Instructions for scaffolding **MIRA** 454 contigs using 454 paired-end data and **BAMBUS**

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).

Prepare fasta, qual, xml files using paired-end SFF file

Follow the instructions for combined paired-end /shotgun MIRA 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 391429721 Aug 24 16:59 non_0157.fasta.qual
-rw-rw-r-- 1 greg greg 136943770 Aug 24 16:59 non_0157.fasta
```

To check if everything went well., type

```
> less non_0157.xml
```

to get :

```
<?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>
```

gregory.harhay@ars.usda.gov

```
    </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 XML file.

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

```
sff_extract -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 924673698 Aug 24 17:06 non_0157.fasta.qual
-rw-rw-r-- 1 greg greg 322941190 Aug 24 17:06 non_0157.fasta
```

Notice how the file sizes have changed to accommodate the shotgun seqs? Change names to make them mira friendly:

```
>mv non_0157.xml  non_0157_traceinfo_in.454.xml
>mv non_0157.fasta.qual non_0157_in.454.fasta.qual
>mv non_0157.fasta non_0157_in.454.fasta
```

Fire off mira:

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

or

```
mira --project=non_0157 --job=denovo,genome,accurate,454 -GE:not=6:kcim=yes
>&log_assembly &
```

Generates

```
-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

Install AMOS and BAMBUS

Install AMOS (<mark>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</mark>; later AMOS-shotr-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
Config::IniFiles perl module
GraphViz package

Just to be safe I also pre-installed

GraphViz perl module

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

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.

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
```

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

Generate scaffolds

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**

Setup bambus configuration file

Read the 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.

Configuration file is **non_O157_out.conf,** contents below

4

```
# 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
```

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$
```

Run BAMBUS

```
goBambus -c non_0157_out.contig -m non_0157_out.mates -C non_O157_out.conf -o
non_O157_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_O157_bambus.detective.xml
-rw-rw-r-- 1 greg greg 103560781 Aug 26 18:19 non_O157_bambus.evidence.xml
-rw-rw-r-- 1 greg greg        19 Aug 26 18:19 non_O157_bambus.lib
-rw-rw-r-- 1 greg greg   4036295 Aug 26 18:19 non_O157_bambus.inp
-rw-rw-r-- 1 greg greg        47 Aug 26 18:19 non_O157_bambus.grommit.conf
-rw-rw-r-- 1 greg greg       315 Aug 26 18:19 non_O157_bambus.default.conf
-rw-rw-r-- 1 greg greg   3570691 Aug 26 18:20 non_O157_bambus.out.xml
-rw-rw-r-- 1 greg greg     79516 Aug 26 18:20 non_O157_bambus.sum
-rw-rw-r-- 1 greg greg       633 Aug 26 18:20 non_O157_bambus.stats
-rw-rw-r-- 1 greg greg    197215 Aug 26 18:20 non_O157_bambus.oo
-rw-rw-r-- 1 greg greg   1217009 Aug 26 18:20 non_O157_bambus.dot
-rw-rw-r-- 1 greg greg   4213072 Aug 26 18:20 non_O157_bambus.details
-rw-rw-r-- 1 greg greg      3221 Aug 26 18:20 goBambus.log
```

If you wanta 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. 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_O157_bambus.evidence.xml -s non_O157_bambus.out.xml -l
non_O157_bambus.lib -f non_0157_out.unpadded.fasta -merge -o
non_O157_bambus_scaffold -page -dot

-rw-rw-r-- 1 greg greg       633 Aug 27 08:31 non_O157_bambus_scaffold.stats
-rw-rw-r-- 1 greg greg   6837159 Aug 27 08:31 non_O157_bambus_scaffold.fasta
-rw-rw-r-- 1 greg greg   1217042 Aug 27 08:31 non_O157_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.

untangle.pl (added 3/10)

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 Fountation.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";
```

Hope this helps:

Greg Harhay
Animal Health Research Unit

gregory.harhay@ars.usda.gov

USDA-ARS-Meat Animal Research Center
Clay Center, NE 6890

gregory.harhay@ars.usda.gov