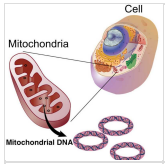


Aug 25, 2021

Mitogenome Assembly from NGS Genome Skimming Data

Avery S Hiley[1]

[1]UCSD - Scripps Institution of Oceanography

Avery S Hiley: seahatch3@gmail.com | ahiley@ucsd.edu | http://www.spineless.info/avery-hiley.html

dx.doi.org/10.17504/protocols.io.bkbqksmw

Avery Hiley
UCSD - Scripps Institution of Oceanography

ABSTRACT

This protocol provides thorough instructions for how to assemble and annotate full mitochondrial genomes from NGS genome skimming data (specifically paired-end, FASTQ, gzipped reads). The steps are categorized in the following sections, which include detailed explanations throughout (e.g. why specific steps are executed in the manner instructed):

I. Download Programs and Dependencies
II. Choose Working Directory
III. Simple Stats with SeqKit
IV. Clean and Trim Reads with Trimmomatic
V. Downsample Reads with MITObim
VI. Reformat Reads to Non-Interleaved with BBMap
VII. Mitogenome Assembly and Annotation: MitoFinder with MetaSPAdes
VIII. Mitogenome Assembly with NOVOPlasty

Furthermore, there are supplementary tutorial videos for each of the aforementioned protocol sections, which show all of the corresponding steps being executed. These videos may be used for extra guidance while running the protocol, and are especially helpful to see how long the commands take to execute and successfully terminate in real-time with real NGS genome skimming data.

DOI

dx.doi.org/10.17504/protocols.io.bkbqksmw

PROTOCOL CITATION

Avery S Hiley 2021. Mitogenome Assembly from NGS Genome Skimming Data. **protocols.io** https://dx.doi.org/10.17504/protocols.io.bkbqksmw

KEYWORDS

Mitogenome, Mitochondrial Genome, Genome Skimming, NGS, Mitogenome Assembly, Mitogenome Annotation, Bioinformatics

CREATED

Aug 25, 2020

LAST MODIFIED

Aug 25, 2021

OWNERSHIP HISTORY

Aug 25, 2020	Marina Mccowin	UCSD- Scripps Institution of Oceanography
Aug 28, 2020	Avery Hiley	UCSD - Scripps Institution of Oceanography

PROTOCOL INTEGER ID

41040

GUIDELINES

The current version of this protocol is streamlined for Mac operating systems. However, alterations to the protocol for other operating systems (e.g. Windows OS) should be relatively simple, mainly requiring the use of different apps with equivalent applications to the macOS apps specified, and different download links for installation of the required programs and dependencies.

MATERIALS TEXT

Detailed instructions on how to correctly download the required programs and their corresponding dependencies are provided in section *I. Download Programs and Dependencies*, along with links and citations. Therefore, it is not necessary to download the following programs beforehand; they are merely listed here for your reference.

1. Xcode Version 12.1
2. Homebrew
3. SeqKit v0.13.2
4. Trimmomatic Version 0.39
5. MITObim Version 1.9.1
6. BBMap Version 38.87
7. MitoFinder v1.4
8. NOVOPlasty Version 4.2

YouTube links for the supplementary tutorial videos are pasted below for your reference. Additionally, these videos are directly embedded in the protocol text at the beginning of each corresponding section.

I. Download Programs and Dependencies: https://youtu.be/ytmsjSH6oog
II. Choose Working Directory: https://youtu.be/FFiJ2KK0ndg
III. Simple Stats with SeqKit: https://youtu.be/s8liMiFt9h4
IV. Clean and Trim Reads with Trimmomatic: https://youtu.be/7uPZXl11PaI
V. Downsample Reads with MITObim: https://youtu.be/Yhn1Urrmf94
VI. Reformat Reads to Non-Interleaved with BBMap: https://youtu.be/kgpXPVhqI4g
VII. Mitogenome Assembly and Annotation: MitoFinder with MetaSPAdes: https://youtu.be/v2miYTNJaLg
VIII. Mitogenome Assembly with NOVOPlasty: https://youtu.be/0FKIEGltRvs

BEFORE STARTING

Download the NGS genome skimming data (paired-end, FASTQ, gzipped reads) for the "species" of interest. There should be 2 files total per gDNA extraction sequenced, typically ending in 1.fq.gz and 2.fq.gz if the data was obtained from Novogene. Additionally, you must have a COI Sanger Sequence (FASTA format) obtained from the exact gDNA extraction of the species/sample of interest if you perform section *VIII. Mitogenome Assembly with NOVOPlasty.*

I. Download Programs and Dependencies

1 Optionally, follow along with this supplementary video as you perform section *I. Download Programs and Dependencies:*

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

Download the required programs, including any prerequisites and dependencies, by following the subsequent steps.

Xcode Version 12.1 is for macOS only. **Homebrew** is for macOS or Linux. Furthermore, this entire protocol is structured for Mac computer users. However, alterations to the protocol for other operating systems (e.g. Windows OS and Linux) should be relatively simple, mainly requiring the use of different apps with equivalent applications to the macOS apps specified (e.g. **Terminal** and **Finder**) and different download links for the program/dependency installations.

All programs relevant to mitogenome assembly (bioinformatic softwares and pipelines) and computer apps are set in **bold** throughout the protocol text.

Additionally, specific file and folder names, buttons that you are instructed to click on, keyboard keys that you need to press, and direct text that you need to reference in any given step are set in *italics*.

2 Open up the **App Store** (macOS), and search for **Xcode** (macOS). Download **Xcode Version 12.1**. Minimize the **App Store** window. Proceed to step 3 while the **Xcode** download is in progress, because this may take some time.

Xcode Version 12.1
source by Apple

Apple Inc. (2020). Xcode (Version 12.1) [Computer software]. Mac App Store. https://apps.apple.com/us/app/xcode/id497799835?mt=12

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

3 Download **Homebrew**, and install the **GNU Wget** computer program and the **GNU Command Line Tools** by following the subsequent steps.

Homebrew
source by Max Howell

Howell, M. (2009). Homebrew: The Missing Package Manager for macOS (or Linux) (Version 2.5.9) [Computer software]. Homebrew. https://brew.sh

3.1 Open up a new **Terminal** (macOS) shell window.

All commands used in this protocol are identified by the blue command icon that strikingly resembles the **Terminal** icon. The name of the command is listed to the right of the blue command icon. The associated command text (including any relevant arguments or specifications) to copy is transcribed in the dark grey box. The description of the command is listed underneath the dark grey box. When a command name is referred to in the protocol text (outside of the isolated, blue command icon contents), it will be placed in quotations.

Copy, paste, and execute (click *return* on your keyboard) the following "/bin/bash -c Homebrew" command in order to download **Homebrew**:

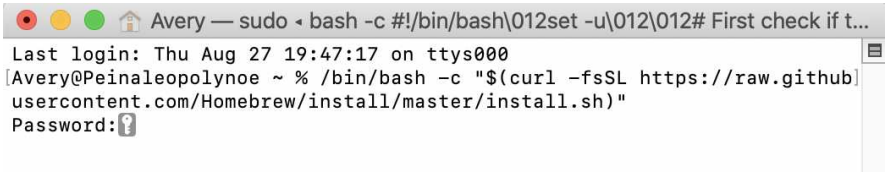
/bin/bash -c Homebrew

/bin/bash -c "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/master/install.sh)"
Installs Homebrew to your computer, which includes several bioinformatic packages you need that Apple (or your Linux system) didn't have automatically.

Howell, M. (2009). Homebrew: The Missing Package Manager for macOS (or Linux) (Version 2.5.9) [Computer software]. Homebrew. https://brew.sh

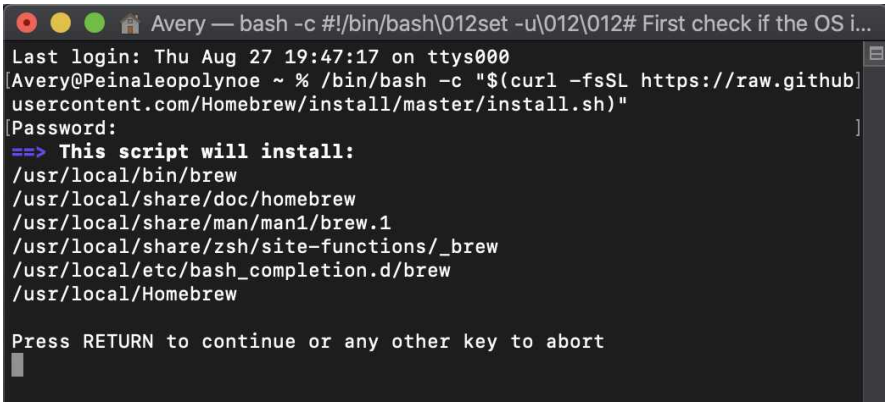
After executing the aforementioned command, you will be prompted to enter your computer password (example screenshot of the **Terminal** shell window pasted below for your reference).

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw



Type in your computer password and click *return* on your keyboard.

When prompted to do so, click *return* on your keyboard again to proceed with the installation (example screenshot of the **Terminal** shell window pasted below for your reference).

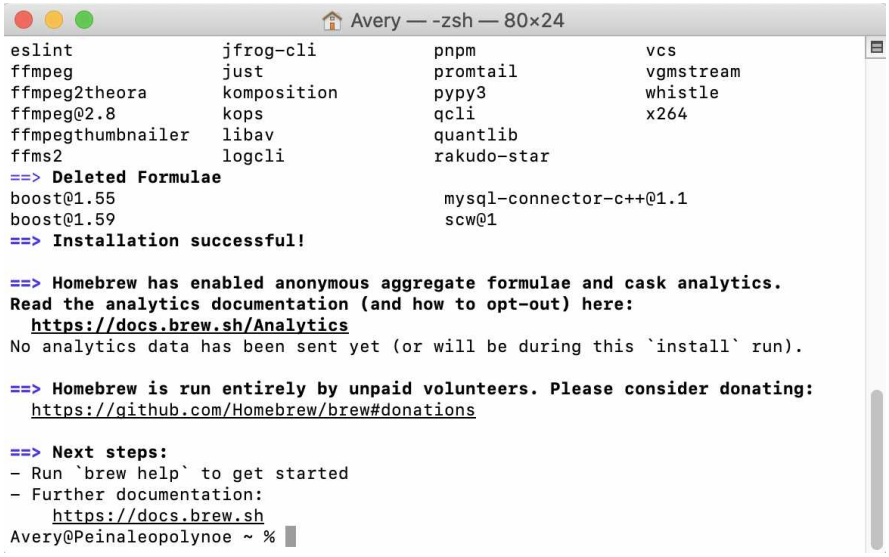


If you are updated to the macOS Catalina software, the default login shell for **Terminal** is the Z shell (zsh). When an executed command has terminated in **Terminal** (zsh), whether or not successfully, you will see a line of text following either of these two formats:

1. *User@Computer ~ %*
   - Appears if a working directory (folder) was not specified in the **Terminal** shell window

2. *User@Computer Name_of_working_directory %*
   - Appears if a working directory (folder) was specified in the **Terminal** shell window

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw



For example, this screenshot of my **Terminal** zsh window shows that the command to download **Homebrew** has successfully terminated, because the last line of text listed follows the aforementioned option 1 format:

Avery@Peinaleopolynoe ~ %
Avery = User
Peinaleopolynoe = Computer

In this example, only the name of my computer is listed after @, because I did not specify a different folder to work within. In section II, you will choose a more specific working directory (folder), and the line of text for command termination will follow the aforementioned option 2 format.

3.2 Once the "/bin/bash -c Homebrew" command has successfully terminated, copy, paste, and execute (click *return* on your keyboard) the following "brew install wget" command into the same **Terminal** shell window in order to install the **GNU Wget** computer program:

brew install wget

brew install wget

Installs the GNU Wget computer program to your system using the general "brew install" Homebrew command.

Niksic, H. (2018). GNU Wget (Version 1.20) [Computer software]. Homebrew Formulae.
https://formulae.brew.sh/formula/wget#default

3.3 Once the "brew install wget" command has successfully terminated, copy, paste, and execute (click *return* on your keyboard) the following "brew install coreutils" command into the same **Terminal** shell window in order to install the **GNU Command Line Tools**:

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

brew install coreutils

brew install coreutils

Installs the GNU Command Line Tools to your system using the general "brew install" Homebrew command.

MacKenzie, D., & Meyering, J. (2020). GNU Coreutils (Version 8.32) [Computer software]. Homebrew Formulae. https://formulae.brew.sh/formula/coreutils

4 Download **SeqKit v0.13.2** by following the subsequent steps.

SeqKit v0.13.2

source by doi:10.1371/journal.pone.0163962

Shen, W. (2020). SeqKit (Version 0.13.2) [Computer software]. GitHub. https://github.com/shenwei356/seqkit/releases/tag/v0.13.2

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. Plos One, 11(10), E0163962. http://10.1371/journal.pone.0163962

4.1 Check if your Mac's processor is 32-bit or 64-bit by viewing *Processor* under *About This Mac*. "If the processor is an Intel Core Solo or Intel Core Duo, it's 32-bit only" (Gamet 2018).

Gamet, J. (2018). How to Tell if Your Mac is 32-bit or 64-bit. theMacObserver. https://www.macobserver.com/tips/how-to/mac-32-bit-64-bit/

4.2 Click on the following link: https://github.com/shenwei356/seqkit/releases/tag/v0.13.2

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

4.3 Download the correct **SeqKit** file for your corresponding OS X processor (example screenshot from the GitHub web page pasted below for your reference). Most likely your Mac's processor is 64-bit, so step 4.4 will assume this condition. Please adjust this subsequent step accordingly if your Mac's processor is 32-bit; only the file name in your *Downloads* should be different, but the instructions are the same.

OS	Arch	File, 中国镜像	Download Count
Linux	32-bit	seqkit_linux_386.tar.gz, 中国镜像	downloads@latest 189 [seqkit_linux_386.tar.gz]
Linux	**64-bit**	**seqkit_linux_amd64.tar.gz**, 中国镜像	downloads@latest 1.9k [seqkit_linux_amd64.tar.gz]
OS X	32-bit	seqkit_darwin_386.tar.gz, 中国镜像	downloads@latest 41 [seqkit_darwin_386.tar.gz]
OS X	**64-bit**	**seqkit_darwin_amd64.tar.gz**, 中国镜像	downloads@latest 425 [seqkit_darwin_amd64.tar.gz]
Windows	32-bit	seqkit_windows_386.exe.tar.gz, 中国镜像	downloads@latest 55 [seqkit_windows_386.exe.tar.gz]
Windows	**64-bit**	**seqkit_windows_amd64.exe.tar.gz**, 中国镜像	downloads@latest 411 [seqkit_windows_amd64.exe.tar.gz]

4.4 If you downloaded the link for OS X 64-bit (*seqkit_darwin_amd64.tar.gz*), open the *seqkit_darwin_amd64.tar* file in your *Downloads* folder.

4.5 Open up a new **Finder** (macOS) window and go to your *Applications* folder. Transfer the resulting *seqkit* (Unix executable) file from your *Downloads* to your *Applications* folder.

5 Download **Trimmomatic Version 0.39** by following the subsequent steps.

Trimmomatic Version 0.39

source by https://doi.org/10.1093/bioinformatics/btu170

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114-2120. http://10.1093/bioinformatics/btu170

Bolger, A. M., Lohse, M., & Usadel, B. (2019). Trimmomatic (Version 0.39) [Computer software]. USADELLAB.org. http://www.usadellab.org/cms/?page=trimmomatic

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

5.1 Click on the following link: http://www.usadellab.org/cms/?page=trimmomatic

5.2 Click on the **Trimmomatic Version 0.39** *binary* link to start the download.

5.3 Transfer the resulting *Trimmomatic-0.39* folder from your *Downloads* to your *Applications* folder.

6 Download **MITObim Version 1.9.1** by following the subsequent steps.

MITObim Version 1.9.1
source by doi: 10.1093/nar/gkt371

Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Research, 41(13). http://10.1093/nar/gkt371

Hahn, C., Bachmann, L., & Chevreux, B. (2018). MITObim (Version 1.9.1) [Computer software]. GitHub. https://github.com/chrishah/MITObim

6.1 Click on the following link: https://github.com/chrishah/MITObim

6.2 Click the green *Code* button on the right hand side of the page.

6.3 Click *Download ZIP*.

6.4 Transfer the resulting *MITObim-master* folder from your *Downloads* to your *Applications* folder.

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

6.5 Under *Prerequisites* in the *README.md MITObim - mitochondrial baiting and iterative mapping* document (on the GitHub web page), it states that **GNU** utilities, **Perl**, and a running version of **MIRA** must be installed on your system.

GNU utilities were already installed in step 3.

Mac OS X is based on a UNIX operating system and thus comes with a **Perl** interpreter already installed.

Click on the *download here* link listed after *MIRA 4.0.2 (for the use with MITObim 1.8 and newer)*. Then click the large, green *Download Latest Version* button on the SourceForge web page. Open the *mira_4.0.2_darwin13.1.0_x86_64_static.tar.bz2* file in your *Downloads* folder. Transfer the resulting *mira_4.0.2_darwin13.1.0_x86_64_static* folder from your *Downloads* to your *Applications* folder.

Chevreux, B., Weber, J., Hörster, A., & Dlugosch, K. (2014). MIRA (Version 4.0.2) [Computer software]. SourceForge. https://sourceforge.net/projects/mira-assembler/files/MIRA/stable/

7 Download **BBMap Version 38.87** by following the subsequent steps.

BBMap Version 38.87
source by Brian Bushnell

Bushnell, B. (2020). BBMap (Version 38.87) [Computer software]. SourceForge. http://sourceforge.net/projects/bbmap/files

7.1 Click on the following link: https://sourceforge.net/projects/bbmap/files/

7.2 Click the large, green *Download Latest Version* button, or click the blue link to download the *BBMap_38.87.tar.gz* file.

7.3 Open the *BBMap_38.87.tar* file in your *Downloads* folder.

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

7.4 Transfer the resulting *bbmap* folder from your *Downloads* to your *Applications* folder.

8 Download **MitoFinder v1.4** and install the dependencies by following the subsequent steps.

MitoFinder v1.4
source by https://doi.org/10.1111/1755-0998.13160

MitoFinder is one of the most all-inclusive and user-friendly bioinformatic pipelines for assembling mitochondrial genomes from various types of NGS data. **MitoFinder** was designed to capture mitochondrial genomes from smaller NGS datasets, namely from target enrichment phylogenomics. However, **MitoFinder** is also applicable to (and successful in analyzing NGS data obtained from) other sequence capture methods, transcriptomic data, and whole genome shotgun sequencing in diverse taxa. Although **MetaSPAdes** has a computation time on average twice that of the other two metagenomic assemblers (**MEGAHIT** and **IDBA-UD**) that may be used with **MitoFinder**; Allio et al. (2020) determined that it is still the most effective in accurately assembling both UCE and mtDNA contigs, notably in the assembly of full mitochondrial genomes in a single contig.

Upon completion of analyzing NGS datasets through the **MitoFinder** pipeline in combination with the **MetaSPAdes** assembler, **MitoFinder** generates a series of extremely organized and useful output files, including but not limited to:

- Fasta file (.fasta) containing the amino acid translations for the 13 PCGs (protein coding genes).
- Fasta file (.fasta) containing the nucleotide sequences for the 15 total mitochondrial genes (13 PCGs and 2 rRNAs).
- Fasta file (.fasta) containing the full length mitochondrial genome in a single nucleotide contig.
- Genbank files (.gb and .tbl) for the resulting mitogenome that are ready for upload to NCBI-- in addition to the full length, nucleotide mitogenome contig, this file contains the annotations of all the gene features (tRNAs and their products, PCGs and their corresponding translations, and rRNAs and their products).
- GFF file (.gff) containing the gene annotations, which can be easily dragged into Geneious and mapped onto the mitogenome contig. This gives a great visual for the gene order. Additionally, if MitoFinder wasn't able to assemble all 15 genes for some reason, you may look for gaps in the gene order. If any gaps exist, you can map sequences of the missing gene from closely related animals to the annotated mitogenome contig. If you had a lot of raw reads to begin with, often these sequences will successfully map to one of the gaps in your gene order.
- Information file (.infos) containing the length of your final contig, the mitogenome coverage, the GC content, and the circularization.
- Fasta file in the metaspades output directory titled "contigs.fasta" containing a series of nuclear contigs that were assembled in addition to the mitochondrial genome contig, along with their nucleotide lengths and associated coverage in the dataset.

Allio, R., Schomaker-Bastos, A., Romiguier, J., Prosdocimi, F., Nabholz, B., & Delsuc, F. (2020). MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. Molecular Ecology Resources, 20(4), 892-905. http://10.1111/1755-0998.13160

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

Allio, R., Schomaker-Bastos, A., Romiguier, J., Prosdocimi, F., Nabholz, B., & Delsuc, F. (2020). MitoFinder (Version 1.4) [Computer software]. GitHub. https://github.com/RemiAllio/MitoFinder

8.1 Click on the following link: https://github.com/RemiAllio/MitoFinder

8.2 Scroll down on the GitHub web page until you see the *README.md MitoFinder version 1.4* document.

8.3 Under *Requirements*, it states that **automake**, **autoconf**, and **gcc** must be installed on your system. Copy, paste, and execute (click *return* on your keyboard) the following "brew install automake autoconf gcc" command into the same **Terminal** shell window:

brew install automake autoconf gcc

brew install automake autoconf gcc
Installs automake, autoconf, and gcc to your system using the general "brew install" Homebrew command.

MacKenzie, D., Elliston, B., & Demaille, A. (2012). GNU Autoconf (Version 2.69) [Computer software]. Homebrew Formulae. https://formulae.brew.sh/formula/autoconf

MacKenzie, D., Tromey, T., Duret-Lutz, A., Wildenhues, R., & Lattarini, S. (2020). GNU Automake (Version 1.16.2) [Computer software]. Homebrew Formulae. https://formulae.brew.sh/formula/automake

Stallman, R. M., & GCC Developer Community (2020). GCC, the GNU Compiler Collection (Version 10.2) [Computer software]. Homebrew Formulae. https://formulae.brew.sh/formula/gcc

Under the *Table of contents*, select the *MAC* link for *Installation guide for MitoFinder*.

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

8.4

8.5 Once the "brew install automake autoconf gcc" command has successfully terminated, copy, paste, and execute (click *return* on your keyboard) the following "cd Applications" command into the same **Terminal** shell window:

cd Applications

cd /Applications

Chooses the Applications folder as your new working directory in Terminal using the general "cd" command.

Neagu, C. (2020). Command Prompt: 11 basic commands you should know (cd, dir, mkdir, etc.). Digital Citizen. http://www.digitalcitizen.life/command-prompt-how-use-basic-commands

8.6 Once the "cd Applications" command has successfully terminated, copy, paste, and execute (click *return* on your keyboard) the following "wget MitoFinder" command into the same **Terminal** shell window:

wget MitoFinder

wget https://github.com/RemiAllio/MitoFinder/archive/master.zip

Installs the master MitoFinder folder to your computer using the general "wget" GNU command.

Free Software Foundation, Inc. (2020). GNU Wget. GNU Operating System. https://www.gnu.org/software/wget/

8.7 Once the "wget MitoFinder" command has successfully terminated, copy, paste, and execute (click *return* on your keyboard) the following "unzip master.zip" command into the same **Terminal** shell window:

unzip master.zip

unzip master.zip

Unzips the downloaded, master MitoFinder folder.

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

Akamai Technologies, Inc. (2020). The “unzip” command. Akamai: NetStorage - User Guide. https://learn.akamai.com/en-us/webhelp/netstorage/netstorage-user-guide/GUID-80C7B749-F9BB-4271-A138-AEEED1070D11.html

8.8 Once the "unzip master.zip" command has successfully terminated, copy, paste, and execute (click *return* on your keyboard) the following "mv MitoFinder" command into the same **Terminal** shell window:

mv MitoFinder

mv MitoFinder-master MitoFinder

The mv command moves files and directories from one directory to another or renames a file or directory. In this specific instance, it changes the name of the folder from MitoFinder-master to MitoFinder.

Akamai Technologies, Inc. (2020). The “mv” command. Akamai: NetStorage - User Guide. https://learn.akamai.com/en-us/webhelp/netstorage/netstorage-user-guide/GUID-1D8171AD-9B63-4344-964F-94D77EE83F3F.html

Apple Inc. (2020). Move and copy files in Terminal on Mac. Apple Support. https://support.apple.com/guide/terminal/move-and-copy-files-apddfb31307-3e90-432f-8aa7-7cbc05db27f7/mac

8.9 Once the "mv MitoFinder" command has successfully terminated, copy, paste, and execute (click *return* on your keyboard) the following "cd MitoFinder" command into the same **Terminal** shell window:

cd MitoFinder

cd /Applications/MitoFinder

Chooses the MitoFinder folder as your new working directory in Terminal using the general "cd" command.

Neagu, C. (2020). Command Prompt: 11 basic commands you should know (cd, dir, mkdir, etc.). Digital Citizen. http://www.digitalcitizen.life/command-prompt-how-use-basic-commands

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

8.10 Once the "cd MitoFinder" command has successfully terminated, copy, paste, and execute (click *return* on your keyboard) the following "./install.sh" command into the same **Terminal** shell window:

```
./install.sh
```

./install.sh

The Centrify agent installation script ("./install.sh") upgrades the corresponding softwares and/or programs within the current working directory specified.

Centrify Corporation (2016). Using the install.sh shell script to update packages. Centrify. https://docs.centrify.com/en/css/suite2017-html/#page/Upgrade_and_compatibility/Using_the_install.sh_shell_script_to_update_pack

8.11 In order for your computer to have access to execute **MitoFinder**, copy, paste, and execute (click *return* on your keyboard) the following "chmod +x mitofinder" command into the same **Terminal** shell window, once the "./install.sh" command has successfully terminated:

```
chmod +x mitofinder
```

chmod +x /Applications/MitoFinder/mitofinder

chmod +x on a file (your script) means that you'll make it executable, in this case mitofinder.

Ask Ubuntu (2014). What does "chmod +x" do and how do I use it?. Ask Ubuntu. https://askubuntu.com/questions/443789/what-does-chmod-x-filename-do-and-how-do-i-use-it

Delete the *master.zip* file in your *Applications*, as it no longer serves a purpose.

8.12 Expand the *MitoFinder* folder in your *Applications* and drag the *Mitofinder.config* file into a text editor, such as **BBEdit**, **TextWrangler**, or **TextEdit** (macOS).

8.13 In **Finder**, right click (or *control click* on your Mac keyboard) on the *megahit* folder within the *MitoFinder* folder. Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "megahit" as Pathname*.

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

8.14 In your *Mitofinder.config* file, delete *default* after *megahitfolder =* and paste in the pathname you copied. Add a forward slash to the end of this text, resulting in *megahitfolder = /Applications/MitoFinder/megahit/*

8.15 Copy and paste *ftp://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST/* into your **Safari** Internet browser. You must include the *ftp://* preceding the link; do not just click the blue link. Select *Allow* in response to *Do you want to allow this page to open "Finder"?*

NCBI (2020). NCBI BLAST executables (Version 2.11.0) [Computer software]. Index of /blast/executables/LATEST. http://ftp://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST/

8.16 Connect to the server as a *Guest*.

8.17 Double click on the *ncbi-blast-2.11.0+-x64-macosx.tar.gz* file to download the associated **BLAST** binaries required for **MitoFinder** to run.

8.18 Transfer the resulting *ncbi-blast-2.11.0+* folder from your *Downloads* to your *Applications* folder.

8.19 Expand the *ncbi-blast-2.11.0+* folder in your *Applications*. Right click (or *control click* on your Mac keyboard) on the *bin* folder. Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "bin" as Pathname*.

8.20 In your *Mitofinder.config* file, delete *default* after *blastfolder =* and paste in the pathname you copied. Add a forward slash to the end of this text, resulting in *blastfolder = /Applications/ncbi-blast-2.11.0+/bin/*

8.21 Under the *Assemblers MetaSPAdes* section on the GitHub web page, click the blue link to *download the pre-compiled binaries*.

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P.A. (2017). MetaSPAdes: A new versatile metagenomic assembler. Genome Research, 27(5), 824-834. http://10.1101/gr.213959.116

8.22 Double click the *SPAdes-3.14.0-Darwin.tar* file in your *Downloads*. Transfer the resulting *SPAdes-3.14.0-Darwin* folder from your *Downloads* to your *Applications* folder.

8.23 Expand the *SPAdes-3.14.0-Darwin* folder in your *Applications*. Right click (or *control click* on your Mac keyboard) on the *bin* folder. Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "bin" as Pathname*.

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

8.24 In your *Mitofinder.config* file, delete *default* after *metaspadesfolder =* and paste in the pathname you copied. Add a forward slash to the end of this text, resulting in *metaspadesfolder = /Applications/SPAdes-3.14.0-Darwin/bin/*

8.25 In order for **MitoFinder** to successfully annotate tRNA, you must install **MiTFi**, **tRNAscan-SE**, and **ARWEN**.

Chan, P.P., & Lowe, T.M. (2019). tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. In: Kollmar M. (eds) Gene Prediction. Methods in Molecular Biology, 1962, 1-14. http://10.1007/978-1-4939-9173-0_1

Jühling, F., Pütz, J., Bernt, M., Donath, A., Middendorf, M., Florentz, C., & Stadler, P.F. (2011). Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. Nucleic Acids Research, 40(7), 2833-2845. http://10.1093/nar/gkr1131

Laslett, D., & Canbäck, B. (2007). ARWEN: A program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics, 24(2), 172-175. http://10.1093/bioinformatics/btm573

The **MiTFi** and **tRNAscan-SE** packages were simultaneously installed by **MitoFinder** when executing the "./install.sh" command. However, you must make sure that an updated version of **Java** is installed on your Mac in order for **MiTFi** and **tRNAscan-SE** to run properly. Open up *System Preferences* and click on the *Java* icon. This will open up the *Java Control Panel* in a separate window. If it states that a *critical Java security update is available* under the *Update* tab, then click the *Update Now* button. Alternatively, if it states that *your system has the recommended version of Java* under the *Update* tab, then you are good to go and may close out of the *Java Control Panel.*

Under the *tRNA annotation ARWEN* section on the GitHub web page, it states that you must compile the **ARWEN** source code on your MAC OS system using **gcc**. Copy, paste, and execute (click *return* on your keyboard) the following commands into the same **Terminal** shell window (one at a time); once a command has successfully terminated, copy, paste, and execute the subsequent command. After completion of the last command, you should see a series of command usages for **ARWEN** in the **Terminal**.

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

cd arwen

cd /Applications/MitoFinder/arwen/

Chooses the arwen folder as your new working directory in Terminal using the general "cd" command.

gcc arwen1.2.3.c

gcc arwen1.2.3.c

GNU Compiler Collection (gcc) performs the compilation step to build a program (ARWEN in this case), and then it calls other programs to assemble the program and to link the program's component parts into an executable program that you can run.

mv a.out arwen

mv a.out arwen

Moves the GNU compiler traditional/conventional default program compilation result name (a.out) into the arwen folder.

./arwen -h

./arwen -h

Expands the help directory of arwen, presenting a series of command usages for ARWEN in the Terminal shell window.

After the last command terminates, close this **Terminal** shell window.

8.26 Expand the *MitoFinder > trnascanSE* folder in your *Applications*. Right click (or *control click* on your Mac keyboard) on the *tRNAscan-SE-2.0* folder. Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "tRNAscan-SE-2.0" as Pathname*.

8.27 In your *Mitofinder.config* file, delete *default* after *trnascanfolder =* and paste in the pathname you copied. Add a forward slash to the end of this text, resulting in *trnascanfolder = /Applications/MitoFinder/trnascanSE/tRNAscan-SE-2.0/*

8.28 Expand the *MitoFinder* folder in your *Applications*. Right click (or *control click* on your Mac keyboard) on the *mitfi* folder. Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "mitfi" as Pathname*.

8.29 In your *Mitofinder.config* file, delete *default* after *mitfifolder =* and paste in the pathname you copied. Add a forward slash to the end of this text, resulting in *mitfifolder = /Applications/MitoFinder/mitfi/*

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

8.30 Save (do not save as or change the file name) the edited *Mitofinder.config* file in its original location (within the *MitoFinder* folder).

8.31 Before analyzing your data, you must also download all the published mitogenomes that are closely related to your animal available on GenBank. Go to NCBI (https://www.ncbi.nlm.nih.gov/). Search the *Nucleotide* database for *"Mitochondrion" AND "Annelida"* (or your respective animal group if it's not Annelida). Click *RefSeq* on the left-hand side of the resulting search page. This filters out the incomplete mitogenomes and results in only those that are complete and have been properly annotated. Click *Send to:* in the upper right-hand corner. Then select the following options: *Complete Record*, *File* under *Choose Destination*, and *GenBank* under *Format*. Then click *Create File*. This is the file that you will use for **MetaSPAdes** to run properly, which utilizes the mitochondrial genomes of closely related animals to the species of interest as a reference for proper gene identification and annotation. Make sure you rename this *sequence.gb* file and transfer it from your *Downloads* to the working directory (aka folder) for your mitogenome assemblies.

NCBI. National Center for Biotechnology Information.
https://www.ncbi.nlm.nih.gov/

***Never use spaces, punctuation, symbols, or special characters (periods, commas, parentheses, ampersands, asterisks, etc.) in your custom folder and file names! "Some of these symbols are used in operating systems to perform certain tasks, such as to identify folder levels in Microsoft products and Mac operating systems. Periods are used to identify file formats such as .jpg and .doc" (Bethcron 2017).

***Use hyphens and/or underscores instead of spaces, punctuation, symbols, or special characters.

Bethcron (2017). Best Practices for File Naming. National Archives: Records Express.
https://records-express.blogs.archives.gov/2017/08/22/best-practices-for-file-naming/

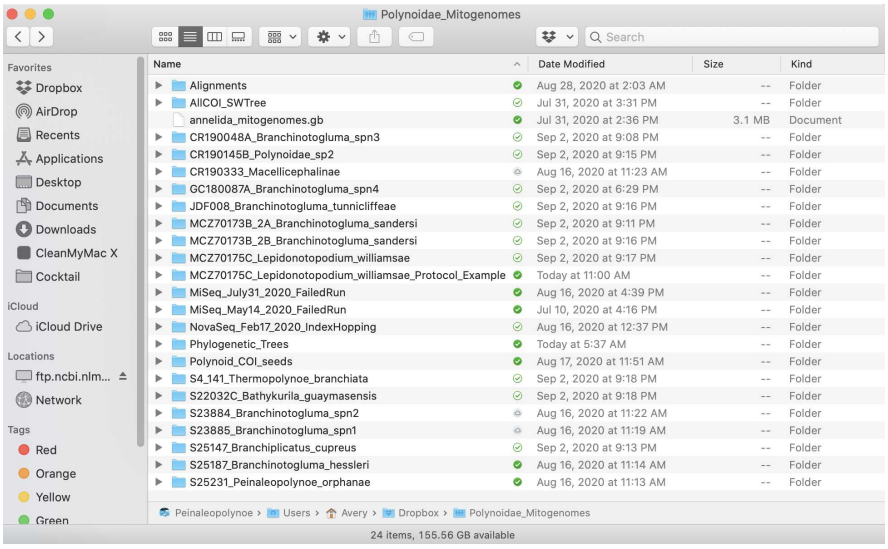
You must create your own working directory (aka folder) where you would like to store, organize, and analyze your genome skimming datasets within. I recommend utilizing a location within **Dropbox**, to ensure that your data is always backed up and to save storage space on your hard drive with the **Smart Sync** feature.

Dropbox. Dropbox Business.
http://www.dropbox.com/?landing=dbv2

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

Dropbox. Smart Sync lets you work without limits.
https://www.dropbox.com/smart-sync

For example, the pathway to my general working directory is */Users/Avery/Dropbox/Polynoidae_Mitogenomes*, and a screenshot showing the contents of the corresponding *Polynoidae_Mitogenomes* folder is attached below for your reference. The *annelida_mitogenomes.gb* file is the renamed file downloaded from GenBank.



Ultimately, you may organize your data according to your own personal preference, but you must ensure that it is easy to navigate within your working directory! You will most likely have several analyses for each species, which is why I recommend creating subfolders for each species or unique dataset, containing its corresponding raw paired-end reads and all resulting output files from the multiple analyses executed.

9 Download **NOVOPlasty Version 4.2** by following the subsequent steps.

NOVOPlasty Version 4.2
source by doi: 10.1093/nar/gkw955

Dierckxsens, N., Mardulyn, P., & Smits, G. (2016). NOVOPlasty: De novo assembly of organelle genomes from whole genome data. Nucleic Acids Research, 45(4), E18.
http://10.1093/nar/gkw955

Dierckxsens, N., Mardulyn, P., & Smits, G. (2020). NOVOPlasty (Version 4.2) [Computer software]. GitHub.
https://github.com/ndierckx/NOVOPlasty

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

Due to **MitoFinder**'s initial design for use with smaller, lower coverage NGS datasets, and the higher computation time for **MetaSPAdes**; the **MitoFinder** pipeline tends to get overloaded from paired-end input datasets larger than 7 M reads each (14 M reads total when the paired-end data is interleaved)-- based on personal experience running on a Mac desktop with a 4.2 GHz Quad-Core Intel Core i7 processor and 32 GB 2400 MHz DDR4 memory. Consequently, we downsample our data in order to efficiently run **MitoFinder** with **MetaSPAdes**, which may result in sub-optimal organelle coverage (typically we hope to achieve around 100X coverage). Therefore, it may also be useful to assemble the mitogenome with **NOVOPlasty** using your full dataset, to determine the maximum mitogenome coverage that can be obtained from your original reads (not downsampled)-- and to see if the resulting **NOVOPlasty** contig does or doesn't align perfectly to the **MitoFinder** contig, allowing you to compare two different pipelines for assembling the mitochondrial genome. Hence, we will proceed to download **NOVOPlasty Version 4.2**.

In this protocol, we are utilizing **NOVOPlasty** solely for the two aforementioned reasons. However, **NOVOPlasty** pales in comparison to **MitoFinder**, especially when looking at accuracy of the resulting mitogenome assembly and organization & utility of output files. Most importantly, **NOVOPlasty** does not identify and annotate the 37 mitochondrial genes (13 protein coding genes [PCGs], 22 tRNAs, and 2 rRNAs) as **MitoFinder** with **MetaSPAdes** does. Furthermore, the majority of handy **MitoFinder** output files mentioned in step 8 are not generated with **NOVOPlasty**, which only results in the full-length, unannotated fasta sequence of the mitogenome in addition to a few stats-- such as the bp length of the contig, average library insert size, total reads analyzed, number of aligned and assembled reads, and the average organelle coverage. That being said, you will most likely use the results generated from **MitoFinder** in your official publication, while also reporting the maximum organelle coverage determined via **NOVOPlasty** as supplementary data-- only if the resulting mitogenome contigs from **MitoFinder** and **NOVOPlasty** are identical.

***If you have significant mitogenome coverage with **MitoFinder** (which is more often the case), you do not have to use **NOVOPlasty** at all and can completely bypass section VIII.

9.1 Click on the following link: https://github.com/ndierckx/NOVOPlasty

9.2 Click the green *Code* button on the right hand side and *Download ZIP*.

9.3 Transfer the resulting *NOVOPlasty-master* folder from your *Downloads* to your *Applications* folder.

II. Choose Working Directory

10 Optionally, follow along with this supplementary video as you perform section *II. Choose Working Directory:*

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

Open up a new text editor window (e.g. **BBEdit**, **TextWrangler**, or **TextEdit**), in which you will edit all your commands prior to pasting and executing them in **Terminal**. For more complex commands, it is easier to both edit and check for mistakes in the corresponding commands in a text editor prior to pasting them into **Terminal**.

Save this text file in the corresponding working directory for the species/sample of interest. Label the command name above each revised command, which will be tailored specifically to your data. Reference this text file when writing up the methods in your manuscript, as it contains all the bioinformatic steps executed (e.g. overall workflow, including the parameters and any adjustments from default settings) for the given species.

10.1 Copy and paste the following "cd working directory" command into the text editor window.

cd working directory

cd /path/to/folder/of/interest

The cd command, also known as chdir, is a command-line shell command used to change the current working directory in various operating systems.

Neagu, C. (2020). Command Prompt: 11 basic commands you should know (cd, dir, mkdir, etc.). Digital Citizen. http://www.digitalcitizen.life/command-prompt-how-use-basic-commands

10.2 In **Finder**, navigate to (but do not expand the contents of) the folder that you created and named, where you want the resulting output files for the current sample you are working on to be saved.

***Reminder: Never use spaces, punctuation, symbols, or special characters (periods, commas, parentheses, ampersands, asterisks, etc.) in your custom folder and file names-- use hyphens

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

and/or underscores instead!

If you followed the recommended file organization presented in step 8.31, then navigate to and expand your general working directory folder in **Dropbox** (in the given example, this is the *Polynoidae_Mitogenomes* folder).

10.3 Right click (or *control click* on your Mac keyboard) on the associated folder for the species you are analyzing (e.g. *MCZ70175C_Lepidonotopodium_williamsae_Protocol_Example* located within the general *Polynoidae_Mitogenomes* working directory folder). Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "Folder_Name" as Pathname*.

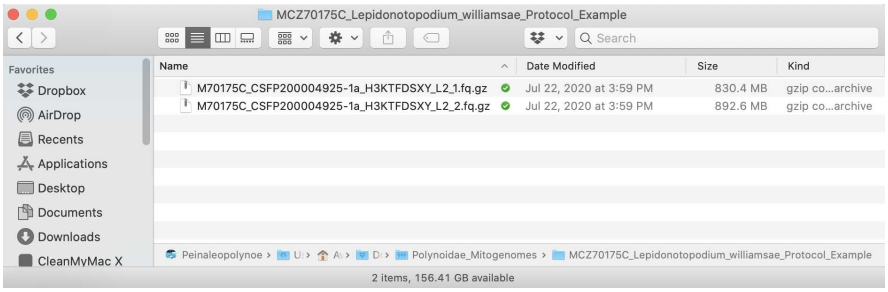
10.4 In the text editor window, delete */path/to/folder/of/interest* and paste in the true pathname you just copied.

Select and copy the revised command text.

10.5 Open up a new **Terminal** shell window.

Do not close this **Terminal** shell window until instructed to do so!

10.6 Paste and execute (click *return* on your keyboard) the revised "cd working directory" command in the **Terminal** shell window in order to officially choose your working directory for the remaining steps. Before proceeding, open/expand the chosen folder in **Finder**, and double check that you have saved the corresponding paired-end reads for this sample within the folder. A screenshot of the NGS data files is attached below for your reference:



If the NGS genome skimming, paired-end data was obtained from Novogene, there should be two files total (ending in *1.fq.gz* and *2.fq.gz*).

III. Simple Stats with SeqKit

11 Optionally, follow along with this supplementary video as you perform section *III. Simple Stats with SeqKit:*

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

Check the data statistics with **SeqKit**, namely the number and length of reads, by following the subsequent steps.

11.1 Copy and paste the following "seqkit stats" command into the text editor window:

seqkit stats

/Applications/seqkit stats /path/to/file/of/interest

View the simple statistics of FASTA/Q files

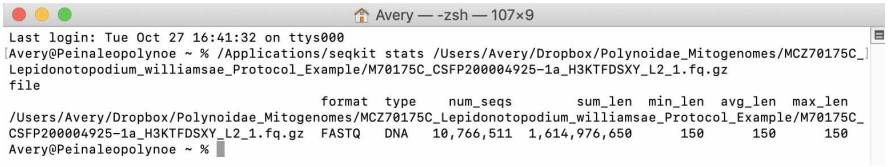
Shen, W. (2020). SeqKit (Version 0.13.2) [Computer software]. GitHub. https://github.com/shenwei356/seqkit/releases/tag/v0.13.2

11.2 In **Finder**, right click (or *control click* on your Mac keyboard) on either of the paired-end, gzipped, fastq files (.fq.gz) inside of the working directory folder you specified in section II. Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "File_Name.fq.gz" as Pathname*.

11.3 In the text editor window, delete */path/to/file/of/interest* and paste in the true pathname you just copied.

11.4 Copy, paste and execute (click *return* on your keyboard) the revised "seqkit stats" command in the **Terminal** shell window. A screenshot of the resulting output for *Lepidonotopodium williamsae* is attached below for your reference:

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw



Resulting output:

- format: should be FASTQ
- type: should be DNA
- num_seqs: number of reads (paired)
- sum_len: total length of all reads
- min_len: length of shortest read
- avg_len: average length of reads
- max_len: length of longest read

11.5 Create (and save) a spreadsheet to record these statistics, which are relevant for use in your manuscript. Make sure that the first three column headers are:

1. *Sample ID;* this is the lab code for the original gDNA extraction (e.g. *MCZ70175C*)
2. *Specimen Voucher;* this is the official catalog number for the exact specimen that was sequenced, which will most likely be deposited at SIO-BIC
3. *Species;* e.g. *Lepidonotopodium williamsae*

SIO-BIC. Benthic Invertebrate Collection. Scripps Institution of Oceanography.
https://sioapps.ucsd.edu/collections/bi/

The most important column headers to add for the corresponding **SeqKit** output are the following:

4. *Total Number of Raw Reads (Paired-End)*; this corresponds to the output for *num_seqs*
5. *Total Length of Raw Reads (Paired-End)*; this corresponds to the output for *sum_len*
6. *Individual Length of Raw Reads*; this corresponds to the output for *avg_len*

As you analyze additional samples using this protocol, add their corresponding **SeqKit** statistics to this spreadsheet.

IV. Clean and Trim Reads with Trimmomatic

12 Optionally, follow along with this supplementary video as you perform section *IV. Clean and Trim Reads with Trimmomatic:*

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

Clean and trim your raw reads with **Trimmomatic** by following the subsequent steps.

12.1 Copy and paste the following "Trimmomatic Paired End Mode" command into the text editor window.

Trimmomatic Paired End Mode

java -jar /Applications/Trimmomatic-0.39/trimmomatic-0.39.jar PE -phred33 /path/to/1.fq.gz/file /path/to/2.fq.gz/file sample_P1.fastq.gz sample_U1.fastq.gz sample_P2.fastq.gz sample_U2.fastq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

"Trimmomatic is a fast, multithreaded command line tool that can be used to trim and crop Illumina (FASTQ) data as well as to remove adapters. These adapters can pose a real problem depending on the library preparation and downstream application... For paired-end data, two input files, and 4 output files are specified, 2 for the 'paired' output where both reads survived the processing, and 2 for corresponding 'unpaired' output where a read survived, but the partner read did not" (Bolger et al. 2014).

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114-2120.
http://10.1093/bioinformatics/btu170

Bolger, A. M., Lohse, M., & Usadel, B. (2019). Trimmomatic (Version 0.39) [Computer software]. USADELLAB.org.
http://www.usadellab.org/cms/?page=trimmomatic

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

Descriptions for the trimming steps & parameters:

- PE: specifies that the data is paired-end
- -phred33: specifies the base quality encoding
- sample_P1.fastq.gz: the forward paired output file generated
- sample_U1.fastq.gz: the forward unpaired output file generated
- sample_P2.fastq.gz: the reverse paired output file generated
- sample_U2.fastq.gz: the reverse unpaired output file generated
- ILLUMINACLIP:TruSeq3-PE.fa:2:30:10: removes adapters
- LEADING:3: removes leading low quality or N bases (below quality 3)
- TRAILING:3: removes trailing low quality or N bases (below quality 3)
- SLIDINGWINDOW:4:15: scans the read with a 4-base wide sliding window, cutting when the average quality per base drops below 15
- MINLEN:36: drops reads below the 36 bases long

12.2 In **Finder**, right click (or *control click* on your Mac keyboard) on the 1.fq.gz file. Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "file_name_1.fq.gz" as Pathname*. In the text editor window, delete */path/to/1.fq.gz/file* and paste in the true pathname you just copied.

In **Finder**, right click (or *control click* on your Mac keyboard) on the 2.fq.gz file. Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "file_name_2.fq.gz" as Pathname*. In the text editor window, delete */path/to/2.fq.gz/file* and paste in the true pathname you just copied.

12.3 In the text editor window, replace the four occurrences of *sample* in *sample_P1.fastq.gz sample_U1.fastq.gz sample_P2.fastq.gz sample_U2.fastq.gz* with your true sample ID (e.g. *MCZ70175C*).

12.4 Copy the revised "Trimmomatic Paired End Mode" command in the text editor window. Then paste and execute (click *return* on your keyboard) it in the **Terminal** shell window.

12.5 Check the output for the number & percentage of reads (*Both Surviving*) retained after the "Trimmomatic Paired End Mode" command terminates. Optimally, this should be around 95% or above; the quality of your data is questionable if this percentage is much lower (e.g. 80%). See the screenshot below for your reference:

```
Input Read Pairs: 10766511 Both Surviving: 10429969 (96.87%) Forward Only Surviv
ing: 258332 (2.40%) Reverse Only Surviving: 57095 (0.53%) Dropped: 21115 (0.20%)
TrimmomaticPE: Completed successfully
Avery@Peinaleopolynoe ~ %
```

Both Surviving from the **Trimmomatic** output text in **Terminal** means that the same exact reads (complementary to one another) are chosen from both P1 (forward) and P2 (reverse) files.

12.6 Add the column header *Number of Reads Post-Trimmomatic (Paired-End)* to the spreadsheet you created in step 11.5 and record the output for *Both Surviving*. This is important because **MitoFinder** will be executed with the trimmed/cleaned reads as opposed to the raw reads, so this statistic must be included in your manuscript.

V. Downsample Reads with MITObim

13 Optionally, follow along with this supplementary video as you perform section *V. Downsample Reads with MITObim:*

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

Conditional: MitoFinder has difficulty processing more than approximately 7 million reads paired-end (14 million reads interleaved). If your cleaned and trimmed output from step 12.5 is greater than 7 million reads, you should downsample your data with **MITObim**'s downsample.py function to achieve around this number (or even less) of paired-end reads.

14 Copy and paste the following "MITObim's downsample.py" command into the text editor window:

MITObim's downsample.py

/Applications/MITObim-master/misc_scripts/downsample.py -s 50 --interleave -r /path/to/sample_P1.fastq.gz/file -r /path/to/sample_P2.fastq.gz/file | gzip > sample_50p_Interleaved.fastq.gz

MITObim's downsample.py function randomly downsamples paired-end reads to a specified percentage of the inputted P1 and P2 reads. It produces an interleaved gzipped fastq file.

Optional: Reads are randomly downsampled using a seed. You can try changing the seed for the random number generator to re-do this analysis if needed (e.g. if by chance you don't get enough mitochondrial reads). To do this, define a seed for the random number generator to allow for replication of the process (downsampling again at the exact same percentage of reads you specified) by appending *--rand -421039* (you can choose any 6 digit number) to the aforementioned "MITObim's downsample.py" command.

*You will not know whether or not to try defining a random seed until you have tried to run **MitoFinder** at least once with the downsampled reads from **MITObim**.

Hahn, C., Bachmann, L., & Chevreux, B. (2018). MITObim (Version 1.9.1) [Computer software]. GitHub. https://github.com/chrishah/MITObim

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

14.1 In the text editor window, delete *50* and replace it with the percentage of your choice that you would like to downsample your dataset to.

In **Finder**, right click (or *control click* on your Mac keyboard) on the sample_P1.fastq.gz file. Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "sample_P1.fastq.gz" as Pathname*. In the text editor window, delete */path/to/sample_P1.fastq.gz/file* and paste in the true pathname you just copied.

In **Finder**, right click (or *control click* on your Mac keyboard) on the sample_P2.fastq.gz file. Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "sample_P2.fastq.gz" as Pathname*. In the text editor window, delete */path/to/sample_P2.fastq.gz/file* and paste in the true pathname you just copied.

Delete *sample* and *50* in *sample_50p_Interleaved.fastq.gz* and replace with your true sample ID and the percentage you decided to downsample to. For example, 50p stands for 50%.

14.2 Copy the revised "MITObim's downsample.py" command in the text editor window. Then paste and execute (click *return* on your keyboard) it in the **Terminal** shell window.

14.3 Once the "MITObim's downsample.py" command has successfully terminated, copy and paste the "seqkit stats" command again into the text editor window:

seqkit stats

/Applications/seqkit stats /path/to/file/of/interest

View the simple statistics of FASTA/Q files

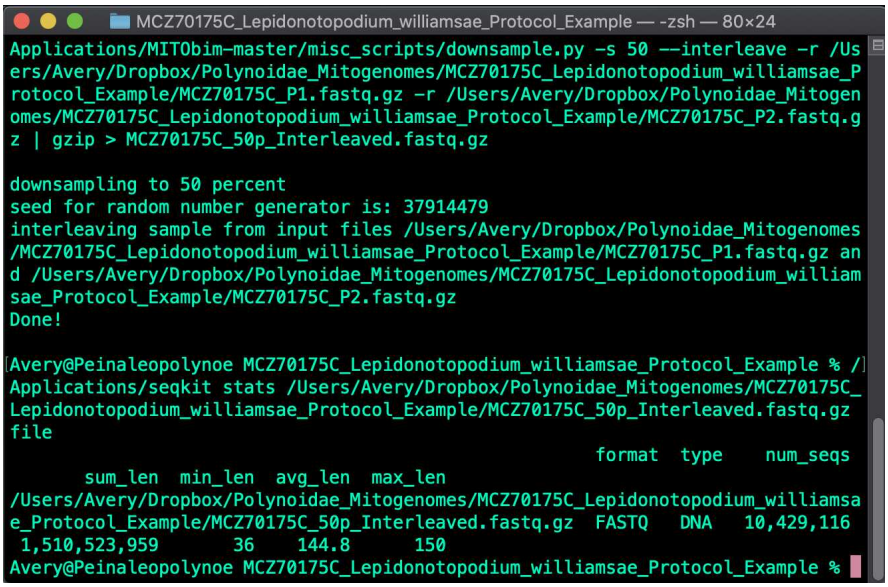
Shen, W. (2020). SeqKit (Version 0.13.2) [Computer software]. GitHub. https://github.com/shenwei356/seqkit/releases/tag/v0.13.2

14.4 In **Finder**, right click (or *control click* on your Mac keyboard) on the Interleaved.fastq.gz file created by **MITObim**'s downsample.py function. Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "sample...Interleaved.fastq.gz" as Pathname*.

14.5 In the text editor window, delete */path/to/file/of/interest* and paste in the true pathname you just copied.

14.6 Copy, paste and execute (click *return* on your keyboard) the revised "seqkit stats" command in the **Terminal** shell window. A screenshot of the resulting output for *Lepidonotopodium williamsae* is attached below for your reference:

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw



In the example screenshot, the number of reads I will report on my spreadsheet will be *5,214,558* (as opposed to *10,429,116*).

The **SeqKit** results for the interleaved file will present the number of sequences as double what the real value is when analyzing the reads as paired-end. Therefore, divide the output for *num_seqs* by 2, and report the resulting number in step 14.7 (see below).

*Optionally, you can confirm this for extra command line practice after section VI by repeating "seqkit stats" again on one of the two (either forward or reverse) files generated from **BBMap**'s reformat.sh function.

14.7 Add the column header *Number of Reads Post-Downsampling (Paired-End)* to the spreadsheet you created in step 11.5. Record the number of paired-end reads surviving after downsampling (e.g. *5,214,558 (50% of cleaned/trimmed reads)*). Make sure to specify the percentage of the cleaned/trimmed reads that you downsampled to, because you may have to repeat section V and troubleshoot with different percentages if **MitoFinder** fails initially.

*This statistic is critical to transcribe, because you must correctly report the amount of reads you analyzed for successful mitogenome assembly in your manuscript! Since you decided to downsample, that means you will be using a different total number of reads to run through the **MitoFinder** pipeline than your original raw data set contained.

Polynoidae NGS Genome Skimming Data Statistics

Sample ID	Specimen Voucher	Species	Total Number of Raw Reads (Paired-End)	Total Length of Raw Reads (Paired-End)	Individual Length of Raw Reads	Number of Reads Post-Trimmomatic (Paired-End)	Number of Reads Post-Downsampling (Paired-End)
MCZ70175C	MCZ 70175	*Lepidonotopodium williamsae*	10,766,511	1,614,976,650	150	10,429,969 (96.87% of total no. raw reads)	5,214,558 (50% of cleaned/ trimmed reads)

Example screenshot of the data statistics for *MCZ70175C* recorded in my *Polynoidae NGS Genome Skimming Data Statistics* spreadsheet.

VI. Reformat Reads to Non-Interleaved with BBMap

15 Optionally, follow along with this supplementary video as you perform section *VI. Reformat Reads to Non-Interleaved with BBMap:*

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

Conditional: In the previous section, your reads had to be interleaved in order to correctly sample the same percentage of matching, complementary forward (P1) and reverse (P2) reads. **MitoFinder** requires non-interleaved, paired-end reads as input, so reformat your reads back into separate forward and reverse files with **BBMap**'s reformat.sh function.

Section VI only needs to be done if the conditional section V was also completed. However, if you bypassed section V, you may skip ahead to section VII.

16 Copy and paste the following "BBMap's reformat.sh" command into the text editor window:

BBMap's reformat.sh

/Applications/bbmap/reformat.sh in=/path/to/Interleaved.fastq.gz/file out1=sample_50p_For.fastq.gz out2=sample_50p_Rev.fastq.gz

Reformats an interleaved fastq file (input 1 file) back into separate paired-end fastq files (outputs 2 files).

Bushnell, B. (2020). BBMap (Version 38.87) [Computer software]. SourceForge.
http://sourceforge.net/projects/bbmap/files

16.1 In **Finder**, right click (or *control click* on your Mac keyboard) on the Interleaved.fastq.gz file. Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "sample...Interleaved.fastq.gz" as Pathname*. In the text editor window, delete */path/to/Interleaved.fastq.gz/file* and paste in the true pathname you just copied.

Delete *sample* and *50* in both *sample_50p_For.fastq.gz* and *sample_50p_Rev.fastq.gz*. Replace *sample* with your true sample ID. Replace *50* with the percentage chosen in section V.

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

16.2 Copy the revised "BBMap's reformat.sh" command in the text editor window. Then paste and execute (click *return* on your keyboard) it in the **Terminal** shell window.

As mentioned in step 14.6, you may optionally execute one of the **BBMap** output files (either the For.fastq.gz or Rev.fastq.gz file) with "seqkit stats" after the "BBMap's reformat.sh" command successfully terminates, to confirm that the number of reads is indeed half of what was stated for the interleaved file.

*This is mainly for additional practice.

VII. Mitogenome Assembly and Annotation: MitoFinder with MetaSPAdes

17 Optionally, follow along with this supplementary video as you perform section *VII. Mitogenome Assembly and Annotation: MitoFinder with MetaSPAdes:*

If you are analyzing paired-end input datasets larger than approximately 4 M reads each (8 M reads total when the paired-end data is interleaved), then I would recommend performing section VII on one of the three Rouse Lab Mac desktop computers, which have 64-bit processors and 32 GB memory-- allowing you to analyze up to ~7 M reads paired-end (14 M reads interleaved). I will teach you privately how to log in to one of these computers remotely, so you may still work from home.

*This is another reason why it is convenient to store, organize, and analyze your data in a **Dropbox** folder; you'll be able to access your **Dropbox** folder from any computer and set it as your working directory.

However, if your personal computer has 32 GB memory as well, disregard this note.

Furthermore, Marina is currently in the process of coordinating Rouse Lab use of the UCSD supercomputers. Once she finishes setting up the interface and downloading the required programs and dependencies, you will also be able to optionally run **MitoFinder** with the total trimmed/cleaned reads (no downsampling necessary) from section IV, because the UCSD supercomputers have much greater computing power than even the Rouse Lab Mac desktops do.

***However, if you are already achieving more than 100X mitogenome coverage from running **MitoFinder** on your downsampled, trimmed/cleaned reads, it is not necessary to analyze your total trimmed/cleaned reads on the supercomputer, nor is it necessary to complete the **NOVOPlasty** section VIII below.

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

Copy and paste the following "MitoFinder with MetaSPAdes assembler" command into the text editor window:

MitoFinder with MetaSPAdes assembler

/Applications/MitoFinder/mitofinder --metaspades -t mitfi -j sample_MitoFinder_50p -1 /path/to/For.fastq.gz/file -2 /path/to/Rev.fastq.gz/file -r /path/to/gb/file -o 5

Assembles and annotates the mitochondrial genome using the input of two cleaned/trimmed, paired-end fastq files and a reference GenBank file containing mitogenomes (complete and annotated) of closely related animals to the species of interest.

Allio, R., Schomaker-Bastos, A., Romiguier, J., Prosdocimi, F., Nabholz, B., & Delsuc, F. (2020). MitoFinder (Version 1.4) [Computer software]. GitHub. https://github.com/RemiAllio/MitoFinder

17.1 In the text editor window, delete *sample* and *50* in *sample_MitoFinder_50p*. Replace *sample* with your true sample ID, and replace *50* with the actual percentage you downsampled your reads to. In the given example, this results in *MCZ70175C_MitoFinder_50p*, since I did indeed downsample to 50%.

In **Finder**, right click (or *control click* on your Mac keyboard) on the For.fastq.gz file. Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "sample...For.fastq.gz" as Pathname*. In the text editor window, delete */path/to/For.fastq.gz/file* and paste in the true pathname you just copied.

In **Finder**, right click (or *control click* on your Mac keyboard) on the Rev.fastq.gz file. Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "sample...Rev.fastq.gz" as Pathname*. In the text editor window, delete */path/to/Rev.fastq.gz/file* and paste in the true pathname you just copied.

In **Finder**, navigate to the .gb file (containing the downloaded full mitogenomes of closely related species) that you created and right click (or *control click* on your Mac keyboard) on the file name. Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "file_name.gb" as Pathname*. In the text editor window, delete */path/to/gb/file* and paste in the true pathname you just copied.

The *5* following *-o* refers to the genetic code used for accurate translation of your 13 mitochondrial PCGs (e.g. 5 is the code for Invertebrate Mitochondrial DNA; 9 is the code for Echinoderm and Flatworm Mitochondrial DNA). Check the following link in case your code varies: https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi, but you will most likely be using *5* in the Rouse Lab.

Elzanowski, A., & Ostell, J. (2019). The Genetic Codes. NCBI. https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi

tRNA annotation step

Second, you can choose the tool for the tRNA annotation step of MitoFinder using the -t option:

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

-t mitfi (default, MiTFi: slower but really efficient)
-t arwen (ARWEN: faster)
-t trnascan (tRNAscan-SE)

In the "MitoFinder with MetaSPAdes assembler" command, the *-t mitfi* option is selected as the tRNA annotation step, because it is really efficient. However, the previous version of **MitoFinder** used **ARWEN** by default, which also worked great with my data. If the "MitoFinder with MetaSPAdes assembler" command hasn't terminated within two days, you may experiment with changing *-t mitfi* to *-t arwen* since it executes faster.

Allio, R., Schomaker-Bastos, A., Romiguier, J., Prosdocimi, F., Nabholz, B., & Delsuc, F. (2020). MitoFinder (Version 1.4) [Computer software]. GitHub.
https://github.com/RemiAllio/MitoFinder

17.2 Copy the revised "MitoFinder with MetaSPAdes assembler" command in the text editor window. Then paste and execute (click *return* on your keyboard) it in the **Terminal** shell window.

17.3 Optimally, **MitoFinder** with **MetaSPAdes** will find 15 genes (13 PCGs and 2 rRNAs) in the mtDNA contig upon successful termination of the command executed.

The "MitoFinder with MetaSPAdes assembler" command could potentially take a day and a half to terminate, so be patient while you await results. Once the **MitoFinder** with **MetaSPAdes** pipeline has terminated, close the **Terminal** shell window in which you executed the corresponding commands for sections II to VII.

Within your specified working directory, **MitoFinder** will create a folder named after the revised *sample_MitoFinder_50p* text specified after *-j* in the "MitoFinder with MetaSPAdes assembler" command (e.g. *MCZ70175C_MitoFinder_50p*). Within this folder, there will be an additional *Final_Results* folder containing the following output files:

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

OUTPUT FILES

Results' folder

Mitofinder returns several files for each mitochondrial contig found:

- **[Seq_ID]_final_genes_NT.fasta** containing the nucleotides sequences of the final genes selected from all contigs found by MitoFinder
- **[Seq_ID]_final_genes_AA.fasta** containing the amino acids sequences of the final genes selected from all contigs found by MitoFinder
- **[Seq_ID]_mtDNA_contig.fasta** containing a mitochondrial contig
- **[Seq_ID]_mtDNA_contig.gff** containing the final annotation for a given contig (GFF3 format)
- **[Seq_ID]_mtDNA_contig.tbl** containing the final annotation for a given contig (Genbank submission format)
- **[Seq_ID]_mtDNA_contig.gb** containing the final annotation for a given contig (Genbank format for visualization)
- **[Seq_ID]_mtDNA_contig_genes_NT.fasta** containing the nucleotide sequences of annotated genes for a given contig
- **[Seq_ID]_mtDNA_contig_genes_AA.fasta** containing the amino acids sequences of annotated genes for a given contig
- **[Seq_ID]_mtDNA_contig.png** schematic representation of the annotation of the mtDNA contig
- **[Seq_ID]_mtDNA_contig.infos** containing the initial contig name, the length of the contig and the GC content

Additionally, **MitoFinder** will create a *MitoFinder.log* file and a *.input* file in your specified working directory. The *MitoFinder.log* file is especially handy for future reference, as it contains all the bioinformatic steps executed through the **MitoFinder** pipeline.

Allio, R., Schomaker-Bastos, A., Romiguier, J., Prosdocimi, F., Nabholz, B., & Delsuc, F. (2020). MitoFinder (Version 1.4) [Computer software]. GitHub. https://github.com/RemiAllio/MitoFinder

17.4 **File Cleanup:** Before proceeding, open the chosen working directory from section II (e.g. *MCZ70175C_Lepidonotopodium_williamsae_Protocol_Example*) in **Finder**.

If you have to troubleshoot and run several **MitoFinder** analyses for the species/sample of interest-- with different parameters and/or input data sets (downsampled to various percentages)-- then file organization can get messy if you do not follow these suggested file cleanup steps.

Transfer the two aforementioned files into the *sample_MitoFinder_...p* folder, in order to keep all the output files associated with this specific **MitoFinder** analysis in one location.

17.5 Double check that you have saved the text file containing all the revised commands executed from section II to section VII for the species/sample analyzed. In **Finder**, transfer this text file into the

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

sample_MitoFinder_...p folder since it is specific to this analysis.

*Reminder: Reference this text file when writing up the methods in your manuscript, as it contains all the **MitoFinder** bioinformatic steps executed (e.g. overall workflow, including the parameters, arguments, and any adjustments from default settings) for the given species.

17.6 In **Finder**, transfer the *sample...Interleaved.fastq.gz*, *sample...For.fastq.gz*, and *sample...Rev.fastq.gz* files into the *sample_MitoFinder_...p* folder since they are also specific to this analysis.

17.7 Check the basic mitogenome statistics obtained from the **MitoFinder** analysis:

In **Finder**, open the *sample_MitoFinder_...p* folder in your working directory. Then open the *Final_Results* folder. Drag the *sample_MitoFinder_...p.infos* file into a text editor, in which you'll be able to view the length of the contig (bp), coverage, GC content, and circularization.

On your own time, create a new spreadsheet to record the aforementioned **MitoFinder** statistics (along with others, such as gene order and number of total genes identified/annotated) for all samples/species analyzed in a given project.

VIII. Mitogenome Assembly with NOVOPlasty

18 Optionally, follow along with this supplementary video as you perform section *VIII. Mitogenome Assembly with NOVOPlasty:*

Conditional/Optional: If you obtain significant mitogenome coverage with **MitoFinder** (100 or greater), you may skip this section. Alternatively, if your assembly had lower than optimal mitogenome coverage, you may try either or all of the following options:

- Downsample your trimmed/cleaned reads to a greater percentage, resulting in a maximum of ~7 million reads paired-end (14 million reads interleaved), and rerun the **MitoFinder** analysis with this revised data set on a computer with a 64-bit processor and 32 GB memory (e.g. Rouse Lab Mac desktops).
- Do not downsample your trimmed/cleaned reads at all, and rerun the **MitoFinder** analysis with the **Trimmomatic** paired-end output files (*sample_P1.fastq.gz* and *sample_P2.fastq.gz*) on the UCSD supercomputer, which has even greater computing power than the Rouse Lab Mac desktops. This will determine the maximum mitogenome coverage that can be obtained from ~95% of your original data (~5% was lost to **Trimmomatic**).
- Downsample your trimmed/cleaned reads to the same percentage, but append to the "MITObim's downsample.py"

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

command a defined, 6 digit seed for the random number generator, in case you originally obtained a smaller than optimal percentage of mitochondrial reads by chance. Rerun the **MitoFinder** analysis with this revised data set.
- Assemble the mitogenome with **NOVOPlasty** using your raw reads as input, to determine the maximum mitogenome coverage that can be obtained from your original data.

Furthermore, if you would like to practice using a different pipeline and/or are genuinely curious to compare the resulting mitogenome contigs from **MitoFinder** and **NOVOPlasty**, then proceed to section VIII.

In **Finder**, open the chosen working directory from section II (e.g. *MCZ70175C_Lepidonotopodium_williamsae_Protocol_Example*). Create a new folder within this directory, named *sample_NOVOPlasty* (e.g. *MCZ70175C_NOVOPlasty*).

You must have a COI Sanger Sequence (FASTA format) obtained from the exact gDNA extraction of the species/sample of interest if you perform section *VIII. Mitogenome Assembly with NOVOPlasty.*

19 Open a new **Finder** window, and open the *NOVOPlasty-master* folder in your *Applications*.

20 Copy the *config.txt* file and paste it into the *sample_NOVOPlasty* folder that you just created.

Rename this file as *sample_config.txt* (e.g. *MCZ70175C_config.txt*).

20.1 Drag the *sample_config.txt* file into a text editor. Change the *Project name* from *Test* to *sample_species* (e.g. *MCZ70175C_Lepidonotopodium_williamsae*).

Delete */path/to/reference_file/reference.fasta (optional)* and */path/to/chloroplast_file/chloroplast.fasta (only for "mito_plant" option)*.

20.2 In **Finder**, navigate to the location of the COI sequence (FASTA format) for the corresponding species/sample of interest. Right click (or *control click* on your Mac keyboard) on the .fasta file for your COI sequence. Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "file_name.fasta" as Pathname*.

In the *sample_config.txt* file, delete */path/to/seed_file/Seed.fasta* for the *Seed Input* and paste in the true pathname you just copied.

20.3 In the *sample_config.txt* file, make sure your *Read Length* is set to the number transcribed under the *Individual Length of Raw Reads* header in your data statistics spreadsheet (this is most likely *150*).

20.4 In **Finder**, navigate to the location of the raw reads for the corresponding species/sample of interest. Right click (or *control click* on your Mac keyboard) on the *1.fq.gz* file. Hold down *option* on your Mac keyboard while simultaneously clicking *Copy "file_name_1.fq.gz" as Pathname*.

In the *sample_config.txt* file, delete */path/to/reads/reads_1.fastq* for the *Forward reads* and paste in the true pathname you just copied.

20.5 In **Finder**, navigate to the location of the raw reads for the corresponding species/sample of interest. Right click (or *control click* on your Mac keyboard) on the 2.*fq.gz* file. Hold down *option* on your Mac

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

keyboard while simultaneously clicking *Copy "file_name_2.fq.gz" as Pathname*.

In the *sample_config.txt* file, delete */path/to/reads/reads_2.fastq* for the *Reverse reads* and paste in the true pathname you just copied.

Save this revised *sample_config.txt* file.

21 Open up a new **Terminal** shell window. Copy and paste the following "cd working directory" command into **Terminal** and delete */path/to/folder/of/interest:*

cd working directory

cd /path/to/folder/of/interest

The cd command, also known as chdir, is a command-line shell command used to change the current working directory in various operating systems.

Neagu, C. (2020). Command Prompt: 11 basic commands you should know (cd, dir, mkdir, etc.). Digital Citizen. http://www.digitalcitizen.life/command-prompt-how-use-basic-commands

In **Finder**, click once on the *sample_NOVOPlasty* folder. Drag the folder name into **Terminal** and execute (click *return* on your keyboard) the revised command.

22 Copy and paste the following "NOVOPlasty: Mitogenome Assembly" command into the same **Terminal** shell window and delete */path/to/sample_config.txt/file*:

NOVOPlasty: Mitogenome Assembly

perl /Applications/NOVOPlasty-master/NOVOPlasty4.2.1.pl -c /path/to/sample_config.txt/file

Assembles the mitochondrial genome with the input of 2 paired-end, FASTQ files (raw reads; not trimmed or downsampled) and a suitable seed-- including a single read from the dataset that originates from the organelle genome (e.g. mitochondrial COI sequence), an organelle sequence derived from the same or a related species, or a complete organelle sequence of a more distant species (recommended when there is no closely related sequence available).

Dierckxsens, N., Mardulyn, P., & Smits, G. (2020). NOVOPlasty (Version 4.2) [Computer software]. GitHub. https://github.com/ndierckx/NOVOPlasty

In **Finder**, click once on the *sample_config.txt* file. Drag the file name into **Terminal** and execute (click *return* on your keyboard) the revised command. Once it has terminated, you may close the **Terminal** shell window.

23 Upon successful termination of the "NOVOPlasty: Mitogenome Assembly" command, **NOVOPlasty** will create three

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw

output files within your specified working directory, the most important two being:

1. The *Contigs_1_sample_species.fasta* file (e.g. *Contigs_1_MCZ70175C_Lepidonotopodium_williamsae.fasta*) will contain the full mitogenome contig assembled. This is not annotated as in **MitoFinder**, but is simply the full length sequence of the mitogenome (most likely ranging from 12,000 to 25,000 bp long).
2. The *log_sample_species.txt* file (e.g. *log_MCZ70175C_Lepidonotopodium_williamsae.txt*) will contain the input parameters specified in the *sample_config.txt* file and the results of the mitogenome assembly, most importantly the length of the contig (bp) and the average organelle coverage (optimally 100 or more).

Proceed to align the resulting contigs from **MitoFinder** and **NOVOPlasty** to check whether or not they are identical. If the two mitogenome assemblies match, you may report the average organelle coverage obtained from **NOVOPlasty** (in addition to all the **MitoFinder** results) in your manuscript.

However, as mentioned at the beginning of this section, you may also try troubleshooting **MitoFinder** with the suggested options. If any of these **MitoFinder** reruns result in at least 100X mitogenome coverage, it is not required to run **NOVOPlasty** whatsoever.

Citation: Avery S Hiley (08/25/2021). Mitogenome Assembly from NGS Genome Skimming Data. https://dx.doi.org/10.17504/protocols.io.bkbqksmw