Reads2Map

Developing best practices for genotyping-by-sequencing analysis using linkage maps as benchmarks

Cristiane Taniguti

chtaniguti@tamu.edu

TEXAS A&M UNIVERSITY

STATISTICAL GENETICS LAB
ESALQ / USP

NC STATE UNIVERSITY

TOOLS FOR POLYPLOIDS

This project is funded by USDA NIFA Specialty Crop Research Initiative Award # 2020-51181-32156 (09/01/20 - 08/31/24)

Motivation: GBS data

STATISTICAL GENETICS ·LAB·
ESALQ / USP

Issues while building linkage maps

- Computational intensive
- Time consuming
- Wrong grouping
- Wrong ordering
- Inflated linkage maps

OneMap

- Since 2007 - 52k downloads
- Diploid species
- Bi-parental populations
- Backcross, RILs, F2 and outcrossing
- Biallelic and Multiallelic markers

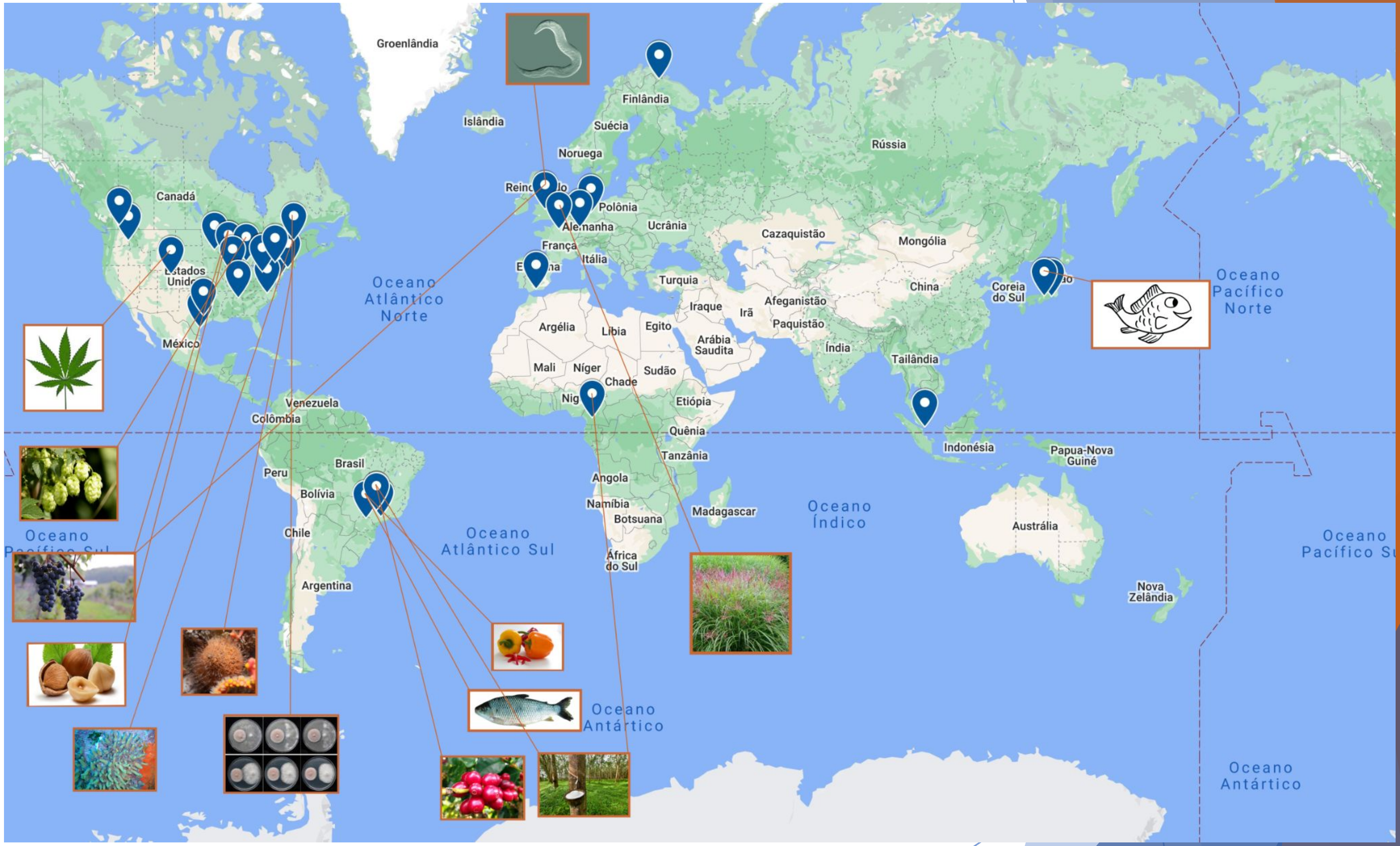
Maintainer since 2017
Updates in version 3.0

Augusto Garcia
Marcelo Mollinari
Gabriel Margarido

Motivation

Users feedbacks

OneMap



TOOLS FOR POLYPLOIDS

Motivation: GBS data

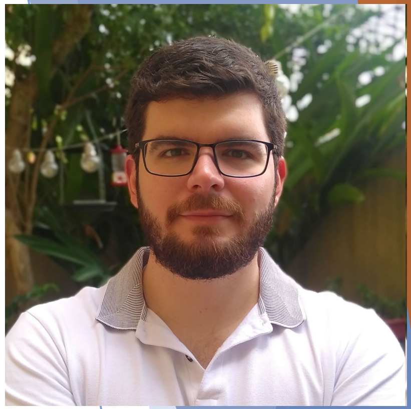
Issues while building linkage maps

- Computational intensive
- Time consuming
- Wrong grouping
- Wrong ordering
- Inflated linkage maps

Tips and Tricks:

RGC11 - Poster

Augusto Garcia



Gabriel Gesteira

MAPpoly

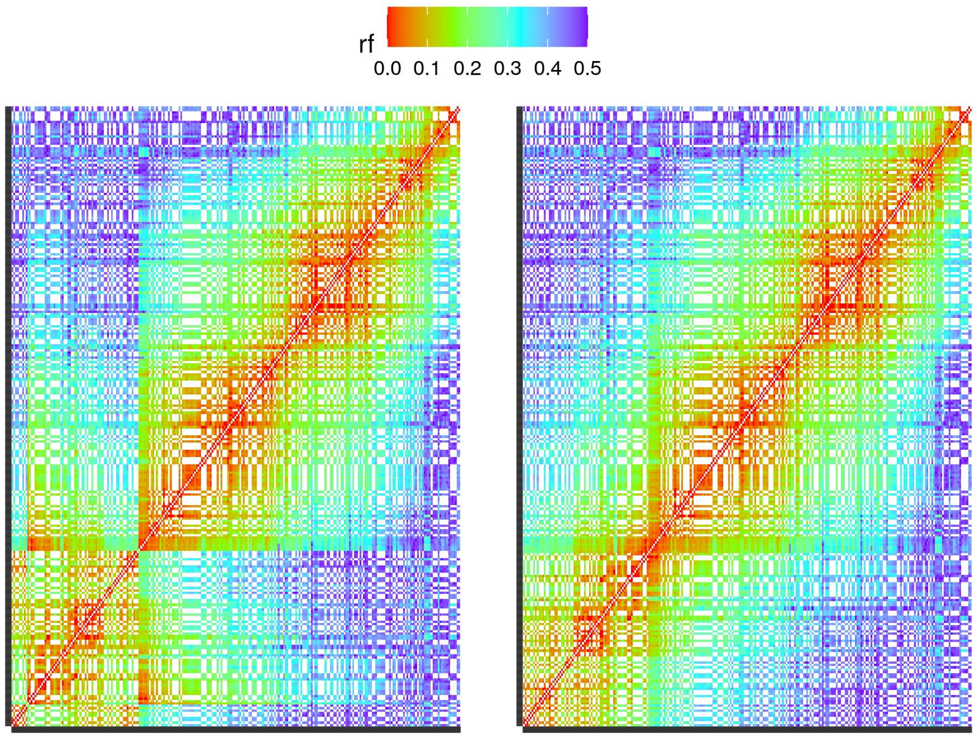
- Since 2018
- Diploid and polyploid species
- Bi-parental populations
- Outcrossing
- All dosages markers
- Updates by Marcelo and Gabriel

Marcelo Mollinari

Recombination fraction matrix as benchmarks

Inversion

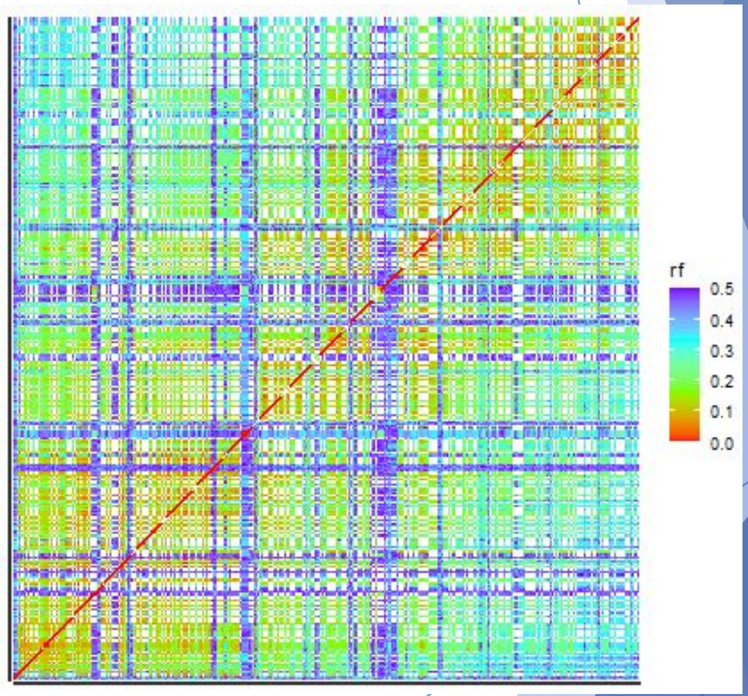
- Aspen chromosome 12



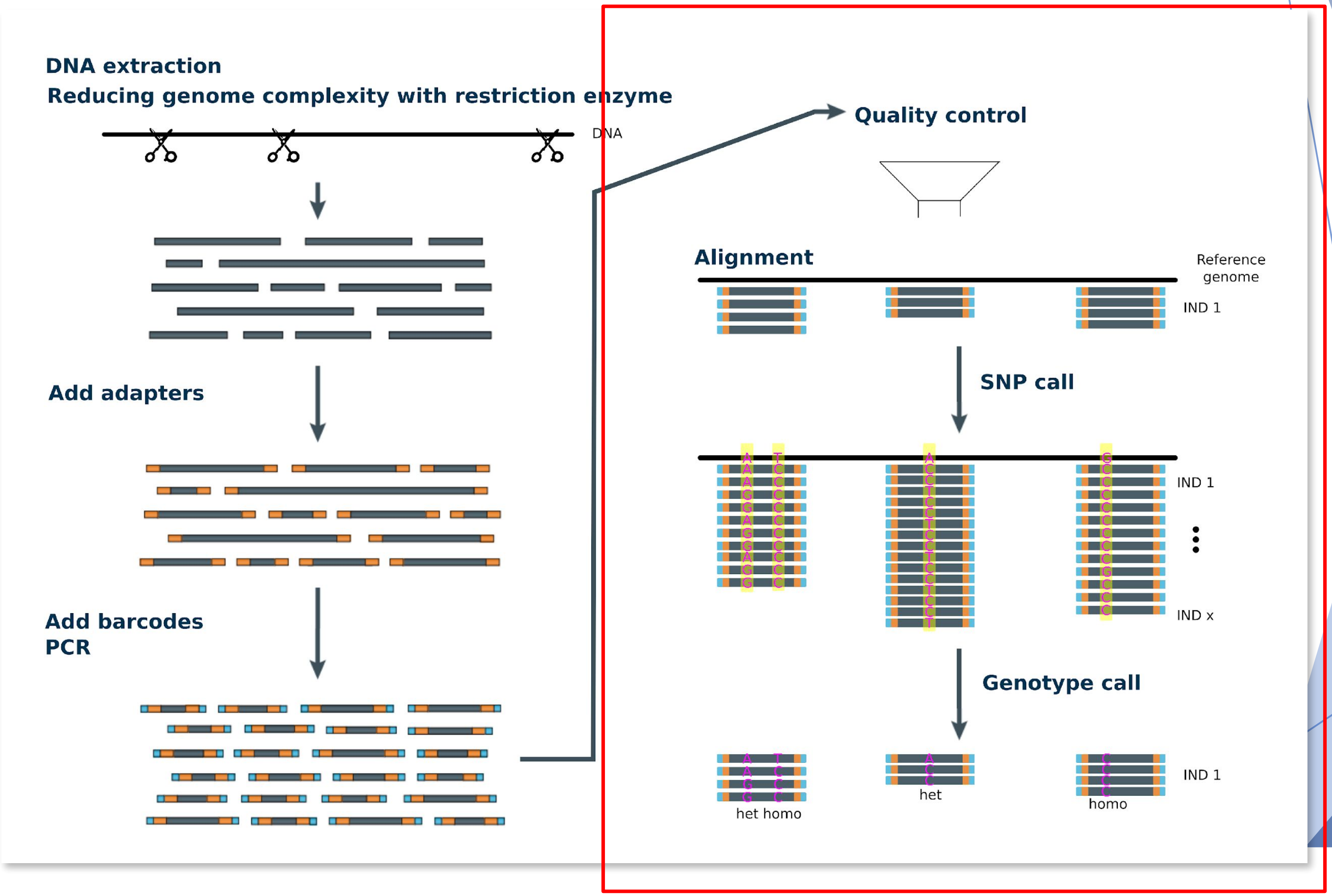
Genomic order

Order fixed

Most likely bad data



GBS Overview



DNA extraction

Reducing genome complexity with restriction enzyme

Add adapters

Add barcodes
PCR

Quality control

Alignment

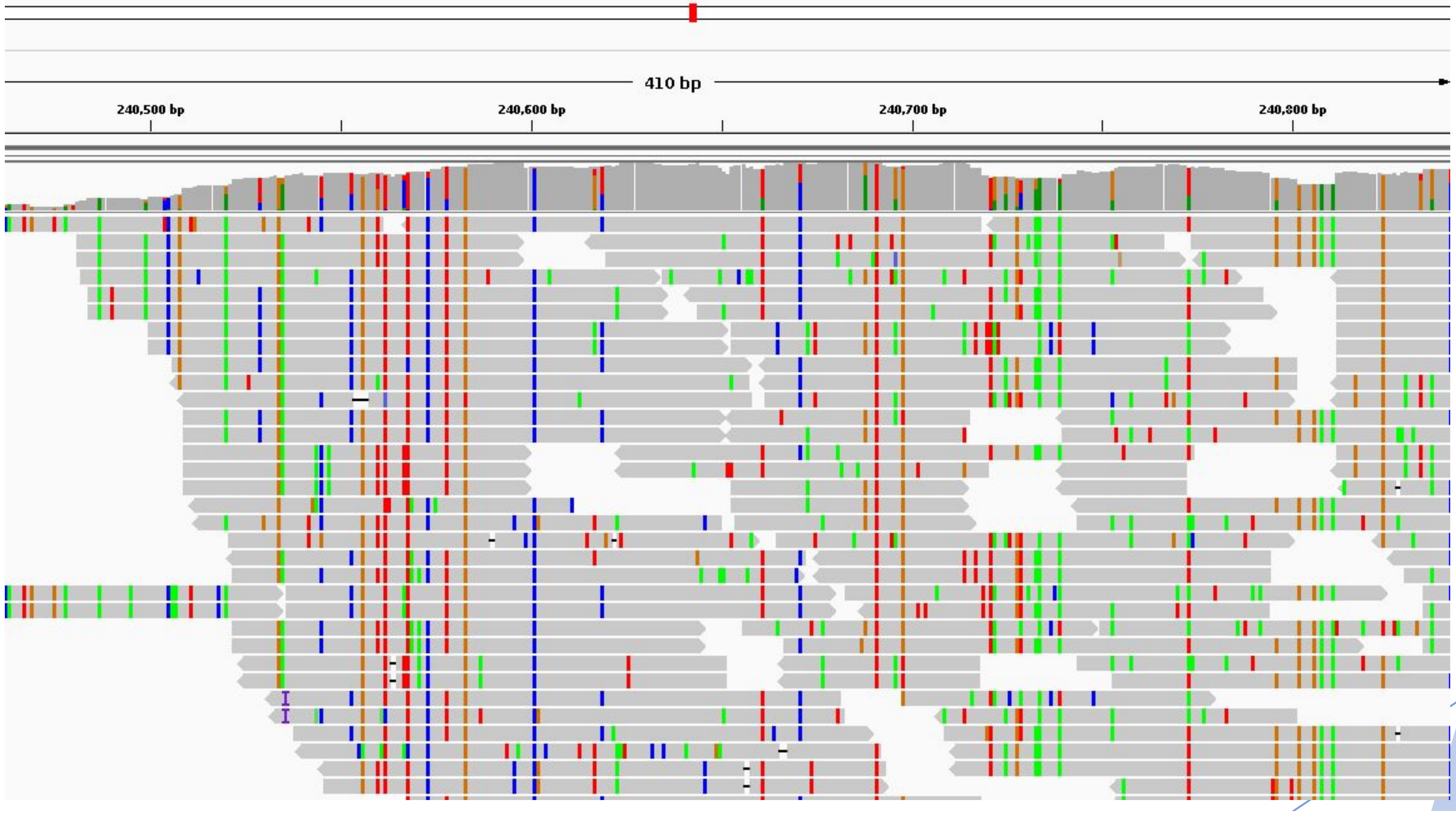
SNP call

Genotype call

SNP Calling

- Whole Genome Sequencing (WGS)

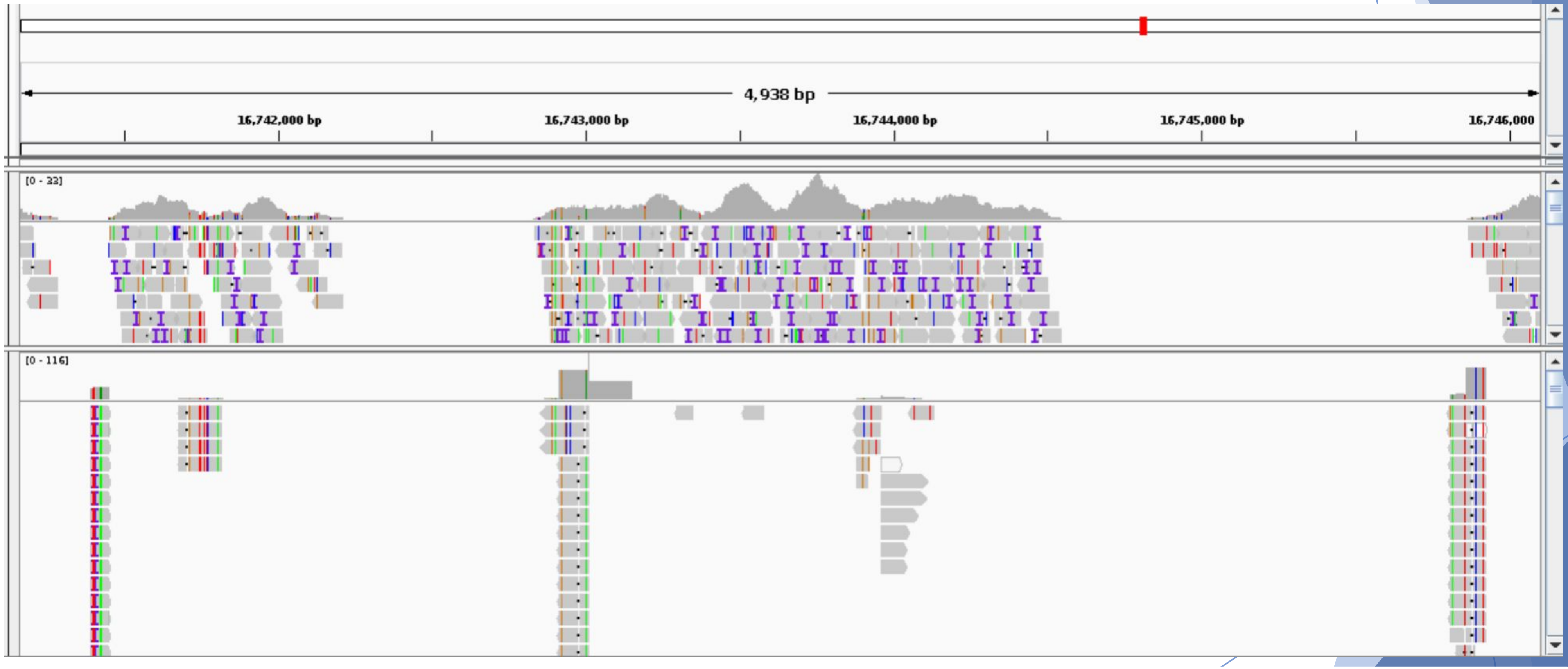
Image: IGV



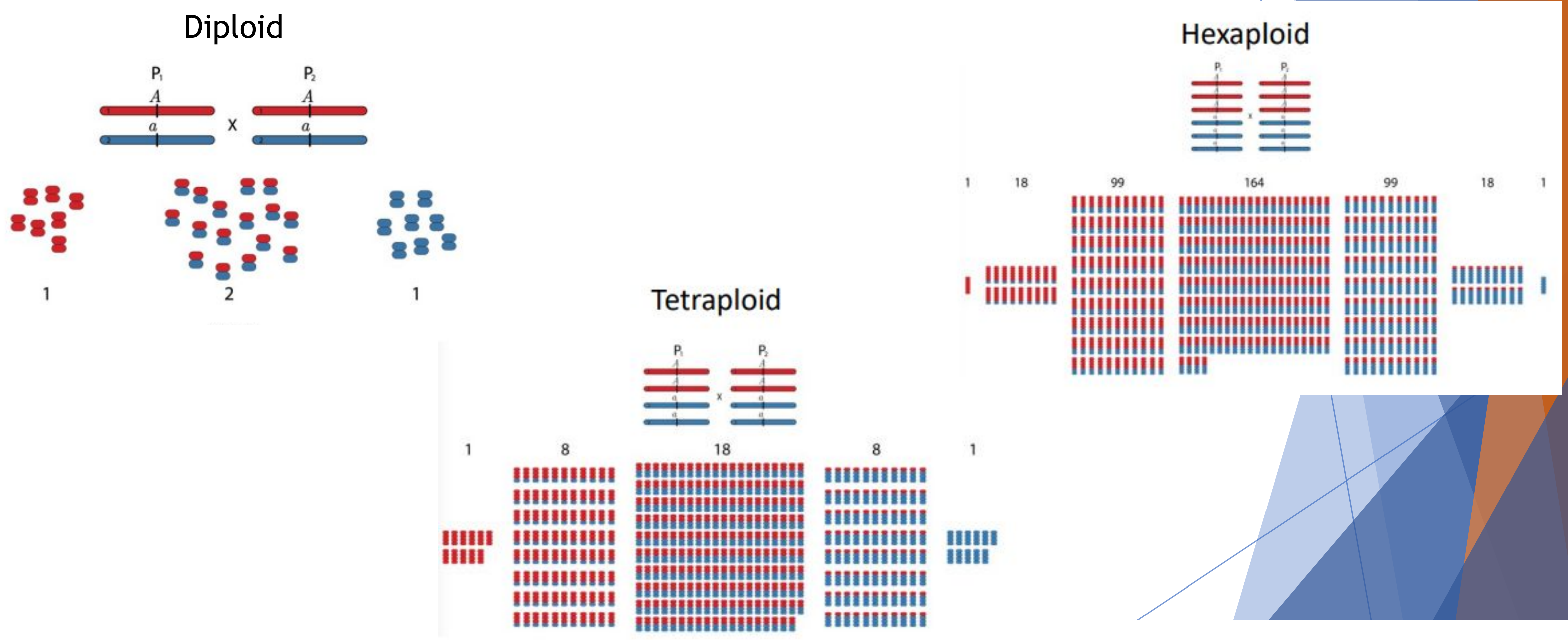
SNP Calling

- Exome sequencing (top) and Genotyping-by-Sequencing (bottom)

Image: IGV



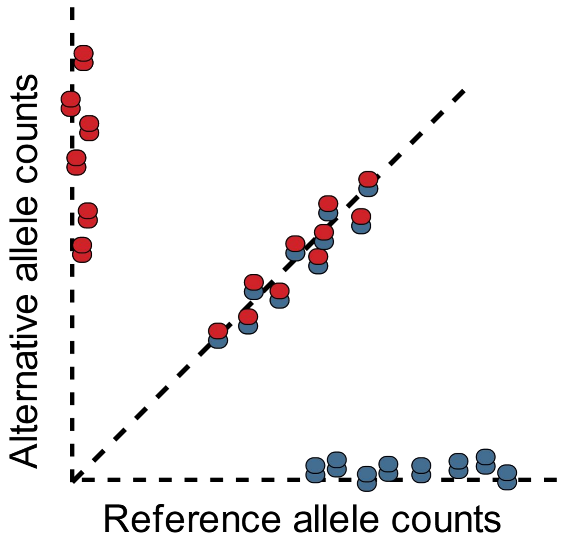
Dosage calling



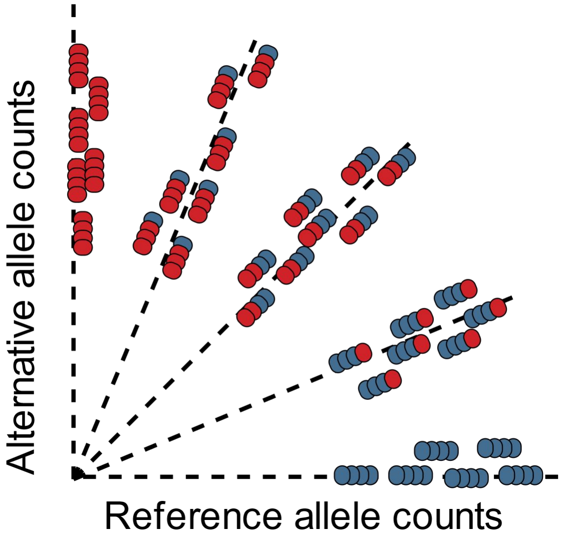
Dosage Calling

- The theory

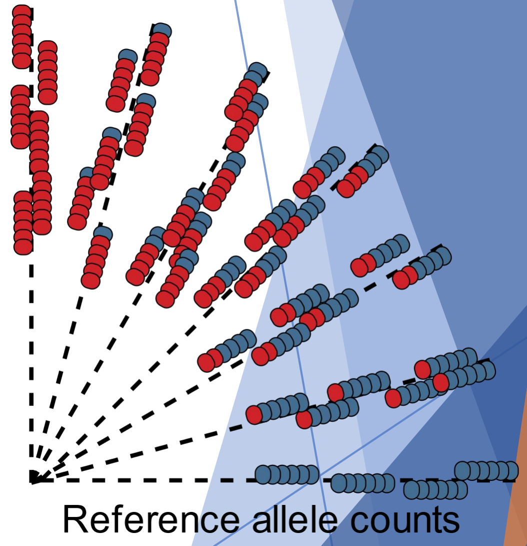
Diploid



Tetraploid



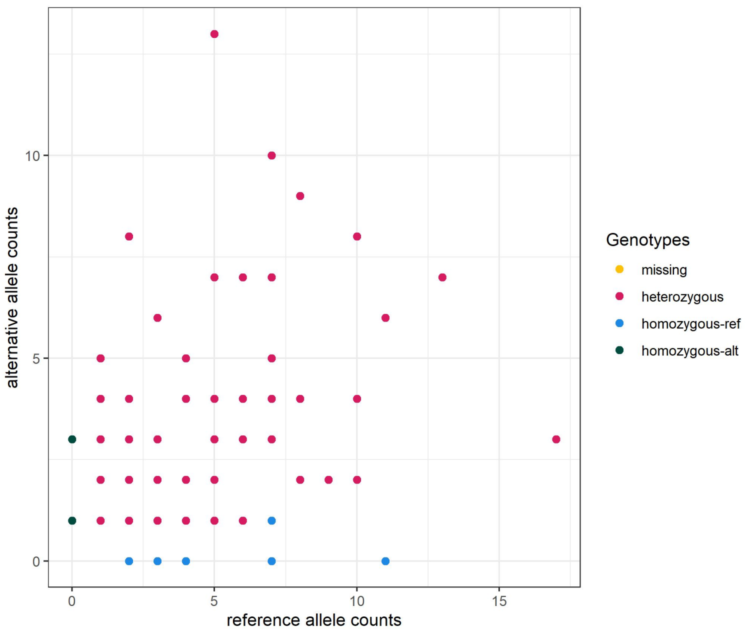
Hexaploid



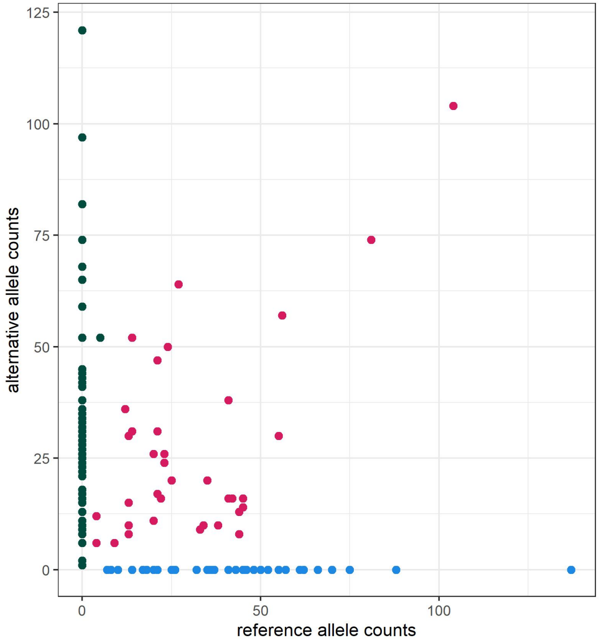
Dosage Calling

- The reality

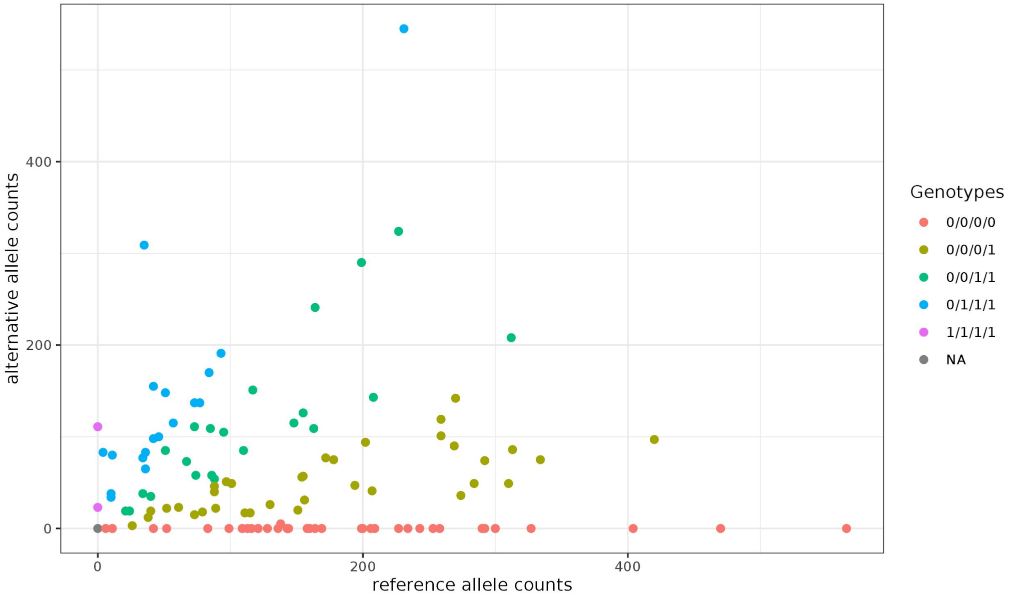
Diploid (mean depth 6)
N = 200
Aa x Aa



Diploid (mean depth 96)
N = 138
Aa x Aa



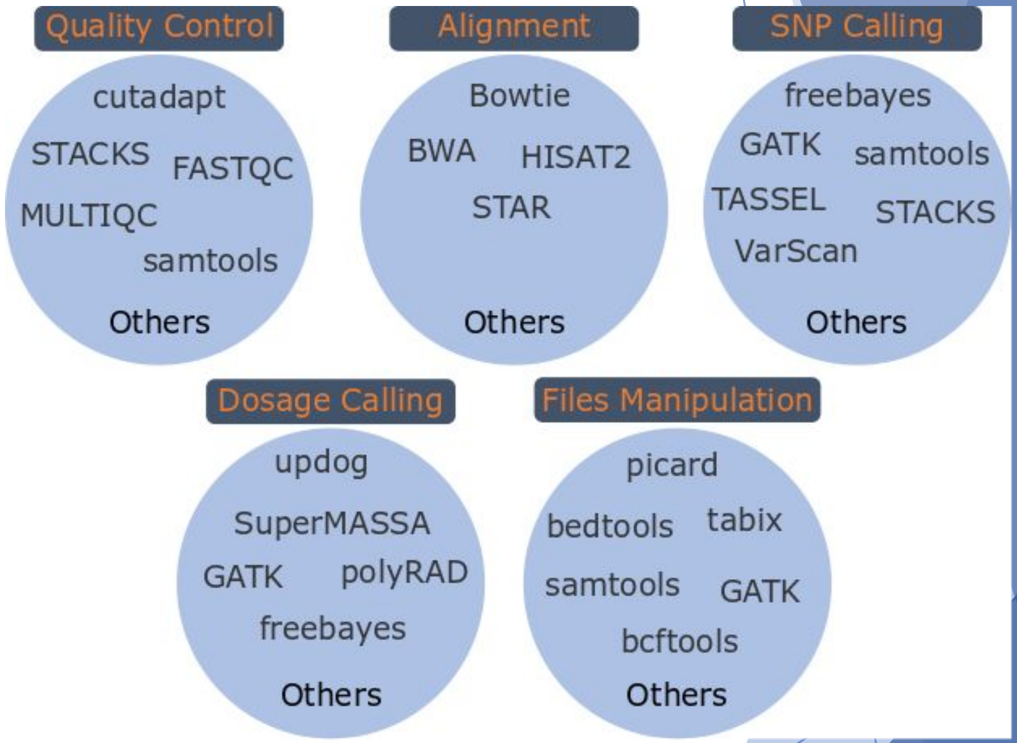
Tetraploid (mean depth 83)
N = 114
AAaa x AAaa



Sequencing Data - Technical Difficulties

- Large files

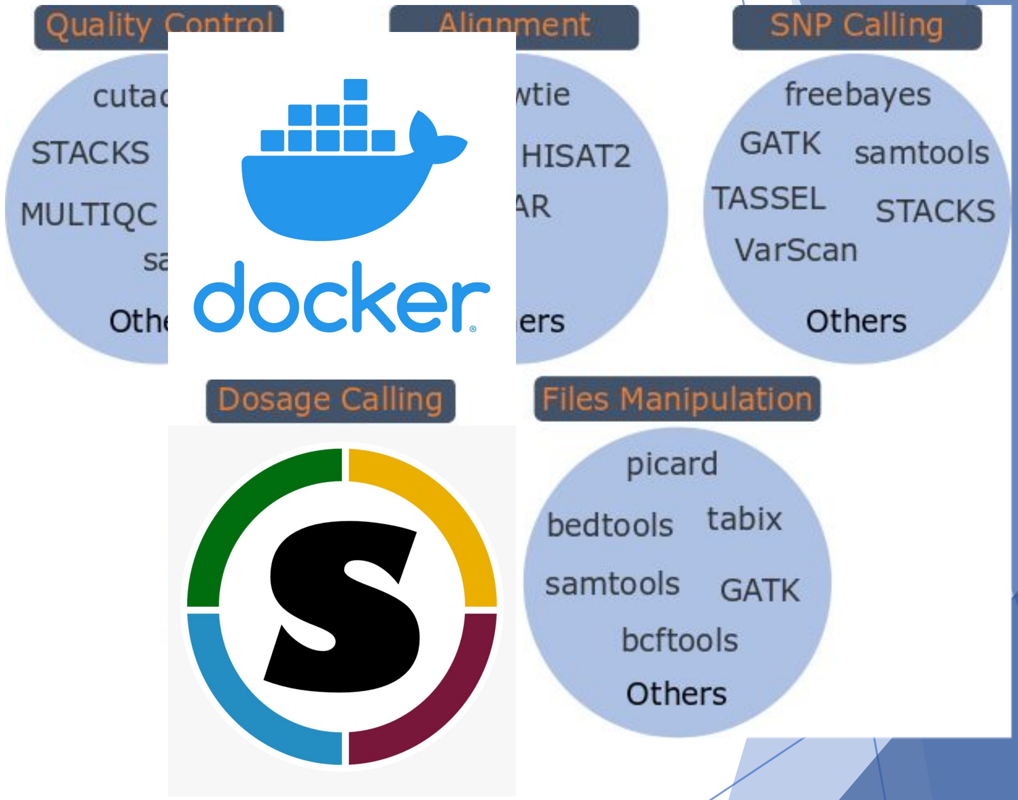
- Many software
- Many programming languages
- Different Operational Systems
- Updates



Sequencing Data - Technical Difficulties

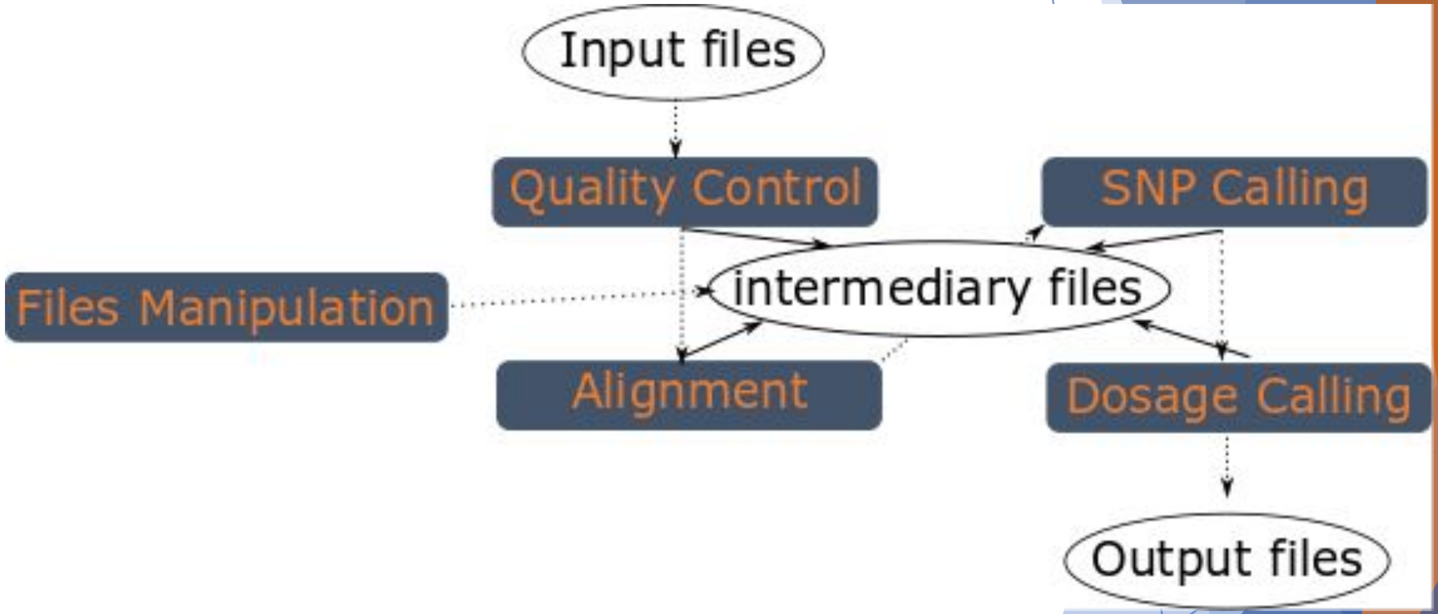
- Large files
  - High Performance Computing (HPC)
  - Management systems (SLURM, SGE)
  - Cloud (Google, Amazon)

- Many software
- Many programming languages
- Different Operational Systems
- Updates
  - Containers
    - Docker
    - Singularity (usually available in HPC)
    - BioContainers



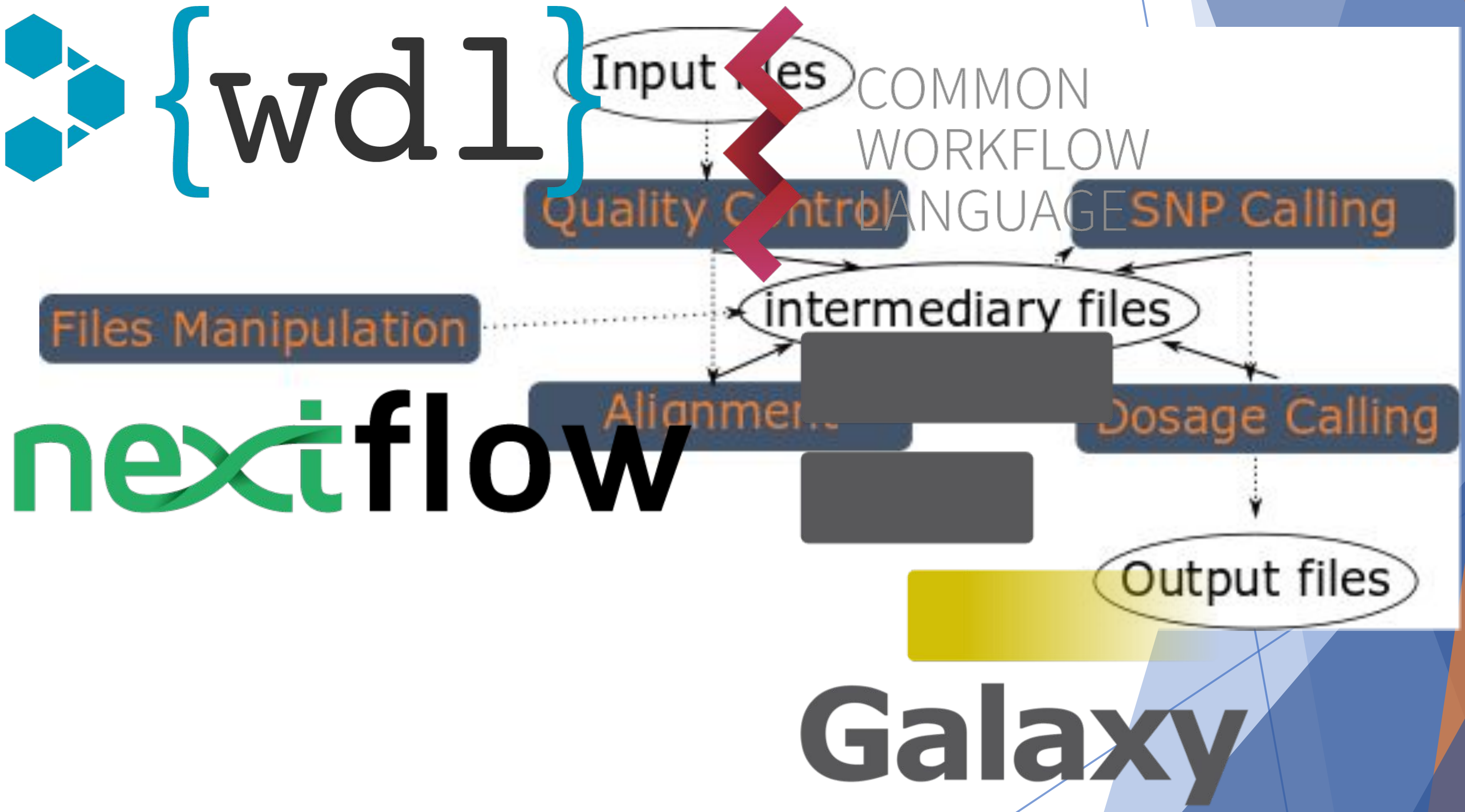
Sequencing Data - Technical Difficulties

- Many steps
- Many file formats



Sequencing Data - Technical Difficulties

- Many steps
- Many file formats
  - Workflows systems
    - Galaxy
    - Nextflow
    - Snakemake
    - CWL
    - WDL
  - Workflows repositories
    - [Dockerstore]()
    - [WorkflowHub]()
  - Run workflows on Cloud
    - Galaxy
    - DNAnexus
    - Terra
    - AnVIL
    - SevenBridges

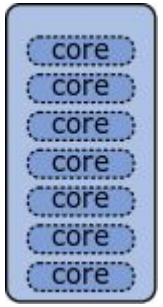


Sequencing Data - Technical Difficulties

- Resources optimization
  - Time
  - Cores
  - Nodes
  - RAM memory

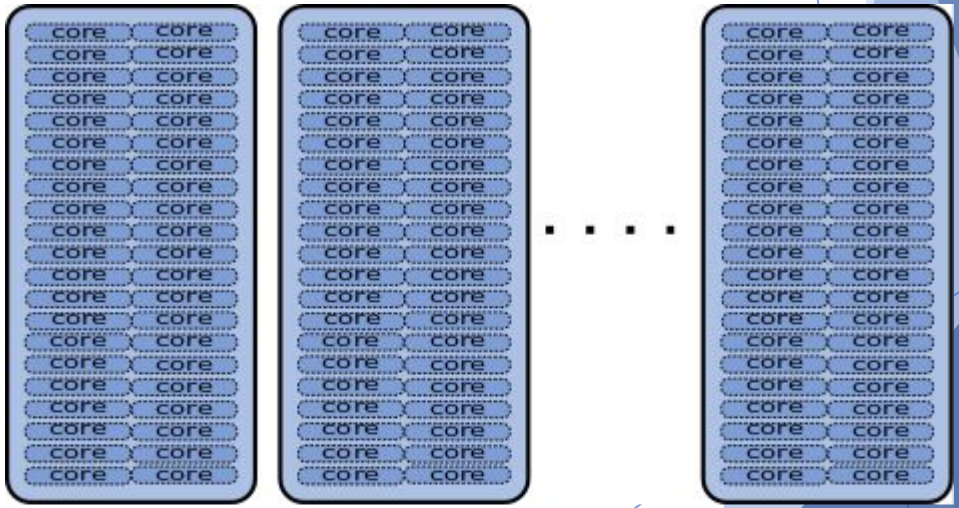
Personal Computer:

4GB RAM; 8 cores; 1 node



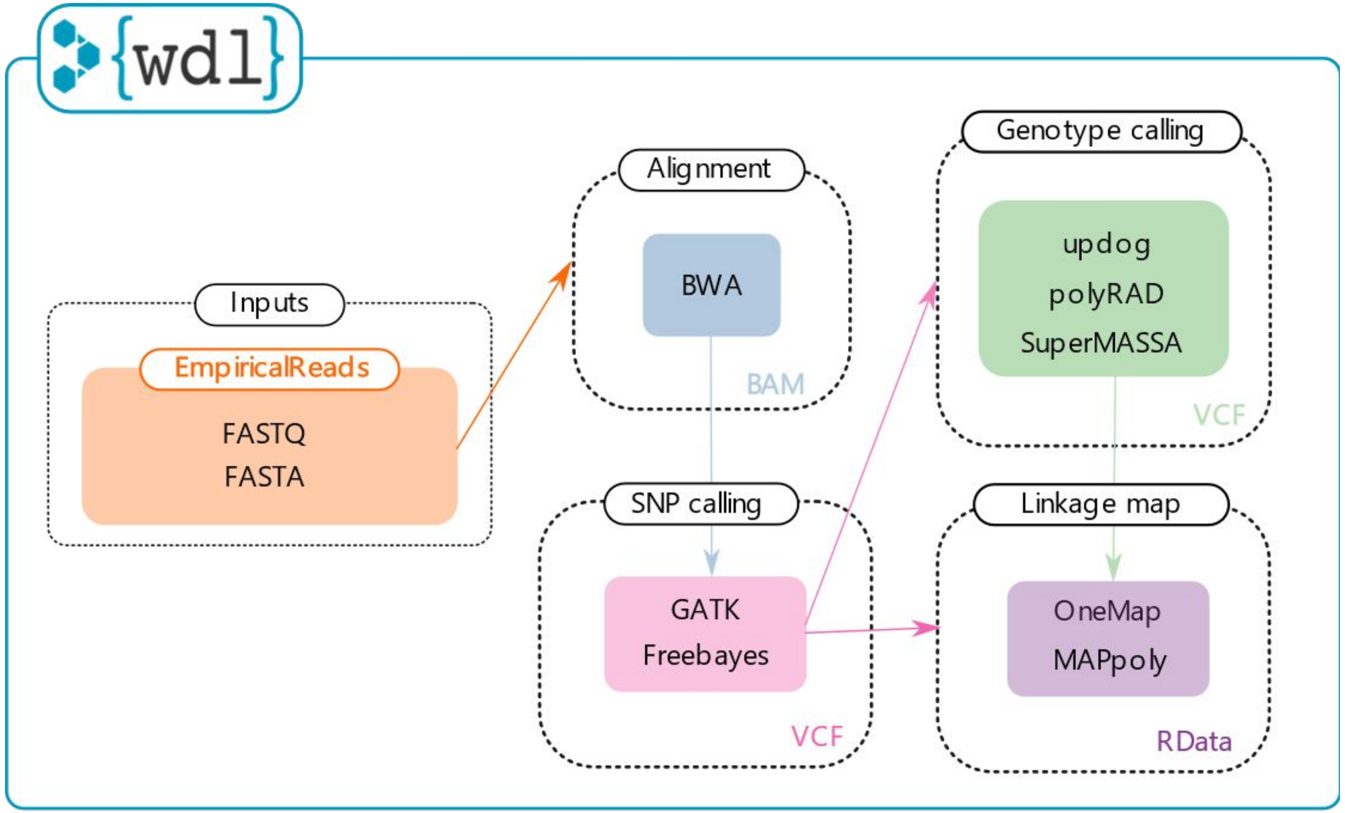
High Performance Computing (Texas A&M):

384GB; 48 cores per node; 900 nodes



TOOLS FOR
POLYPLOIDS

How we solved it: Reads2Map

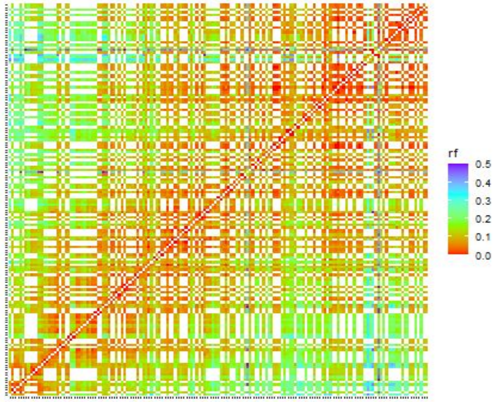


Available in Github, Dockerstore and WorkflowHub

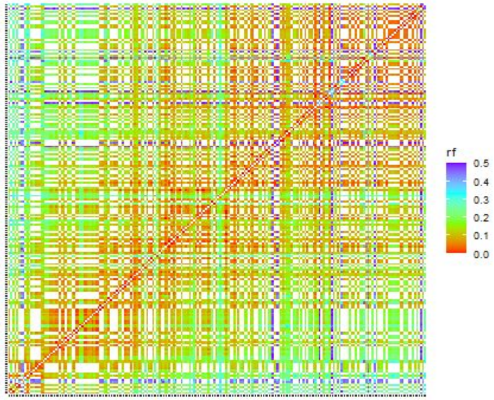
Reads2Map results - Diploid roses

- 37% of chromosome 1
- ~ 38 cM
- Sequencing depth ~ 94X

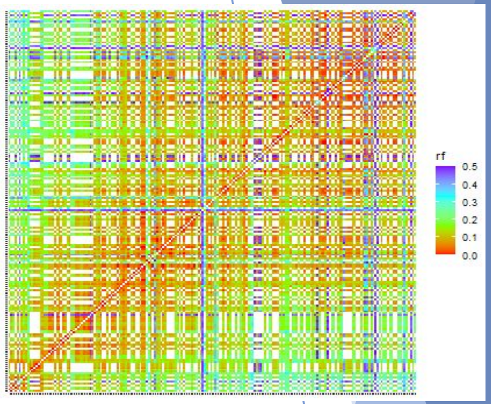
GATK



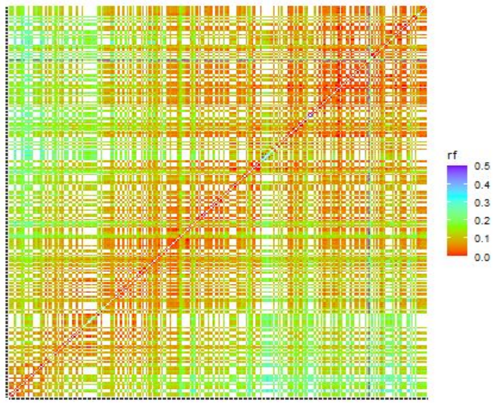
GATK + polyRAD



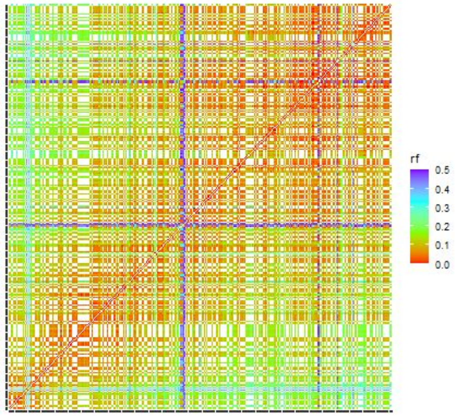
GATK + updog



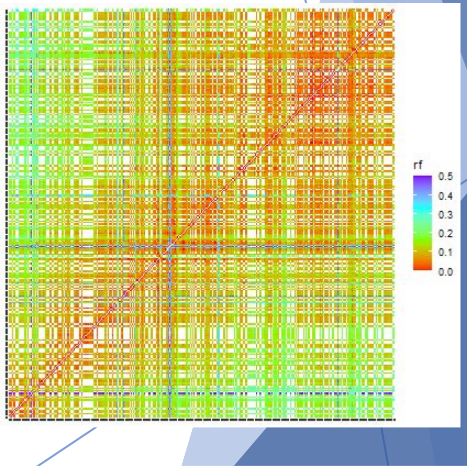
freebayes



freebayes + polyRAD

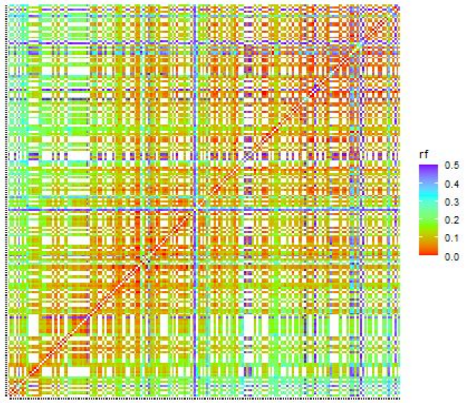


freebayes + updog

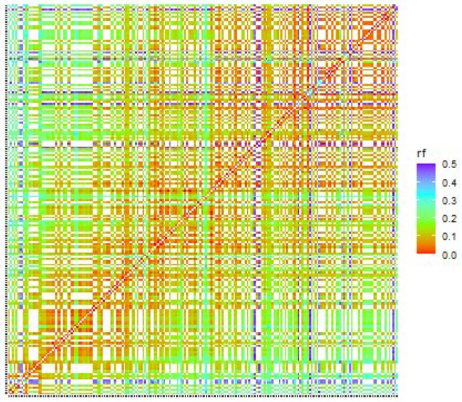


This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

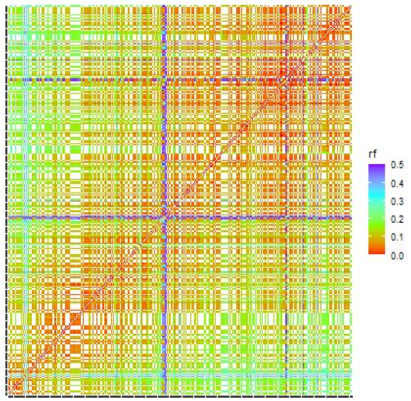
GATK + updog



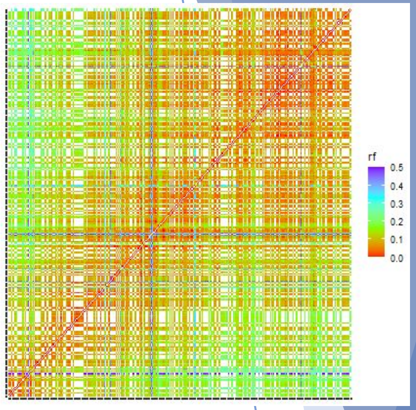
GATK + polyRAD



freebayes + polyRAD



freebayes + updog

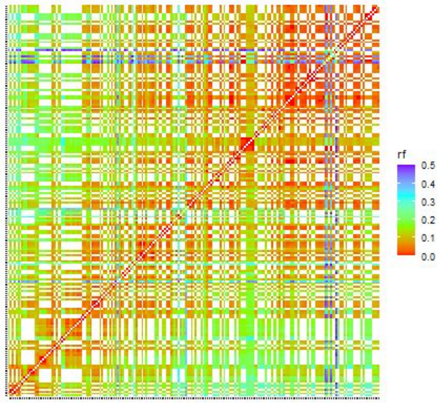


Simulation studies: 10,880 maps
Empirical studies: 816 maps

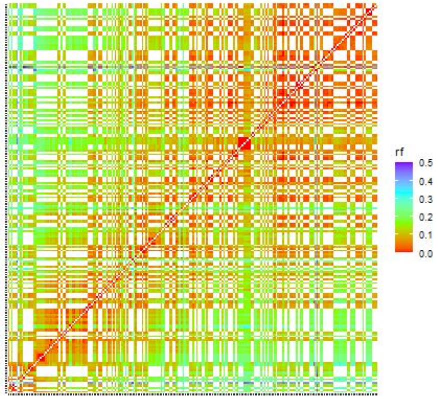
Filters:

- genotype probabilities
- non-informative markers
- replace AD by missing when GT is missing

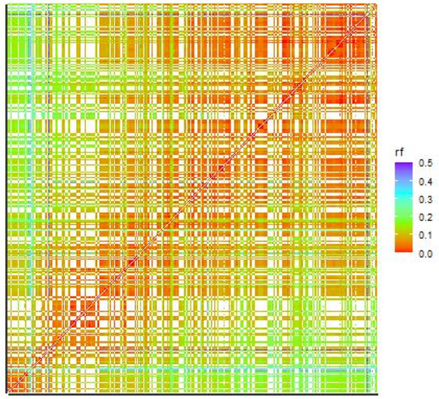
GATK + updog



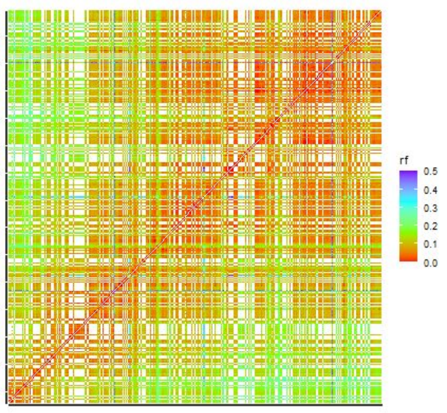
GATK + polyRAD



freebayes + polyRAD



freebayes + updog

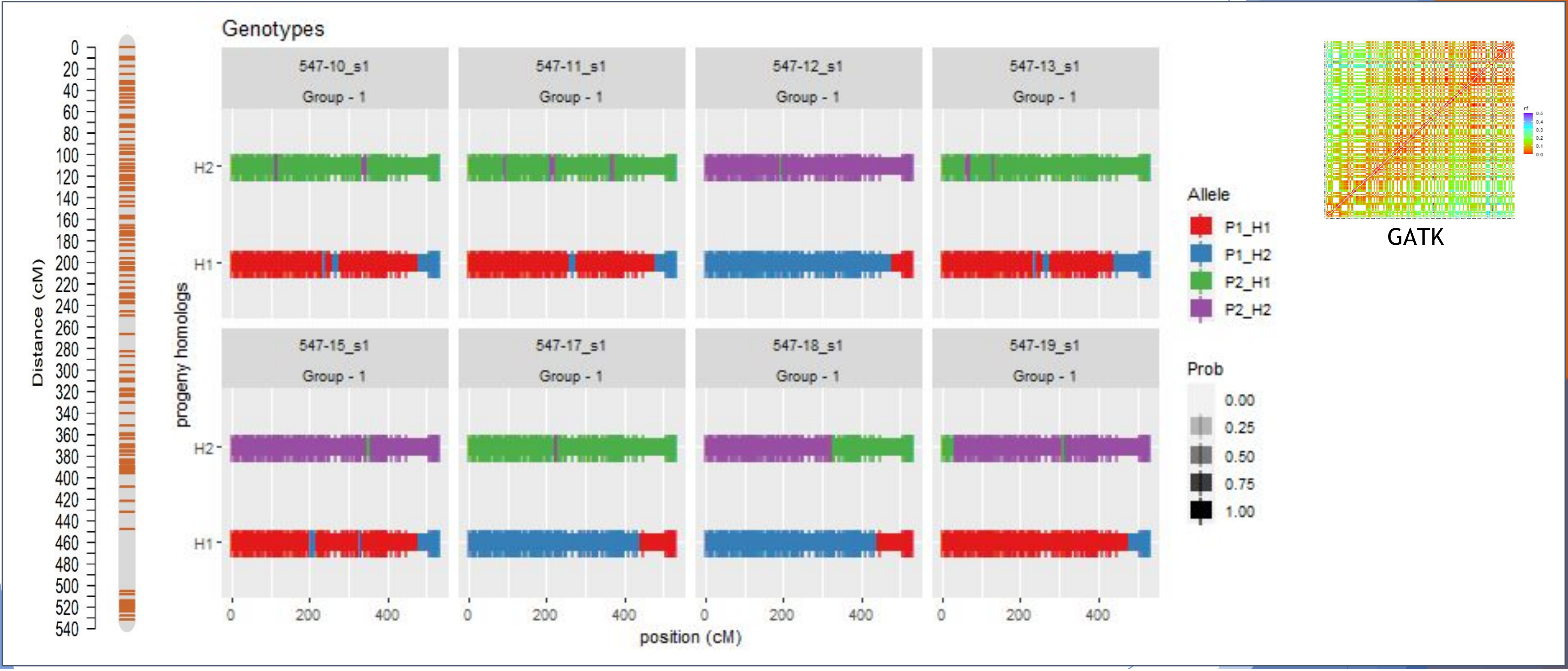


TOOLS FOR POLYPLOIDS

This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

Map size

37% of chromosome 1 ~38cM



This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

Map size

- Hidden Markov Model Emission Function
  - global error rate
  - genotype probabilities (PL)

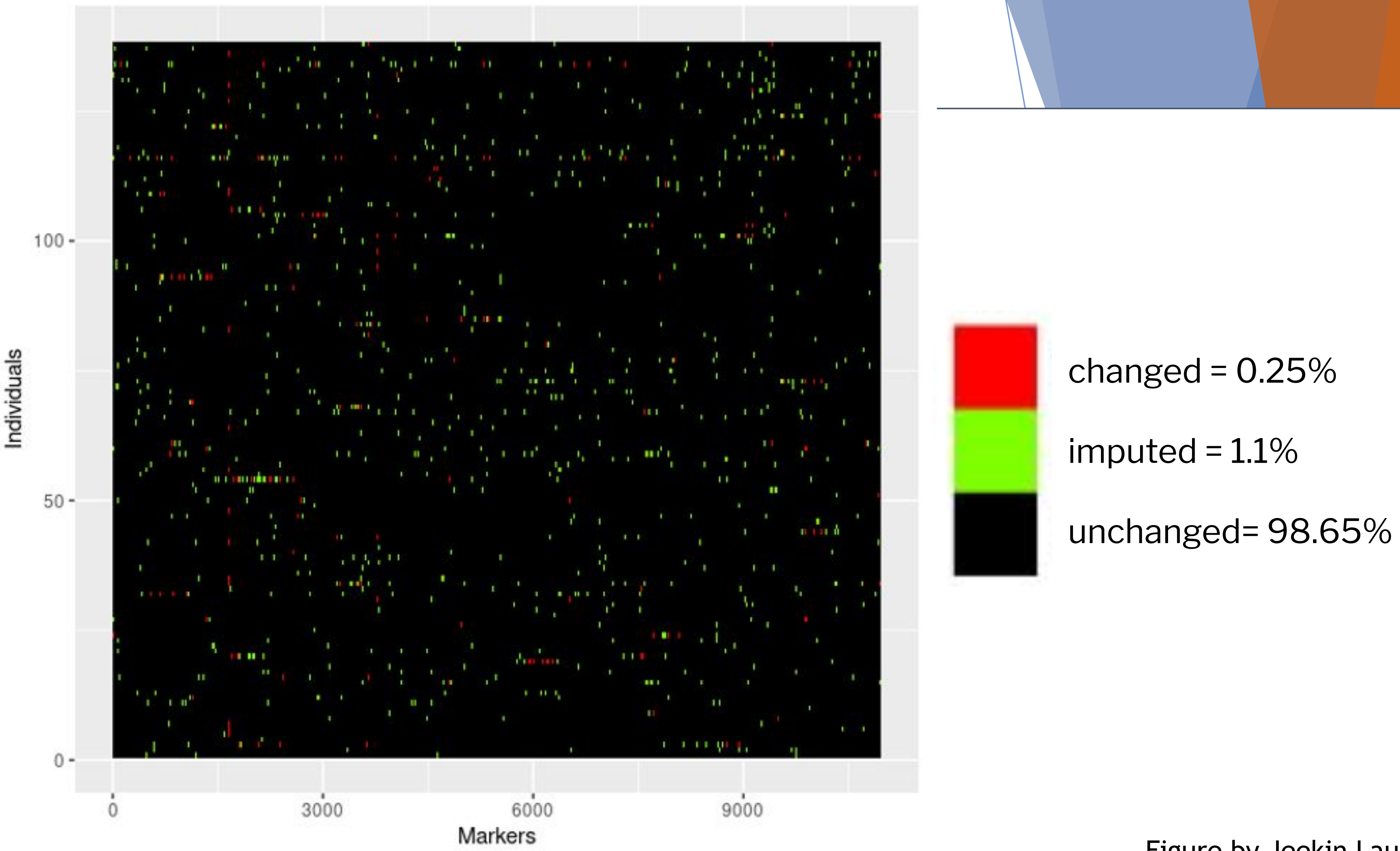


Figure by Jeekin Lau

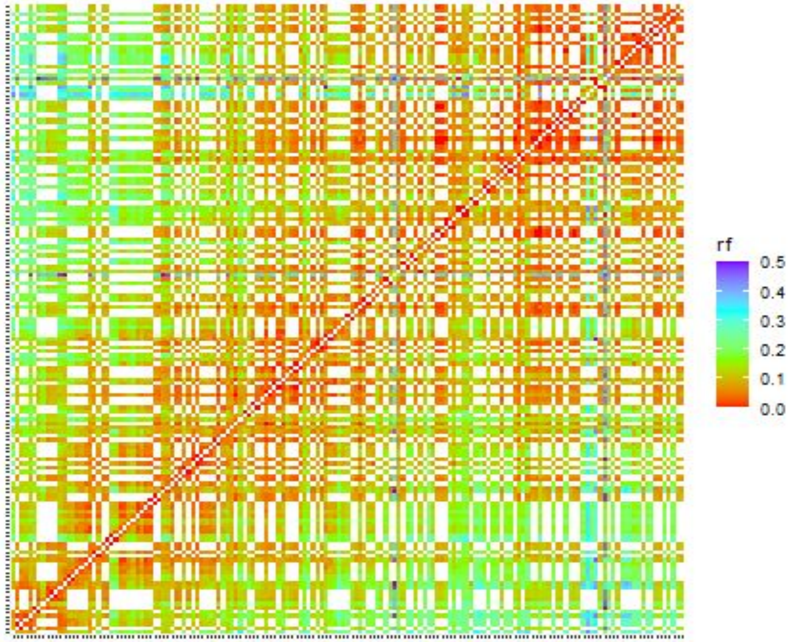
TOOLS FOR POLYPLOIDS

This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

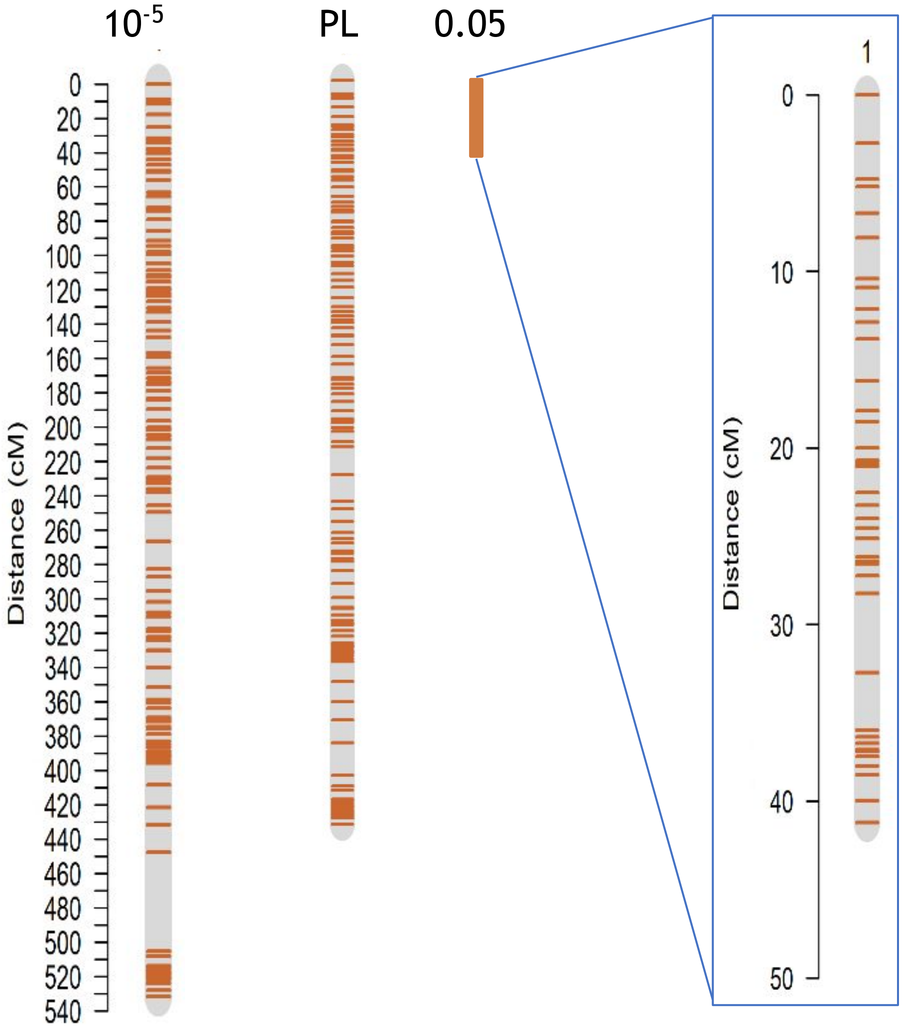
Map size

37% of chromosome 1 ~38cM

- Hidden Markov Model Emission Function
    - global error rate
    - genotype probabilities (PL)



GATK

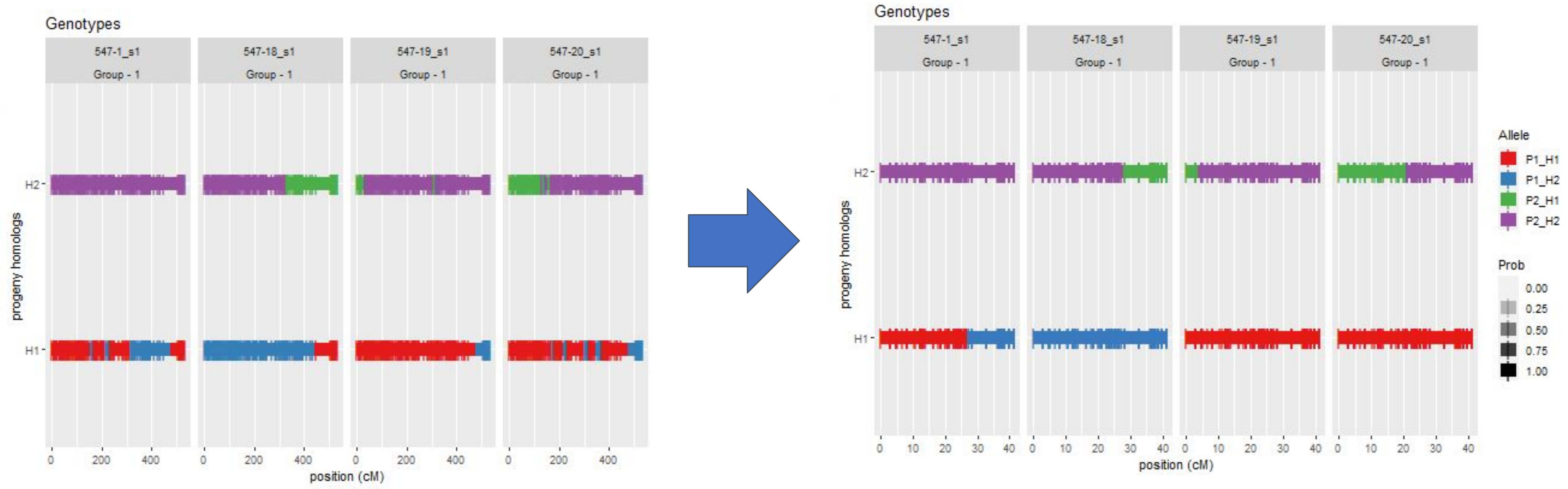


TOOLS FOR POLYPLOIDS

This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

Map size

37% of chromosome 1 ~38cM

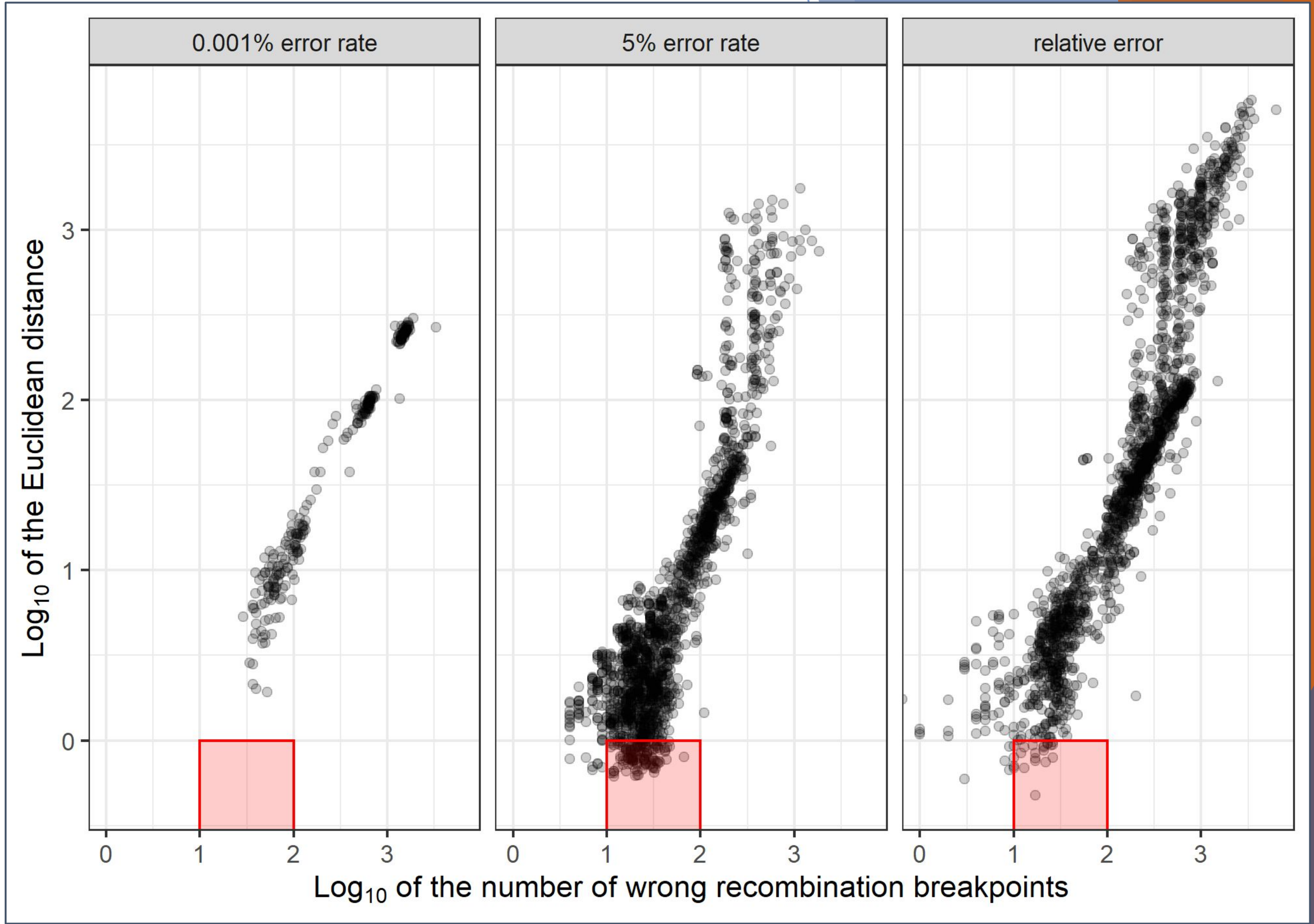


This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

Simulation study

- Match recombinations breakpoints
  - Large maps - always bad
  - Small maps - not always good

- Other tested scenarios with:
  - Segregation distortion
  - Contaminants samples
  - Multiallelic markers

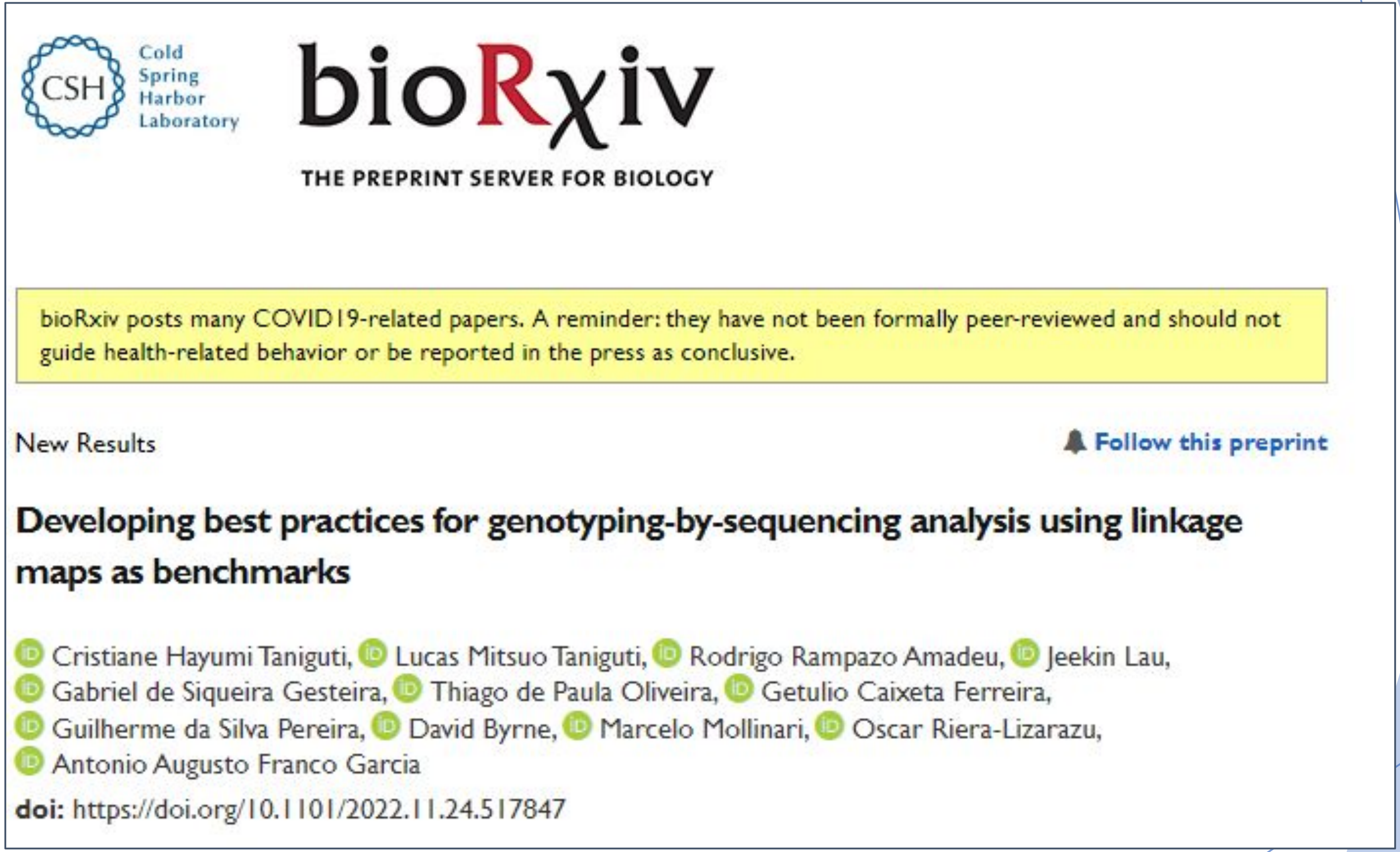


Red square: no inflated size (1 or less Euclidean distance) but have from 10 to 100 wrong recombination breakpoints

TOOLS FOR POLYPLOIDS

This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

Preprint



Cold Spring Harbor Laboratory
bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

bioRxiv posts many COVID19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

New Results

Follow this preprint

Developing best practices for genotyping-by-sequencing analysis using linkage maps as benchmarks

Cristiane Hayumi Taniguti, Lucas Mitsuo Taniguti, Rodrigo Rampazo Amadeu, Jeekin Lau, Gabriel de Siqueira Gesteira, Thiago de Paula Oliveira, Getulio Caixeta Ferreira, Guilherme da Silva Pereira, David Byrne, Marcelo Mollinari, Oscar Riera-Lizarazu, Antonio Augusto Franco Garcia

doi: https://doi.org/10.1101/2022.11.24.517847

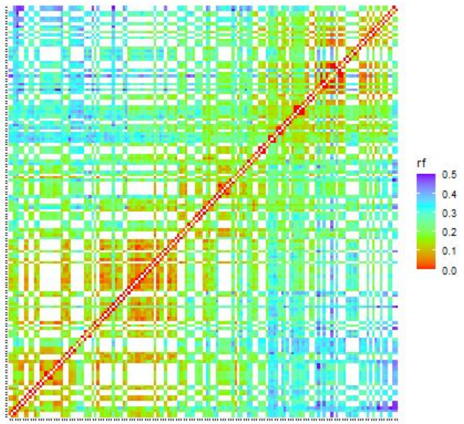
This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

Diploid Aspen

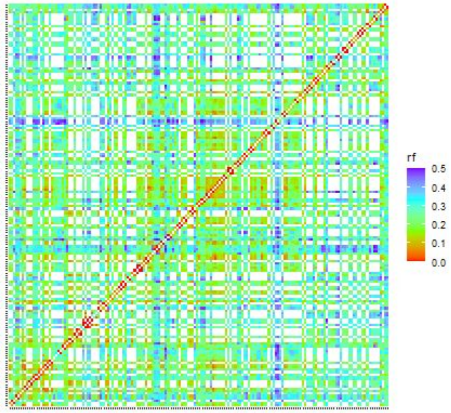
- 37% of chromosome 10
- Sequencing depth ~ 6X

OneMap

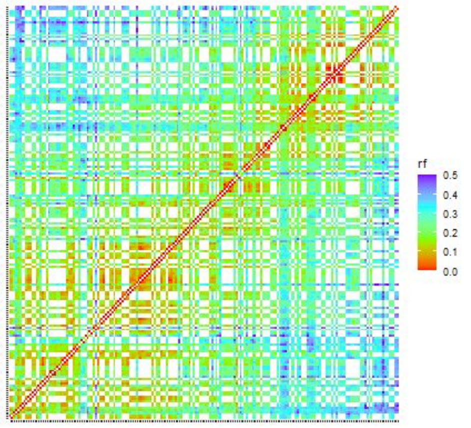
GATK



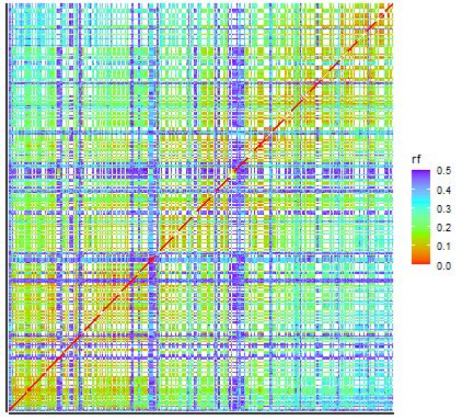
GATK + polyRAD



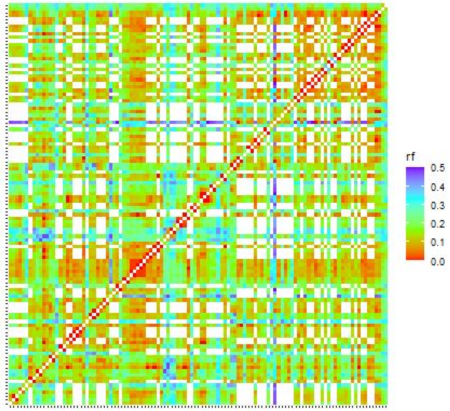
GATK + updog



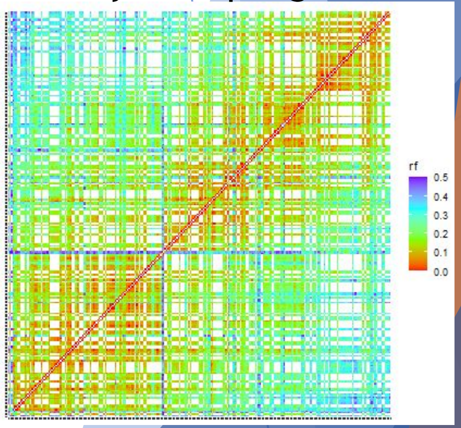
freebayes



freebayes + polyRAD



freebayes + updog



TOOLS FOR POLYPLOIDS

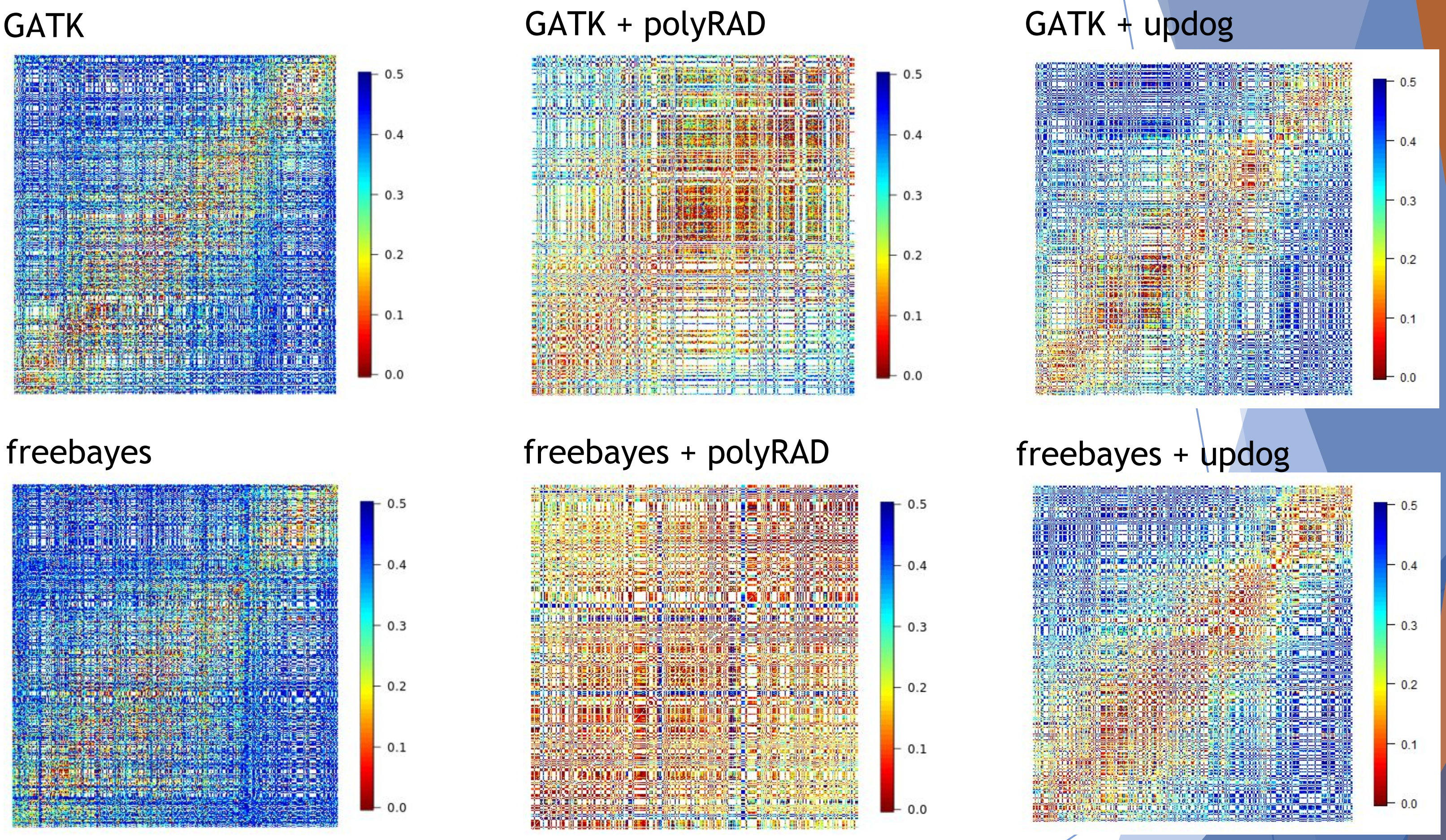
This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

TEXAS A&M UNIVERSITY

Tetraploid rose

- Chromosome 2
- Sequencing depth ~ 50X

MAPpoly



GATK

GATK + polyRAD

GATK + updog

freebayes

freebayes + polyRAD

freebayes + updog

This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

Hands-on Workshop!!

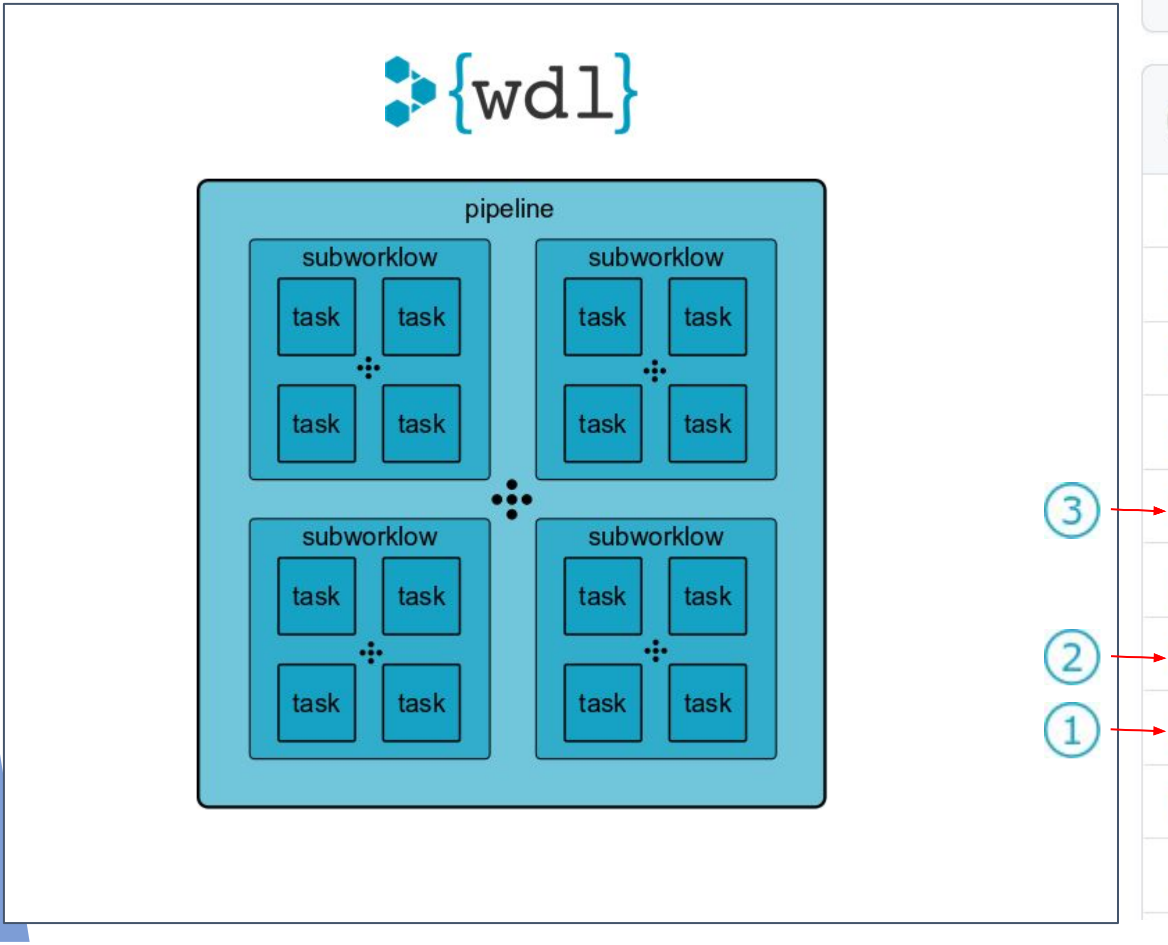
https://github.com/Cristianetaniguti/Reads2Map

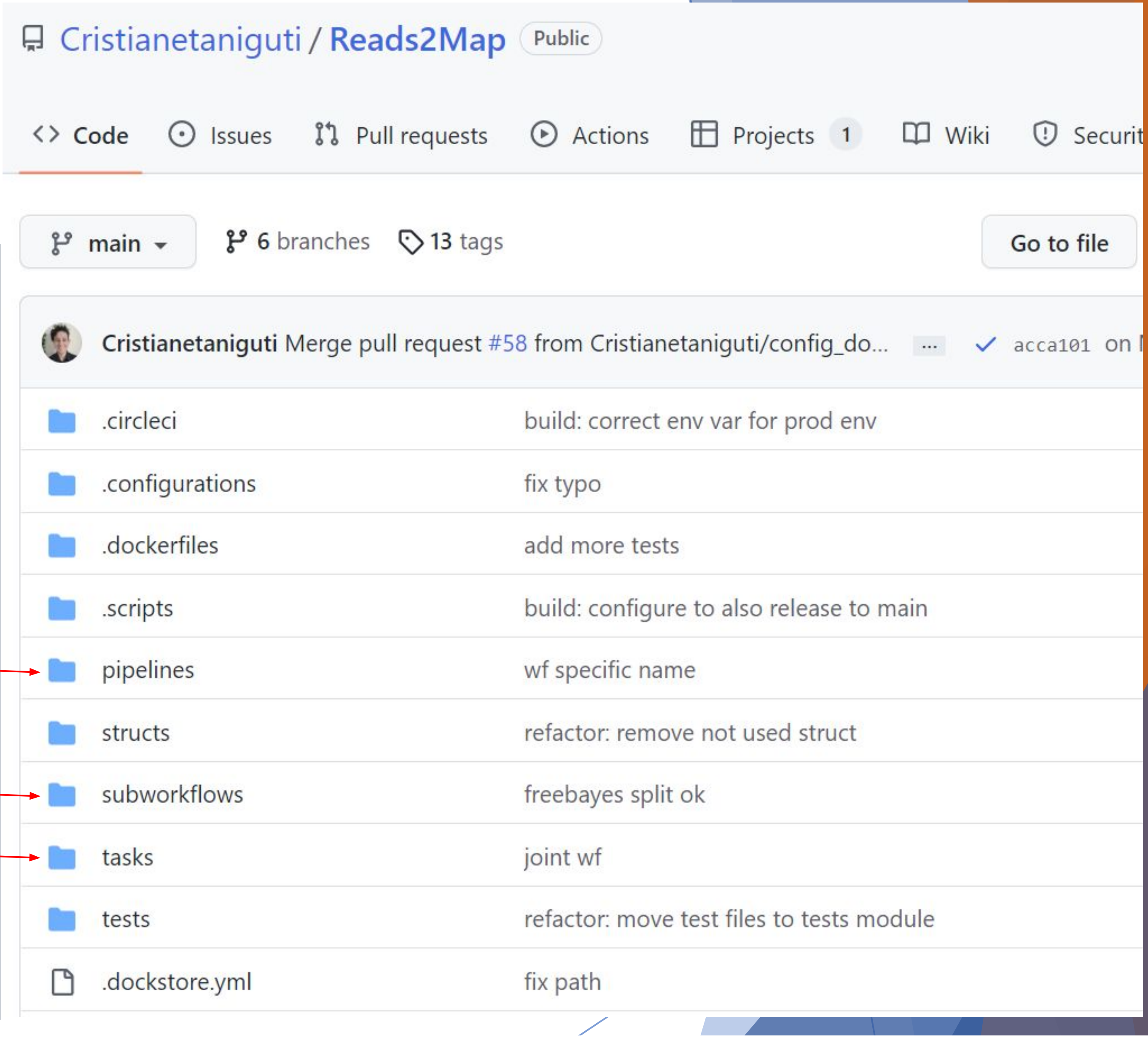
Tutorial:

bit.ly/Reads2Map2023

TOOLS FOR POLYPLOIDS

Reads2Map




Cristianetaniguti / Reads2Map Public

6 branches 13 tags

Cristianetaniguti Merge pull request #58 from Cristianetaniguti/config_do... … ✓ acca101 on

.circleci	build: correct env var for prod env
.configurations	fix typo
.dockerfiles	add more tests
.scripts	build: configure to also release to main
pipelines	wf specific name
structs	refactor: remove not used struct
subworkflows	freebayes split ok
tasks	joint wf
tests	refactor: move test files to tests module
.dockstore.yml	fix path

TOOLS FOR POLYPLOIDS

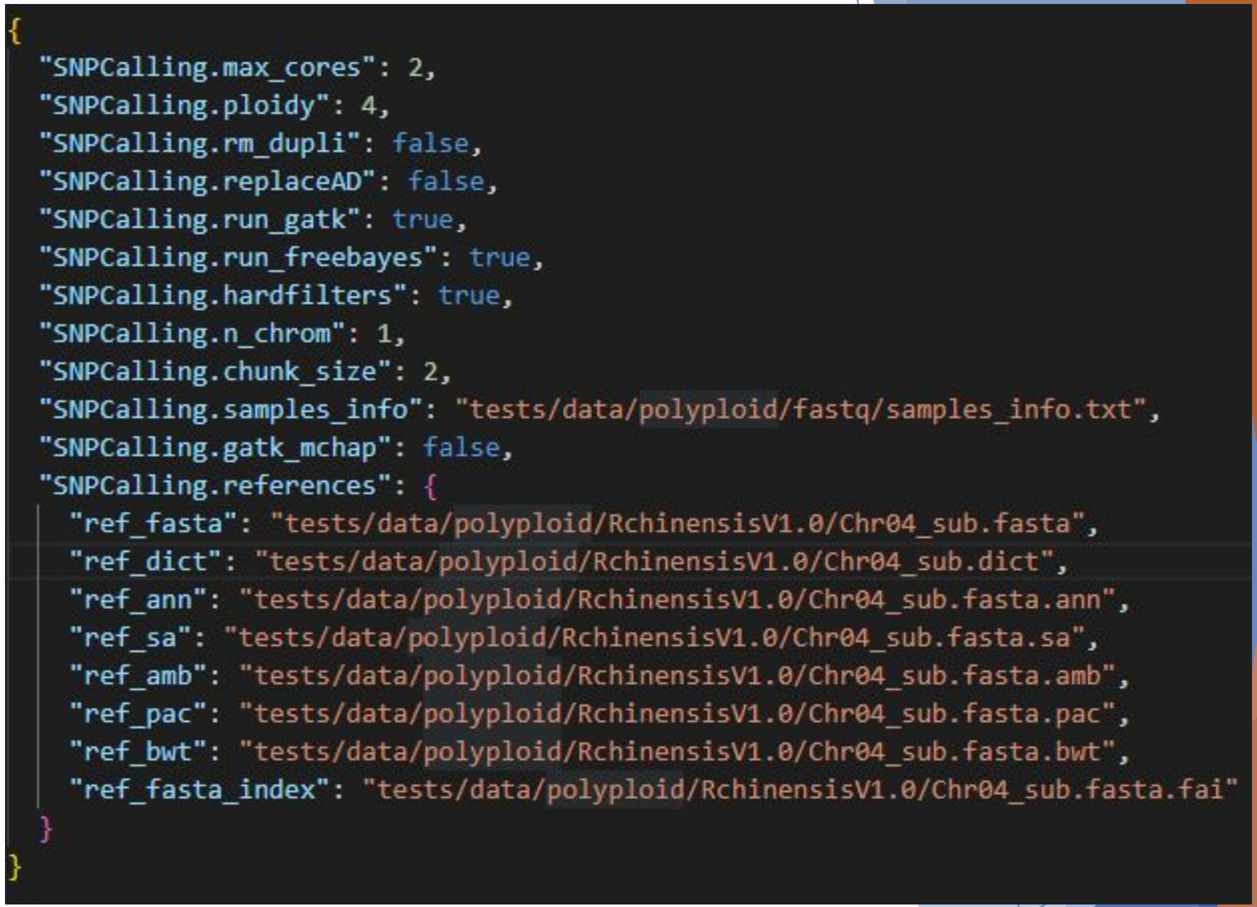
This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

TEXAS A&M UNIVERSITY

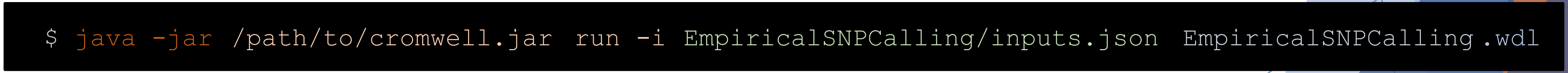
Reads2Map

- Cloud environments
  - terra.bio
- HPC
  - Cromwell
  - MiniWDL
  - dxWDL

inputs.json

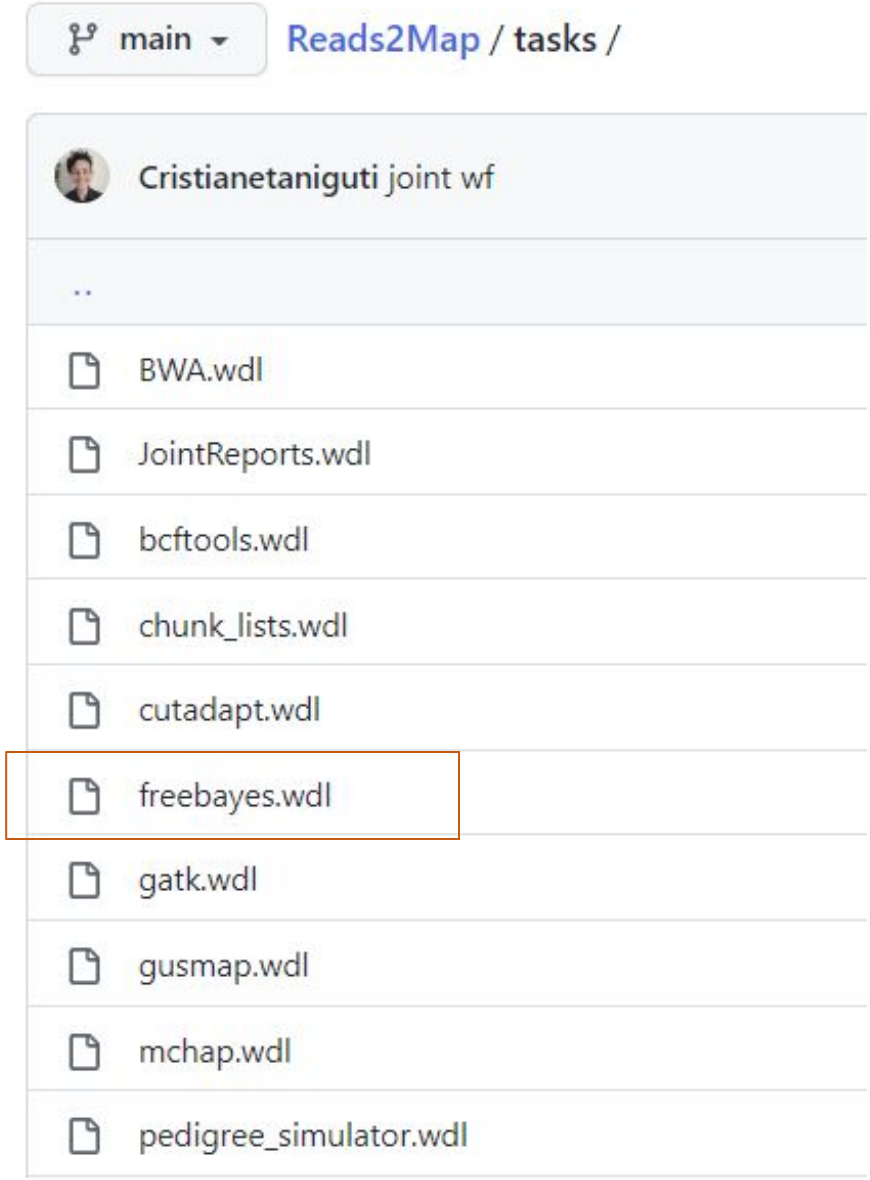


```
{
  "SNPCalling.max_cores": 2,
  "SNPCalling.ploidy": 4,
  "SNPCalling.rm_dupli": false,
  "SNPCalling.replaceAD": false,
  "SNPCalling.run_gatk": true,
  "SNPCalling.run_freebayes": true,
  "SNPCalling.hardfilters": true,
  "SNPCalling.n_chrom": 1,
  "SNPCalling.chunk_size": 2,
  "SNPCalling.samples_info": "tests/data/polyploid/fastq/samples_info.txt",
  "SNPCalling.gatk_mchap": false,
  "SNPCalling.references": {
    "ref_fasta": "tests/data/polyploid/RchinensisV1.0/Chr04_sub.fasta",
    "ref_dict": "tests/data/polyploid/RchinensisV1.0/Chr04_sub.dict",
    "ref_ann": "tests/data/polyploid/RchinensisV1.0/Chr04_sub.fasta.ann",
    "ref_sa": "tests/data/polyploid/RchinensisV1.0/Chr04_sub.fasta.sa",
    "ref_amb": "tests/data/polyploid/RchinensisV1.0/Chr04_sub.fasta.amb",
    "ref_pac": "tests/data/polyploid/RchinensisV1.0/Chr04_sub.fasta.pac",
    "ref_bwt": "tests/data/polyploid/RchinensisV1.0/Chr04_sub.fasta.bwt",
    "ref_fasta_index": "tests/data/polyploid/RchinensisV1.0/Chr04_sub.fasta.fai"
  }
}
```



```
$ java -jar /path/to/cromwell.jar run -i EmpiricalSNPCalling/inputs.json EmpiricalSNPCalling.wdl
```

task



main / Reads2Map / tasks /

Cristianetaniguti joint wf

..

BWA.wdl

JointReports.wdl

bcftools.wdl

chunk_lists.wdl

cutadapt.wdl

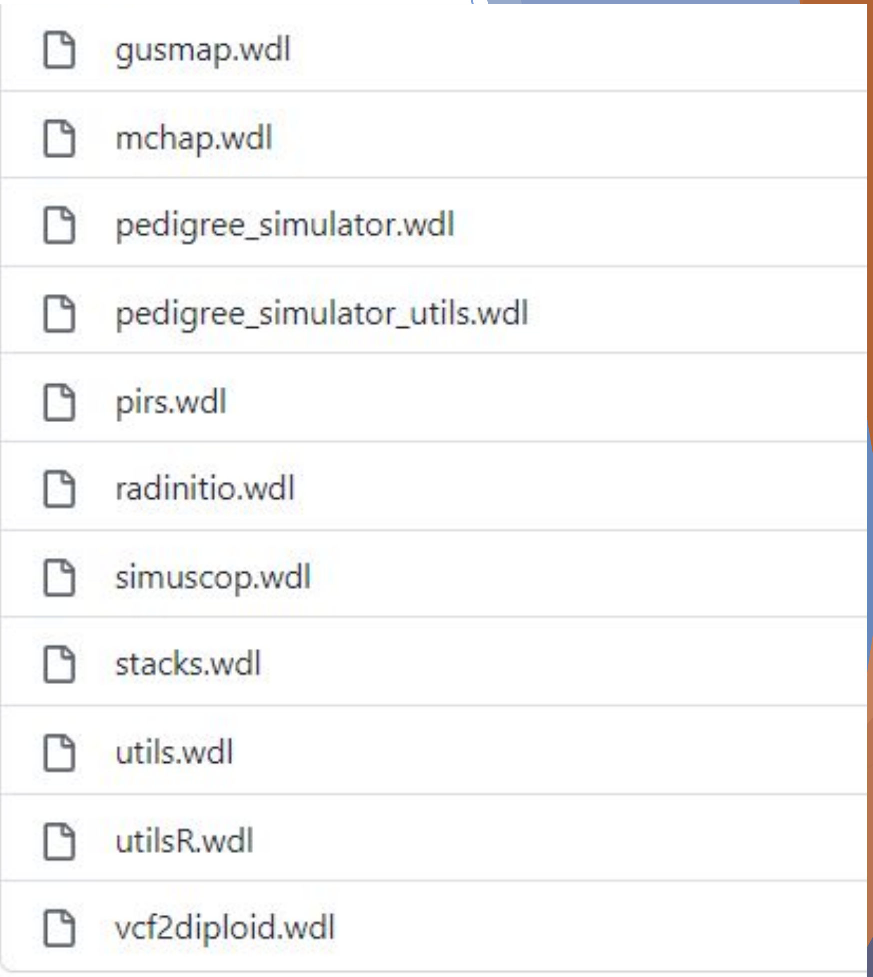
freebayes.wdl

gatk.wdl

gusmap.wdl

mchap.wdl

pedigree_simulator.wdl



gusmap.wdl

mchap.wdl

pedigree_simulator.wdl

pedigree_simulator_utils.wdl

pirs.wdl

radinitio.wdl

simuscop.wdl

stacks.wdl

utils.wdl

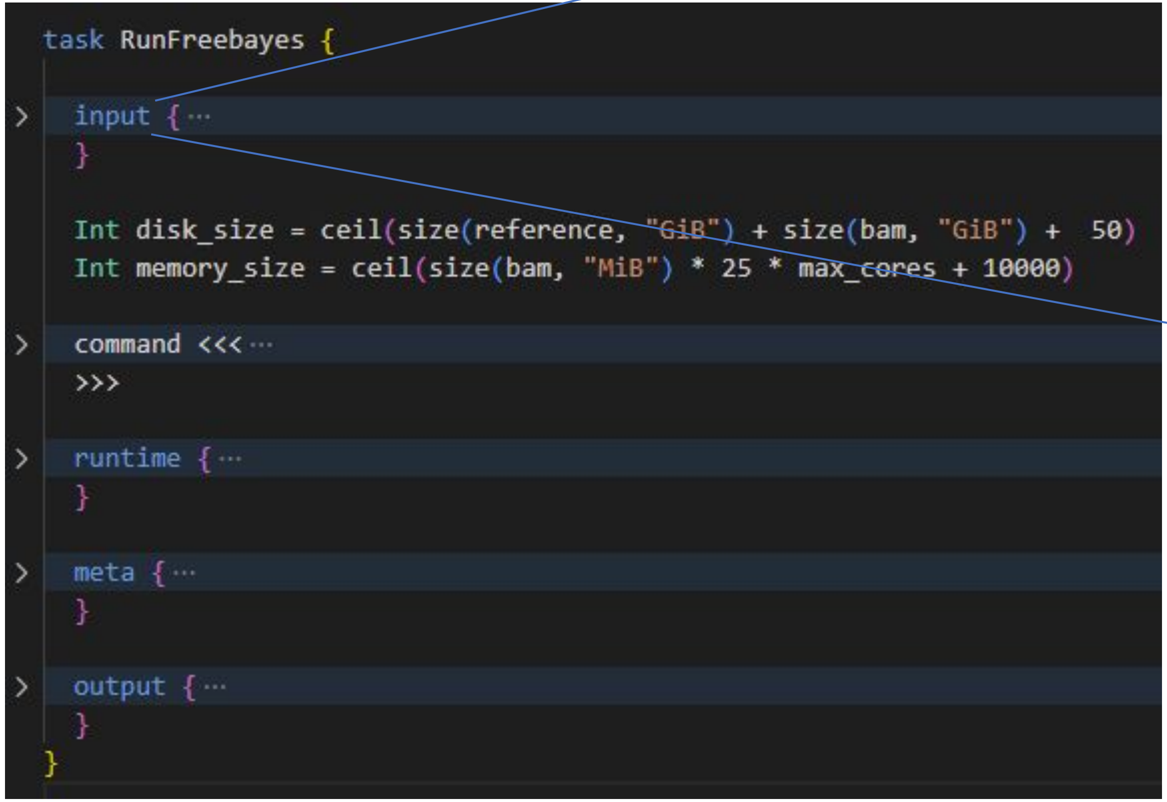
utilsR.wdl

vcf2diploid.wdl

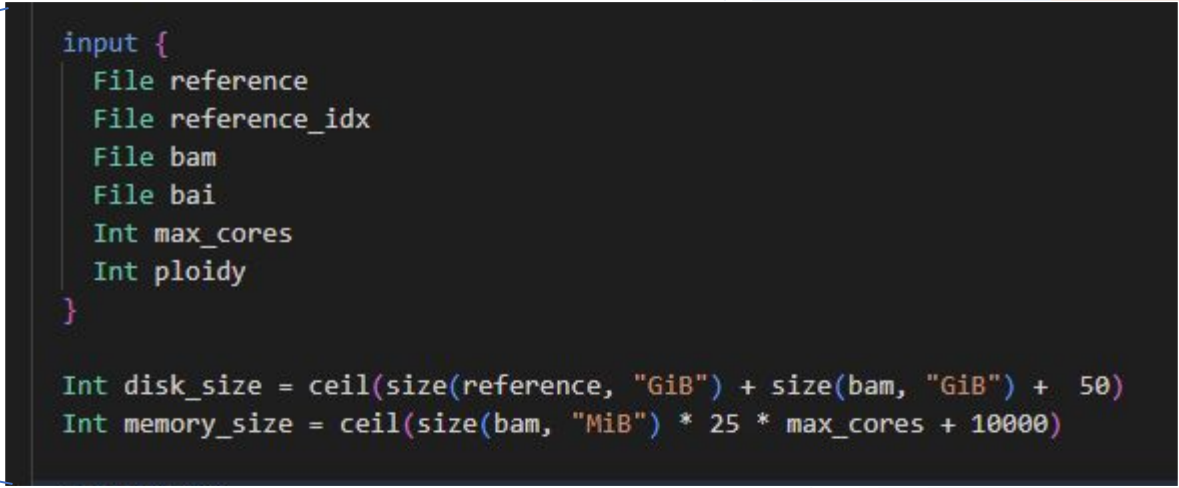
This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

task

example: freebayes.wdl

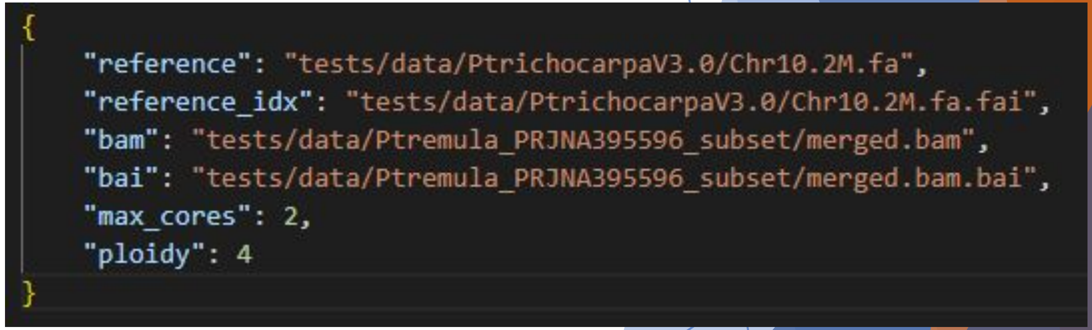


```
task RunFreebayes {

  input {…
  }

  Int disk_size = ceil(size(reference, "GiB") + size(bam, "GiB") +  50)
  Int memory_size = ceil(size(bam, "MiB") * 25 * max_cores + 10000)

  command <<<…
  >>>

  runtime {…
  }

  meta {…
  }

  output {…
  }
}
```



```
input {
  File reference
  File reference_idx
  File bam
  File bai
  Int max_cores
  Int ploidy
}

Int disk_size = ceil(size(reference, "GiB") + size(bam, "GiB") +  50)
Int memory_size = ceil(size(bam, "MiB") * 25 * max_cores + 10000)
```

inputs.json

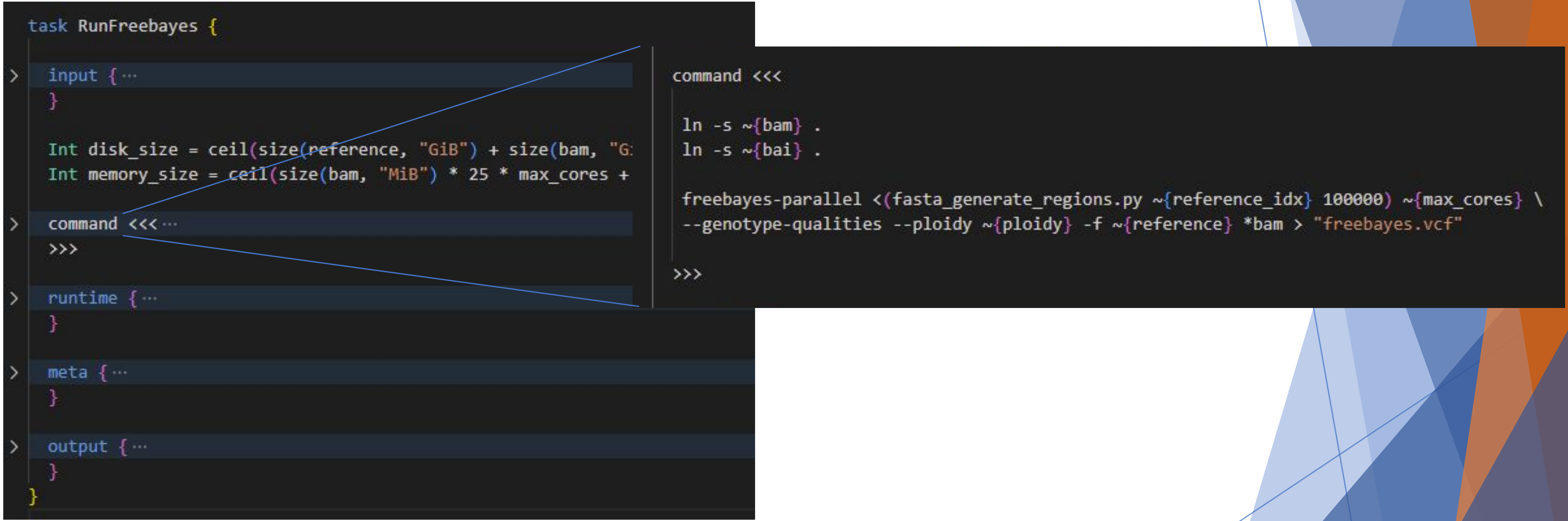


```
{
    "reference": "tests/data/PtrichocarpaV3.0/Chr10.2M.fa",
    "reference_idx": "tests/data/PtrichocarpaV3.0/Chr10.2M.fa.fai",
    "bam": "tests/data/Ptremula_PRJNA395596_subset/merged.bam",
    "bai": "tests/data/Ptremula_PRJNA395596_subset/merged.bam.bai",
    "max_cores": 2,
    "ploidy": 4
}
```

This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

task

example: freebayes.wdl



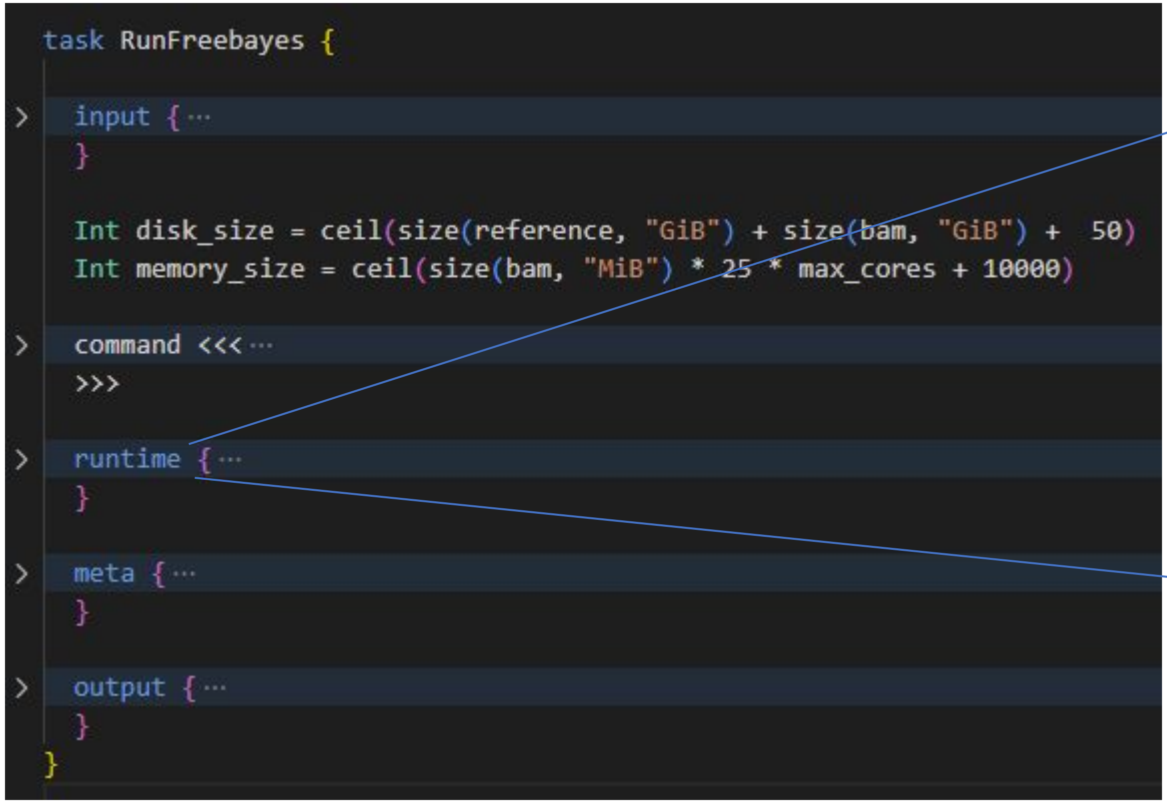
```
task RunFreebayes {

  input {···
  }

  Int disk_size = ceil(size(reference, "GiB") + size(bam, "G
  Int memory_size = ceil(size(bam, "MiB") * 25 * max_cores +

  command <<<···
  >>>

  runtime {···
  }

  meta {···
  }

  output {···
  }
}
```

```
command <<<

 ln -s ~{bam} .
 ln -s ~{bai} .

 freebayes-parallel <(fasta_generate_regions.py ~{reference_idx} 100000) ~{max_cores} \
 --genotype-qualities --ploidy ~{ploidy} -f ~{reference} *bam > "freebayes.vcf"

>>>
```

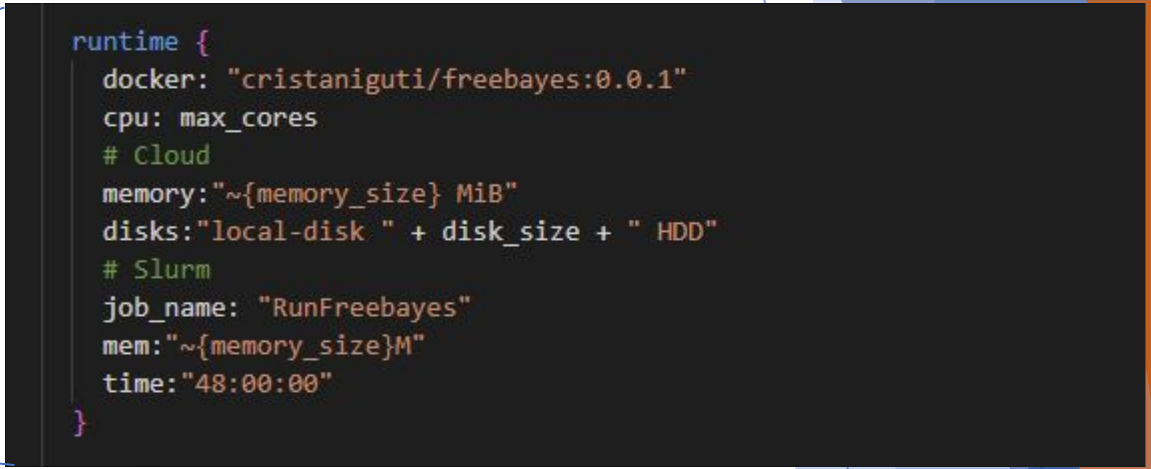
This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

task

example: freebayes.wdl



```
task RunFreebayes {

  input {···
  }

  Int disk_size = ceil(size(reference, "GiB") + size(bam, "GiB") +  50)
  Int memory_size = ceil(size(bam, "MiB") * 25 * max_cores + 10000)

  command <<<···
  >>>

  runtime {···
  }

  meta {···
  }

  output {···
  }
}
```



```
runtime {
  docker: "cristaniguti/freebayes:0.0.1"
  cpu: max_cores
  # Cloud
  memory:"~{memory_size} MiB"
  disks:"local-disk " + disk_size + " HDD"
  # Slurm
  job_name: "RunFreebayes"
  mem:"~{memory_size}M"
  time:"48:00:00"
}
```

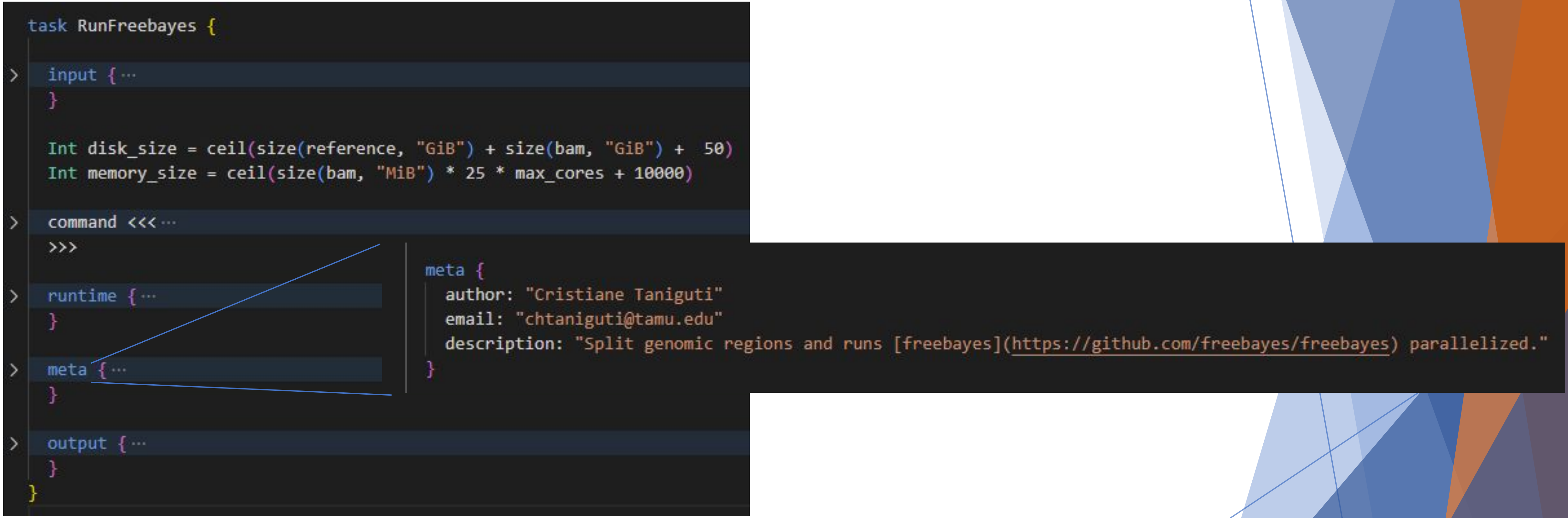
TOOLS FOR POLYPLOIDS

This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

TEXAS A&M UNIVERSITY

task

example: freebayes.wdl



```
task RunFreebayes {

  input {…
  }

  Int disk_size = ceil(size(reference, "GiB") + size(bam, "GiB") +  50)
  Int memory_size = ceil(size(bam, "MiB") * 25 * max_cores + 10000)

  command <<<…
  >>>

  runtime {…
  }

  meta {…
  }

  output {…
  }
}
```

```
meta {
  author: "Cristiane Taniguti"
  email: "chtaniguti@tamu.edu"
  description: "Split genomic regions and runs [freebayes](https://github.com/freebayes/freebayes) parallelized."
}
```

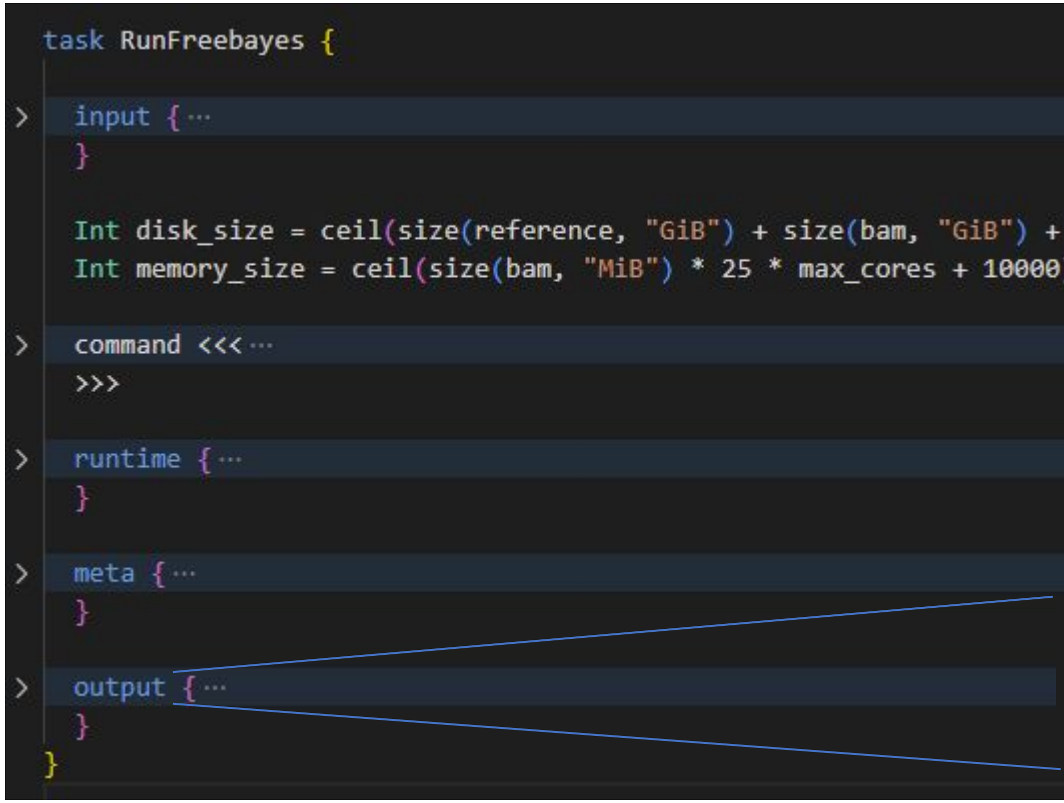
TOOLS FOR POLYPLOIDS

This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

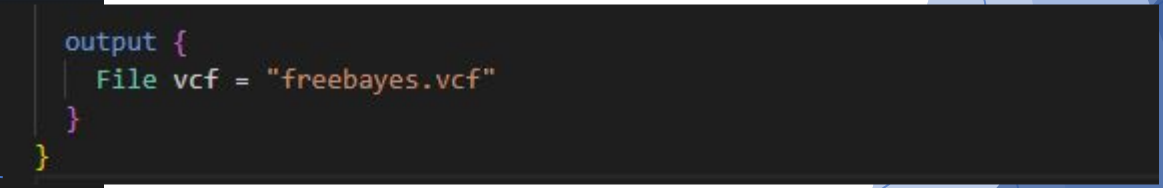
TEXAS A&M UNIVERSITY

task

example: freebayes.wdl

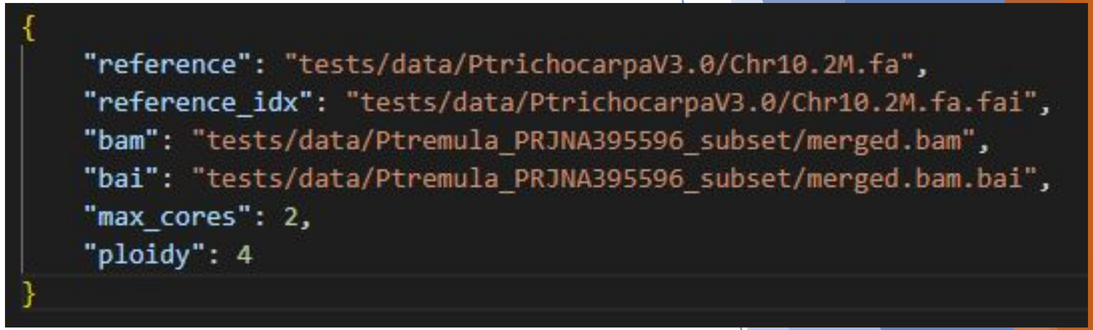


```
task RunFreebayes {

  input {…
  }

  Int disk_size = ceil(size(reference, "GiB") + size(bam, "GiB") +  50)
  Int memory_size = ceil(size(bam, "MiB") * 25 * max_cores + 10000)

  command <<<…
  >>>

  runtime {…
  }

  meta {…
  }

  output {…
  }
}
```

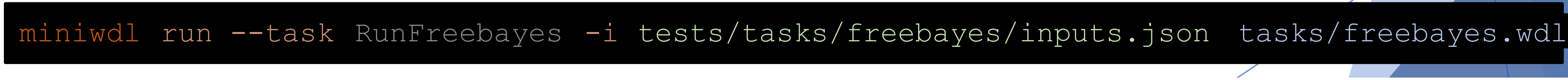


```
  output {
    File vcf = "freebayes.vcf"
  }
}
```

inputs.json



```
{
    "reference": "tests/data/PtrichocarpaV3.0/Chr10.2M.fa",
    "reference_idx": "tests/data/PtrichocarpaV3.0/Chr10.2M.fa.fai",
    "bam": "tests/data/Ptremula_PRJNA395596_subset/merged.bam",
    "bai": "tests/data/Ptremula_PRJNA395596_subset/merged.bam.bai",
    "max_cores": 2,
    "ploidy": 4
}
```

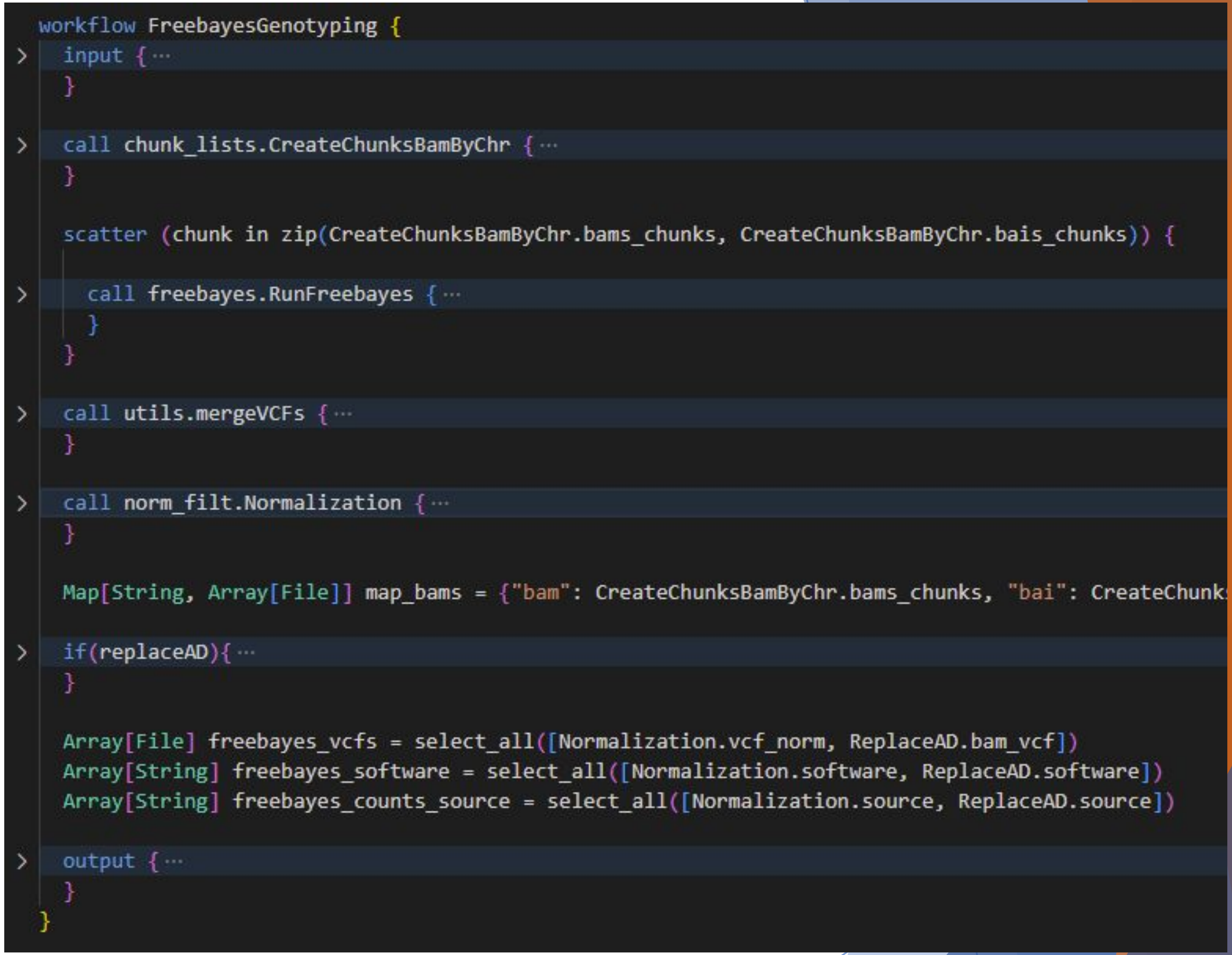


```
miniwdl run --task RunFreebayes -i tests/tasks/freebayes/inputs.json  tasks/freebayes.wdl
```

This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

subworkflow

example

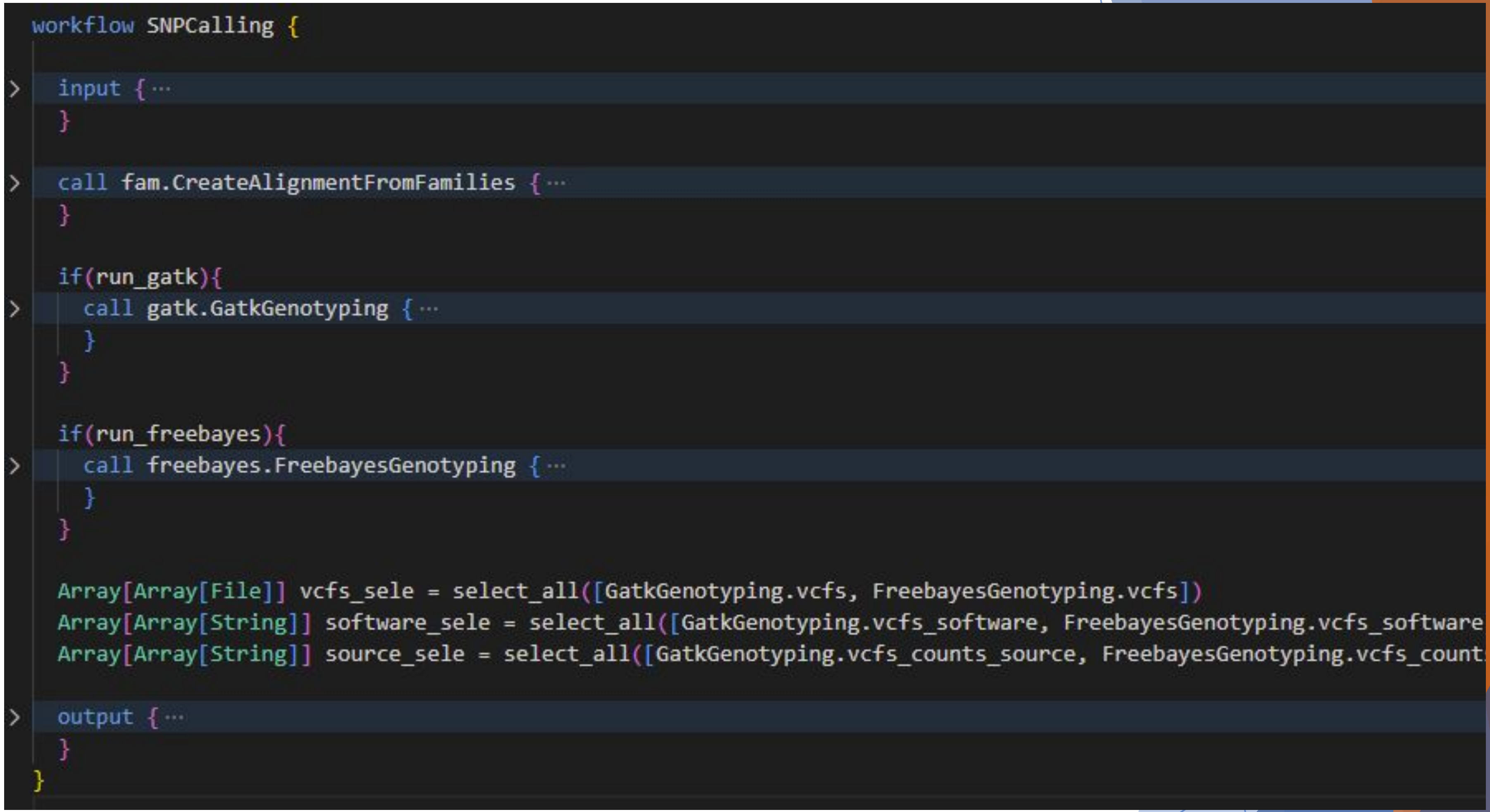


```
workflow FreebayesGenotyping {
  input {…
  }

  call chunk_lists.CreateChunksBamByChr {…
  }

  scatter (chunk in zip(CreateChunksBamByChr.bams_chunks, CreateChunksBamByChr.bais_chunks)) {

    call freebayes.RunFreebayes {…
    }
  }

  call utils.mergeVCFs {…
  }

  call norm_filt.Normalization {…
  }

  Map[String, Array[File]] map_bams = {"bam": CreateChunksBamByChr.bams_chunks, "bai": CreateChunk

  if(replaceAD){…
  }

  Array[File] freebayes_vcfs = select_all([Normalization.vcf_norm, ReplaceAD.bam_vcf])
  Array[String] freebayes_software = select_all([Normalization.software, ReplaceAD.software])
  Array[String] freebayes_counts_source = select_all([Normalization.source, ReplaceAD.source])

  output {…
  }
}
```

```
$ java -jar /path/to/cromwell.jar run -i freebayes_genotyping/inputs.json
freebayes_genotyping.wdl
```

This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

pipeline
example



```
workflow SNPCalling {

  input {···
  }

  call fam.CreateAlignmentFromFamilies {···
  }

  if(run_gatk){
    call gatk.GatkGenotyping {···
    }
  }

  if(run_freebayes){
    call freebayes.FreebayesGenotyping {···
    }
  }

  Array[Array[File]] vcfs_sele = select_all([GatkGenotyping.vcfs, FreebayesGenotyping.vcfs])
  Array[Array[String]] software_sele = select_all([GatkGenotyping.vcfs_software, FreebayesGenotyping.vcfs_software
  Array[Array[String]] source_sele = select_all([GatkGenotyping.vcfs_counts_source, FreebayesGenotyping.vcfs_count

  output {···
  }
}
```

```
$ java -jar /path/to/cromwell.jar run -i EmpiricalSNPCalling/inputs.json EmpiricalSNPCalling.wdl
```

This project is funded by USDA NIFA Specialty Crop Research Initiative
Award # 2020-51181-32156 (09/01/20 - 08/31/24)

Thank you

TEXAS A&M UNIVERSITY

Oscar Riera-Lizarazu
David Byrne
Jeekin Lau
Tessa Hochhaus

STATISTICAL GENETICS ·LAB·
ESALQ / USP

Augusto Garcia
Rodrigo Amadeu
Getulio Caixeta

NC STATE UNIVERSITY

Marcelo Mollinari
Gabriel Gesteira

Mendelics

Lucas Taniguti

UNIVERSIDADE FEDERAL DE VIÇOSA

Guilherme Pereira

ROSLIN

Thiago Oliveira

TOOLS FOR POLYPLOIDS

Thank you!!

- Susan Thomson
- Cecilia Deng
- Ben Warren

Plant & Food™
Research
Rangahau Ahumāra Kai

Project Members

TEXAS A&M UNIVERSITY

Plant & Food™ Research
Rangahau Ahumāra Kai

Cornell University

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

PennState

WAGENINGEN
UNIVERSITY & RESEARCH

WASHINGTON STATE UNIVERSITY

UNIVERSITY OF MINNESOTA
Driven to Discover®

NC STATE UNIVERSITY

UNIVERSITY OF ARKANSAS

1865 THE UNIVERSITY OF MAINE

Oregon State University

TOOLS FOR POLYPLOIDS

This project is funded by USDA NIFA Specialty Crop Research Initiative Award # 2020-51181-32156 (09/01/20 - 08/31/24)

Other Collaborators

Genetwister

DAICEL arbor biosciences

VacCAP

Neuhouse Farms

bejo

Spring Meadow™

Royal Van Zanten

American Garden Rose Selections™

Plant Sciences Inc. PSI
Advancing Agriculture through Science

IITA
Transforming African Agriculture

CIAT

CIP
INTERNATIONAL POTATO CENTER

Potatoes® USA

Woolf Roses L.L.C.

star®
ROSES and PLANTS

Dekker
chrysanten

BAILEY®

CSS FARMS

Weeks Roses

DÜMMEN ORANGE

illumina®

PACIFIC BERRY
BREEDING

DELIFLOR
Chrysanten

genVATIONS

Driscoll's

This project is funded by USDA NIFA Specialty Crop Research Initiative Award # 2020-51181-32156 (09/01/20 - 08/31/24)

Links for Tutorials

- polyRAD tutorial
- updog tutorial
- fitPoly tutorial
- (TASSEL) Variant and Genotype Calling in Highly Duplicated Genomes (Lindsay Clark)
- Using Reads2Map workflows for SNP and dosage calling in polyploid sequencing data

More cool tools related to Workflows systems:

- cromwell-cli (by my brother)
- cromwell server
- womtools

References

- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A.; Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. Molecular Ecology, 22(11), 3124–3140. https://doi.org/10.1111/mec.12354
- Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q.; Buckler, E. S. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS ONE, 9(2), 1–11. https://doi.org/10.1371/journal.pone.0090346
- Garrison, E.; Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. ArXiv E-Prints, 9. https://doi.org/1207.3907
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M.; DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

This project is funded by USDA NIFA Specialty Crop Research Initiative Award
# 2020-51181-32156 (09/01/20 - 08/31/24)

References

- Gerard, D., Ferrão, L. F. V., Garcia, A. A. F.,& Stephens, M. (2018). Genotyping Polyploids from Messy Sequencing Data. Genetics, 210(3), 789-807. doi: 10.1534/genetics.118.301468.
- Wadl, P. A., Olukolu, B. A., Branham, S. E., Jarret, R. L., Yencho, G. C.; Jackson, D. M. (2018). Genetic Diversity and Population Structure of the USDA Sweetpotato (Ipomoea batatas) Germplasm Collections Using GBSpoly. Frontiers in Plant Science, 9, 1166. https://doi.org/10.3389/fpls.2018.01166
- Serang, O., Mollinari, M.; Garcia, A. A. F. (2012). Efficient exact maximum a posteriori computation for bayesian SNP genotyping in polyploids. PLoS ONE, 7(2), 1–13. https://doi.org/10.1371/journal.pone.0030906
- Clark, L. v., Lipka, A. E.; Sacks, E. J. (2019). polyRAD: Genotype Calling with Uncertainty from Sequencing Data in Polyploids and Diploids. G3: Genes|Genomes|Genetics, 9(March), g3.200913.2018. https://doi.org/10.1534/g3.118.200913

This project is funded by USDA NIFA Specialty Crop Research Initiative Award
# 2020-51181-32156 (09/01/20 - 08/31/24)