

Isomorph
0.2

Generated by Doxygen 1.8.6

Tue Jun 16 2015 14:14:17

Contents

1	Main Page	1
2	Hierarchical Index	3
2.1	Class Hierarchy	3
3	Class Index	5
3.1	Class List	5
4	File Index	7
4.1	File List	7
5	Class Documentation	9
5.1	isomorph::CountEstimator Class Reference	9
5.1.1	Detailed Description	9
5.1.2	Member Function Documentation	9
5.1.2.1	estimate_abundances	9
5.2	isomorph::EMEstimator Class Reference	10
5.2.1	Detailed Description	10
5.2.2	Member Function Documentation	10
5.2.2.1	estimate_abundances	10
5.3	isomorph::Estimator Class Reference	11
5.3.1	Detailed Description	11
5.3.2	Member Function Documentation	11
5.3.2.1	estimate_abundances	11
5.4	isomorph::FastAData Struct Reference	11
5.4.1	Detailed Description	12
5.5	isomorph::FastQData Struct Reference	12
5.5.1	Detailed Description	12
5.6	isomorph::PairedRead Struct Reference	12
5.6.1	Detailed Description	13
5.6.2	Member Data Documentation	13
5.6.2.1	id	13
5.6.2.2	left_fa_id	13

5.6.2.3	left_phred	13
5.6.2.4	left_seq	13
5.6.2.5	pi_x_n	13
5.6.2.6	right_fa_id	13
5.6.2.7	right_phred	13
5.6.2.8	right_seq	14
5.7	isomorph::Read Struct Reference	14
5.7.1	Detailed Description	14
5.8	isomorph::Reader Class Reference	14
5.8.1	Detailed Description	14
5.8.2	Member Function Documentation	14
5.8.2.1	read_fasta	14
5.8.2.2	read_fastq	15
5.8.2.3	read_sam	15
5.9	isomorph::SamData Struct Reference	15
5.9.1	Detailed Description	15
5.10	isomorph::SingleRead Struct Reference	15
5.10.1	Detailed Description	16
5.10.2	Member Data Documentation	16
5.10.2.1	fa_id	16
5.10.2.2	id	16
5.10.2.3	phred	16
5.10.2.4	pi_x_n	16
5.10.2.5	seq	16
6	File Documentation	17
6.1	src/count_estimator.h File Reference	17
6.1.1	Detailed Description	17
6.2	src/em_estimator.cpp File Reference	17
6.2.1	Detailed Description	18
6.3	src/em_estimator.h File Reference	18
6.3.1	Detailed Description	18
6.4	src/estimator.h File Reference	19
6.4.1	Detailed Description	19
6.5	src/main.cpp File Reference	19
6.5.1	Detailed Description	20
6.6	src/paired_read.h File Reference	20
6.6.1	Detailed Description	20
6.7	src/read.h File Reference	21
6.7.1	Detailed Description	21

6.8	src/single_read.h File Reference	21
6.8.1	Detailed Description	21
6.9	src/utility.cpp File Reference	22
6.9.1	Detailed Description	22
6.10	src/utility.h File Reference	22
6.10.1	Detailed Description	23

Chapter 1

Main Page

Isomorph is a software for RNA transcript abundance estimation. It works with de novo assembled transcripts and supports different modes of operation.

At its core, it contains an EM algorithm that tries to optimize and find parameters of a robust statistical model proposed in the papers

"RNA-Seq gene expression estimation with read mapping uncertainty"

<http://bioinformatics.oxfordjournals.org/content/26/4/493.fullm>

and "RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome"

<http://www.biomedcentral.com/1471-2105/12/323>.

Isomorph, besides the above described statistical model,

also supports simple read count models described in the associated master thesis.

It requires seqan (<http://www.seqan.de/>), an open source C++ library for more efficient handling of input and output data.

Chapter 2

Hierarchical Index

2.1 Class Hierarchy

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

isomorph::Estimator	11
isomorph::CountEstimator	9
isomorph::EMEstimator	10
isomorph::FastAData	11
isomorph::FastQData	12
isomorph::Read	14
isomorph::PairedRead	12
isomorph::SingleRead	15
isomorph::Reader	14
isomorph::SamData	15

Chapter 3

Class Index

3.1 Class List

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

isomorph::CountEstimator	9
isomorph::EMEstimator	10
isomorph::Estimator	11
isomorph::FastAData	11
isomorph::FastQData	12
isomorph::PairedRead	12
isomorph::Read	14
isomorph::Reader	14
isomorph::SamData	15
isomorph::SingleRead	15

Chapter 4

File Index

4.1 File List

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

src/ count_estimator.h	
Declares the count estimator class	17
src/ em_estimator.cpp	
CountEstimator class methods are implemented here	17
src/ em_estimator.h	
Declares the em estimator class	18
inherited by other estimator classes	19
src/ main.cpp	
Parses input parameters and runs the appropriate estimator	19
src/ paired_read.h	
Declares paired_end read header	20
src/ read.h	
Declares a generic read header	21
src/ single_read.h	
Declares single_end read header	21
src/ utility.cpp	
Utility data and functions used by other classes are implemented here	22
src/ utility.h	
Utility data and functions used by other classes are declared here	22

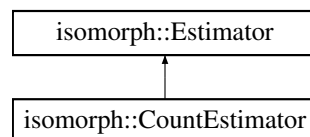
Chapter 5

Class Documentation

5.1 isomorph::CountEstimator Class Reference

```
#include <count_estimator.h>
```

Inheritance diagram for isomorph::CountEstimator:



Public Member Functions

- virtual void [estimate_abundances](#) (seqan::CharString reads, seqan::CharString transcripts, seqan::CharString pairs="")

5.1.1 Detailed Description

[CountEstimator](#) class.

This class estimates transcript abundances using very simple ideas of counting number of reads mapped to a transcript as well as the number of bases mapped to a certain isoform and reports the result.

5.1.2 Member Function Documentation

5.1.2.1 void isomorph::CountEstimator::estimate_abundances (seqan::CharString *reads*, seqan::CharString *transcripts*, seqan::CharString *pairs* = " ") [virtual]

Implementation of estimate_abundances virtual method.

The method estimates transcript abundances according to the strategy of this class.

Parameters

<i>reads</i>	path to the reads fastq file
<i>transcripts</i>	path to the reconstructed transcripts fasta file
<i>pairs</i>	path to the paired end file, if paired reads are used

Implements [isomorph::Estimator](#).

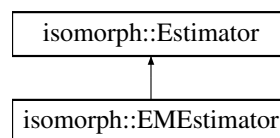
The documentation for this class was generated from the following files:

- [src/count_estimator.h](#)
- [src/count_estimator.cpp](#)

5.2 isomorph::EMEstimator Class Reference

```
#include <em_estimator.h>
```

Inheritance diagram for isomorph::EMEstimator:

**Public Member Functions**

- virtual void [estimate_abundances](#) (seqan::CharString left_pairs, seqan::CharString right_pairs, seqan::CharString transcripts)

5.2.1 Detailed Description

[EMEstimator](#) class.

This class estimates transcript abundances using a robust statistical model by applying the EM algorithm to obtain maximum likelihood estimates of the model parameters of the model.

5.2.2 Member Function Documentation

5.2.2.1 void isomorph::EMEstimator::estimate_abundances (seqan::CharString left_pairs, seqan::CharString right_pairs, seqan::CharString transcripts) [virtual]

Implementation of estimate_abundances virtual method.

The method estimates transcript abundances according to the strategy of this class.

Parameters

<i>reads</i>	path to the reads fastq file
<i>transcripts</i>	path to the reconstructed transcripts fasta file

<i>pairs</i>	path to the paired end file, if paired reads are used
--------------	---

Implements [isomorph::Estimator](#).

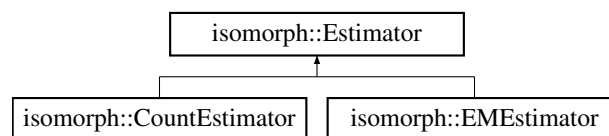
The documentation for this class was generated from the following files:

- [src/em_estimator.h](#)
- [src/em_estimator.cpp](#)

5.3 isomorph::Estimator Class Reference

```
#include <estimator.h>
```

Inheritance diagram for isomorph::Estimator:



Public Member Functions

- virtual void [estimate_abundances](#) (seqan::CharString reads, seqan::CharString transcripts, seqan::CharString pairs="")=0

5.3.1 Detailed Description

[Estimator](#) class.

Abstract base class all estimator classes must inherit.

5.3.2 Member Function Documentation

5.3.2.1 virtual void [isomorph::Estimator::estimate_abundances](#) (seqan::CharString *reads*, seqan::CharString *transcripts*, seqan::CharString *pairs* = " ") [pure virtual]

This method must be implemented in the classes inheriting this one.

Parameters

<i>reads</i>	path to the reads fastq file
<i>transcripts</i>	path to the transcripts fasta file
<i>pairs</i>	path to the paired end fastq file, if used

Implemented in [isomorph::CountEstimator](#), and [isomorph::EMEstimator](#).

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

- [src/estimator.h](#)

5.4 isomorph::FastAData Struct Reference

```
#include <utility.h>
```

Public Attributes

- `seqan::StringSet`
 < `seqan::CharString` > **ids**
- `seqan::StringSet`
 < `seqan::Dna5String` > **seqs**

5.4.1 Detailed Description

[FastADData](#) struct. Encapsulates seqan fasta file data classes for easier handling.

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

- [src/utility.h](#)

5.5 isomorph::FastQData Struct Reference

```
#include <utility.h>
```

Public Attributes

- `seqan::StringSet`
 < `seqan::CharString` > **ids**
- `seqan::StringSet`
 < `seqan::Dna5String` > **seqs**
- `seqan::StringSet`
 < `seqan::CharString` > **phred**

5.5.1 Detailed Description

[FastQData](#) struct. Encapsulates seqan fastq file data classes for easier handling.

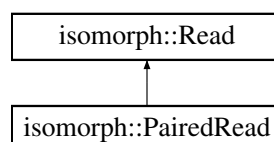
The documentation for this struct was generated from the following file:

- [src/utility.h](#)

5.6 isomorph::PairedRead Struct Reference

```
#include <paired_read.h>
```

Inheritance diagram for `isomorph::PairedRead`:



Public Attributes

- `int` **id**
- `seqan::CharString` **left_fa_id**

- seqan::Dna5String [left_seq](#)
- seqan::CharString [left_phred](#)
- seqan::CharString [right_fa_id](#)
- seqan::Dna5String [right_seq](#)
- seqan::CharString [right_phred](#)
- std::vector< std::tuple< int, int, int, seqan::CharString, seqan::CharString > > [pi_x_n](#)

5.6.1 Detailed Description

Paired read class.

All the data necessary for paired read analysis is placed here.

5.6.2 Member Data Documentation

5.6.2.1 int isomorph::PairedRead::id

[Read](#) id given to it by isomorph for easier indexing.

5.6.2.2 seqan::CharString isomorph::PairedRead::left_fa_id

The full read ID string from first fastq file.

5.6.2.3 seqan::CharString isomorph::PairedRead::left_phred

Corresponding phred quality symbols..

5.6.2.4 seqan::Dna5String isomorph::PairedRead::left_seq

Sequence of mate from the first fastq file.

5.6.2.5 std::vector<std::tuple<int, int, int, int, seqan::CharString, seqan::CharString> > isomorph::PairedRead::pi_x_n

The reduced space of transcripts this read maps well to together with mapped read sequence from SAM file.

5.6.2.6 seqan::CharString isomorph::PairedRead::right_fa_id

The full ID string of the second mate.

5.6.2.7 seqan::CharString isomorph::PairedRead::right_phred

Corresponding phred quality symbols.

5.6.2.8 seqan::Dna5String isomorph::PairedRead::right_seq

Sequence of the second mate.

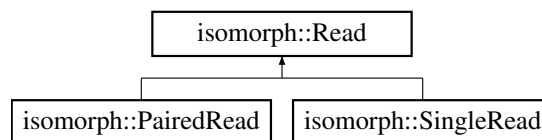
The documentation for this struct was generated from the following file:

- [src/paired_read.h](#)

5.7 isomorph::Read Struct Reference

```
#include <read.h>
```

Inheritance diagram for isomorph::Read:



5.7.1 Detailed Description

A generic read class.

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

- [src/read.h](#)

5.8 isomorph::Reader Class Reference

```
#include <utility.h>
```

Public Member Functions

- [int read_sam](#) (seqan::CharString filename, [SamData](#) *data)
- [int read_fasta](#) (seqan::CharString filename, [FastAData](#) *data)
- [int read_fastq](#) (seqan::CharString filename, [FastQData](#) *data)

5.8.1 Detailed Description

[Reader](#) class that encapsulates methods for data input and output.

5.8.2 Member Function Documentation

5.8.2.1 int isomorph::Reader::read_fasta (seqan::CharString filename, FastAData * data)

Reads a fasta file.

Parameters

<i>filename</i>	path to the file
<i>data</i>	pointer to fastadata struct that will be filled.

5.8.2.2 `int isomorph::Reader::read_fastq (seqan::CharString filename, FastQData * data)`

Reads a fastq file.

Parameters

<i>filename</i>	path to the file
<i>data</i>	pointer to fastqadat struct that will be filled.

5.8.2.3 `int isomorph::Reader::read_sam (seqan::CharString filename, SamData * data)`

Reads SAM alignment file.

Parameters

<i>filename</i>	path to the file
<i>data</i>	Pointer to samdata struct that will be filled.

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

- [src/utility.h](#)
- [src/utility.cpp](#)

5.9 isomorph::SamData Struct Reference

```
#include <utility.h>
```

Public Attributes

- seqan::BamHeader **header**
- std::vector
 < seqan::BamAlignmentRecord > **records**

5.9.1 Detailed Description

[SamData](#) struct. Encapsulates seqan fsam file data classes for easier handling.

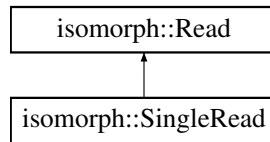
The documentation for this struct was generated from the following file:

- [src/utility.h](#)

5.10 isomorph::SingleRead Struct Reference

```
#include <single_read.h>
```

Inheritance diagram for isomorph::SingleRead:



Public Attributes

- int [id](#)
- seqan::CharString [fa_id](#)
- seqan::Dna5String [seq](#)
- seqan::CharString [phred](#)
- std::vector< std::tuple< int, int > > [pi_x_n](#)

5.10.1 Detailed Description

A single-end read class.

All the data necessary for single-read analysis is placed here.

5.10.2 Member Data Documentation

5.10.2.1 seqan::CharString isomorph::SingleRead::fa_id

Full read ID from the fastq file.

5.10.2.2 int isomorph::SingleRead::id

[Read](#) id given to it by isomorph for easier indexing.

5.10.2.3 seqan::CharString isomorph::SingleRead::phred

Associated phred quality symbols.

5.10.2.4 std::vector<std::tuple<int, int> > isomorph::SingleRead::pi_x_n

The reduced space of the transcripts this read maps well to.

5.10.2.5 seqan::Dna5String isomorph::SingleRead::seq

[Read](#) sequence from the fastq file.

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

- [src/single_read.h](#)

Chapter 6

File Documentation

6.1 src/count_estimator.h File Reference

Declares the count estimator class.

```
#include <vector>
#include <unordered_map>
#include "utility.h"
#include "estimator.h"
#include "single_read.h"
#include "paired_read.h"
```

Classes

- class [isomorph::CountEstimator](#)

6.1.1 Detailed Description

Declares the count estimator class.

Author

Pavlovic:Dario

Version

Revision 0.2

The count estimator header class is declared here.

Date

Tuesday, June 16, 2015

6.2 src/em_estimator.cpp File Reference

CountEstimator class methods are implemented here.

```
#include <ctype.h>
#include <cstdio>
#include <cstdlib>
#include <iostream>
#include <vector>
#include <unordered_set>
#include <fstream>
#include <seqan/stream.h>
#include <seqan/basic.h>
#include <seqan/file.h>
#include <seqan/sequence.h>
#include "em_estimator.h"
#include "utility.h"
#include "count_estimator.h"
#include "read.h"
#include "single_read.h"
#include "paired_read.h"
```

6.2.1 Detailed Description

CountEstimator class methods are implemented here. EMEstimator class methods are implemented here.

Author

Pavlovic:Dario

Version

Revision 0.2

Date

Tuesday, June 16, 2015

6.3 src/em_estimator.h File Reference

Declares the em estimator class.

```
#include <vector>
#include <unordered_map>
#include "utility.h"
#include "estimator.h"
#include "read.h"
#include "single_read.h"
#include "paired_read.h"
```

Classes

- class [isomorph::EMEstimator](#)

6.3.1 Detailed Description

Declares the em estimator class.

Author

Pavlovic:Dario

Version

Revision 0.2

The em estimator header class is declared here. This is the core header of isomorph.

Date

Tuesday, June 16, 2015

6.4 src/estimator.h File Reference

Declares a generic abstract estimator base class
inherited by other estimator classes.

```
#include <sean/stream.h>
#include <sean/basic.h>
```

Classes

- class [isomorph::Estimator](#)

6.4.1 Detailed Description

Declares a generic abstract estimator base class
inherited by other estimator classes.

Author

Pavlovic:Dario

Version

Revision 0.2

This is the abstract base class header that other estimators
must include and inherit to keep everything according to the strategy pattern.

Date

Tuesday, June 16, 2015

6.5 src/main.cpp File Reference

Parses input parameters and runs the appropriate estimator.

```
#include <iostream>
#include <seqan/arg_parse.h>
#include <seqan/stream.h>
#include <seqan/basic.h>
#include <seqan/file.h>
#include <seqan/sequence.h>
#include "count_estimator.h"
#include "utility.h"
#include "em_estimator.h"
```

Functions

- `int main (int argc, char **argv)`

6.5.1 Detailed Description

Parses input parameters and runs the appropriate estimator.

Author

Pavlovic:Dario

Version

Revision 0.2

This file starts the program, parses input flags by using seqan's `arg_parse` module and runs the selected estimator on the input.

Date

Tuesday, June 16, 2015

6.6 `src/paired_read.h` File Reference

Declares `paired_end` read header.

```
#include <seqan/seq_io.h>
#include <seqan/bam_io.h>
#include "read.h"
```

Classes

- struct [isomorph::PairedRead](#)

6.6.1 Detailed Description

Declares `paired_end` read header.

Author

Pavlovic:Dario

Version

Revision 0.2

Class modeling paired-end reads is declared here.

Date

Tuesday, June 16, 2015

6.7 src/read.h File Reference

Declares a generic read header.

Classes

- struct [isomorph::Read](#)

6.7.1 Detailed Description

Declares a generic read header.

Author

Pavlovic:Dario

Version

Revision 0.2

A generic class modeling a read is declared here.

Date

Tuesday, June 16, 2015

6.8 src/single_read.h File Reference

Declares single_end read header.

```
#include "read.h"
```

Classes

- struct [isomorph::SingleRead](#)

6.8.1 Detailed Description

Declares single_end read header.

Author

Pavlovic:Dario

Version

Revision 0.2

Class modeling single-end reads is declared here.

Date

Tuesday, June 16, 2015

6.9 src/utility.cpp File Reference

Utility data and functions used by other classes are implemented here.

```
#include <ctype.h>
#include <cstdio>
#include <cmath>
#include <iostream>
#include <string>
#include <vector>
#include <seqan/bam_io.h>
#include <seqan/seq_io.h>
#include <seqan/sequence.h>
#include "utility.h"
```

6.9.1 Detailed Description

Utility data and functions used by other classes are implemented here.

Author

Pavlovic:Dario

Version

Revision 0.2

Date

Tuesday, June 16, 2015

6.10 src/utility.h File Reference

Utility data and functions used by other classes are declared here.

```
#include <cmath>
#include <string>
#include <unordered_map>
#include <vector>
#include <seqan/seq_io.h>
#include <seqan/bam_io.h>
```

Classes

- struct [isomorph::FastQData](#)
- struct [isomorph::FastAData](#)
- struct [isomorph::SamData](#)
- class [isomorph::Reader](#)

Functions

- std::string **isomorph::execute_command** (const char *cmd)
- double **isomorph::prob_normal** (double mean, double stdev, double x)
- void **isomorph::estimate_insert_size** (const SamData &alignments, std::pair< double, double > ¶ms)
- void **isomorph::run_alignment** (const std::string &reads, const std::string &pairs, const std::string &transcripts, const std::string &output_dir, const bool paired_end)
- void **isomorph::print_sam_alignment_records** (const std::vector< seqan::BamAlignmentRecord > &records)
- char **isomorph::reverse_complement** (const char &c)

Variables

- const double **PI** = 3.14159265358979

6.10.1 Detailed Description

Utility data and functions used by other classes are declared here.

Author

Pavlovic:Dario

Version

Revision 0.2

Date

Tuesday, June 16, 2015