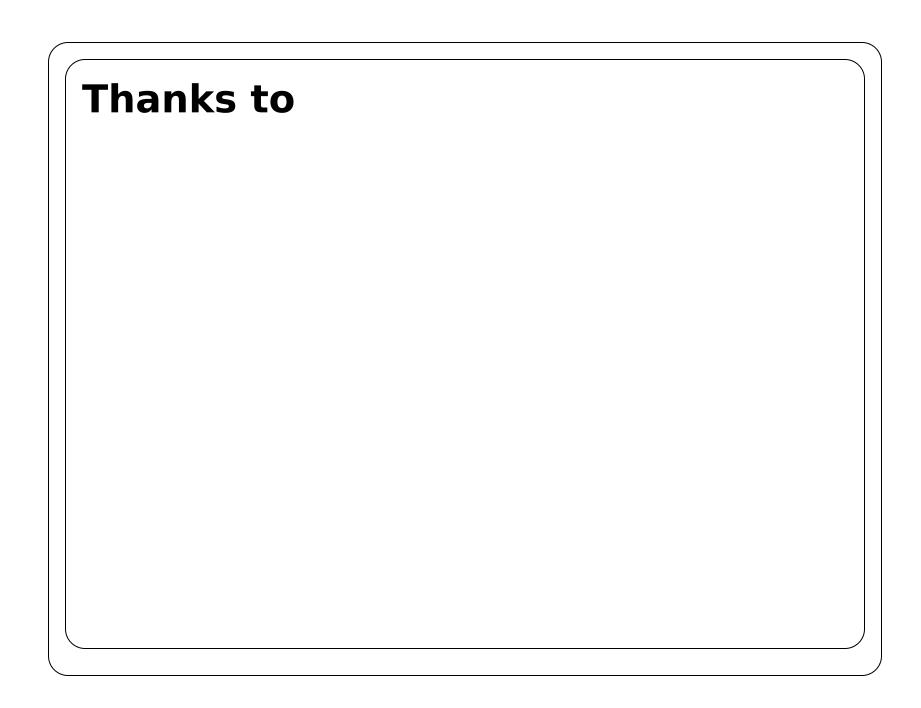


SNP Pipeline Code Review

Hugh Rand and Yar	N Luo <i>A I A I</i> 201 <i>A</i>		
nugn Kanu anu far	Luo 4/4/2014		



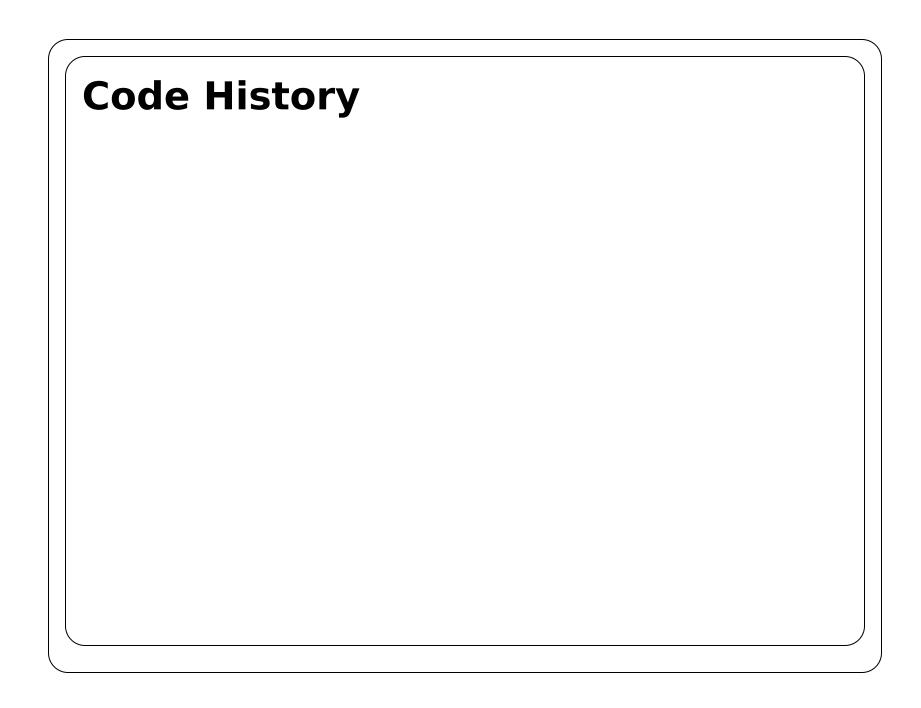
•	án - graciously letting me rearrange her code.	

Jamie - for help testing and troubleshooting.	

John I for an idea about how do one thing.

_						
	• Errol - fo	r prodding m	ne along.			

Fish - for prodding on code reviews.					
	• Fish - for	prodding on code re	eviews.		



========= Errol - 1st version Yan - 2nd version Jamie and Yan - use for publication

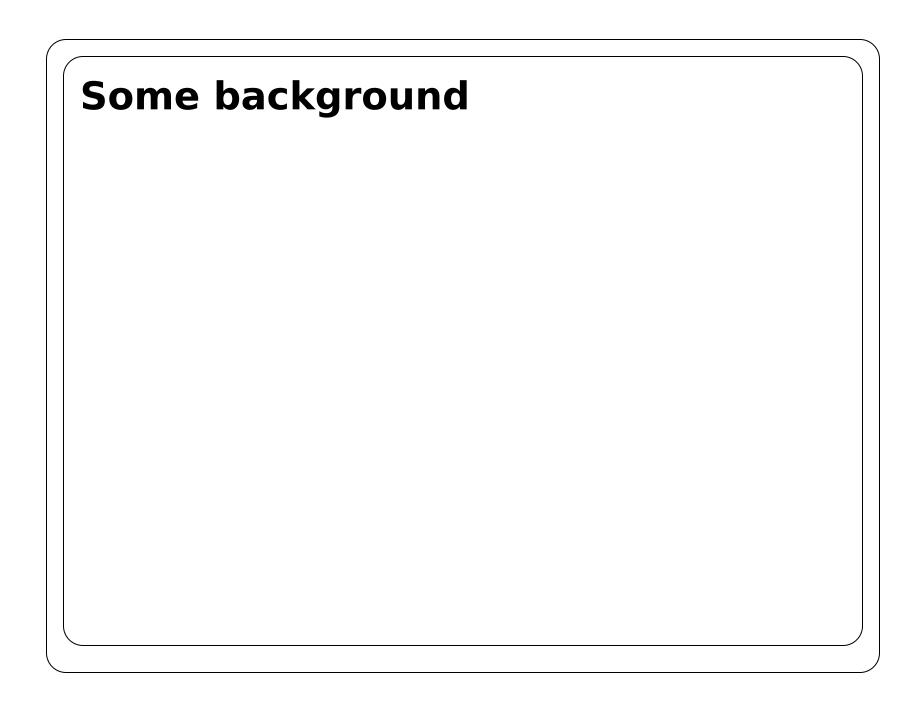
Errol, use fo	or outbreak investigation CDC pipeline (Lee Katz) https://github.com/lskatz/lyve-





===== Make it publishable - CDC person - CVM persor	ו

Mako it a packago Pousoablo codo Facily installed Locked down Tostable	
Make it a package Reuseable code Easily installed Locked down Testable	



==========	Running with anaconca Using python 2.7.6 PyVCF (If you want to

get good at writing, then read a fair bit.) (But, be judicious about what you read.) reasonably
get good at writing, then read a rail bit.) (but, be judicious about what you read.) reasonably

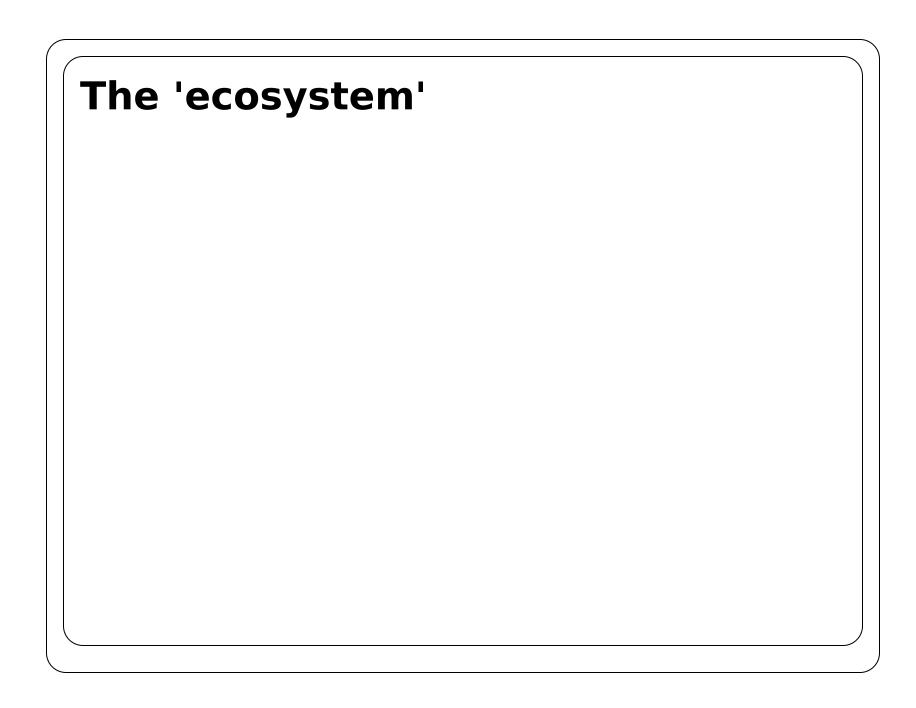
well writte	n. does something	useful for me.	\$python -m ea	sy_install PyVCF	- 131 Searchir	ng for
				-		

PyVCF	= 132 Best ma	atch: PyVCF (0.6.7 133 Pro	cessing PyV0	CF-0.6.7-py2.	7-linux-x86_	64.egg 134

PyVCF 0.6.7	' is already the a	ctive version in	easy-install.pth	135 Installing v	vcf_filter.py so	ript to

_	
/	home/hugh.rand/anaconda/bin 136 Installing vcf_melt script to /home/hugh.rand/anaconda/





Python instalation - Anaconda IDE - Spyder Code standards PEP8 Google Code Testing pylint
Tython mistalation 7 maconau ibe spyder code standards i er o doogle code lesting pyline

unittest doctest Distribution distutils setuptools Documentation PEP8 sphinx Source Code	

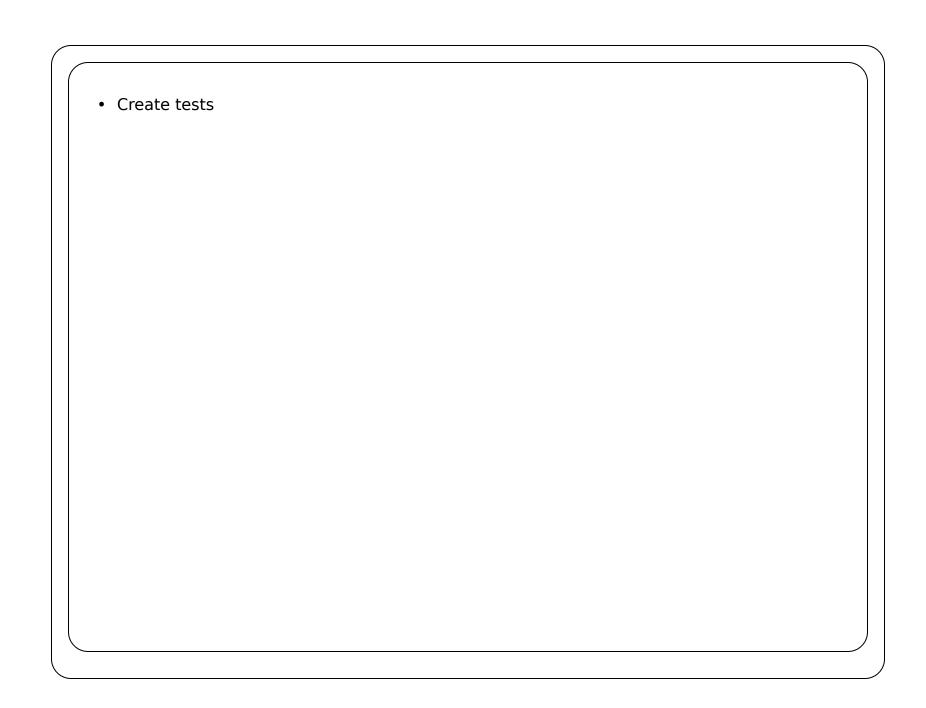
Control - git Presentations	

What I did / What is to do



•	Broke much of c	ode out as func	tions		

. D		و و اوا و اورون و اورون و		
• Pythonify	names of function	ns and variables		



	ariables gone		
- All global V	anabies gone		

• Brea	k up code into	'main' and utilit	ies		

Move main 'script' into fur	nction		

. Undata are	umant naccina ac	okogo (2.6 × 2.7)		
• Opdate arg	gument passing pa	ickage (2.6->2.7)		

• 0	rganize code as	a python packag	ge		

• Move from three	ds to processes	
 Move from threa 	as to processes	

General cleanup			

Add use of PyVCF	

 Improve pylint score 		

• Ac	dd release ver	sioning and a	all under vers	sion control		

•	Add new flag	gs (includeRe	ference, vari	ous paramete	ers).	

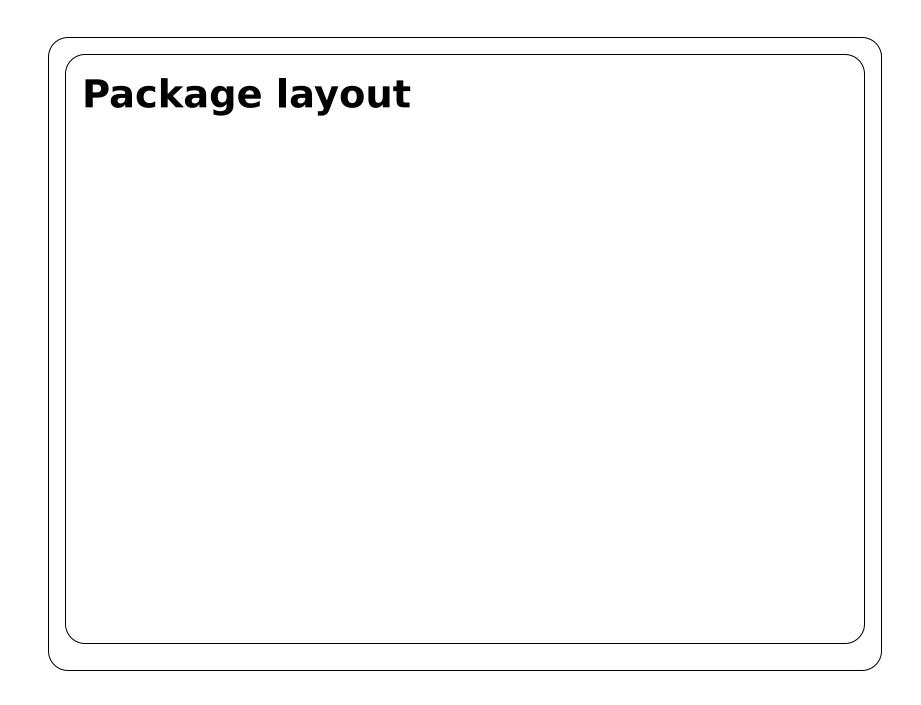
•	Add new f	ags (verbos	se, useOldPi	leups).		

• Trap keyboa	ard interupts so can halt co	ode cleanly.	

Turn it into an "Egg"	

Get it out on gitHub	

 Move to SVN 		



9	snppipeline build dist doc LICENSE.txt MANIFEST.in notes presentations README.txt setup.py
_	

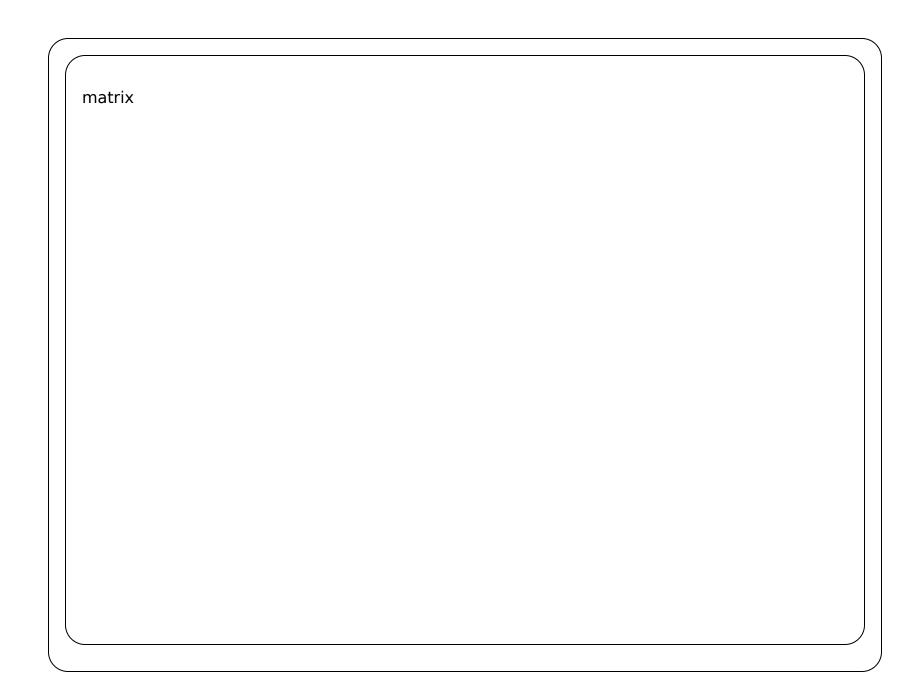
_	
	snppipeline README_developmentNotes snppipeline.py utils.py test codeComparisonFiles

toct/aona//O//	tastlambda\/irus tast	conningling by test	utilany	
testAgonaMOM t	testLambdaVirus test	_snppipeline.py test	_utils.py	

snppipeline.egg-info	

Functions (no classes)

(egrep 'def \"\"' snppipeline/snppipeline.py def run_snp_pipeline(options_dict): """Create SNP
_	



_	
ϵ	egrep 'def \"\"' snppipeline/utils.py def pileup_wrapper(args): """Wraps pileup to use multiple
_	

_	
а	arguments with multiprocessing package. def pileup(filePath, options_dict): """Run samtools to
_	

generate pileup. def get_consensus_base_from_pileup(base, length, da	ata): """Call the base for

_	
_	anch SND position dof croate, consensus dist/piloup filo path), """Croate a dist based on the
_	each SNP position def create_consensus_dict(pileup_file_path): """Create a dict based on the
_	

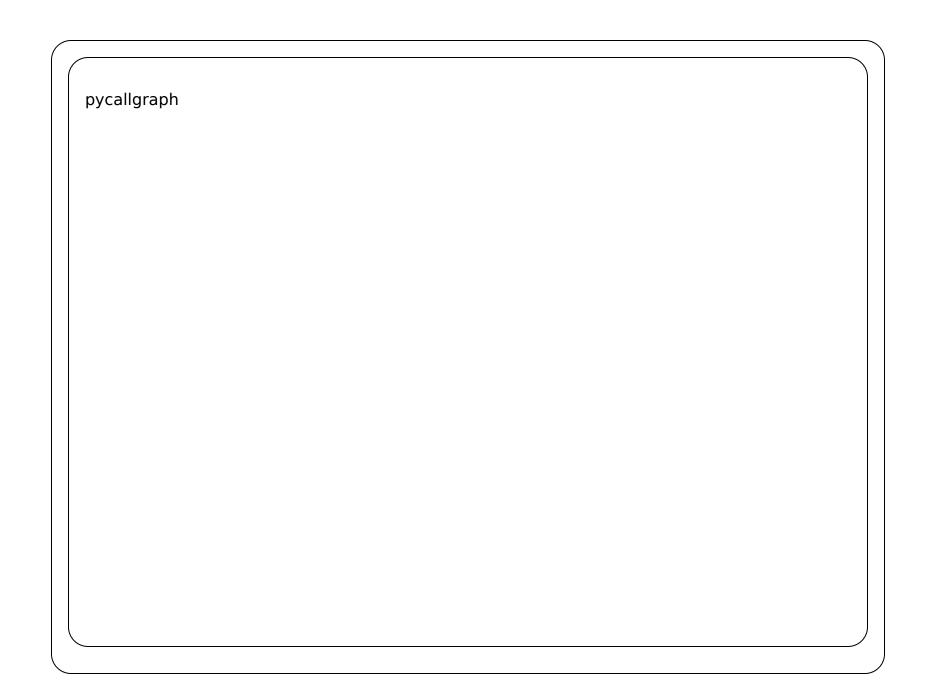
i	nformation in a pileup file. def write_list_of_snps(file_path, snp_list_dict):
_	

	$\overline{}$
"""Write out list of snps for all samples to a single file. def	

,	write_reference_snp_file(reference_file_path, snp_list_file_path, """Write out the snp fasta file
_	

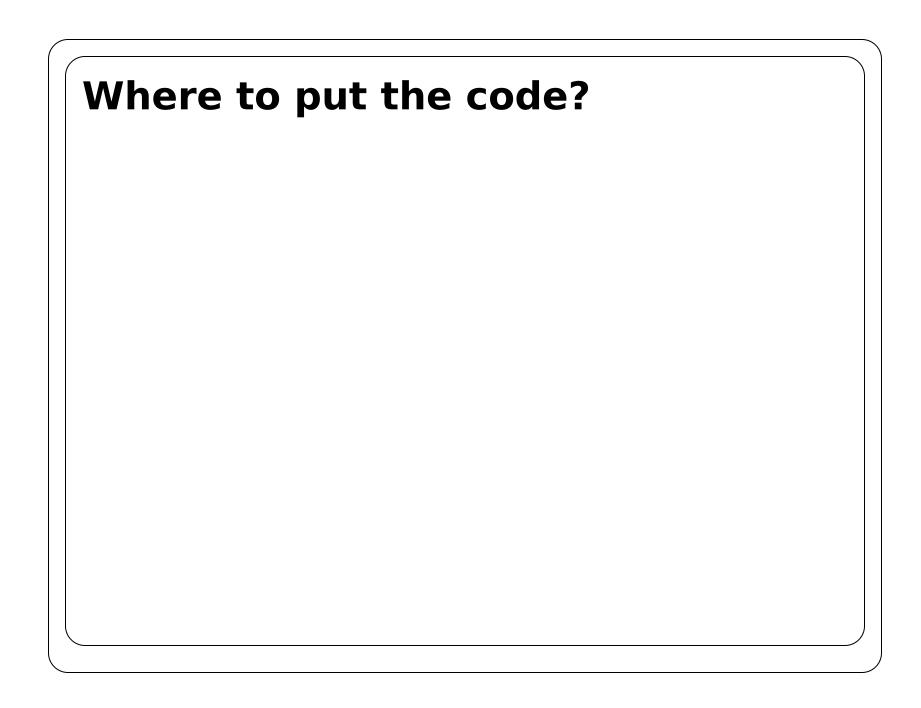
for the reference.fas	ta using the snp		

Function call graph

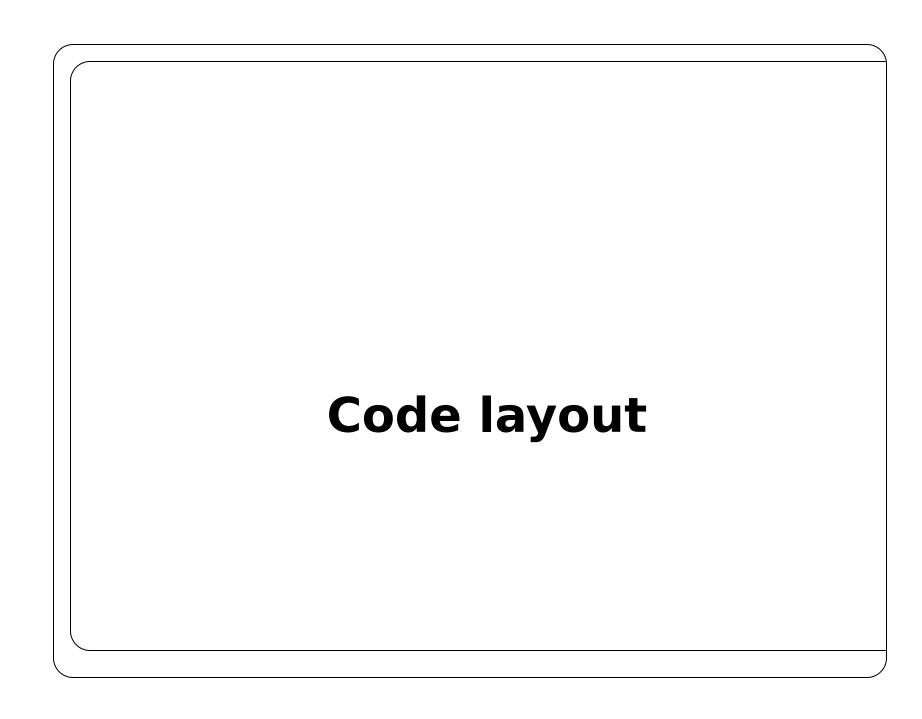


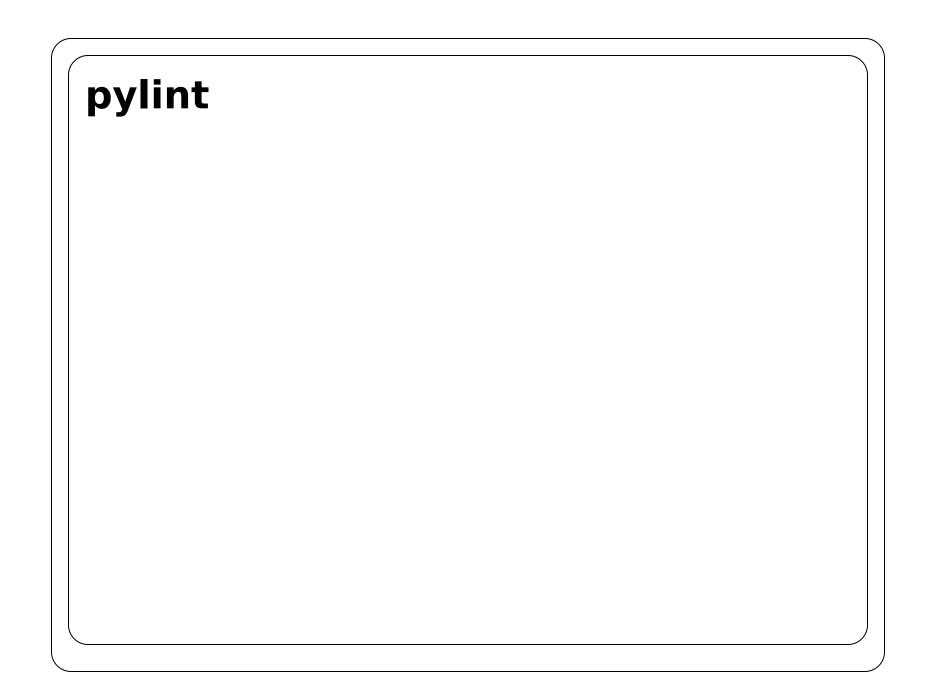
Testing environment

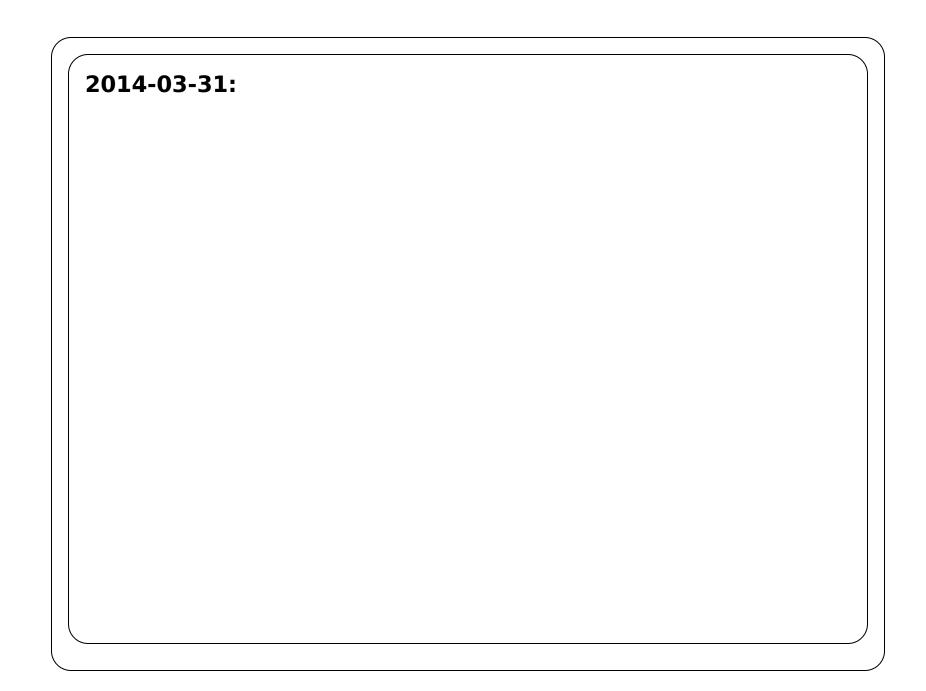
V	rirtual machine vmware - vmplayer ubuntu install process samtools pyvcf????



svn: https://x	serve19.fda.go	v/svn/bioin/ma	pping?		

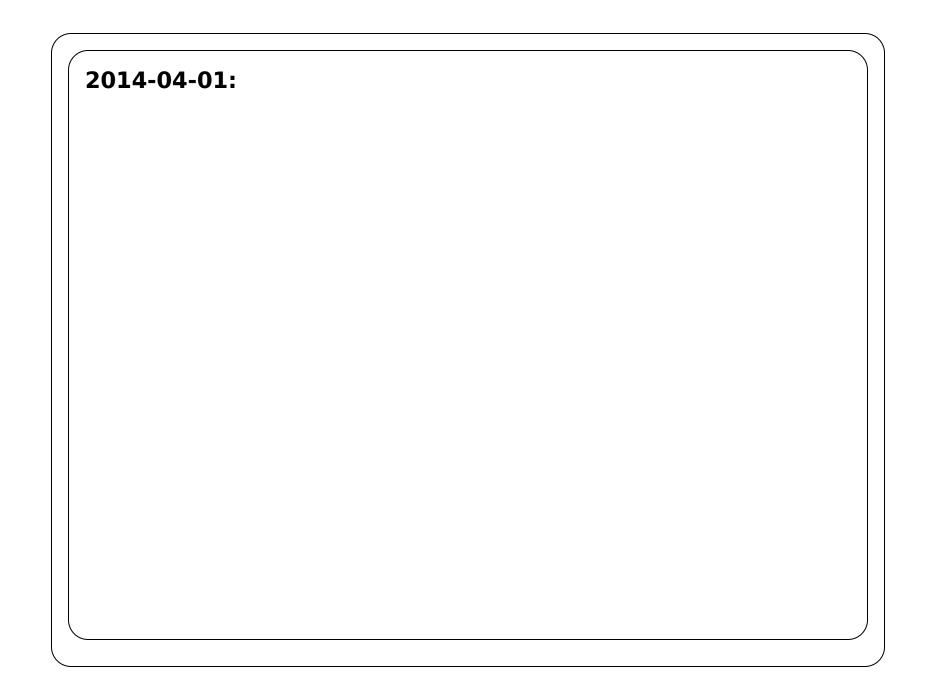






	P / P			42/10	
pylint snppipe	line/snppipeline.p	y Your code has	been rated at 2.	42/10	

pylint snp	pipeline/utils.	py Your code.	has been rat	ed at 6.15/10		



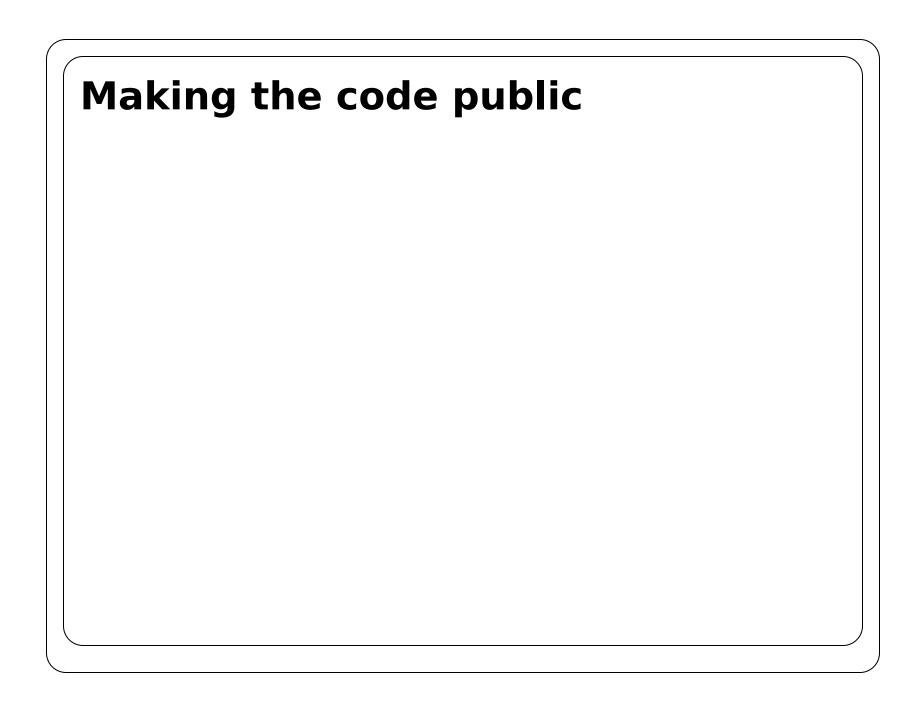
pylir	nppipeline/utils.py grep "rated" No config file found, using default con	figuration Your

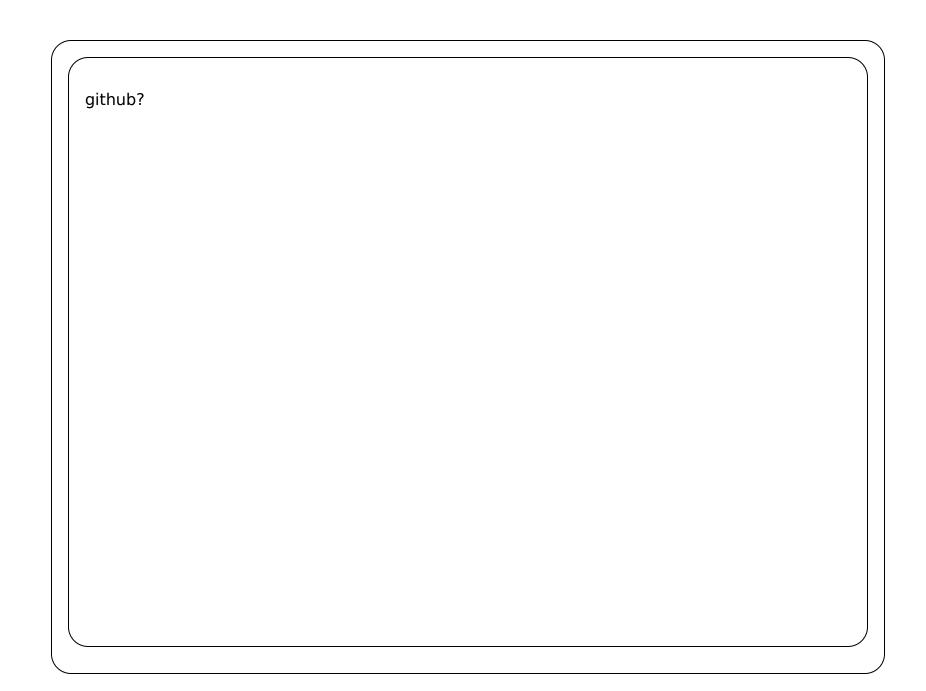
codo bac ba	on rated at 7.72/	10 (provious rur	s. 7 72/10\	
code has bee	en rated at 7.72/	10 (previous rur	1: 7.72/10)	

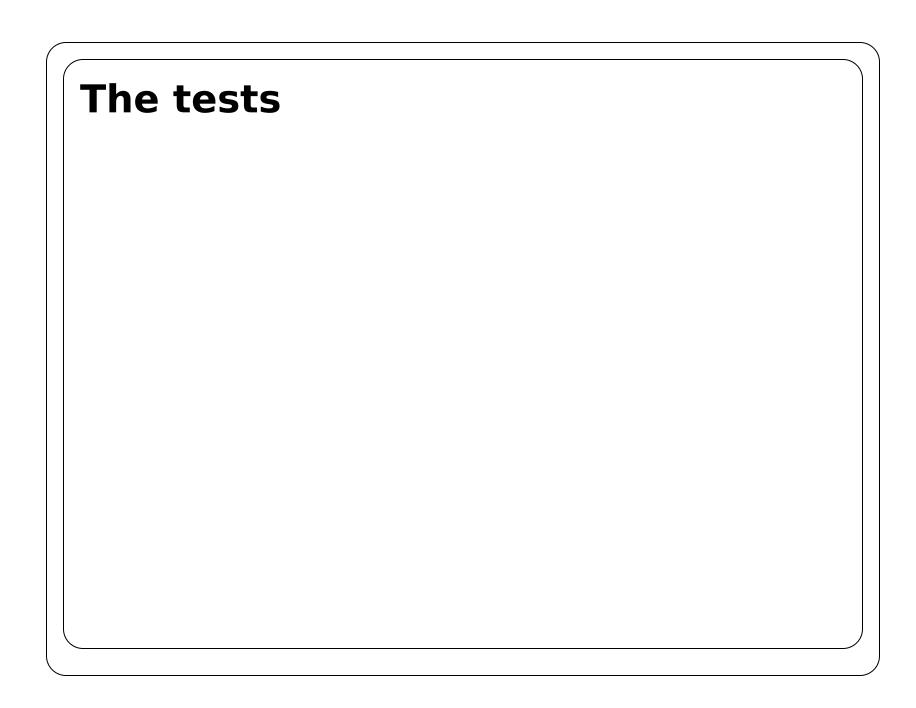
ı	pylint snppipeline/snppipeline.py grep "rated" No config file found, using default

configuration Your code has been rated at 5.68/10 (previous run: 5.68/10)	
configuration Your code has been rated at 5.68/10 (previous run: 5.68/10)	

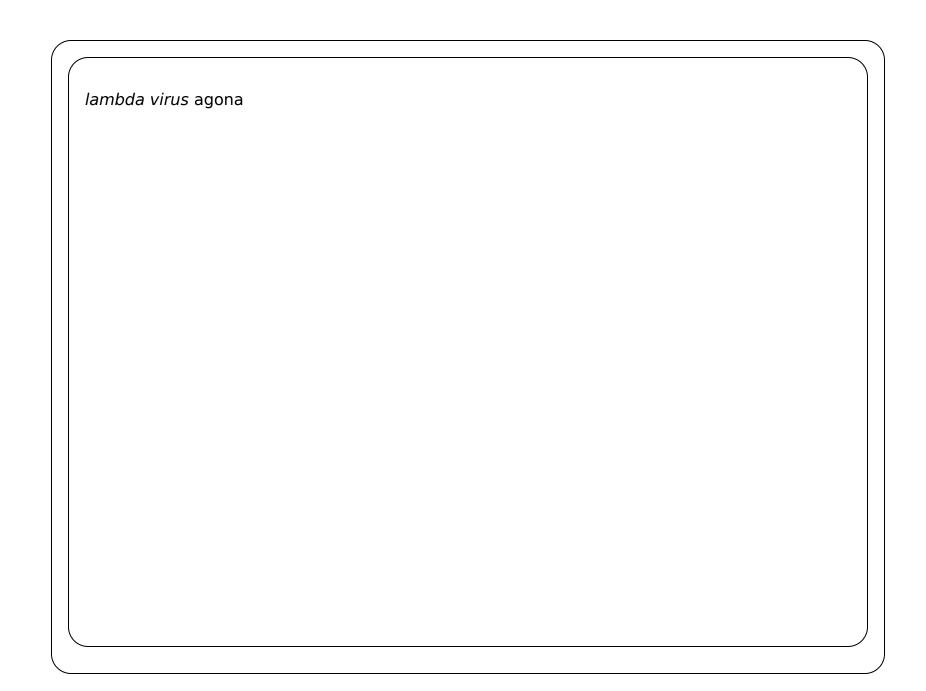
The git log

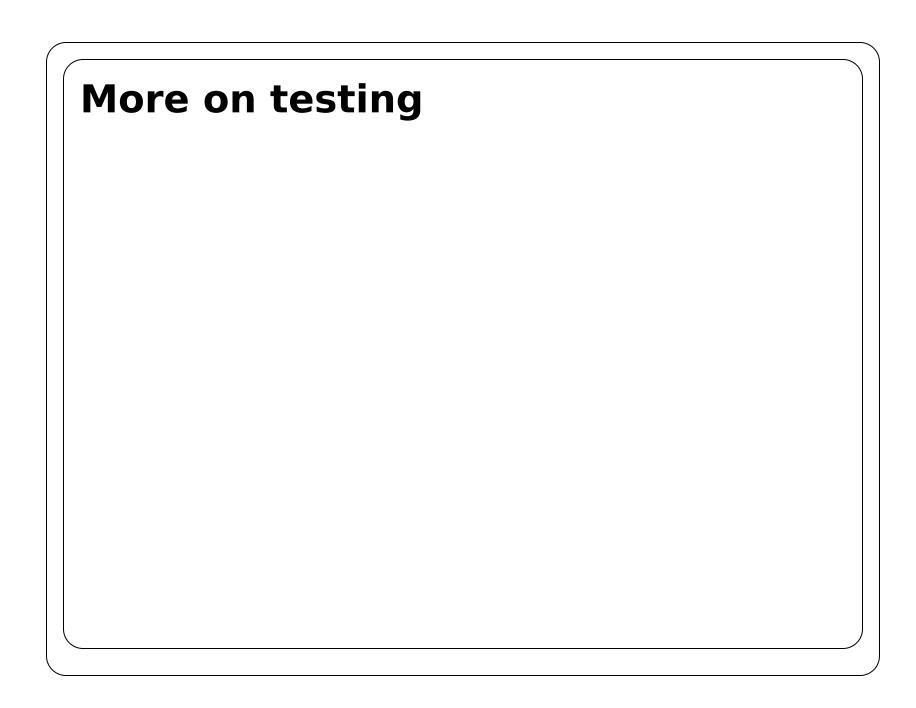






====== * Packages unittest doctest * Some simple unit tests * Two integration tests
======= * Packages unittest doctest * Some simple unit tests * Two integration tests
======= * Packages unittest doctest * Some simple unit tests * Two integration tests
======= * Packages unittest doctest * Some simple unit tests * Iwo integration tests



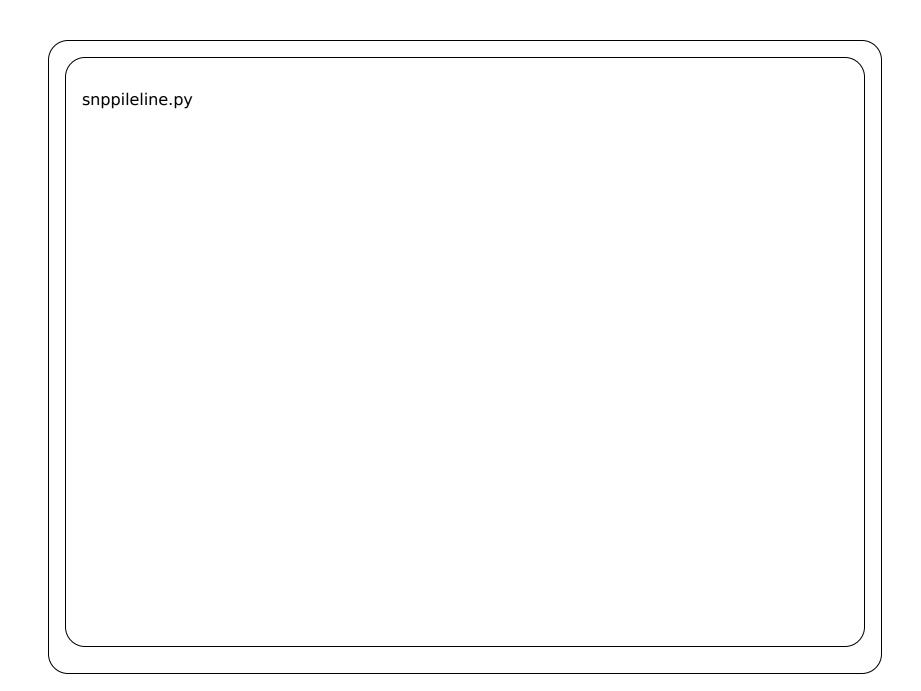


		- 1 1				
./test/test_ı	utils.py -v ok	5 items had no	o tests: utils util	s.pileup utils.pi	leup_wrapper	

utils.write list	of_snps utils.write	_reference snp f	ïle 2 items passe	d all tests: 3 tests	in
			·		

utils.	.create_cons	sensus_dict 9 t	tests in utils.g	jet_consensus	_base_from_pi	leup 12 tests	in 7

items. 12 passed	and 0 failed. Test p	assed. ok	

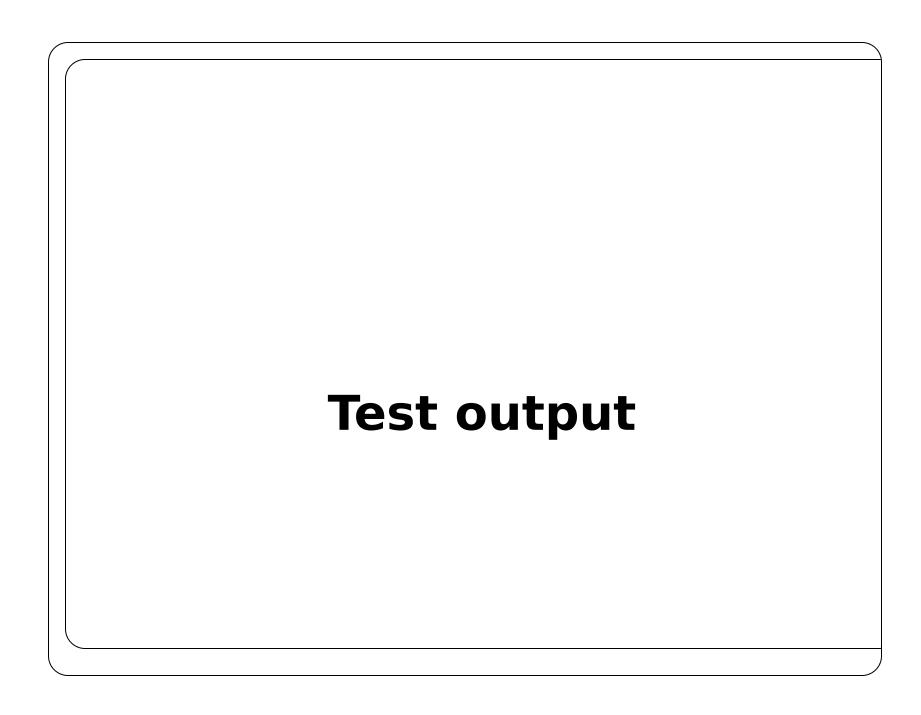


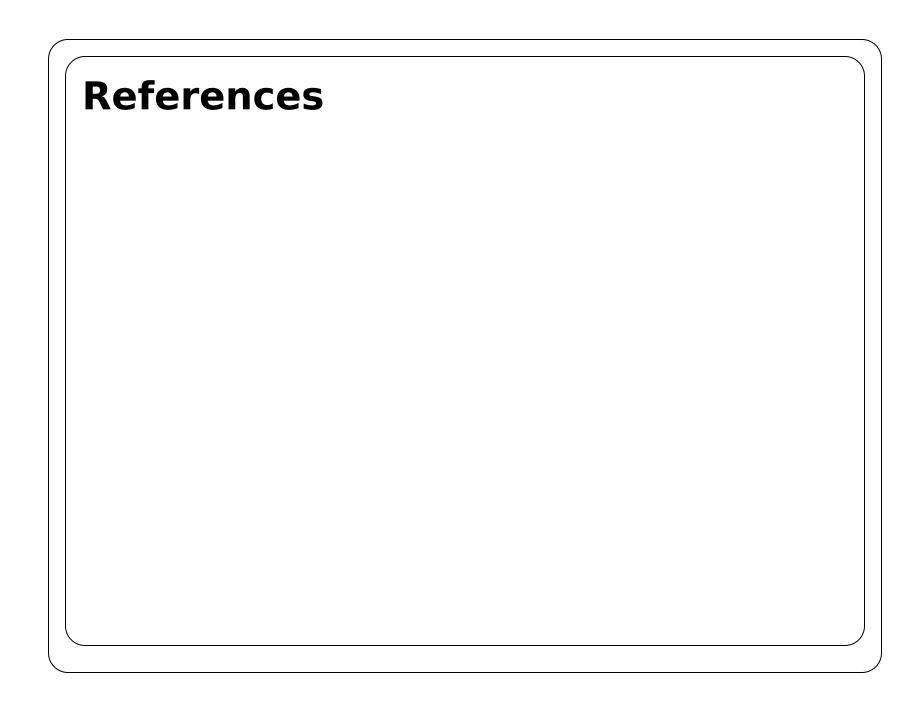
./test/test_snppipeline.py -v	test snopipeline agona	(main.Test) Run sopping	eline with agona ^s
, rest, rest_stippipelitie.py	cest_s.ippipee_agona	(main rest) rian shippip	sime with agona s

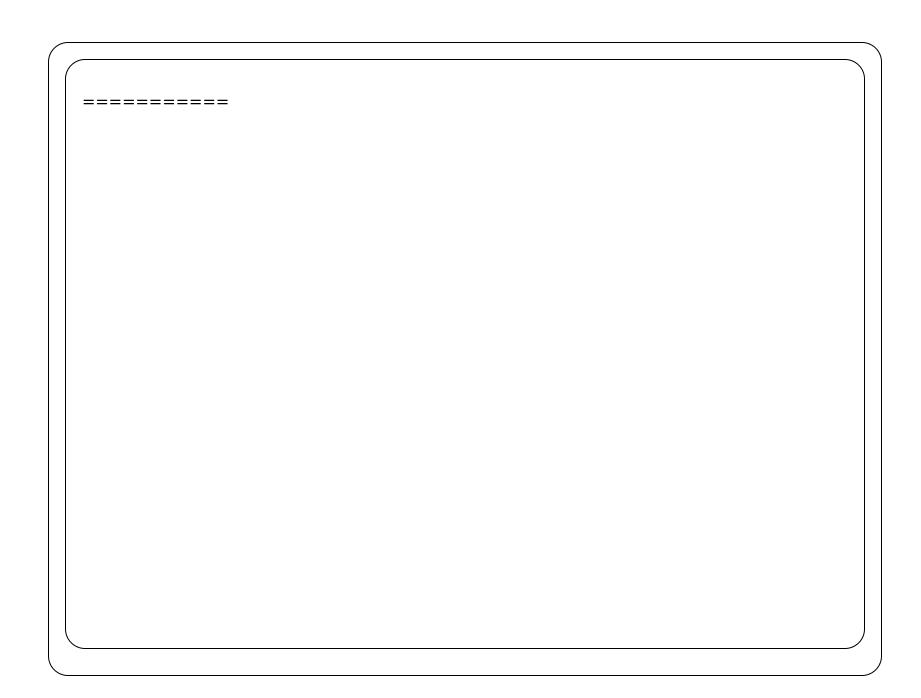
_		_
sa	ples example [mpileup] 1 samples in 1 input files Match, Mismatch, Errors: 8, 0, 0	1

ok test_snppipeline_lambda_virus (main .Test) Run snppipeline with synthetic virus example

[mpileup] 1 sar	nples in 1 input file	s Match, Misma	tch, Errors: 7, 0,	0 ok	







Python Eggs

A "Python egg" is a logica	al structure embodyina t	he release of a specific v	version of a Python
a Tython egg 13 a logica	in structure embodying t	ne release of a specific v	reform of a rythorn

pr	roject	t, com	prisir	ng its	code,	resou	ırces,	and r	netada	ta. Th	nere ai	re mult	iple fo	ormats	that o	can be

used to physically encode a Python egg, and others can be develop	oed. However, a key

principle	of Python eggs is that they should be discoverable and importable. That is, it sho	ould

_	
k	be possible for a Python application to easily and efficiently find out what eggs are present on
-	

a syste	em, and to e	nsure that th	e desired eg	gs' contents	are import	able.	

The .egg format is well-suited to distribution and the easy uninstallation or upgrades of code,

_	
S	since the project is essentially self-contained within a single directory or file, unmingled with
_	

any ot	her projects' code or resources. It also makes it possible to have multiple versions of a

proje	t simultaneously installed, such that individual programs can select the versions they	

wish to use.		

(http://stackoverflow.com/questions/2051192/what-is-a-python-egg)	

Building Python Eggs

Do alcomo a allatenti	lla aakumkaala Ca-la	antun nu		
Packages distut	ls setuptools Code	setup.py		