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**
Fasta alignments editing and filtering for subsequent analyses**												
El mujtar et al. 2014	change_dash_to_questionmark.pl	perl prog arg	names of input files	Produces new file in which "---" at the beginning and end will be changed into "???".			*.aln fasta files	*.aln fasta files	no	NA		defined by the user
El mujtar et al. 2014	split_aln_multi.pl	perl prog arg	nb of bp after which each file is split	looks for all *.aln/*.fas files in same folder (add folder in path otherwise), will create split files in a new folder according to arg and rename them with corresponding suffix,					yes	aln_splitted	useful for very large contigs with many reads wher the average depth is low, allows to speed up SNP-statistic programs afterwards	~/SNP_statistic
Lang et al.	split_aln_one.pl	perl prog arg filename	name of input file	same prog than above for one file			one fasta file		no			
"	print_source-delete_files.pl			run delete empty files (via printing txt files and then run source delete.txt)	Function (should be printed to txt files and then run source) delete_empty_files.pl	ace-only, diploid-ab1, diploid-ace, diploidraw-ab1, haploid-ab1, haploid-ace, haploid-ace, 454						
"	delete_empty_files.pl			delete empty fasta alignments (including ones with only consensus (name started with Contig)) in the current directory, need to run source delete.txt produced by program above		"	delete.txt					
El mujtar et al. 2014	delete_empty_seqs.pl	perl prog	none	Looks for *.aln/*.fas files. Delete read lines which only contain "???" Or "---"(empty seqs, for ex after maskin steps) and produces new files without those lines in a new folder. Names them "OriginalSuffix".del.aln/fas.			*.aln fasta files	*.aln fasta files	yes	aln_del_empty_seq	Prog will produce lots of warning on the screen. So use > log.txt if you don't want them to be on the screen.	~/SNP_statistic
Brousseau et al. 2014, El mujtar et al. 2014	**get_consensus.pl**	perl prog	none	Looks for *.aln files. Produces as many files as fasta alignments only including the first line (assumed to be the consensus) in a new folder. Names them "OriginalSuffix_consensus.aln". Also gives all first lines in one fas file.			*.aln fasta files	*.aln fasta files	yes	aln_consensus		defined by the user
El mujtar et al. 2014	homomaskAln.pl	perl prog arg > log.txt	nb of bp before or after the insertion for which the deletions/insertions are masked, works also for SNPs with false deletions in those homopolymer regions	Looks for *.aln files. Produces files in which some deletions are masked (replaced by "?") in a new folder. The deletions to be masked occur in homopolymers regions for a defined number of repeated nucleotides (the arg). Names the files "OriginalSuffix.homomasked.aln".			*.aln fasta files	*.aln fasta files	yes	aln_homomask	Use with > log.txt to get all std output onscreen so it is easier to stop with CtrlZ & run it in the background (type bg after Ctrl Z). Warning: do not relaunch the same prog if an aln_homomask* folder already present, move the aln_homomask folder up first.	~/SeqQual
Brousseau et al. 2014, El mujtar et al. 2014	make_consensus_IUPAC/_maxallele.pl	perl prog	none	Looks for *.aln files. Produces files with new consensus (either using IUPAC codes when there is a SNP (for make_consensus_IUPAC.pl) or using the allele in highest frequency (make_consensus_maxallele.pl) in a new folder. Name them "OriginalSuffix.nc.aln".			*.aln fasta files	*.aln fasta files	yes	aln_newconsensus		defined by the user
Brousseau et al. 2014	maskAln.pl	perl prog --man --man_if_depth --maf --indel_only	maf= minimum allele frequency, man=minimum allele number, man_if_depth=minimum depth for which masking is on for the man given, indel_only=if yes only performs masking on indels, if no does it across all variants	according to either maf or man, will mask (=replace by "?") any nucleotide which is a variant below the given maf or man.			*.aln fasta files	*.aln fasta files	yes	**aln_clean --> to change in aln_mask**	maf and man can't be used together, man-if-depth is used with man only. to use the program in batch across *.aln fasta file, type: for file in *.aln; do perl ~/folder.where.the.program.is/maskAln.pl --maf=0.125 --file=$file > $file.cleaned.aln ; done; Usage example: perl maskAln.pl --man=1 --man_if_depth=10 --maf=0.05 --indel_only=no; default man=1, man_if_depth=15, maf=0.05, indel_only=no; to use the program in batch across *.aln fasta file, type: for file in *.aln; do perl ~/path where the program is/maskAln.pl --maf=0.125 --file=$file > $file.cleaned.aln ; done;	~/SeqQual
Brousseau et al. 2014	pick-seq.pl	perl prog arg *.aln	arg indicates an optional suite of characters for selecting in batch in a series of fasta files (*.aln) another series including only the reads containing this character chain. The character chain needs to be at the start of the reads ID ; ex: pop1 pop2	looks for set of *.aln fasta files. Select in each of them a number of reads identified by the arg and produces another set of *.aln files in a separate folder			*.aln fasta files	*.aln fasta files	yes	aln_pick		defined by the user
Brousseau et al. 2014, El mujtar et al. 2014	remove_consensus.pl	perl prog	none	Looks for *.aln files. Produces files without the first line (assumed to be the consensus) in a new folder. Names them "OriginalSuffix.removeconsensus.aln".			*.aln fasta files	*.aln fasta files	yes	aln_removeconsensus		defined by the user
Brousseau et al. 2014	print_source-remove1.pl	perl prog inputfile > source-remove1.txt	inputfile is the folder name where all fasta files *.aln are located	Produces a text file to source for running the remove1-bad-pos_aln.pl program	remove1-bad-pos_aln.pl		text file names inputfile with folder list	text file to source	no	NA		~/SeqQual
Brousseau et al. 2014	remove1-bad-pos_aln.pl	from source-remove1.txt file above, perl prog	none	Needs first print_source-remove1.pl to run as it is launched from the source-remove1.txt file, assumes programs are located under the home/SeqQual folder (~/SeqQual/ but this can be changed in the code). Looks for a set of fasta files in the folder or list of folders located below the current folder, and in each file, removes the positions which contains only "?" and "-" if the consensus sequence is not an "-".			*.aln fasta files	*.aln fasta files	yes	aln_remove	if the folder aln_remove already existe, a warning is given and the program stops. You need to delete or rename the old folder	~/SeqQual
Fasta alignments post-treatment for genetic analyses and SNP detection**												
Brousseau et al. 2014, El mujtar et al. 2014	print_source-SNP_statistic_haplo.pl	perl prog inputfile arg> source-SNP_statistic.txt	inputfile = text file including folder name where **haploid data** fasta files (*.aln) are located; arg= prefix in sequence/read names identifying different groups, for ex: group1 group2; only one group identifier can be used; arg can be ignored if all reads are treated as a single group	script to run first for sourcing SNP_statistic0/1/2-haplo.pl program command lines for none or up to 2 groups **(see example shell file)**	SNP-statistic0/1/2-haplo.pl	TO COMPLETE	text file named inputfile with folder list					~/SeqQual
Brousseau et al. 2014, El mujtar et al. 2014	SNP_statistic0/1/2-haplo.pl	perl prog arg > resultfile.txt	arg= same than for the print_source-SNP_statistic_haplo.pl, but programs can be run separately	either run it from the source-SNP_statistic.txt file produced by script above, or to use separately use, computes in batch counts of different base/indels for any polymorphism, depth, maf, exclusive & shared alleles, and if 2 groups: divergence statistics (Gst, Gst'), etc...		TO COMPLETE	*.aln fasta files	tabulated resultfile.txt (default name) as the program runs in the same folder + output file printed at the end of the script giving the list of alignments treated, their for each alignment/contig the reads number, total polymorphisms number etc...	no, output files in the same folder than the script	NA		~/SeqQual
lang et al.	print_source-SNP_statistic.pl	perl prog inputfile arg > source-SNP_statistic.txt	inputfile = text file including the folder name where **diploid data** fasta files (*.aln) are located; arg= prefix in sequence/read names identifying no, 1 or 2 different groups, for ex: group1 group2; arg can be ignored if all reads are treated as a single group	To run first for sourcing SNP_statistic0/1/2.pl program command lines for none or up to 2 groups (see example shell file)	SNP-statistic0/1/2.pl	TO COMPLETE	text file named inputfile with folder list	text file to source	no	NA		
"	SNP_statistic0/1/2.pl	perl prog arg > resultfile.txt	arg= same than for the print_source-SNP_statistic.pl, but programs can be run separately	either run it from the source-SNP_statistic.txt file produced by script above, or to use separately use, computes in batch counts of different base/indels for any polymorphism, depth, maf, exclusive & shared alleles, and if 2 groups: divergence statistics (Gst, Gst'), etc...		TO COMPLETE	*.aln fasta files	tabulated resultfile.txt (default name) as the program runs in the same folder	no, output files in the same folder than the script	NA	transform first *.aln fasta files into haploid phase unknown fasta to get the other output file giving the list of alignments treated, their number of reads and sums or means of all statistic across alignments counts	~/SeqQual
"	print_source-fasta.pl	print_source-fasta.pl inputfile > source-fastadealing.txt	file with list of folder names	create working directories and files	checkdir_mydata.pl	fasta-diploid-data, fasta-haploid-data						
"	checkdir_mydata.pl											
"	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						
"	fasta2snp_no_first.pl											
"	print_source-write_SNP-fasta_aln.pl			write SNP alignment from fasta	fasta2snp.pl	fasta-diploid-data, fasta-haploid-data						
"	fasta2snp.pl											
"	print_source-take_SNP-fasta2snp.pl			take SNP alignment to Output folder	take_SNP-fasta2snp_to_o	fasta-diploid-data, fasta-haploid-data						
"	take_SNP-fasta2snp_to_output.pl											
"	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						
"	take_SNP_to_output.pl											
"	print_source-write_halotypealn_nofirst.pl			write phase unknown alignment from diploid alignment without consensus sequence (assumed to be the first line)	write_haplotype_phase_u	diploid-ab1, diploid-ace, diploidraw-ab1, fasta-diploid-data						
"	write_haplotype_phase_unknown_multinput_nofirst.pl											

"	print_source-write_halotypealn.pl		write phase unknown alignment from diploid alignment	write_haplotype_phase_u	diploid-ab1, diploid-ace, diploidraw-ab1, fasta-diploid-data
"	write_haplotype_phase_unknown_multinput.pl				
"	print_source-take_haplotypealn.pl		take haplotype alignment Output folder	take_haplotypealn_to_ou	diploid-ace, fasta-diploid-data
"	take_haplotypealn_to_output.pl				
	print_source-arlequin-diploid.pl	up to 10/20 (checks with Tiange) group names	produces arlequin arp files from diploid fasta alignments, both type, either the diploid format with IUPAC codes (*.dip.arpthat can be used for phasing haplotypes in Arlequin, and also the haploid format with haplotype phase unknown that can be used for computing sumarry statistics, also produces the arb files with the list of arp files allowing batch analyses in Arlequin (put into folder arlequin_input0 and arlequin_input1 --> summarizes by saying thes go into arlequin folder)	write_arlequin_input_dip loid-genotypicdata0_multinp ut.pl, write_arlequin_input_dip loid-genotypicdata1_multinp ut.pl	diploid-ab1, diploid-ace, diploidraw-ab1, fasta-diploid-data
	print_source-take_arp_diploid.pl take_arp_to_output_diploid.pl		take diploid arlequin input files (two types) to Output folder, create arb file	take_arp_to_output_diplo	diploid-ab1, diploid-ace, fasta-diploid-data
	print_source-take_arp_haploid.pl take_arp_to_output_haploid.pl		take haploid arlequin input files without to Output folder, create arb file	take_arp_to_output_hapl	ace-only, fasta-haploid-data, haploid-ab1, haploid-ace, 454

in blue: programs modified since original publications

* 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)

All print_source scripts work by printing a txt file that needs to be sourced to launch another prog for batch treatment of fasta files