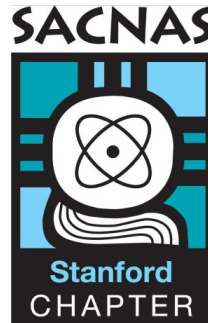


Remembering what you did: Automated reports with R Markdown

Sur Herrera Paredes
2022-04-23

Data management for publishing
Jasper Ridge Biological Preserve



A (hopefully not) familiar organizational structure



analysis



analysis2



analysis_final



analysis_final2



analysis_v2



combined_analysis



combined_data



data



data_old



full_data



full_data_new

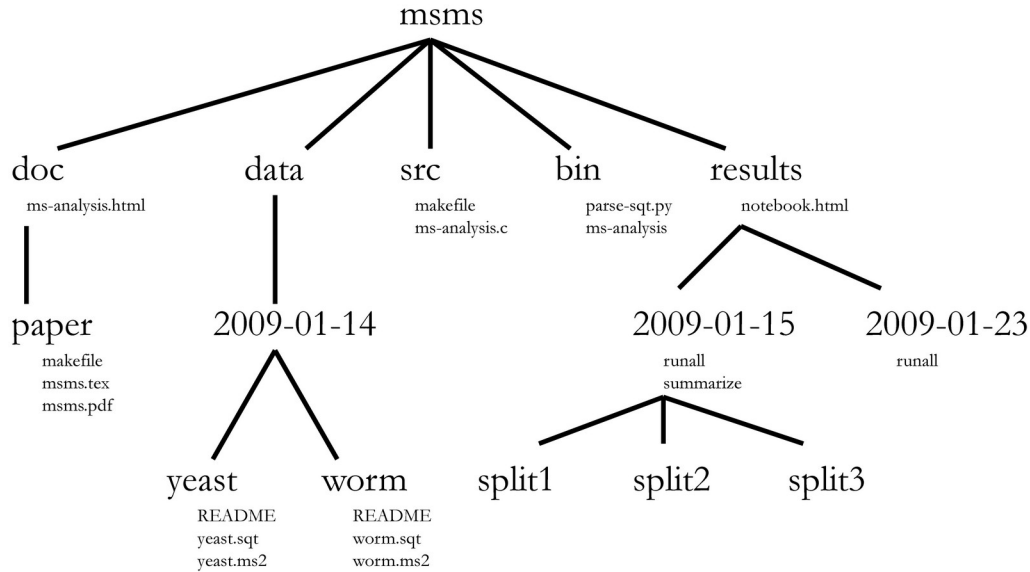


method_benchmark



new_data

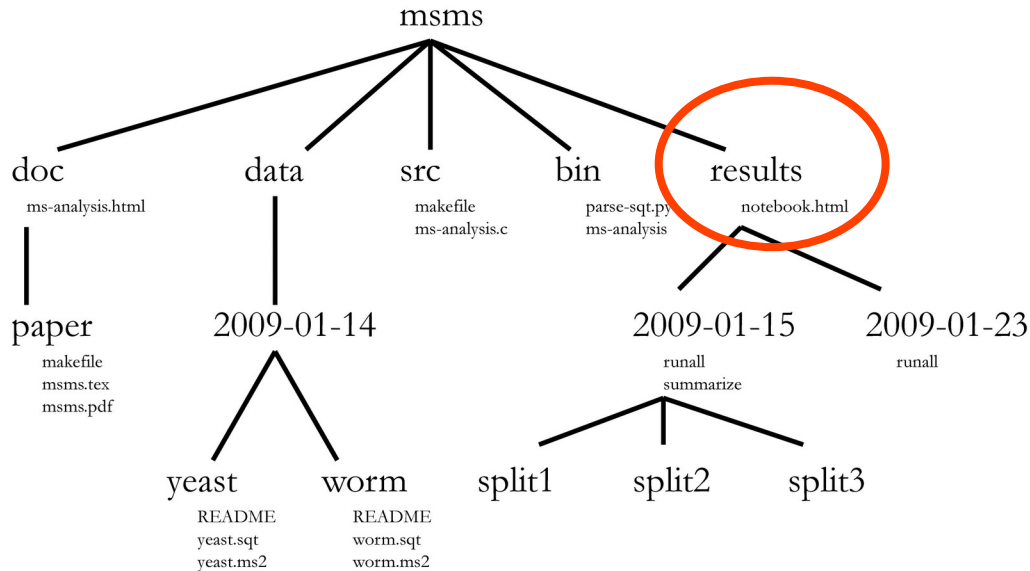
Thinking about organization: types of files



- Keep different types of digital objects (e.g. data, code, executables, analysis) separate.
- Establish rules for different types of digital objects (i.e. files)
- Be consistent!

Noble, «A quick guide to organizing computational biology projects».

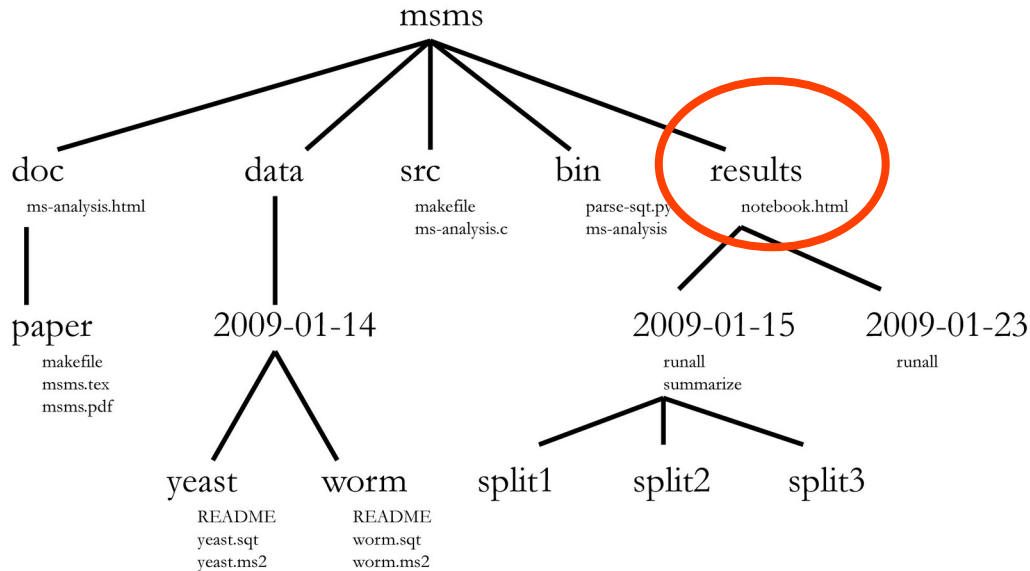
Thinking about organization: types of files



- Keep different types of digital objects (e.g. data, code, executables, analysis) separate.
- Establish rules for different types of digital objects (i.e. files)
- Be consistent!

Noble, «A quick guide to organizing computational biology projects».

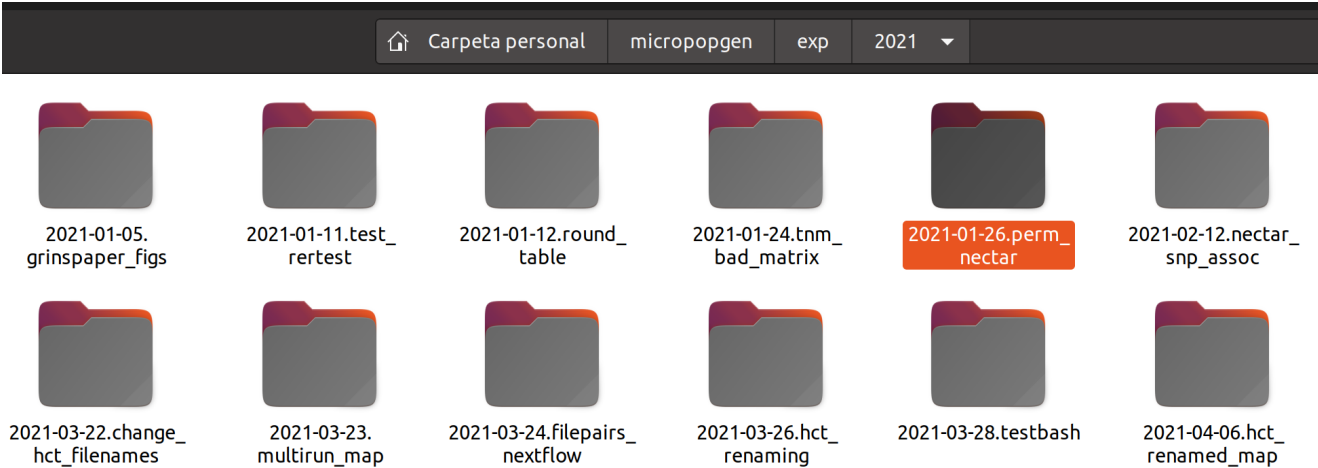
Organize your analysis chronologically



- You never know where your results will take you...yay Science!
- Organizing your results chronologically ensures that you can keep the organization consistent

Noble, «A quick guide to organizing computational biology projects».

Create a digital lab notebook: chronological order



Don't forget to add a small informational name to the directory names

Create a digital lab notebook: A self-contained record of every analysis



all.genmap.max.
miss.1.recode.vcf.
012



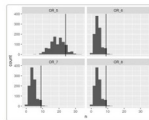
all.genmap.max.
miss.1.recode.vcf.
012.indv



all.genmap.max.
miss.1.recode.vcf.
012.pos



all.vcf



an_neutral_enrich_
perm.jpeg



assocs_AN_lph.tsv



assocs_AN_neutral.
tsv



assocs_lph_neutral.
tsv



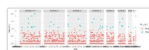
enrich_AN_lph.tsv



enrich_AN_neutral.
tsv



enrich_lph_neutral.
tsv



manhatt.jpeg



OR_analysis.png



OR_analysis.svg



README.md

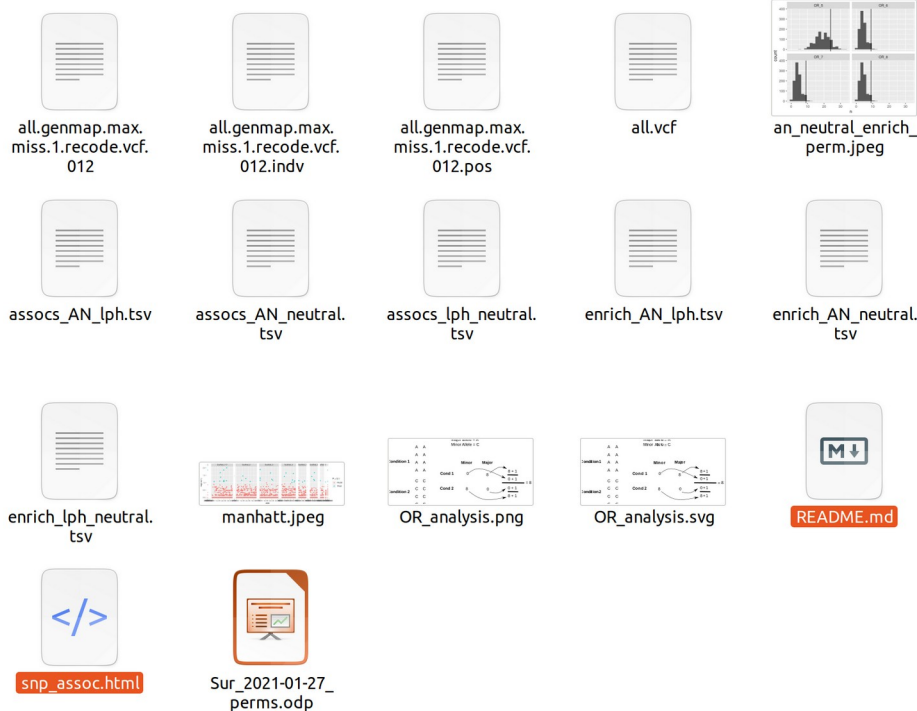


snp_assoc.html



Sur_2021-01-27_
perms.odp

Create a digital lab notebook: save the code and output together



```
an_neutral <- test_condition_pair(selected_conditions = c("neutral", "AN"),
                                  geno = geno, meta = meta, Perms = Perms)
write_tsv(lph_neutral$assocs, "assocs_AN_neutral.tsv")
write_tsv(lph_neutral$enrich, "enrich_AN_neutral.tsv")
an_neutral
```

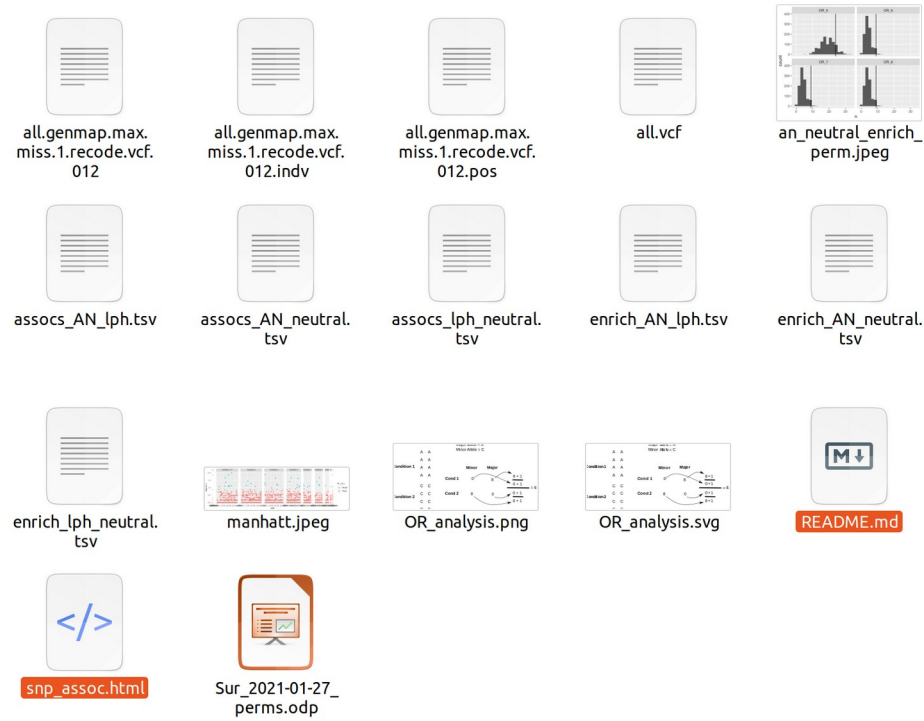
```
## $assocs
## # A tibble: 26,317 x 7
##   snp      major_c1 minor_c1 major_c2 minor_c2 OR.lap      P
##   <chr>      <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 X11989          8         0         4         4 0.00999
## 2 X19677          0         8         4         4 0.111 0.0130
## 3 X10423          0         8         4         4 0.0140
## 4 X13824          8         0         4         4 0.0140
## 5 X19445          8         0         4         4 0.0140
## 6 X1627           0         8         4         4 0.111 0.0210
## 7 X7805           0         8         4         4 0.111 0.0210
## 8 X22988          8         0         5         3 0.0220
## 9 X2992           8         0         5         3 0.0220
## 10 X3263          3         5         0         8 0.0240
## # ... with 26,307 more rows
##
## $enrich
## # A tibble: 8 x 2
##   OR.thres pval
##   <int> <dbl>
## 1     1 0.23
## 2     2 0.212
## 3     3 0.673
## 4     4 0.106
## 5     5 0.106
## 6     6 0.007
## 7     7 0.007
## 8     8 0.007
```

```
an_lph <- test_condition_pair(selected_conditions = c("AN", "lph"),
                              geno = geno, meta = meta, Perms = Perms)
write_tsv(lph_neutral$assocs, "assocs_AN_lph.tsv")
write_tsv(lph_neutral$enrich, "enrich_AN_lph.tsv")
an_lph
```

```
## $assocs
## # A tibble: 26,317 x 7
##   snp      major_c1 minor_c1 major_c2 minor_c2 OR.lap      P
##   <chr>      <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 X14333          6         2         0         8 21. 0.00300
## 2 X14565          4         4         8         0 0.111 0.00899
## 3 X22988          4         4         8         0 0.111 0.0110
## 4 X2992          4         4         8         0 0.111 0.0110
## 5 X3282          4         4         8         0 0.111 0.0110
## 6 X3219          4         4         8         0 0.111 0.0130
## 7 X13824          5         3         8         0 0.167 0.0250
## 8 X16220          8         0         5         3 0.0250
## 9 X19445          5         3         8         0 0.167 0.0250
## 10 X13756          8         0         5         3 0.0350
## # ... with 26,307 more rows
##
```

That's where that table came from

Create a digital lab notebook: the README is your notebook



In the wet lab nothing is more important than having a lab notebook. The same is true in the dry lab

Create a digital lab notebook: a simple approach

- Organize analysis chronologically
- Add a simple tag to each analysis directory
- Save the code together with the output that it produced.
- Keep a README inside every analysis
- Recreate don't overwrite

surh / Rmarkdown_workshop (Public)

<> Code Issues Pull requests Actions Projects Wiki Security Insights

master 2 branches 0 tags Go to file Code

surh	clean repetitive info	
data	New dataset.	
figs	Reorganizingin files in main dir	
misc	Exctracting dataset	
.gitignore	Momentarily ignoreing html output	
LICENSE	adding params, updating license	22 days ago
README.md	clean repetitive info	22 days ago
Resources.md	adding RMarkdown for Scientists	22 days ago
learning.md	Adding format & basic text	22 days ago
r_commands.r	SAving code	22 days ago
references.bib	Reorganizingin files in main dir	23 days ago
report.Rmd	adding params, updating license	22 days ago
report.r	SAving code	22 days ago

Clone

HTTPS GitHub CLI

https://github.com/surh/Rmarkdown_work

Use Git or checkout with SVN using the web URL.

Download ZIP

https://github.com/surh/Rmarkdown_workshop

