eXpress

0.90 BETA

http://bio.math.berkeley.edu/eXpress

Generated by Doxygen 1.7.3

Thu Sep 15 2011 22:50:16

Contents

Chapter 1

Class Index

1.1 Class Hierarchy

This inheritance list is sorted roughly, but not completely, alphabetically:

Chapter 2

Class Index

2.1 Class List

Here are the classes, structs, unions and interfaces with brief descriptions:

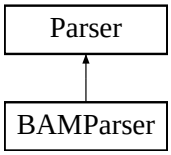
Chapter 3

Class Documentation

3.1 BAMParser Class Reference

`#include <mapparser.h>`

Inheritance diagram for BAMParser:



Public Member Functions

- BAMParser (BamTools::BamReader *reader)
- bool next_fragment (Fragment &f)

3.1.1 Detailed Description

The BAMParser class fills Fragment objects by parsing an input file in BAM format.

Author

Adam Roberts

Date

2011 Artistic License 2.0

Definition at line 46 of file mapparser.h.

3.1.2 Constructor & Destructor Documentation

3.1.2.1 BAMParser::BAMParser (BamTools::BamReader * *reader*)

BAMParser constructor opens the file

Definition at line 117 of file mapparser.cpp.

3.1.3 Member Function Documentation

3.1.3.1 bool BAMParser::next_fragment (Fragment & *f*) [virtual]

a member function that loads all mappings of the next fragment

Parameters

f	the empty Fragment to add mappings to

Returns

true if more reads remain in the BAM file, false otherwise

Implements Parser.

Definition at line 133 of file mapparser.cpp.

The documentation for this class was generated from the following files:

- src/mapparser.h
- src/mapparser.cpp

3.2 BiasBoss Class Reference

`#include <biascorrection.h>`

Public Member Functions

- BiasBoss (double alpha)
- void update_expectations (const Transcript &trans)
- void update_observed (const FragMap &frag, double mass)
- double get_transcript_bias (std::vector< double > &start_bias, std::vector< double > &end_bias, const Transcript &trans) const
- std::string to_string () const
- void append_output (std::ofstream &outfile) const

3.2.1 Detailed Description

The BiasBoss class keeps track of sequence-specific and positional bias. It allows for the bias associated with a given fragment end to be calculated, and for the bias parameters to be updated based on additional observations. All stored and returned values are in log space.

Author

Adam Roberts

Date

2011 Artistic License 2.0

Definition at line 208 of file biascorrection.h.

3.2.2 Constructor & Destructor Documentation

3.2.2.1 BiasBoss::BiasBoss (double *alpha*)

BiasBoss Constructor

Parameters

alpha	a double specifying the strength of the uniform prior (logged pseudo-counts for each paramater)

Definition at line 218 of file biascorrection.cpp.

3.2.3 Member Function Documentation

3.2.3.1 void BiasBoss::append_output (std::ofstream & *outfile*) const

a member function that outputs the positional and sequence-specific bias parameter matrices

Parameters

outfile	the file to append to

3.2.3.2 double BiasBoss::get_transcript_bias (std::vector< double > & *start_bias*, std::vector< double > & *end_bias*, const Transcript & *trans*) const

a member function that returns the 5' and 3' bias values at each position in a given transcript based on the current bias parameters

Parameters

start_bias	a vector containing the logged bias for each 5' start site in the transcript
end_bias	a vector containing the logged bias for each 3' end site in the transcript
trans	the transcript for which to calculate the logged bias

Returns

the product of the average 5' and 3' bias (logged)

Definition at line 284 of file biascorrection.cpp.

3.2.3.3 string BiasBoss::to_string ( ) const

a member function that returns a string containing the observed positional nucleotide probabilities (non-logged) in column-major order (A,C,G,T)

Returns

the string representation of the observed probabilities

Definition at line 313 of file biascorrection.cpp.

3.2.3.4 void BiasBoss::update_expectations (const Transcript & *trans*)

a member function that updates the expectation parameters (sequence-specific and positional) assuming uniform expression of and accross the transcript's sequence

Parameters

trans	the transcript to measure expected counts from

Definition at line 225 of file biascorrection.cpp.

3.2.3.5 void BiasBoss::update_observed (const FragMap & *frag,* double *mass*)

a member function that updates the observed parameters (sequence-specific and positional) given a fragment mapping to a transcript and its logged probabilistic assignment

Parameters

frag	the fragment mapping
mass	the logged probabality of the mapping, which is the amount to update the observed counts by

Definition at line 245 of file biascorrection.cpp.

The documentation for this class was generated from the following files:

- src/biascorrection.h

- src/biascorrection.cpp

3.3 FLD Class Reference

`#include <fld.h>`

Public Member Functions

- FLD (double alpha, size_t max_val, size_t mean, size_t std_dev)
- size_t max_val () const
- double mean () const
- void add_val (size_t len, double mass)
- double pdf (size_t len) const
- double tot_mass () const
- std::string to_string () const
- void append_output (std::ofstream &outfile) const

3.3.1 Detailed Description

The FLD class keeps track of the observed fragment length distribution. It starts with a Gaussian prior with parameters specified by the arguments. A small "Gaussian" kernel is added for each observation. All mass values and probabilities are stored and returned in log space (except in to_string).

Definition at line 21 of file fld.h.

3.3.2 Constructor & Destructor Documentation

3.3.2.1 FLD::FLD (double *alpha*, size_t *max_val*, size_t *mean*, size_t *std_dev*)

FLD Constructor

Parameters

alpha	double that sets the average pseudo-counts (logged)
max_val	an integer that sets the maximum allowable FragMap length
mean	a size_t for the mean of the prior gaussian dist
std_dev	a size_t for the std dev of the prior gaussian dist

Definition at line 21 of file fld.cpp.

3.3.3 Member Function Documentation

3.3.3.1 void FLD::add_val (size_t *len*, double *mass*)

a member function that updates the distribution based on a new FragMap observation

Parameters

len	an integer for the observed FragMap length
mass	a double for the mass (logged) of the observed FragMap

Definition at line 46 of file fld.cpp.

3.3.3.2 void FLD::append_output (std::ofstream & *outfile*) const

a member function that appends the FLD parameters to the end of the given file

Parameters

outfile	the file to append to

3.3.3.3 size_t FLD::max_val ( ) const

a member function that returns the maximum allowed FragMap length

Returns

max allowed FragMap length

Definition at line 41 of file fld.cpp.

3.3.3.4 double FLD::mean ( ) const

a member function that returns the mean FragMap length

Returns

mean observed FragMap length

Definition at line 74 of file fld.cpp.

3.3.3.5 double FLD::pdf (size_t *len*) const

a member function that returns the (logged) probability of a given FragMap length

Parameters

len	an integer for the FragMap length to return the probability of

Returns

(logged) probability of observing the given FragMap length

Definition at line 62 of file fld.cpp.

3.3.3.6 string FLD::to_string ( ) const

a member function that returns a string containing the current distribution

Returns

space-separated string of probabilities ordered from length 0 to max_val (non-logged)

Definition at line 79 of file fld.cpp.

3.3.3.7 double FLD::tot_mass ( ) const

a member function that returns the (logged) number of observed fragmaps (including pseudo-counts)

Returns

number of observed fragments

Definition at line 69 of file fld.cpp.

The documentation for this class was generated from the following files:

- src/fld.h
- src/fld.cpp

3.4 FragMap Struct Reference

`#include <fragments.h>`

Public Member Functions

- int length () const
- PairStatus pair_status () const

Public Attributes

- std::string name
- TransID trans_id
- Transcript * mapped_trans

- std::string seq_l
- std::string seq_r
- int left
- int right
- int mate_l
- bool left_first

3.4.1 Detailed Description

The FragMap struct stores the information for a single (multi-)mapping of a fragment.

Author

Adam Roberts

Date

2011 Artistic License 2.0

Definition at line 36 of file fragments.h.

3.4.2 Member Function Documentation

3.4.2.1 int FragMap::length ( ) const [inline]

a member function returning the length of the fragment according to this mapping note, that this result will be invalid if the fragment is single-end

Returns

int length of fragment mapping

Definition at line 94 of file fragments.h.

3.4.2.2 PairStatus FragMap::pair_status ( ) const [inline]

a member function returning whether the mapping is PAIRED, LEFT_ONLY, or RIGHT_-ONLY LEFT_ONLY denotes that the single read is not reverse complemented => its left end is the left fragment end RIGHT_ONLY denotes that the single read is reverse complemented => its right end is the right fragment end

Returns

PairStatus the pair status of the mapping

Definition at line 105 of file fragments.h.

3.4.3 Member Data Documentation

3.4.3.1 int FragMap::left

a public int containing the 0-based leftmost coordinate mapped to in the transcript valid only if PairStatus is PAIRED or LEFT_ONLY

Definition at line 67 of file fragments.h.

3.4.3.2 bool FragMap::left_first

a public bool specifying that the "right" (second according to SAM flag) is reverse complemented when true and the "left" (first according to SAM flag) is reverse complemented when false in other words, the "left" read is truly left of the "right" read in transcript coordinate space when true

Definition at line 87 of file fragments.h.

3.4.3.3 Transcript* FragMap::mapped_trans

a public pointer to the transcript mapped to

Definition at line 51 of file fragments.h.

3.4.3.4 int FragMap::mate_l

a public int containing the left position for the mate of the first read read in from the SAM file 0 if single-end fragment this is temporarily used to help find the mate, but is not important later on

Definition at line 80 of file fragments.h.

3.4.3.5 std::string FragMap::name

a public string for the SAM "Query Template Name" (fragment name)

Definition at line 41 of file fragments.h.

3.4.3.6 int FragMap::right

a public int containing the position following the 0-based rightmost coordinate mapped to in the transcript valid only if PairStatus is PAIRED or RIGHT_ONLY

Definition at line 73 of file fragments.h.

3.4.3.7 std::string FragMap::seq_l

a public string containing the "left" read sequence (first according to SAM flag)

Definition at line 56 of file fragments.h.

3.4.3.8 std::string FragMap::seq_r

a public string containing the "right" read sequence (second according to SAM flag)

Definition at line 61 of file fragments.h.

3.4.3.9 TransID FragMap::trans_id

a public TransID for the transcript mapped to

Definition at line 46 of file fragments.h.

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

- src/fragments.h

3.5 Fragment Class Reference

`#include <fragments.h>`

Public Member Functions

- ~Fragment ()
- bool add_map_end (FragMap *f)
- const std::string name () const
- const size_t num_maps () const
- const std::vector< FragMap * > & maps () const

3.5.1 Detailed Description

The Fragment class stores information for all multi-mappings of a single fragment. By design, only paired-end mappings of paired-end reads will be accepted. All mappings of single-end reads will be accepted.

Author

Adam Roberts

Date

2011 Artistic License 2.0

Definition at line 123 of file fragments.h.

3.5.2 Constructor & Destructor Documentation

3.5.2.1 Fragment::~Fragment ( )

Fragment destructor deletes all FragMap objects pointed to by the Fragment

Definition at line 14 of file fragments.cpp.

3.5.3 Member Function Documentation

3.5.3.1 bool Fragment::add_map_end (FragMap * f)

a member function that adds a new FragMap (single read at this point) to the Fragment if it is the first FragMap, it sets the Fragment name and is added to _open_mates, if the fragment is not paired, it is added to _frag_maps, otherwise, add_open_mate is called

Parameters

f	the FragMap to be added

Definition at line 27 of file fragments.cpp.

3.5.3.2 const std::vector<FragMap*>& Fragment::maps ( ) const [inline]

a member function that returns FragMap multi-mappings of the fragment

Returns

a vector containing pointers to the FragMap multi-mappings

Definition at line 181 of file fragments.h.

3.5.3.3 const std::string Fragment::name ( ) const [inline]

a member function that returns the SAM "Query Template Name" (fragment name)

Returns

the string SAM "Query Template Name" (fragment name)

Definition at line 169 of file fragments.h.

3.5.3.4 const size_t Fragment::num_maps ( ) const [inline]

a member function that returns the number of multi-mappings for the fragment

Returns

number of multi-mappings for fragment

Definition at line 175 of file fragments.h.

The documentation for this class was generated from the following files:

- src/fragments.h
- src/fragments.cpp

3.6 FrequencyMatrix Class Reference

`#include <frequencymatrix.h>`

Public Member Functions

- FrequencyMatrix ()
- FrequencyMatrix (size_t m, size_t n, double alpha)
- double operator() (size_t i, size_t j) const
- double operator() (size_t k) const
- void increment (size_t i, size_t j, double incr_amt)
- void increment (size_t k, double incr_amt)
- double arr (size_t k) const
- double row (size_t i) const

3.6.1 Detailed Description

The FrequencyMatrix class keeps track of the frequency parameters in order to allow for constant-time probability look-ups and updates. The table is rectangular to allow for multiple distributions to be stored in one FrequencyMatrix. Rows are distributions. All values are stored and returned in log space.

Author

Adam Roberts

Date

2011 Artistic License 2.0

Definition at line 24 of file frequencymatrix.h.

3.6.2 Constructor & Destructor Documentation

3.6.2.1 FrequencyMatrix::FrequencyMatrix ( ) [inline]

dummy constructor

Definition at line 51 of file frequencymatrix.h.

3.6.2.2 FrequencyMatrix::FrequencyMatrix (size_t *m*, size_t *n*, double *alpha*)

FrequencyMatrix constructor initializes the matrix based on the log of the given pseudo-counts

Parameters

m	a size_t specifying the number of distributions (rows)
n	a size_t specifying the number of values in each distribution (columns)
alpha	a double specifying the intial psuedo-counts (un-logged)

Definition at line 16 of file frequencymatrix.cpp.

3.6.3 Member Function Documentation

3.6.3.1 double FrequencyMatrix::arr (size_t *k*) const [inline]

a member function that returns the raw value stored at a given position of the flattened matrix

Parameters

k	the array position

Returns

a double specifying the raw value stored at the given position of the flattened matrix

Definition at line 96 of file frequencymatrix.h.

3.6.3.2 void FrequencyMatrix::increment (size_t *i*, size_t *j*, double *incr_amt*)

a member function to increase the mass of a given position in the matrix

Parameters

i	the distribution (row)
j	the value (column)
incr_amt	the logged amount to increase the mass by

Definition at line 37 of file frequencymatrix.cpp.

3.6.3.3 void FrequencyMatrix::increment (size_t *k*, double *incr_amt*)

a member function to increase the mass of a given position in the flattened matrix

Parameters

k	the array position

incr_amt	the logged amount to increase the mass by

Definition at line 45 of file frequencymatrix.cpp.

3.6.3.4 double FrequencyMatrix::operator() (size_t *k*) const

a member function to extract the logged probability of a given position in the flattened matrix

Parameters

k	the array position

Returns

a double specifying the logged probability of the given position in the flattened matrix

Definition at line 31 of file frequencymatrix.cpp.

3.6.3.5 double FrequencyMatrix::operator() (size_t *i*, size_t *j*) const

a member function to extract the logged probability of a given position in the matrix

Parameters

i	the distribution (row)
j	the value (column)

Returns

a double specifying the logged probability of the given value in the given distribution

Definition at line 24 of file frequencymatrix.cpp.

3.6.3.6 double FrequencyMatrix::row (size_t *i*) const [inline]

a member function that returns the raw row sum

Parameters

i	the distribution (row)

Returns

a double specifying the raw row sum for the given distribution

Definition at line 103 of file frequencymatrix.h.

The documentation for this class was generated from the following files:

- src/frequencymatrix.h
- src/frequencymatrix.cpp

3.7 MismatchTable Class Reference

`#include <mismatchmodel.h>`

Public Member Functions

- MismatchTable (double alpha)
- void activate (bool active=true)
- double log_likelihood (const FragMap &f) const
- void update (const FragMap &, double mass)
- std::string to_string () const
- void append_output (std::ofstream &outfile) const

3.7.1 Detailed Description

The MismatchTable class is used to store and update mismatch (error) parameters using a first-order Markov model based on nucleotide and position in a ride and to return likelihoods of mismatches in given reads. All values are stored and returned in log space.

Author

Adam Roberts

Date

2011 Artistic License 2.0

Definition at line 24 of file mismatchmodel.h.

3.7.2 Constructor & Destructor Documentation

3.7.2.1 MismatchTable::MismatchTable (double *alpha*)

MismatchTable constructor initializes the model parameters using the specified (non-logged) pseudo-counts.

Parameters

alpha	a double containing the non-logged pseudo-counts for parameter initialization

Definition at line 18 of file mismatchmodel.cpp.

3.7.3 Member Function Documentation

3.7.3.1 void MismatchTable::activate (bool *active* = `true`) [inline]

member function that 'activates' the table to allow its values to be used in calculating log-likelihoods when it is sufficiently burned-in

Parameters

active	a boolean specifying whether to activate (true) or deactivate (false)

Definition at line 59 of file mismatchmodel.h.

3.7.3.2 void MismatchTable::append_output (std::ofstream & *outfile*) const

a member function that outputs the final model parameters in a tab-separated file the file has 1 row for each read position and the parameters are in columns indexed as (ref, prev, obs) in base 4 with A,C,G,T encoded as 0,1,2,3.

Parameters

file	stream to append to

Definition at line 85 of file biascorrection.cpp.

3.7.3.3 double MismatchTable::log_likelihood (const FragMap & *f*) const

member function returns the log likeihood of mismatches in the mapping given the current error model paramaters

Parameters

f	the fragment mapping to calculate the log likelihood for

Returns

the log likelihood of the mapping based on mismatches

Definition at line 25 of file mismatchmodel.cpp.

3.7.3.4 string MismatchTable::to_string ( ) const

member function that returns a string containing a collapsed confusion matrix based on the model parameters for the first read

Returns

a space-separated string for the flattened, collapsed confusion matrix in row-major

format (observed value as rows)

Definition at line 113 of file mismatchmodel.cpp.

3.7.3.5 void MismatchTable::update (const FragMap & *f*, double *mass*)

member function that updates the error model parameters based on a mapping and its (logged) mass

Parameters

f	the fragment mapping
mass	the logged mass to increase the parameters by

Definition at line 72 of file mismatchmodel.cpp.

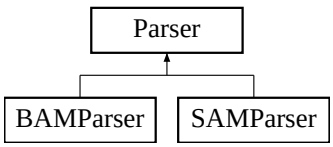
The documentation for this class was generated from the following files:

- src/mismatchmodel.h
- src/biascorrection.cpp
- src/fld.cpp
- src/mismatchmodel.cpp

3.8 Parser Class Reference

`#include <mapparser.h>`

Inheritance diagram for Parser:



Public Member Functions

- virtual bool next_fragment (Fragment &f)=0

3.8.1 Detailed Description

The Parser class is an abstract class that can be a SAMParser or BAMParser.

Author

Adam Roberts

Date

2011 Artistic License 2.0

Definition at line 27 of file mapparser.h.

3.8.2 Member Function Documentation

3.8.2.1 virtual bool Parser::next_fragment (Fragment & *f*) [pure virtual]

a member function that loads all mappings of the next fragment

Parameters

f	the empty Fragment to add mappings to

Returns

true if more reads remain in the SAM/BAM file/stream, false otherwise

Implemented in BAMParser, and SAMParser.

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

- src/mapparser.h

3.9 ParseThreadSafety Struct Reference

`#include <mapparser.h>`

Public Attributes

- Fragment * next_frag
- boost::mutex proc_lk
- boost::mutex parse_lk

3.9.1 Detailed Description

The ParseThreadSafety struct stores objects to allow for parsing to safely occur on a separate thread from processing.

Author

Adam Roberts

Date

2011 Artistic License 2.0

Definition at line 126 of file mapparser.h.

3.9.2 Member Data Documentation

3.9.2.1 Fragment* ParseThreadSafety::next_frag

a pointer to the next Fragment to be processed by the main thread

Definition at line 131 of file mapparser.h.

3.9.2.2 boost::mutex ParseThreadSafety::parse_lk

a mutex to lock the parsing thread when the next_frag pointer should be not modified

Definition at line 141 of file mapparser.h.

3.9.2.3 boost::mutex ParseThreadSafety::proc_lk

a mutex to lock the main (processing) thread when next_frag has not yet been updated

Definition at line 136 of file mapparser.h.

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

- src/mapparser.h

3.10 PosWeightTable Class Reference

`#include <biascorrection.h>`

Public Member Functions

- PosWeightTable (const std::vector< size_t > &len_bins, const std::vector< double > &pos_bins, double alpha)
- const std::vector< size_t > & **len_bins** () const
- const std::vector< double > & **pos_bins** () const
- void increment_expected (size_t len, double pos)
- void increment_expected (size_t l, size_t p)
- void increment_observed (size_t len, double pos, double normalized_mass)
- void increment_observed (size_t l, size_t p, double normalized_mass)
- double get_weight (size_t len, double pos) const
- double get_weight (size_t l, size_t p) const
- void append_output (std::ofstream &outfile) const

3.10.1 Detailed Description

The PosWeightTable class keeps track of fractional position bias parameters in log space. It allows for the bias associated with a given fractional position to be calculated, and for the bias parameters to be updated based on additional fragment observations.

Author

Adam Roberts

Date

2011 Artistic License 2.0

Definition at line 101 of file biascorrection.h.

3.10.2 Constructor & Destructor Documentation

3.10.2.1 PosWeightTable::PosWeightTable (const std::vector< size_t > & *len_bins*, const std::vector< double > & *pos_bins*, double *alpha*)

PosWeightTable Constructor

Parameters

len_bins	a vector of unsigned integers specifying the bin ranges for transcript lengths
pos_bins	a vector of doubles specifying the bin ranges for fractional positions
alpha	a double specifying the strength of the uniform prior (logged pseudo-counts for each paramater)

Definition at line 123 of file biascorrection.cpp.

3.10.3 Member Function Documentation

3.10.3.1 void PosWeightTable::append_output (std::ofstream & *outfile*) const

a member function that outputs the fractional position probabilities in matrix format with length bins as rows and fractional position bins as columns

Parameters

outfile	the file to append to

3.10.3.2 double PosWeightTable::get_weight (size_t *len*, double *pos*) const

a member function that return the bias weight (logged) of a fractional transcript position

Parameters

len	the transcript length
pos	the fractional transcript position

Returns

the logged bias weight for the fractional transcript position

Definition at line 156 of file biascorrection.cpp.

3.10.3.3 double PosWeightTable::get_weight (size_t *l*, size_t *p*) const

a member function that return the bias weight (logged) of a fractional transcript position bin

Parameters

l	the transcript length bin
p	the fractional transcript position bin

Returns

the logged bias weight for the fractional transcript position

Definition at line 55 of file biascorrection.cpp.

3.10.3.4 void PosWeightTable::increment_expected (size_t *l*, size_t *p*)

a member function that increments the expected counts for the given fractional position bin by 1 (logged)

Parameters

l	the transcript length bin
p	the fractional transcript position bin

Definition at line 137 of file biascorrection.cpp.

3.10.3.5 void PosWeightTable::increment_expected (size_t *len*, double *pos*)

a member function that increments the expected counts for the given fractional position by 1 (logged)

Parameters

len	the transcript length
pos	the fractional transcript position

Definition at line 130 of file biascorrection.cpp.

3.10.3.6 void PosWeightTable::increment_observed (size_t *len*, double *pos*, double *normalized_mass*)

a member function that increments the observed counts for the given fragment position by some mass (logged)

Parameters

len	the transcript length
pos	the fractional transcript position
normalized_mass	the mass (logged probabilistic assignment) of the fragment normalized by its estimated expression

Definition at line 143 of file biascorrection.cpp.

3.10.3.7 void PosWeightTable::increment_observed (size_t *l*, size_t *p*, double *normalized_mass*)

a member function that increments the observed counts for the given fragment position bin by some mass (logged)

Parameters

l	the transcript length bin
p	the fractional transcript position bin
normalized_mass	the mass (logged probabilistic assignment) of the fragment normalized by its estimated expression

Definition at line 44 of file biascorrection.cpp.

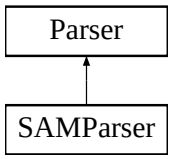
The documentation for this class was generated from the following files:

- src/biascorrection.h
- src/biascorrection.cpp

3.11 SAMParser Class Reference

`#include <mapparser.h>`

Inheritance diagram for SAMParser:



Public Member Functions

- SAMParser (std::istream *in)
- bool next_fragment (Fragment &f)

3.11.1 Detailed Description

The SAMParser class fills Fragment objects by parsing an input in SAM format. The input may come from a file or stdin.

Author

Adam Roberts

Date

2011 Artistic License 2.0

Definition at line 85 of file mapparser.h.

3.11.2 Constructor & Destructor Documentation

3.11.2.1 SAMParser::SAMParser (std::istream * *in*)

SAMParser constructor removes the header, and parses the first line

Definition at line 183 of file mapparser.cpp.

3.11.3 Member Function Documentation

3.11.3.1 bool SAMParser::next_fragment (Fragment & *f*) [virtual]

a member function that loads all mappings of the next fragment when the next fragment is reached, the current alignment is left in the _frag_buff for the next call

Parameters

f	the empty Fragment to add mappings to

Returns

true if more reads remain in the SAM file, false otherwise

Implements Parser.

Definition at line 211 of file mapparser.cpp.

The documentation for this class was generated from the following files:

- src/mapparser.h
- src/mapparser.cpp

3.12 SeqWeightTable Class Reference

```
#include <biascorrection.h>
```

Public Member Functions

- SeqWeightTable (size_t window_size, double alpha)
- void increment_expected (char c)
- void increment_observed (std::string &seq, double normalized_mass)
- double get_weight (const std::string &seq, int i) const
- std::string to_string () const
- void append_output (std::ofstream &outfile) const

3.12.1 Detailed Description

The SeqWeightTable class keeps track of sequence-specific bias parameters. It allows for the bias associated with a given sequence to be calculated, and for the bias parameters to be updated based on additional observations. All values stored in log space.

Author

Adam Roberts

Date

2011 Artistic License 2.0

Definition at line 29 of file biascorrection.h.

3.12.2 Constructor & Destructor Documentation

3.12.2.1 SeqWeightTable::SeqWeightTable (size_t *window_size*, double *alpha*)

SeqWeightTable Constructor

Parameters

window_size	an unsigned integer specifying the size of the bias window surrounding fragment ends
alpha	a double specifying the strength of the uniform prior (logged pseudo-counts for each paramater)

Definition at line 29 of file biascorrection.cpp.

3.12.3 Member Function Documentation

3.12.3.1 void SeqWeightTable::append_output (std::ofstream & *outfile*) const

a member function that outputs the positional nucleotide probabilities in matrix format with nucleotides (A,C,G,T) as rows and window position as columns

Parameters

outfile	the file to append to

3.12.3.2 double SeqWeightTable::get_weight (const std::string & *seq*, int *i*) const

a member function that calculates the bias weight (logged) of a bias window

Parameters

seq	the transcript sequence the fragment maps to
i	the fragment end point (the central point of the bias window)

Returns

the bias weight for the bias window which is the product of the individual nucleotide bias weights

3.12.3.3 void SeqWeightTable::increment_expected (char *c*)

a member function that increments the expected counts for the given nucleotide by 1 (logged)

Parameters

c	a char representing a nucleotide that has been observed in the transcriptome

Definition at line 34 of file biascorrection.cpp.

3.12.3.4 void SeqWeightTable::increment_observed (std::string & *seq*, double *normalized_mass*)

a member function that increments the observed counts for the given fragment sequence by some mass (logged)

Parameters

seq	a string of nucleotides in the bias window for the sequenced fragment end
normalized_mass	the mass (logged probabilistic assignment) of the fragment normalized by its estimated expression

3.12.3.5 string SeqWeightTable::to_string ( ) const

a member function that returns a string containing the positional nucleotide probabilities in column-major order (A,C,G,T)

Returns

the string representation of the positional nucleotide probabilities

Definition at line 68 of file biascorrection.cpp.

The documentation for this class was generated from the following files:

- src/biascorrection.h
- src/biascorrection.cpp

3.13 ThreadedMapParser Class Reference

`#include <mapparser.h>`

Public Member Functions

- ThreadedMapParser (std::string input_file)
- void threaded_parse (ParseThreadSafety *thread_safety, TranscriptTable *trans_table)

3.13.1 Detailed Description

The ThreadedMapParser class is meant to be run on as a separate thread from the main processing. Once started, this thread will read input from a file or stream in SAM/BAM format, parse, and collect read alignments into fragment alignments, and fragment alignments into fragments, which are placed on a buffer for the processing thread. Once the processing thread copies the fragment address from the buffer, the parser is unlocked to load the next fragment. The process stops when EOF is reached

Author

Adam Roberts

Date

2011 Artistic License 2.0

Definition at line 154 of file mapparser.h.

3.13.2 Constructor & Destructor Documentation

3.13.2.1 ThreadedMapParser::ThreadedMapParser (std::string *input_file*)

ThreadedMapParser constructor determines what format the input is in and initializes the correct parser.

Definition at line 51 of file mapparser.cpp.

3.13.3 Member Function Documentation

3.13.3.1 void ThreadedMapParser::threaded_parse (ParseThreadSafety * *thread_safety*, TranscriptTable * *trans_table*)

a member function that drives the parse thread when all valid mappings of a fragment have been parsed, its mapped transcripts are found and the information is passed in a Fragment object to the processing thread through the ParseThreadSafety struct

Parameters

thread_safety	a pointer to the struct containing shared locks and data with the processing thread
trans_table	a pointer to the table of Transcript objects to lookup the mapped transcripts

Definition at line 80 of file mapparser.cpp.

The documentation for this class was generated from the following files:

- src/mapparser.h
- src/mapparser.cpp

3.14 Transcript Class Reference

`#include <transcripts.h>`

Public Member Functions

- Transcript (const std::string &name, const std::string &seq, double alpha, const FLD *fld, const BiasBoss *bias_table, const MismatchTable *mismatch_table)
- const std::string & name () const
- TransID id () const
- const std::string & seq () const
- size_t length () const
- double mass () const
- double var () const
- size_t tot_counts () const
- size_t uniq_counts () const
- void add_mass (double p, double mass)
- void incr_bundle_counts (size_t incr_amt=1)
- size_t bundle_counts ()
- double log_likelihood (const FragMap &frag) const
- double effective_length () const
- double est_effective_length () const
- double unbiased_effective_length () const
- void update_transcript_bias ()

3.14.1 Detailed Description

The Transcript class is used to store objects for the transcripts being mapped to. Besides storing basic information about the object (id, length), it also stores a mass based on the number of fragments mapping to the object. To help with updating this number, it returns the likelihood that a given fragment originated from it. These values are stored and returned in log space.

Author

Adam Roberts

Date

2011 Artistic License 2.0

Definition at line 57 of file transcripts.h.

3.14.2 Constructor & Destructor Documentation

3.14.2.1 Transcript::Transcript (const std::string & *name*, const std::string & *seq*, double *alpha*, const FLD * *fld*, const BiasBoss * *bias_table*, const MismatchTable * *mismatch_table*)

Transcript Constructor

Parameters

name	a string that stores the transcript name
seq	a string that stores the transcript sequence
alpha	a double that specifies the intial pseudo-counts (non-logged)
fld	a pointer to the global Fragment Length Distribution (FLD) object
bias_table	a pointer to the global BiasBoss object
mismatch_table	a pointer to the global MismatchTable object

Definition at line 24 of file transcripts.cpp.

3.14.3 Member Function Documentation

3.14.3.1 void Transcript::add_mass (double *p*, double *mass*)

a member function that increases the expected fragment counts and variance by a given (logged) fragment mass

Parameters

p	a double for the (logged) probability that the fragment was generated by this transcript
mass	a double specifying the (logged) mass of the fragment being mapped

Definition at line 41 of file transcripts.cpp.

3.14.3.2 size_t Transcript::bundle_counts ( ) [inline]

a member function that returns the counts mapping to the bundle this transcript is in the total bundle counts is the sum of this value for all transcripts in the bundle

Returns

a portion of the counts mapping to the bundle this transcript is in

Definition at line 231 of file transcripts.h.

3.14.3.3 double Transcript::effective_length ( ) const

a member function that calcualtes and returns the effective length of the transcript (non-logged)

Returns

the effective length of the transcript calculated as $\tilde{l} = \sum_{l=1}^{L(t)} \sum_{i=1}^{L(t)} D(l) b_5[i] * b_3[i+l]$

Definition at line 110 of file transcripts.cpp.

3.14.3.4 double Transcript::est_effective_length ( ) const

a member function that calcualtes and returns the estimated effective length of the transcript (non-logged) using the avg bias

Returns

the estimated effective length of the transcript calculated as $\tilde{l} = \bar{bias} \sum_{l=1}^{L(t)} D(l)(L(t) - l + 1)$

Definition at line 91 of file transcripts.cpp.

3.14.3.5 TransID Transcript::id ( ) const [inline]

a member function that returns the transcript id

Returns

TransID transcript ID

Definition at line 172 of file transcripts.h.

3.14.3.6 void Transcript::incr_bundle_counts (size_t *incr_amt* = 1) [inline]

a member function that increases the counts mapping to the bundle this transcript is in the total bundle counts is the sum of this value for all transcripts in the bundle

Parameters

incr_amt	a size_t to increase the counts by

Definition at line 221 of file transcripts.h.

3.14.3.7 size_t Transcript::length ( ) const [inline]

a member function that returns the transcript length

Returns

transcript length

Definition at line 183 of file transcripts.h.

3.14.3.8 double Transcript::log_likelihood (const FragMap & *frag*) const

a member function that returns (a value proportional to) the log likelihood the given fragment originated from this transcript

Parameters

frag	a FragMap to return the likelihood of being originated from this transcript

Returns

(a value proportional to) the log likelihood the given fragment originated from this transcript

Definition at line 57 of file transcripts.cpp.

3.14.3.9 double Transcript::mass ( ) const [inline]

a member function that returns the current (logged) fragment mass

Returns

logged mass

Definition at line 189 of file transcripts.h.

3.14.3.10 const std::string& Transcript::name ( ) const [inline]

a member function that returns the transcript name

Returns

string containing transcript name

Definition at line 166 of file transcripts.h.

3.14.3.11 const std::string& Transcript::seq ( ) const [inline]

a member function that returns the transcript sequence

Returns

string containing transcript sequence

Definition at line 177 of file transcripts.h.

3.14.3.12 size_t Transcript::tot_counts ( ) const [inline]

a member function that returns the current count of fragments mapped to this transcript (uniquely or ambiguously)

Returns

total fragment count

Definition at line 201 of file transcripts.h.

3.14.3.13 double Transcript::unbiased_effective_length ( ) const

a member function that calcualtes and returns the effective length of the transcript (non-logged) ignoring bias and using the prior FLD distribution

Returns

the effective length of the transcript calculated as $\tilde{l} = \sum_{l=1}^{L(t)} D_{prior}(l)(L(t) - l + 1)$

Definition at line 105 of file transcripts.cpp.

3.14.3.14 size_t Transcript::uniq_counts ( ) const [inline]

a member function that returns the current count of fragments uniquely mapped to this transcript

Returns

unique fragment count

Definition at line 207 of file transcripts.h.

3.14.3.15 void Transcript::update_transcript_bias ( )

a member function that causes the transcript bias to be re-calculated by the _bias_table based on curent parameters

Definition at line 134 of file transcripts.cpp.

3.14.3.16 double Transcript::var ( ) const [inline]

a member function that returns the current (logged) variance

Returns

logged mass variance

Definition at line 195 of file transcripts.h.

The documentation for this class was generated from the following files:

- src/transcripts.h
- src/transcripts.cpp

3.15 TranscriptTable Class Reference

`#include <transcripts.h>`

Public Member Functions

- TranscriptTable (const std::string &trans_fasta_file, double alpha, const FLD *fld, BiasBoss *bias_table, const MismatchTable *mismatch_table)
- ~TranscriptTable ()
- Transcript * get_trans (TransID id)
- size_t size () const
- void update_covar (TransID trans1, TransID trans2, double covar)
- double get_covar (TransID trans1, TransID trans2)
- size_t covar_size () const
- TransID get_trans_rep (TransID trans)
- TransID merge_bundles (TransID rep1, TransID rep2)
- size_t num_bundles ()
- void output_bundles (std::string output_dir)
- void threaded_bias_update ()
- void output_results (std::string output_dir, size_t tot_counts, bool output_varcov)

3.15.1 Detailed Description

The TranscriptTable class is used to keep track of the Transcript objects for a run. The constructor parses a fasta file to generate the Transcript objects and store them in a map that allows them to be looked up based on their string id.

Author

Adam Roberts

Date

2011 Artistic License 2.0

Definition at line 285 of file transcripts.h.

3.15.2 Constructor & Destructor Documentation

3.15.2.1 TranscriptTable::TranscriptTable (const std::string & *trans_fasta_file*, double *alpha*, const FLD * *fld*, BiasBoss * *bias_table*, const MismatchTable * *mismatch_table*)

TranscriptTable Constructor

Parameters

trans_fasta_file	a string storing the path to the fasta file from which to load transcripts
alpha	a double that specifies the intial pseudo-counts for each bp of the transcripts (non-logged)
fld	a pointer to the global Fragment Length Distribution (FLD) object
bias_table	a pointer to the global BiasBoss object
mismatch_table	a pointer to the global MismatchTable object

Definition at line 142 of file transcripts.cpp.

3.15.2.2 TranscriptTable::~TranscriptTable ( )

TranscriptTable Destructor deletes all of the transcript objects in the table

Definition at line 199 of file transcripts.cpp.

3.15.3 Member Function Documentation

3.15.3.1 size_t TranscriptTable::covar_size ( ) const [inline]

a member function that returns the number of pairs of transcripts with non-zero covariance

Returns

the number of transcript pairs with non-zero covariance

Definition at line 368 of file transcripts.h.

3.15.3.2 double TranscriptTable::get_covar (TransID *trans1*, TransID *trans2*)

a member function that returns the covariance between two transcripts these returned value will be the log of the negative of the true value

Parameters

trans1	one of the transcripts in the pair
trans2	the other transcript in the pair

Returns

the log of the negative of the pair's covariance

Definition at line 248 of file transcripts.cpp.

3.15.3.3 Transcript * TranscriptTable::get_trans (TransID *id*)

a member function that returns a pointer to the transcript with the given id

Parameters

id	of the transcript queried

Returns

pointer to the transcript wit the given id

Definition at line 227 of file transcripts.cpp.

3.15.3.4 TransID TranscriptTable::get_trans_rep (TransID *trans*)

a member function that returns the bundle representative of the given transcript in the partitioning

Parameters

trans	the TransID of the transcript whose representative is requested

Returns

the TransID of the representative for the bundle the given transcript is in

Definition at line 262 of file transcripts.cpp.

3.15.3.5 TransID TranscriptTable::merge_bundles (TransID *rep1*, TransID *rep2*)

a member function that merges the bundles represented by the two given transcripts

Parameters

rep1	the TransID of the first bundle representative
rep2	the TransID of the second bundle representative

Returns

the TransID of the representative for the new merged bundle

Definition at line 268 of file transcripts.cpp.

3.15.3.6 size_t TranscriptTable::num_bundles ( )

a member function that returns the number of bundles in the partition

Returns

the number of bundles in the partition

Definition at line 277 of file transcripts.cpp.

3.15.3.7 void TranscriptTable::output_bundles (std::string *output_dir*)

a member function that outputs the bundles of the partition in a tab-delimited file called 'bundles.tab' in the given output directory each line contains a space-separated list of transcripts in a single bundle

Parameters

output_dir	the directory to output the bundle file to

Definition at line 288 of file transcripts.cpp.

3.15.3.8 void TranscriptTable::output_results (std::string *output_dir*, size_t *tot_counts*, bool *output_varcov*)

a member function that outputs the final expression data in a file called 'results.xprs' and (optionally) the variance-covariance matrix in 'varcov.xprs' in the given output directory

Parameters

output_dir	the directory to output the expression file to
tot_counts	the total number of observed mapped fragments
output_varcov	boolean specifying whether to also output the variance-covariance matrix

Definition at line 368 of file transcripts.cpp.

3.15.3.9 size_t TranscriptTable::size ( ) const [inline]

a member function that returns the number of transcripts in the table

Returns

number of transcripts in the table

Definition at line 344 of file transcripts.h.

3.15.3.10 void TranscriptTable::threaded_bias_update ( )

a member function for driving a thread that continuously updates the transcript bias values

Definition at line 308 of file transcripts.cpp.

3.15.3.11 void TranscriptTable::update_covar (TransID *trans1*, TransID *trans2*, double *covar*)

a member function that increases the covariance between two transcripts by the specified (logged) amount these values are stored positive even though they are negative

Parameters

trans1	one of the transcripts in the pair
trans2	the other transcript in the pair
covar	a double specifying the (logged) amount to increase the pair's covariance by

Definition at line 235 of file transcripts.cpp.

The documentation for this class was generated from the following files:

- src/transcripts.h
- src/transcripts.cpp

Index