alias yolo='git push --force'

Chaochih Liu and Paul Hoffman

Applied Bioinformatics Discussion

June 24th, 2016

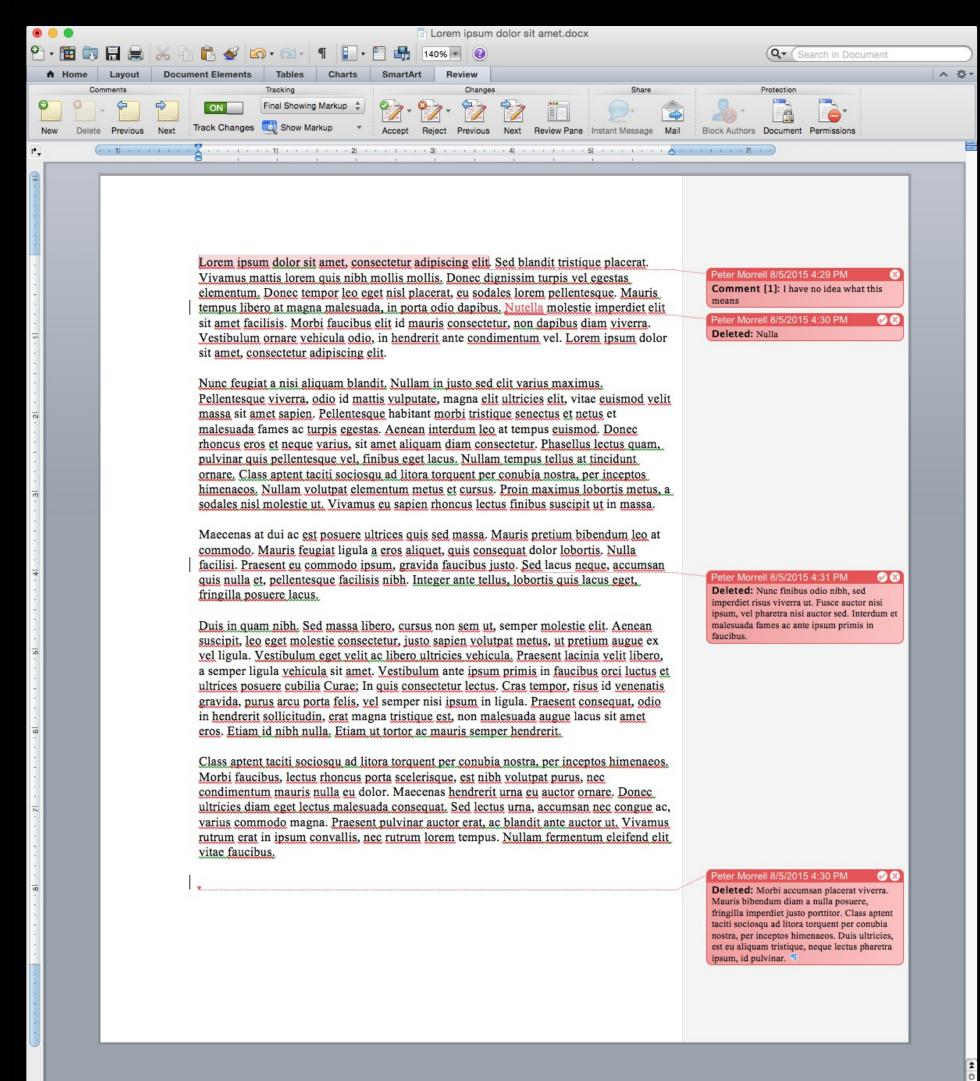
What is Git?

Version Control System

- Keep track of changes
- Use specific version of code
- Free software

Git Vs. MS Word

MS Word Track Changes



Git Track Changes

18 Config			
<u>†</u> ∡	@@ -23,11 +23,6 @@ PLATFORM=		
23	# What reference genome are we using?	23	# What reference genome are we using?
24	REF_GEN=	24	REF_GEN=
5		25	
7	-# Where is the list of processed BAM samples?		
8	-# This is used for Coverage_Map and Indel_Realignment		
29	<pre>-# If using SAM_Processing, then leave as is -BAM_SAMPLES=\${OUT_DIR}/SAM_Processing/\${PROJECT}_Processed_BAM.txt</pre>		
30	- BANI_SANIPLES=\${OUI_DIR}/SANI_PROCESSING/\${PROJECT}_PROCESSEU_BANI.CXC		
31	***************************************	26	***************************************
32	######## Quality_Assessment ########	27	######## Quality_Assessment ########
33	***************************************	28	***************************************
E	@@ -121,28 +116,28 @@ RM_QSUB="mem=8gb,nodes=1:ppn=8,walltime=16:00:00"		
21	#	116	#
	TRIMMED_LIST=\${OUT_DIR}/Adapter_Trimming/\${PROJECT}_trimmed_adapters.txt		TRIMMED_LIST=\${OUT_DIR}/Adapter_Trimming/\${PROJECT}_trimmed_adapters.txt
22	# If you're using the list from Quality_Trimming	117	# If you're using the list from Quality_Trimming
23	#	118	#
	TRIMMED_LIST=\${OUT_DIR}/Quality_Trimming/\${PROJECT}_trimmed_quality.txt		TRIMMED_LIST=\${OUT_DIR}/Quality_Trimming/\${PROJECT}_trimmed_quality.txt
24	-TRIMMED_LIST=	119	+TRIMMED_LIST=\${OUT_DIR}/Adapter_Trimming/\${PROJECT}_trimmed_adapters.txt
25		120	
26	# How are trimmed forward samples named?	121	# How are trimmed forward samples named?
27	# If using the Adapter_Trimming handler	122	# If using the Adapter_Trimming handler
.28	# FORWARD_TRIMMED=_Forward_ScytheTrimmed.fastq.gz	123	# FORWARD_TRIMMED=_Forward_ScytheTrimmed.fastq.gz
.29	# If using the Quality_Trimming handler	124	# If using the Quality_Trimming handler
30	# FORWARD_TRIMMED=_R1_trimmed.fastq.gz	125	# FORWARD_TRIMMED=_R1_trimmed.fastq.gz
31 32	-FORWARD_TRIMMED=	126 127	+FORWARD_TRIMMED=_Forward_ScytheTrimmed.fastq.gz
.33	# How are trimmed reverse samples named?	128	# How are trimmed reverse samples named?
34	# If using the Adapter_Trimming handler	129	# If using the Adapter_Trimming handler
.35	# REVERSE_TRIMMED=_Reverse_ScytheTrimmed.fastq.gz	130	# REVERSE_TRIMMED=_Reverse_ScytheTrimmed.fastq.gz
136	# If using the Quality_Trimming handler	131	# If using the Quality_Trimming handler
37		132	# REVERSE_TRIMMED=_R2_trimmed.fastq.gz
138	-FORWARD_TRIMMED=	133	+FORWARD_TRIMMED=_Reverse_ScytheTrimmed.fastq.gz
39		134	
40	# How are trimmed single-end samples named?	135	# How are trimmed single-end samples named?
141	# If using the Adapter_Trimming handler	136	# If using the Adapter_Trimming handler
142	# SINGLES_TRIMMED=_Single_ScytheTrimmed.fastq.gz	137	# SINGLES_TRIMMED=_Single_ScytheTrimmed.fastq.gz
L43	# If using the Quality_Trimming handler	138	# If using the Quality_Trimming handler
L44	# SINGLES_TRIMMED=_single_trimmed.fastq.gz	139	# SINGLES_TRIMMED=_single_trimmed.fastq.gz
.45	-SINGLES_TRIMMED=	140	+SINGLES_TRIMMED=_Single_ScytheTrimmed.fastq.gz
46		141	
47	# BWA mem parameters; below are the defaults for BWA mem	142	# BWA mem parameters; below are the defaults for BWA mem
48	# How many threads are we using?	143	# How many threads are we using?
#₹	@@ -262,6 +257,11 @@ REF_ANN=		
62	# Below are the recommended settings	257	# Below are the recommended settings
63	<pre>IR_QSUB="mem=15gb, nodes=1:ppn=4,walltime=48:00:00"</pre>	258	<pre>IR_QSUB="mem=15gb, nodes=1:ppn=4, walltime=48:00:00"</pre>
64		259	
		260	+# Where are the processed BAM files located?
		261	+# This should be the full file path to their directory
		262	+# If using SAM_Processing, leave as is
		263 264	+BAM_DIRECTORY="\${OUT_DIR}/SAM_Processing/finished_BAM" .
		264	+ What is the full file math for the CATV tan file?
65	# What is the full file math for the CATV day file?	()	# What is the full file path for the GATK jar file?
	# What is the full file path for the GATK jar file?		CATY JAP
.65 .66	<pre># What is the full file path for the GATK jar file? GATK_JAR=</pre>	266 267	GATK_JAR=

Git Basics

To create a new repository

git init [directory]

Check Git's status of files in repository.

git status

See commit history.

git log --graph

Git Basics

To move and rename files:

git mv fileName1 fileName2 git mv filepath1/fileName1 filepath3/fileName1

To remove files:

git rm filename

For more git commands, refer to Git Cheat Sheet

Git Basics

Add files to be tracked by Git

git add

Commit a file and it's changes to Git

git commit

.gitignore

Special file that tells Git to not track certain files

- Data files
- Temporary files
- Byte/compiled code

Git Remotes allow for a backup of your repository on some other devices

- Can be hosted service like BitBucket or GitHub
- Can be on private server

Obtain a repository from an existing URL

git clone [url]

Synchronize Changes

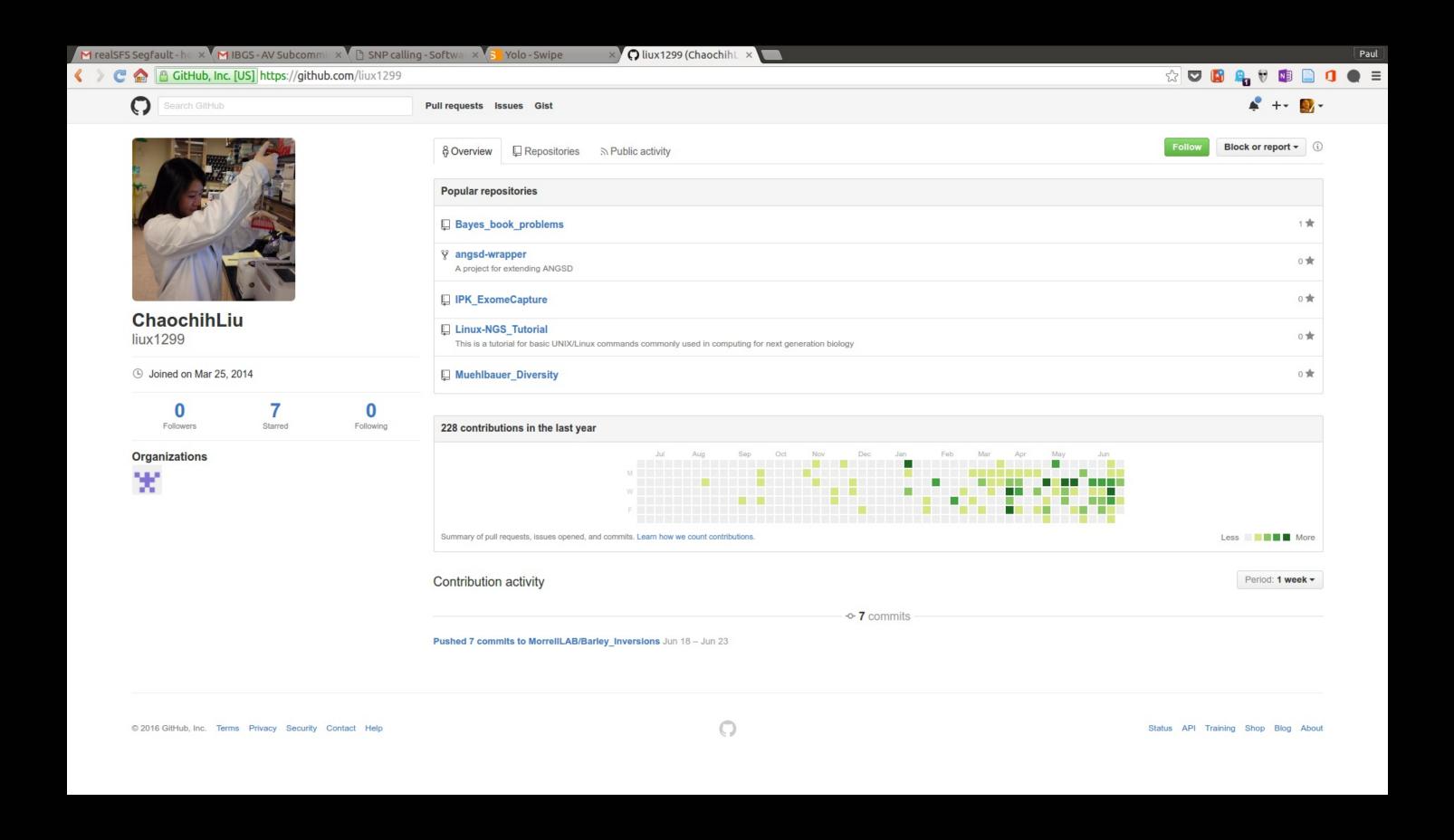
- Downloads history and updates changes
- You must be in the repository before you run this command

git pull

- Uploads all local changes (use all three commands in order)
- You add the change, commit the change, and push the change

git add git commit -m "commit message" git push

Common examples of Git Remotes are GitHub and BitBucket



Authenticating Git Remotes

Two methods

- SSH
- HTTPS

VINCE BUFFALO IS WRONG

It is recommended to use HTTPS over SSH when working with Git, especially GitHub

- HTTPS is simpler to set up than SSH
- Git's credential.helper only works with HTTPS, not SSH
- HTTPS is an open port on all networks, SSH may be blocked
- HTTPS has the same, or better, level of security as SSH
 - Users unfamiliar with SSH best practices will find SSH less secure due to improper management of private keys

Intro to GitHub

Collaboration

- Review changes
- Comment on code
- Report issues

Sharing of code

Get feedback

User friendliness

- Web-based graphical interface
- READMEs and Wiki pages

An account is not necessary to view and download public repositories.

Collaborating with GitHub

Working with Your Collaborators

Caution!

Always pull before you start editing a file!!

What (NOT) to Use GitHub For?

Using GitHub

GitHub is used to store code and documentation

GitHub can also be used for writing in Markdown/LaTeX

Website hosting through GitHub Pages is an acceptable use of GitHub

Abusing GitHub

GitHub is **not** used for data storage

• Example datasets are acceptable

GitHub is **not** used for compiled machine code or interpreted byte code

Python .pyc files should not be uploaded

GitHub is **not** used for non-human readable files

• Binary files should **not** be uploaded

See Bryan J et al. (2016) Good enough practices for scientific computing.

Possibly Abusing GitHub

Figure sharing on GitHub should only be done if the figure is being used elsewhere on GitHub

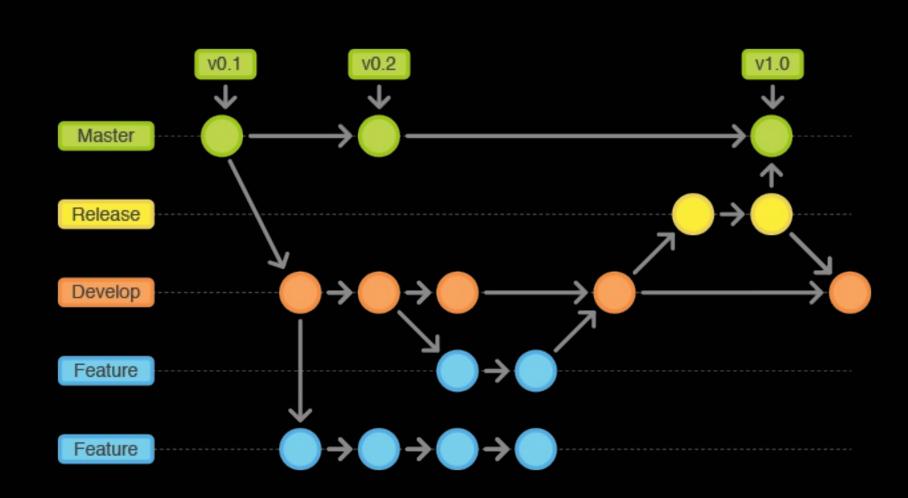
- Wiki pages/READMEs
- GitHub Pages
- Writings hosted on GitHub
 - Example: ANGSD-wrapper manuscript

Working with Past Commits

Undo and Edit Commits

git commit --amend git reset --hard git checkout --force

Git Branches, Forks, and Pull Requests



Merge Errors

Chaochih's Favorite!

More Advanced Git Topics

Submodules

Searching revisions with git grep

Rebasing git rebase

Exercise: Test the waters!

Get into groups of 3-5. Make sure **at least 1** person has a computer with Git installed on it.

If you need a reference for Git commands, take a look at GitHub's cheat sheet.