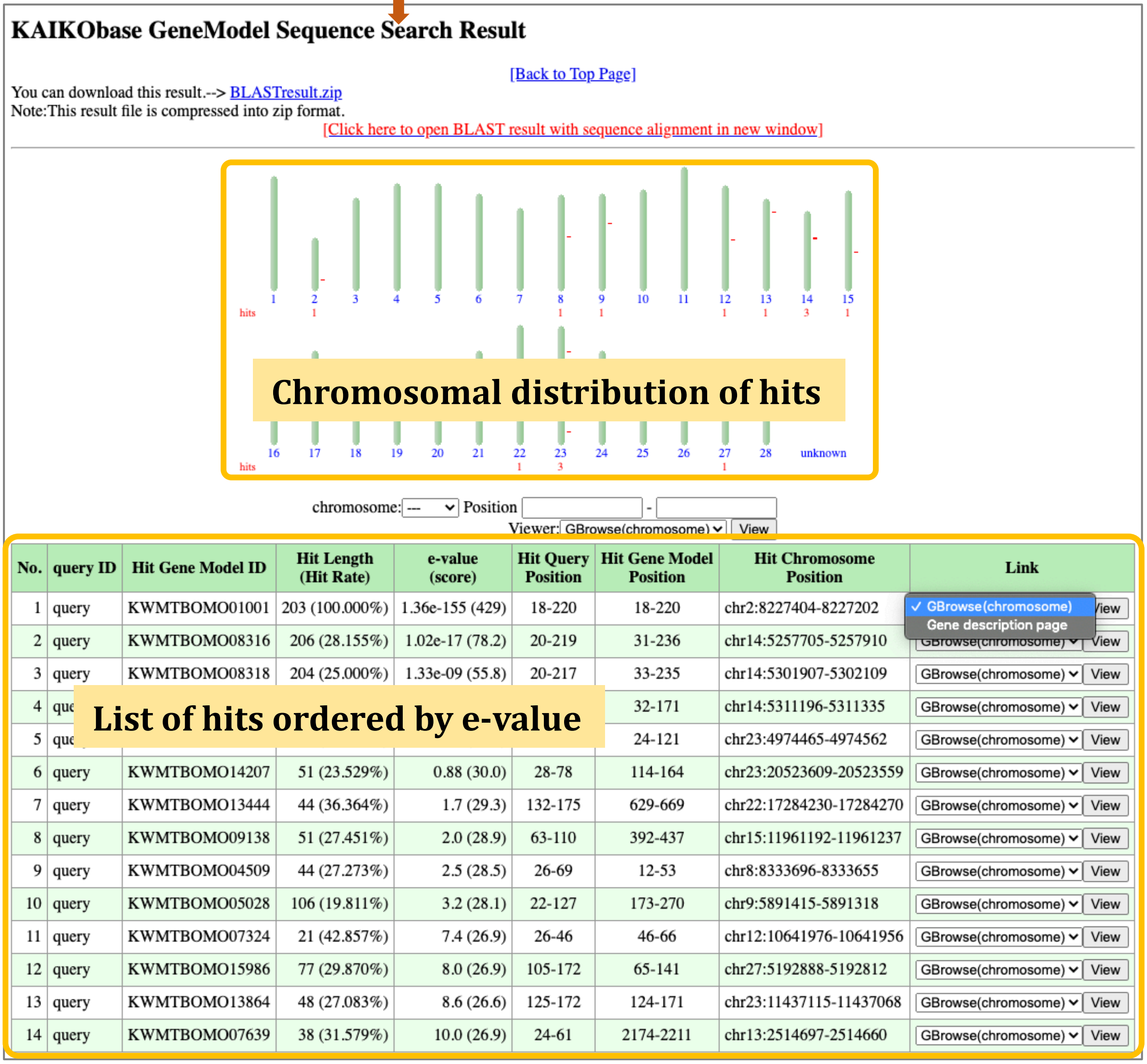
Now your search request is accepted.

```
Program Blastp
DataLib KWMTBOMO.prot
Expect  10
WordSize        6
Filter  2
Descriptions
Alignments

Begin
>query
MLARCLAVAA VAVLASAGPP SPIYRPCYLD DYKCISDHLA AISKCIPGRG
QIPSQYEIPV FQFEIPYFNA TYVDHNLITR NHDKCRVSEF YDNVRTLKTV
LTVDCPWLNF ESNRTLAQHM SFKEDVVLSF YINGSYPLIR LTTVFDKGNN
FDLCSAFTFA DLAGGLPIFH INPNDQRTAQ WLSKDLTLLH IYEREHIFGK
RNWLARSFIS RTLCDFGCHH

//
```

Please check following URL.
/KAIKObase/blast_search.php?sid=pate4sry303w uqirl4d67pi3ig1lt7&dtype=GeneModel&program=blastp



KAIKObase GeneModel Sequence Search Result

[Back to Top Page]

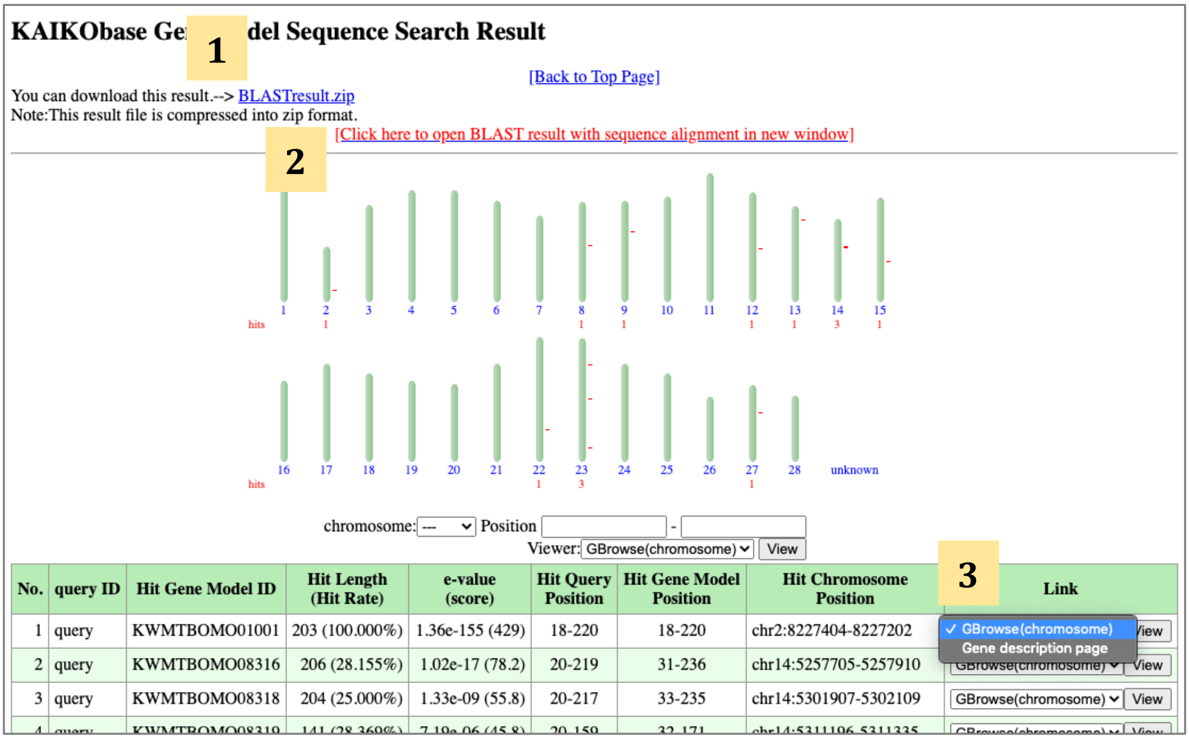
You can download this result.--> BLASTresult.zip
Note:This result file is compressed into zip format.
[Click here to open BLAST result with sequence alignment in new window]

Chromosomal distribution of hits

chromosome: --- Position - Viewer: GBrowse(chromosome) View

List of hits ordered by e-value

No.	query ID	Hit Gene Model ID	Hit Length (Hit Rate)	e-value (score)	Hit Query Position	Hit Gene Model Position	Hit Chromosome Position	Link
1	query	KWMTBOMO01001	203 (100.000%)	1.36e-155 (429)	18-220	18-220	chr2:8227404-8227202	✓ GBrowse(chromosome) Gene description page View
2	query	KWMTBOMO08316	206 (28.155%)	1.02e-17 (78.2)	20-219	31-236	chr14:5257705-5257910	GBrowse(chromosome) View
3	query	KWMTBOMO08318	204 (25.000%)	1.33e-09 (55.8)	20-217	33-235	chr14:5301907-5302109	GBrowse(chromosome) View
4	que					32-171	chr14:5311196-5311335	GBrowse(chromosome) View
5	que					24-121	chr23:4974465-4974562	GBrowse(chromosome) View
6	query	KWMTBOMO14207	51 (23.529%)	0.88 (30.0)	28-78	114-164	chr23:20523609-20523559	GBrowse(chromosome) View
7	query	KWMTBOMO13444	44 (36.364%)	1.7 (29.3)	132-175	629-669	chr22:17284230-17284270	GBrowse(chromosome) View
8	query	KWMTBOMO09138	51 (27.451%)	2.0 (28.9)	63-110	392-437	chr15:11961192-11961237	GBrowse(chromosome) View
9	query	KWMTBOMO04509	44 (27.273%)	2.5 (28.5)	26-69	12-53	chr8:8333696-8333655	GBrowse(chromosome) View
10	query	KWMTBOMO05028	106 (19.811%)	3.2 (28.1)	22-127	173-270	chr9:5891415-5891318	GBrowse(chromosome) View
11	query	KWMTBOMO07324	21 (42.857%)	7.4 (26.9)	26-46	46-66	chr12:10641976-10641956	GBrowse(chromosome) View
12	query	KWMTBOMO15986	77 (29.870%)	8.0 (26.9)	105-172	65-141	chr27:5192888-5192812	GBrowse(chromosome) View
13	query	KWMTBOMO13864	48 (27.083%)	8.6 (26.6)	125-172	124-171	chr23:11437115-11437068	GBrowse(chromosome) View
14	query	KWMTBOMO07639	38 (31.579%)	10.0 (26.9)	24-61	2174-2211	chr13:2514697-2514660	GBrowse(chromosome) View



KAIKObase Ge del Sequence Search Result

[Back to Top Page]

You can download this result.--> BLASTresult.zip
Note:This result file is compressed into zip format.

[Click here to open BLAST result with sequence alignment in new window]

chromosome: --- Position -
Viewer: GBrowse(chromosome) View

No.	query ID	Hit Gene Model ID	Hit Length (Hit Rate)	e-value (score)	Hit Query Position	Hit Gene Model Position	Hit Chromosome Position	Link
1	query	KWMTBOMO01001	203 (100.000%)	1.36e-155 (429)	18-220	18-220	chr2:8227404-8227202	GBrowse(chromosome) View
2	query	KWMTBOMO08316	206 (28.155%)	1.02e-17 (78.2)	20-219	31-236	chr14:5257705-5257910	GBrowse(chromosome) View
3	query	KWMTBOMO08318	204 (25.000%)	1.33e-09 (55.8)	20-217	33-235	chr14:5301907-5302109	GBrowse(chromosome) View

The Search Result page provides several links:

1 Results are summarized to the file "**BLASTresult.zip**" for download, including following two files:

- result_rawdata.txt: BLAST result in format 0
- restult.txt: tab-delimited text file of the table below the Search Result page

2 Link to the **graphic summary and alignment of BLAST result**

The alignment page further provides BLASTresult_rawdata.zip (containing result_rawdata.txt) for download and links to genome browser for each hit.

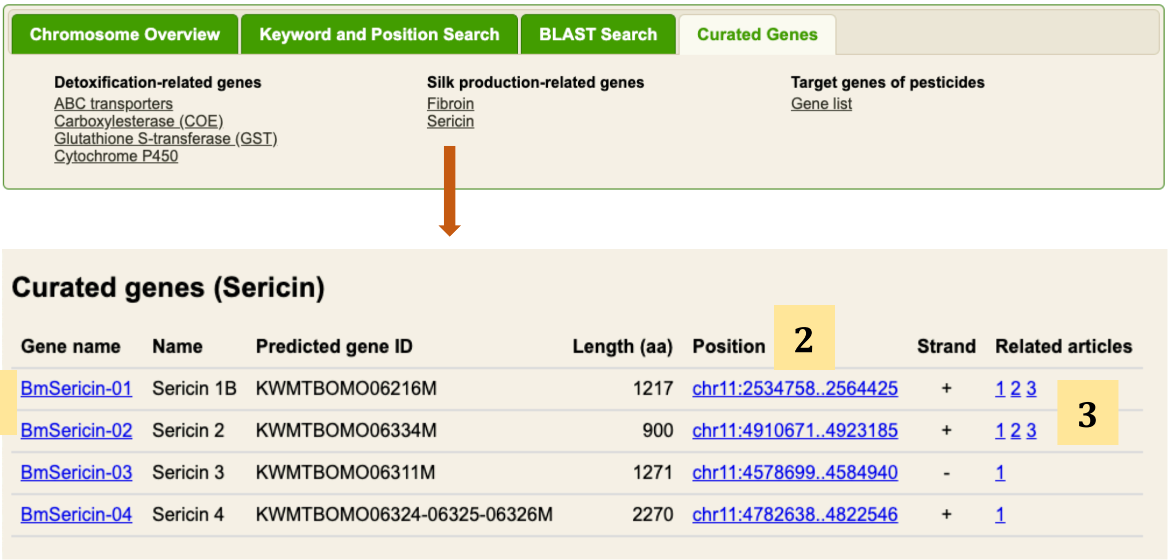
3 Links to **genome browser** (for all data sets) and **gene description page** (only for New gene model)

2.4 Curated Genes

“Curated Genes” provides lists of manually curated genes for users.

Currently three types of genes were curated:
1. Detoxification-related genes
2. Silk production-related genes
3. Target genes of pesticides

Click the link to check the list of genes.



Curated genes (Sericin)

Gene name	Name	Predicted gene ID	Length (aa)	Position	Strand	Related articles
BmSericin-01	Sericin 1B	KWMTBOMO06216M	1217	chr11:2534758..2564425	+	1 2 3
BmSericin-02	Sericin 2	KWMTBOMO06334M	900	chr11:4910671..4923185	+	1 2 3
BmSericin-03	Sericin 3	KWMTBOMO06311M	1271	chr11:4578699..4584940	-	1
BmSericin-04	Sericin 4	KWMTBOMO06324-06325-06326M	2270	chr11:4782638..4822546	+	1

“Predicted gene ID” shows the gene ID of New gene model in the same chromosomal position. ID which ends with an “M” shows that the intron-exon structure of the curated gene is different from that of the New gene model.

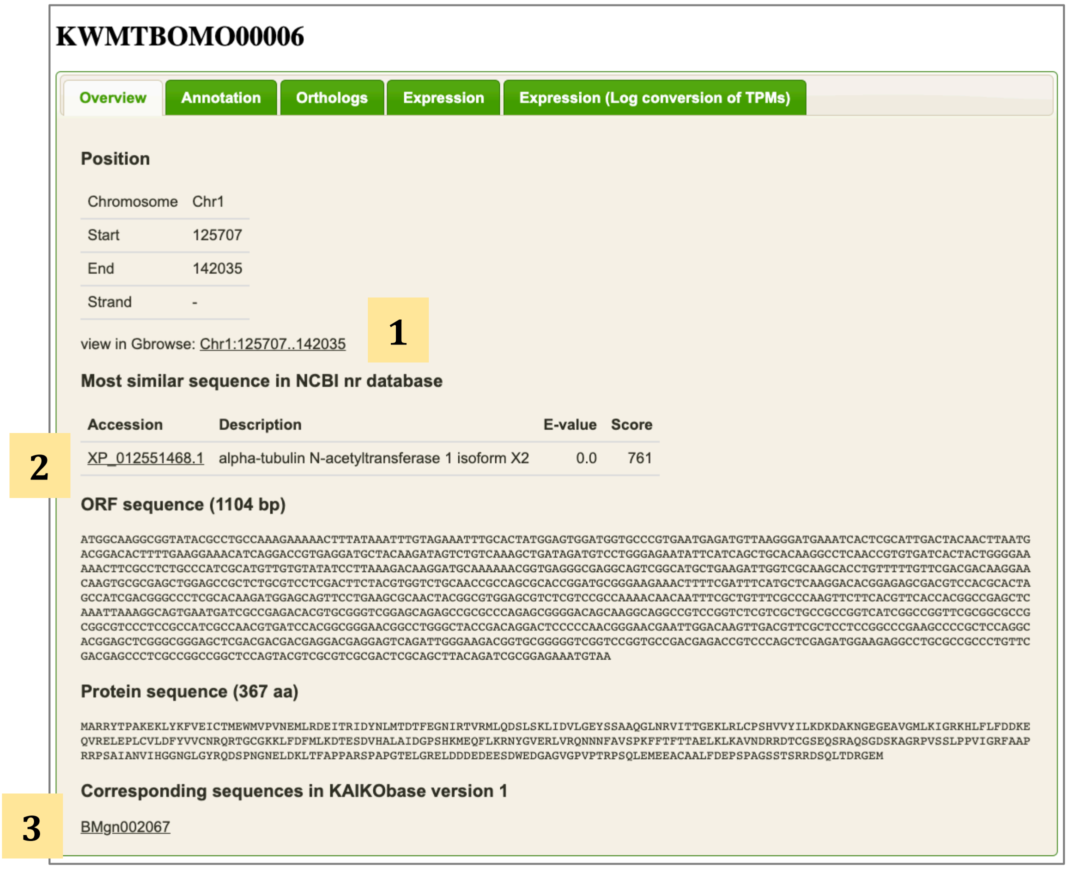
The page of gene list provides links to:

1 Link to **description page**
2 Link to **genome browser**
3 Links to **related articles**

3 Description page

“Description page” shows genetic information of each gene, including chromosomal position, nucleotide and amino acid sequences, most similar gene in nr database, corresponding gene(s) in old KAIKObase (if any), functional domains and motifs, orthologs, and expression data in several tissues.

Here ”KWMTBOMO00006” is used as an example. The description page shows “Overview” by default.



KWMTBOMO00006

Position

Chromosome	Chr1
Start	125707
End	142035
Strand	-

view in Gbrowse: Chr1:125707..142035

Most similar sequence in NCBI nr database

Accession	Description	E-value	Score
XP_012551468.1	alpha-tubulin N-acetyltransferase 1 isoform X2	0.0	761

ORF sequence (1104 bp)

```
ATGGCAAGGCGGTATACGCCTGCCAAAGAAAAACTTTATAAATTTGTAGAAATTTGCACTATGGAGTGGATGGTGCCCGTGAATGAGATGTTAAGGGATGAAATCACTCGCATTGACTACAACTTAATG
ACGGACACTTTTGAAGGAAACATCAGGACCGTGAGGATGCTACAAGATAGTCTGTCAAAGCTGATAGATGTCCTGGGAGAATATTCATCAGCTGCACAAGGCCTCAACCGTGTGATCACTACTGGGGAA
AAACTTCGCCTCTGCCCATCGCATGTTGTGTATATCCTTAAAGACAAGGATGCAAAAAACGGTGAGGGCGAGGCAGTCGGCATGCTGAAGATTGGTCGCAAGCACCTGTTTTTGTTCGACGACAAGGAA
CAAGTGCGCGAGCTGGAGCCGCTCTGCGTCCTCGACTTCTACGTGGTCTGCAACCGCCAGCGCACCGGATGCGGGAAGAAACTTTTCGATTTCATGCTCAAGGACACGGAGAGCGACGTCCACGCACTA
GCCATCGACGGGCCCTCGCACAAGATGGAGCAGTTCCTGAAGCGCAACTACGGCGTGGAGCGTCTCGTCCGCCAAAACAACAATTTCGCTGTTTCGCCCAAGTTCTTCACGTTCACCACGGCCGAGCTC
AAATTAAAGGCAGTGAATGATCGCCGAGACACGTGCGGGTCGGAGCAGAGCCGCGCCCAGAGCGGGGACAGCAAGGCAGGCCGTCCGGTCTCGTCGCTGCCGCCGGTCATCGGCCGGTTCGCGGCGCCG
CGGCGTCCCTCCGCCATCGCCAACGTGATCCACGGCGGGAACGGCCTGGGCTACCGACAGGACTCCCCCAACGGGAACGAATTGGACAAGTTGACGTTCGCTCCTCCGGCCCGAAGCCCCGCTCCAGGC
ACGGAGCTCGGGCGGGAGCTCGACGACGACGAGGACGAGGAGTCAGATTGGGAAGACGGTGCGGGGGTCGGTCCGGTGCCGACGAGACCGTCCCAGCTCGAGATGGAAGAGGCCTGCGCCGCCCTGTTC
GACGAGCCCTCGCCGGCCGGCTCCAGTACGTCGCGTCGCGACTCGCAGCTTACAGATCGCGGAGAAATGTAA
```

Protein sequence (367 aa)

```
MARRYTPAKEKLYKFVEICTMEWMVPVNEMLRDEITRIDYNLMTDTFEGNIRTVRMLQDSLSKLIDVLGEYSSAAQGLNRVITTGEKLRLCPSHVVYILKDKDAKNGEGEAVGMLKIGRKHLFLFDDKE
QVRELEPLCVLDFYVVCNRQRTGCGKKLFDFMLKDTESDVHALAIDGPSHKMEQFLKRNYGVERLVRQNNNFAVSPKFFTFTTAELKLKAVNDRRDTCGSEQSRAQSGDSKAGRPVSSLPPVIGRFAAP
RRPSAIANVIHGGNGLGYRQDSPNGNELDKLTFAPPARSPAPGTELGRELDDDEDEESDWEDGAGVGPVPTRPSQLEMEEACAALFDEPSPAGSSTSRRDSQLTDRGEM
```

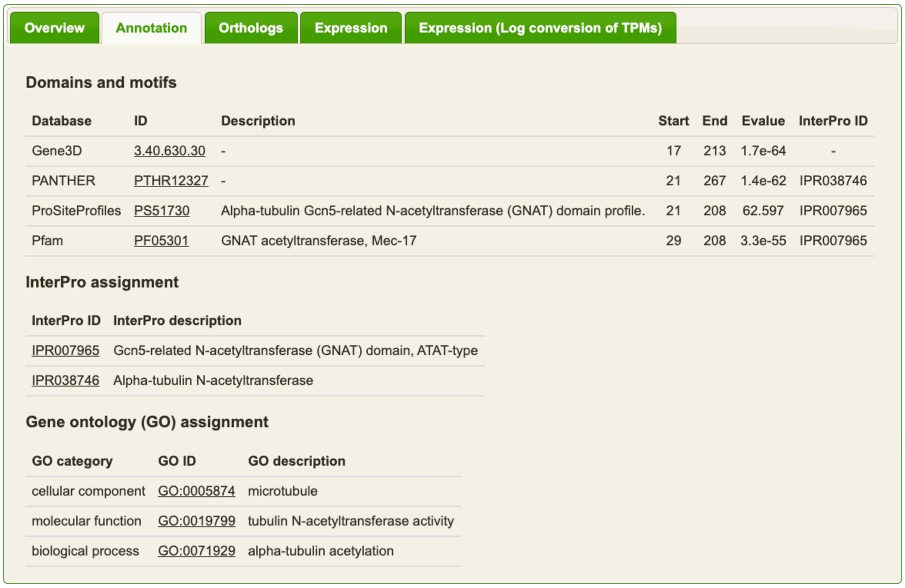
Corresponding sequences in KAIKObase version 1

BMgn002067

The page of gene list provides links to:

1 Link to **genome browser**
2 Link to **NCBI**
3 Link(s) to **description page of old KAIKObase**

“Annotation” shows the functional domains and motifs. Click the URL of each ID to check its detailed information.



Domains and motifs

Database	ID	Description	Start	End	Evalue	InterPro ID
Gene3D	3.40.630.30	-	17	213	1.7e-64	-
PANTHER	PTHR12327	-	21	267	1.4e-62	IPR038746
ProSiteProfiles	PS51730	Alpha-tubulin Gcn5-related N-acetyltransferase (GNAT) domain profile.	21	208	62.597	IPR007965
Pfam	PF05301	GNAT acetyltransferase, Mec-17	29	208	3.3e-55	IPR007965

InterPro assignment

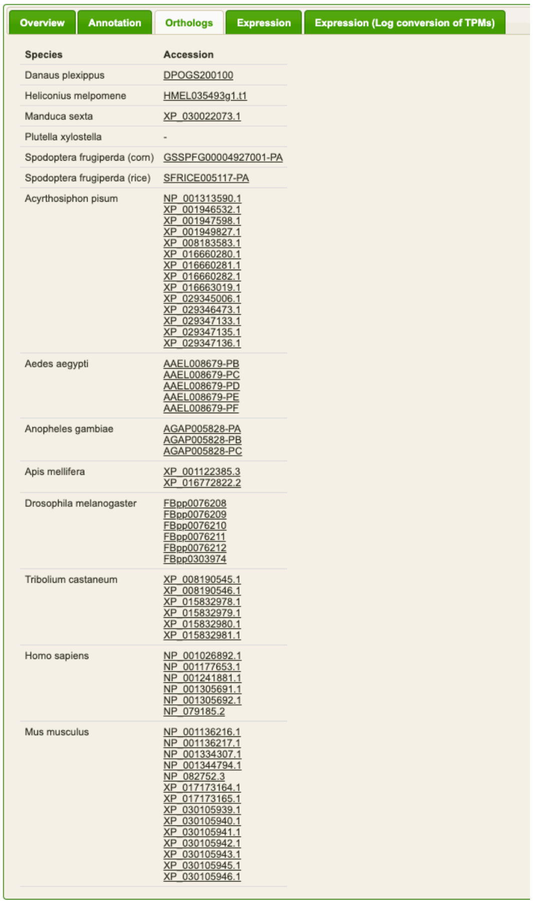
InterPro ID	InterPro description
IPR007965	Gcn5-related N-acetyltransferase (GNAT) domain, ATAT-type
IPR038746	Alpha-tubulin N-acetyltransferase

Gene ontology (GO) assignment

GO category	GO ID	GO description
cellular component	GO:0005874	microtubule
molecular function	GO:0019799	tubulin N-acetyltransferase activity
biological process	GO:0071929	alpha-tubulin acetylation

“Orthologs” provides a list of the orthologs in 5 lepidopterans, 6 model insects, and human and mouse.

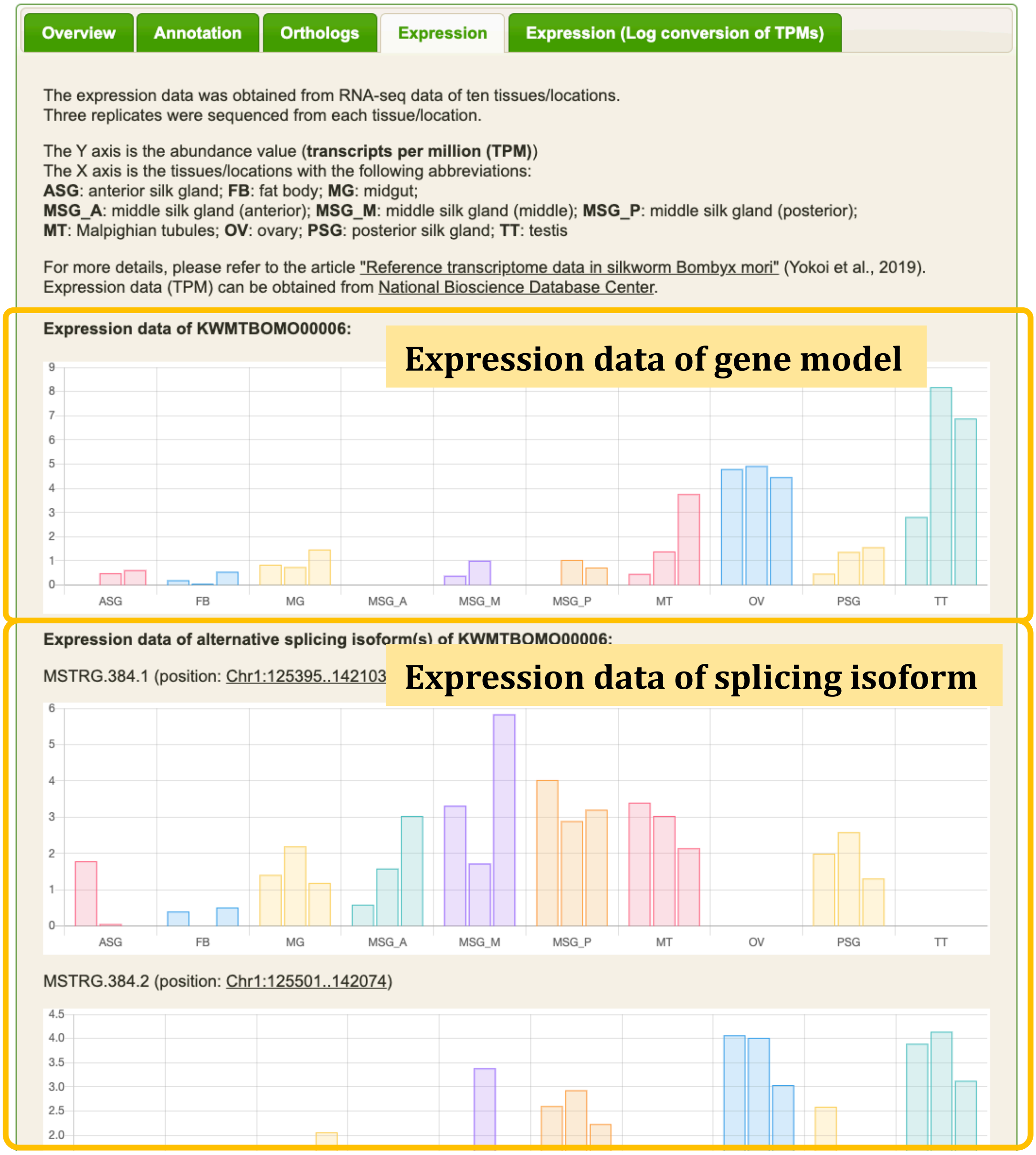
Click the URL of each accession to check its detailed information.



Species	Accession
Danaus plexippus	DPOGS200100
Heliconius melpomene	HMEL035493g1.t1
Manduca sexta	XP_030022073.1
Plutella xylostella	-
Spodoptera frugiperda (corn)	GSSPFG00004927001-PA
Spodoptera frugiperda (rice)	SFRICE005117-PA
Acyrthosiphon pisum	NP_001313590.1 XP_001946532.1 XP_001947598.1 XP_001949827.1 XP_008183583.1 XP_016660280.1 XP_016660281.1 XP_016660282.1 XP_016663019.1 XP_029345006.1 XP_029346473.1 XP_029347133.1 XP_029347135.1 XP_029347136.1
Aedes aegypti	AAEL008679-PB AAEL008679-PC AAEL008679-PD AAEL008679-PE AAEL008679-PF
Anopheles gambiae	AGAP005828-PA AGAP005828-PB AGAP005828-PC
Apis mellifera	XP_001122385.3 XP_016772822.2
Drosophila melanogaster	FBpp0076208 FBpp0076209 FBpp0076210 FBpp0076211 FBpp0076212 FBpp0303974
Tribolium castaneum	XP_008190545.1 XP_008190546.1 XP_015832978.1 XP_015832979.1 XP_015832980.1 XP_015832981.1
Homo sapiens	NP_001026892.1 NP_001177653.1 NP_001241881.1 NP_001305691.1 NP_001305692.1 NP_079185.2
Mus musculus	NP_001136216.1 NP_001136217.1 NP_001334307.1 NP_001344794.1 NP_082752.3 XP_017173164.1 XP_017173165.1 XP_030105939.1 XP_030105940.1 XP_030105941.1 XP_030105942.1 XP_030105943.1 XP_030105945.1 XP_030105946.1

"Expression" shows the expression data from 10 different tissues/locations. Below the expression data of the gene model shows the expression data of the alternative splicing isoforms (if any).

Click the Link of each alternative splicing isoform to check its position on chromosome in genome browser.



The expression data was obtained from RNA-seq data of ten tissues/locations.
Three replicates were sequenced from each tissue/location.

The Y axis is the abundance value (**transcripts per million (TPM)**)
The X axis is the tissues/locations with the following abbreviations:
ASG: anterior silk gland; **FB**: fat body; **MG**: midgut;
MSG_A: middle silk gland (anterior); **MSG_M**: middle silk gland (middle); **MSG_P**: middle silk gland (posterior);
MT: Malpighian tubules; **OV**: ovary; **PSG**: posterior silk gland; **TT**: testis

For more details, please refer to the article "Reference transcriptome data in silkworm Bombyx mori" (Yokoi et al., 2019).
Expression data (TPM) can be obtained from National Bioscience Database Center.

Expression data of KWMTBOMO00006:

Expression data of gene model

Expression data of alternative splicing isoform(s) of KWMTBOMO00006:

MSTRG.384.1 (position: Chr1:125395..142103

Expression data of splicing isoform

MSTRG.384.2 (position: Chr1:125501..142074)

"Expression (log conversion of TPMs)" shows the same expression data as "Expression" page, but the TPM values in Y axis are shown with $\log_{10}$ conversion for better comparison between transcripts.

TPM value with an offset of 1 was used for $\log_{10}$ conversion to avoid negative values.

The expression data was obtained from RNA-seq data of ten tissues/locations.
Three replicates were sequenced from each tissue/location.

The Y axis is the abundance value with $\log_{10}$ conversion (value of **transcripts per million (TPM) plus 1** is used to avoid negative output)
The X axis is the tissues/locations with the following abbreviations:
ASG: anterior silk gland; **FB**: fat body; **MG**: midgut;
MSG_A: middle silk gland (anterior); **MSG_M**: middle silk gland (middle); **MSG_P**: middle silk gland (posterior);
MT: Malpighian tubules; **OV**: ovary; **PSG**: posterior silk gland; **TT**: testis

For more details, please refer to the article "Reference transcriptome data in silkworm Bombyx mori" (Yokoi et al., 2019).
Expression data (TPM) can be obtained from National Bioscience Database Center.

Expression data of KWMTBOMO00006:

Expression data of alternative splicing isofo

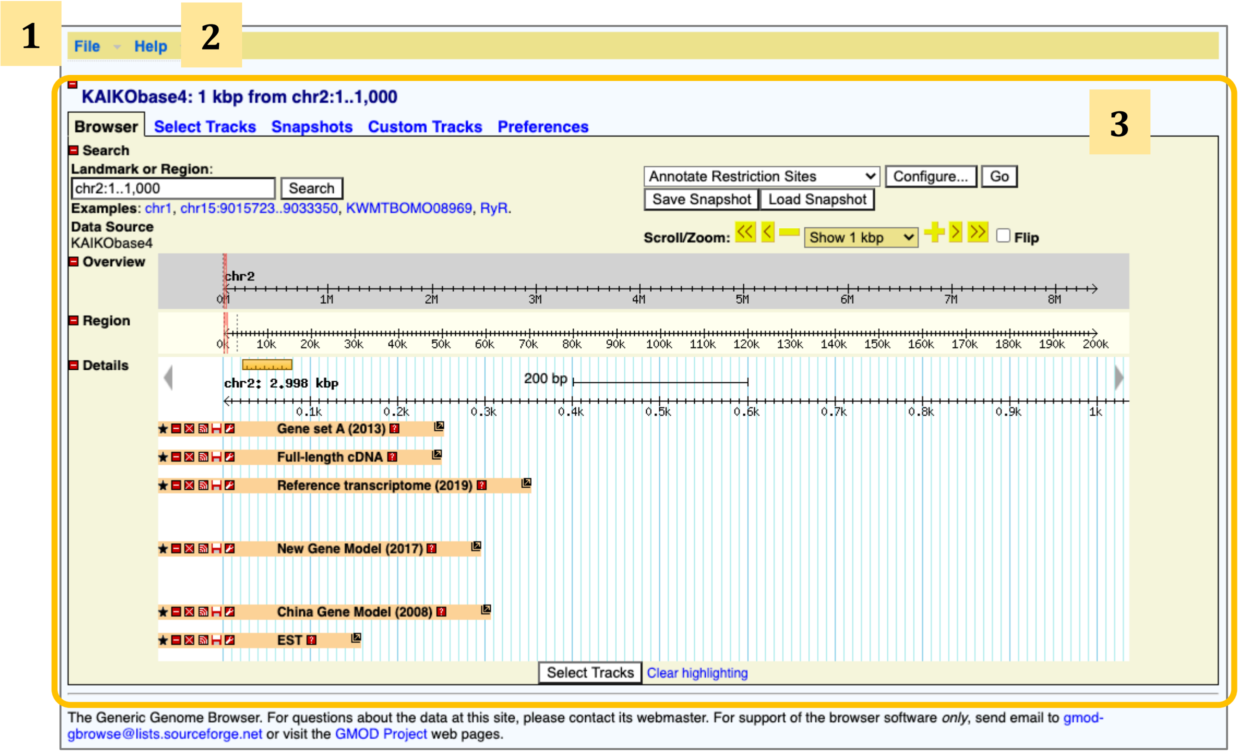
MSTRG.384.1 (position: Chr1:125395..142103

MSTRG.384.2 (position: Chr1:125501..142074)

4 Genome browser

The genome browser (GBrowse) is accessible from the top page. It can also be accessed through links provided in description pages, BLAST results, keyword search results, etc.

The genome browser contains three parts: **1. File**; **2. Help**; and **3. browser and settings**.



KAIKObase4: 1 kbp from chr2:1..1,000

Landmark or Region: chr2:1..1,000

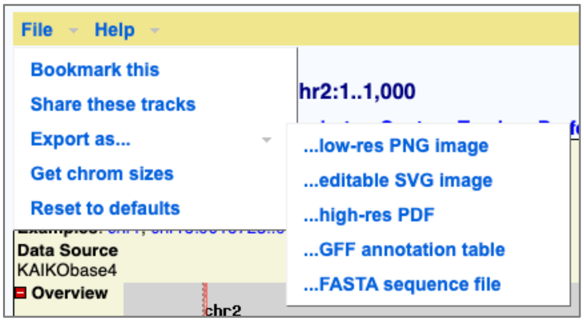
Examples: chr1, chr15:9015723..9033350, KWMTBOMO08969, RyR.

Data Source: KAIKObase4

chr2: 2.998 kbp

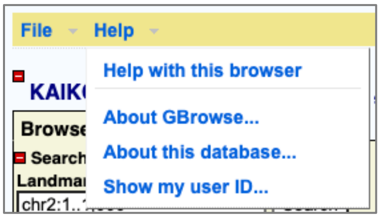
The Generic Genome Browser. For questions about the data at this site, please contact its webmaster. For support of the browser software *only*, send email to gmod-gbrowse@lists.sourceforge.net or visit the GMOD Project web pages.

1. File: A pull-down menu shows up when the pointer gets close. Information within the region shown below can be exported to various file format under the “Export as...” option



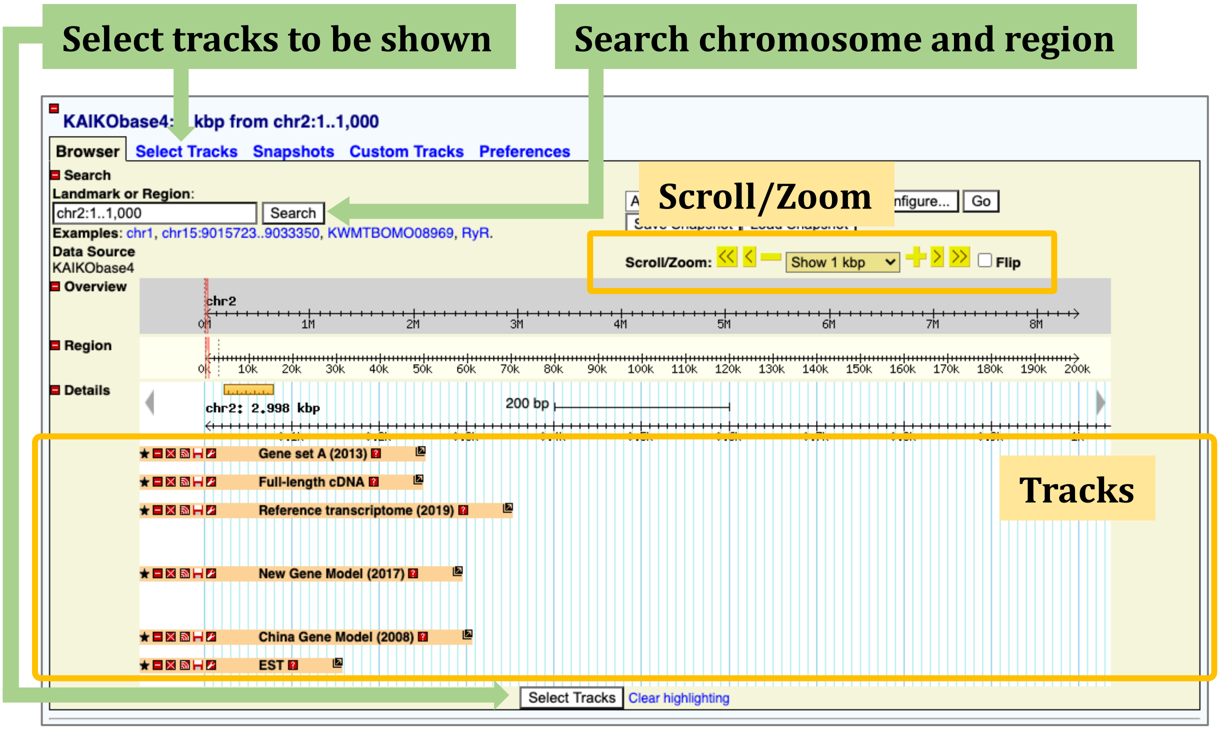
---

2. Help: A pull-down menu shows up when the pointer gets close. "Help with this browser" will show tutorial about how to use Gbrowse.



The tutorial is also accessible through the following URL:
http://kaikobase.dna.affrc.go.jp/gbrowse2/general_help.html

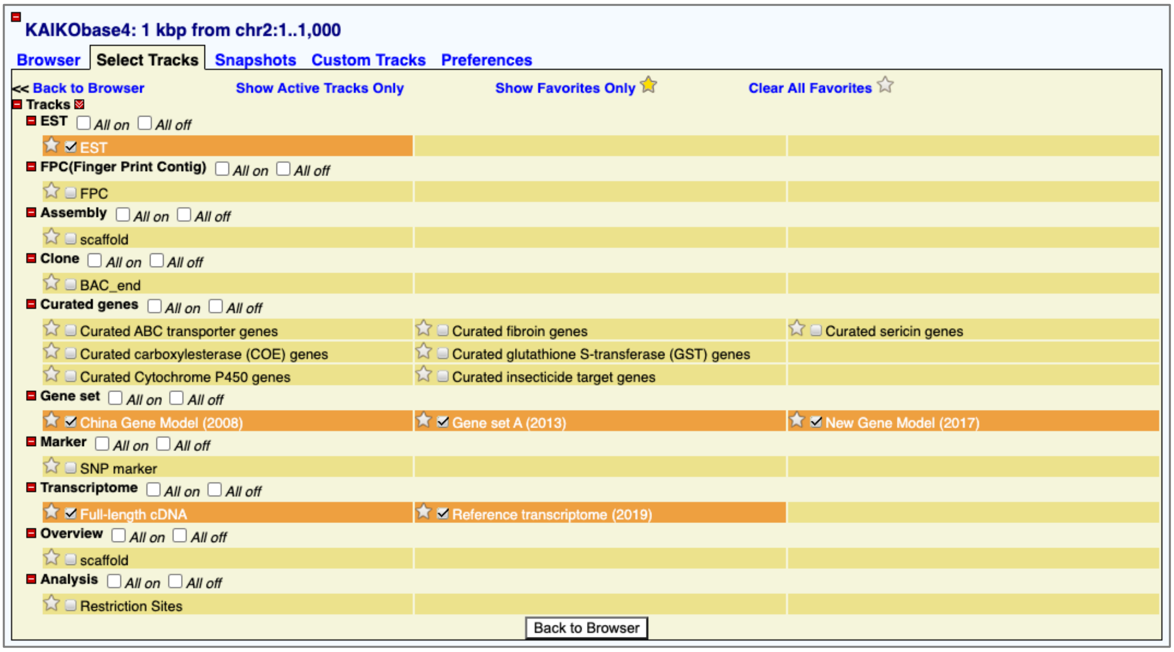
3. browser and settings: "Tracks" shows different gene sets within selected chromosomal region. Default gene sets are as shown here.



"Scroll/Zoom" bottoms allow users to check the up- and down-stream regions of of current region. The pull-down menu provides options from "Show 100 bp" to "Show 1 Mbp" for the users to select the optimal length of the region for viewing the genes.

”Select Tracks” shows the available tracks. Besides the default gene sets, the following data sets are also visible:

- **Genetic markers** (finger print contigs, BAC-end sequences, SNP markers)
- **Assembly (scaffolds of old genome assembly)**
- **Curated genes**
- **Restriction sites**


KAIKObase4: 1 kbp from chr2:1..1,000

Tracks
EST ☐ All on ☐ All off
☆ ☑ EST
FPC(Finger Print Contig) ☐ All on ☐ All off
☆ ☐ FPC
Assembly ☐ All on ☐ All off
☆ ☐ scaffold
Clone ☐ All on ☐ All off
☆ ☐ BAC_end
Curated genes ☐ All on ☐ All off
☆ ☐ Curated ABC transporter genes ☆ ☐ Curated fibroin genes ☆ ☐ Curated sericin genes
☆ ☐ Curated carboxylesterase (COE) genes ☆ ☐ Curated glutathione S-transferase (GST) genes
☆ ☐ Curated Cytochrome P450 genes ☆ ☐ Curated insecticide target genes
Gene set ☐ All on ☐ All off
☆ ☑ China Gene Model (2008) ☆ ☑ Gene set A (2013) ☆ ☑ New Gene Model (2017)
Marker ☐ All on ☐ All off
☆ ☐ SNP marker
Transcriptome ☐ All on ☐ All off
☆ ☑ Full-length cDNA ☆ ☑ Reference transcriptome (2019)
Overview ☐ All on ☐ All off
☆ ☐ scaffold
Analysis ☐ All on ☐ All off
☆ ☐ Restriction Sites

5 Contact

If you have any question or find any problem about KAIKObase, please contact us through the following email:

SGPwebmaster@ml.affrc.go.jp