

## **3rd party documentation to MIRA 4**

---

**Things "The Definitive Guide" missed and other people  
kindly contributed**

---

This documentation is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

**COLLABORATORS**

	<i>TITLE :</i> 3rd party documentation to MIRA 4		
<i>ACTION</i>	<i>NAME</i>	<i>DATE</i>	<i>SIGNATURE</i>
WRITTEN BY		February 2, 2014	

**REVISION HISTORY**

NUMBER	DATE	DESCRIPTION	NAME

# Contents

<b>1</b>	<b>Scaffolding MIRA 454 contigs using 454 paired-end data and BAMBUS</b>	<b>1</b>
1.1	Installing BAMBUS	1
1.1.1	Installing AMOS	1
1.1.2	Installing BAMBUS	1
1.1.3	Troubleshooting	2
1.1.3.1	For Ubuntu 9 users	2
1.1.3.2	For SUSE 11.2 users	2
1.1.3.3	Problem with untangle.pl	2
1.2	Assemble with MIRA	3
1.2.1	Prepare paired-end data for assembly	3
1.2.2	Prepare shotgun data for assembly	3
1.2.3	Assemble	4
1.3	Scaffolding MIRA results with BAMBUS	4
1.3.1	Generate AMOS contigs	4
1.3.2	Setup BAMBUS configuration file	4
1.3.3	Setup mates file	5
1.3.4	Run BAMBUS	5

## Chapter 1

# Scaffolding MIRA 454 contigs using 454 paired-end data and BAMBUS

Gregory Harhay 2010Gregory Harhay

## 1.1 Installing BAMBUS

Note that you'll have to do this only once.

### 1.1.1 Installing AMOS

---

**Note** Download and use [amos-2.0.8](#), because the **ace2contig** script created by the installer is necessary to convert MIRA .ace file into a contig file that can be used by BAMBUS; later AMOS-short-tgz versions of AMOS don't create this script.

---

Follow directions for installing amos 2.0.8. I am installing on a 64-bit Redhat Enterprise Linux 5. The configure script complains that Qt doesn't work correctly and GUI apps won't run properly. I ignore this because I'm really only interested in the ace2contig script. For BAMBUS to run correctly, the BAMBUS team recommends you should pre-install:

- XML::Parser perl module from <http://www.cpan.org/>
- Config::IniFiles perl module from <http://www.cpan.org/>
- GraphViz package from <http://www.graphviz.org/>

Just to be safe I also pre-installed

- GraphViz perl module from <http://www.graphviz.org/>

### 1.1.2 Installing BAMBUS

Download bambus-2.33 and decompress. Edit Makefile file to set environment variables (within script) for BASEDIR & PERL.



#### Warning

There is a bug in the installation source scripts goBabmus.pl

bambus-2.33/src/goBabmus.pl line 22 \$ENV{PERLLIB} .=":\$BAMBUS\_BASE/lib"; should change PERLLIB to PERL5LIB

\$ENV{PERL5LIB} =":\$BAMBUS\_BASE/lib";

---

Follow the directions for BAMBUS installation, don't forget to edit the Makefile BASEDIR variable and to set the environmental variable PERL as directed in [http://sourceforge.net/apps/mediawiki/amos/index.php?title=Bambus\\_Manual](http://sourceforge.net/apps/mediawiki/amos/index.php?title=Bambus_Manual).

Run the test case as directed in the manual to make sure BAMBUS is installed properly.

### 1.1.3 Troubleshooting

#### 1.1.3.1 For Ubuntu 9 users

(anonymous user contribution, added 3/10)

There are problems making the grommit executable. They are remedied by the following procedure: in the bambus-2.33/src/TIGR\_Foundation\_CC directory apply the following changes

**ConfigFile.hh** add the line: `#include <string.h>`

**Logger.hh** add the line: `#include <stdlib.h>`

**OptionResult.hh** add the line: `#include <stdlib.h>`

**Options.hh** add the line: `#include <string.h>`

After changes, remake bambus from the the bambus-2.33 directory run

```
make clean
make all
make install
```

#### 1.1.3.2 For SUSE 11.2 users

(from Lionel Guy [guy.lionel@gmail.com](mailto:guy.lionel@gmail.com), added 7/27/10)

To be able to compile, had to further modify the following files in the bambus-2.33/src/TIGR\_Foundation\_CC directory:

**FileSystem.cc** in function `FileSystem::isCreatableFile` (line 58), changed

```
char * end_of_path = strrchr(filename, PATH_DELIMITER); to const char * end_of_p
ath = strrchr(filename, PATH_DELIMITER);
```

**Options.cc** add the line: `#include <cstdio>`

#### 1.1.3.3 Problem with untangle.pl

It has been reported that in some installations that the untangle script doesn't run properly and generates the following error:

```
Can't locate TIGR/Foundation.pm in @INC to locate Foundation.pm type:
```

```
locate Foundation.pm
```

the output of which will be the path to your Foundation.pm. In my case, it is

```
/usr/local/share/apps/bambus-2.33/lib/TIGR/Foundation.pm
```

Include this path at the beginning of the untangle.pl script as in

```
use lib "/usr/local/share/apps/bambus-2.33/lib/TIGR/Foundation.pm";
```

## 1.2 Assemble with MIRA

Before you begin you should have a fasta file of the paired-end (PE) adaptor sequence, the PE and shotgun SFF files, and know the mean and standard deviation of PE construct (get from lab, likely a sizing gel). Then follow the [instructions](#) of the MIRA manual for combined paired-end /shotgun MIRA assembly, of which a simplified version is summarized below.

### 1.2.1 Prepare paired-end data for assembly

Issue something the something like the following command:

```
sff_extract -l ../paired_end_adaptor.fasta F04Y2AI01.sff -i "insert_size:25000, ←  
insert_stdev:9000" -o non_0157
```

to create the following files:

```
-rw-rw-r-- 1 greg greg 125438250 Aug 24 16:59 non_0157.xml  
-rw-rw-r-- 1 greg greg 136943770 Aug 24 16:59 non_0157.fastq
```

To check if everything went well, type

```
>less non_0157.xml
```

to get something like:

```
<?xml version="1.0"?>  
<trace_volume>  
  <trace>  
    <trace_name>F04Y2AI01AIS9Z.fn</trace_name>  
    <insert_stdev>9000</insert_stdev>  
    <insert_size>25000</insert_size>  
    <clip_vector_left>5</clip_vector_left>  
    <clip_vector_right>58</clip_vector_right>  
  </trace>  
  <trace>  
    <trace_name>F04Y2AI01AJLJM.r</trace_name>  
    <insert_stdev>9000</insert_stdev>  
    <insert_size>25000</insert_size>  
    <clip_vector_right>2</clip_vector_right>  
    <template_id>F04Y2AI01AJLJM</template_id>  
    <trace_end>r</trace_end>  
  </trace>  
  <trace>  
    <trace_name>F04Y2AI01AJLJM.f</trace_name>  
    <insert_stdev>9000</insert_stdev>  
    <insert_size>25000</insert_size>  
    <clip_vector_right>130</clip_vector_right>  
    <template_id>F04Y2AI01AJLJM</template_id>  
    <trace_end>f</trace_end>  
  </trace>
```

Looks good, PE data in the XML file.

### 1.2.2 Prepare shotgun data for assembly

Now, work on shotgun SFF files and append (use -a) to PE data in fasta, qual, xml files:

```
sff_extract -Q -a FAPPU2401.sff FA79DFG01.sff E95K6EU01.sff -o non_0157
```

to get something like

```
-rw-rw-r-- 1 greg greg 230827474 Aug 24 17:06 non_0157.xml
-rw-rw-r-- 1 greg greg 322941190 Aug 24 17:06 non_0157.fastq
```

Notice how the file sizes have changed to accommodate the shotgun seqs? Now change file names to make them MIRA friendly:

```
>mv non_0157.xml non_0157_traceinfo_in.454.xml
>mv non_0157.fastq non_0157_in.454.fastq
```

### 1.2.3 Assemble

Fire off MIRA with something like this:

```
mira --project=non_0157 --job=denovo,genome,accurate,454 >&log_assembly &
```

which generates - amongst others - result files like these:

```
-rw-rw-r-- 1 greg greg 1612330844 Aug 25 04:49 non_0157_out.caf
-rw-rw-r-- 1 greg greg 19702684 Aug 25 04:49 non_0157_out.wig
-rw-rw-r-- 1 greg greg 20298482 Aug 25 04:49 non_0157_out.unpadded.fasta.qual
-rw-rw-r-- 1 greg greg 6820644 Aug 25 04:49 non_0157_out.unpadded.fasta
-rw-rw-r-- 1 greg greg 20342658 Aug 25 04:49 non_0157_out.padded.fasta.qual
-rw-rw-r-- 1 greg greg 6841175 Aug 25 04:49 non_0157_out.padded.fasta
-rw-rw-r-- 1 greg greg 708184237 Aug 25 04:50 non_0157_out.ace
```

You'll be using the .ace file with BAMBUS.

## 1.3 Scaffolding MIRA results with BAMBUS

### 1.3.1 Generate AMOS contigs

Convert MIRA .ace file to contig file for BAMBUS using AMOS's ace2contig:

```
ace2contig -i non_0157_out.ace -o non_0157_out.contig
```

to produce contig file non\_0157\_out.contig

### 1.3.2 Setup BAMBUS configuration file

Read the BAMBUS [configuration file help](#).

---

**Note** The mingroupsize switch/parameter doesn't appear to work. This switch is supposed to control the minimum size scaffold the application returns. Currently, it returns all scaffolds with lots of cruft. Bug report is in about this.

---

The configuration file is non\_0157\_out.conf, contents below:

```
# Priorities
priority ALL 1

# The following lines can be un-commented to specify certain
# per-library settings

# Redundancies
# redundancy lib_some 1
```

```
# allowed error
# error MUMmer 0.5

# min group size
mingroupsize 50000

# Global redundancy
redundancy 2

# overlaps allowed
# overlaps MUMmer Y
```

### 1.3.3 Setup mates file

Follow [directions](#) for generating mates file. Mates file is non\_0157\_out.mates (tab delimited!), contents below

```
library twentyfiveKB      9000      40000      (.....). *
pair      (.)\f$
```

### 1.3.4 Run BAMBUS

```
goBambus -c non_0157_out.contig -m non_0157_out.mates -C non_0157_out.conf -o ↩
non_0157_bambus
```

Once this has completed, BAMBUS will have produce a bunch of files looking similar to this:

```
-rw-rw-r-- 1 greg greg  91717540 Aug 26 18:18 non_0157_bambus.detective.xml
-rw-rw-r-- 1 greg greg 103560781 Aug 26 18:19 non_0157_bambus.evidence.xml
-rw-rw-r-- 1 greg greg      19 Aug 26 18:19 non_0157_bambus.lib
-rw-rw-r-- 1 greg greg  4036295 Aug 26 18:19 non_0157_bambus.inp
-rw-rw-r-- 1 greg greg    47 Aug 26 18:19 non_0157_bambus.grommit.conf
-rw-rw-r-- 1 greg greg   315 Aug 26 18:19 non_0157_bambus.default.conf
-rw-rw-r-- 1 greg greg  3570691 Aug 26 18:20 non_0157_bambus.out.xml
-rw-rw-r-- 1 greg greg   79516 Aug 26 18:20 non_0157_bambus.sum
-rw-rw-r-- 1 greg greg    633 Aug 26 18:20 non_0157_bambus.stats
-rw-rw-r-- 1 greg greg  197215 Aug 26 18:20 non_0157_bambus.oo
-rw-rw-r-- 1 greg greg  1217009 Aug 26 18:20 non_0157_bambus.dot
-rw-rw-r-- 1 greg greg  4213072 Aug 26 18:20 non_0157_bambus.details
-rw-rw-r-- 1 greg greg    3221 Aug 26 18:20 goBambus.log
```

If you want a fasta file produced of the scaffolds, read [http://sourceforge.net/apps/mediawiki/amos/index.php?title=Bambus\\_Manual#Getting\\_more\\_.28or\\_less.29\\_information\\_from\\_the\\_output](http://sourceforge.net/apps/mediawiki/amos/index.php?title=Bambus_Manual#Getting_more_.28or_less.29_information_from_the_output). A fasta file of the contig sequences (unpadded) generated by MIRA is required. This file is non\_0157\_out.unpadded.fasta, and is used in the command below:

```
printScaff
-e non_0157_bambus.evidence.xml
-s non_0157_bambus.out.xml
-l non_0157_bambus.lib
-f non_0157_out.unpadded.fasta
-merge -o non_0157_bambus_scaffold -page -dot
```

which gives you this

```
-rw-rw-r-- 1 greg greg      633 Aug 27 08:31 non_0157_bambus_scaffold.stats
-rw-rw-r-- 1 greg greg  6837159 Aug 27 08:31 non_0157_bambus_scaffold.fasta
-rw-rw-r-- 1 greg greg  1217042 Aug 27 08:31 non_0157_bambus_scaffold.dot
```

You can edit the fasta and dot files to keep the desired scaffolds and get rid of the cruft. In this case, 1 giant 5.4 MB scaffold was produced (this is what we want ) and hundreds sub-1 KB less desirable shorter scaffolds of little interest. You can use GraphVis's dot and dotty to visualize the scaffolding dot plots.