RE: Error in implementing MultiGWA

19 messages

Paula Helena Reyes Herrera <phreyes@agrosavia.co> Fri, Apr 8, 2022 at 8:28 AM
To: lgarreta <lgarreta@gmail.com>
Cc: RN Sarma <ramendra.sarma@aau.ac.in>

Hi Luis,

Could you address this issue?
Thanks

Paula

De: RN Sarma <ramendra.sarma@aau.ac.in>
Enviado: viernes, 8 de abril de 2022 6:34 a. m.
Para: Paula Helena Reyes Herrera <phreyes@agrosavia.co>
Asunto: Error in implementing MultiGWA

RNsarma-GBS_Final_filtered_0.9.recode_ajt.vcf

Hi,
Greetings from AAU Jorhat!

While running jmultigwas, noo results are displayed with my VC file and phenotypic file with chromosome number 24 (diploid). All files are empty.

Please suggest.
--
Warm regards,

Dr R N Sarma
Professor
Department of Plant Breeding and Genetics
Assam Agricultural University
Jorhat-785013
Assam, India
Orcid id: 0000-0002-9106-7926
web: www.aau.ac.in; Phone: +91- 9435350529(M)

HJ

AGROSAVIA
Corporación colombiana de investigación agropecuaria

Paula Helena Reyes Herrera
Investigador Ph.D.
Coordinación de Innovación Regional
C.I. Tibaitatá · Mosquera, Cundinamarca

Luis Garreta <lgarreta@gmail.com> Fri, Apr 8, 2022 at 9:31 AM
To: Paula Helena Reyes Herrera <phreyes@agrosavia.co>

Listo. Lo voy a revisar más tarde.
[Quoted text hidden]

RN Sarma <ramendra.sarma@aau.ac.in> Sun, Apr 10, 2022 at 9:02 PM
To: lgarreta@gmail.com

Hi,
I need your help to use MULTIGWAS. PLease check my files.
[Quoted text hidden]

2 attachments

phenotype.csv
16K

multiGWAS.config
1K

Luis Garreta <lgarreta@gmail.com> Mon, Apr 11, 2022 at 9:59 AM
To: RN Sarma <ramendra.sarma@aau.ac.in>

Hi Mr: Sarma:
I'm the developer of MultiGWAS, I have both your phenotype and config files, but you didn't send your genotype (VCF format) as I read in the config file. Could you send the genotype to run MultiGWAS in my environment?

Thanks,

Luis Garreta

RN Sarma <ramendra.sarma@aau.ac.in> Mon, Apr 11, 2022 at 10:30 AM
To: lgarreta@gmail.com

The vcf file is shared.
[Quoted text hidden]

2 attachments

phenotype.csv
16K

multiGWAS.config
1K

RN Sarma <ramendra.sarma@aau.ac.in> Mon, Apr 11, 2022 at 7:50 PM
To: Luis Garreta <lgarreta@gmail.com>

Hi Mr Garreta,

Thank you for your email. Now I am sending the VCF file. It is a diploid species with 2n=24. Kindly check and suggest.

RNsarma-GBS_Final_filtered_0.9.recode_ajt.vcf

[Quoted text hidden]

Luis Garreta <lgarreta@gmail.com> Tue, Apr 19, 2022 at 8:39 AM
To: RN Sarma <ramendra.sarma@aau.ac.in>

Hi Mr. Sarma:
I'm sending you some MultiGwas results from your data, a GWAS analysis for additive model, naive approach, without filters.

In the following days I will write to you about the MultiGwas changes and how to update them. Also, I will copy all the results to a repository for download.

Some issues I found were:
1. Your phenotypes were not formatted in CSV (Comma Separated Values) format but in TSV (Table Separated Values) format. MultiGwas only supports CSV for phenotypes.
2. Your phenotypes have a group column that needs to be removed. MultiGwas phenotypes contain only trait ID and phenotype columns.
3. There were problems running the Gapit functions from MultiGwas, now they are running but I need to check further or update it to Gapit3.

Other issues that were not critical but they helped me to improve the code were the following:
1. Your genotype is very large, so I implemented a cache system for converting the genotype to the tools format only once and not for each phenotype.

Other questions that were not critical but that helped me improve the code were the following:
1. Your genotype is very large, so I implemented a caching system to convert the genotype to tools format only once and not for every phenotype.
2. I am parallelizing the execution by tool, but I think it is better to parallelize by trait. Not yet implemented.

Finally, thanks for using MultiGwas.

Sincerely,
Luis Garreta

[Quoted text hidden]

multigwas-results-additive-naive.zip
19968K

RN Sarma <ramendra.sarma@aau.ac.in> Tue, Apr 19, 2022 at 8:54 AM
To: Luis Garreta <lgarreta@gmail.com>

Dear Luis,
Thank you for your email. I'm thankful to you in pointing out of the error in phenotype data. I need to run big files of genotypic data. Do I need to reinstall the cache system by updating MultiGwas? I'm waiting for your email about the updates done by you and the way to use it. How to parallelize multicast?

On Tue, Apr 19, 2022, 7:09 PM Luis Garreta <lgarreta@gmail.com> wrote:

> Hi Mr. Sarma:
> I'm sending you some MultiGwas results from your data, a GWAS analysis for additive model, naive approach, without filters.
>
> In the following days I will write to you about the MultiGwas changes and how to update them. Also, I will copy all the results to a repository for download. What codes to be run for parallelization?
> [Quoted text hidden]
> [Quoted text hidden]

RN Sarma <ramendra.sarma@aau.ac.in> Fri, May 6, 2022 at 4:23 AM
To: Luis Garreta <lgarreta@gmail.com>

Hi Luis,
Have you updated the MultiGwas on Github, since I want to use it again??

I am looking for a reply from you.
[Quoted text hidden]

Luis Garreta <lgarreta@gmail.com> Fri, May 6, 2022 at 12:45 PM
To: RN Sarma <ramendra.sarma@aau.ac.in>

Hi RN,
The multigwas github has been updated with minor modifications. I also send you some test that I have run with part of your genotype and phenotype. These tests are one for each tool (four in total) and one for all tools. All of them finished successfully.

Now, multigwas results include also the scores for each trait (e.g. DM-scores.csv) and the scores for all markers (SCORES-ALL.csv). The scores for the best markers was renamed to SCORES-BEST.csv.

The SCORES-ALL.csv file includes a column called GSSCORE used to select the best scores. It is a own score that takes into account for each SNP its Genomic Control Factor (GC), its significance, and if it was obtained for more than one tool.

I attach the test configuration files and their full results:

Any problem, do not hesitate to write to me.

Luis Garreta

[Quoted text hidden]

10 attachments

GWASPOLY-test.config
1K

SHESIS-test.config
1K

GAPIT-test.config
1K

ALLTOOLS-test.config
1K

TASSEL-test.config
1K

out-GWASPOLY-test.zip
1130K

out-TASSEL-test.zip
1228K

out-SHESIS-test.zip
1147K

out-GAPIT-test.zip
1167K

out-ALLTOOLS-test.zip
4126K

RN Sarma <ramendra.sarma@aau.ac.in> Sat, May 7, 2022 at 12:38 AM
To: Luis Garreta <lgarreta@gmail.com>

Dear Luis,
Thank you for your tips. When I checked the Github (https://github.com/agrosavia-bioinformatics/MultiGWAS), I found that all files and directories are old and not reflected any modification that you have done. Should I have to re-install multiGWAS on my Linux system again with new updates and if so how? Is it ready to use for large data files than you you shown with test data?
What is GC in column D and Score in column I of SCORES-BEST.csv file?
Can you please explain GSCORE a bit more with some reference?

Regards,

Ramen

[Quoted text hidden]

Luis Garreta <lgarreta@gmail.com> Sat, May 7, 2022 at 12:05 PM
To: RN Sarma <ramendra.sarma@aau.ac.in>

Hi,
You are using a wrogn github address. The multigwas github is:
https://github.com/agrosavia-bioinfo/multiGWAS

If you downloaded multiGWAS from this repo, I think you should just have to do a git pull. Another option is to replace the R source files in the $MULTIGWAS_HOME/main with these ones that I'm attaching to this mail. An the last option is to download again the new 1.3 version and run the installer, but if the libraries were previously installed, the installer just installs the sources and set the environment variable.

Now, the column GC is for Genomic Control factor (https://en.wikipedia.org/wiki/Genomic_control) and it shows how inflated could be the p_values or the scores (log10 de p_value). Literature shows that GC must be closer to one, you can see it in the manhattan plots.

As multiGWAS integrates four gwas tools, each tool produces a set of SNPs with their own p_value or score (log10 of pvalue). For one tool one SNPs has low p_value (high score) but for other this pvalue can be different. The problem is how you can determinate what SNPs are really important, and for this reason I defines the GSCORE. The GSSCORE is an own score (no literature) that tries to rank SNPs importance according to three elements: the GC score (80%), Significance (10%) (if the SNPs was significant or not), and if the SNPs is reported by more than one tool (10%).

In the paper of multiGWAS (https://onlinelibrary.wiley.com/doi/full/10.1002/ece3.7572) there is a section (2.3 Integration Stage, 2.3.1 Selection of best gene action model) of a first GSCORE version, very close to this new version.

Any problem, do not hesitate to write to me.
[Quoted text hidden]

upd-sources.zip
175K

RN Sarma <ramendra.sarma@aau.ac.in> Sat, May 7, 2022 at 8:01 PM
To: Luis Garreta <lgarreta@gmail.com>

Dear Luis,
Thank you for your explanation. I shall update my system as per your suggestion.
Genomic control (GC) is to control population stratification, then is it redundant with PCA in Tassel?
Where from Q matrix is available to Tassel in Multigwas, since standalone Tassel requires to feed with a Q matrix fil?

Ramen
[Quoted text hidden]

RN Sarma <ramendra.sarma@aau.ac.in> Wed, May 11, 2022 at 2:13 AM
To: Luis Garreta <lgarreta@gmail.com>

Hi Luis,
I have a problem as I was using a wrong github address as your mail, I removed the old directory ad I did as follows:
rnsarma@rns-pbg-ws:~$ git clone https://github.com/agrosavia-bioinfo/multiGWAS.git ## It was successful.
rnsarma@rns-pbg-ws:~/multiGWAS/install$ sh install-linux-packages.sh ## It was successful.
rnsarma@rns-pbg-ws:~/multiGWAS/install$ Rscript install-R-libraries.R ## It was successful with a message that "Close the terminal to finish the installation process
Then, open a new terminal and write: multiGWAS
In a new terminal, I did as follows;
rnsarma@rns-pbg-ws:~$ source ~/.bashrc
rnsarma@rns-pbg-ws:~$
This terminal was closed and the new terminal was opened. typed as follows:
rnsarma@rns-pbg-ws:~$ jmultigwas
jmultigwas: command not found
rnsarma@rns-pbg-ws:~$

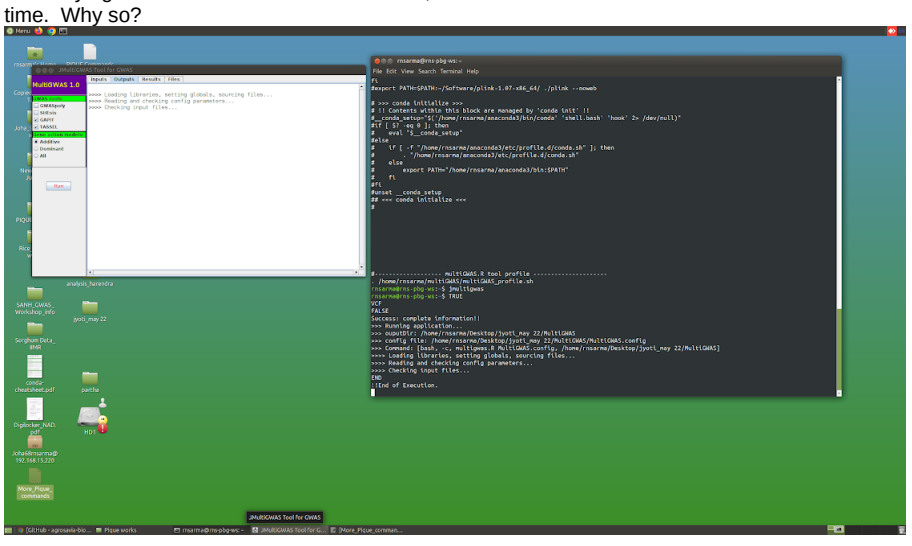
Why is itis behaving so? Please suggest.

Ramen

[Quoted text hidden]

RN Sarma <ramendra.sarma@aau.ac.in> Wed, May 11, 2022 at 6:19 AM
To: Luis Garreta <lgarreta@gmail.com>

While trying MULTIGWAS in another machine, it is like this as shown in the screenshot in which the command prompt says "end of execution" while the java screen says "checking the input files for a long time. Why so?



[Quoted text hidden]

Luis Garreta <lgarreta@gmail.com> Mon, May 16, 2022 at 1:22 PM
To: RN Sarma <ramendra.sarma@aau.ac.in>

Hi RN,
I updated the multiGWAS github: https://github.com/agrosavia-bioinfo/multiGWAS.

Take care of the input parameters in the configuration file or in the GUI interface, mainly the ploidy (e.g. 2, in your case) and the genotype format (e.g. "VCF", in your case). I ran some tests with your data, and I found that your genotype has problems with the TASSEL tool, but it's OK with the others (Gwaspoly, Shesis, Gapit). I can't discover the problem with TASSEL and your genotype. I need to perform more tests.

I ran a test with the full genotype and the first five phenotypes and It ran OK, although it took a long time as your genotype has more than 100k markers.

You can run a test with few data to check if multiGWAS runs without problems, I attach you a test with 10000 genotypes and two phenotypes, and the configuration file to do the test. Open a terminal, change to "test-10kGenos-2Phenos" directory, and write the command:
$ multigwas test-10kG-2P.config
At the end, results are saved into the "out-test-10kG-2P" directory. The report for each trait is in its corresponding html file. Also, there are the scores file, and outputs (pdfs, pngs, tables) are in the "report" dir for each trait.

Remember that your phenotype matrix has a group column that must be removed.

I attach the results of the test and if you have any problem, do not hesitate to write to me.

[Quoted text hidden]

test-fullGeno-5Phenos.zip
5921K

Luis Garreta <lgarreta@gmail.com> Mon, May 16, 2022 at 1:24 PM
To: RN Sarma <ramendra.sarma@aau.ac.in>

Sorry, I forgot the test files.
[Quoted text hidden]

2 attachments

test-fullGeno-5Phenos.zip
5921K

test-10kGenos-2Phenos.zip
3707K

RN Sarma <ramendra.sarma@aau.ac.in> Mon, May 16, 2022 at 8:53 PM
To: Luis Garreta <lgarreta@gmail.com>

[Quoted text hidden]

RN Sarma <ramendra.sarma@aau.ac.in> Mon, Aug 1, 2022 at 4:49 AM
To: Luis Garreta <lgarreta@gmail.com>

Hi Luis,
Greetings from Assam Agricultural University!

I remove the old version of multiGWAS by "rm -r multiGWASpre". Then I followed your information as given below without installing R (as I have the R package already):

```
1 Download or clone the MultiGWAS repository
    git clone https://github.com/agrosavia-bioinfo/multiGWASpre.git
2. Change to install directory:
    cd install
3. Run the bash script to install the necessary linux packages (it needs sudo privileges).
    sh install-linux-packages.sh
4. Run the R script to install MultiGWAS
    Rscript INSTALL.R


But when I tried to execute MultiGWAS GUI with jmultigwas or MultiGWAS CLI with multigwas in
the new terminal. the response is as follows


jmultigwas

jmultigwas: command not found


What to do now?
```

On Tue, Apr 19, 2022 at 7:09 PM Luis Garreta <lgarreta@gmail.com> wrote:
[Quoted text hidden]