

Reference	SeqQual pipeline / program (prog) name*, <del>⚠</del>	usage**	arguments (arg)	description	using programs (other prog)	Shell examples **	Input file(s)	Output folder/file(s)	log file available\$
<b>file/folder management***</b>									
lang et al.	clean_all.pl	perl prog arg	( text file with list of folder names, e.g. inputile) = <b>(B)</b>	remove all "mydata" sub-folders in each folder of the inputfile list		clean_all.sh	see arg		
"	clean.pl	perl prog arg	<b>(B)</b>	remove older mydata sub-folders in each folder in the input list, keep the last one		clean.sh	see arg		
"	checkdir_mydata.pl	perl prog		create new folder mydata, if one already exists--> change it to mydata_current_date&time first				empty folder mydata	
"	checkdir_Output.pl	perl prog		create new folder Output, if it already exists, change it to Output_current_date&time first		1.2-haploid- ab1.sh, 1.1- diploid-ab1.sh		empty folder Output	
"	checkinput.pl	perl prog arg	<b>(B)</b>	check if the folders listed in inputfile exist, give a warning if not			see arg	screen	
"	<b>print_source</b> -delete_files.pl <del>⚠</del>	<b>(perl prog arg &gt; out source out ) = <b>(A)</b></b>	<b>(B)</b>	( run <b>other prog</b> in batch on a set of folder given in arg , when sourcing the "out" text file) = <b>(C)</b>	delete_empty_files.pl	1.2-haploid- ab1.sh, 1.1- diploid-ab1.sh	see arg		
"	delete_empty_files.pl <del>⚠⚠</del>	perl prog		delete empty fasta alignments (including ones with only consensus (names started with "Contig", can be changed in code) in the current directory				delete.txt	
"	<b>print_source</b> -take_aln.pl <del>⚠</del>	<b>(A)</b>	<b>(B)</b>	<b>(C)</b>	take_aln_to_output.pl	1.2-haploid- ab1.sh, 1.1- diploid-ab1.sh	see arg		<b>X</b>
"	take_aln_to_output.pl	perl prog arg	<b>(B)</b>	look for files with extension ".aln"; rename as filenamenumber.aln, take file to Output/aln folder					
<b>Processing ab1/scf/abd files with phd*/poly* files</b>									
"	<b>print_source</b> -aln-diploid-ab1.pl	<b>(A)</b>	<b>(B)</b> , polyphred score, polyphred quality, 18 parameters from phrap (see User doc & example *.sh)	<b>(C)</b> , create working directories and files; run phredPhrap to create poly and phd files, run phrap, run polyphred	checkdir_mydata.pl, <del>\$\$</del> , phd2fasta <del>\$\$</del> , phrap_all_lang.pl, polyphred <del>\$\$</del>	1.1-diploid- ab1.sh	see arg, ab1/scf/abd files		<b>X</b>
"	phrap_all_lang.pl	perl prog	file with list of file names, and 18 parameters for Phrap	runs phrap in batch and get ace files	phrap <del>\$\$</del>		see arg	ace files from Phrap	
"	<b>print_source</b> -userphd-diploid- ab1.pl	<b>(A)</b>	<b>(B)</b>	copy user poly files and phd files into working direcotry		1.1-diploid- ab1.sh	see arg, phd/poly files in same folder than ab1 files		<b>X</b>
"	<b>print_source</b> -renamephd.pl	<b>(A)</b>	<b>(B)</b>	<b>(C)</b>	rename_phd-ab1.pl, rename_phd-abd.pl, rename_phd-scf.pl	1.1-diploid- ab1.sh	see arg, phd/poly files in same folder than ab1 files		<b>X</b>
"	rename_phd-ab1.pl	perl prog		rename filename.phd.1 as filename.ab1.phd.1				rename-ab1.txt	
"	rename_phd-abd.pl	perl prog		rename filename.phd.1 as filename.abd.phd.1				rename-abd.txt	
"	rename_phd-scf.pl	perl prog		rename filename.phd.1 as filename.scf.phd.1				rename-scf.txt	
"	<b>print_source</b> -write_aln- diploid.pl	<b>(A)</b>	<b>(B)</b> , phred score, genotype score	<b>(C)</b> , moves edited fasta alignments into output folder	write_acealn- heter_multinput.pl	1.1-diploid- ab1.sh			<b>X</b>

"	write_acealn-heter_multinput.pl	perl prog arg	phred score, genotype score	look for file(s) with extension ".ace"; write diploid alignment(s) with heterozygote IUPAC codes from polyphred.out files (in phd_dir folder), integrates quality by accepting nucleotides as valid only if their phd score is >= to the arg value given (non-valid ones considered as missing data and coded by "?")			*.ace files	folder aln with fasta files	
"	<b>print_source</b> -aln-haploid-ab1.pl	(A)	(B) , 18 parameters from phrap	(C) , create working directories and files	checkdir_mydata.pl, phred <sup>\$\$</sup> , phd2fasta <sup>\$\$</sup> , phrap_all_lang.pl	1.2-haploid-ab1.sh	see arg, ab1/scf/abd files		X
"	<b>print_source</b> -userphd-haploid-ab1.pl	(A)	(B)	copy user phd files into working direcotry		1.2-haploid-ab1.sh	see arg, phd/poly files in same folder than ab1 files		X
"	<b>print_source</b> -write_aln-haploid.pl	(A)	(B), phred score	(C) , moves edited fasta alignments into output folder	write_acealn-onlyqual_multinput.pl	1.2-haploid-ab1.sh, 2.3-ace-qual.sh	see arg, phd/poly files in same folder than ab1 files	aln_final/*.aln fasta files	X
"	write_acealn-onlyqual_multinput.pl	perl prog arg	phred score	look for file(s) with extension ".ace"; write alignment(s) with quality from phd files in folder phd_dir, thuz accepting nucleotide as valid <b>only if</b> its phred score is >= to the arg value given, whether or not it is the same than the consensus sequence.			*.ace files	aln/*.aln fasta files	

- \*

all shell scripts can be run by typing "source \*.sh"
- ⌘

All print\_source scripts work by printing a txt file that needs to be sourced to launch other programs for batch treatment of files located in one or more folders. They also require a particular folder structure for printing results files (see start of example \*.sh files for details)
- ⌘⌘

example of one other program, which can be used also independently to the print\_source script
- \*\*

To run print\_source\*.pl scripts, most other scripts are assumed to be located under "home/SeqQual" but this can easily be changed in the code
- \$

see the description and list of log related scripts in **SeqQual-log-related-scripts.pdf**
- \$\$

from the phredphrap suite or polyphred programs that need to be installed (see User documentation)
- \*\*\*

these programs are needed if the print\_source programs are used