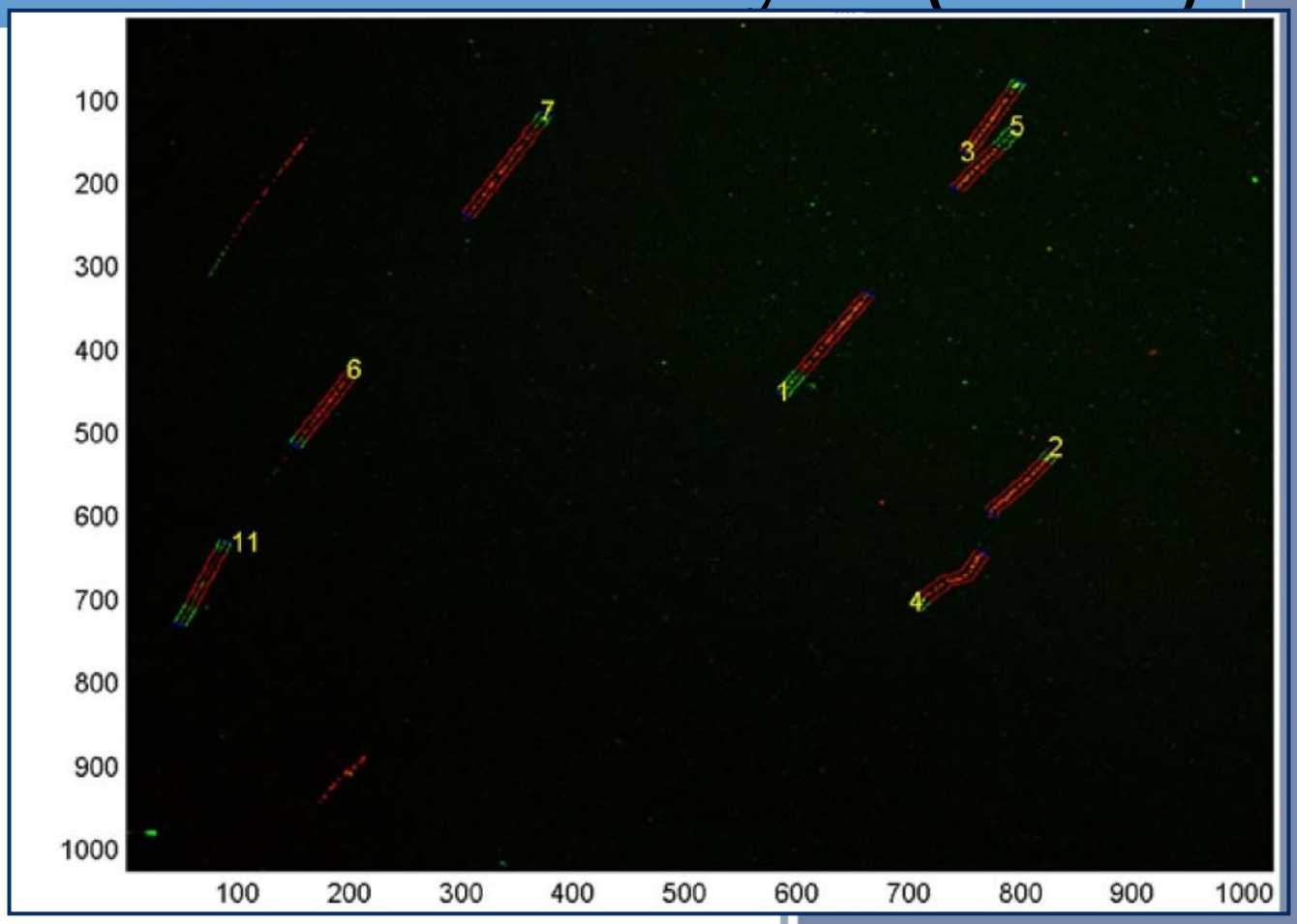
2017

Computer Assisted Scoring & Analysis (CASA)



Paul Chastain

DNA Fiber Analysis

4/17/2017

Contents

Table of Figures

Step 1: Installation

`

A. Double-Click Installation Package

Figure 1: Image of the installer program

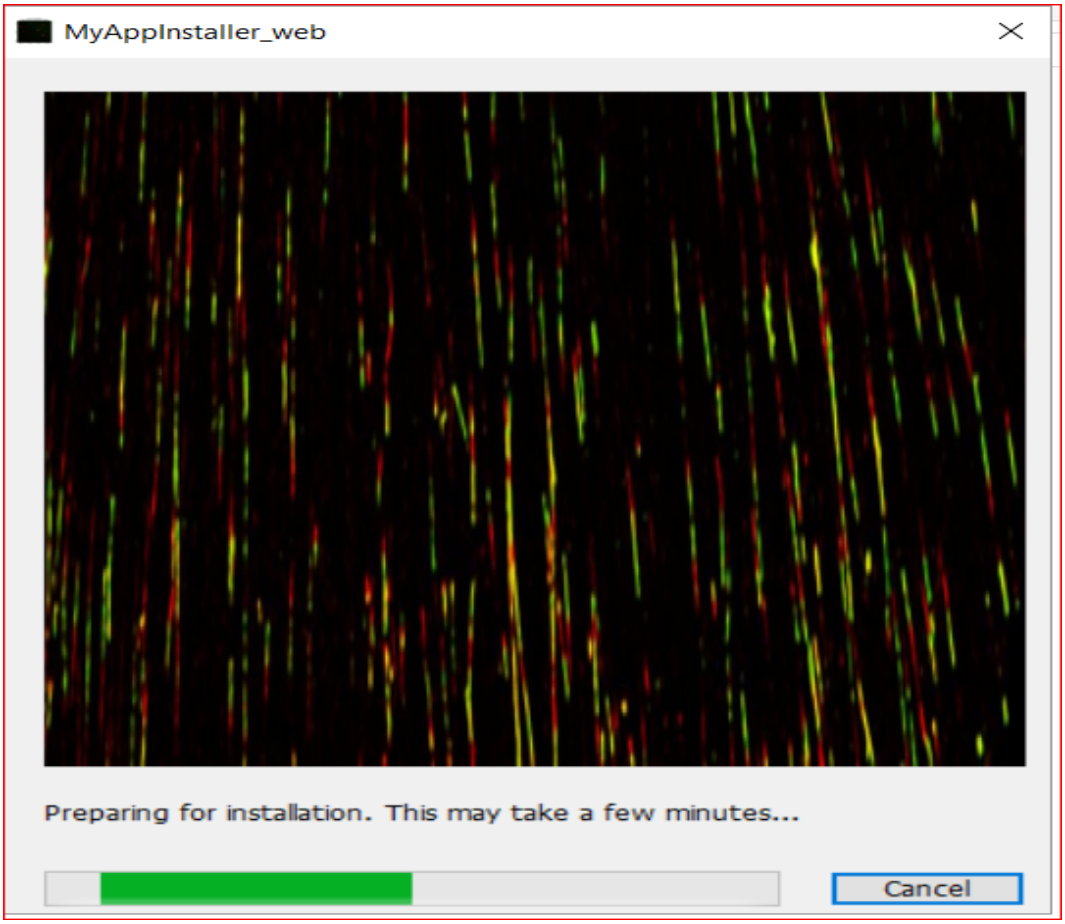


Figure 2 Preparation for installation splash screen

B. During the Installation Process

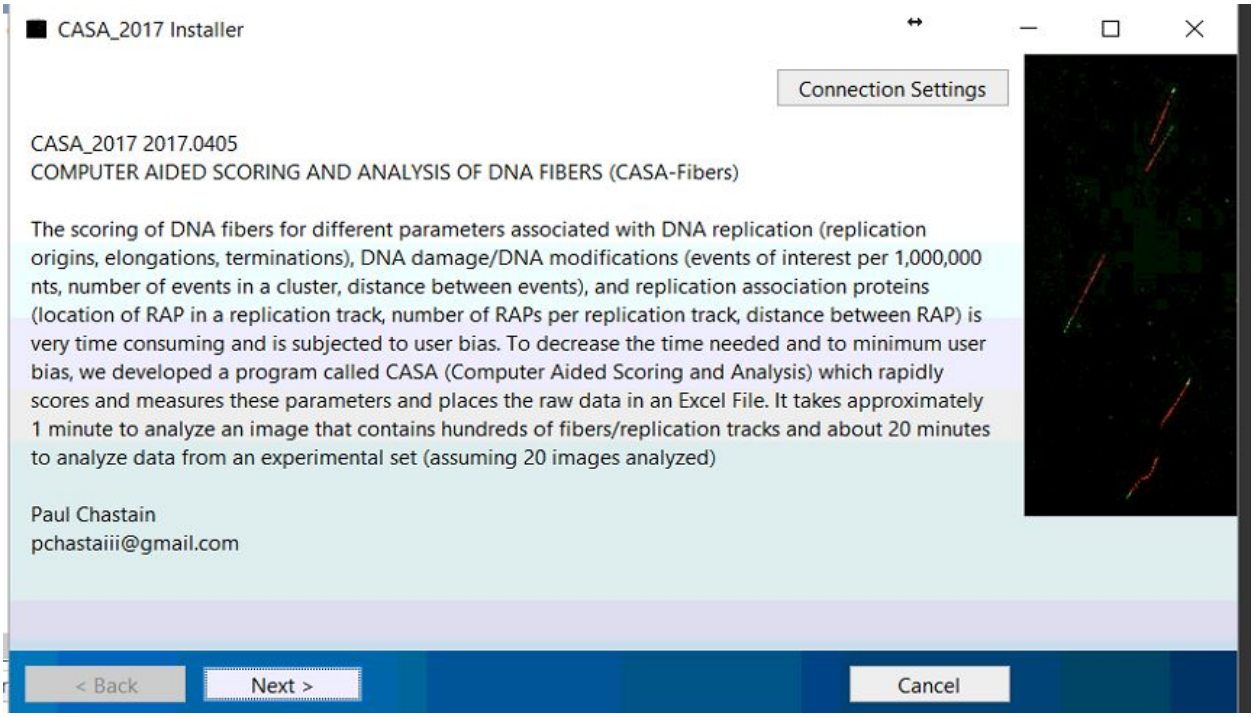


Figure 3: Installation Begins

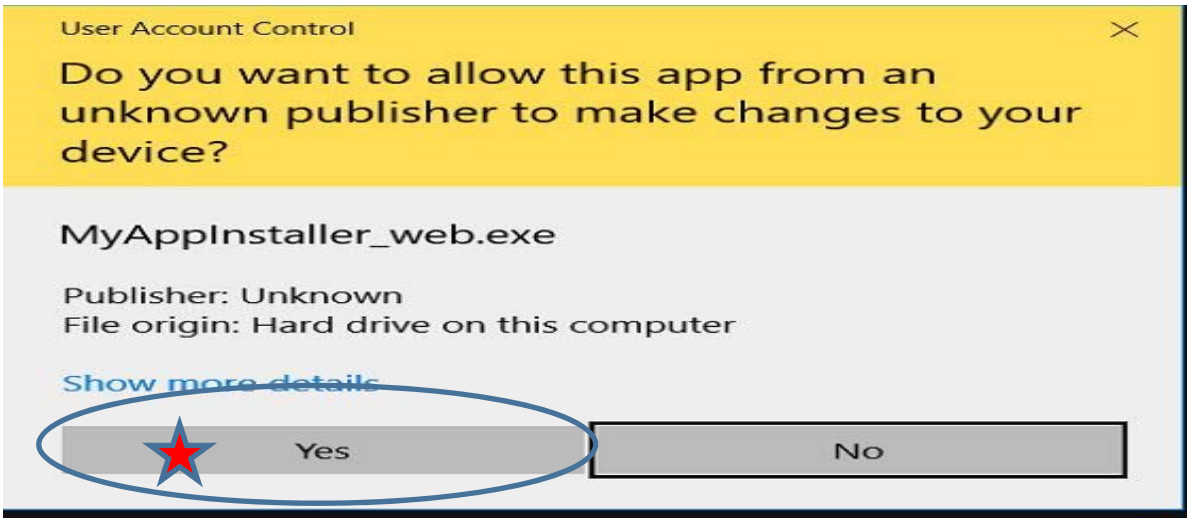


Figure 4: Windows 10 warning for installation of our program. Just click on yes.

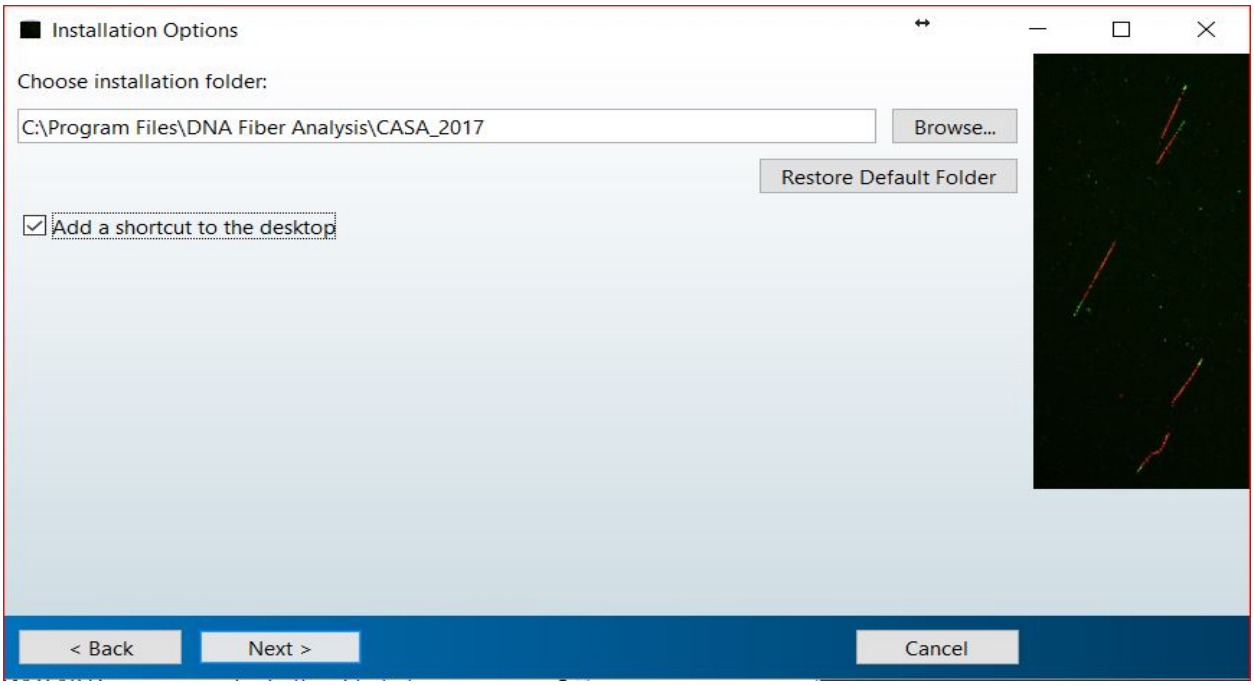


Figure 5: Installation Options - ***Note clicking on the box next to Add a shortcut to the desktop will add shortcut to the desktop. We suggest you click to add shortcut.*** *You can designate the location of the program (this is where the program lives as well as the excel file, license, and how to install and run the program*

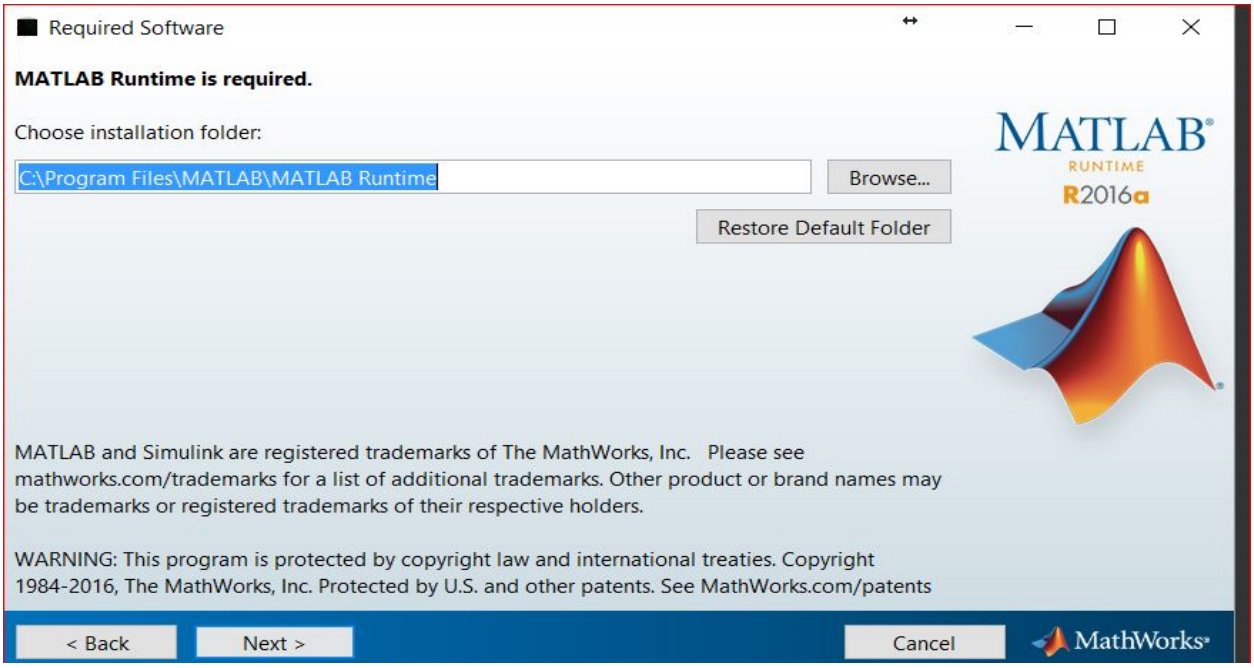


Figure 6: MATLAB (the core image analysis engine) needs to install its runtime/compiler. It is about 216 MB. If you have trouble installing this (or if you do not have access to the internet), I can send you an installation package that has the MATLAB Runtime embedded.

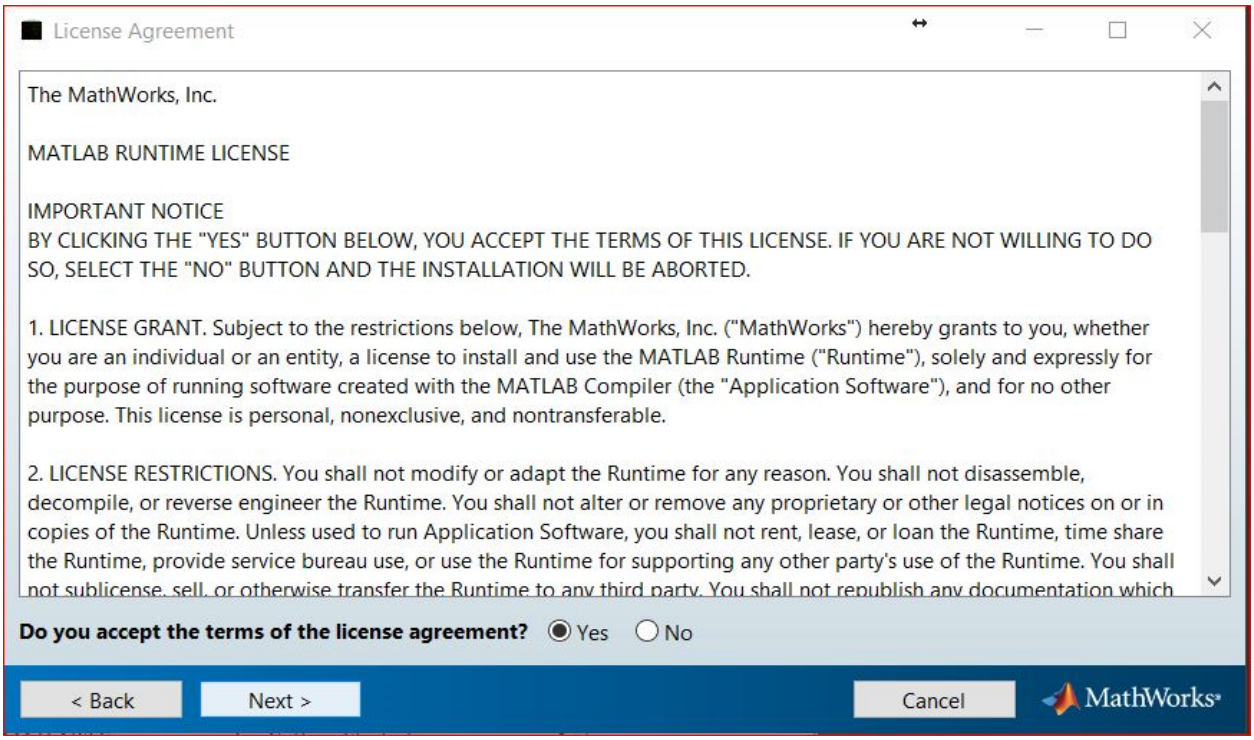


Figure 7: License Agreement. You need to agree to this license agreement otherwise you cannot install the software. Also, CASA program has its own license agreement. By clicking on yes, you agree to both.

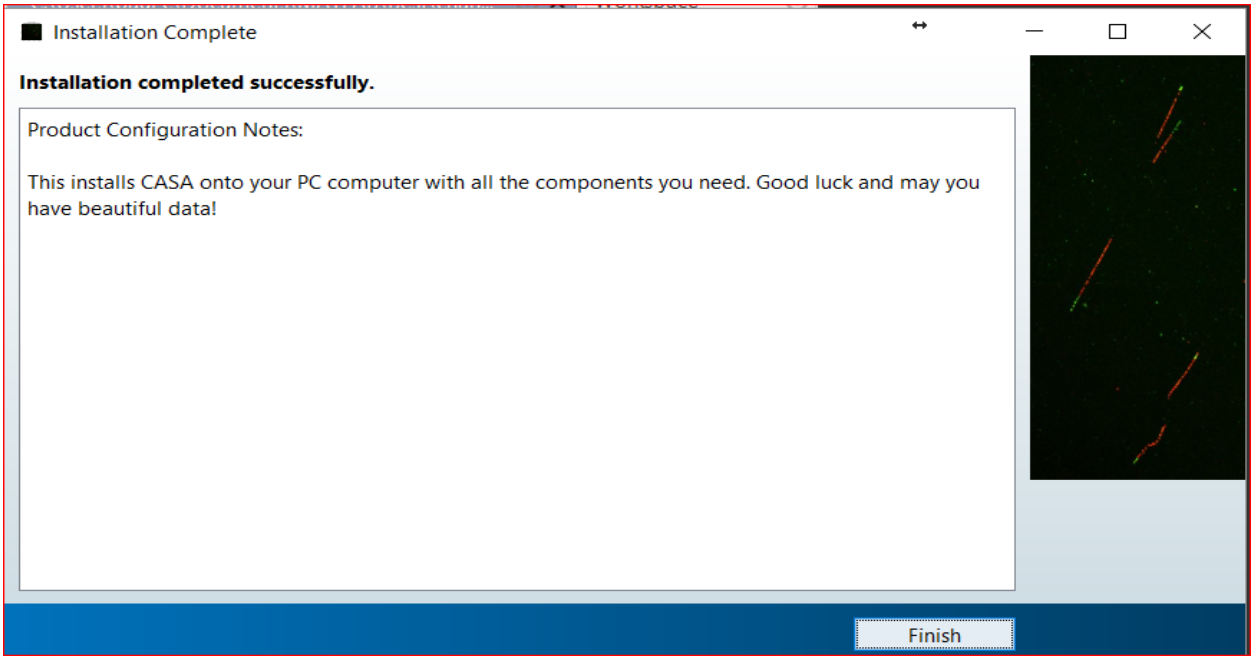


Figure 8: Installation is complete.

C. After Installation

1. ***A shortcut should be located on your desktop (if you pressed the checkbox during the installation – See Figure 5)***

Figure 9: Image of the CASA icon on the desktop.

2. ***Excel Files, License, and Program are now located on your computer.***

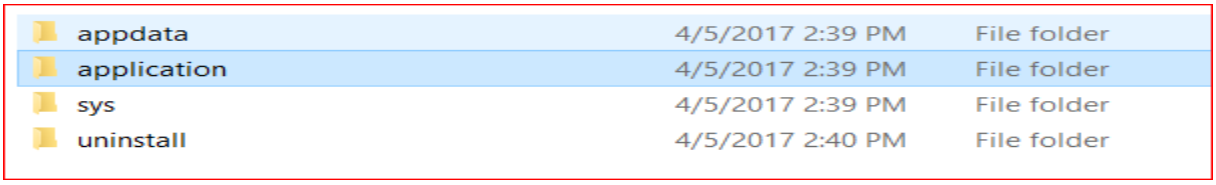


Figure 10: Files and folders that were placed onto your computer

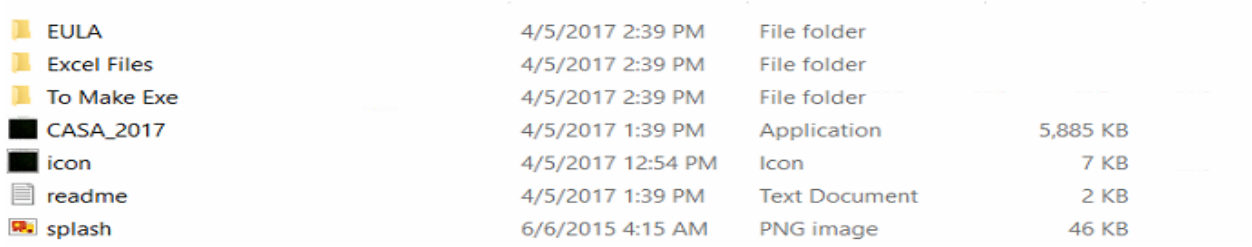


Figure 11: Folders and Programs in Application Folder.

Location of the Programs/Folders

CASA Program

- Program Files-> DNA Fiber Analysis -> CASA_2017->Application

Excel File Template

- Where all the data from the images will be placed
- DNA Fiber Analysis -> CASA_2017->Application ->Excel Files

Matlab compiler/runtime

- Program Files-> MATLAB -> MATLAB Runtime

Uninstaller

- Program Files-> DNA Fiber Analysis -> CASA_2017

Step 2: To Run Program

A. Press Icon

Figure 12: Icon for CASA Program.

- It takes a few minutes for the program to start (it has to acquire enough memory to run)

Figure 13: Log file is created to record program activities.

- As it opens, the program will place a log file in the location where the CASA program is located. This log file records all the activities performed by the program. The name given to the log file is CASA_info_Date, where Date refers to the date the program is run.
    - If you run the program multiple times during the day, it will overwrite the log file (since they all have the same date). Therefore, it is recommended that you rename this log file prior running the program again (i.e., for the example above, I would change CASA_Info_05-Apr-2017 to CASA_Info_05-Apr-2017_Run1. I would rename run 2 to be CASA_Info_05-Apr-2017_Run).

B. Select Excel File

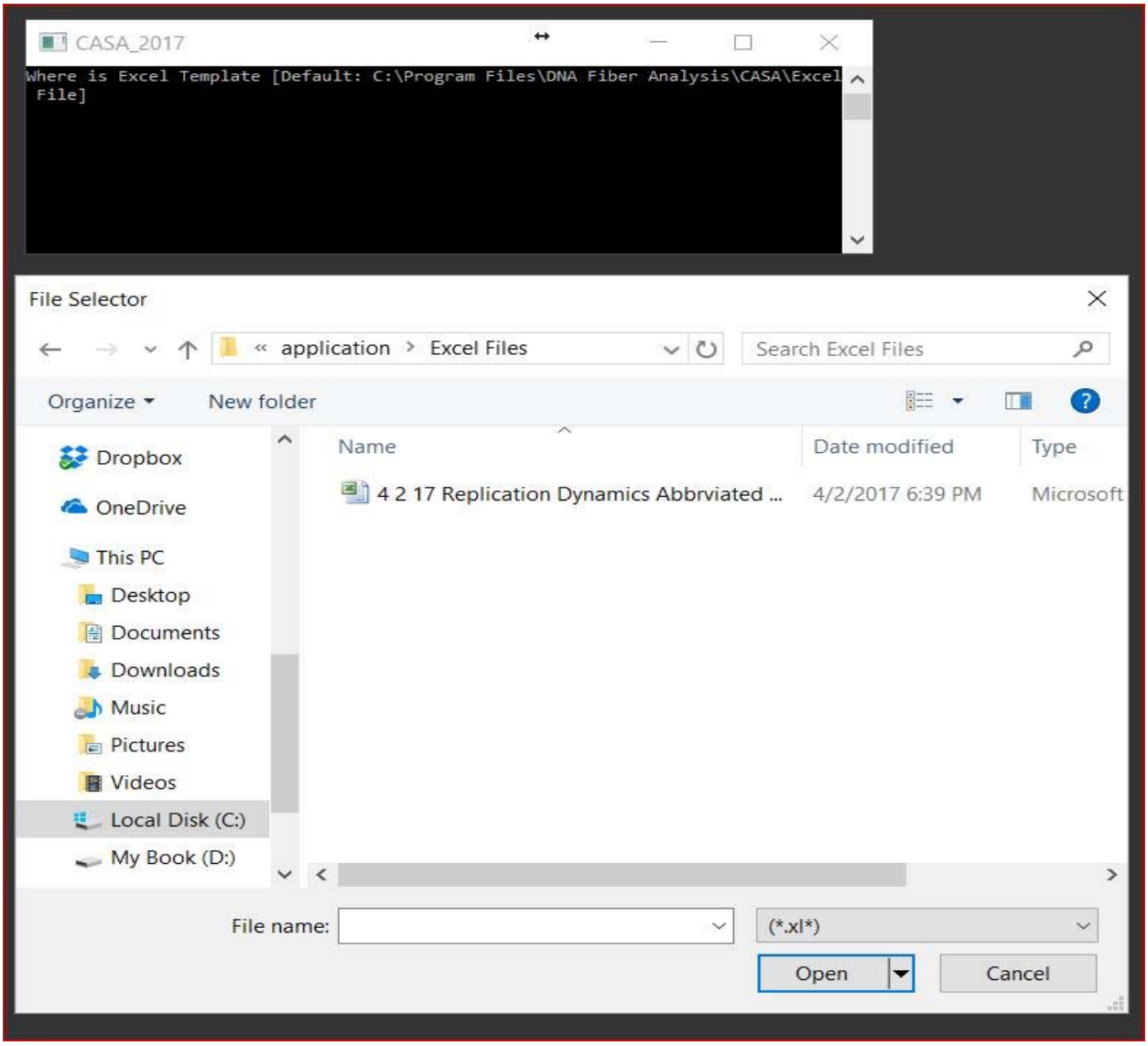
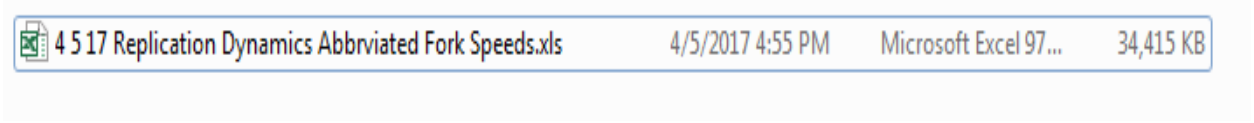


Figure 14: Picking Excel File

- A dialog box should open up and ask you to select an excel file. During the installation process, the Excel file was placed into your Program File Folder. (It is also located in your dropbox folder.)
  - DNA Fiber Analysis -> CASA_2017->Application ->Excel Files



- Click on Excel File and then select the file.
- Next, you will need to enter a set of parameters (i.e., values) that the program will use to detect, select, and analyze your fibers.
  -

C. Parameters

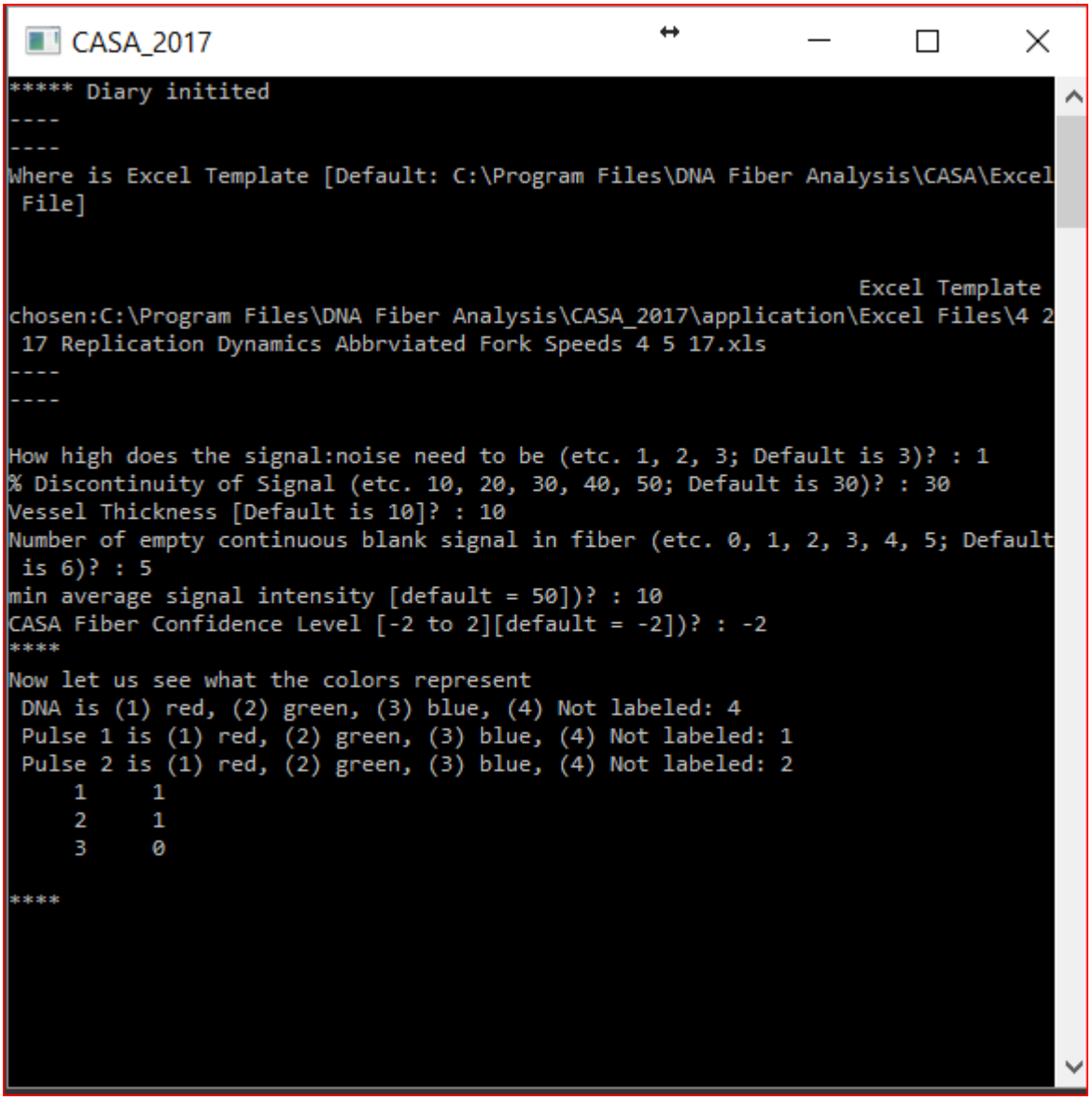


Figure 15: Parameters for CASA

Signal to Noise Ratio *(how much above background does the fiber have to be).*

- I typically use between 1 and 3.
- The algorithm takes the line tracing as the center of the line and takes each value along that tracing as a "middle point" of the signal. For each middle point of the signal, it determines a line that is perpendicular to that middle point (using the line tracing as a frame of reference) and uses the middle point of the signal plus the first two adjacent points along the perpendicular line (both above and below) as total Signal. For "Noise," the program sums the next three points along the perpendicular line (above and below). For the S/N, CASA divides the Signal by the Noise (giving more weight to the signal than the noise).

Discontinuity *(how much signal loss along a fiber is acceptable)*

- I typically use 30.
- Discontinuity is the percentage of the fiber without any signal (either there is no signal, or the signal is below the minimum average signal allowed).
- So "30" means that 30% of the fiber does not contain any "signal."

Number of empty continuous blank signal

- I typically use either 4 or 5.
- This parameter enables you to define how much of a gap in the signal you are okay within your fibers.
- Let us say a fiber has 30% discontinuity. That loss of signal can be throughout the fiber or a long, continuous stretch within the fiber. If the fiber has a long, continuous stretch without a signal, it may mean that CASA thought that two closely spaced fibers were one fiber.

Vessel (i.e., fiber) Thickness

- I usually use 8 for DNA fibers and 10 for chromatin fibers.
- Vessel Thickness is how thick (wide) a fiber is.
- Typically, DNA fibers are about 6 - 8 in thickness, chromatin is 8 - 10 in thickness, and bundles are 11 or so.

Minimum Ave Signal (Intensity) of Fiber

- I usually use 10, 25, or 50.
- This parameter reflects how much average signal the fiber has across its entire length.

CASA Fiber Confidence Level (Note: Experimental Parameter)

- For right now, we pick -2.
- This experimental parameter tries to figure out how confident the CASA program is with a fiber. This parameter is based on the percent continuity, number of empty continuous blank signals, and average signal of the fiber. We have not come up with a value that means this is a "lovely" fiber versus this is a "terrible" fiber. Hence, why we select -2 for the value that is acceptable – as of right now, all fibers are lovely to us.
- Currently, we use this parameter to see how one set of fibers compare to another. For instance, if one set of fibers has a confidence value of 0.6 and another set has a confidence level of 0.2, then you may feel more confident about the first fiber analysis than the second analysis.

What the colors represent

- Since the program does not know what it is analyzing, CASA will ask you for that information.
- Currently, this version of the program only utilizes two colors for its analysis of replication tracks
- Since our IdU is usually red and the first pulse and CldU is usually green and the second pulse. We do not label the DNA.
    - We input, 4 for DNA, Pulse 1 – (1) red, Pulse 2 – (2) green.
- As an aside, we have another version that can do three colors for analysis. We have used this program to do the following: Quantify DNA damage within fluorescently labeled replication tracks, quantifying inter-origin distances between replication tracks along a stained DNA fiber, characterizing replication patterns within Fiber Fish, characterizing the relationship between DNA damage associated proteins and stalled/stopped replication forks along chromatin fibers.
    - Currently, it is not commercially available as we are still optimizing this program. With that being said, if you want to help us optimize or beta test the program, please let us know.

Multi RG Tracks (more than 3 color segments) – Note: This pop-ups in a separate window

- Yes/No
- We usually allow the computer to analyze fibers that have red-only, green-only, red-green, red-green-red, and green-red-green replication track patterns. Depending on what we are interested in measuring, we may also analyze more complex replication patters (replication tracks with multiple red and green segments).

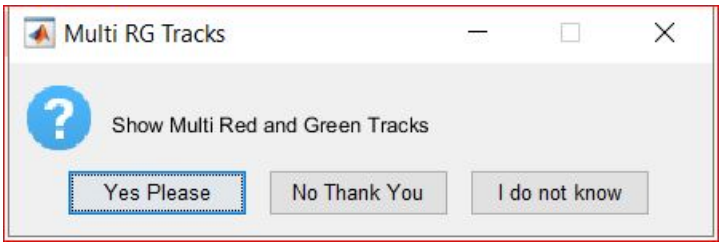


Figure 16: Multi RG Popup Window.

Min size for one color track

- I usually have 10 as my limit
- If you have a high background (or background dots), you may want to have 20 as your value. Typically, this will stop CASA from trying to connect the background dots into lines.

Min size for two color track

- I usually have 20 as my limit
- In the excel sheet, you can further define the min size of each segment within the two color lines.

Min size for three color track

- I usually have 30 as my limit
- In the excel sheet, you can further define the min size of each segment within the three color lines.

D. The program will ask you where the folder container the images is located.

- ***Images have to be in tif format*** (if your image is in a different format, please let me know and I will either 1) change CASA's computer code so will analyze those images as well, or 2) write an ImageJ/FIJI macro to convert your images to Tif – free of charge).

- ***Image files have to be in a subfolder*** (a folder within a folder). The main folder can have as many subfolders that you want to be analyzed (it analyzes all of the subfolders (but not any folders within the subfolders)).

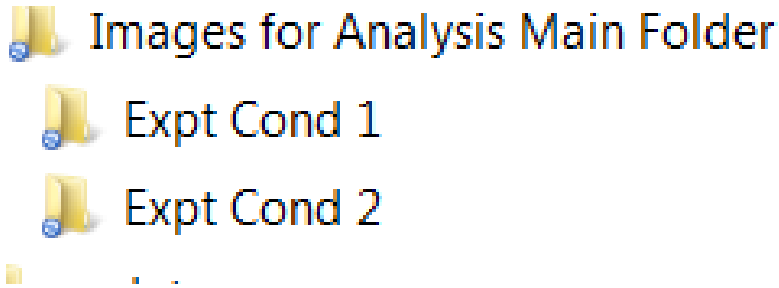


Figure 7: Folder Structure Needed for Program to Work

Step 3: What Happens When CASA Runs.

As the program runs, it opens up each image and assesses/analyzes the images for DNA fibers. Once a fiber has been found, CASA will sequentially “walk along” the fiber length. With each step, the program determines the color of that step and then moves along to the next step. Once it has walked along the entire length of the fiber, it then assigns a color to the line. If the line only has one color, then the color of the line is that color. If the line has multiple colors, it will divide the line into different segments with each segment one color.

After it processes the image, it will save a markup image of the image it analyzed in a folder called CASA Images. Each fiber in the markup image is enumerated (numbered) and has a box around it. The top and bottom lines show you how CASA color coded each segment of the line (i.e., did it assign that area to be red or green). The edges represent the beginning and end of the fiber. It also places the name of the image on the image as well as the parameters it used for its fiber analysis. The raw data that is generated by CASA is moved into an Excel File that allows all the data to be aggregated into one file (the file can be found in CASA results).

Below is a quick overview of additional steps that CASA performs.

A. CASA converts image intensities into a 3D "relief map."

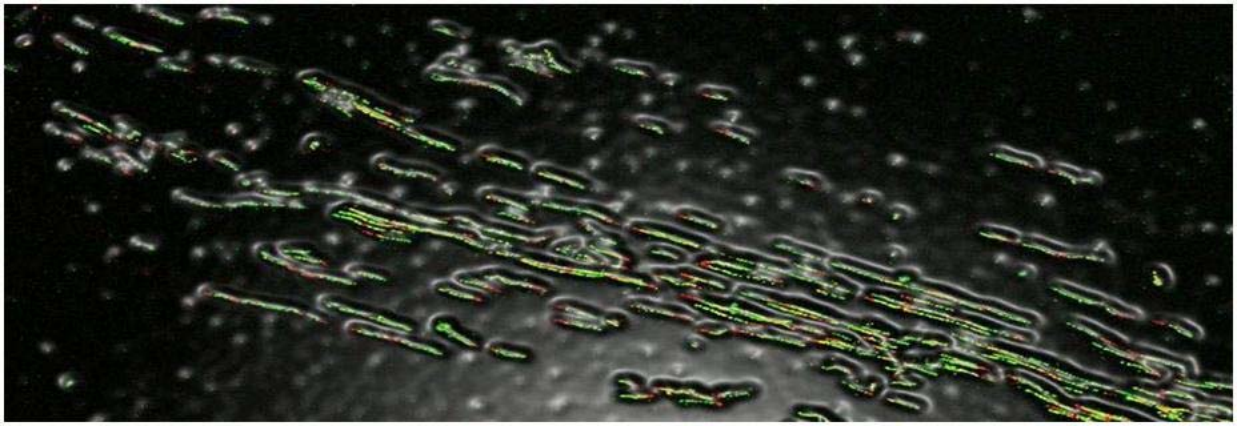


Figure 17: a Relief map of signal intensities within an image.

B. CASA marks up the image with its assessment of the red and green tracks along with a unique identifier for each fiber tracing.

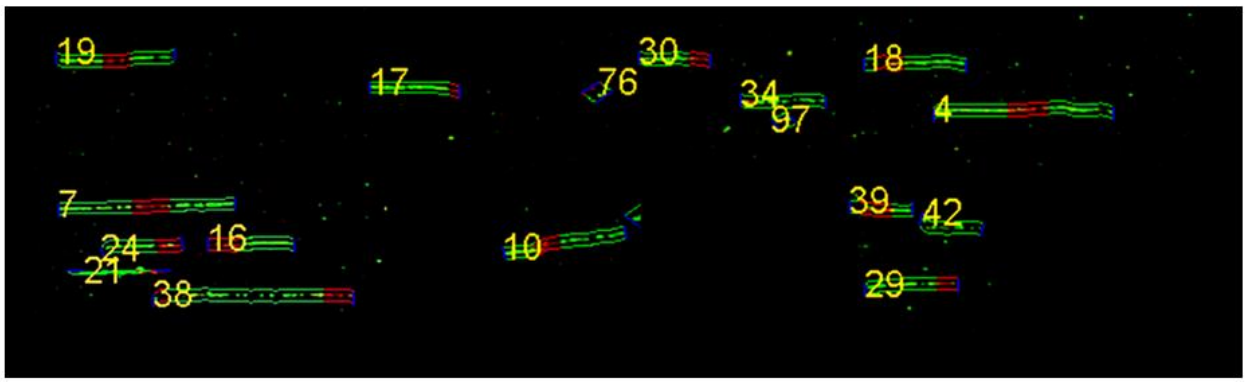


Figure 18: Section of Image that was analyzed by CASA. Fibers are enumerated, traced, and color segments identified.

C. CASA places raw data into an Excel file

The raw data for up to 20 images are located in one excel file. The location of each image data is in a different Excel tab. The raw data from Image 1 (the first image it analyzes) is placed in an Image1 tab, Image 2 in Image 2 tab, etc. (maximum of 20 images per Excel File). The first tab of the Excel file contains a summary of the raw data in Image1, Image2, Image3, etc.

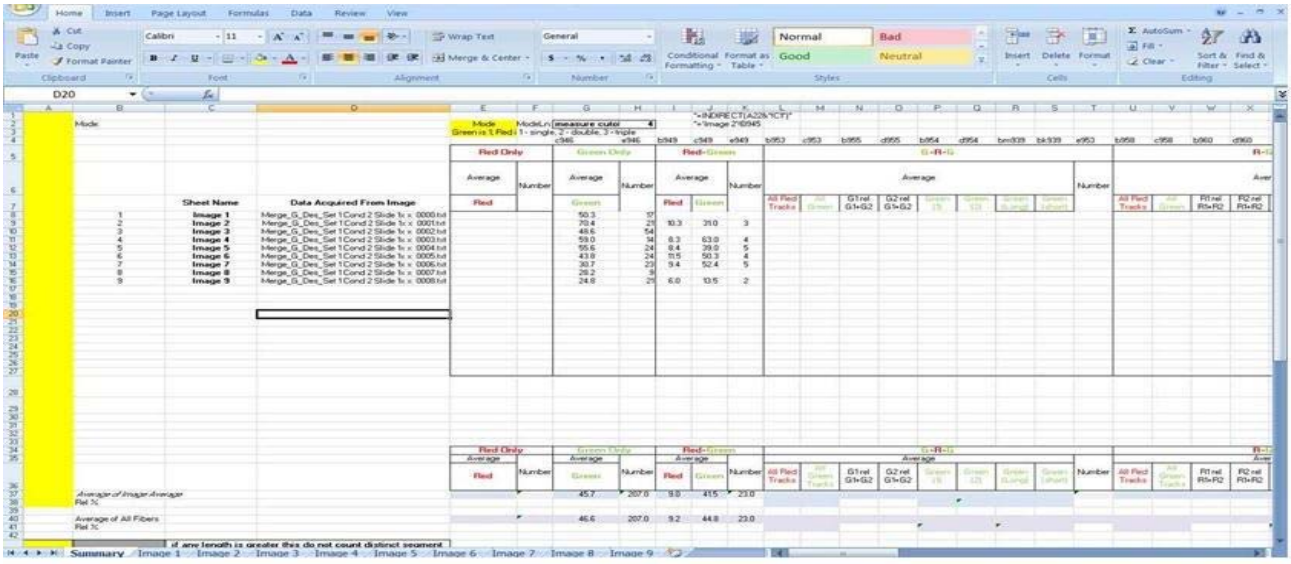


Figure 19: Excel data containing raw data and summary of images analyzed.

The Data is searchable so one can focus on red and green patterns of interest

ID	green (small)	red9	green (large)
39	6	21	15
18	9	18	41
57	9	5	19
56	9	5	13
103	6	4	16
10	22	15	43
19	29	18	31
7	43	24	48
4	43	26	50
80	60	5	66

Figure 20: Raw data within each image tab can be restricted based on various parameters.

Click on the triangle after each column label, and a pull-down tab will have all the ways in which that column can be searched. To search for a specific parameter just click on the corresponding box and it will be colored in (if you click on it again the box will be deselected. You can search multiple column parameters at one time).

The parameters can be further refined within the Excel file.

For instance, one can make sure that small tracks are not measured.

Parameter Settings		
max length	1000	max Length of Total Fiber
min size	5	Fiber Length is Below this do not count
R or G	5	if length is less than this, don't count it
R:G	5	if any segment is smaller than this length, do not count fiber
3 color	5	if any segment is smaller than this length, do not count fiber
Discontinuity of Signal Tolerated:	50%	Less than ths % of non-continuous signal, score and analyze
Level above background Needed:	0	Greater than this ratio, score and analyze
Vessel Thick Max:	10	Less than this value, it is counted
Sig Intensity:	0	If Segment Average Signal Intensity is Lower, do not count fiber
Level of Confidence	-2	
All Fibers Scored (including multi):	688.0	
All Fibers Scored (excluding multi):	631.0	

Figure 21: Parameters that can be refined in Excel File.

Pulse Lengths for each nucleotide precursor can be entered.

The replication rates (and number of replication forks per MB) are based on the pulse durations of the nucleotide analogs. To adjust the rates according to your pulse lengths, please enter those values as well.

Red Pulse = (min)	10	min
Green Pulse = (min)	20	min

Figure 22: Pulse Times for Red and Green Incubations.

Objective Lenses Used.

The replication rates values are based on Fiber FISH probes hybridized to DNA fibers that were analyzed using a 63X objective. To "correct" for when people use different objectives, we ask you to input your objective where it says Magnification (where the 40 is located below). Please do not enter it in the Magnification for conversion (as this the magnification we used in our original measurements).

Magification	40	X
Magnification for conversion	60	X

Figure 23: Objective Magnification of Images.

Number of Replication Forks per Replication Pattern

To determine the number of replication forks per Mb we decided upon the following:

A. A Red-only track only needed one 1 replication fork to make this track.
B. A Green-only track needed to have 1.5 replication forks to make this track.
    a. Our reasoning is that this track could be due to either a replication fork that was stalled before the first pulse but then continued after the second; or if it was an origin that initiated during the second pulse.
C. A Red-Green only needed 1 replication fork to make this track.
D. A Red-Green-Red track needed two replication forks to make this pattern.
E. A Green-Red-Green track needed two replication forks to make this pattern.

Note: These values are based on pulsing with red then with green. If you pulsed with green then red, then the values for red-only and green-only need to be manually switched in the excel file.

Replication Synthesis Assumptions		
Red Only Track	1	active forks
Green Track	1.5	active forks
Red-Green Track	1	active forks
Green-Red-Green	2	active forks
Red-Green-Red Tracks	2	active forks
Red-Green Clusters	1.5	active forks
Replication Rates		
bases replicated	(pixesls * /2.2)*3862))	
Replication Rates per fork	ave bases replicated (really ave pixes replicated)/ pulse time	

These values are when Red is pulsed first
and green is pulsed second
If the opposite is true then r-only should be changed to 1.5,
Green-only to 1

Figure 24: Number of Replication Forks per replication pattern.

Errors:

Q1: Why do I get an error saying "Undefined function or variable 'matlabrc"?

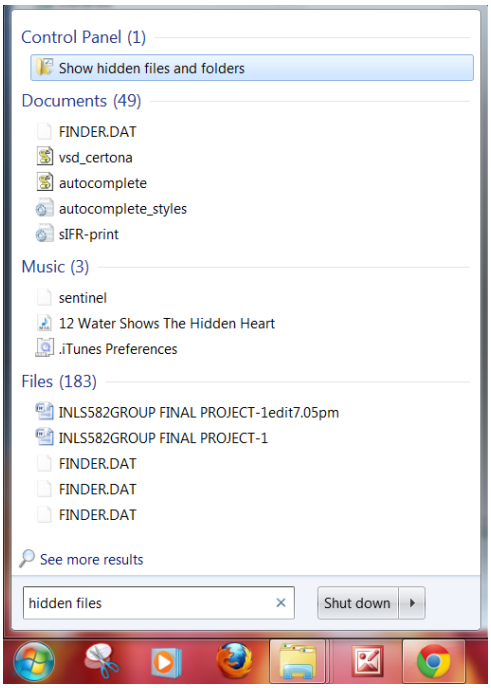
A1: Sometimes Matlab gets confused, and the compiler gets corrupt (people at Matlab do not know why).

How Do I fix this?
Delete the following folder: mcrCache7.13

How do I do this?
My computer's name is 603A1, so I go to the location of the location which contains the folder (for me it is the following: \Users\603A1\AppData\Local\Temp\603A1\mcrCache7.13) and delete it.

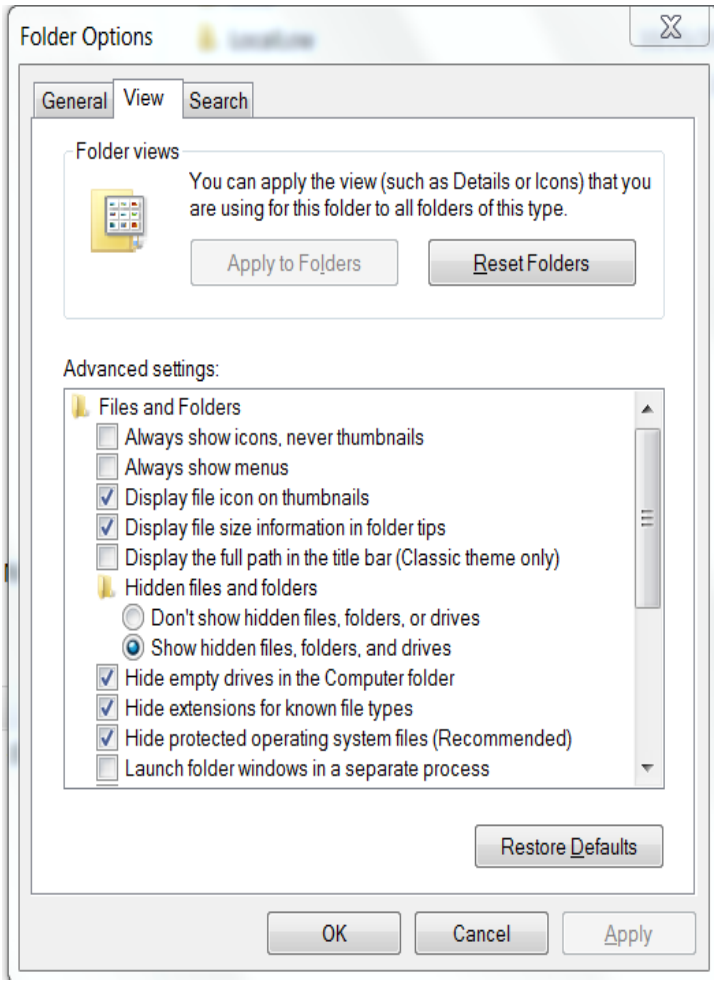
For you, the location of the file should be the same, except 603A1 is replaced by your computer's name.

So if your computer is named TAMU, then it would be located in the following location: C:\Users\TAMU\AppData\Local\Temp\TAMU\mcrCache7.13



Okay, but I cannot see AppData.

You need to search for hidden files with Window's search. Or, go to Control Panel\Appearance and Personalization and find an option called Show hidden files and folders.



After that, press the radial button which says “Show hidden files, folders, and drives).

Then you should be apple to find *AppData*

Q2: Why do I receive an error regarding missing mclmcrrt7x.dll when I run my stand-alone application compiled with MATLAB Compiler?

A2: Sometimes the installer places "the this is where the program engine" (i.e., the dll) at the wrong location in the "environmental variable" and CASA cannot "see it."

To fix this, you need to delete the text which contains the text -

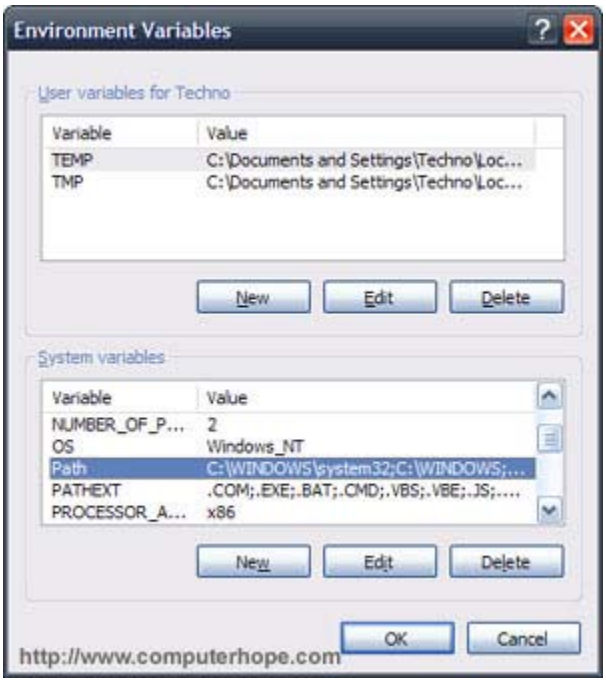
C:\Program Files (x86)\MATLAB\MATLAB Compiler Runtime\v713\runtime\win32;

To get to the environmental variables do the following:

How to set the path and environment variables in Windows

Windows Vista and Windows 7 users

1. From the Desktop, right-click **My Computer** and click **Properties**.
2. Click **Advanced System Settings** link in the left column.
3. In the System Properties window click the **Environment Variables** button.
4. In the Environment Variables window (as shown below), highlight the **Path**variable in the Systems Variable section and click the **Edit** button. Add or modify the path lines with the paths you want the computer to access. Each different directory is separated with a semicolon as shown below.



5. Finally, delete the text that has MATLAB\MATLAB Compiler Runtime\v713\runtime\win32; in it (between the ";"s).

C:\Program Files (x86)\NVIDIA Corporation\PhysX\Common;C:\PROGRAM FILES (X86)\INTEL\ICLS CLIENT\;C:\PROGRAM FILES\INTEL\ICLS CLIENT\;%SYSTEMROOT%\SYSTEM32;%SYSTEMROOT%;%SYSTEMROOT%\SYSTEM32\WBEM;%SYSTEMROOT%\SYSTEM32\WINDOWSPOWERSHELL\V1.0\;C:\PROGRAM FILES\INTEL\WIFI\BIN\;C:\PROGRAM FILES\COMMON FILES\INTEL\WIRELESSCOMMON\;C:\PROGRAM FILES\INTEL\INTEL(R) MANAGEMENT ENGINE COMPONENTS\DAL;C:\PROGRAM FILES\INTEL\INTEL(R) MANAGEMENT ENGINE COMPONENTS\IPT;C:\PROGRAM FILES (X86)\INTEL\INTEL(R) MANAGEMENT ENGINE COMPONENTS\DAL;C:\PROGRAM FILES (X86)\INTEL\INTEL(R) MANAGEMENT ENGINE COMPONENTS\IPT;;C:\PROGRAM FILES\DELL\DELL DATA PROTECTION\ACCESS\ADVANCED\WAVE\GEMALTO\ACCESS CLIENT\V5\;C:\PROGRAM FILES (X86)\SECURITY INNOVATION\SITSS\BIN\;C:\Program Files\WIDCOMM\Bluetooth Software\;C:\Program Files\WIDCOMM\Bluetooth Software\syswow64;;C:\Program Files (x86)\Intel\OpenCL SDK\2.0\bin\x86;C:\Program Files (x86)\Intel\OpenCL SDK\2.0\bin\x64;C:\Program Files (x86)\QuickTime\QTSystem\;

So the text I want to find has MATLAB v713 runtime and win32 as part of the text…MATLAB\MATLAB Compiler Runtime\v713\runtime\win32;

Which would be this "sentence" – remember sentences start with hard drive name and end with;

C:\Program Files (x86)\MATLAB\MATLAB Compiler Runtime\v713\runtime\win32;

Next, just delete the appropriate sentence. ***Make sure you only delete the sentence and the (from the beginning to the ";") otherwise major problems can occur.***

So this is what my environmental variable now are…

C:\Program Files (x86)\NVIDIA Corporation\PhysX\Common;C:\PROGRAM FILES (X86)\INTEL\ICLS CLIENT\;C:\PROGRAM FILES\INTEL\ICLS CLIENT\;%SYSTEMROOT%\SYSTEM32;%SYSTEMROOT%;%SYSTEMROOT%\SYSTEM32\WBEM;%SYSTEMROOT%\SYSTEM32\WINDOWSPOWERSHELL\V1.0\;C:\PROGRAM FILES\INTEL\WIFI\BIN\;C:\PROGRAM FILES\COMMON FILES\INTEL\WIRELESSCOMMON\;C:\PROGRAM FILES\INTEL\INTEL(R) MANAGEMENT ENGINE COMPONENTS\DAL;C:\PROGRAM FILES\INTEL\INTEL(R) MANAGEMENT ENGINE COMPONENTS\IPT;C:\PROGRAM FILES (X86)\INTEL\INTEL(R) MANAGEMENT ENGINE COMPONENTS\DAL;C:\PROGRAM FILES (X86)\INTEL\INTEL(R) MANAGEMENT ENGINE COMPONENTS\IPT;;C:\PROGRAM FILES\DELL\DELL DATA PROTECTION\ACCESS\ADVANCED\WAVE\GEMALTO\ACCESS CLIENT\V5\;C:\PROGRAM FILES (X86)\SECURITY INNOVATION\SITSS\BIN\;C:\Program Files\WIDCOMM\Bluetooth Software\;C:\Program Files\WIDCOMM\Bluetooth Software\syswow64;;C:\Program Files (x86)\Intel\OpenCL SDK\2.0\bin\x86;C:\Program Files (x86)\Intel\OpenCL SDK\2.0\bin\x64;C:\Program Files (x86)\QuickTime\QTSystem\;

Sometimes you need to place that statement at the beginning of the variables.