

tatajuba

0.1

Generated by Doxygen 1.8.13

Contents

1	Tatajuba source code documentation	2
2	Data Structure Index	2
2.1	Data Structures	2
3	File Index	2
3.1	File List	2
4	Data Structure Documentation	3
4.1	__bmc2_kstring_t Struct Reference	3
4.2	arg_parameters Struct Reference	3
4.3	cont_hist_ptr_t Struct Reference	3
4.4	context_histogram_struct Struct Reference	4
4.4.1	Field Documentation	4
4.5	g_tract_t Struct Reference	5
4.6	g_tract_vector_struct Struct Reference	6
4.7	genome_set_struct Struct Reference	7
4.8	genomic_context_list_struct Struct Reference	8
4.9	hopo_counter_struct Struct Reference	8
4.10	hopo_element Struct Reference	9
4.11	tatajuba_options_t Struct Reference	9
5	File Documentation	10
5.1	context_histogram.h File Reference	10
5.1.1	Detailed Description	11
5.2	genome_set.h File Reference	11
5.2.1	Detailed Description	12
	Index	13

1 Tatajuba source code documentation

This is the doxyfile-generated documentation for `tatajuba`.

This documentation relates to structures, functions, etc. used by the library.

2 Data Structure Index

2.1 Data Structures

Here are the data structures with brief descriptions:

__bmc2_kstring_t	3
arg_parameters	3
cont_hist_ptr_t	3
context_histogram_struct	4
g_tract_t	5
g_tract_vector_struct	6
genome_set_struct	7
genomic_context_list_struct	8
hopo_counter_struct	8
hopo_element	9
tatajuba_options_t	9

3 File Index

3.1 File List

Here is a list of all documented files with brief descriptions:

context_histogram.h	
Homopolymer counter	10
genome_set.h	
Set of homopolymer counters (set of single end or paired end fastq files)	11
kseq.h	??

4 Data Structure Documentation

4.1 __bmc2_kstring_t Struct Reference

Data Fields

- unsigned **l**
- unsigned **m**
- char * **s**

The documentation for this struct was generated from the following file:

- kseq.h

4.2 arg_parameters Struct Reference

Data Fields

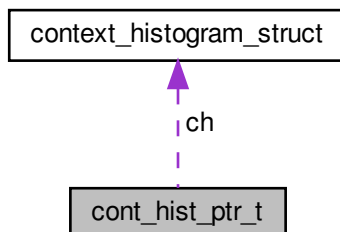
- struct arg_lit * **help**
- struct arg_lit * **version**
- struct arg_lit * **paired**
- struct arg_int * **kmer**
- struct arg_int * **minsize**
- struct arg_int * **minread**
- struct arg_int * **maxdist**
- struct arg_int * **leven**
- struct arg_int * **threads**
- struct arg_file * **gff**
- struct arg_file * **fna**
- struct arg_file * **fastq**
- struct arg_end * **end**
- void ** **argtable**

The documentation for this struct was generated from the following file:

- main.c

4.3 cont_hist_ptr_t Struct Reference

Collaboration diagram for cont_hist_ptr_t:



Data Fields

- [context_histogram_t](#) **ch**
- int **location**
- int **integral**

The documentation for this struct was generated from the following file:

- `__unused.c`

4.4 context_histogram_struct Struct Reference

Data Fields

- uint64_t * **context**
- int8_t [base](#)
now a vector since we store all within distance
- char * [name](#)
homopolymer base (AT or CG)
- int [n_context](#)
context name is flanking kmers with tract base in the middle
- int [integral](#)
vector size (of neighbourhood)
- int [location](#)
sum of frequencies
- int [coverage](#)
genomic location(s) of context
- int **n_tracts**
- int [mode_context_count](#)
- int [mode_context_length](#)
frequency of reads, defining "best homopolymer+context"
- int [mode_context_id](#)
tract length of best homopolymer+context
- int * [tmp_count](#)
which context (from neighbourhood) has best homopolymer+context
- int * **tmp_length**
- int **index**
- empfreq **h**
- gff3_fields **gfeature**
- int [ref_counter](#)
could be a list (even for a single position on a single genome) but here we store first belonging

4.4.1 Field Documentation

4.4.1.1 mode_context_count

```
int context_histogram_struct::mode_context_count
```

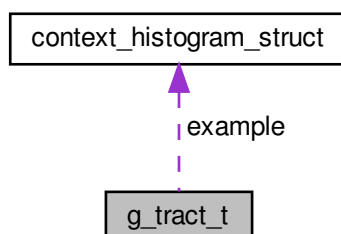
values related to genome (not histogram)

The documentation for this struct was generated from the following file:

- [context_histogram.h](#)

4.5 g_tract_t Struct Reference

Collaboration diagram for g_tract_t:



Data Fields

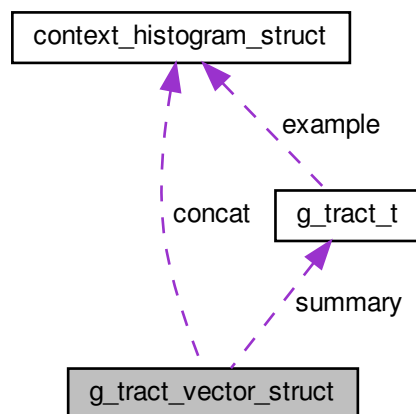
- int **location**
- int **n_dist**
- int **lev_distance**
- int **id_in_concat**
- double * **tab0**
- double * **d1**
- double * **d2**
- double * **gentab** [N_SUMMARY_TABLES]
- double **reldiff** [N_SUMMARY_TABLES]
- int **n_genome_total**
- int **n_genome_id**
- int * **genome_id**
- [context_histogram_t](#) **example**

The documentation for this struct was generated from the following file:

- [genome_set.h](#)

4.6 g_tract_vector_struct Struct Reference

Collaboration diagram for g_tract_vector_struct:



Data Fields

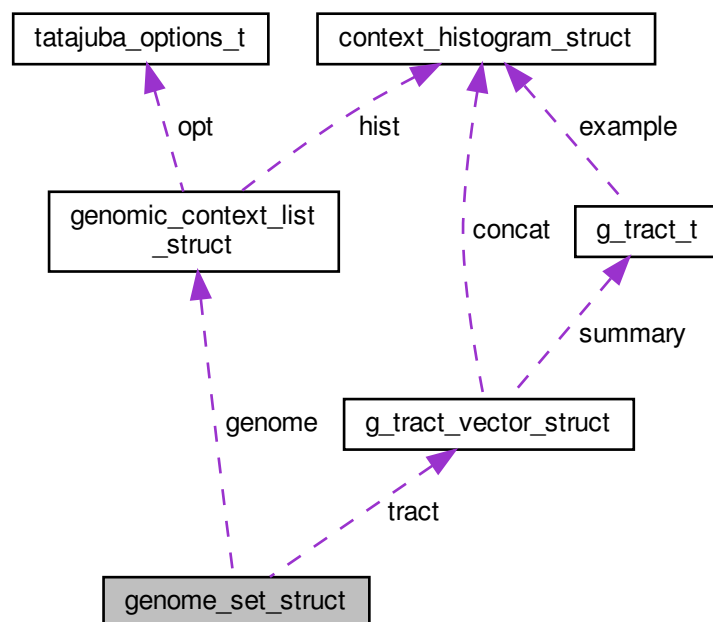
- [g_tract_t](#) * **summary**
- [context_histogram_t](#) * **concat**
- int **n_summary**
- int **n_concat**

The documentation for this struct was generated from the following file:

- [genome_set.h](#)

4.7 genome_set_struct Struct Reference

Collaboration diagram for genome_set_struct:



Data Fields

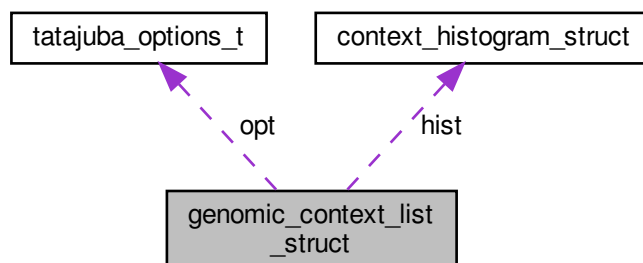
- [genomic_context_list_t](#) * **genome**
- [g_tract_vector_t](#) **tract**
- int **n_genome**
- double **secs** [3]
- int **ref_counter**

The documentation for this struct was generated from the following file:

- [genome_set.h](#)

4.8 genomic_context_list_struct Struct Reference

Collaboration diagram for genomic_context_list_struct:



Data Fields

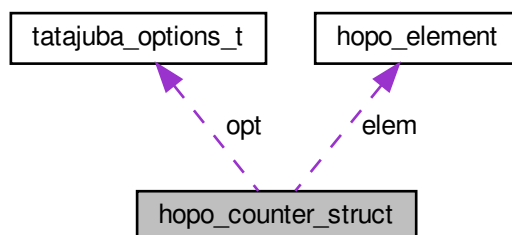
- [context_histogram_t](#) * **hist**
- char * **name**
- [tatajuba_options_t](#) **opt**
- int **n_hist**
- int **coverage**
- int **ref_start**

The documentation for this struct was generated from the following file:

- [context_histogram.h](#)

4.9 hopo_counter_struct Struct Reference

Collaboration diagram for hopo_counter_struct:



Data Fields

- [hopo_element](#) * **elem**
- char * **name**
- int **n_elem**
- int **n_alloc**
- int **kmer_size**
- int **coverage**
- int * **idx**
- int **n_idx**
- [tatajuba_options_t](#) **opt**
- int **ref_counter**

The documentation for this struct was generated from the following file:

- [context_histogram.h](#)

4.10 hopo_element Struct Reference

Data Fields

- uint64_t **context** [2]
- int64_t **base**:2
- int64_t **length**:16
- int64_t **count**:32

The documentation for this struct was generated from the following file:

- [context_histogram.h](#)

4.11 tatajuba_options_t Struct Reference

Data Fields

- char * **reference_fasta_filename**
- bool **paired_end**
- gff3_t **gff**
- int **max_distance_per_flank**
- int **kmer_size**
- int **min_tract_size**
- int **levenshtein_distance**
- int **min_coverage**
- int **n_threads**

The documentation for this struct was generated from the following file:

- [context_histogram.h](#)

5 File Documentation

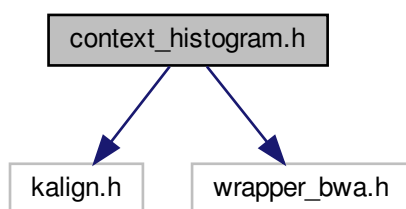
5.1 context_histogram.h File Reference

homopolymer counter

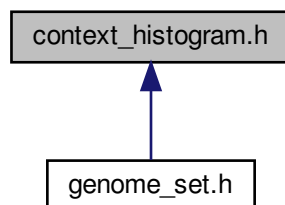
```
#include <kalign.h>
```

```
#include <wrapper_bwa.h>
```

Include dependency graph for context_histogram.h:



This graph shows which files directly or indirectly include this file:



Data Structures

- struct [tatajuba_options_t](#)
- struct [hopo_element](#)
- struct [hopo_counter_struct](#)
- struct [context_histogram_struct](#)
- struct [genomic_context_list_struct](#)

Typedefs

- typedef struct [hopo_counter_struct](#) * [hopo_counter](#)
- typedef struct [context_histogram_struct](#) * [context_histogram_t](#)
- typedef struct [genomic_context_list_struct](#) * [genomic_context_list_t](#)

Functions

- void **print_tatajuba_options** ([tatajuba_options_t](#) opt)
- [hopo_counter](#) **new_or_append_hopo_counter_from_file** ([hopo_counter](#) hc, const char *filename, [tatajuba_options_t](#) opt)
- void **del_hopo_counter** ([hopo_counter](#) hc)
- void **del_context_histogram** ([context_histogram_t](#) ch)
- void **print_debug_genomic_context_hist** ([genomic_context_list_t](#) genome)
- [genomic_context_list_t](#) **new_genomic_context_list** ([hopo_counter](#) hc)
- void **del_genomic_context_list** ([genomic_context_list_t](#) genome)
- void **finalise_genomic_context_hist** ([genomic_context_list_t](#) genome)
- int **distance_between_context_histograms** ([context_histogram_t](#) c1, [context_histogram_t](#) c2, double *result)
- bool **context_histograms_overlap** ([context_histogram_t](#) c1, [context_histogram_t](#) c2, int *distance, [tatajuba_options_t](#) opt)

Variables

- [uint8_t](#) **dna_in_2_bits** [256][2]
- char **bit_2_dna** []

5.1.1 Detailed Description

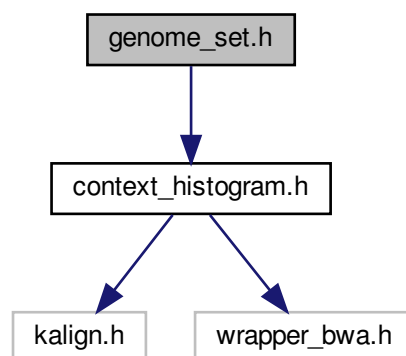
homopolymer counter

5.2 genome_set.h File Reference

set of homopolymer counters (set of single end or paired end fastq files)

```
#include "context_histogram.h"
```

Include dependency graph for genome_set.h:



Data Structures

- struct [g_tract_t](#)
- struct [g_tract_vector_struct](#)
- struct [genome_set_struct](#)

Macros

- `#define N_SUMMARY_TABLES 5`

Typedefs

- typedef struct [genome_set_struct](#) * **genome_set_t**
- typedef struct [g_tract_vector_struct](#) * **g_tract_vector_t**

Functions

- [genome_set_t](#) **new_genome_set_from_files** (const char **filenames, int n_filenames, [tatajuba_options_t](#) opt)
- void **del_genome_set** ([genome_set_t](#) g)
- void **print_selected_g_tract_vector** ([genome_set_t](#) g)
- void **print_debug_g_tract_vector** ([genome_set_t](#) g)

5.2.1 Detailed Description

set of homopolymer counters (set of single end or paired end fastq files)

Index

[__bmc2_kstring_t](#), [3](#)

[arg_parameters](#), [3](#)

[cont_hist_ptr_t](#), [3](#)

[context_histogram.h](#), [10](#)

[context_histogram_struct](#), [4](#)
 [mode_context_count](#), [4](#)

[g_tract_t](#), [5](#)

[g_tract_vector_struct](#), [6](#)

[genome_set.h](#), [11](#)

[genome_set_struct](#), [7](#)

[genomic_context_list_struct](#), [8](#)

[hopo_counter_struct](#), [8](#)

[hopo_element](#), [9](#)

[mode_context_count](#)
 [context_histogram_struct](#), [4](#)

[tatajuba_options_t](#), [9](#)