Publication	SeqQual pipeline / program (prog) name	usage*	arguments (arg)	description	using programs	Shell examples (with print_source perl scripts)*	Input file(s)	Output file(s)	new folder created	name of new folder created	Comments	folder where program files should be located**	todo
*	for all scripts, see example shell column												
**	most program locations are assumed to be under home/SeqQual but this can easily be changed in the code												
***	all programs can be used separately with your own batch command lines on fasta files (defined as above)												
	Miscellenious / file management												
lang et al.	clean_all.pl	perl clean_all.pl arg	text file with list of folder names	remove all mydata sub-folders in each folder in the input list			NA	NA					
lang et al.	clean.pl	perl clean.pl arg	text file with list of folder names	remove old mydata sub-folders in each folder in the input list, keep the lastest one			NA	NA					
	Program Name		Argument(s)	Function				Output					
lang et al.	delete_empty_files.pl			delete empty alignments (including ones with only consensus which begins with Contig), need to run source delete.txt				delete.txt					
"	change_ace_name.pl			change filename.ace to filename.ace.1 in the working older									
"	checkdir_mydata.pl			create folder mydata, if already exist, change to mydata_current_date&time first then create				empty folder mydata					
"	checkdir_Output.pl			create folder Output, if already exist, change to Output_current_date&time first then create				empty folder Output					
"	checkinput.pl		file with list of folder names	check if the folders listed in inputfile exist, give a warning if not				screen					
"	clean_all.pl		file with list of folder names	remove all mydata folders in each folder in the input list									
"	clean.pl		file with list of folder names	remove old mydata folders in each folder in the input list, keep the lastest one									
"	fasta2snp_no_first.pl			pick SNP alignment files from fasta files, not count the first alignment				folder SNP with files					
"	fasta2snp.pl			pick SNP alignment files from fasta files				folder SNP with files					
"	phrap_all_lang.pl		file with list of file names, and 18 parameters for Phrap	do phrap in batch and get ace files				ace files from Phrap					
"	polybayes_all_lang.pl		file with list of ace file names	do polybayes in batch and get polybayes.ace files				polybayes.ace files from Polybayes					
"	polyphred_all_lang.pl		file with list of ace file names, polyphred score, polyphred quality	do polyphred in batch and get polyphred.out files				updated ace files from polyphred, and polyphred.out					
"	remove-bad-pos_aln.pl			look for file with name .aln; remove the column with only "?" and "-"				folder aln_remove with files					
"	rename_phd-ab1.pl			rename filename.phd.1 as filename.ab1.phd.1				rename-ab1.txt					
"	rename_phd-abd.pl			rename filename.phd.1 as filename.abd.phd.1				rename-abd.txt					
"	rename_phd-scf.pl			rename filename.phd.1 as filename.scf.phd.1				rename-scf.txt					
"	replace-1-nucleotide_aln.pl			look for file with name .aln; replace single isolated nuclotides surrounded by "?" with "?"				folder aln_replace1 with files					
"	replace-2-nucleotide_aln.pl			look for file with name .aln; replace single and two neighbor isolated nuclotides surrounded by "?" with "?"				folder aln_replace2 with files					
"	replace-3-nucleotide_aln.pl			look for file with name .aln; replace single, two neighbor and three neighbor isolated nuclotides surrounded by "?" with "?"				folder aln_replace3 with files					
"	separate454ace.pl		name of ace file	splite big ace file into small ace files which have only one ace assembly				ace files					
"	separate454phd.pl		name of phd file	splite big phd file into small phd files which have only one phd assembly				phd files					
"	take_aln_to_output.pl		file with list of folder names	look for file with name filename ".aln"; rename as filenamenumber.aln, take file to Output/aln folder									
"	take_arp_to_output_diploid.pl		file with list of folder names	look for file with name filename ".arp"; rename as filenamenamenumber.dip(or hap).arp (depends on whether filename contains "genotypicdata0" or "genotypic data1"), take file to Output/arlequin folder									
"	take_arp_to_output_haploid.pl		file with list of folder names	look for file with name filename ".arp"; rename as filename.aln.arp, take file to Output/arlequin folder									
"	take_haplotypealn_to_output.pl		file with list of folder names	look for file with name filename ".haplotype.aln"; rename as filenamenumber.haplotype.aln, take file to Output/aln_haplotype folder									
"	take_log_to_output.pl			take log.txt to folder Output, check name to log_data&time.txt									
"	take_SNP-fasta2snp_to_output.pl		file with list of folder names	look for file with name filename ".aln.snp"; rename as filename.snp.aln, take file to Output/SNP folder									
"	take_SNP_to_output.pl		file with list of folder names	look for file with name filename ".snp.aln"; take file to Output/SNP folder									
"	take_unaln_to_output.pl		file with list of folder names	look for file with name filename ".unaln"; take file to Output/unaln folder									
"	trunc_aln.pl		missing data number in one column from which truncate will be done	look for file with name ".aln"; truncate start and end if the number of missing data from that column is above the argument number				folder aln_trunc with files					
"	write_acealn-heteraw_multinput.pl		phred score, heterzygotes score	look for file with name ".ace"; write alignment with quality and heterzygote by folder phd_dir/poly_dir which has phd/poly files. The range for poly score is 0.5-1. 1 means only 1 peak, 0.5 means perfect double peak.				folder aln with files					
"	write_acealn-heter_multinput.pl		phred score, genotype score	look for file with name ".ace"; write alignment with quality and heterzygote by folder phd_dir which has phd files and by file polyphred.out				folder aln with files					
"	write_acealn-nonqual_multinput.pl			look for file with name ".ace"; write alignment				folder aln with files					

"	write_acealn-qual_multinput.pl	phred score	look for file with name ".ace"; write alignment with quality by folder phd_dir which has phd files			folder aln with files
"	write_acesnp-heter_multinput.pl	phred score, genotype score	look for file with name ".ace"; write SNP alignment with quality and heterzygote by folder phd_dir which has phd files and by file polyphred.out			folder SNP with files
"	write_aceunaln-heteraw_multinput.pl	phred score, heterzygotes score	look for file with name ".ace"; write unaligned sequences with quality and heterzygote by folder phd_dir/poly_dir which has phd/poly files			folder unaln with files
"	write_aceunaln-heter_multinput.pl	phred score, genotype score	look for file with name ".ace"; write unaligned sequences with quality and heterzygote by folder phd_dir which has phd files and by file polyphred.out			folder unaln with files
"	write_aceunaln-qual_multinput.pl	phred score	look for file with name ".ace"; write unaligned sequences with quality by folder phd_dir which has phd files			folder unaln with files
"	write_unaln-nonqual.pl	file with list of folder names	look for file with name ".aln"; write unaligned sequences without quality			folder unaln with files
"	write_arlequin_input_diploid-genotypicdata0_multinput.pl	up to 20 group names	look for file with name ".aln"; write phase unknow haplotypic type arlequin input file for diplotypic type alignment			folder arlequin_input0 with files
"	write_arlequin_input_diploid-genotypicdata1_multinput.pl	up to 20 group names	look for file with name ".aln"; write phase unknow diplotypic type arlequin input file for diplotypic type alignment			folder arlequin_input1 with files
"	write_arlequin_input_multinput.pl	up to 20 group names	look for file with name ".aln"; write haplotypic type arlequin input file for haplotypic type alignment			folder arlequin_input with files
"	write_haplotype_phase_unknown_multinput_nofirst.pl		look for file with name ".aln"; write haplotypic type phase unknown alignment for diplotypic alignment without the first sequence (commonly consensus sequence)			folder aln_haplotype with files
"	write_haplotype_phase_unknown_multinput.pl		look for file with name ".aln"; write haplotypic type phase unknown alignment for diplotypic alignment			folder aln_haplotype with files
	Program Name	**Argument(s) other than file with list of folder names**	**Description (should be printed to txt files and then run source)**	**Using program**	**Being used in Shell (print_source*)**	
"	print_source-delete_files.pl		run delete empty files	delete_empty_files.pl	ace-only, diploid-ab1, diploid-ace, diploidraw-ab1, haploid-ab1, haploid-ace, haploid-ace, 454	
"	print_source-454ace_changename.pl		use change_name_ace.pl	change_ace_name.pl	454	
"	print_source-454ace_separate.pl		use separate454ace.pl and move old ace file into oldace folder	separate454ace.pl	454	
"	print_source-454phd_separate.pl		use separate454phd.pl	separate454phd.pl	454	
"	print_source-acedealing-diploid.pl		create working directories and files; run polyphred, change name.ace to name.fasta.ace	checkdir_mydata.pl, polyphred_all_lang.pl, change_acename.pl	diploid-ace	
"	print_source-acedealing-haploid.pl		create working directories and files;	checkdir_mydata.pl, polypayes_all_lang.pl	haploid-ace, 454	
"	print_source-aceonly.pl		create working directories and files;	checkdir_mydata.pl, polypayes_all_lang.pl	ace-only	
"	print_source-aln-diploid-ab1.pl	polyphred score, polyphred quality, 18 parameters from phrap	create working directories and files; run phredPhrap to create poly and phd files, run phrap, run polyphred	checkdir_mydata.pl, phrap_all_lang.pl	diploid-ab1	
"	print_source-aln-diploidraw-ab1.pl	18 parameters from phrap	create working directories and files; run phred, phd2fasta, phrap	checkdir_mydata.pl, phrap_all_lang.pl	diploidraw-ab1	
"	print_source-aln-haploid-ab1.pl	18 parameters from phrap	create working directories and files; run phred, phd2fasta, phrap	checkdir_mydata.pl, phrap_all_lang.pl,	haploid-ab1	
"	print_source-arlequin-diploid.pl	up to 10 group names	run write arlequin diploid, put into folder arlequin_input0 and arlequin_input1	write_arlequin_input_diploid-genotypicdata0_multinput.pl, write_arlequin_input_diploid-genotypicdata1_multinput.pl	diploid-ab1, diploid-ace, diploidraw-ab1, fasta-diploid-data	
"	print_source-arlequin-haploid.pl	up to 10 group names	run write arlequin haploid, put into folder arlequin_input	write_arlequin_input_multinput.pl	ace-only, fasta-haploid-data, haploid-ab1, haploid-ace, 454	
"	print_source-fasta.pl		create working directories and files	checkdir_mydata.pl	fasta-diploid-data, fasta-haploid-data	
"	print_source-remove.pl		run remove bad position, rename folder aln_final if existed, put result files into folder aln_final	remove-bad-pos_aln.pl	diploid-ab1, diploid-ace, diploidraw-ab1, haploid-ab1, haploid-ace, haploid-ace, 454	
"	print_source-renamephd.pl		rename phd files from ab1, abd, scf	rename_phd-ab1.pl, rename_phd-abd.pl, rename_phd-scf.pl	diploid-ab1, diploidraw-ab1, haploid-ab1	
"	print_source-replace.pl	number show which replace program used	run replace isolated nucleotides, rename folder aln_final if existed, put result files into folder aln_final	replace-nucleotide_aln.pl	diploid-ab1, diploid-ace, diploidraw-ab1, haploid-ab1, haploid-ace, haploid-ace, 454	
"	print_source-take_aln.pl		take alignment Output folder, change file name	take_aln_to_output.pl	ace-only, diploid-ab1, diploid-ace, haploid-ab1, haploid-ace, 454	
"	print_source-take_arp_diploid.pl		take diploid arlequin input files (two types) to Output folder, create arb file	take_arp_to_output_diploid.pl	diploid-ab1, diploid-ace, fasta-diploid-data	
"	print_source-take_arp_haploid.pl		take haploid arlequin input files without to Output folder, create arb file	take_arp_to_output_haploid.pl	ace-only, fasta-haploid-data, haploid-ab1, haploid-ace, 454	

"	print_source-take_haplotypealn.pl		take haplotype alignment Output folder	take_haplotypealn_to_output.pl	diploid-ace, fasta-diploid-data
"	print_source-take_SNP-fasta2snp.pl		take SNP alignment to Output folder	take_SNP-fasta2snp_to_output.pl	fasta-diploid-data, fasta-haploid-data
"	print_source-take_SNP.pl		take SNP alignment to Output folder	take_SNP_to_output.pl	ace-only, diploid-ace, haploid-ace, haploid-ace, 454
"	print_source-take_unaln.pl		take unaligned fasta sequence to Output folder	take_unaln_to_output.pl	diploid-ace, haploid-ace, haploid-ace, 454
"	print_source-truncate.pl	missing data number in one column from which truncate will be done	run truncate alignment, rename folder aln_final if exist, put result files into folder aln_final	trunc_aln.pl	ace-only, diploid-ab1, diploid-ace, diploidraw-ab1, haploid-ab1, haploid-ace, haploid-ace, 454
"	print_source-userphd-diploid-ab1.pl		copy user poly files and phd files into working direcotry		diploid-ab1, diploidraw-ab1
"	print_source-userphd-haploid-ab1.pl		copy user phd files into working direcotry		haploid-ab1
"	print_source-write_aln-aceonly.pl	phred score	write alignment from ace	write_acealn-nonqual-multinput.pl	ace-only
"	print_source-write_aln-diploid.pl	phred score, genotype score	write alignment from ace, phd, polyphred.out	write_acealn-heter_multinput.pl	diploid-ab1, diploid-ace
"	print_source-write_aln-diploidraw.pl	phred score, heterzygotes score	write alignment from ace, phd, poly	write_acealn-heteraw_multinput.pl	diploidraw-ab1
"	print_source-write_aln-haploid.pl	phred score	write alignment from ace, phd	writea_acealn-qual_multinput.pl	haploid-ab1, haploid-ace, 454
"	print_source-write_halotypealn_nofirst.pl		write phase unknown alignment from diploid alignment without consensus sequence	write_haplotype_phase_unknown_multinput_nofirst.pl	diploid-ab1, diploid-ace, diploidraw-ab1, fasta-diploid-data
"	print_source-write_halotypealn.pl		write phase unknown alignment from diploid alignment	write_haplotype_phase_unknown_multinput.pl	diploid-ab1, diploid-ace, diploidraw-ab1, fasta-diploid-data
"	print_source-write_SNP-diploid.pl	phred score, genotype score	write SNP alignment from ace, polyphred.out	write_acesnp-heter_multinput.pl	diploid-ab1, diploid-ace
"	print_source-write_SNP-fasta_aln-no_first.pl		write SNP alignment from fasta without consensus	fasta2snp_no_first.pl	fasta-diploid-data, fasta-haploid-data
"	print_source-write_SNP-fasta_aln.pl		write SNP alignment from fasta	fasta2snp.pl	fasta-diploid-data, fasta-haploid-data
"	print_source-write_unaln-diploid.pl	phred score, genotype score	write unaligned fasta from ace, phd, polyphred.out	write_aceunaln-heter_multinput.pl	diploid-ab1, diploid-ace
"	print_source-write_unaln-diploidraw.pl	phred score, heterzygotes score	write unaligned fasta from ace, phd, poly	write_aceunaln-heteraw_multinput.pl	diploidraw-ab1
"	print_source-write_unaln-haploid.pl	phred score	write unaligned fasta from ace, phd	write_aceunaln-qual_multinput.pl	haploid-ab1, haploid-ace, 454
"	print_source-write_unaln-fasta.pl		write unaligned fasta from aligned fasta	write_unaln-nonqual.pl	fasta-diploid-data, fasta-haploid-data