

---

# High-level Storage Overview

The meeting will begin soon.  
Please mute your microphone.

[email] [biohpc-help@utsouthwestern.edu](mailto:biohpc-help@utsouthwestern.edu)

# Outline

---

- Storage resources
  - Quotas and backups
- Reviewing multi-tier storage
- Tape storage
- Choosing a storage system
- Using storage effectively
- Data sharing methods

## Storage resources - Quotas and backups

Mount point	Quota (default)	Backup (default)	Speeds	Suggestions
/home2 '/home2/you'	50 GB / user	Mirror backup twice per week	50 GB/s	User environment, small docs, scripts
/project '/project/dept/shared' '/project/dept/lab/shared' '/project/dept/lab/you'	5 TB / group	No backup	400 GB/s	Data processing, large files, sharing
/work '/work/dept/you'	5 TB / user	Mirror backup once per week	30-40 GB/s	Data processing
/archive '/archive/dept/shared' '/archive/dept/lab/shared' '/archive/dept/lab/you'	5 TB / group	No backup	30-40 GB/s	Long term storage, offload location

- /project, /archive, /work quota can be increased from PI's request w/ Dept Chair approval (and backup frequency)
- /archive discount - usage is multiplied by 2/3
- Data on /home2 counts thrice and on work counts twice because of backup

## Storage resources – Cloud storage

---

- Lamella, web-based, backup weekly
  - 100 GB of separate space
  - Separate from home/project/work/archive quotas
  - Potential issues with uploading files that are more than 4 GB
- Cloud (External), web-based, backup daily
  - Similar to Lamella, but 50 GB of space
  - Can't mount internal BioHPC storage for information security restrictions
- Flash is /archive quota
- Thunder is /project quota

## Multi-tier /project storage

---

- /project
  - Hot pool - High performance (NVMe)
  - Cold pool - Storage (HDD)
- Storage process
  1. User saves a file to /project
  2. File is stored in NVMe (counted towards quota)
  3. File is mirrored to HDD after 10 mins of inactivity (counted towards quota)
  4. If NVMe pool usage is high, the file in NVMe pool will be purged
- 1.2x initial quota adjustment

## Multi-tier /archive: Tape storage

---

- ETA: Late Fall / early Winter
- 100 PB total
- Integrates into /archive
- If data isn't interacted with in a year, it moves to tape storage (cold pool) and becomes free
  - Quota is based on data in the hot pool usage
  - The 2/3 discount will still be in effect!

## Using storage effectively - Suggestions

---

- Data in your user directory
- Data in /shared
- Less directories and files in a single directory (< 1000)
- I/O – cache in memory, use when needed, /tmp, ID bottlenecks
- Backup data
  - PI needs to manage /project and /archive backups
  - Mirror / incremental
  - Frequency
- Compressing data
- Move data from /project to /archive after processing

## Managing data – Finding large files

- /home2
  - Files in ~/.cache/\* or ~/.conda/pkgs/\*
- `du -hs $(ls -A) | sort -rh`
- `biohpc_quota`

```
[s212972@Nucleus005 ~]$ biohpc_quota
Current BioHPC Storage Quotas:

```

FILE SYSTEM	SPACE USAGE			NUMBER OF FILES		
	USED	SOFT	HARD	USED	SOFT	HARD
-----						
User quotas for s212972						
-----						
home2	1515M	51200M	71680M	16698	0	0
work	98.41G	5T	7T	265	0	0
-----						
Group quotas for biohpc_admin						
-----						
project	47.23T	0k	0k	23820968	0	0
archive	25.02T	30T	40T	14534743	0	0
-----						

```
*/home2 are backed up twice per week (Mon and Wed)
*/work are backed up weekly
*/project and /archive have no backup, PI can request backup
```



## Managing data – Finding large files

- `du -hs $(ls -A) | sort -rh`
- **Command may take a long time. Contact us if you need help**

```
[s212972@Nucleus005 ~]$ du -hs $(ls -A) | sort -rh
811M    .cache
417M    .conda
166M    scenicplus
67M     .mozilla
54M     HPE_2M2827055D_20230124.ahs
1.2M    .local
776K    .rstudio
452K    .nv
404K    .config
96K     Desktop
80K     portal_jobs
68K     .schrodinger
36K     .dbus
28K     work_files
24K     .vnc
```

## Managing data - Compression

- Use **tar** to package up files into a tarball file (.tar)
- Compress with **bzip2**, a popular **lossless** compression technique
  - Rule of thumb: the more potentially efficient a compression algorithm is, the more CPU it requires.
  - Don't compress on login nodes

- Create .tar archive file: `tar cvf dir2.tar /dir1/dir2`
  - `c` – Creates a new .tar archive file
  - `v` – Verbosely show the .tar file progress
  - `f` – File name type of the archive file
- Create a .tar.bz2 archive file: `tar cvfj dir2.tar.bz2 /dir1/dir2`
  - `j` – Creates a highly compressed file
- Using bzip2:
  - `bzip2 file_to_compress`

## Managing data – Moving data to different directory

---

- Lamella
  - Simple, web based
  - Smaller files
- rsync
  - Command line
  - Lots of files, ~TB
  - Preferred for incremental transfer and unreliable connection

# Managing data – rsync

---

- Before running **rsync**:
  - Make sure there is enough space in the destination filesystem
  - Allocate a 32GB node on the cluster through Slurm or the web visualization
  - Choose the right flags for the **rsync**

```
rsync -aAvh /project/biohpcadmin/s183990/ /archive/biohpcadmin/s183990/project_bak/
```

-a: archive mode  
-A: preserve ACLS  
-v: increase verbosity  
-h: output numbers in a human-readable format

50 – 100 Mbytes/sec

# Access via Network Drives

## Mapping BioHPC Storage Directories to your local environment:

- ▶ Via our load-balanced Samba filesharing server

- ▶ **PC:**

Computer -> Map Network Drive

Folder is: `\\lamella.biohpc.swmed.edu\username` *home2*  
`\\lamella.biohpc.swmed.edu\project\department\group` **project**  
`\\lamella.biohpc.swmed.edu\work\department\group` **work**

- ▶ **MAC:**

Finder -> Go -> Connect to Server

Folder is: `smb://lamella.biohpc.swmed.edu/username` *home2*  
`smb://lamella.biohpc.swmed.edu/project/department/group` **project**  
`smb://lamella.biohpc.swmed.edu/work/department/group` **work**

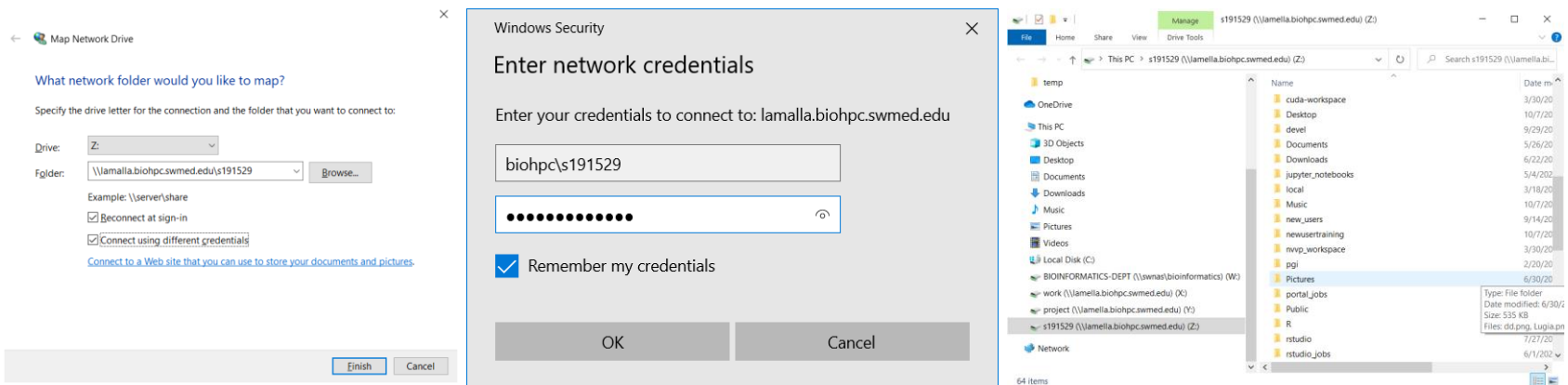
# Accessing BioHPC Storage Directly from Windows

Computer -> Map Network Drive

Folder is: <\\lamella.biohpc.swmed.edu\username> (home dir)  
<\\lamella.biohpc.swmed.edu\project>  
<\\lamella.biohpc.swmed.edu\work>

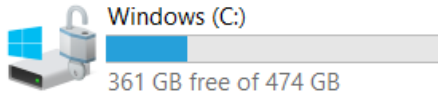
Check 'Connect using different credentials'

Enter your BioHPC username and password when prompted.

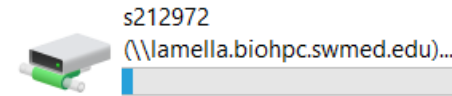
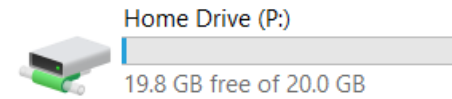
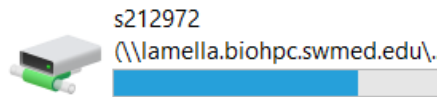
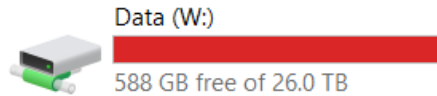
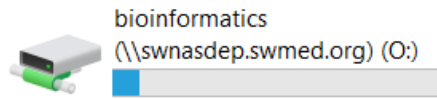
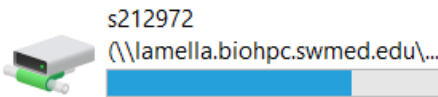
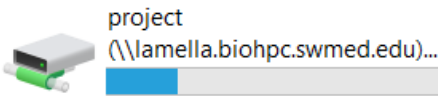
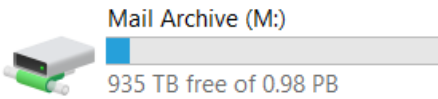


# Access via Network Drives

Devices and drives (1)



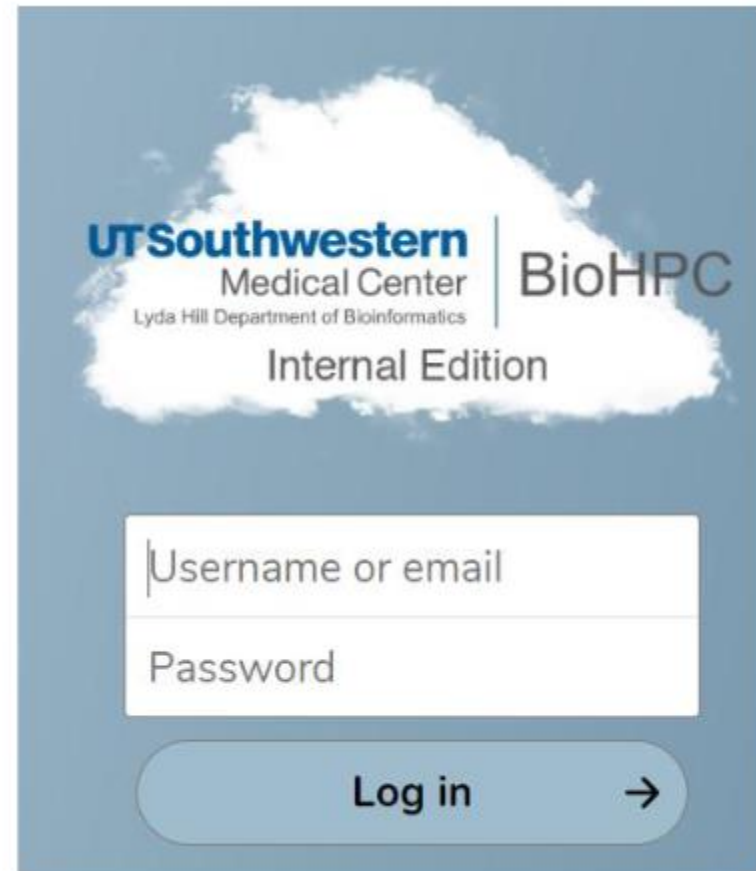
Network locations (8)



# Access via Web Interface

## ❖ Web Access

- Lamella.biohpc.swmed.edu
  - Storage Gateway/Files sharing
  - **/home2** mounted by default (“Home”)
  - **/project, /work, /archive** require additional configurations



UTSouthwestern  
Medical Center  
Lyda Hill Department of Bioinformatics | BioHPC  
Internal Edition

Username or email

Password

Log in →

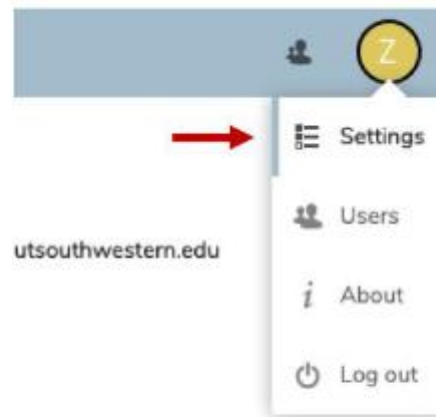


# Access via Web Interface

## ► External Storage Mount:



1.



2.



3.

# Access via Web Interface

## External storages

External storage enables you to mount external storage service external storage services.

Folder name	External storage
home	BioHPC/Lysosome
Cloud	Cloud
<any name>	<b>Add storage</b> BioHPC/Lysosome Cloud

4.


project

Username and password ▾

dept / lab / uid

uid

.....

... 

5.

6.

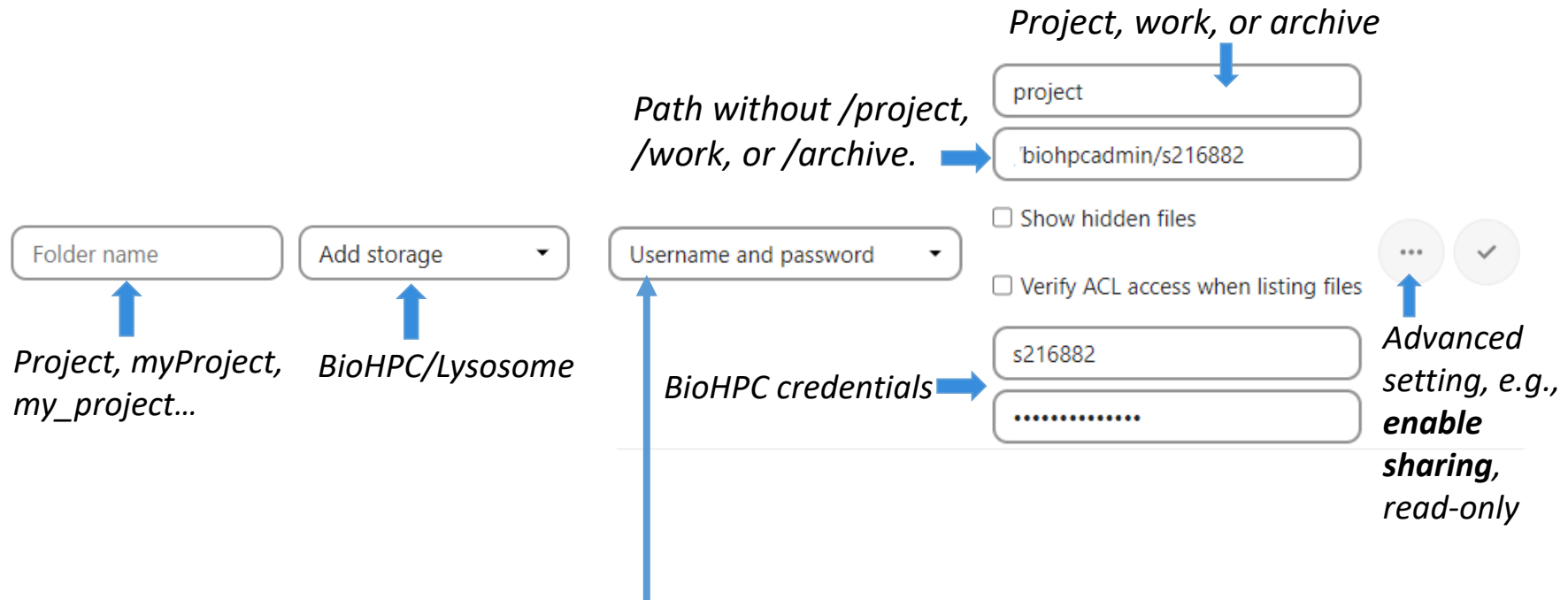
Username and password ▾

- Enable previews
- Enable sharing

7. (optional: select enable sharing is required to share files with others)

# Access via Web Interface

Top-Right Initials icon -> Settings -> External storage on the left

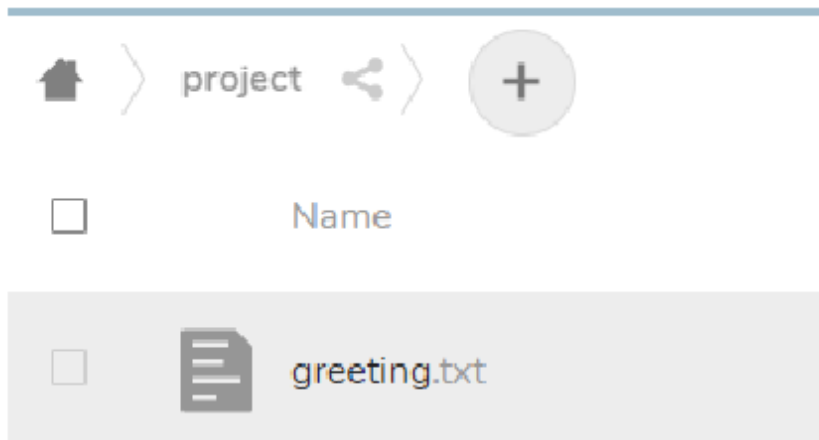


**Log-in credentials, save in session** uses the BioHPC login credentials and only saved in the user session, giving increased security. The drawback is that sharing is disabled, as lamella has no access to the cluster storage credentials.

**Username and password** mechanism requires a manually-defined username and password. Remember to click the three-dots icon and enable sharing.

## Access via Web Interface

```
[s179389@Nucleus006 s179389]$ cat /project/biohpcadmin/s179389/greeting.txt  
Hello World and Good Morning!
```



The screenshot shows a web file manager interface. At the top, there is a breadcrumb navigation path: a home icon, a right-pointing chevron, the word 'project', another right-pointing chevron, and a circular button with a plus sign. Below this, there is a table with two columns: a checkbox and a 'Name' header. The table contains one row with a checkbox and a file icon followed by the text 'greeting.txt'.

<https://lamella.biohpc.swmed.edu/index.php/>

**greeting.txt**

```
1 | Hello World and Good Morning!  
2
```

## Filesharing

---

- Among BioHPC users
  - './shared/..' directories
- Non-BioHPC collaborators (at UTSW)
  - Can't access cluster components
  - Use Thunder or Lamella
- External collaborators (outside of UTSW)
  - Blocked from UTSW internal network
  - Use Flash or Cloud

# Filesharing

## Sharing Among BioHPC Users

- ▶ Sharing with other departments (**made upon request**)
  - ▶ ``/archive/shared`` & ``/project/shared``
- ▶ Sharing within your department:
  - ▶ ``/archive/<dept>/shared`` & ``/project/<dept>/shared``
- ▶ Sharing within your lab:
  - ▶ ``/archive/<dept>/<lab>/shared`` & ``/project/<dept>/<lab>/shared``



# Filesharing


Internal shares	External shares
<b>Lamella</b> <ul style="list-style-type: none"><li>Files that are &lt; 100 GB</li><li>lamella.biohpc.swmed.edu</li></ul>	<b>Cloud (External file exchange)</b> <ul style="list-style-type: none"><li>Files that are &lt; 50 GB</li><li>cloud.biohpc.swmed.edu</li></ul>
<b>Thunder FTP</b> <ul style="list-style-type: none"><li>Files that are &gt; 100 GB</li><li>thunder.biohpc.swmed.edu</li></ul>	<b>Flash FTP</b> <ul style="list-style-type: none"><li>Files that are &gt; 50 GB</li><li>flash.biohpc.swmed.edu</li></ul>

# Filesharing - Lamella

---

Sharing with Non-BioHPC Collaborators: < 100GB






lamella.biohpc.swmed.edu



- ▶ `home` (/home2/<you>) -> shareable!
- ▶ Mapped `/project` & `/work` -> manually enable sharing
- ▶ Look for the  icon
- ▶ Instruct collaborator to access via web URL







# Filesharing - Lamella


Sharing with Non-BioHPC Collaborators: < 100GB



		Size	Modified
	...	Pending	a month ago
	...	1.4 MB	2 months ago
	...	Pending	6 months ago
1. 	...	Pending	22 days ago
	...	Pending	2 months ago


 My\_Scripts 

★ < 1 KB, 22 days ago 

 Activity  Comments  Sharing

Name... 

 Share link 2. 

 Add to a project  
Connect items to a project to make them easier New share link

# Filesharing - Lamella

## Sharing with Non-BioHPC Collaborators: < 100GB

- ▶ Grant appropriate permissions!
  - ▶ Stick to principle of least privilege
- ▶ Password protect!
- ▶ Set Expiration date!

The screenshot shows the 'My\_Scripts' file sharing interface. At the top, it displays 'My\_Scripts' with a star icon, '< 1 KB, 22 days ago', and a 'Tags' icon. Below this are three tabs: 'Activity', 'Comments', and 'Sharing'. A red arrow points to the 'Sharing' tab. A red circle highlights the 'Share link' icon in the top right corner of the sharing menu. A red arrow points to the 'Share link' option in the menu. A red arrow points to the 'Add to a private Connect item' option. A red arrow points to the 'Read only' option. A red arrow points to the 'Password protect' option. A red arrow points to the 'Set expiration date' option. A red arrow points to the 'Choose a password for' input field. A red arrow points to the '22-10-2019' date input field. A red arrow points to the 'Note to recipient' option. A red arrow points to the 'Delete share link' option. A red arrow points to the 'Add another link' option. A red arrow points to the 'Copy link to clipboard' text.

\*Copy link to clipboard

Share link

Add to a private Connect item

Read only

Allow upload and editing

File drop (upload only)

Hide download

Password protect

Choose a password for →

Set expiration date

22-10-2019

Note to recipient

Delete share link

+ Add another link

# Filesharing - Thunder

Sharing with Non-BioHPC Collaborators: > 100GB

thunder.biohpc.swmed.edu


- ▶ ``/project/thunder_ftp/<uid>``  
-> share folder
- ▶ Upload limit = `/project quota*`
- ▶ If size exceeds quota -> Contact us!
- ▶ Access via **web** –or– **FileZilla** (if collaborator wishes to upload)



# Filesharing – Thunder – Creating guest access

## Sharing with Non-BioHPC Collaborators: > 100GB


### Guests: Create and manage guest FTP accounts

Share with UTSW collaborator: 

 Add new Guest

You currently have not allowed any guest accounts to access your ftp folder form.

### Guests: Create and manage guest FTP accounts



Share with UTSW collaborator: 

Email address \*

Note

Directory Name\*

Email Content

guest created successfully 

Guestname	Expiration	Email	Note	directory	Actions
guest94202Szhi	Oct. 19, 2019, 7:53 a.m.	zengxing.pang@utsouthwestern.edu		/project/thunder_ftp /z pang1 /another_test	  

# Filesharing – Thunder – Sharing guest details

## Sharing with Non-BioHPC Collaborators: > 100GB

BioHPC user:

z pang1 has shared his BioHPC ftp folder with you with a message:

**UTSW employee: z pang1 has shared his ftp folder with you, please click on below link to view your ftp credentials**

please click the link below to view your temporary message:

[https://thunder.biohpc.swmed.edu/manage/view\\_guest\\_account/?q=9297aa0cf02f82c464c47b2d5e16673b86146c691b4f442facf666e0aeeb8c4a153db883fd05b9a6f1ad04c716c946c5](https://thunder.biohpc.swmed.edu/manage/view_guest_account/?q=9297aa0cf02f82c464c47b2d5e16673b86146c691b4f442facf666e0aeeb8c4a153db883fd05b9a6f1ad04c716c946c5)

### Guest Account Session:

Thanks for using BioHPC, your temporary account and password are as follows, please save them, this account automatically expires in three days:

USERNAME: **guest94202Szhl**

PASSWORD: **zf5Eh7b3HtFV**

URL: **ftp://guest94202Szhl:zf5Eh7b3HtFV@thunder.biohpc.swmed.edu**

# Filesharing – Guest account FTP access with Thunder

Host: flash.biohpc.swmed.edu

Username: <username>

Password: <password>

Port: Leave Empty

Sharing with Non-BioHPC Collaborators: > 100GB

The screenshot shows the Thunder FTP client interface. At the top, the connection details are: Host: thunder.biohpc.swrr, Username: guest94202Szr, Password: [masked], Port: [empty]. A 'Quickconnect' button is visible. Below this, a status log shows: Connection established, waiting for welcome message...; Insecure server, it does not support FTP over TLS.; Logged in; Retrieving directory listing...; Calculating timezone offset of server...; Timezone offset of server is 0 seconds.; Directory listing of "/" successful.

The interface is split into two panes. The left pane shows the 'Local site' at /Users/zeng/Desktop/Work/. The right pane shows the 'Remote site' at /. Below the panes, there are two file listings. The left listing shows the local directory structure with folders like .vim, .vnc, .vscode, Applications, Desktop, Oct\_16\_19, Temps, and Work. The right listing shows the remote directory contents:

Filename	Filesize	Filetype	Last modified
..			
limit_rate.png	35,833	png-file	10/15/2019 1...
snow capped mountains.jpg	14,089,0...	jpg-file	10/15/2019 1...
test_statement.md	6	Visual Stu...	09/26/2019 1...
test_statement_back.md	6	Visual Stu...	10/15/2019 1...

The bottom pane shows a detailed file listing for the local site:

Filename	Filesize	Filetype	Last modified
..			
Artifacts		Directory	10/01/2019
Meetings		Directory	02/27/2019
Pre_Docs		Directory	10/01/2019
RedTape		Directory	03/02/2019
.DS_Store	10,244	File	10/08/2019

# Filesharing – External conclusion

---

Sharing with External Collaborators:

< 50GB -> BioHPC External File Exchange

[cloud.biohpc.swmed.edu](http://cloud.biohpc.swmed.edu)

> 50GB -> Flash FTP Service

[flash.biohpc.swmed.edu](http://flash.biohpc.swmed.edu)

- ▶ External Equivalents of Lamella & Thunder (internal)
  - ▶ Identical operations

## BioHPC Acknowledgement

---

Please always acknowledge our contribution in your publication:

*This research was supported in part by the computational resources provided by the BioHPC supercomputing facility located in the Lyda Hill Department of Bioinformatics, UT Southwestern Medical Center.*

Your acknowledgment holds significant value in enabling us to secure funding for the advancement of next-generation hardware, enhancement of user support services, and furthering our research and development efforts in the field of HPC.



# Thank you!

---

- Email: [biohpc-help@utsouthwestern.edu](mailto:biohpc-help@utsouthwestern.edu)
- More info:
  - Tiered /project
    - <https://portal.biohpc.swmed.edu/pages/important-news/the-new-two-tier-high-performance-project-filesystem/>
  - Storage cheat sheet
    - <https://portal.biohpc.swmed.edu/content/guides/storage-cheat-sheet/>
  - Quota
    - <https://portal.biohpc.swmed.edu/content/faqs/>