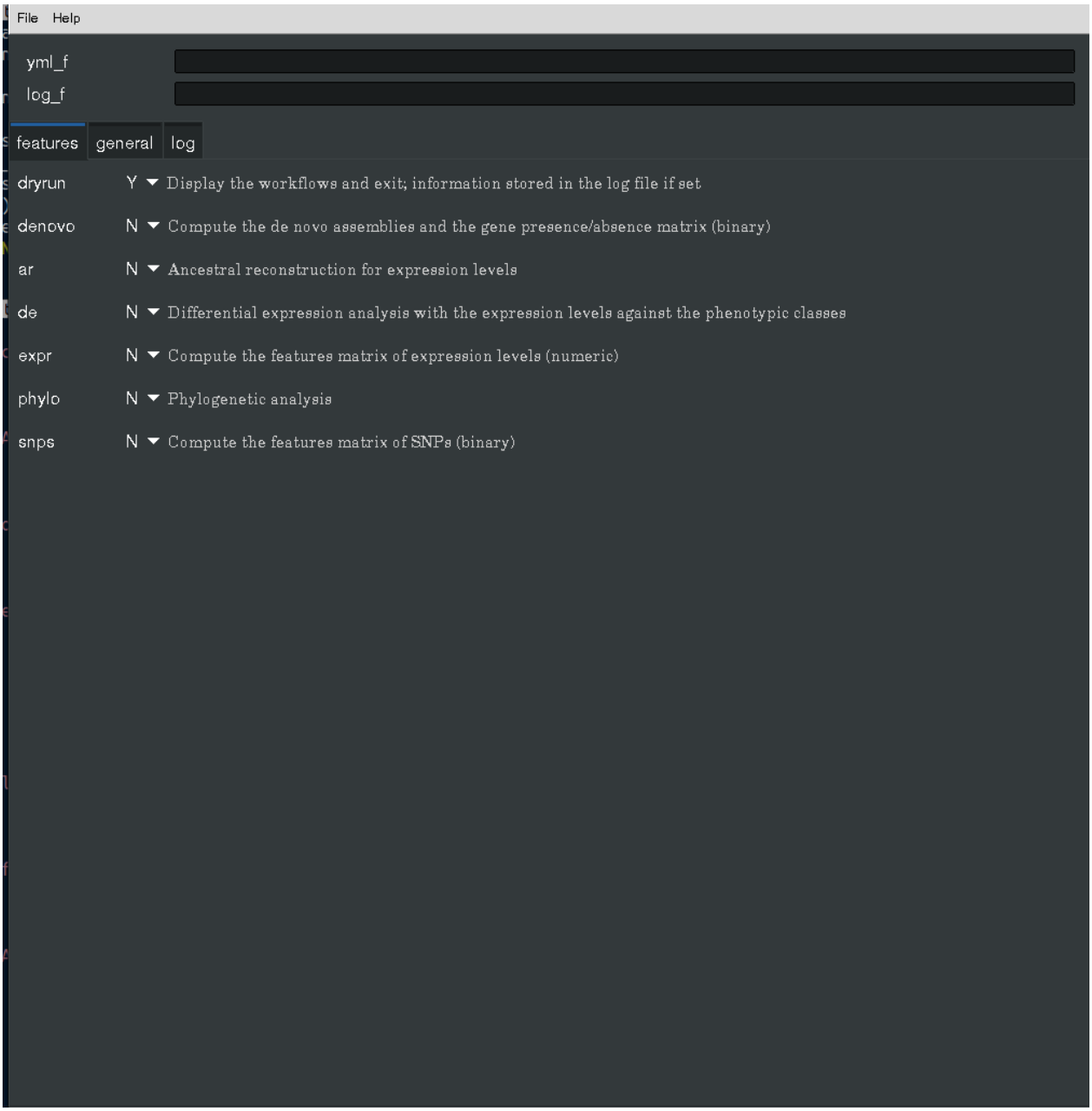
The manual for the graphical user interface of Seq2Geno: the first stage of Seq2Geno2Pheno



In the menu bar, *File* includes the options to read or save files. *Load yaml* will read a yaml file that was formatted for Seq2Geno, and the values in the graphical interface will be updated accordingly. By clicking *Save yaml,* a dialogue window will be opened for determining the filename of the yaml file to write the argument settings. *Load log* opens the Seq2Geno log file, displays the content in the *log* subpanel (next to features and general) and updates the values in the interface accordingly. *Run* will allow the user to save the yaml file and subsequently launch Seq2Geno in the background with the saved settings; to learn the status, use *Load log* to view the information.

panels			**description**
yml_f			The yaml file as the input of Seq2Geno, where the arguments described in the main panel will be written
log_f			If specified, the log file will store messages that are redirected from stdout and stderr from Seq2Geno. For dryrun mode, the procedures will be printed.
main			The informations that are described in the Seq2Geno input yaml file
	features		The workflows for computing features
		denovo	Compute the de novo assemblies and the gene presence/absence matrix (binary)
		snps	Compute the features matrix of SNPs (binary)
		expr	Compute the features matrix of expression levels (numeric)
		phylo	Phylogenetic analysis
		ar	Ancestral reconstruction for expression levels
		de	Differential expression analysis with the expression levels against the phenotypic classes
		dryrun	Display the workflows and exit; information stored in the log file if set
	general		Input data and other files
		cores	number of cpus
		mem_mb	memory size (mb)
		old_config	re-use the procedure-specific config files previously generated in the project folder (wd)
		dna_reads	The list of DNA-seq data (paired-end reads)
		wd	Working directory
		phe_table	The list of phenotypes
		ref_fa	The fasta file of reference genome; only ONE sequence should be contained
		ref_gbk	The genbank file of reference genome; only ONE chromosome should be contained
		ref_gff	The gff file of reference genome; only ONE chromosome should be contained
		rna_reads	The list of RNA-seq data (short reads)
		adaptor	The fasta file of adaptor sequences for trimming reads

Example :

Step 1. decompress the example dataset and follow the instruction to prepare an yaml file and the input data.

Step 2. launch Seq2Geno_gui and from the *Menu* bar select *File>Load yaml*, and import the example yaml file

Step 3. determine the filename of log file and review all the settings. You might want to opt dryrun if feel concerned, while the procedure-specific environments won't be installed. Alternatively, retain and run only one function one by one might be more flexible and error-proof.

Step 4. From the *Menu* bar, select *Load log* and find the previously specified log file. Then, click the *log* subpanel. The information may include error messages. For *dryrun* mode, the log file still includes all planned procedures although they were not conducted.