TAMMiCol User Guide

Hayden Tronnolone[*1]**, Jennifer M. Gardner**[2]**, Joanna F. Sundstrom**[2]**, Vladimir Jiranek**[2]**,
Stephen G. Oliver**[3]**, and Benjamin J. Binder**[1]

[1]School of Mathematical Sciences, University of Adelaide, Adelaide, SA 5005, Australia
[2]Department of Wine and Food Science, University of Adelaide, Adelaide, SA 5005, Australia
[3]Cambridge Systems Biology Centre and Department of Biochemistry, University of Cambridge,
Cambridge CB2 1GA, United Kingdom

August 14, 2018

[*]Corresponding author: hayden.tronnolone@adelaide.edu.au

Contents

1   Introduction

Tool for Analysis of the Morphology of Microbial Colonies (TAMMiCol) is an application for efficiently and automatically analysing images of microbial colonies. The software converts colony images to binary, which separates the colony from the background and any other extraneous features. The data may be analysed by exporting the binary images or by using the in-built statistics.

This guide provides instructions for using TAMMiCol, including preparing and managing saved images, converting these to binary, and computing statistics. The guide is divided into five broad sections: Installation, Input, Basic Processing, Advanced Settings, and Output. Most users will only require Basic Processing, while Advanced Settings allows users to exert more control over the method, output, and statistics computed.

TAMMiCol is provided free under the GNU General Public License v3.0. Source code and binaries are available from github.com/HaydenTronnolone/TAMMiCol. Bug reports may be submitted to hayden.tronnolone@adelaide.edu.au.

A full explanation of the method used to threshold the images and produce the statistics is given by Tronnolone, Gardner, Sundstrom, Jiranek, Oliver and Binder (2018).

2 Installation

2.1 Basic Installation

TAMMiCol is written in MATLAB using the App Designer package; however, this software does not require a MATLAB license in order to operate. To install TAMMiCol, open the installation file and follow the prompts. The installation will download and install MATLAB Runtime, which contains necessary shared libraries. While an internet connection is required during the installation, TAMMiCol will run offline following installation.

Separate installation files are provided for systems running macOS and Windows; however, the installation process is similar for both. Users running macOS 10.12 Sierra should see the following subsection for specific steps.

2.2 Installation on macOS 10.12 Sierra

On macOS 10.12 Sierra, Gatekeeper may prevent TAMMiCol from installing correctly. Gatekeeper may be disabled through the Terminal using the command `sudo spctl --master-disable`. The installation file will also need to be placed in the installation location.

3   Input

While not required, it is recommended that images are saved using a consistent format that provides meaningful information about the experimental conditions. TAMMiCol is able to interpret four pieces of information in each filename: the dataset name, sample number, observation time, and image magnification. The dataset name should uniquely identify the dataset and could include information such as the strain and nutrient level. The sample number is expected to be a positive integer, but these numbers need not be sequential. The time may be given as the number of hours or number of days, but all times will be converted to hours when images or data are exported.

In order for this information to be interpreted correctly, filenames are expected to have the form:

[name] s[sample] [time]h [magnification]X.[extension],

where quantities in brackets represent the actual values. For example, a tif image of unspecified magnification from sample 5 that is part of the dataset named AWRI 796 observed after 233 hours would have the filename

AWRI 796 s5 233h.tif.

If the observation time is measured in days, this should have the form d[day] rather than [time]h. For example, an image taken after 3 days of growth would use d3. The images may be stored in a single directory and do not need to be sorted; however, they may be kept in subdirectories if so desired.

When saving binary images or statistics, TAMMiCol will use the information contained in the filename to sort the output. Images will be named as per the format given above, while any tables of statistics produced will also contain the name, sample number, time, and magnification. Missing values are replaced by 'NaN'. In addition, saved images will be sorted into directories based on this information. Each dataset will be allocated a new directory, named '[name] Binary', which will contain directories for each sample, named 'Sample [sample]'. All observations of a given sample are saved in the same directory.

The sample number, time, and magnification may be omitted from the filename. When this occurs, any directories and files created with have 'Nan' in place of the missing values.
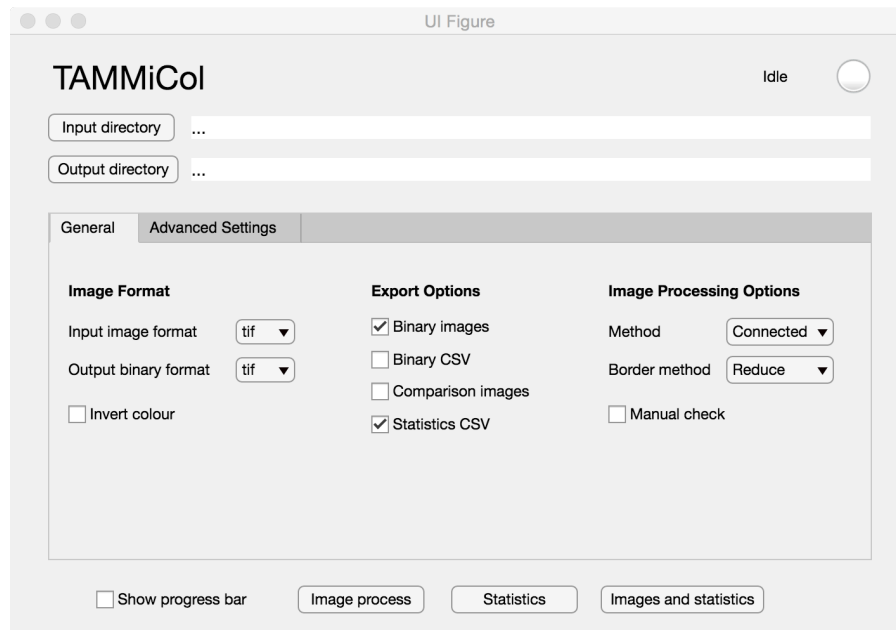
Figure 1: The general tab open in TAMMiCol.

4 Basic Processing

4.1 Overview

TAMMiCol opens with the general tab displayed, as shown in Figure 1. The general tab has three sections: Image Format, Export Options and Image Processing Options. After processing, all settings used for processing are saved as a CSV file in the same location as the saved data.

4.2 Input and Output

The input and output fields are viewable on all tabs. These fields control where the raw images are located, and where output from TAMMiCol will be saved.

Input   Specify the input directory, which contains the images, or subdirectories of images, to be processed.

Output   Specify the output location. If this is not set, it will be taken to be the same location as the input directory.

4.3 Image Format

The user must specify the format of the images to be analysed. Only one format may be chosen for each batch processing job. Any files in the input directory or subdirectories that do not

match the specified input format will be ignored. This means that other file types in these directories, such as files containing experimental details, need not be removed before processing.

Image Format   Specify the format of the images to be processed.

Output binary format   Specify the format for any binary images produced.

Invert colour   TAMMiCol expects the colony to be darker than the background. If this is not the case, then the colour can be inverted by selecting 'Invert colour'.

4.4 Export Options

There are four options under 'Export Options' that may each be selected independently. All of the output selected will be saved in the directory structure created by TAMMiCol.

Binary images   Save binary images of the colonies in the selected image format.

Binary CSV   Save CSV files of the binary image data.

Comparison images   Produce copies of the original images with the colony pixels identified by TAMMiCol coloured green.

Statistics CSV   Save a CSV file containing the in-built statistics, provided these are computed. The statistics are described in subsection 6.1.

4.5 Image Processing Options

'Image Processing Options' control how images are converted to binary. The default options are best for unwashed colonies. Washed colonies comprising unconnected pieces generally require the 'Unconnected' method with the 'Reduce' border pixels method and manual checks on the threshold level.

Method   Specify whether the colony is one connected region ('Connected') or multiple unconnected regions ('Unconnected').

Border Method   Specify how pixels on the border of the binary image should be handled. By default, the threshold is reduced until no border pixels are selected, referred to as the 'Reduce' method. This may be changed to 'Remove', in which any connected component that includes a border pixel is removed from the binary image. This may be done if there is no chance that a piece of the colony lies near the boundary.

Manual check   The comparison image of each colony is displayed, as illustrated in Figure 2, and the user has the option to adjust the threshold level.
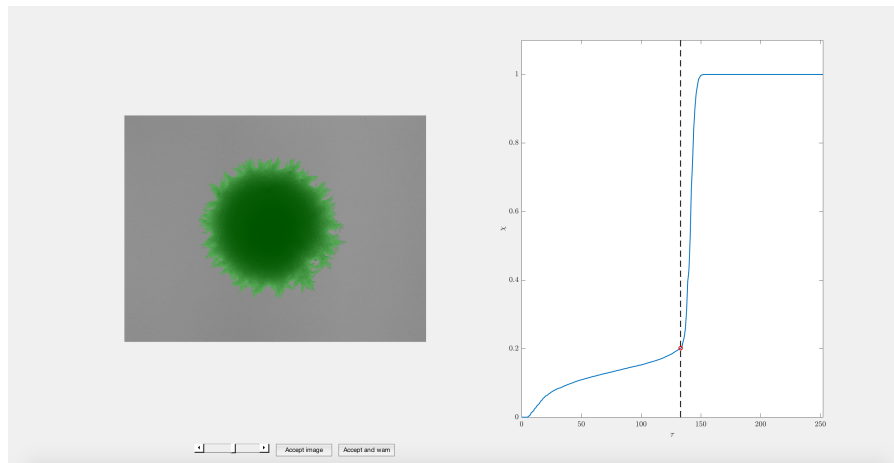
Figure 2: The manual control window.

4.6 Processing Options

At the bottom of the window are options for processing.

Image process	Produces binary versions of the images. No statistics are produced using the processed images. The input data should be raw images.
Statistics	Computes the statistics only. The input data should be binary images. If you wish to compute the statistics for raw images, you must select 'Images and statistics' instead.
Images and statistics	Produces binary images and computes the associated statistics. The input data should be a directory of raw images.
Show progress bar	Shows a graphical display of the proportion of all images processed.

4.7 Warnings

TAMMiCol has a number of checks to ensure the image processing is performed as expected. If there are errors during this process, a warning containing the image metadata and a description of the problem is added to the log. A description of each warning is given below.

Critical index not found	The critical tolerance at which the entire image is selected could not be identified. This value is set to be the maximum tolerance. The original image may be corrupted and the binary image should be reviewed.

|	
Threshold index not found	The error in the linear fit for the proportions could not be minimised and no threshold could be identified. The tolerance is set to zero. It is likely that parts of the colony will be missed in the binary image.
Reduced to maximum threshold N	The selected tolerance produced an image proportion above the maximum N allowed. The tolerance is reduced in order to give a binary image with the maximum permitted proportion. It is possible that the colony extended outside of the original image. The binary image should be reviewed.
Image flagged by user	The user flagged the image as problematic during the manual check. The binary image should be reviewed.
Border pixels identified	Selected pixels were detected on the border of the image and an attempt was made to remove them. If parts of the colony were touching the boundary in the original image, these may have been removed.
Removing N of M isolated pixels (P %)	Isolated pixels were detected in the image and were removed. The number N of isolated pixels and the total number M of pixels are given, along with the percentage of isolated pixels removed P.
Border pixels selected	Selected pixels were detected on the border of the image that were unable to be removed. This may be because the colony extends outside the original image or there is some contamination in the image that could not be removed. The binary image should be reviewed.
Image threshold below N	The selected tolerance was below the selected level N. The image may not have been processed correctly and the binary image should be reviewed.
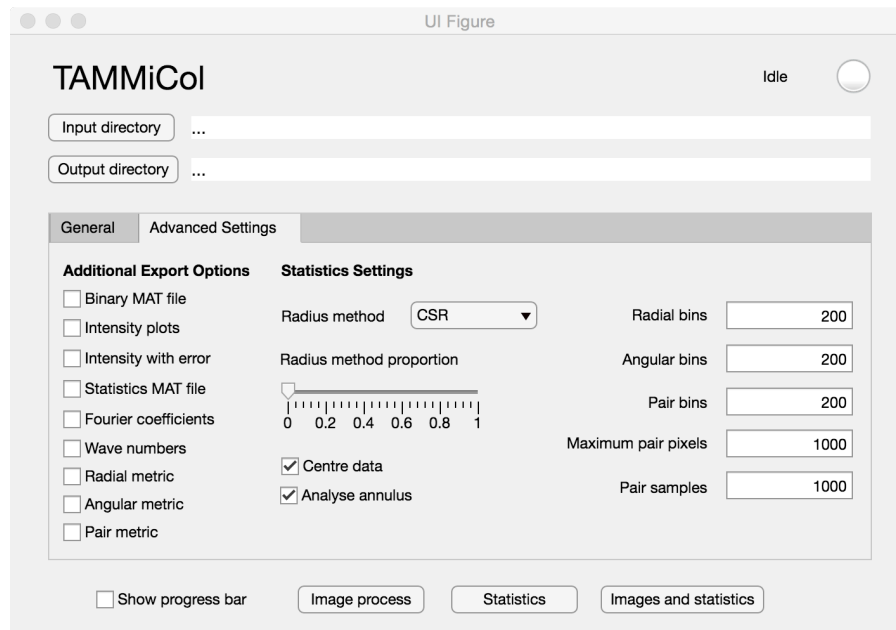
Figure 3: The advanced tab open in TAMMiCol.

5 Advanced Options

5.1 Advanced Options

The 'Advanced Options' tab, shown in Figure 3, contains a number of extra settings that provide additional output or adjust the methods used. These are not needed for basic image processing or for computing the default statistics.

5.2 Additional Export Options

These options save additional output that may be used to examine the image processing step or to export the data in a different format. A full explanation of the metrics is given by Tronnolone *et al.* (2018).

Binary MAT file   Produces a MAT file for use in Matlab containing the binary image data. The variables produced are described in subsection 6.6.

Intensity plots   Produces EPS and Matlab fig files showing the pixel proportion and selected tolerance for each image.

Intensity with error   Produces EPS and Matlab fig files of the intensity plot with the error from the linear fit added.

Statistics MAT file   Produces a MAT file for use in Matlab containing the statistics. The variables produced are described in subsection 6.6.

Fourier coefficients	Produces a txt file containing the coefficients from the Fourier transform of the pair correlation function.
Wavenumbers	Produces a txt file containing the wavenumbers ordered by the absolute value of the corresponding Fourier coefficient.
Radial metric	Produces a txt file containing the radial metrics.
Angular metric	Produces a txt file containing the angular metrics.
Pair metric	Produces a txt file containing the pair metrics.

5.3   Statistics Settings

The remaining settings control the statistics computed. The different radius methods are illustrated in Figure 4.

|	
Radius method	The data can be filtered using either the complete-spatial-randomness radius or by removing the largest connected disk.
Radius proportion	Use some proportion of the selected radius method to compute the angular and pair quantities. Setting the radius proportion to zero means the entire colony is analysed, while a setting of one means the entire minimum radius is used to filter the data.
Centre data	By default, the statistics are computed using the centroid of the selected pixels. Uncheck this option to instead use the centre of the image.
Analyse annulus	Use either the entire colony or the selected annulus.
Radial bins	The number of bins used to group the data for the radial count.
Angular bins	The number of bins used to group the data for the angular count.
Pair bins	The number of bins used to group the data for the pair correlation count.
Pair samples	The number of samples averaged to compute the pair correlation function.
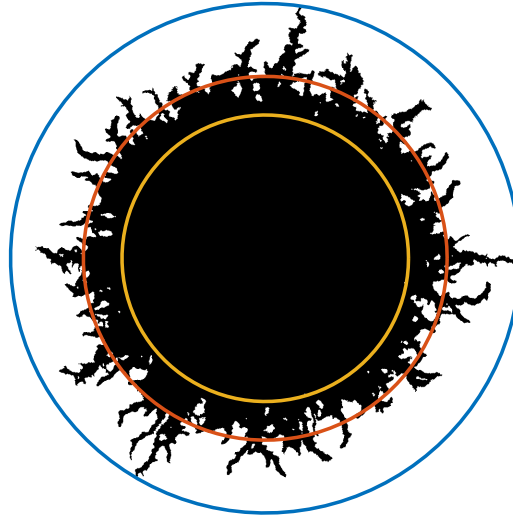
Figure 4: The radii used to analyse the data represented by circles drawn over a typical colony (black). The maximum radius $R_{\max}$ (blue) represents the largest radius measured from the colony centre. The minimum radius $R_{\min}$ is measured in two ways. The punch radius (yellow) represents the smallest circle with the same centre as the colony that contains only occupied cells. The CSR radius (red) represents the distance at which the chance of finding an occupied pixel is equal to the corresponding probability from a random distribution of pixels.

6 Output

6.1 Statistics

The statistics are saved in a text file called '[name] Statistics.txt', which contains the statistics computed from each image, as described in the following table. These are also saved in the MATLAB data file '[name] Statistics.mat', described in subsection 6.6.

6.2 Fourier Coefficients

The Fourier coefficients for the pair metric for each image are saved in a text file called '[name] Fourier Coefficients.txt'. The columns contain the dataset, sample, and time, with the nth coefficient stored in the column labelled 'Fn'. This variable is labelled `ffTheta` in the MATLAB data file.

6.3 Wavenumbers

The Wavenumbers ordered by the magnitude of the associated Fourier coefficient corresponding to the Fourier transform of the pair metric for each image are saved in a text file called '[name] Wavenumbers.txt'. The columns contain the dataset, sample and time, with the nth wavenumber in the column labelled 'Wavenumber n'. This variable is labelled `k` in the MATLAB data file.

Heading	Symbol	Description
Dataset	–	The name of the dataset
Sample	–	Sample number
Time	–	Observation time in hours
Magnification	–	Image magnification
Ir	I_r	A value between 0 and 1 that measures variation in the radial direction. Larger values indicate greater variation.
Itheta	I_θ	A value between 0 and 1 that measures variation in the angular direction using the angular metric. Larger values indicate greater variation.
ITheta	I_Θ	A value between 0 and 1 that measures local aggregation in the angular direction. Larger values indicate greater aggregation.
Icsr	I_{CSR}	A value between 0 and 1 that measures angular variation using the pair correlation metric. Larger values indicate greater variation.
Rmin	$R_{\min}$	The distance from the centre to the inner radius used to filter the data. This value is reported even if this option is not selected.
Rmax	$R_{\max}$	The radius of the colony in pixels as measured from the centre to the most distant point.
Area	A	The area of the colony in pixels.

6.4   Radial, Angular and Pair Metrics

The files '[name] Radial Metric.txt', '[name] Angular Metric.txt' and '[name] Pair Metric.txt' contain the scaled bin counts from each of the metrics. The columns contain the dataset, sample and time, with the nth bin in the column 'Bin n'. For radial, angular and pair metrics with bin counts n_r. n_θ and n_Θ, respectively, the corresponding bin widths are $R_{\max}/n_r$, $2\pi/n_\theta$ and π/n_Θ. In MATLAB, the radial, angular and pair metrics are labelled `fr`, `ftheta` and `fTheta`, respectively.

6.5   Settings

The file '[name] Settings.txt' contains the input settings used in TAMMiCol to analyse the data. This file is created automatically.

6.6   MATLAB Variables

All statistics variables in the MATLAB output are saved as arrays in which the first dimension represents the time and the second the sample. For example, the value of the radial index I_r at the second observation time and fifth sample, which is stored in the variable `Ir`, would be accessed by `Ir(2,5)`. For variables with more than one value at each observation time and

sample, such as the metrics, the values are stored in the third dimension. For example, the radial metric at the second observation time and fifth sample, which is stored in `fr`, would be accessed by `fr(2,5,:)`.

Binary image data stored in MATLAB is also saved as an array, with the y-data stored along the first dimension and the x-data along the second.

6.7 File Names for Multiple Datasets

When processing multiple datasets at the same time, the output filenames will start with 'Multiple Datasets [timestamp]', where [timestamp] gives the date and time of processing in the format 'YYYYMMDDhhmm'.
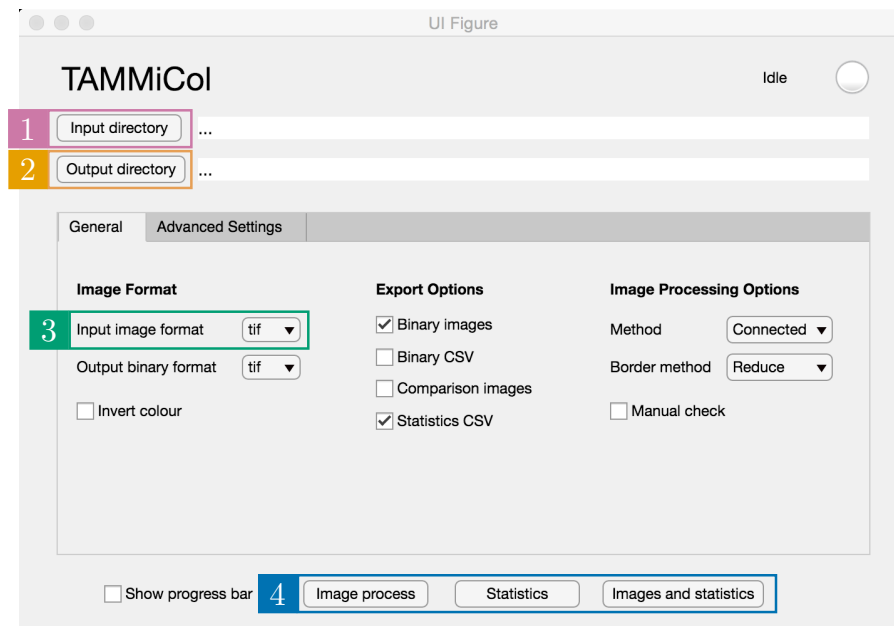
Figure 5: The basic steps for using TAMMiCol.

7 Basic Example

The following steps describe how to undertake a basic analysis of images using TAMMiCol. These steps are illustrated in figure 5

1. Select the directory containing the images to be processed.

2. Select the desired output location for the processed images.

3. Select the format of the input images.

4. For binary images only, click 'Image process'; for binary images and statistics, click 'Images and statistics'. If the input images are already binary, statistics may be generated by selecting 'Statistics'.

8   Version History

Number	Release Date	Description
1.0	Unreleased	Basic functionality
2.0	Unreleased	Major change to GUI.
2.1		CSV and MAT file binary export options added. Warnings now exported.

References

Tronnolone, H., Gardner, J. M., Sundstrom, J. F., Jiranek, V., Oliver, S. G. and Binder, B. J. (2018), TAMMiCol: Tool for analysis of the morphology of microbial colonies. In preparation.