



Malin Pinsky <malin.pinsky@gmail.com>

---

## bash lines to get a "SNP join"

2 messages

---

**Bastiaan Star** <bastiaan.star@ibv.uio.no>  
To: Malin Pinsky <malin.pinsky@rutgers.edu>

Wed, Oct 12, 2016 at 11:58 AM

Hi Malin,

Assuming you do this per LG and you have 2 column files with LG in the first, and position in the other:

```
cat <files_of_interest> |awk '{print $2}' |sort -n |uniq -c |awk '$1==<X> {print  
""$1""""\t"$2}"' > Position_file_of_SNPs_found_in_each_subset.txt
```

<X> being the number of files that you supply.

This essentially allows you to take the join of an unlimited set of files. You may have to sort the final dataset on column two again. NOTE that this: ""\$1"""" allows you to introduce a variable into awk from the bash environment. This line come from a script whereby \$1 is LG01, LG02 etc are provided in a loop.

Let me know if you get stuck, but I'm sure you'll manage.

Cheers,

Bastiaan

---

**Malin Pinsky** <malin.pinsky@rutgers.edu>  
To: Bastiaan Star <bastiaan.star@ibv.uio.no>

Wed, Oct 12, 2016 at 11:26 PM

Cool. Thank you!

[Quoted text hidden]

--

Malin Pinsky  
Assistant Professor  
Department of Ecology, Evolution, and Natural Resources  
and Institute of Earth, Ocean, and Atmospheric Sciences  
14 College Farm Rd.  
Rutgers University  
(848) 932-8242  
[malin.pinsky@rutgers.edu](mailto:malin.pinsky@rutgers.edu)  
<http://pinsky.marine.rutgers.edu>