

# **UT Genome Browser**

## **User's Manual**

### **Version 1.0**

**August 17, 2004**

**UT Genome Browser Development Team**

# Table of Contents

- 1 Welcome
- 2 Keyword Search
  - 2.1 Gene List
  - 2.2 Mapping Position List
- 3 Online Sequence Alignment
- 4 Main Window
  - 4.1 Scrolling and Zooming the Main Window
  - 4.2 Functions of Buttons
  - 4.3 Addition, Deletion, and Rearrangement of Tracks
  - 4.4 Setting for Tracks
- 5 Pre-defined Tracks
  - 5.1 Overview Tracks
  - 5.2 Ruler and Zooming Tracks
  - 5.3 Base Color Track
  - 5.4 Gap Track
  - 5.5 GC Content Track
  - 5.6 Mapped Gene Track
  - 5.7 Genscan Track
  - 5.8 Comparative Genomics Track

**Ramen Assembler / UT Genome Browser Team Members**

**Acknowledgements**

# 1. Welcome

Welcome to UT Genome Browser (University of Tokyo Genome Browser). With this browser you can browse genomic information of a target species (at present Medaka only) from several aspects. For example, the following tasks can be performed with different combinations.

- ✓ You can search gene, by gene name, clone name, cluster name and display its surroundings gene groups.
- ✓ You can map a genomic sequence to any genome and observe its surroundings.
- ✓ You can save the sequence itself on your system from here.

Questions, suggestions or comments are welcome to [webmaster@utgenome.org](mailto:webmaster@utgenome.org).

## 2. Keyword Search



Keyword Search	
Keyword	MF01SSB031K0 <input type="button" value="Search"/>
Species	medaka
Revision	200406
Keykinds	<input checked="" type="checkbox"/> Clone ID <input checked="" type="checkbox"/> Ensemble ID <input checked="" type="checkbox"/> GI Number <input checked="" type="checkbox"/> Genbank Acc <input checked="" type="checkbox"/> LocusLink ID <input checked="" type="checkbox"/> Symbol <input checked="" type="checkbox"/> Unigene Cluster ID

The top window displays the keyword search engine, as above picture illustrates. Inputting any keyword at the query box, we can see the detail information about the query. More about keyword search can be found in the help, written on the same window. In addition, if the keyword is a part of targets, then all the targets will be displayed. For example, the keyword “BCD” will match to “ABCD”, “BCDE”, “BCD” and “BCDF” etc. (this is worked out with using a suffix array.)

## 2.1. Gene List

UT Genome Browser (Medaka) Search

14 hits.

keyword: MF01SSB031K0 Search

(back) (clear)

	Keyword	Length	Match Cnt	Target	Match Range	Match Ratio	Cover Ratio	
	MF01SSB031K09	665	1	scaffold6774	23561 - 26434	0.97	1	MF01SSB031K09
(*)	MF01SSB031K09	617	1	scaffold6774	23714 - 27077	0.985	1	MF01SSB031K09
					70116			

When there are many genes for one keyword search, all genes will be displayed on a list. Pressing the (\*) button on the left side, you can go to the mapped region for the gene. This list appears when multiple candidate entries are found, typically when class names or part of a gene number are input as a keyword. If a Genebank accession number is given, the number must be unique, and hence this list does not appear.

## 2.2. Mapping Position List

UT Genome Browser (Medaka) Search

gene [BJ540397] is mapped to 3 locations

keyword: MF01SSB031K0 Search

(back) (top) (clear)

	Target	Range	Cover ratio	Match ratio
(*)	scaffold8621	15616 - 15994	0.58	0.99
(*)	scaffold15029	13229 - 13606	0.58	1.00
(*)	scaffold1161	35775 - 36187	0.50	0.87

If the query gene has mapped to several regions on the genomic sequence, the list of all positions will be shown. Pressing the (\*) button on the left side displays the mapped region of the gene. If there is only one mapped region of the gene, the list is skipped to display.

### 3. Sequence Alignment



Sequence mapping is possible by inputting a sequence in the online mapping frame (above image) at the bottom of the main window. After selecting the species and its revision, press the search button on the bottom. The request is sent to the alignment server ALPS to align the given sequence to the selected genome. Afterwards, the list of alignment results is returned if the system finds some alignments.

```

Sequence
tccctgaatccaagatgggctgagacatcagacacaacaaggaccgcaaggtgcacaggaaggagccaagagccaggat
atctcaactgaggctcctgggtaagtgttacaggttcttgtctcggcgtaccaatgctcctttcaacaaggtggtctcgaa
ggggtgttcatgagcaggccaacaggcctcctatcgaatctcccgctctgatctgtaaaatgaaacttccggccgtg
ataataaaatgctgtgtgtcgtgggaacagtcaactgatgatgtcaggatcaggagatccccaaactcaagatctgcgct
ctgaaagttaactgatggtgcacgtcgcaggatcctgaaatcagggggacaagtgatgacctttgaccagctggctttggc
cactccccaaaggacatggcacagtgctgctgtcaggaccgccaaagccagagaggtttacaggcactttggaaggcca
ctggaacccctcacagccac
    
```

#### Alignment Result Summary

scaffold429: target position=63382-65343 query position=1-500 match ratio=96.8% cover ratio=100%

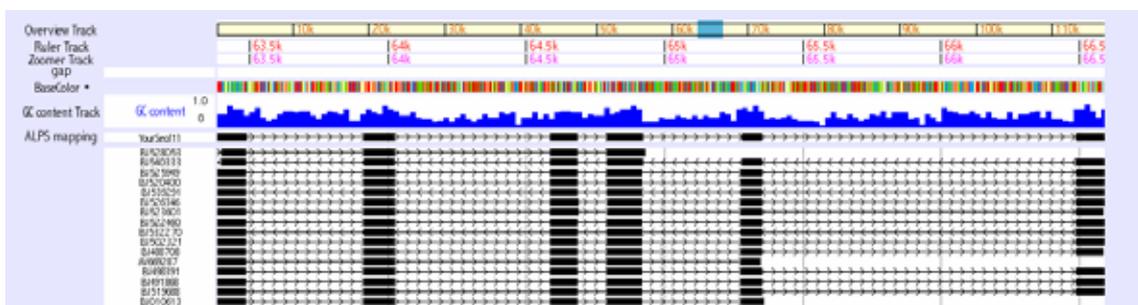
#### Alignment Result [ 1 ] [view 63382-65343@scaffold429](#)

```
[+]
target=scaffold429
```

type	target position	query position	length	direction	match ratio	coverage	intron motif
exon	63382 - 63485	1 - 104	104	+	89.42% (93/104)	20.8%	
TTGTTTTTTTCTGTAGGGCGTAGACATCAGACACAACAAGGACCCGCAAG 063820061							
XXX XXX X XX XX XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX							
TCCCTGAATCCAAGATGGGCGTAGACATCAGACACAACAAGGACCCGCAAG 000000050							
GTGCACAGGAAGGAGCCAAAGAGCCAGGATATCTACCTGAGGCTCTTGCT 063820111							
XX							
GTGCACAGGAAGGAGCCAAAGAGCCAGGATATCTACCTGAGGCTCTTGCT 000000100							
TAAG 063820115							
TAAG 000000104							
intron	63486 - 63913		(428)	+			full
exon	63914 - 64021	105 - 212	108	+	100% (108/108)	21.6%	
intron	64022 - 64586		(565)	+			full
exon	64587 - 64683	213 - 309	97	+	97.94% (95/97)	19.4%	
intron	64684 - 64791		(108)	+			none
exon	64792 - 64917	310 - 435	126	+	97.62% (123/126)	25.2%	
intron	64918 - 65278		(361)	+			full
exon	65279 - 65343	436 - 500	65	+	100% (65/65)	13%	
total	63382 - 65343	1 - 500	500	+	96.8% (484/500)	100%	



From the list, you can select one alignment by clicking the result button as illustrated above. Then, the aligned result will appear on the browser. The ALPS mapping track in following snapshot displays the added mapping status.



## 4. Main Window

The above picture shows the main screen for UT Genome Browser. It is possible to see the targeted region of genomic sequence by many tracks. This browser provides the following functions to manipulate tracks on this image screen.

- ✓ Shifting, zooming in and out, and reversing of the target region. It facilitates zooming in the sequence until bases can be seen, while providing a panoramic, structural view for the sequence.
- ✓ Adding, deleting and reordering the tracks.
- ✓ Customizing of each track's display.
- ✓ Retrieving the present displayed sequence in Fasta format.
- ✓ Continuing keyword search.

In what follows, we describe the detail about each function.

## 4.1 Scrolling and zooming the main window

On main window, it is possible to shift and to zoom in/out the present observing sequence. These tasks can be carried out by:

- ✓ Inputting value in the input box method.
- ✓ Using shifting button method.
- ✓ Using overview, ruler and zooming track method.

### Inputting value in the input box method

species:  revision:  target:  start:  end:  width:

(scaffold123 length: 155085bp) (193.86 bp/pixel)

You can input values into the input boxes (like above image) inside the main window. The details about this method are as follows,

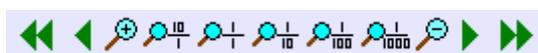
- ✓ **Species:** Select the species to display. The selectable revisions would be changed with respect to the species. At present, only Medaka is possible to be selected.
- ✓ **Revision:** Select the revision of the target specie's genome. For Medaka, the revision name format is yyyyymm. The selecting topics will change with respect to the selected revision.
- ✓ **Target:** Select the target region to be displayed, from the selected species and selected revision. In the case of 200406 revision of Medaka, only scaffold name (e.g. scaffold123) can be selected but not the chromosome name (e.g. chr10). Chromosome numbers will be available in our later revision.
- ✓ **Start:** Select the start position from your targeted sequence to display. The number should be 1-origin (i.e. the starting point will be 1).
- ✓ **End:** Select the end point from your targeted sequence to display. The number should be 1-origin. If the ending point is greater than the starting point then the

display will be normal. Otherwise, the display will be reverse. In both cases, the ending base will be included in the displaying sequence.

- ✓ **Width:** Input the width in terms of pixels for output track display. Input the value that is suitable for your vision. Normally, it is assumed that width is equal to the screen width. It is also possible to input the width as 10 times bigger than the screen width. In that case, it seems to be convenient to use the scroll of the browser.

After manipulating the previous items, press the “Apply” button to shifting to the desired position of the sequence.

## Using shifting button method



Moreover, there is another method by using the button (like above image). Normally, the input box method is useful only when it is important to have the correct position of the sequence. As number must be supplied in this method, sometime this process is quit cumbersome. So we provide more intuitive methods for your convenience. So the buttons related to these methods are as follows,

- ✓ **Shifting** ( $\ll, \leftarrow, \rightarrow, \gg$ ):  $\ll, \leftarrow, \rightarrow, \gg$  are buttons to shifting the sequence. The  $\ll$  button scrolls sequence up to one screen size where the  $\leftarrow$  button scrolls half screen to the left. The inverse buttons are for shifting to the right.
- ✓ **Scaling** ( $\otimes, \otimes, \otimes_{\frac{10}{1}}, \otimes_{\frac{1}{10}}, \otimes_{\frac{1}{100}}, \otimes_{\frac{1}{1000}}$ ): The  $\otimes$  button does scaling down and the display will be half of the present size where  $\otimes$  button does scaling up and the display will be 2 times bigger than present size. The fractional form button specifies the scale directly. The scaling unit is 1bp/pixel. For example,  $\otimes_{\frac{1}{10}}$  means 1 bp is displayed by 10 pixels, whereas  $\otimes_{\frac{10}{1}}$  means 10 bp is displayed by 1 pixel. So  $\otimes_{\frac{10}{1}}$  means the scale that enlarges the display up to the base level.  $\otimes_{\frac{1}{1000}}$  is a scale ratio useful in displaying 1M bps on the screen. At the time of scaling up or down the view points of the present display will be remained unchanged.

## Using overview, ruler and zooming track method

In addition, there are some methods to shifting. These are overview, ruler, zooming trucks. You can shift the viewpoint by clicking any position of these tracks. For details, please refer to the respective details about the each track.

## 4.2. Functions of buttons

- ✓ **(rev)** You can display the sequence in reverse mode by UT Genome Browser. Normally a sequence is displayed from left to right but you can reverse it with the (rev) button. An alternative way to display in reverse way is to input “start” value greater than “end” value in the Input Box. The same task can be done easily by clicking the (rev) button. As the (rev) button reverses start and end points of the present displayed sequence so this can change the display mode from normal to reverse or vice versa. If you press the (fasta) button on reverse mode, then the complementary strand of displayed strand will be retrieved.
- ✓ **(top)** By pressing this button you can return to the main screen. Information about present displayed range, track and each track setting etc. will be preserved.
- ✓ **(clear)** By pressing this button you can return to the main screen. Information about present displayed range, track and each track setting etc. will be initialized.
- ✓ **(Track)** Addition, deletion, and rearranging can be performed by this button. Details description will be given later.
- ✓ **(fasta)** Present displaying sequence can be retrieved with this button in fasta format. You will be instructed to save as a file name. For example, Medaka-200406-scaffold429 (64587-66612). fasta

## 4.3. Addition, Deletion, and Rearrangement of Tracks

displaying tracks

name	comment	operation
Overview Track		up   down   remove
Ruler Track	The ruler track.	up   down   remove
Zoomer Track		up   down   remove
BaseColor		up   down   remove
gap		up   down   remove
GC content Track		up   down   remove
mappedGene	medaka	up   down   remove
genscan		up   down   remove

removed tracks

name	comment	operation
FuguScaffold		unremove

add new track

Track URL

When the (track) button of the main screen is pressed, the above track-editing screen will pop up. Here you can do addition, deletion and rearrangement of the tracks. This editing screen can be divided into three parts - displaying tracks, removed tracks, and adding new track.

- ✓ **Displaying track:** Here, tracks' deletion and rearrangement can be done to the present displayed screen. The deleted tracks will be replaced to next removed track.
- ✓ **Removed track:** This is the list of all tracks, which are deleted at Displaying Tracks. It is something like garbage bin. Undo-processing is possible for the tracks that are displayed here.
- ✓ **Add new track:** New tracks, which are not available at present display, can be added by this track. By setting the URL of the new tracks, push add button. If the new tracks' URL is read properly, the new tracks will be added after the present displayed tracks.

#### 4.4. Setting for Tracks

Config - [mappedGene] Close	
style select setting	<input type="radio"/> full <input checked="" type="radio"/> pack <input type="radio"/> small <input type="radio"/> dense <input type="button" value="Apply"/>
fromSpecies disp setting	<input checked="" type="checkbox"/> medaka <input checked="" type="checkbox"/> fugu <input checked="" type="checkbox"/> zebrafish <input type="button" value="Apply"/>
fromSpecies color setting	medaka: Black <input type="button" value="v"/> fugu: Green <input type="button" value="v"/> zebrafish: Cyan <input type="button" value="v"/> <input checked="" type="radio"/> use this color <input type="button" value="Apply"/>
matchRatio gradation setting	min[0.7] color: Default <input type="button" value="v"/> max[1.0] color: Default <input type="button" value="v"/> <input checked="" type="radio"/> use this color <input type="button" value="Apply"/>
matchRatio ulbound setting	matchRatio must be: <input type="text"/> <= matchRatio <= <input type="text"/> <input type="button" value="Apply"/>
coverRatio gradation setting	min[0.4] color: Default <input type="button" value="v"/> max[1.0] color: Default <input type="button" value="v"/> <input checked="" type="radio"/> use this color <input type="button" value="Apply"/>
coverRatio ulbound setting	coverRatio must be: <input type="text"/> <= coverRatio <= <input type="text"/> <input type="button" value="Apply"/>
stage disp setting	<input checked="" type="checkbox"/> SSA somite_stage <input checked="" type="checkbox"/> SSB somite_stage <input checked="" type="checkbox"/> SDA organogenesis_stage <input checked="" type="checkbox"/> FSA fry_stage <input checked="" type="checkbox"/> FFA fry_stage full_length <input checked="" type="checkbox"/> unknown <input type="button" value="Apply"/>
stage color setting	SSA somite_stage: Default <input type="button" value="v"/> SSB somite_stage: Default <input type="button" value="v"/> SDA organogenesis_stage: Default <input type="button" value="v"/> FSA fry_stage: Default <input type="button" value="v"/> FFA fry_stage full_length: Default <input type="button" value="v"/> unknown: Default <input type="button" value="v"/> <input checked="" type="radio"/> use this color <input type="button" value="Apply"/>

On main screen, when the  button, beside the track name, is pressed, the display of that track can be customized as above image. The customizing contents of the tracks differ from each other. On the other hand, there are some tracks which contents cannot be customized at all. For details, please refer to the individual tracks information.

## 5. Pre-defined Tracks

### 5.1. Overview Track



This track shows the range of present displayed window in the total genome. In above image, for instance, the blue-colored viewpoint shows the range from 125kb to 145kb. Clicking on any position of this overview track moves the viewpoint the middle position. For example, when you click on around 100k positions, the viewpoint will move to 90k-110k.

### 5.2. Ruler and Zooming Tracks



Both the ruler track and the zooming track are displaying same contents with different color. These two show the base position number of the present displayed content with respect to the total sequence. If you click any position of the ruler track, the viewpoint will move to the middle position. For example, if the rightmost end of the ruler track is clicked then the same effect will be happened as the ► button of main screen is pressed. Before using the zoom button on main screen, it is better to have the required region on middle position by using the ruler track to smoothing future enlarge processing.

In addition to the effect of the ruler track, if you click any position on the zooming track then two times zoom will be happened. For example, it is same to click the central button on zoom track as the 🔍 button on main screen. On the other hand, it is also same to clicking the rightmost end of the zooming track as click ► then click 🔍 on main screen. This track is useful to zoom out our interested region on the sequence..

### 5.3. Base Color Track



Base color track changes the sequence to color mode. All four ATGC bases are displayed by converting to the individual color. It is natural to display each base pair by one pixel.

By zooming in, you can see any specific region of the sequence in alphabet symbol (above bottom image).

It is also possible to customize the corresponding base pair color also. For example, if GC base pair are only colored then we can use it as a simple GC content track. To get the sequence itself, it is better to save from fasta track.

## 5.4. Gap Track



Gap Track is a track that displays gaps in a scaffold. By clicking any place except the gap, we can get the whole contig sequence in FASTA format. It is different from the (fasta) button in the main window, because it does not retrieve the displayed sequence only, but saves the whole contig sequence.

## 5.5. GC Content Track



This track displays the ratio of GC contents for 5 pixels. If the pixel number of a base pair is more than 5, the coloring will be done to tell whether it is GC or AT.

## 5.6. Mapped Gene Track



This track displays the mapping result of Fugu, Zebra fish and Medaka's Est to the Medaka genome, done by ALPS (<http://alps.gi.k.u-tokyo.ac.jp>). The mapping region for the gene is displayed by line and the exons are shown in rectangular box. The arrow sign shows the plus or minus strand of the genome. The above image displays Medaka's ESTs in black color, while the green color indicates cDNAs of Fugu. The current data sources of mapped genes are Medaka Unigene Build#10, Zebra fish Unigene Build#71, and Fugu ensemble pufferfish v21.2.c1( 10<sup>th</sup> May, 2004). We will revise the alignment

periodically in response to the update of the data sources.

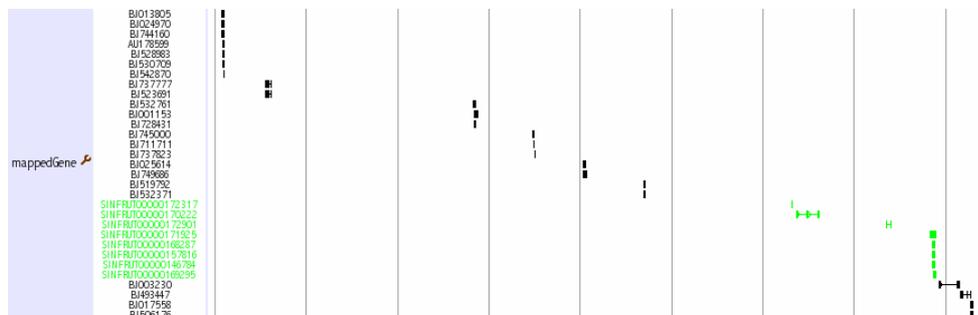
## Display setting

If the button on the left side of Mapped Gene Track is pressed, the display setting windows will pop up. There are three parts of this setting; namely, changing the total display style, alternating the mapping results if a gene will be displayed or not, and modifying the color of a gene. The setting, whether to show or not, will display all which fulfill the conditions. The color setting can be done by “use this color” button beside the track.

### Style select setting

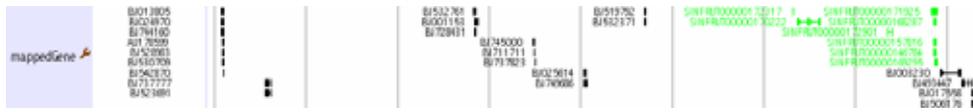
Predicted genes' display setting can be changed here. There are four kinds of display style. These styles are full, pack, small and dense.

- ✓ **Full:** In full style, each gene's position is displayed on one line. On left side the Genebank accession number or Ensemble Gene ID is written.



- ✓ **Pack:** By showing individual alignments together with the names (Genebank Acc or Ensemble Gene ID) of genes on left side in multiple on one line, the pack

style is more compact than full style.



- ✓ **Small:** In small style, the up and down spaces for one gene are reduced to the minimum. Unlike the full style and the pack style, this does not display the names of the genes.



- ✓ **Dense:** In this style, only the positions of aligned exons are displayed on one line.



At present except the full style, if the numbers of the mapped genes are more than 200 in a range of more than 50kbp, the graphical view of the genes will be displayed.

### FormSpecies disp setting

Here we can set which species among Medaka, Zebra fish, Fugu's mapping results are displayed or not. On default, all species results will be displayed. Only the species, which are checked on the Checkbook, will be displayed.

### FromSpecies color setting

The color distinguishing for Medaka, Zebra fish, Fugu's can be done here.

### MatchRatio gradation setting

Gene's color can be changed with respect to its match ratio. Provided colors for the match ratios 0.7 and 1.0 respectively, the system automatically gradates the coloring of the alignments between the lower and upper match ratios.

### MatchRatio ulbound setting

You can specify the range of the mapping ratio by inputting the lower and upper bounds into the corresponding boxes. Alignments of match ratios within the range are only displayed.

### CoverRatio gradation setting

Gene's color can also be changed with respect to its cover ratio. Provided colors for the cover ratios 0.4 and 1.0 respectively, the system automatically gradates the coloring of the alignments between the lower and upper cover ratios.

### CoverRatio ulbound setting

You can specify the range of the cover ratio by inputting the lower and upper bounds into the corresponding boxes. Alignments of cover ratios within the range are only displayed.

### Stage disp setting

Here, you can fix which development stage's expression EST should be displayed.

### Stage color setting

Color distinguishing can be done with respect to development stage.

### Linkage for each gene

In full, pack and small style, each gene can be clicked on. By clicking each gene, you can see the details of this gene.

## 5.7. Genscan Track



This track displays the predicted genes by Genscan. The predicted genes are viewed on lines, where the exons parts are on rectangular shape. Arrow sign presents the strand of the genes. The above image shows the predicted genes in pack style.

### Display settings

If the  mark button on the left side of Genscan Track is clicked, the display setting will pop up. Here are the details of this setting.

### Style selects setting

Predicted genes' display setting can be changed here. There are four kinds of display style. These styles are full, pack, small and dense, and they are similar to those styles in Mapped Gene Track.

### Linkage of each gene

In full, pack and small style each gene can be clicked on. After clicking each gene, the details of prediction by Genscan will be appeared.

## 5.8. Comparative Genomics Track

### Fugu Scaffold track

#### Description

This track shows Fugu/Medaka homologous scaffolds detected by ALPS alignment program. Fugu scaffolds are denoted by boxes connected by arrows. The boxes represent regions of high homology (match ratio > 60%) aligned by ALPS. The arrows represent low homology regions or gaps in Fugu scaffold. The direction of arrows indicates the orientation of Fugu/Medaka alignments. Clicking on a Fugu scaffold will open a new window to display its dotplot with Fugu scaffold sequence provided in the same page.

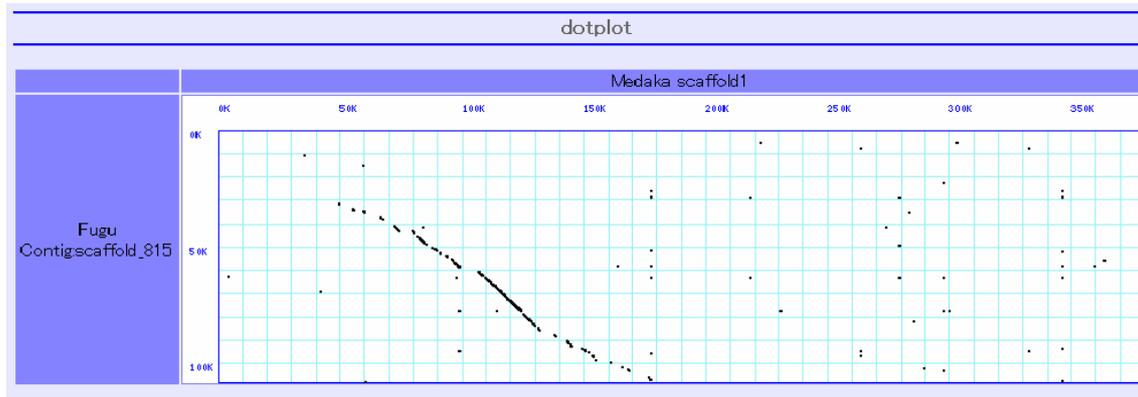
The Fugu sequence (Fugu v.2.0) was downloaded from JGI.

### Method

Fugu scaffold sequences are split into non-overlapping 300mer sequences and these 300mer sequences are mapped to Medaka scaffolds with ALPS. ALPS alignments with match ratio less than 60% are discarded and remaining alignments are chained by longest monotone subsequence algorithm. Chains consist of more than 10 alignments are displayed in the track. Note that inversions or microrearrangements are not shown. Only the longest monotone subsequence is displayed for each Fugu scaffold.

### Linkage of each scaffold

In full, pack and small style each gene can be clicked on. After clicking each scaffold, the detailed alignments will be appeared as like dot plot. Dotplot shows homologous regions of Fugu/Medaka scaffolds as diagonal runs of dots. Each dot, plotted based on sequence similarity score, indicates that significantly many seed matches are found between corresponding regions. The sequence similarity score is defined in such a way that tandem repeats are not assigned high scores while unique sequences are assigned high scores.



## **Ramen Assembler / UT Genome Browser Team Members**

### Ramen Genome Assembler Development Team

- ✓ Development of “Ramen” genome assembler and assembly of medaka genome:  
Masahiro Kasahara and Shin Sasaki
- ✓ Development of “Ramen Viewer” for genome assembly:  
Yukinobu Nagayasu

### UT Genome Browser Development Team

- ✓ Design and development of UT Genome Browser, keyword search function, libraries for describing tracks:  
Yukinobu Nagayasu and Koichiro Doi
- ✓ Online mapping function for query sequences:  
Tomoyuki Yamada
- ✓ Comparative Genomics Track:  
Yoichiro Nakatani and Wei Qu
- ✓ Gene Prediction:  
Ahsan Budrul
- ✓ Mapped Gene Track:  
Yasuhiro Kasai
- ✓ Database access accelerators:  
Takehiro Furudate and Atsushi Mori
- ✓ Overall management:  
Koichiro Doi and Shinichi Morishita

## **Acknowledgements**

This work has been supported by Grant-in-Aid for Scientific Research on Priority Areas (Grant#12209003) to Shinichi Morishita.

Ramen Assembler Development Team members are indebted to Yuji Kohara and Tadasu Shin-i for their technical discussions on the whole genome shotgun assembly.

Members in the UT Genome Browser Development Team are grateful to Kiyoshi Naruse, Daisuke Kobayashi, and Takanori Narita for their valuable input to improve the functions of the browser in a variety of ways.