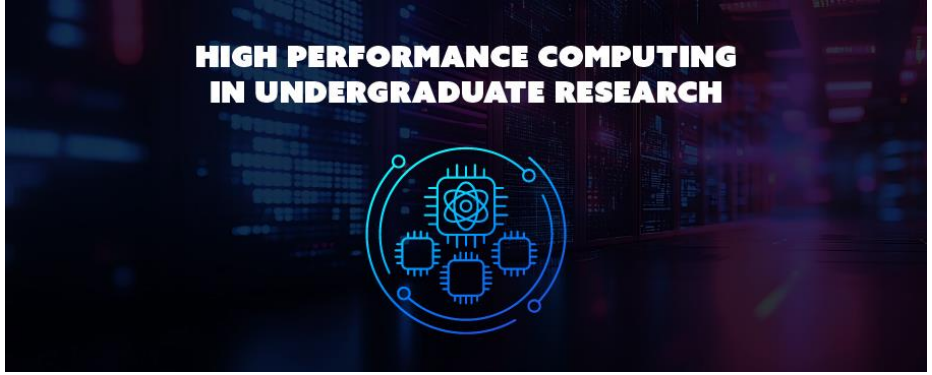


Sheep breed composition assessment using PCA

Dr. Joel D. Leal-Gutierrez
Laboratory of Genomic Dissection
College of Agriculture
Lincoln University of Missouri

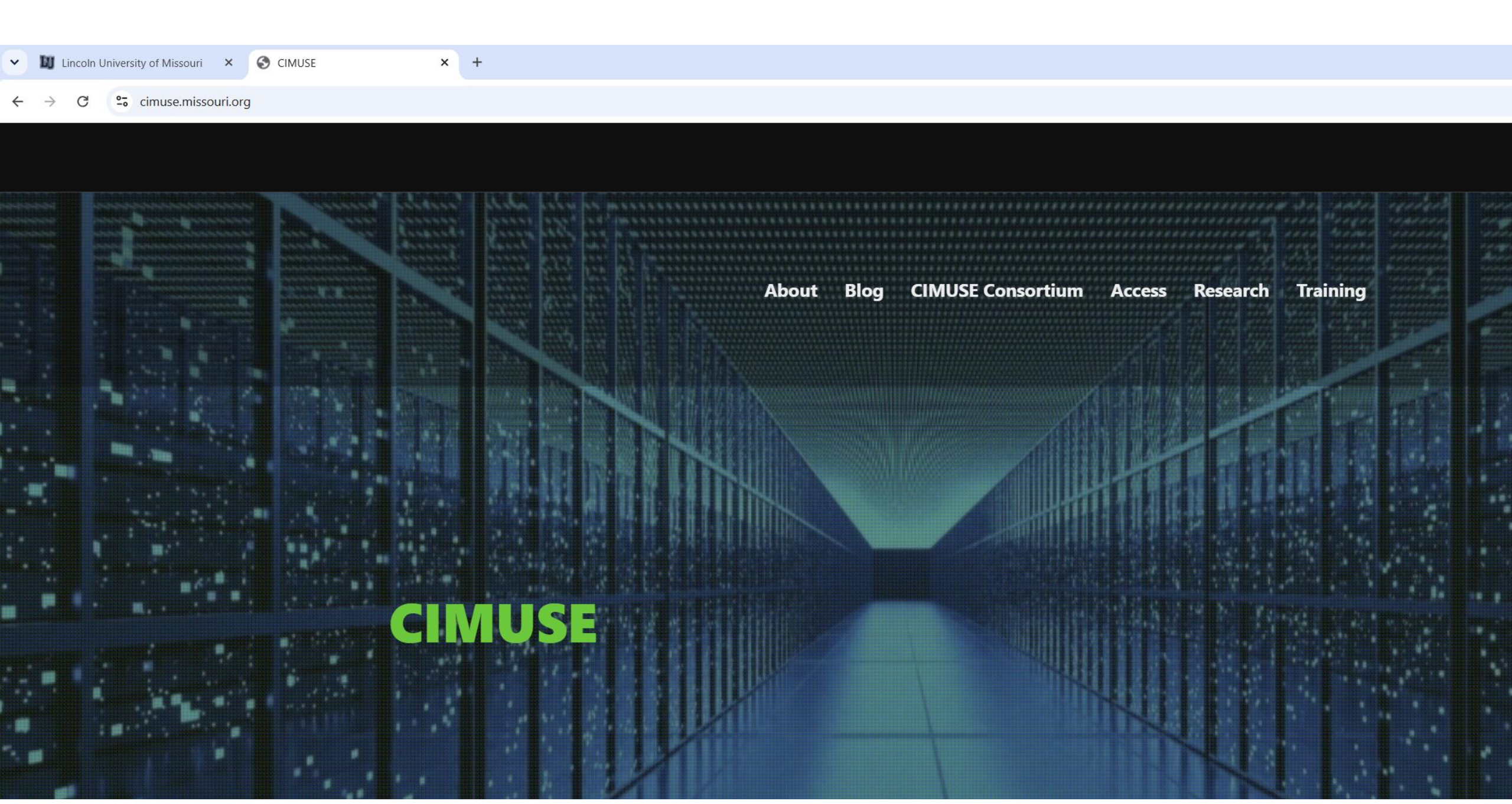
LINCOLN UNIVERSITY
MISSOURI
College of Agriculture

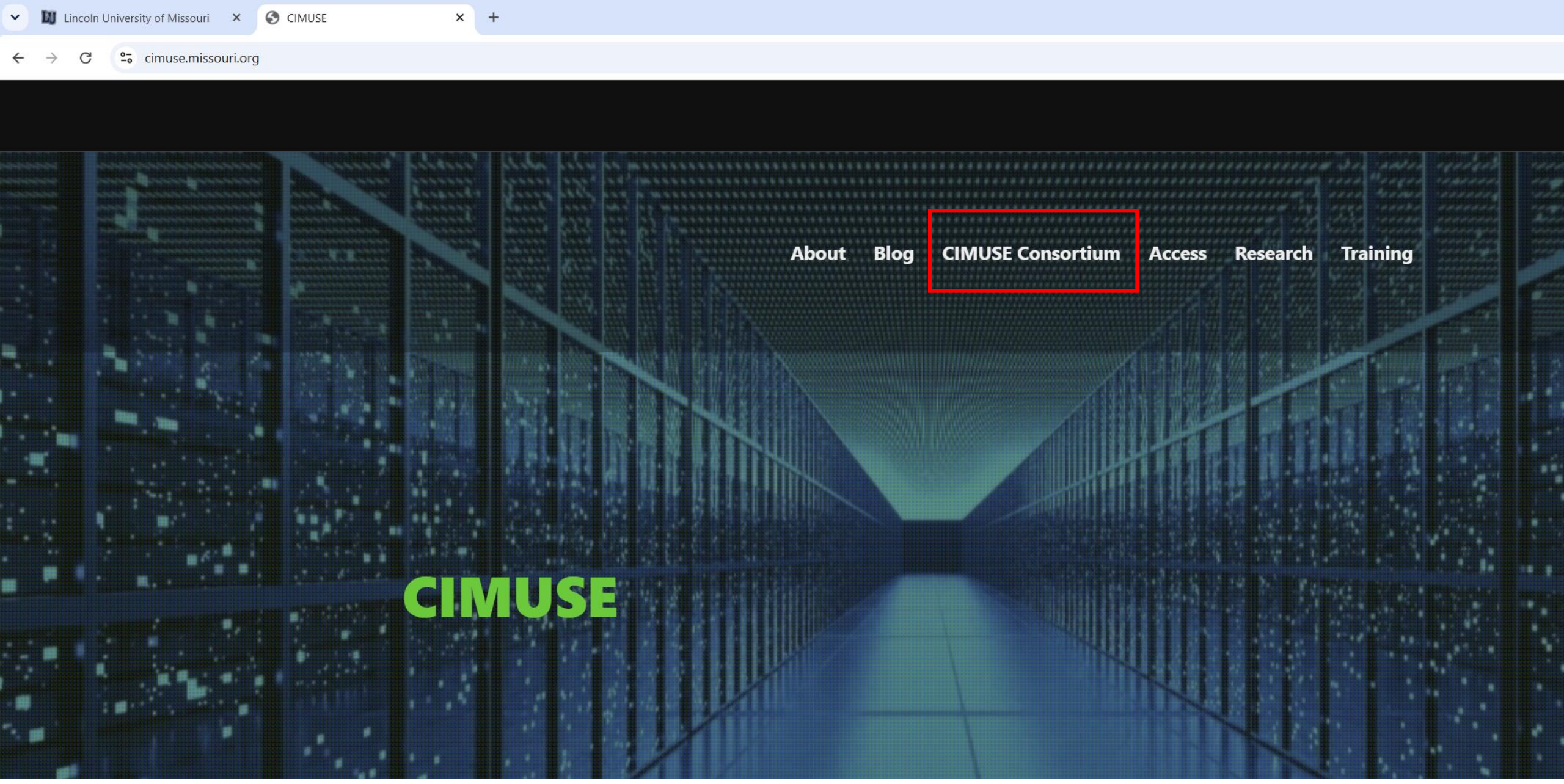
Laboratory of Genomic Dissection
Lincoln University of Missouri

www.LabOfGenomicDissection.com

Research question: Can I use genomic data to determine the breed composition of individuals in a crossbred sheep population using HPC resources?

www.ocj.com


Lincoln University of Missouri
CIMUSE
cimuse.missouri.org
About
Blog
CIMUSE Consortium
Access
Research
Training
CIMUSE


Lincoln University of Missouri
CIMUSE
cimuse.missouri.org
About
Blog
CIMUSE Consortium
Access
Research
Training
CIMUSE

Dr. Joel Leal Gutierrez

Professor of Animal Science

Lincoln University of Missouri

Email

University of Missouri Partners

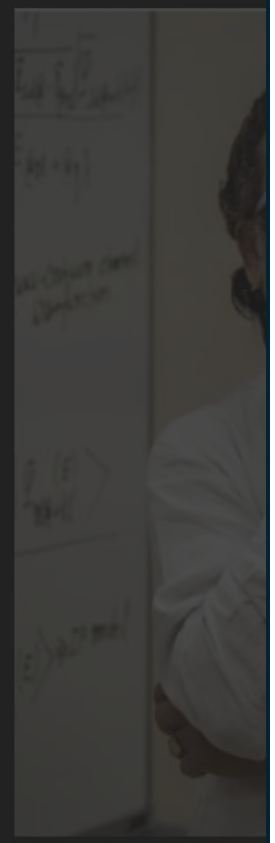
Mr. Matthew Keeler

Dr. Joel Leal Gutierrez

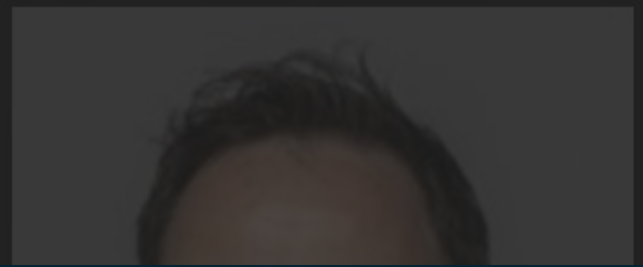
Professor of Animal Science

Lincoln University of Missouri

Email



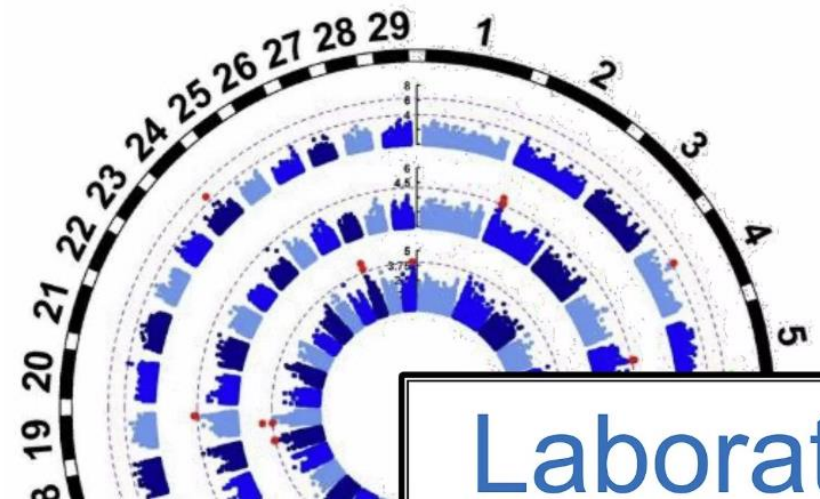
University of Missouri Partners



Mr. Matthew Keeler

www.labofgenomicdissection.com

COME GROW
with us
LINCOLN UNIVERSITY
MISSOURI



Laboratory of

Laboratory of

www.labofgenomicdissection.com

Wet laboratory protocols

Disclaimer: All protocols are based on the availability of resources in the Laboratory of Genomic Dissection.

- Agarose gel electrophoresis
- Blood collection
- Blue Juice 6X loading buffer
- DNA extraction Salting-out from blood
- Genomic DNA Cleaning & Concentration
- Primer re-nature and dilution
- RNA extraction using TRIzol
- RNase free glassware material
- Swab collection for vaginal microbial assay

Dry laboratory protocols

- Breed composition activity

www.labofgenomicdissection.com

Wet laboratory protocols

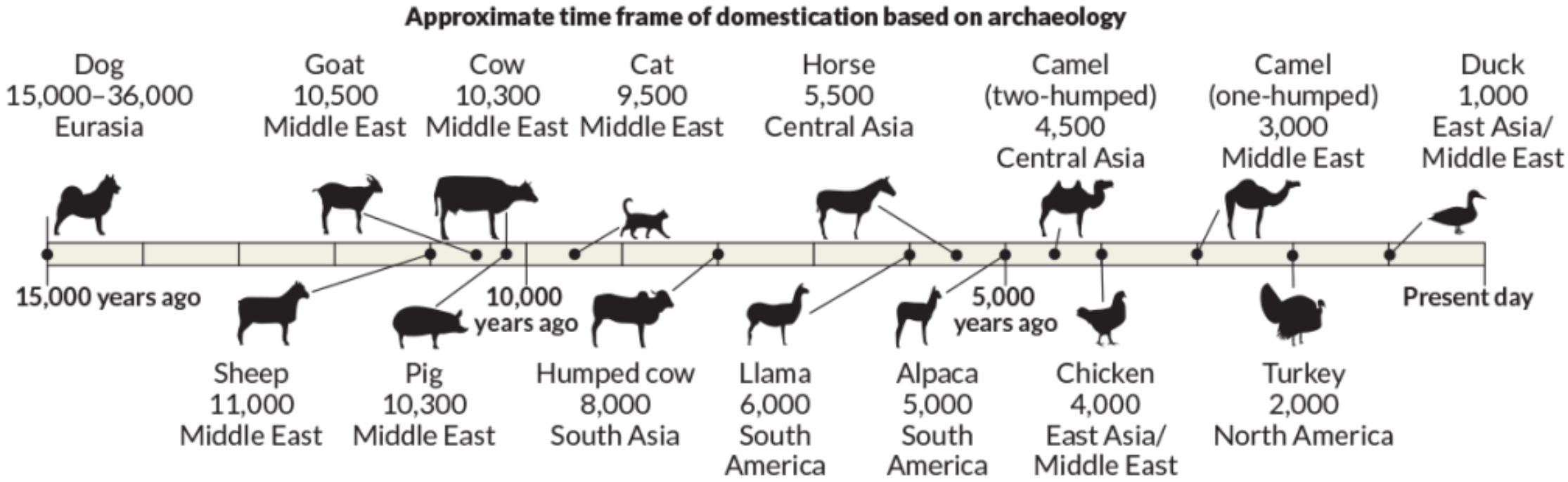
Disclaimer: All protocols are based on the availability of resources in the Laboratory of Genomic Dissection.

- Agarose gel electrophoresis
- Blood collection
- Blue Juice 6X loading buffer
- DNA extraction Salting-out from blood
- Genomic DNA Cleaning & Concentration
- Primer re-nature and dilution
- RNA extraction using TRIzol
- RNase free glassware material
- Swab collection for vaginal microbial assay

Dry laboratory protocols

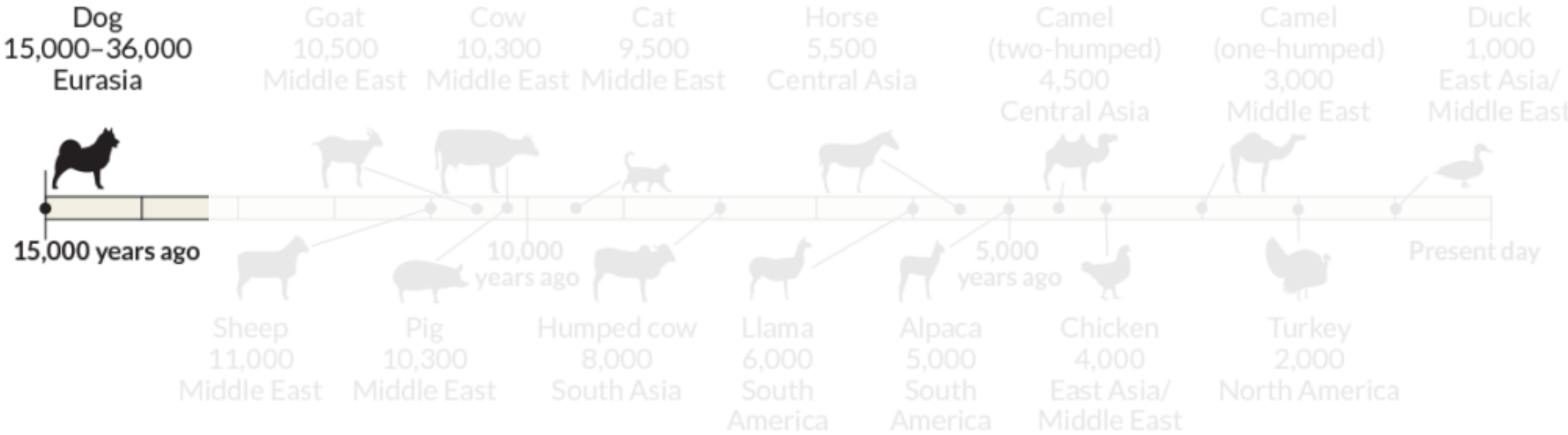
- Breed composition activity

Animal domestication



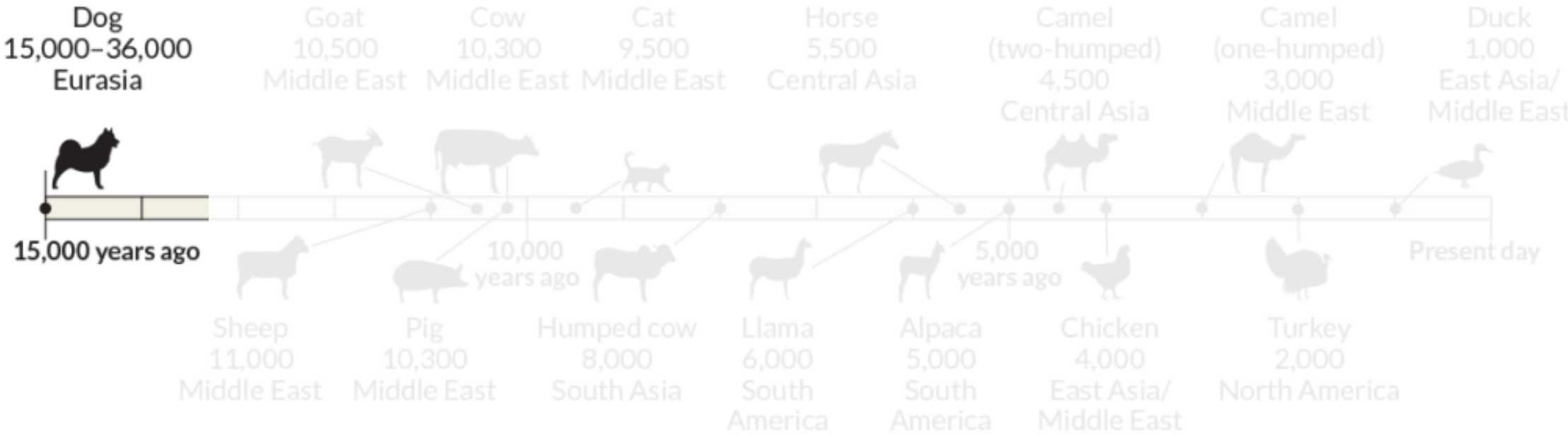
https://www.sciencenews.org/article/dna-evidence-rewriting-domestication-origin-stories

Approximate time frame of domestication based on archaeology

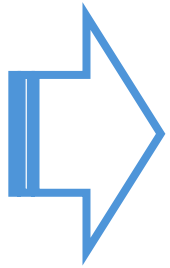


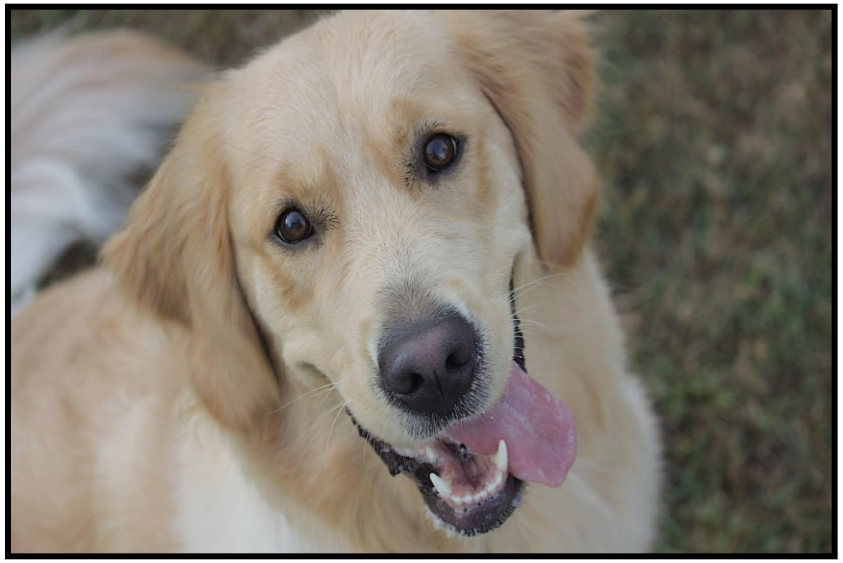
https://www.sciencenews.org/article/dna-evidence-rewriting-domestication-origin-stories

Approximate time frame of domestication based on archaeology

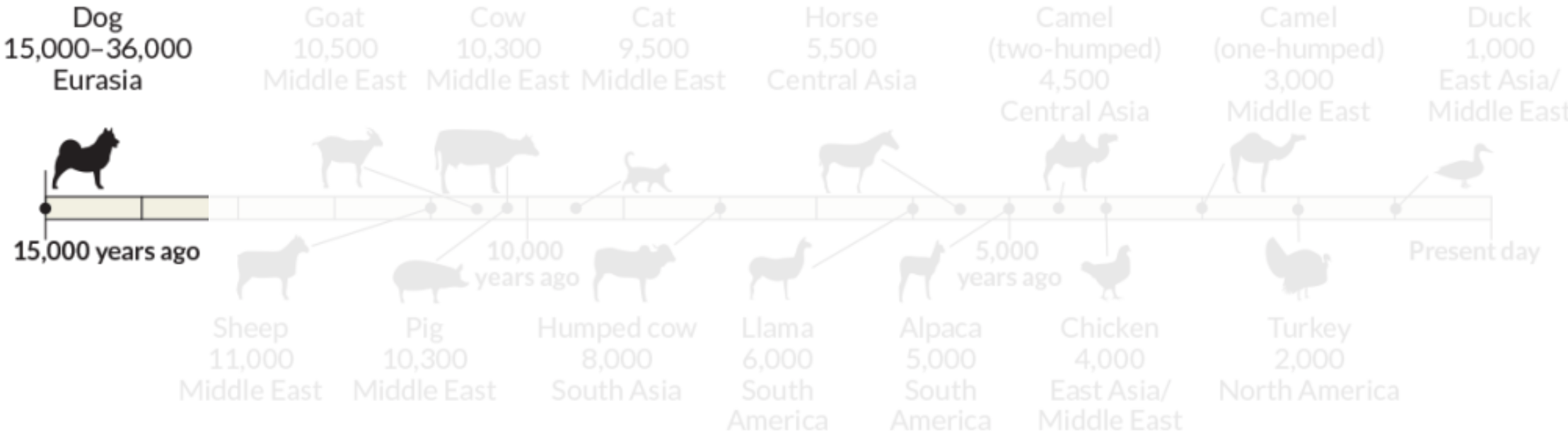


https://www.sciencenews.org/article/dna-evidence-rewriting-domestication-origin-stories



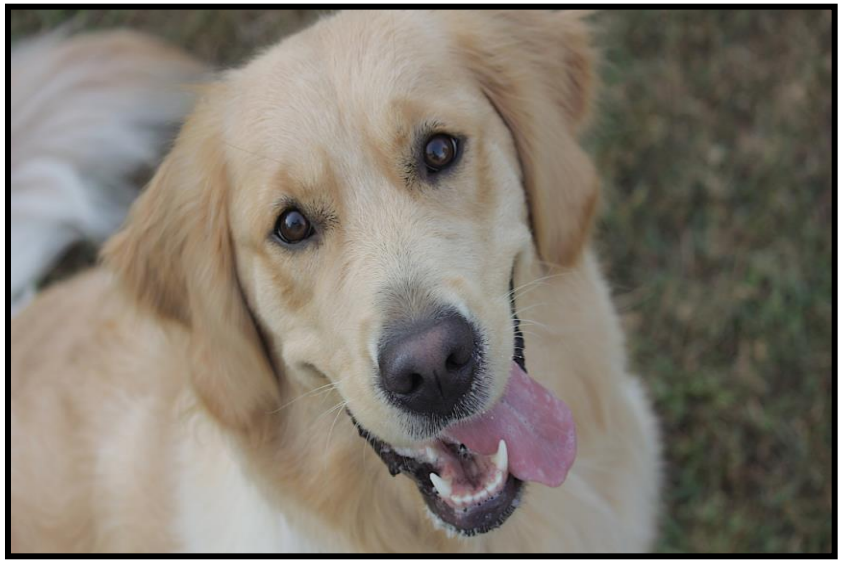


Approximate time frame of domestication based on archaeology

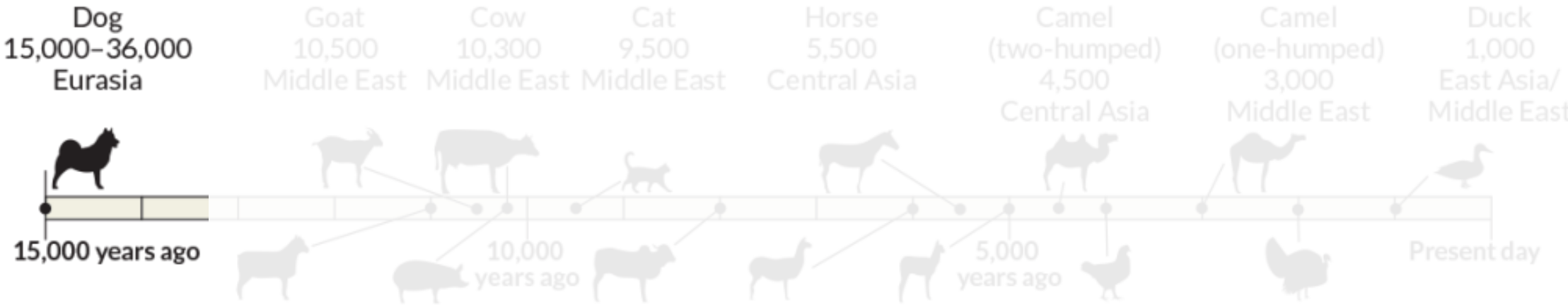


https://www.sciencenews.org/article/dna-evidence-rewriting-domestication-origin-stories

Phenotypic selection

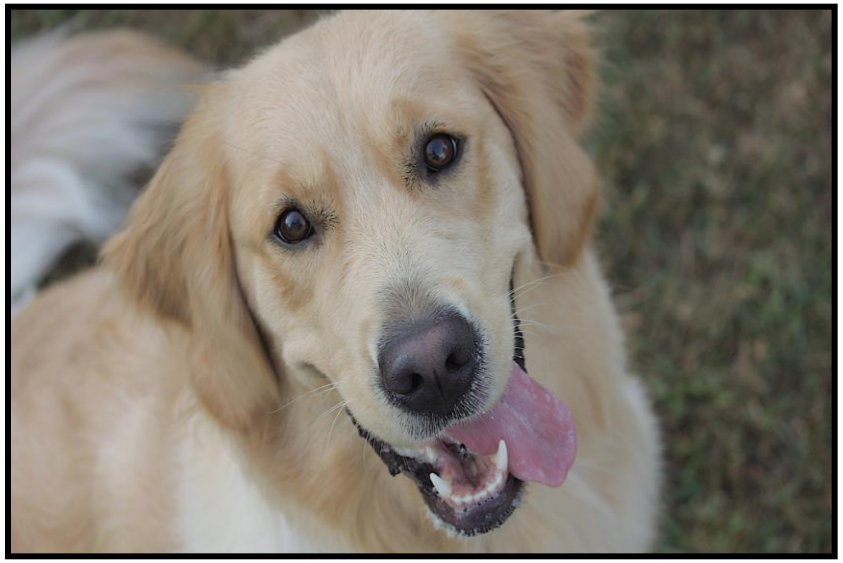


Approximate time frame of domestication based on archaeology



https://www.sciencenews.org/article/dna-evidence-rewriting-domestication-origin-stories

Genomic selection

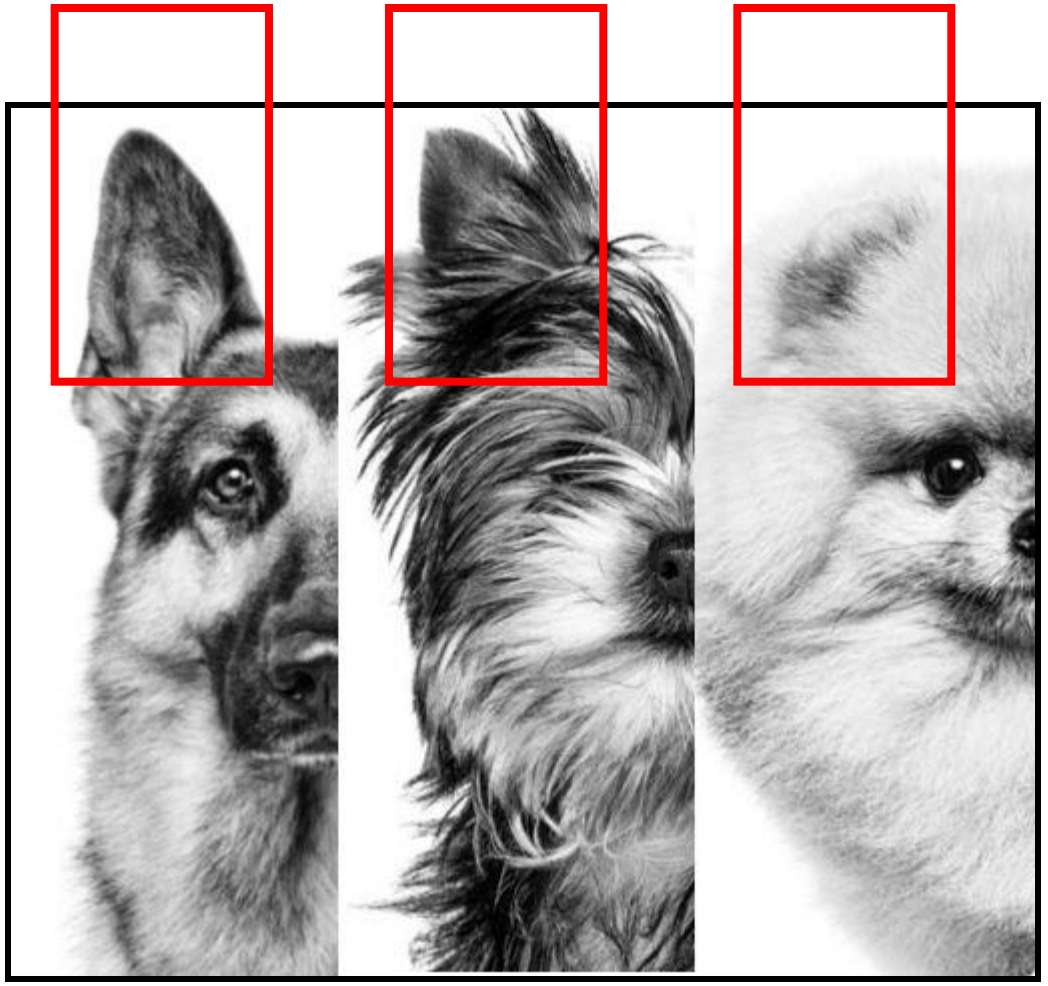


Genomic selection

Breeds

Genomic selection

Ear shape

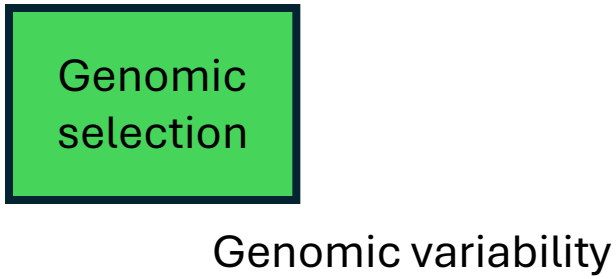


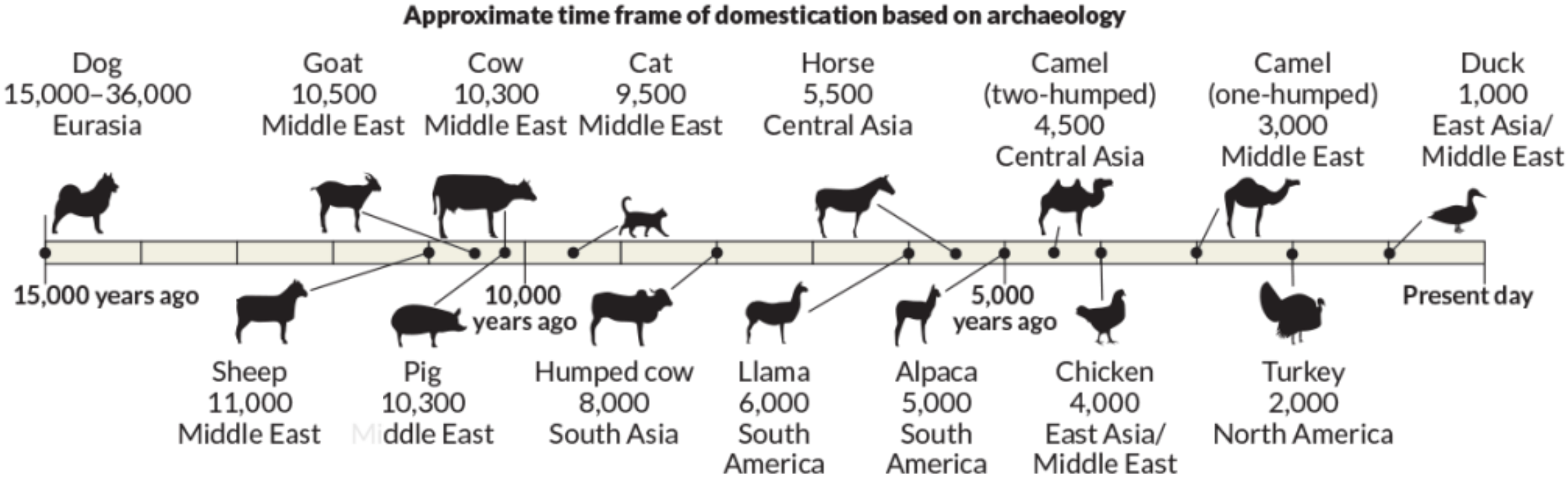
Breeds

pngtree.com, stacker.com, theconversation.com, www.purina.com

pngtree.com, stacker.com, theconversation.com, w

Genomic selection

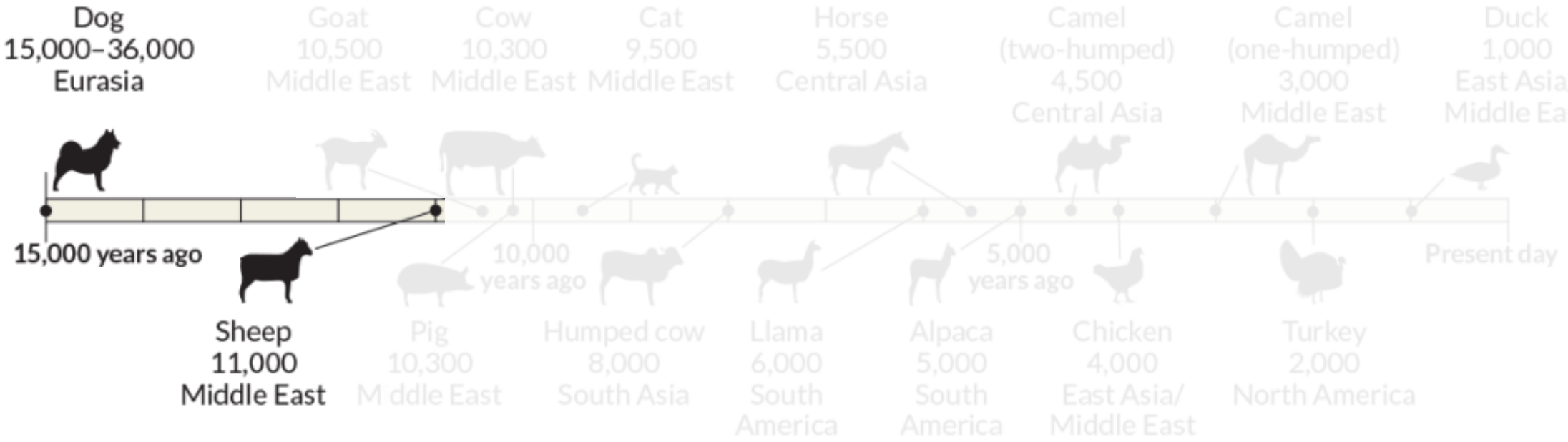

Genomic selection
Genomic variability



https://www.sciencenews.org/article/dna-evidence-rewriting-domestication-origin-stories

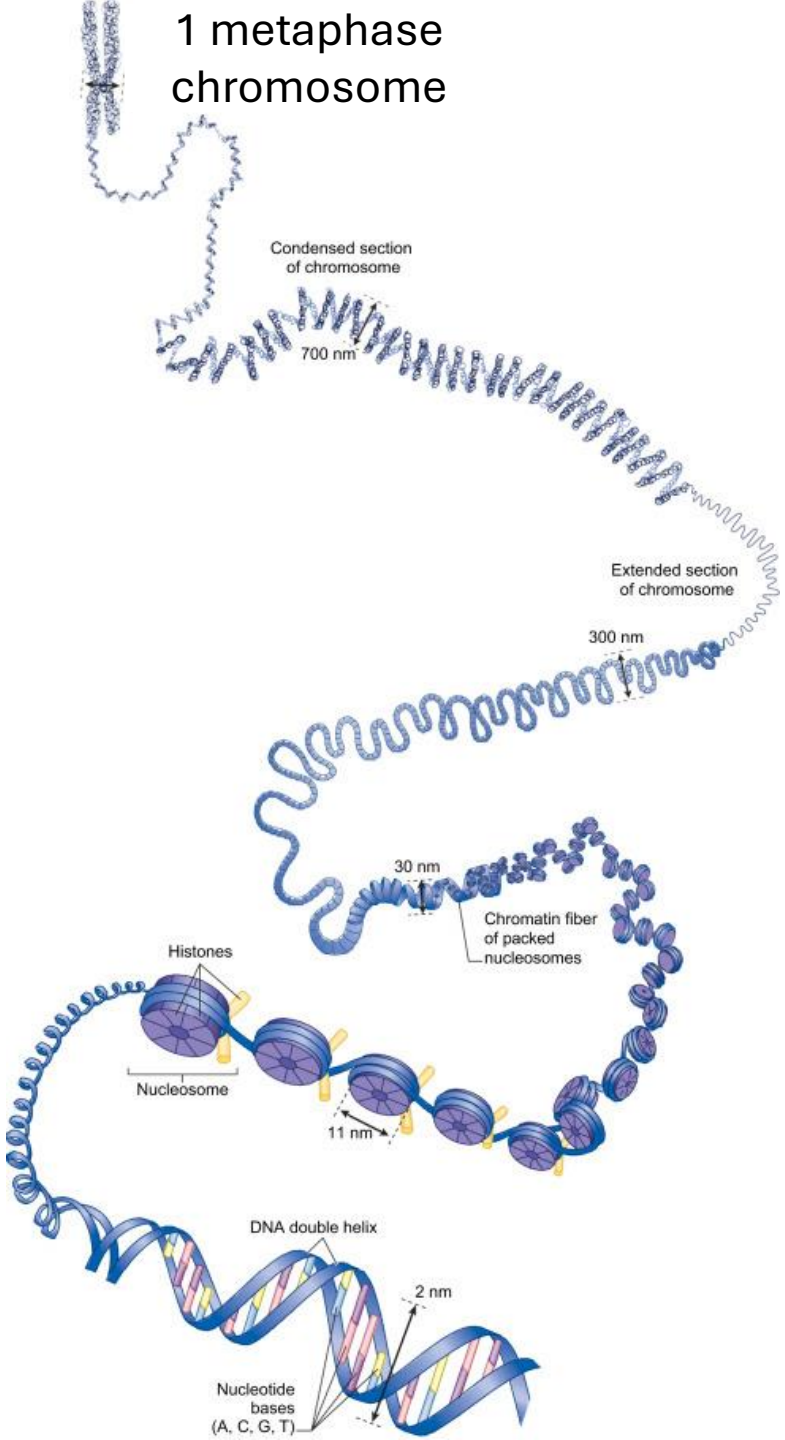
Asiatic wild mouflon (*Ovis gmelini*)

Approximate time frame of domestication based on archaeology



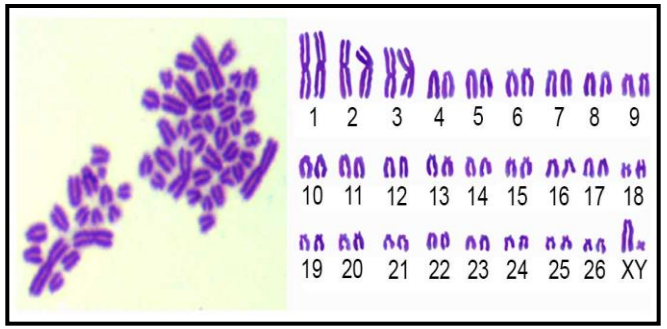
https://www.sciencenews.org/article/dna-evidence-rewriting-domestication-origin-stories

Asiatic wild mouflon (*Ovis gmelini*)



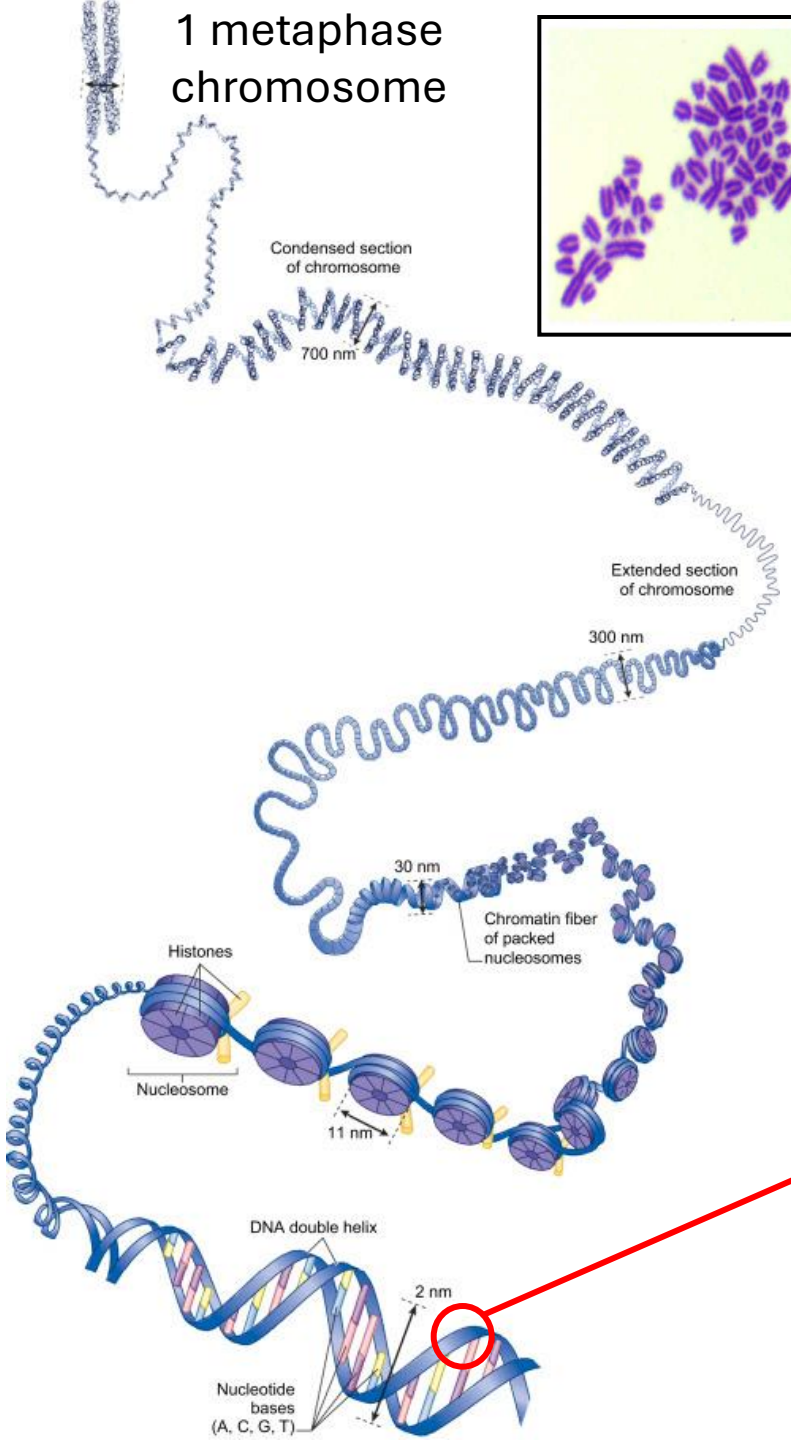
Genomic variability

1 metaphase chromosome



Ji et al., 2016

https://www.britannica.com/



4 nucleotides

a

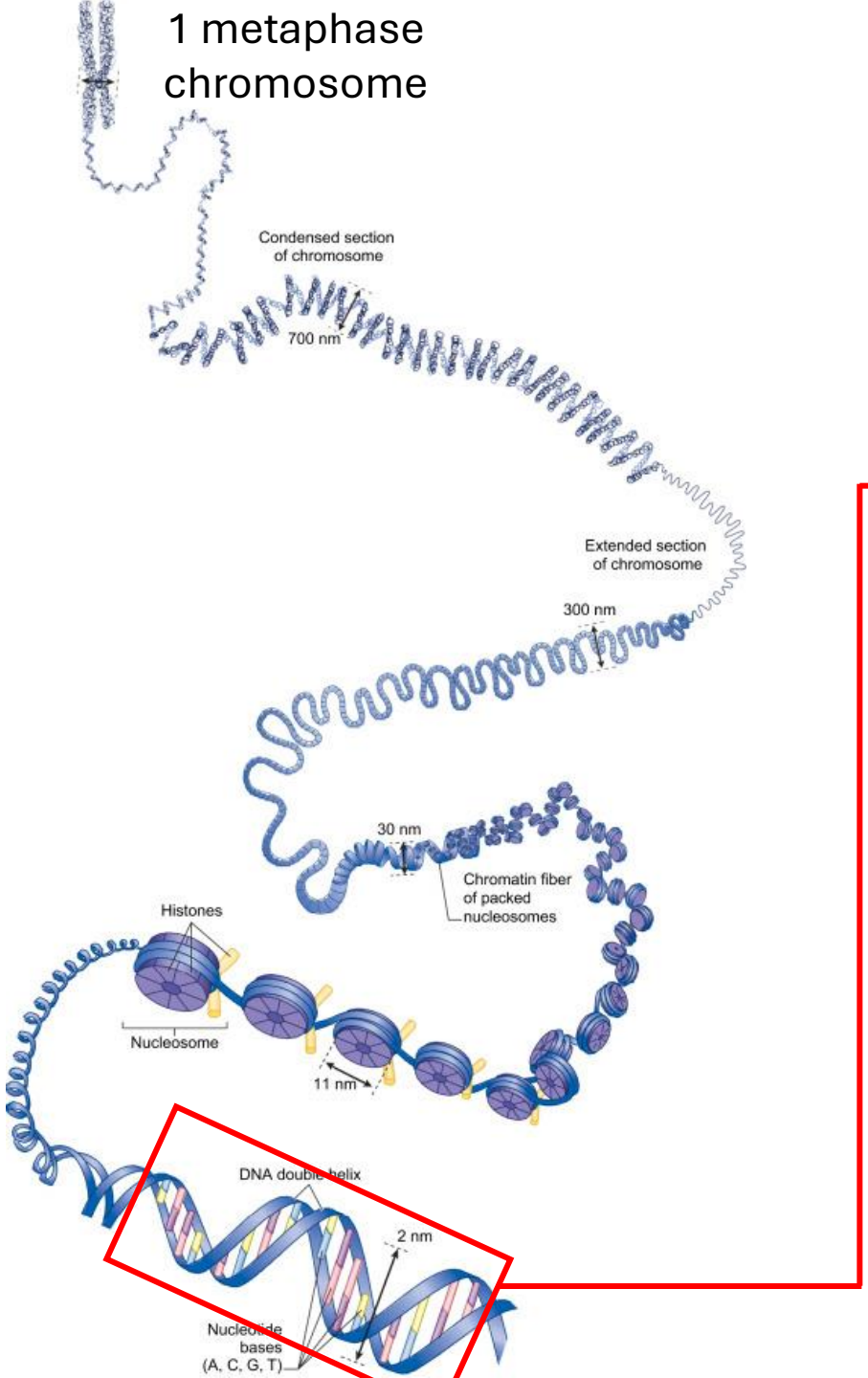
t

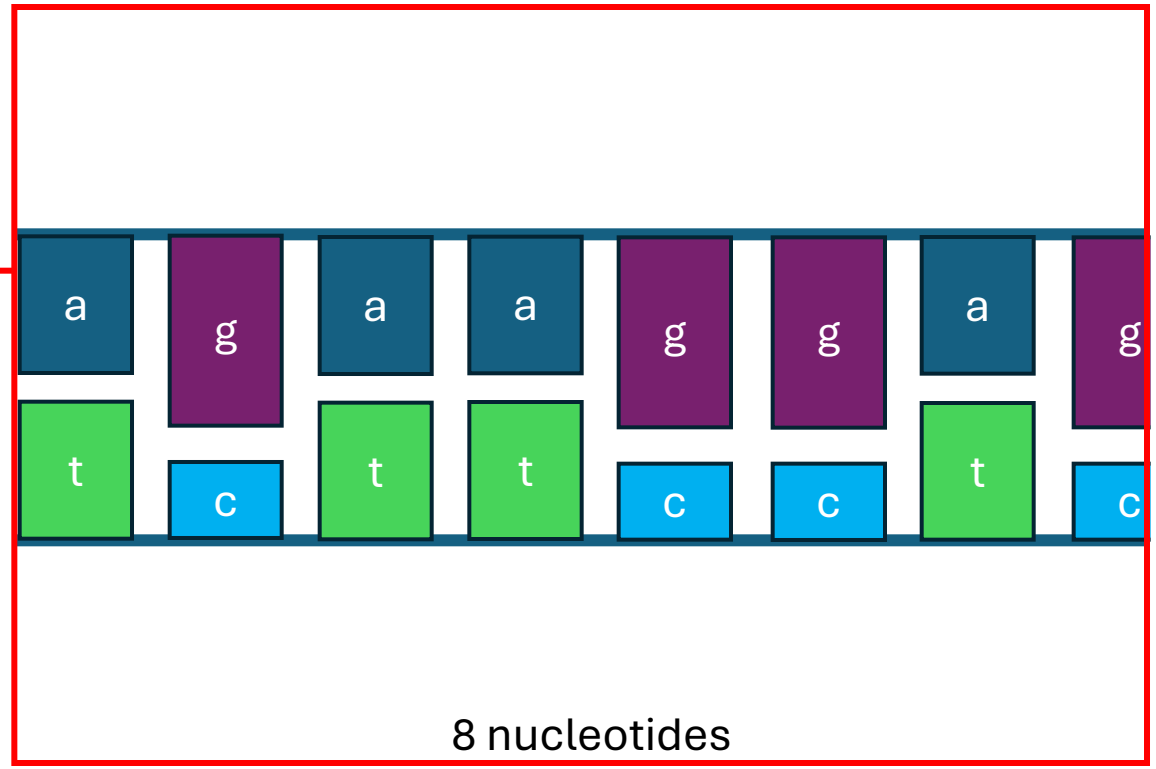
c

g

Genomic variability

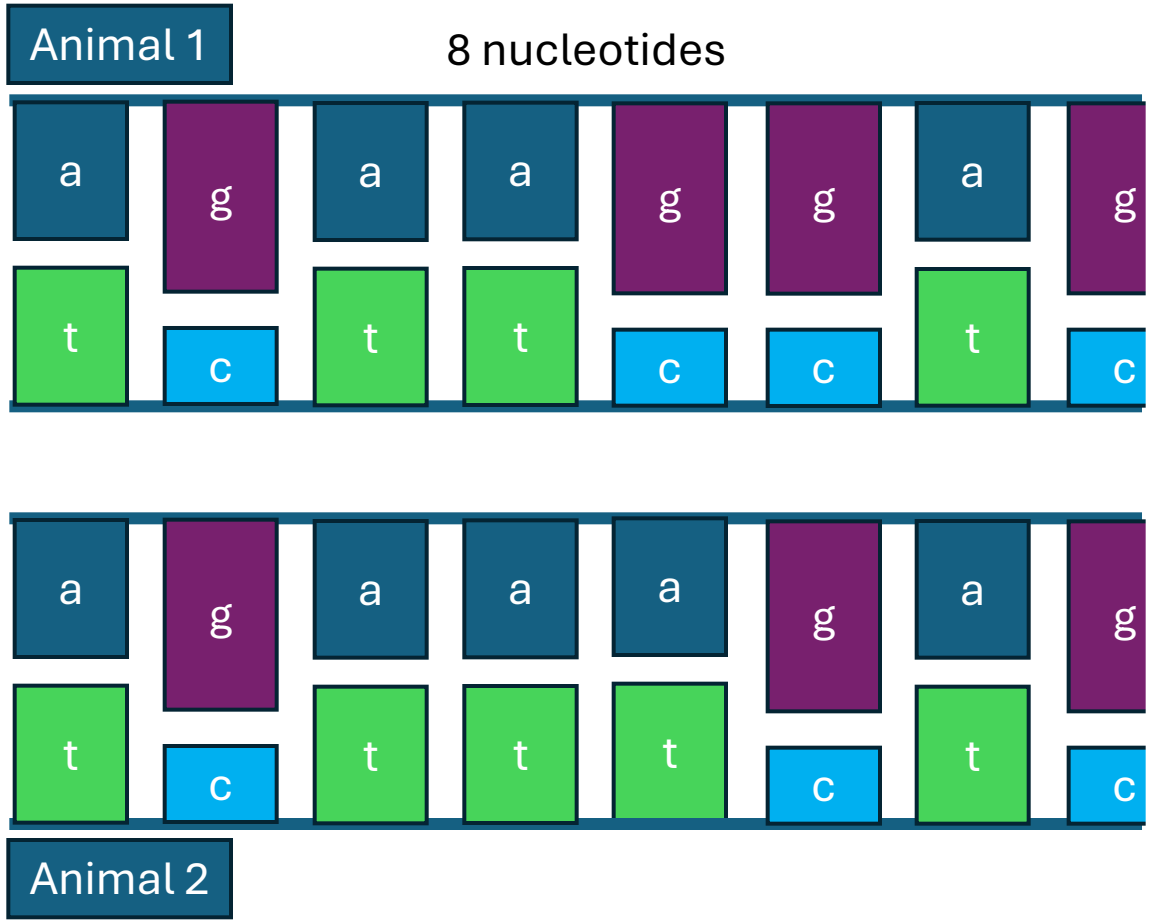
1 metaphase chromosome



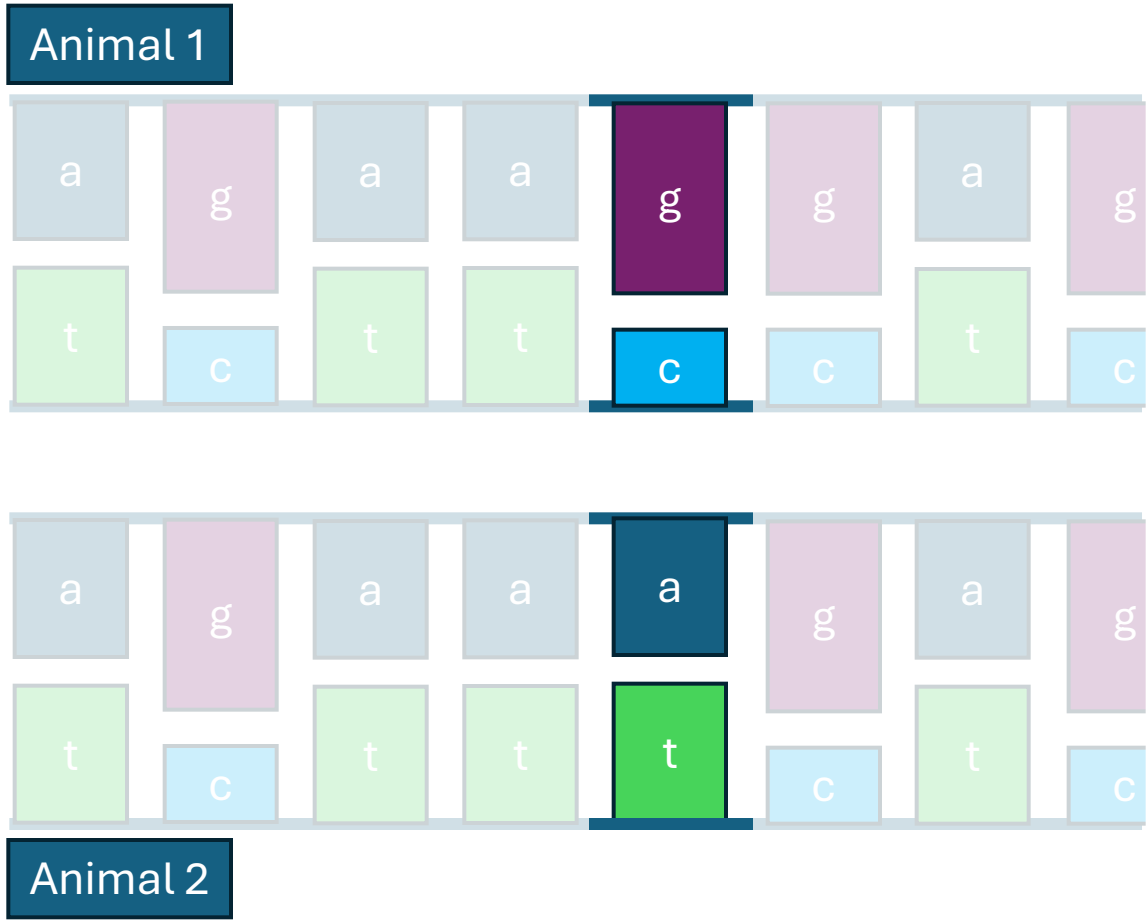


8 nucleotides

Individuals from
the same population

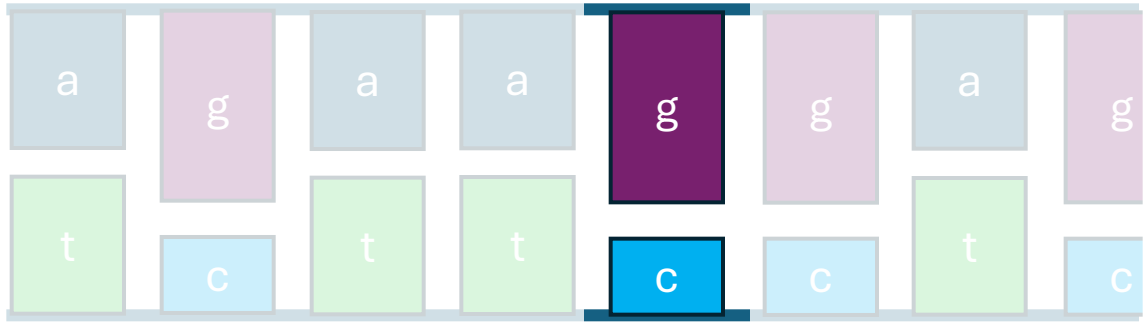


Individuals from
the same population

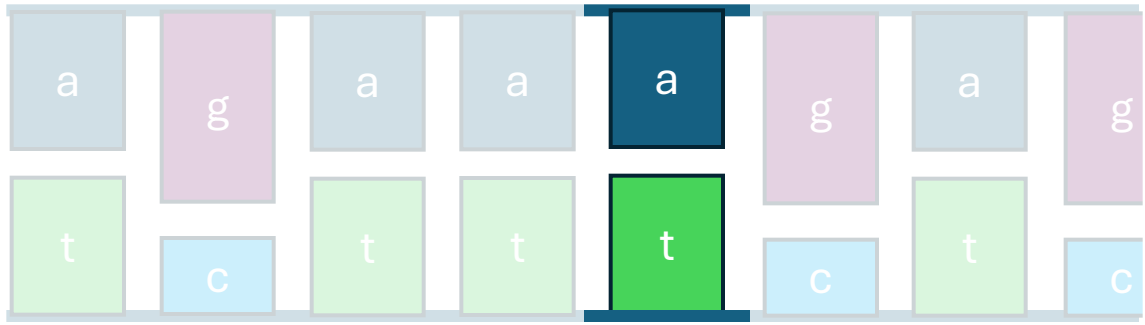


Change G > A in
a known position

Animal 1



Individuals from
the same population



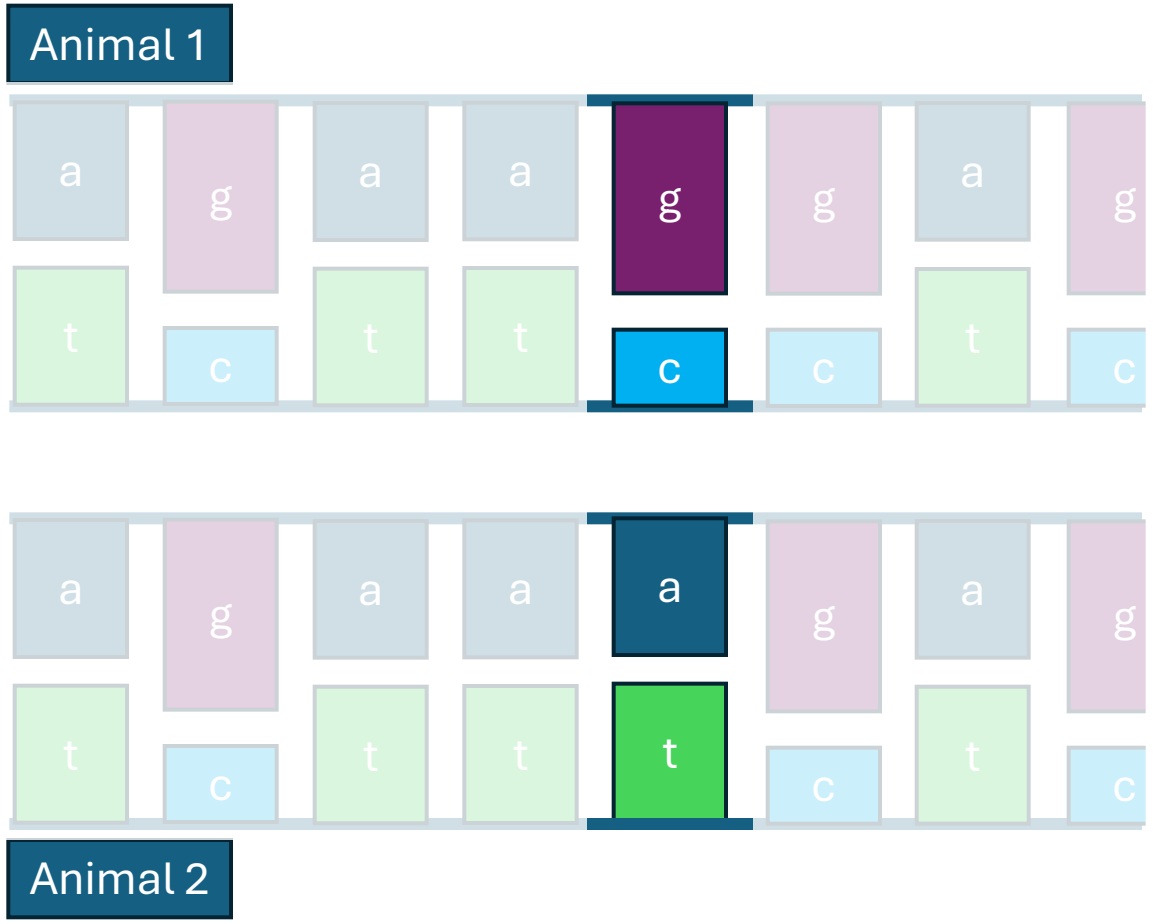
Animal 2

https://discover.hubpages.com/

Sheep genome: **2.5 billion bases**
distributed into 26 autosomes
and one sex chromosome pair

Individuals from
the same population

Change G > A in
a known position



https://discover.hubpages.com/

Sheep genome: **2.5 billion bases**
distributed into 26 autosomes
and one sex chromosome pair

Individuals from
the same population

https://discover.hubpages.com/

Sheep genome: **2.5 billion bases**
distributed into 26 autosomes
and one sex chromosome pair

Individuals from
the same population

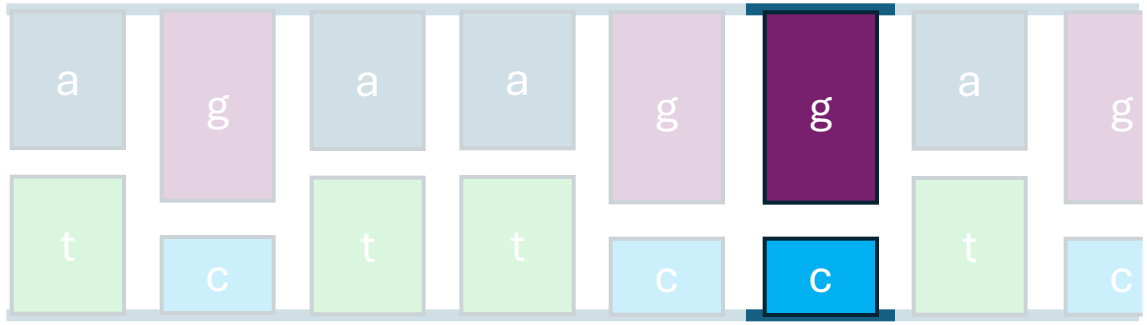
https://discover.hubpages.com/

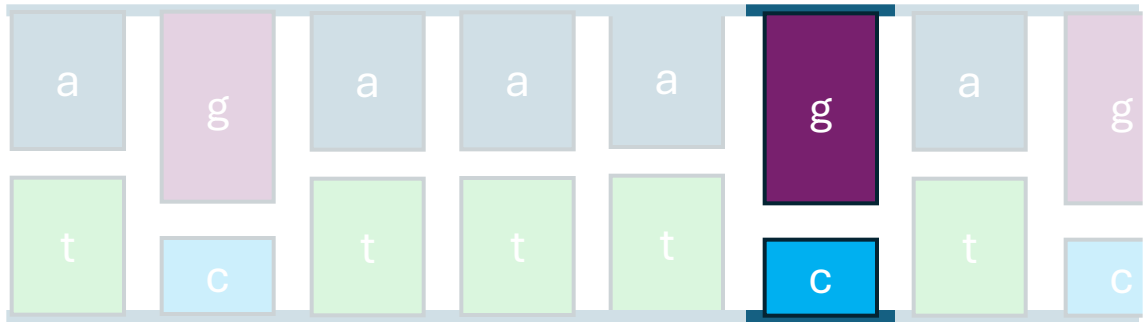
Sheep genome: **2.5 billion bases**
distributed into 26 autosomes
and one sex chromosome pair

Individuals from
the same population

Polymorphism?

Animal 1





Animal 2

Specialized sheep breeds for production

Specialized sheep breeds for production

Suffolk sheep

East Friesian sheep

Specialized sheep breeds for production

Suffolk sheep

East Friesian sheep

Specialized sheep breeds for production

Suffolk sheep

East Friesian sheep

Specialized sheep breeds for production

Suffolk sheep

East Friesian sheep

50:50

Specialized sheep breeds for production

Suffolk sheep

East Friesian sheep

Crossbred:

1. Higher productivity
2. Higher resistance
3. Higher prolificacy
4. Double purpose

Specialized sheep breeds for production

Suffolk sheep

East Friesian sheep

Crossbred:

1. Higher productivity
2. Higher resistance
3. Higher prolificacy
4. Double purpose

Breed composition

Files to be used:

Full folder:
https://figshare.com/s/9148a98043a49e06100e

Genotypic data filtered:
https://figshare.com/ndownloader/files/65762952?private_link=9148a98043a49e06100e

Breed composition data:
https://figshare.com/ndownloader/files/65762934?private_link=9148a98043a49e06100e

R code:
https://figshare.com/ndownloader/files/65762949?private_link=9148a98043a49e06100e

Provided files:

Name	Date modified	Type	Size
.Rhistory	6/19/2026 12:33 P...	RHISTORY File	8 KB
250KUF	6/18/2026 12:02 P...	vCalendar File	450,554 ...
breedComposition	6/19/2026 12:18 P...	Text Document	14 KB
breedComposition	6/19/2026 12:18 P...	Microsoft Exce...	203 KB
genotipicData	6/18/2026 2:04 PM	Text Document	217,225 ...
genotipicDataFiltered	6/19/2026 10:53 A...	Text Document	176,027 ...
genotipicDataRecoded	6/18/2026 1:23 PM	Text Document	216,730 ...
PCA_2PCs	6/19/2026 12:32 P...	Text Document	33 KB
PCA_Breed_Merged	6/19/2026 12:32 P...	Text Document	42 KB
populationStructure	6/19/2026 12:14 P...	PNG File	20 KB
populationStructureCode	6/22/2026 10:03 A...	R File	3 KB

Provided files:

Name	Date modified	Type	Size
.Rhistory	6/19/2026 12:33 P...	RHISTORY File	8 KB
250KUF	6/18/2026 12:02 P...	vCalendar File	450,554 ...
breedComposition	6/19/2026 12:18 P...	Text Document	14 KB
breedComposition	6/19/2026 12:18 P...	Microsoft Exce...	203 KB
genotipicData	6/18/2026 2:04 PM	Text Document	217,225 ...
genotipicDataFiltered	6/19/2026 10:53 A...	Text Document	176,027 ...
genotipicDataRecoded	6/18/2026 1:23 PM	Text Document	216,730 ...
PCA_2PCs	6/19/2026 12:32 P...	Text Document	33 KB
PCA_Breed_Merged	6/19/2026 12:32 P...	Text Document	42 KB
populationStructure	6/19/2026 12:14 P...	PNG File	20 KB
populationStructureCode	6/22/2026 10:03 A...	R File	3 KB

Raw genotypic data per animal	**2,500,000**

Number of animals	**700**

Provided files:

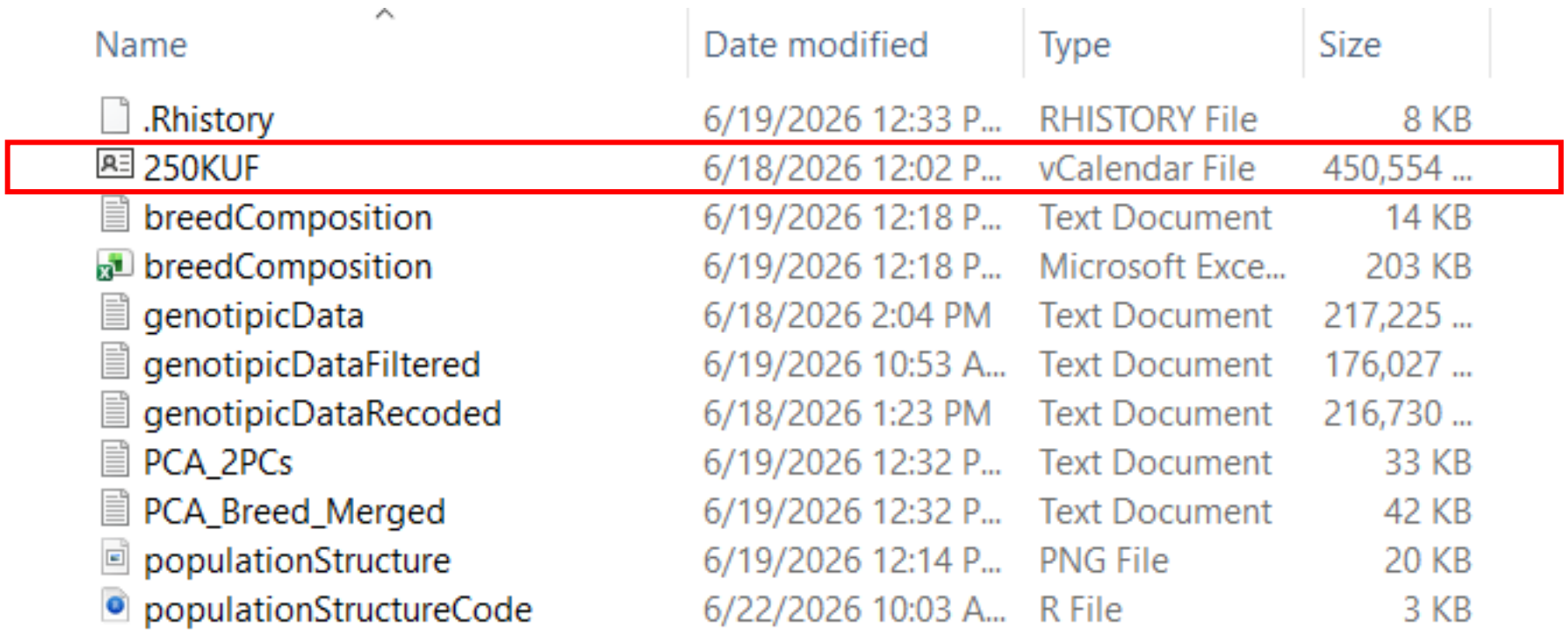
Name	Date modified	Type	Size
.Rhistory	6/19/2026 12:33 P...	RHISTORY File	8 KB
250KUF	6/18/2026 12:02 P...	vCalendar File	450,554 ...
breedComposition	6/19/2026 12:18 P...	Text Document	14 KB
breedComposition	6/19/2026 12:18 P...	Microsoft Exce...	203 KB
genotipicData	6/18/2026 2:04 PM	Text Document	217,225 ...
genotipicDataFiltered	6/19/2026 10:53 A...	Text Document	176,027 ...
genotipicDataRecoded	6/18/2026 1:23 PM	Text Document	216,730 ...
PCA_2PCs	6/19/2026 12:32 P...	Text Document	33 KB
PCA_Breed_Merged	6/19/2026 12:32 P...	Text Document	42 KB
populationStructure	6/19/2026 12:14 P...	PNG File	20 KB
populationStructureCode	6/22/2026 10:03 A...	R File	3 KB

Raw genotypic data per animal: **2,500,000**

Number of animals: **700**

Total data points: **1,750,000,000**

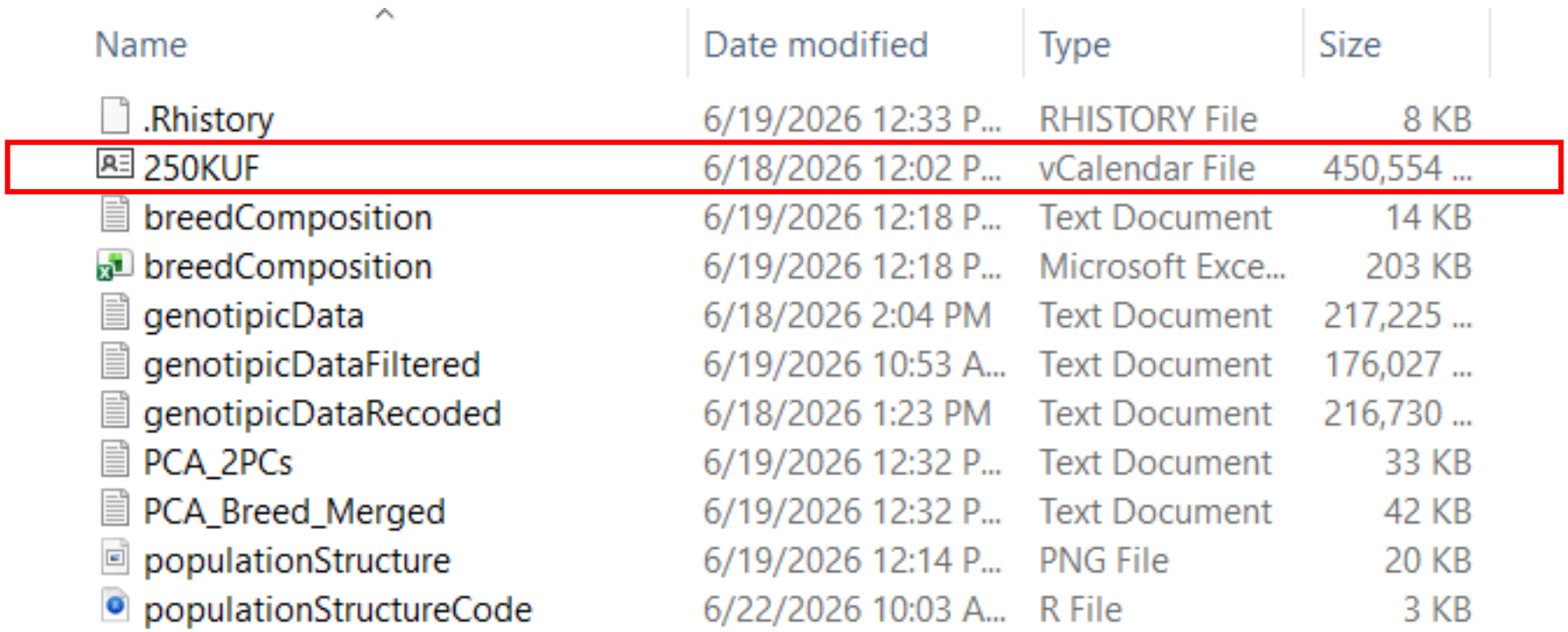
Provided files:



Name	Date modified	Type	Size
.Rhistory	6/19/2026 12:33 P...	RHISTORY File	8 KB
250KUF	6/18/2026 12:02 P...	vCalendar File	450,554 ...
breedComposition	6/19/2026 12:18 P...	Text Document	14 KB
breedComposition	6/19/2026 12:18 P...	Microsoft Exce...	203 KB
genotipicData	6/18/2026 2:04 PM	Text Document	217,225 ...
genotipicDataFiltered	6/19/2026 10:53 A...	Text Document	176,027 ...
genotipicDataRecoded	6/18/2026 1:23 PM	Text Document	216,730 ...
PCA_2PCs	6/19/2026 12:32 P...	Text Document	33 KB
PCA_Breed_Merged	6/19/2026 12:32 P...	Text Document	42 KB
populationStructure	6/19/2026 12:14 P...	PNG File	20 KB
populationStructureCode	6/22/2026 10:03 A...	R File	3 KB

```
#CHROM→POS→ID→REF→ALT→QUAL→FILTER→INFO→FORMAT→2130309→2130326→2130336→2130350→2130352→2130356→2130358→3109369→3119025→3130004→3130012→31
1→19774→rs136249228→A→G→.→.→ALL=[T/C];IL=BOT;CU=TOP;NO=50→GT→0/1→./.→0/1→0/1→./.→0/1→0/1→./.→0/1→0/0→./.→0/1→0/1→./.→0/1→0/1→0/0→0/
1→19832→rs459659610→C→T→.→.→ALL=[T/C];IL=BOT;CU=BOT;NO=50→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/1→0/
1→70600→rs454137042→T→C→.→.→ALL=[T/C];IL=BOT;CU=BOT;NO=-54525902→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/
1→123681→rs209484914→A→C→.→.→ALL=[A/C];IL=TOP;CU=TOP;NO=-427753422→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/1→0/0→0/0→0/1→0/0→0/0→0/0→0/0→0/0→0/
1→123782→rs109505697→T→C→.→.→ALL=[T/C];IL=BOT;CU=BOT;NO=-427753422→GT→1/0→1/0→1/0→1/0→1/0→1/1→1/1→1/0→1/1→1/0→0/0→0/0→0/0→1/0→0/0→1/
1→124921→rs457956940→T→C→.→.→ALL=[T/C];IL=BOT;CU=BOT;NO=50→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/
```

Provided files:



Name	Date modified	Type	Size
.Rhistory	6/19/2026 12:33 P...	RHISTORY File	8 KB
250KUF	6/18/2026 12:02 P...	vCalendar File	450,554 ...
breedComposition	6/19/2026 12:18 P...	Text Document	14 KB
breedComposition	6/19/2026 12:18 P...	Microsoft Exce...	203 KB
genotipicData	6/18/2026 2:04 PM	Text Document	217,225 ...
genotipicDataFiltered	6/19/2026 10:53 A...	Text Document	176,027 ...
genotipicDataRecoded	6/18/2026 1:23 PM	Text Document	216,730 ...
PCA_2PCs	6/19/2026 12:32 P...	Text Document	33 KB
PCA_Breed_Merged	6/19/2026 12:32 P...	Text Document	42 KB
populationStructure	6/19/2026 12:14 P...	PNG File	20 KB
populationStructureCode	6/22/2026 10:03 A...	R File	3 KB

Animal

```
#CHROM→POS→ID→REF→ALT→QUAL→FILTER→INFO→FORMAT→2130309→2130326→2130336→2130350→2130352→2130356→2130358→3109369→3119025→3130004→3130012→31
1→19774→rs136249228→A→G→.→.→ALL=[T/C];IL=BOT;CU=TOP;NO=50→GT→0/1→./.→0/1→0/1→./.→0/1→0/1→./.→0/1→0/0→./.→0/1→0/1→./.→0/1→0/1→0/0→0/
1→19832→rs459659610→C→T→.→.→ALL=[T/C];IL=BOT;CU=BOT;NO=50→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/1→0/
1→70600→rs454137042→T→C→.→.→ALL=[T/C];IL=BOT;CU=BOT;NO=-54525902→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/
1→123681→rs209484914→A→C→.→.→ALL=[A/C];IL=TOP;CU=TOP;NO=-427753422→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/1→0/0→0/0→0/1→0/0→0/0→0/0→0/0→0/0→0/0→0/
1→123782→rs109505697→T→C→.→.→ALL=[T/C];IL=BOT;CU=BOT;NO=-427753422→GT→1/0→1/0→1/0→1/0→1/0→1/1→1/1→1/0→1/1→1/0→0/0→0/0→0/0→1/0→0/0→1/
1→124921→rs457956940→T→C→.→.→ALL=[T/C];IL=BOT;CU=BOT;NO=50→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/
```

Provided files:

Name	Date modified	Type	Size
.Rhistory	6/19/2026 12:33 P...	RHISTORY File	8 KB
250KUF	6/18/2026 12:02 P...	vCalendar File	450,554 ...
breedComposition	6/19/2026 12:18 P...	Text Document	14 KB
breedComposition	6/19/2026 12:18 P...	Microsoft Exce...	203 KB
genotipicData	6/18/2026 2:04 PM	Text Document	217,225 ...
genotipicDataFiltered	6/19/2026 10:53 A...	Text Document	176,027 ...
genotipicDataRecoded	6/18/2026 1:23 PM	Text Document	216,730 ...
PCA_2PCs	6/19/2026 12:32 P...	Text Document	33 KB
PCA_Breed_Merged	6/19/2026 12:32 P...	Text Document	42 KB
populationStructure	6/19/2026 12:14 P...	PNG File	20 KB
populationStructureCode	6/22/2026 10:03 A...	R File	3 KB

Polymorphism

```
#CHROM→POS→ID→REF→ALT→QUAL→FILTER→INFO→FORMAT→2130309→2130326→2130336→2130350→2130352→2130356→2130358→3109369→3119025→3130004→3130012→31
1→19774→rs136249228→A→G→.→.→ALL=[T/C];IL=BOT;CU=TOP;NO=50→GT→0/1→./.→0/1→0/1→./.→0/1→0/1→./.→0/1→0/0→./.→0/1→0/1→./.→0/1→0/1→0/0→0/
1→19832→rs459659610→C→T→.→.→ALL=[T/C];IL=BOT;CU=BOT;NO=50→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/1→0/
1→70600→rs454137042→T→C→.→.→ALL=[T/C];IL=BOT;CU=BOT;NO=-54525902→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/
1→123681→rs209484914→A→C→.→.→ALL=[A/C];IL=TOP;CU=TOP;NO=-427753422→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/1→0/0→0/0→0/1→0/0→0/0→0/0→0/0→0/0→0/
1→123782→rs109505697→T→C→.→.→ALL=[T/C];IL=BOT;CU=BOT;NO=-427753422→GT→1/0→1/0→1/0→1/0→1/0→1/1→1/1→1/0→1/1→1/0→0/0→0/0→0/0→1/0→0/0→1/
1→124921→rs457956940→T→C→.→.→ALL=[T/C];IL=BOT;CU=BOT;NO=50→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/
```

Provided files:

Name	Date modified	Type	Size
.Rhistory	6/19/2026 12:33 P...	RHISTORY File	8 KB
250KUF	6/18/2026 12:02 P...	vCalendar File	450,554 ...
breedComposition	6/19/2026 12:18 P...	Text Document	14 KB
breedComposition	6/19/2026 12:18 P...	Microsoft Exce...	203 KB
genotipicData	6/18/2026 2:04 PM	Text Document	217,225 ...
genotipicDataFiltered	6/19/2026 10:53 A...	Text Document	176,027 ...
genotipicDataRecoded	6/18/2026 1:23 PM	Text Document	216,730 ...
PCA_2PCs	6/19/2026 12:32 P...	Text Document	33 KB
PCA_Breed_Merged	6/19/2026 12:32 P...	Text Document	42 KB
populationStructure	6/19/2026 12:14 P...	PNG File	20 KB
populationStructureCode	6/22/2026 10:03 A...	R File	3 KB

Genotype

```
#CHROM→POS→ID→REF→ALT→QUAL→FILTER→INFO→FORMAT→2130309→2130326→2130336→2130350→2130352→2130356→2130358→3109369→3119025→3130004→3130012→31
1→19774→rs136249228→A→G→.→.→ALL=[T/C];IL=BOT;CU=TOP;NO=50→GT→0/1→./.→0/1→0/1→./.→0/1→0/1→./.→0/1→0/0→./.→0/1→0/1→./.→0/1→0/1→0/0→0/
1→19832→rs459659610→C→T→.→.→ALL=[T/C];IL=BOT;CU=BOT;NO=50→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/1→0/
1→70600→rs454137042→T→C→.→.→ALL=[T/C];IL=BOT;CU=BOT;NO=-54525902→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/
1→123681→rs209484914→A→C→.→.→ALL=[A/C];IL=TOP;CU=TOP;NO=-427753422→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/1→0/0→0/0→0/1→0/0→0/0→0/0→0/0→0/0→0/0→0/
1→123782→rs109505697→T→C→.→.→ALL=[T/C];IL=BOT;CU=BOT;NO=-427753422→GT→1/0→1/0→1/0→1/0→1/0→1/1→1/1→1/0→1/1→1/0→0/0→0/0→0/0→1/0→0/0→1/
1→124921→rs457956940→T→C→.→.→ALL=[T/C];IL=BOT;CU=BOT;NO=50→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/
```

Provided files:

rs1366249228: A/G

Name	Date modified	Type	Size
.Rhistory	6/19/2026 12:33 P...	RHISTORY File	8 KB
250KUF	6/18/2026 12:02 P...	vCalendar File	450,554 ...
breedComposition	6/19/2026 12:18 P...	Text Document	14 KB
breedComposition	6/19/2026 12:18 P...	Microsoft Exce...	203 KB
genotipicData	6/18/2026 2:04 PM	Text Document	217,225 ...
genotipicDataFiltered	6/19/2026 10:53 A...	Text Document	176,027 ...
genotipicDataRecoded	6/18/2026 1:23 PM	Text Document	216,730 ...
PCA_2PCs	6/19/2026 12:32 P...	Text Document	33 KB
PCA_Breed_Merged	6/19/2026 12:32 P...	Text Document	42 KB
populationStructure	6/19/2026 12:14 P...	PNG File	20 KB
populationStructureCode	6/22/2026 10:03 A...	R File	3 KB

Genotype

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	2130309	2130326	2130336	2130350	2130352	2130356	2130358	3109369	3119025	3130004	3130012	31
1	19774	rs136249228	A	G	.	.	ALL=[T/C];IL=BOT;CU=TOP;NO=50	GT	0/1	./.	0/1	0/1	./.	0/1	0/1	./.	0/1	0/0	./.	0/1	0/1	./.	0/1	0/1	0/0	0/
1	19832	rs459659610	C	T	.	.	ALL=[T/C];IL=BOT;CU=BOT;NO=50	GT	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/
1	70600	rs454137042	T	C	.	.	ALL=[T/C];IL=BOT;CU=BOT;NO=-54525902	GT	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/
1	123681	rs209484914	A	C	.	.	ALL=[A/C];IL=TOP;CU=TOP;NO=-427753422	GT	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/0	0/0	0/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/
1	123782	rs109505697	T	C	.	.	ALL=[T/C];IL=BOT;CU=BOT;NO=-427753422	GT	1/0	1/0	1/0	1/0	1/0	1/1	1/1	1/0	1/1	1/0	0/0	0/0	0/0	1/0	0/0	1/
1	124921	rs457956940	T	C	.	.	ALL=[T/C];IL=BOT;CU=BOT;NO=50	GT	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/

Provided files:

Name	Date modified	Type	Size
.Rhistory	6/19/2026 12:33 P...	RHISTORY File	8 KB
250KUF	6/18/2026 12:02 P...	vCalendar File	450,554 ...
breedComposition	6/19/2026 12:18 P...	Text Document	14 KB
breedComposition	6/19/2026 12:18 P...	Microsoft Exce...	203 KB
genotipicData	6/18/2026 2:04 PM	Text Document	217,225 ...
genotipicDataFiltered	6/19/2026 10:53 A...	Text Document	176,027 ...
genotipicDataRecoded	6/18/2026 1:23 PM	Text Document	216,730 ...
PCA_2PCs	6/19/2026 12:32 P...	Text Document	33 KB
PCA_Breed_Merged	6/19/2026 12:32 P...	Text Document	42 KB
populationStructure	6/19/2026 12:14 P...	PNG File	20 KB
populationStructureCode	6/22/2026 10:03 A...	R File	3 KB

rs1366249228: A/G

0/0 = AA

1/1 = GG

Genotype

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	2130309	2130326	2130336	2130350	2130352	2130356	2130358	3109369	3119025	3130004	3130012	31
1	19774	rs136249228	A	G	.	.	ALL=[T/C];IL=BOT;CU=TOP;NO=50	GT	0/1	./.	0/1	0/1	./.	0/1	0/1	./.	0/1	0/0	./.	0/1	0/1	./.	0/1	0/1	0/0	0/
1	19832	rs459659610	C	T	.	.	ALL=[T/C];IL=BOT;CU=BOT;NO=50	GT	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/
1	70600	rs454137042	T	C	.	.	ALL=[T/C];IL=BOT;CU=BOT;NO=-54525902	GT	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/
1	123681	rs209484914	A	C	.	.	ALL=[A/C];IL=TOP;CU=TOP;NO=-427753422	GT	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/0	0/0	0/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/
1	123782	rs109505697	T	C	.	.	ALL=[T/C];IL=BOT;CU=BOT;NO=-427753422	GT	1/0	1/0	1/0	1/0	1/0	1/1	1/1	1/0	1/1	1/0	0/0	0/0	0/0	1/0	0/0	1/
1	124921	rs457956940	T	C	.	.	ALL=[T/C];IL=BOT;CU=BOT;NO=50	GT	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/

Provided files:

Name	Date modified	Type	Size
.Rhistory	6/19/2026 12:33 P...	RHISTORY File	8 KB
250KUF	6/18/2026 12:02 P...	vCalendar File	450,554 ...
breedComposition	6/19/2026 12:18 P...	Text Document	14 KB
breedComposition	6/19/2026 12:18 P...	Microsoft Exce...	203 KB
genotipicData	6/18/2026 2:04 PM	Text Document	217,225 ...
genotipicDataFiltered	6/19/2026 10:53 A...	Text Document	176,027 ...
genotipicDataRecoded	6/18/2026 1:23 PM	Text Document	216,730 ...
PCA_2PCs	6/19/2026 12:32 P...	Text Document	33 KB
PCA_Breed_Merged	6/19/2026 12:32 P...	Text Document	42 KB
populationStructure	6/19/2026 12:14 P...	PNG File	20 KB
populationStructureCode	6/22/2026 10:03 A...	R File	3 KB

rs1366249228: A/G

0/0 = AA

0/1 = ?

1/1 = GG

Genotype

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	2130309	2130326	2130336	2130350	2130352	2130356	2130358	3109369	3119025	3130004	3130012	31
1	19774	rs136249228	A	G	.	.	ALL=[T/C];IL=BOT;CU=TOP;NO=50	GT	0/1	./.	0/1	0/1	./.	0/1	0/1	./.	0/1	0/0	./.	0/1	0/1	./.	0/1	0/1	0/0	0/
1	19832	rs459659610	C	T	.	.	ALL=[T/C];IL=BOT;CU=BOT;NO=50	GT	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/
1	70600	rs454137042	T	C	.	.	ALL=[T/C];IL=BOT;CU=BOT;NO=-54525902	GT	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/
1	123681	rs209484914	A	C	.	.	ALL=[A/C];IL=TOP;CU=TOP;NO=-427753422	GT	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/0	0/0	0/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/
1	123782	rs109505697	T	C	.	.	ALL=[T/C];IL=BOT;CU=BOT;NO=-427753422	GT	1/0	1/0	1/0	1/0	1/0	1/1	1/1	1/0	1/1	1/0	0/0	0/0	0/0	1/0	0/0	1/
1	124921	rs457956940	T	C	.	.	ALL=[T/C];IL=BOT;CU=BOT;NO=50	GT	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/

Provided files:

Name	Date modified	Type	Size
.Rhistory	6/19/2026 12:33 P...	RHISTORY File	8 KB
250KUF	6/18/2026 12:02 P...	vCalendar File	450,554 ...
breedComposition	6/19/2026 12:18 P...	Text Document	14 KB
breedComposition	6/19/2026 12:18 P...	Microsoft Exce...	203 KB
genotipicData	6/18/2026 2:04 PM	Text Document	217,225 ...
genotipicDataFiltered	6/19/2026 10:53 A...	Text Document	176,027 ...
genotipicDataRecoded	6/18/2026 1:23 PM	Text Document	216,730 ...
PCA_2PCs	6/19/2026 12:32 P...	Text Document	33 KB
PCA_Breed_Merged	6/19/2026 12:32 P...	Text Document	42 KB
populationStructure	6/19/2026 12:14 P...	PNG File	20 KB
populationStructureCode	6/22/2026 10:03 A...	R File	3 KB

rs1366249228: A/G

0/0 = AA

0/1 = AG

1/1 = GG

Genotype

```
#CHROM→POS→ID→REF→ALT→QUAL→FILTER→INFO→FORMAT→2130309→2130326→2130336→2130350→2130352→2130356→2130358→3109369→3119025→3130004→3130012→31
1→19774→rs136249228→A→G→.→.→ALT=[T/C];IL=BOT;CU=TOP;NO=50→GT→0/1→./.→0/1→0/1→./.→0/1→0/1→./.→0/1→0/0→./.→0/1→0/1→./.→0/1→0/1→0/0→0/
1→19832→rs459659610→C→T→.→.→ALT=[T/C];IL=BOT;CU=BOT;NO=50→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/1→0/
1→70600→rs454137042→T→C→.→.→ALT=[T/C];IL=BOT;CU=BOT;NO=-54525902→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/
1→123681→rs209484914→A→C→.→.→ALT=[A/C];IL=TOP;CU=TOP;NO=-427753422→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/1→0/0→0/0→0/1→0/0→0/0→0/0→0/0→0/0→0/
1→123782→rs109505697→T→C→.→.→ALT=[T/C];IL=BOT;CU=BOT;NO=-427753422→GT→1/0→1/0→1/0→1/0→1/0→1/1→1/1→1/0→1/1→1/0→0/0→0/0→0/0→1/0→0/0→1/
1→124921→rs457956940→T→C→.→.→ALT=[T/C];IL=BOT;CU=BOT;NO=50→GT→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/0→0/
```

Provided files:

rs1366249228: A/G

0/0 = AA
0/1 = AG
1/1 = GG

0/0 = 0
0/1 = 1
1/1 = 2

Name	Date modified	Type	Size
.Rhistory	6/19/2026 12:33 P...	RHISTORY File	8 KB
250KUF	6/18/2026 12:02 P...	vCalendar File	450,554 ...
breedComposition	6/19/2026 12:18 P...	Text Document	14 KB
breedComposition	6/19/2026 12:18 P...	Microsoft Exce...	203 KB
genotipicData	6/18/2026 2:04 PM	Text Document	217,225 ...
genotipicDataFiltered	6/19/2026 10:53 A...	Text Document	176,027 ...
genotipicDataRecoded	6/18/2026 1:23 PM	Text Document	216,730 ...
PCA_2PCs	6/19/2026 12:32 P...	Text Document	33 KB
PCA_Breed_Merged	6/19/2026 12:32 P...	Text Document	42 KB
populationStructure	6/19/2026 12:14 P...	PNG File	20 KB
populationStructureCode	6/22/2026 10:03 A...	R File	3 KB

Polymorphism	2130309	2130326	2130336	2130350	2130352	2130356	2130358	3109369	3119025	3130004	3130012	3130014	3130020	3130025	3130027	3130029	3130030	3130031	313004
1_19774	1	.	1	1	.	1	1	.	1	0	.	1	1	.	1	1	0	1	.	1	1	1	1	1	1	0	.	1	.	.	1	.	1	1	1	.	1	1	1
1_19832	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	.	0	0	0	0
1_70600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
1_123681	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1_123782	2	2	2	2	2	2	2	2	2	2	0	0	0	2	0	2	0	2	2	0	0	0	0	2	2	2	0	2	2	2	2	0	0	0	2	2	2	0	2
1_124921	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1_264526	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Provided files:

Name	Date modified	Type	Size
.Rhistory	6/19/2026 12:33 P...	RHISTORY File	8 KB
250KUF	6/18/2026 12:02 P...	vCalendar File	450,554 ...
breedComposition	6/19/2026 12:18 P...	Text Document	14 KB
breedComposition	6/19/2026 12:18 P...	Microsoft Exce...	203 KB
genotipicData	6/18/2026 2:04 PM	Text Document	217,225 ...
genotipicDataFiltered	6/19/2026 10:53 A...	Text Document	176,027 ...
genotipicDataRecoded	6/18/2026 1:23 PM	Text Document	216,730 ...
PCA_2PCs	6/19/2026 12:32 P...	Text Document	33 KB
PCA_Breed_Merged	6/19/2026 12:32 P...	Text Document	42 KB
populationStructure	6/19/2026 12:14 P...	PNG File	20 KB
populationStructureCode	6/22/2026 10:03 A...	R File	3 KB

rs1366249228: A/G

0/0 = AA
0/1 = AG
1/1 = GG

0/0 = 0
0/1 = 1
1/1 = 2

Exclusion of non-polymorphic markers

Polymorphism	2130309	2130326	2130336	2130350	2130352	2130356	2130358	3109369	3119025	3130004	3130012	3130014	3130020	3130025	3130027	3130029	3130030	3130031	313004
1_19774	1	.	1	1	.	1	1	.	1	0	.	1	1	.	1	1	0	1	.	1	1	1	1	1	1	0	.	1	.	.	1	.	1	1	1	.	1	1	1
1_19832	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	.	0	0	0	0
1_70600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
1_123681	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1_123782	2	2	2	2	2	2	2	2	2	2	0	0	0	2	0	2	0	2	2	0	0	0	0	2	2	2	0	2	2	2	2	0	0	0	2	2	2	0	2
1_124921	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1_264526	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Provided files:

Name	Date modified	Type	Size
.Rhistory	6/19/2026 12:33 P...	RHISTORY File	8 KB
250KUF	6/18/2026 12:02 P...	vCalendar File	450,554 ...
breedComposition	6/19/2026 12:18 P...	Text Document	14 KB
breedComposition	6/19/2026 12:18 P...	Microsoft Exce...	203 KB
genotipicData	6/18/2026 2:04 PM	Text Document	217,225 ...
genotipicDataFiltered	6/19/2026 10:53 A...	Text Document	176,027 ...
genotipicDataRecoded	6/18/2026 1:23 PM	Text Document	216,730 ...
PCA_2PCs	6/19/2026 12:32 P...	Text Document	33 KB
PCA_Breed_Merged	6/19/2026 12:32 P...	Text Document	42 KB
populationStructure	6/19/2026 12:14 P...	PNG File	20 KB
populationStructureCode	6/22/2026 10:03 A...	R File	3 KB

rs1366249228: A/G

0/0 = AA
0/1 = AG
1/1 = GG

0/0 = 0
0/1 = 1
1/1 = 2

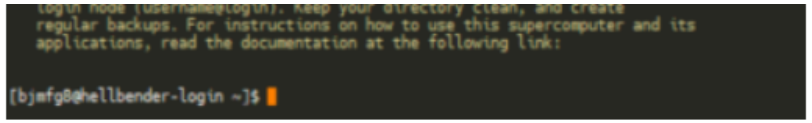
Exclusion of non-polymorphic markers

No genomic variability

Polymorphism	2130309	2130326	2130336	2130350	2130352	2130356	2130358	3109369	3119025	3130004	3130012	3130014	3130020	3130025	3130027	3130029	3130030	3130031	313004
1_19774	1	.	1	1	.	1	1	.	1	0	.	1	1	.	1	1	0	1	.	1	1	1	1	1	1	0	.	1	.	.	1	.	1	1	1	.	1	1	1
1_19832	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	.	0	0	0	0
1_70600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
1_123681	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1_123782	2	2	2	2	2	2	2	2	2	2	0	0	0	2	0	2	0	2	2	0	0	0	0	2	2	2	0	2	2	2	2	0	0	0	2	2	2	0	2
1_124921	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1_264526	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Run the R script for analysis

- Go to https://docs.itrss.umsystem.edu/pub/hpc/hellbender



Browser

See https://docs.itrss.umsystem.edu/pub/hpc/hellbender#open_ondemand on how to access Hellbender via Open OnDemand.

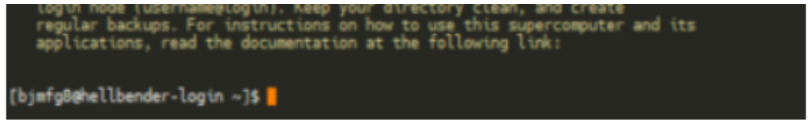
Open OnDemand

- https://ondemand.rnet.missouri.edu - Hellbender Open OnDemand (Researcher)
- https://hb-classes.missouri.edu - Hellbender Classes Open OnDemand (Classes)

OnDemand provides an integrated, single access point for all of your HPC resources. The following apps are currently available on Hellbender's Open Ondemand.

- Jupyter Notebook
- RStudio Server
- Virtual Desktop
- VSCode

- Go to https://docs.itrss.umsystem.edu/pub/hpc/hellbender



Browser

See https://docs.itrss.umsystem.edu/pub/hpc/hellbender#open_ondemand on how to access Hellbender via Open OnDemand.

Open OnDemand

- https://ondemand.rnet.missouri.edu - Hellbender Open OnDemand (Researcher)
- https://hb-classes.missouri.edu - Hellbender Classes Open OnDemand (Classes)

OnDemand provides an integrated, single access point for all of your HPC resources. The following apps are currently available on Hellbender's Open Ondemand.

- Jupyter Notebook
- RStudio Server
- Virtual Desktop
- VSCode

IT Research Support Solutions
University of Missouri

OnDemand provides an integrated, single access point for all of your HPC resources.

Pinned Apps A featured subset of all available apps

Code Server

System Installed App

RStudio Server

System Installed App

Jupyter Notebook

System Installed App

Matlab

System Installed App

Hellbender Desktop

System Installed App

VMD

System Installed App

MatLab (Web)

System Installed App

Hellbender Shell Access

System Installed App

IT Research Support Solutions
University of Missouri

OnDemand provides an integrated, single access point for all of your HPC resources.

Pinned Apps A featured subset of all available apps

Code Server
System Installed App

RStudio Server
System Installed App

Jupyter Notebook
System Installed App

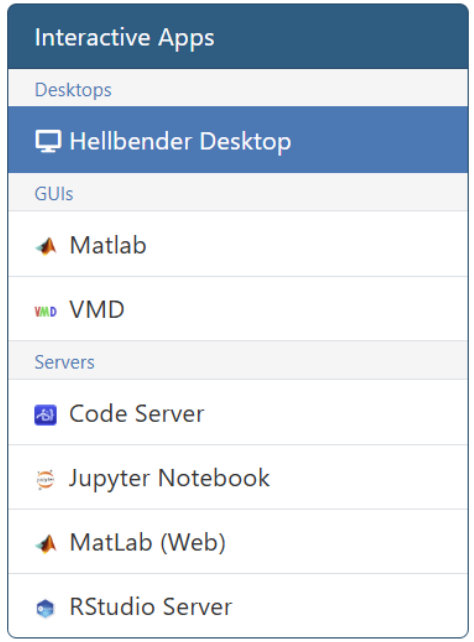
Matlab
System Installed App

Hellbender Desktop
System Installed App

VMD
System Installed App

MatLab (Web)
System Installed App

Hellbender Shell Access
System Installed App



Hellbender Desktop

This app will launch an interactive desktop on one or more compute nodes. You will have full access to the resources these nodes provide. This is analogous to an interactive batch job.

Account

general

Partition

cimuse-class

Number of hours

1

Maximum of 48 hours

Node Type

CPU

Number of Nodes

1

Number of CPUs

1

Memory (GB)

30

Advanced Options

Enter additional SLURM options.

Enter additional SLURM options.

For example, to run in a reservation use `--reservation=my_reservation`. To run an MPI application use `--ntasks` or `--ntasks-per-node`.

Launch

* The Hellbender Desktop session data for this session can be accessed under the data root directory.

Regular maintenance is scheduled for Tuesday, August 11th 2026, from 8:00 AM to 5:00 PM. The Hellbender cluster will be unavailable at this time.

Session was successfully created.

Interactive Apps

Desktops

- Hellbender Desktop

GUIs

- Matlab
- VMD

Servers

- Code Server
- Jupyter Notebook
- MatLab (Web)
- RStudio Server

Hellbender Desktop (15282851) — Queued

Created at: 2026-07-23 15:46:20 CDT

Cancel

Time Requested: 1 hour

Session ID: 9e57c54b-f4ed-4351-8b5c-a587d8409516

Please be patient as your job currently sits in queue. The wait time depends on the number of cores as well as time requested.

Hellbender Desktop (15281835) — Completed | | |

Created at: 2026-07-23 15:06:46 CDT

Delete

Session ID: 4d37fc15-da08-45e1-9ba1-33573a315b2d

For debugging purposes, this card will be retained for 6 more days

Hellbender Desktop (15282851)

1 node | 1 core | Running

Host: >_ c221.mgmt.hellbender

Cancel

Created at: 2026-07-23 15:46:20 CDT

Time Remaining: 59 minutes

Session ID: 9e57c54b-f4ed-4351-8b5c-a587d8409516

Compression

0 (low) to 9 (high)

Image Quality

0 (low) to 9 (high)

Launch Hellbender Desktop

View Only (Share-able Link)

Hellbender Desktop (15282851)

1 node | 1 core | Running

Host: >_ c221.mgmt.hellbender

Cancel

Created at: 2026-07-23 15:46:20 CDT

Time Remaining: 59 minutes

Session ID: 9e57c54b-f4ed-4351-8b5c-a587d8409516

Compression

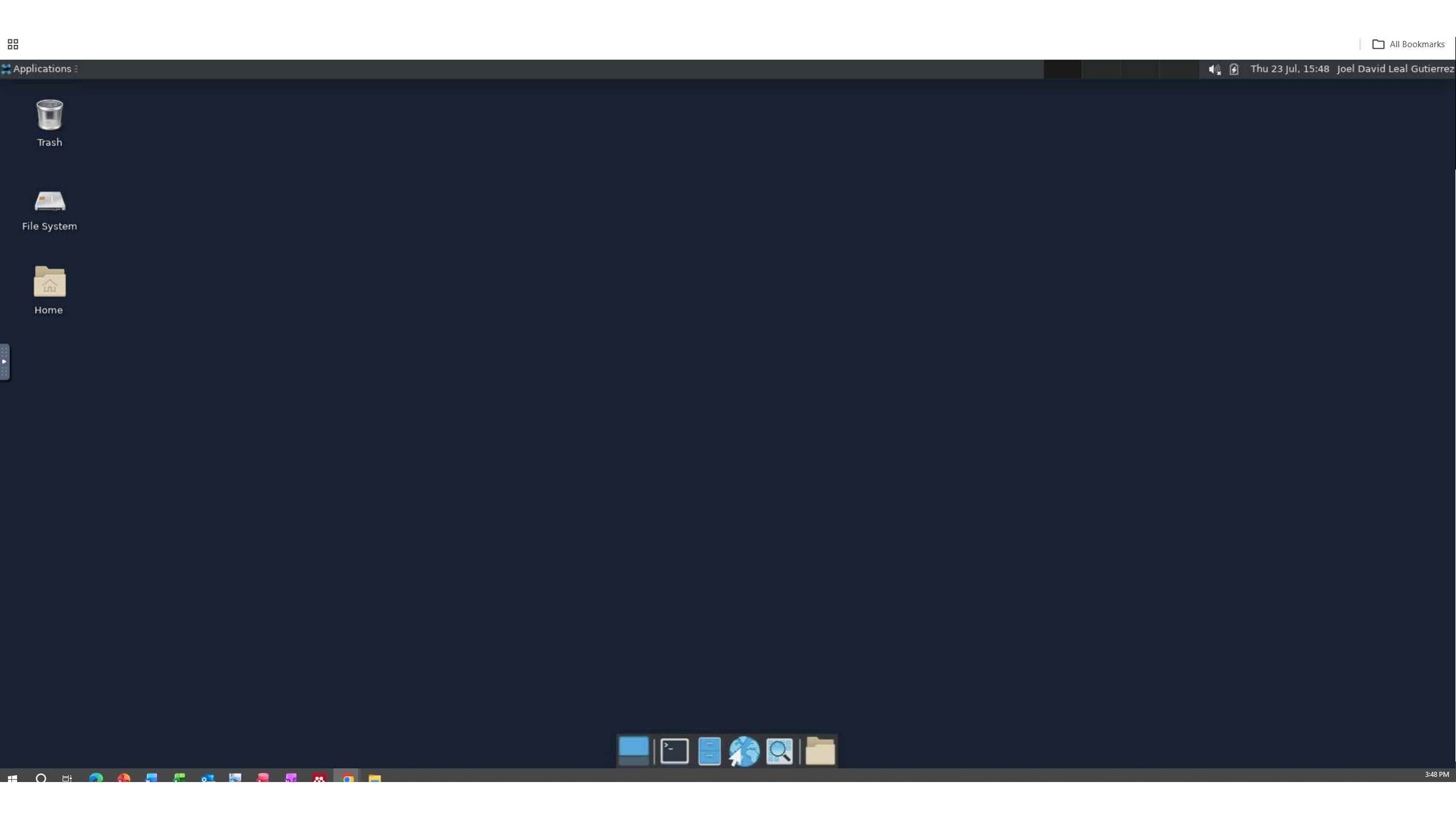
0 (low) to 9 (high)

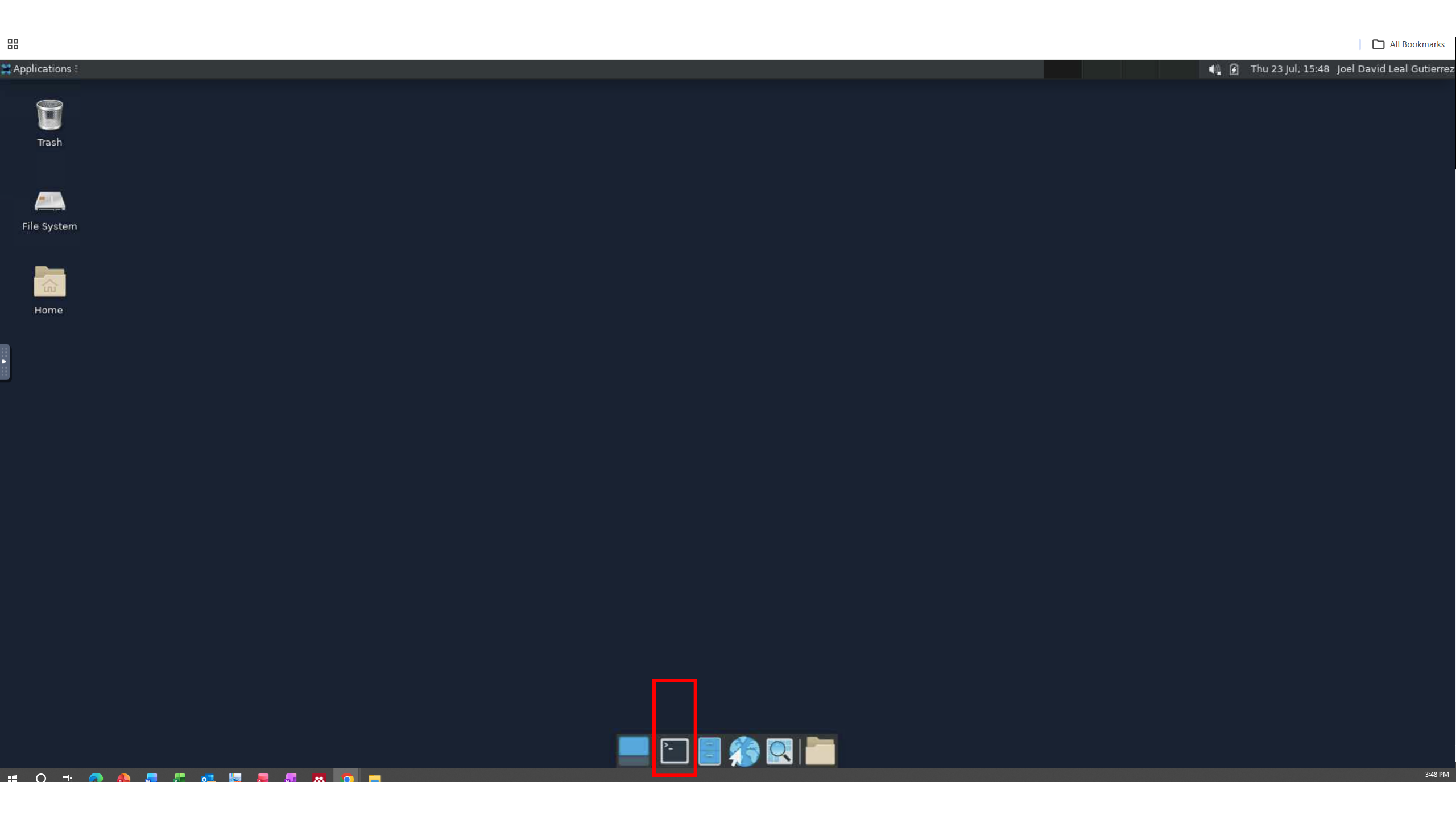
Image Quality

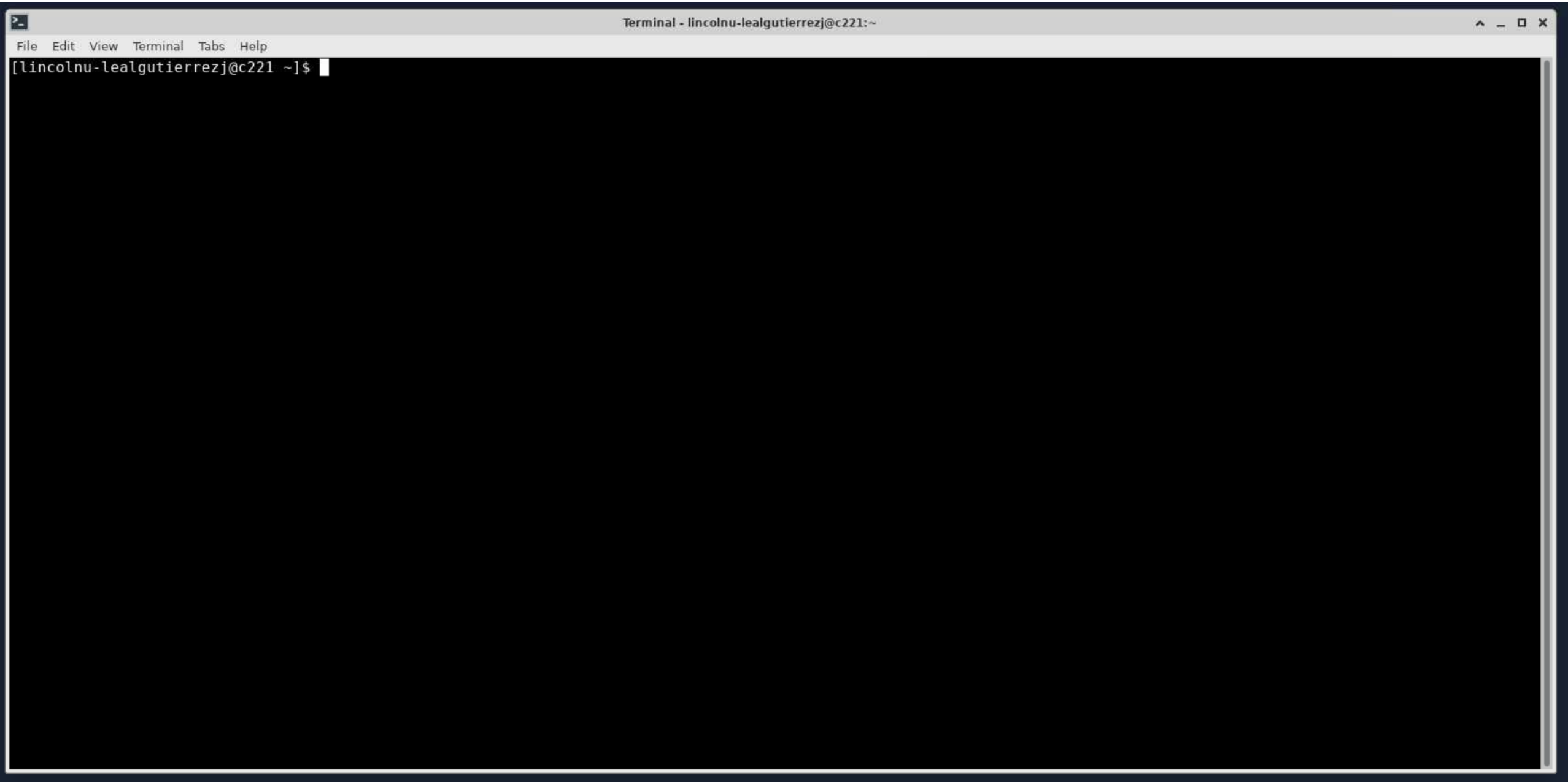
0 (low) to 9 (high)

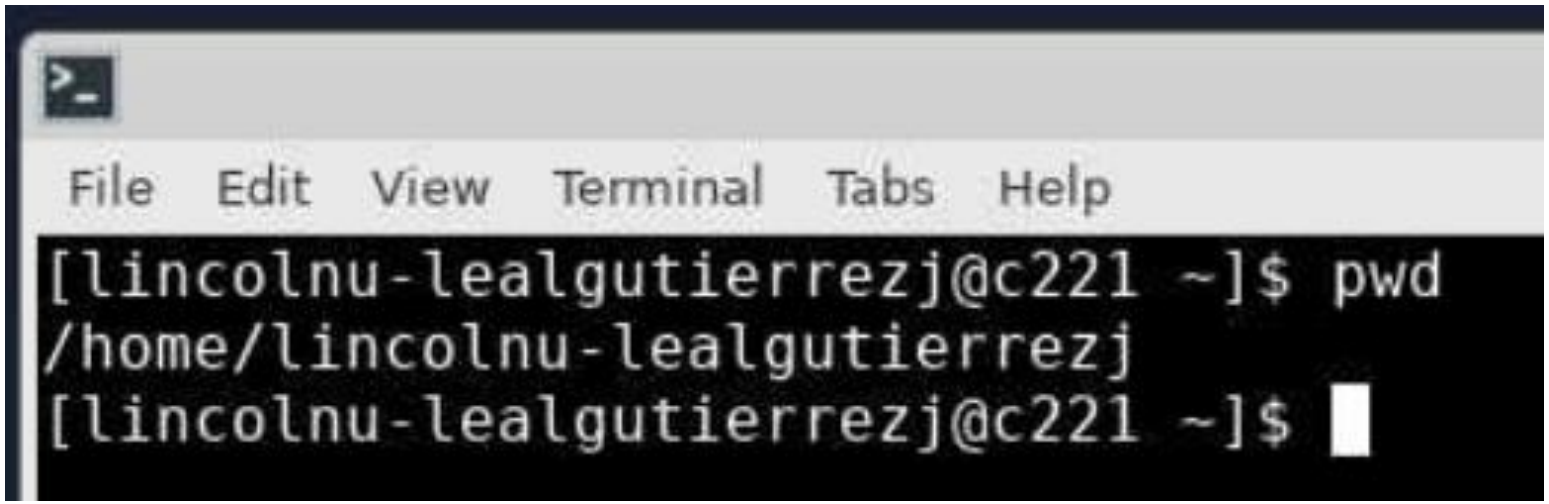
Launch Hellbender Desktop

View Only (Share-able Link)


All Bookmarks
Applications
Thu 23 Jul, 15:48 Joel David Leal Gutierrez
Trash
File System
Home
3:48 PM


All Bookmarks
Applications
Thu 23 Jul, 15:48 Joel David Leal Gutierrez
Trash
File System
Home
3:48 PM


Terminal - lincolnu-lealgutierrezj@c221:~
File Edit View Terminal Tabs Help
[lincolnu-lealgutierrezj@c221 ~]$


File Edit View Terminal Tabs Help
[lincolnu-lealgutierrezj@c221 ~]$ pwd
/home/lincolnu-lealgutierrezj
[lincolnu-lealgutierrezj@c221 ~]$


[lincolnu-lealgutierrezj@c221 ~]$ ls
cimuseWorkshop data Desktop ondemand R
[lincolnu-lealgutierrezj@c221 ~]$

We need the folder “cimuseWorkshop”

```
[lincolnu-lealgutierrezj@c221 ~]$ ls
cimuseWorkshop  data  Desktop  ondemand  R
[lincolnu-lealgutierrezj@c221 ~]$
```

We need the folder “cimuseWorkshop”

```
[lincolnu-lealgutierrezj@c221 ~]$ cp -r /cluster/pixstor/cimuse-workshop/cimuseWorkshop/ .
[lincolnu-lealgutierrezj@c221 ~]$
```

```
[lincolnu-lealgutierrezj@c221 ~]$ ls
cimuseWorkshop  data  Desktop  ondemand  R
[lincolnu-lealgutierrezj@c221 ~]$
```

We need the folder “cimuseWorkshop”

```
[lincolnu-lealgutierrezj@c221 ~]$ cp -r /cluster/pixstor/cimuse-workshop/cimuseWorkshop/ .
[lincolnu-lealgutierrezj@c221 ~]$
```

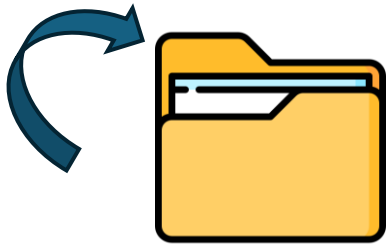
```
[lincolnu-lealgutierrezj@c221 ~]$ ls
cimuseWorkshop  data  Desktop  ondemand  R
[lincolnu-lealgutierrezj@c221 ~]$
```

```
[lincolnu-lealgutierrezj@c221 ~]$ ls
cimuseWorkshop  data  Desktop  ondemand  R
[lincolnu-lealgutierrezj@c221 ~]$
```

We need the folder “cimuseWorkshop”

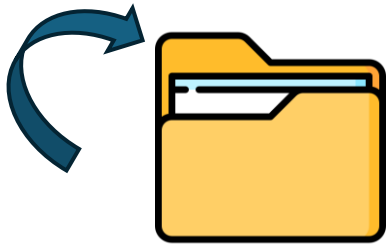
```
[lincolnu-lealgutierrezj@c221 ~]$ cp -r /cluster/pixstor/cimuse-workshop/cimuseWorkshop/ .
[lincolnu-lealgutierrezj@c221 ~]$
```

```
[lincolnu-lealgutierrezj@c221 ~]$ ls
cimuseWorkshop  data  Desktop  ondemand  R
[lincolnu-lealgutierrezj@c221 ~]$
```



cimuseWorkshop

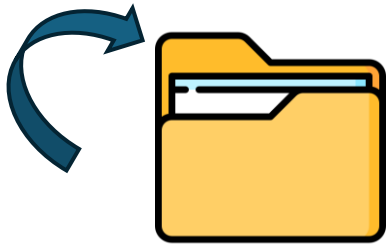
```
[lincolnu-lealgutierrezj@c221 ~]$ cd cimuseWorkshop/
```



cimuseWorkshop

```
[lincolnu-lealgutierrezj@c221 ~]$ cd cimuseWorkshop/
```

```
[lincolnu-lealgutierrezj@c221 cimuseWorkshop]$ ls
 250KUF.vcf                      genotipicDataFilteredPCARun.txt   PCA_Breed_MergedBuying.txt
 breedCompositionBuying.txt      genotipicDataFiltered.txt         PCA_Breed_Merged.txt
 breedComposition.txt            genotipicDataRecoded.txt          populationStructureCode.R
 breedComposition.xls            genotipicData.txt                 populationStructureCode.txt
 cimuseWorkshop                  PCA_2PCs.txt                      populationStructure.png
[lincolnu-lealgutierrezj@c221 cimuseWorkshop]$
```



cimuseWorkshop

```
[lincolnu-lealgutierrezj@c221 ~]$ cd cimuseWorkshop/
```

```
[lincolnu-lealgutierrezj@c221 cimuseWorkshop]$ ls
 250KUF.vcf                      genotipicDataFilteredPCARun.txt   PCA_Breed_MergedBuying.txt
 breedCompositionBuying.txt      genotipicDataFiltered.txt         PCA_Breed_Merged.txt
 breedComposition.txt            genotipicDataRecoded.txt          populationStructureCode.R
 breedComposition.xls            genotipicData.txt                 populationStructureCode.txt
 cimuseWorkshop                  PCA_2PCs.txt                      populationStructure.png
[lincolnu-lealgutierrezj@c221 cimuseWorkshop]$
```

```
[lincolnu-lealgutierrezj@c221 cimuseWorkshop]$ ml cimuse_workshop
[lincolnu-lealgutierrezj@c221 cimuseWorkshop]$ rstudio
```

Lincoln University of Missouri | My Interactive Sessions - RSS C | RStudio Server | Ask Gemini

ondemand.rnet.missouri.edu/rnode/c159.mgmt.hellbender/57973/

File Edit Code View Plots Session Build Debug Profile Tools Help

lincolnu-lealgutierrezj

Go to file/function | Addins | Project: (None)

Console Terminal Jobs

R 4.2.0 · ~/

```
R version 4.2.0 (2022-04-22) -- "Vigorous Calisthenics"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.


> getwd
function ()
.Internal(getwd())
<bytecode: 0x564cfc2cc6a0>
<environment: namespace:base>
> getwd()
[1] "/home/lincolnu-lealgutierrezj"
Connected to your session in progress, last started 2026-Jul-02 13:28:38 UTC (37 minutes ago)
>
```

Environment History Connections Tutorial

Import Dataset | 178 MiB | List

R | Global Environment

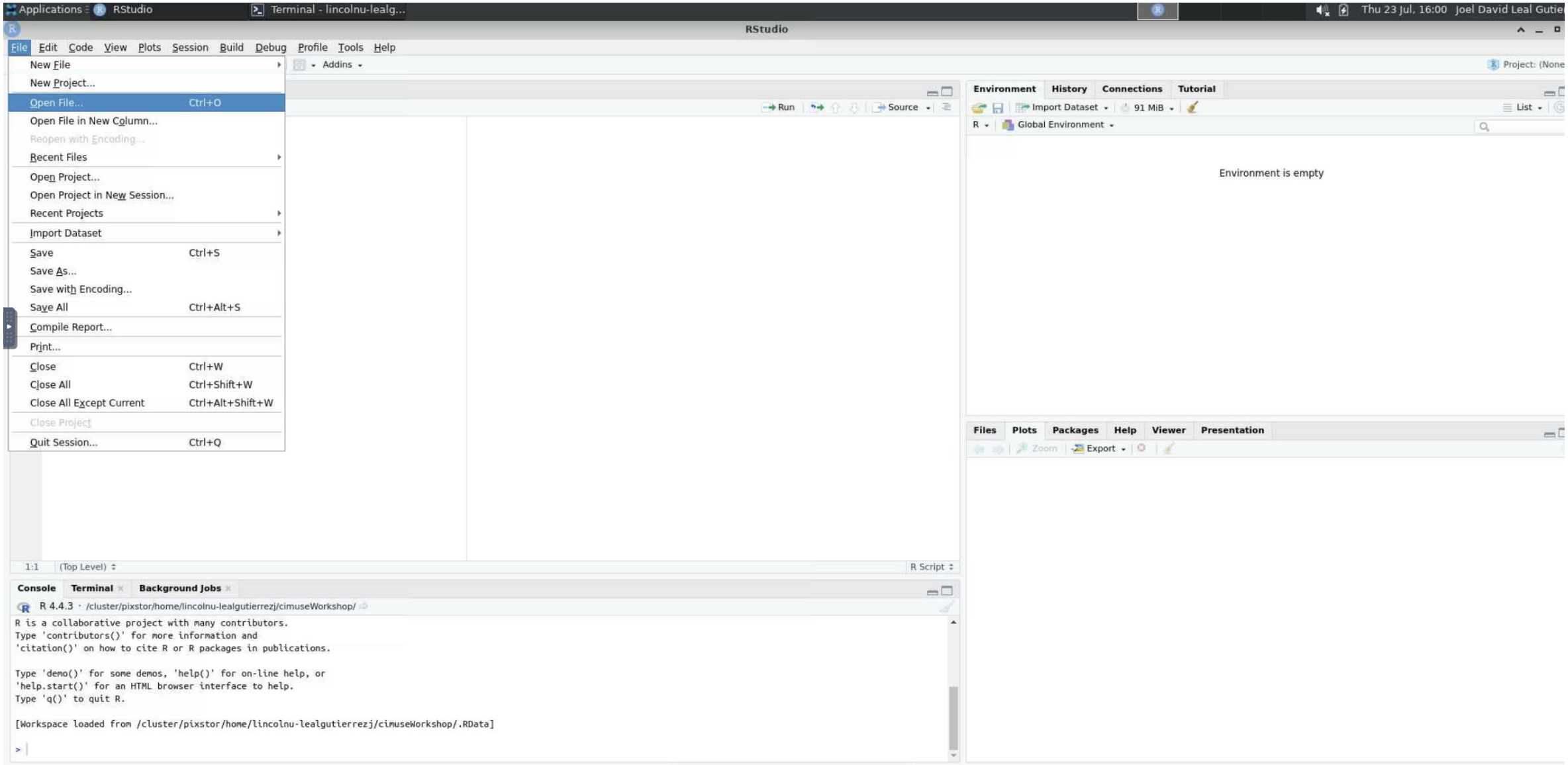
Environment is empty

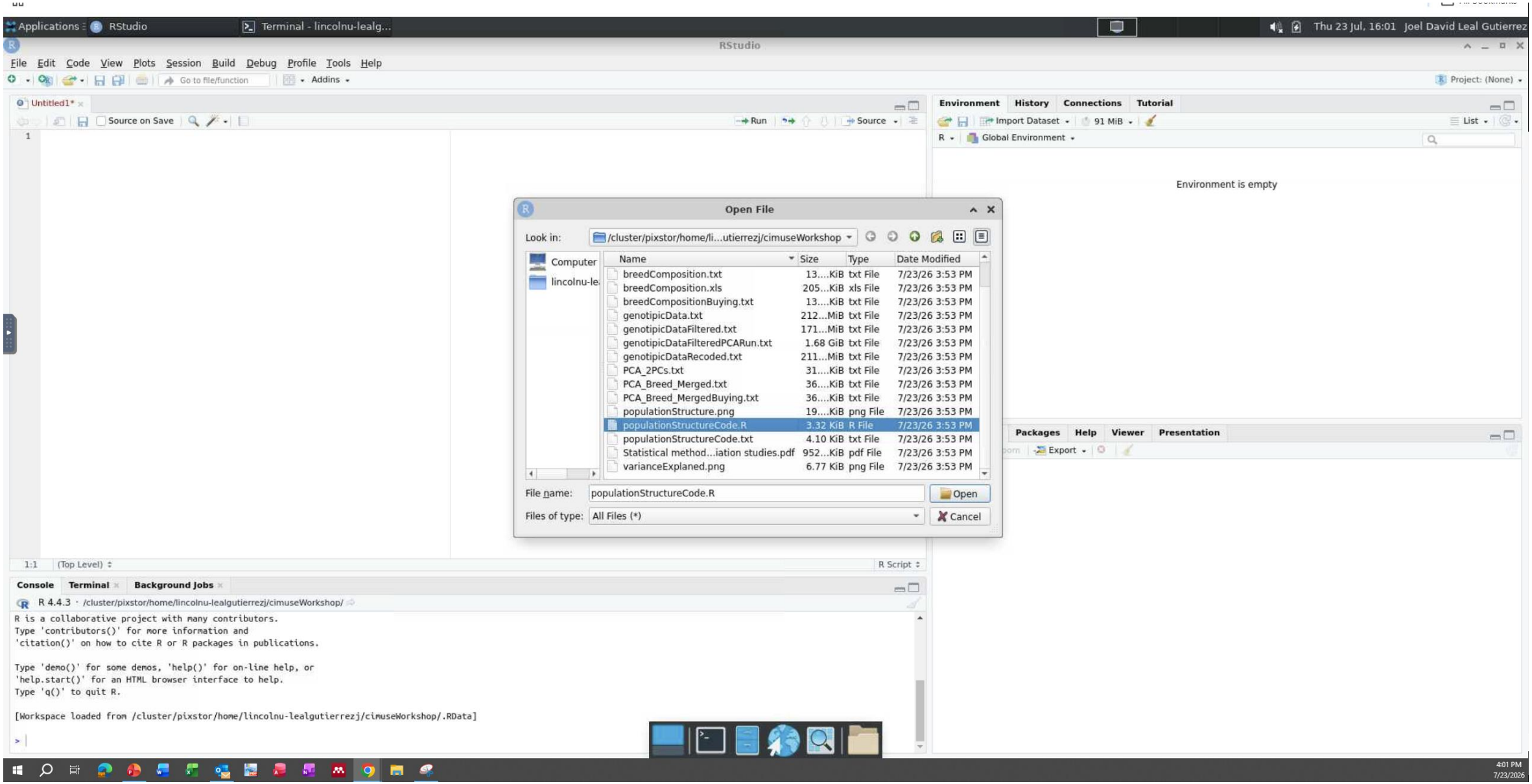
Files Plots Packages Help Viewer

New Folder | New Blank File | Upload | Delete | Rename | More

Home

Name	Size	Modified
data		
ondemand		
R		


Applications
RStudio
Terminal - lincolnu-lealg...
Thu 23 Jul, 16:00 Joel David Leal Gutie
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
New File
New Project...
Open File... Ctrl+O
Open File in New Column...
Reopen with Encoding...
Recent Files
Open Project...
Open Project in New Session...
Recent Projects
Import Dataset
Save Ctrl+S
Save As...
Save with Encoding...
Save All Ctrl+Alt+S
Compile Report...
Print...
Close Ctrl+W
Close All Ctrl+Shift+W
Close All Except Current Ctrl+Alt+Shift+W
Close Project
Quit Session... Ctrl+Q
Addins
Project: (None
Run
Source
Environment History Connections Tutorial
Import Dataset
91 MiB
List
R
Global Environment
Environment is empty
Files Plots Packages Help Viewer Presentation
Zoom
Export
1:1 (Top Level)
R Script
Console Terminal Background Jobs
R 4.4.3 · /cluster/pixstor/home/lincolnu-lealgutierrezj/cimuseWorkshop/
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Workspace loaded from /cluster/pixstor/home/lincolnu-lealgutierrezj/cimuseWorkshop/.RData]
>


Applications
RStudio
Terminal - lincolnu-lealg...
Thu 23 Jul, 16:01 Joel David Leal Gutierrez
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Untitled1*
Source on Save
Run
Source
1
1:1 (Top Level)
R Script
Environment History Connections Tutorial
Import Dataset
91 MiB
List
R
Global Environment
Environment is empty
Packages Help Viewer Presentation
Export
Open File
Look in: /cluster/pixstor/home/li...utierrezj/cimuseWorkshop
Computer
lincolnu-le
Name Size Type Date Modified
breedComposition.txt 13....KiB txt File 7/23/26 3:53 PM
breedComposition.xls 205...KiB xls File 7/23/26 3:53 PM
breedCompositionBuying.txt 13....KiB txt File 7/23/26 3:53 PM
genotipicData.txt 212...MiB txt File 7/23/26 3:53 PM
genotipicDataFiltered.txt 171...MiB txt File 7/23/26 3:53 PM
genotipicDataFilteredPCARun.txt 1.68 GiB txt File 7/23/26 3:53 PM
genotipicDataRecoded.txt 211...MiB txt File 7/23/26 3:53 PM
PCA_2PCs.txt 31....KiB txt File 7/23/26 3:53 PM
PCA_Breed_Merged.txt 36....KiB txt File 7/23/26 3:53 PM
PCA_Breed_MergedBuying.txt 36....KiB txt File 7/23/26 3:53 PM
populationStructure.png 19....KiB png File 7/23/26 3:53 PM
populationStructureCode.R 3.32 KiB R File 7/23/26 3:53 PM
populationStructureCode.txt 4.10 KiB txt File 7/23/26 3:53 PM
Statistical method...iation studies.pdf 952...KiB pdf File 7/23/26 3:53 PM
varianceExplaned.png 6.77 KiB png File 7/23/26 3:53 PM
File name: populationStructureCode.R
Open
Files of type: All Files (*)
Cancel
Console Terminal Background Jobs
R 4.4.3 · /cluster/pixstor/home/lincolnu-lealgutierrezj/cimuseWorkshop/
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Workspace loaded from /cluster/pixstor/home/lincolnu-lealgutierrezj/cimuseWorkshop/.RData]
>
4:01 PM
7/23/2026

populationStructureCode.R*

```r
# Read tabular data into R

#setwd("C:/Users/jguti258/Desktop/cimuseWorkshop/")

# get directory location so you can input this as an string to import and save files
directory <- getwd()

# loading an already installed library to be used downstream, data.table
library(data.table)

# read the file genotipicDataFilteredPCARun.txt, which is available inside directory
# it uses tab as delimiter ("\t"), returns a data.frame object and check.names makes names as valid
tgeno <- fread(file.path(directory, "genotipicDataFilteredPCARun.txt"),
               sep = "\t",
               data.table = FALSE,
               check.names = FALSE)

# assigns rownames as the first row
rownames(tgeno) <- tgeno[[1]]

# reads genotipicDataFiltered.txt file, including the header, tab separated
# and assign first row as name identifiers
tgeno <- read.table(file = file.path(directory, "genotipicDataFiltered.txt"), header = TRUE,
           sep = "\t", dec = ".",
           row.names=1)

# invert table tgeno
geno<- t(tgeno)

# Remove markers with zero variance
geno <- geno[, apply(geno, 2, var, na.rm = TRUE) > 0]

# convert to a matrix
geno <- as.matrix(geno)

# Standardize SNPs, from 0, 1 and 2 to -1, 0 and 1
geno.scaled <- scale(geno)

# perform PCA: orthogonal decomposition of variables
# PCA takes many correlated variables (such as thousands of SNPs or gene expression values)
# and summarizes them into a smaller number of new variables called principal components (PCs).
pca <- prcomp(geno.scaled, center = FALSE, scale. = FALSE)

# Variance explained calculation and plotting
# generates orthogonal variables summarizing the original genomic variability
```

3:2 (Top Level)

```
4
5  # get directory location so you can input this as an string to import and save files
6  directory <- getwd()
7
```

```
4
5  # get directory location so you can input this as an string to import and save files
6  directory <- getwd()
7
```

```
7
8  # loading an already installed library to be used downstream, data.table
9  library(data.table)
10
```

```r
# get directory location so you can input this as an string to import and save files
directory <- getwd()


# loading an already installed library to be used downstream, data.table
library(data.table)


# read the file genotipicDataFilteredPCARun.txt, which is available inside directory
# it uses tab as delimiter ("\t"), returns a data.frame object and check.names makes names as valid
tgeno <- fread(file.path(directory, "genotipicDataFilteredPCARun.txt"),
               sep = "\t",
               data.table = FALSE,
               check.names = FALSE)
tgeno <- tgeno[, -1]
```

```r
# get directory location so you can input this as an string to import and save files
directory <- getwd()
```

```r
# loading an already installed library to be used downstream, data.table
library(data.table)
```

```r
# read the file genotipicDataFilteredPCARun.txt, which is available inside directory
# it uses tab as delimiter ("\t"), returns a data.frame object and check.names makes names as valid
tgeno <- fread(file.path(directory, "genotipicDataFilteredPCARun.txt"),
               sep = "\t",
               data.table = FALSE,
               check.names = FALSE)
tgeno <- tgeno[, -1]
```

It will use the value in “directory” as input to find the file needed

```
5  # get directory location so you can input this as an string to import and save files
6  directory <- getwd()
7
8  # loading an already installed library to be used downstream, data.table
9  library(data.table)
10
11 # read the file genotipicDataFilteredPCARun.txt, which is available inside directory
12 # it uses tab as delimiter ("\t"), returns a data.frame object and check.names makes names as valid
13 tgeno <- fread(file.path(directory, "genotipicDataFilteredPCARun.txt"),
14                sep = "\t",
15                data.table = FALSE,
16                check.names = FALSE)
17 tgeno <- tgeno[, -1]
```

It will use the value in “directory” as input to find the file needed

The file needed is “genotypicDataFilteredPCARun.txt”

```
5  # get directory location so you can input this as an string to import and save files
6  directory <- getwd()
7
8  # loading an already installed library to be used downstream, data.table
9  library(data.table)
10
11 # read the file genotipicDataFilteredPCARun.txt, which is available inside directory
12 # it uses tab as delimiter ("\t"), returns a data.frame object and check.names makes names as valid
13 tgeno <- fread(file.path(directory, "genotipicDataFilteredPCARun.txt"),
14                sep = "\t",
15                data.table = FALSE,
16                check.names = FALSE)
17 tgeno <- tgeno[, -1]
```

It will use the value in "directory" as input to find the file needed

The file needed is "genotypicDataFilteredPCARun.txt"

→ File name and its location

```r
# get directory location so you can input this as an string to import and save files
directory <- getwd()


# loading an already installed library to be used downstream, data.table
library(data.table)


# read the file genotipicDataFilteredPCARun.txt, which is available inside directory
# it uses tab as delimiter ("\t"), returns a data.frame object and check.names makes names as valid
tgeno <- fread(file.path(directory, "genotipicDataFilteredPCARun.txt"),
               sep = "\t",
               data.table = FALSE,
               check.names = FALSE)
tgeno <- tgeno[, -1]
```

It will use the value in "directory" as input to find the file needed

The file needed is "genotypicDataFilteredPCARun.txt"

→ File name and its location

Delete the first column from the genotypic data file. This column included the polymorphism names

```r
# get directory location so you can input this as an string to import and save files
directory <- getwd()


# loading an already installed library to be used downstream, data.table
library(data.table)


# read the file genotipicDataFilteredPCARun.txt, which is available inside directory
# it uses tab as delimiter ("\t"), returns a data.frame object and check.names makes names as valid
tgeno <- fread(file.path(directory, "genotipicDataFilteredPCARun.txt"),
               sep = "\t",
               data.table = FALSE,
               check.names = FALSE)
tgeno <- tgeno[, -1]
```

```r
# get directory location so you can input this as an string to import and save files
directory <- getwd()


# loading an already installed library to be used downstream, data.table
library(data.table)


# read the file genotipicDataFilteredPCARun.txt, which is available inside directory
# it uses tab as delimiter ("\t"), returns a data.frame object and check.names makes names as valid
tgeno <- fread(file.path(directory, "genotipicDataFilteredPCARun.txt"),
               sep = "\t",
               data.table = FALSE,
               check.names = FALSE)
tgeno <- tgeno[, -1]
```

Columns are separated by tabs
The object is not a data.table
It also does not check column names to ensure they are valid during imputation

```
19  # an alternative way to read genotipicDataFilteredPCARun.txt file, including the header, tab separated
20  # and assign first row as name identifiers
21  #tgeno <- read.table(file = file.path(directory, "genotipicDataFilteredPCARun.txt"), header = TRUE,
22  #                    sep = "\t", dec = ".",
23  #                    row.names=1)
24
25  # invert table tgeno
26  geno<- t(tgeno)
27
```

Transpose “tgeno” and store the information in a “geno” variable

```
19  # an alternative way to read genotipicDataFilteredPCARun.txt file, including the header, tab separated
20  # and assign first row as name identifiers
21  #tgeno <- read.table(file = file.path(directory, "genotipicDataFilteredPCARun.txt"), header = TRUE,
22  #                    sep = "\t", dec = ".",
23  #                    row.names=1)
```

```
25  # invert table tgeno
26  geno<- t(tgeno)
```

Transpose “tgeno” and store the information in a “geno” variable

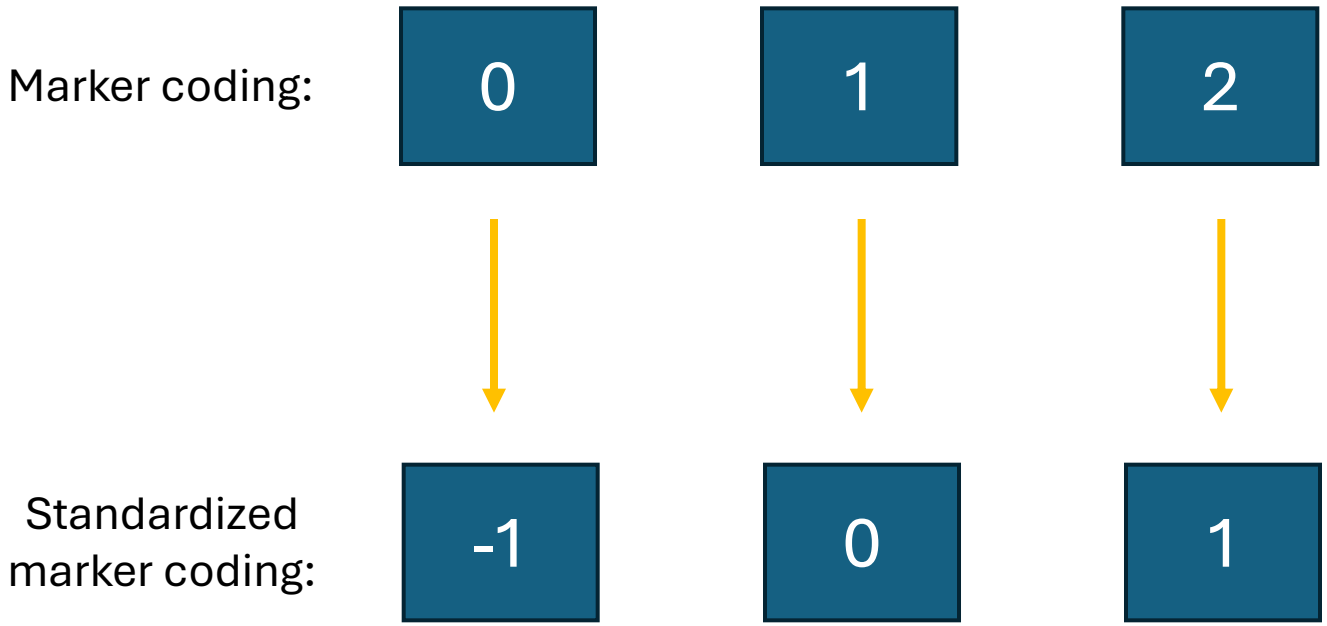
```
28  # Remove SNPs with zero variance
29  geno <- geno[, apply(geno, 2, var, na.rm = TRUE) > 0]
30
```

```
31  # convert to a matrix
32  geno <- as.matrix(geno)
```

```
34  # Standardize SNPs
35  geno.scaled <- scale(geno)
36
```

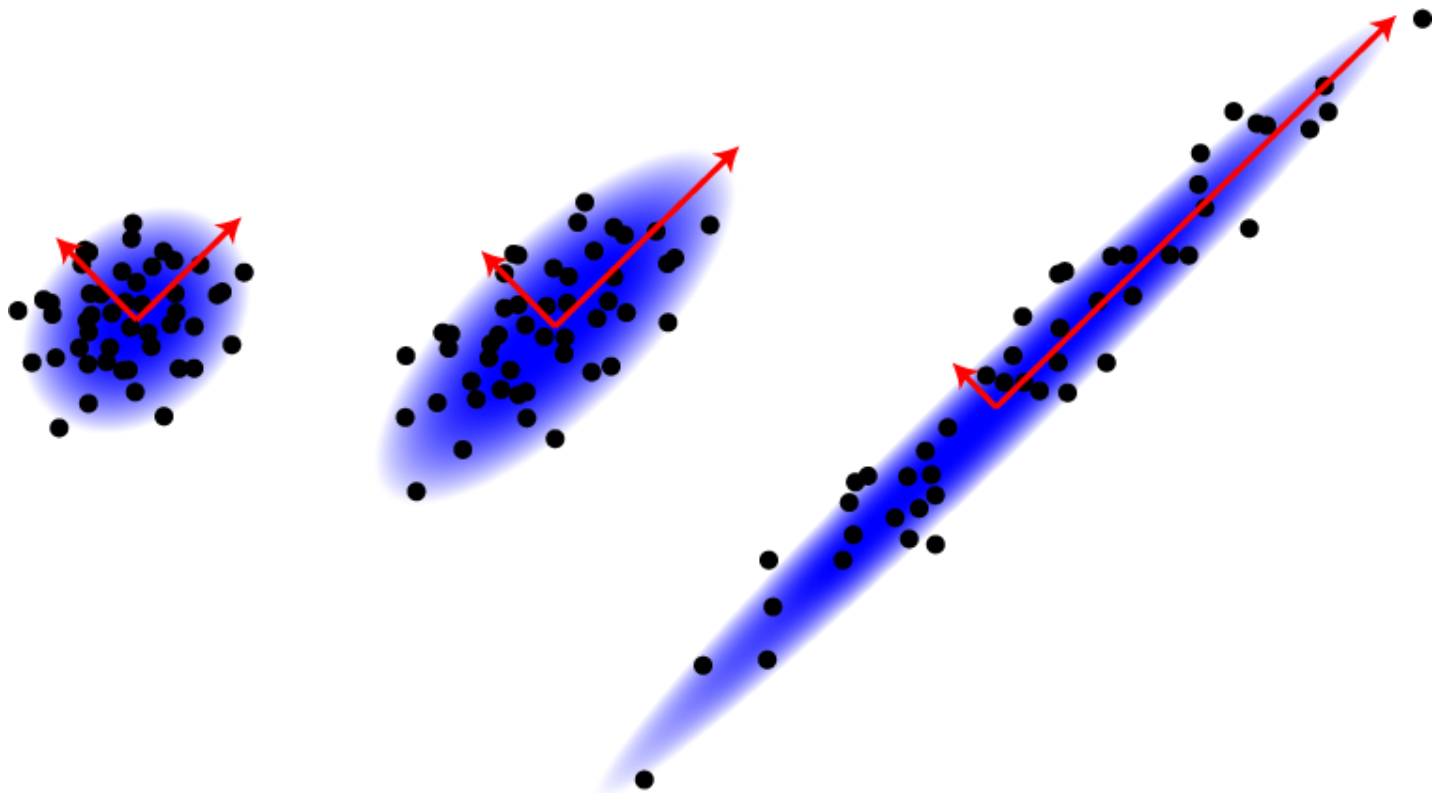
Marker coding:

```
34  # Standardize SNPs
35  geno.scaled <- scale(geno)
36
```



```r
# perform PCA: orthogonal decomposition of variables
# PCA takes many correlated variables (such as thousands of SNPs or gene expression values)
# and summarizes them into a smaller number of new variables called principal components (PCs).
pca <- prcomp(geno.scaled, center = FALSE, scale. = FALSE)
```

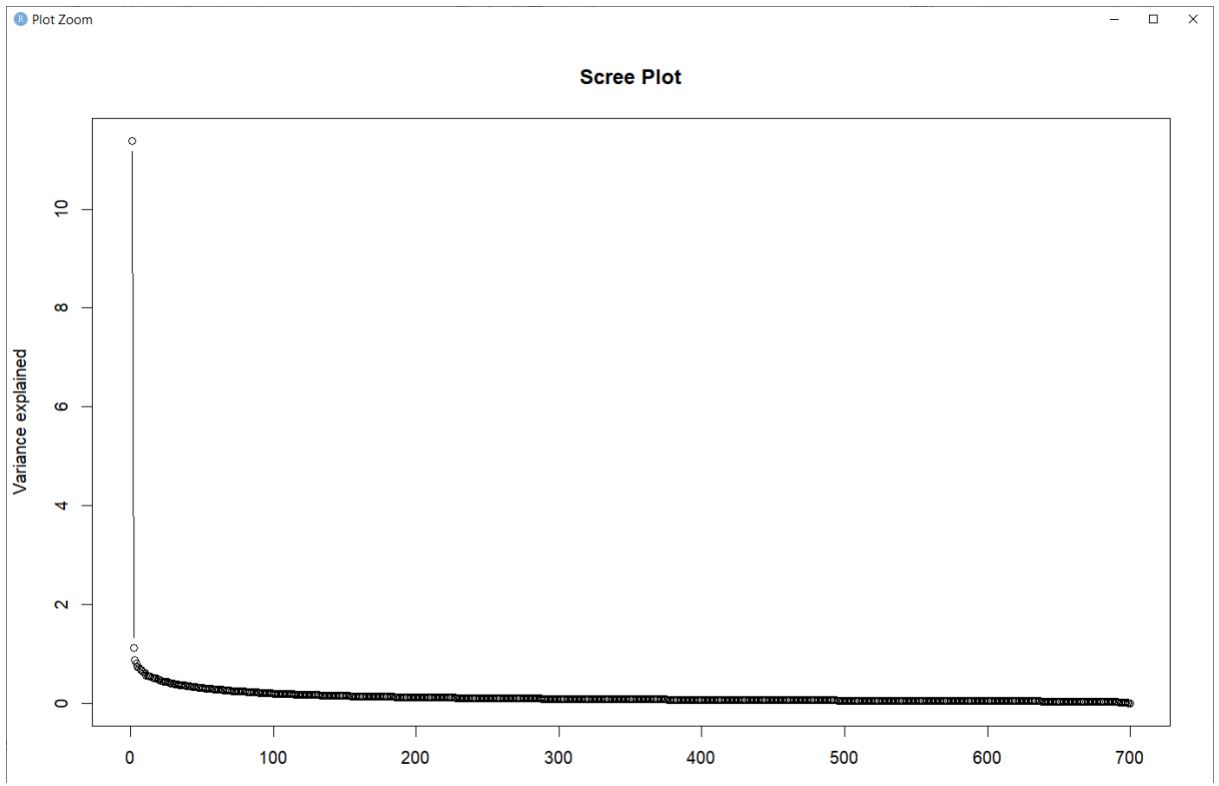
```r
# perform PCA: orthogonal decomposition of variables
# PCA takes many correlated variables (such as thousands of SNPs or gene expression values)
# and summarizes them into a smaller number of new variables called principal components (PCs).
pca <- prcomp(geno.scaled, center = FALSE, scale. = FALSE)
```



https://shapeofdata.wordpress.com/

```r

# Variance explained calculation and plotting
# generates orthogonal variables summarizing the original genomic variability
var.exp <- (pca$sdev^2) / sum(pca$sdev^2) * 100
plot(var.exp, type = "b",
     xlab = "Principal Component",
     ylab = "Variance explained",
     main = "Scree Plot")

```

```r
41
42  # Variance explained calculation and plotting
43  # generates orthogonal variables summarizing the original genomic variability
44  var.exp <- (pca$sdev^2) / sum(pca$sdev^2) * 100
45  plot(var.exp, type = "b",
46       xlab = "Principal Component",
47       ylab = "Variance explained",
48       main = "Scree Plot")
49
```



```r

# Create output table
pca.out <- data.frame(
  ID = rownames(geno),
  pca$x
)

# Create data frame with ID, PC1, and PC2
pca2 <- data.frame(
  ID = rownames(geno),
  PC1 = pca$x[,1],
  PC2 = pca$x[,2]
)

# Save to tab-delimited text file using the original directory variable as input
write.table(
  pca2,
  file = file.path(directory,"PCA_2PCs.txt"),
  sep = "\t",
  quote = FALSE,
  row.names = FALSE
)

```

```
49
50  # Create output table
51  pca.out <- data.frame(
52    ID = rownames(geno),
53    pca$x
54  )
55
56  # Create data frame with ID, PC1, and PC2
57  pca2 <- data.frame(
58    ID = rownames(geno),
59    PC1 = pca$x[,1],
60    PC2 = pca$x[,2]
61  )
62
63  # Save to tab-delimited text file using the original directory variable as input
64  write.table(
65    pca2,
66    file = file.path(directory,"PCA_2PCs.txt"),
67    sep = "\t",
68    quote = FALSE,
69    row.names = FALSE
70  )
71
```

Creates a data frame, keeping row names and including their corresponding PCA1 and PCA2

```
49
50  # Create output table
51  pca.out <- data.frame(
52    ID = rownames(geno),
53    pca$x
54  )
55
56  # Create data frame with ID, PC1, and PC2
57  pca2 <- data.frame(
58    ID = rownames(geno),
59    PC1 = pca$x[,1],
60    PC2 = pca$x[,2]
61  )
62
63  # Save to tab-delimited text file using the original directory variable as input
64  write.table(
65    pca2,
66    file = file.path(directory,"PCA_2PCs.txt"),
67    sep = "\t",
68    quote = FALSE,
69    row.names = FALSE
70  )
71
```

Creates a data frame, keeping row names and including their corresponding PCA1 and PCA2

Write “pca2” in a file using the “directory” path

```r

# Read file PCA_2PCs.txt using the original directory variable as input
pca <- read.table(file.path(directory,"PCA_2PCs.txt"),
                  header = TRUE,
                  sep = "\t",
                  stringsAsFactors = FALSE)

# Read file BreedComposition.txt using the original directory variable as input
breed <- read.table(file.path(directory,"breedComposition.txt"),
                    header = TRUE,
                    sep = "\t",
                    stringsAsFactors = FALSE)

# Merge by animal ID
merged <- merge(pca,
                breed,
                by = "ID")

# Save merged file, PCA_Breed_Merged.txt, using the original directory variable as input
write.table(merged,
            file = file.path(directory,"PCA_Breed_Merged.txt"),
            sep = "\t",
            quote = FALSE,
            row.names = FALSE)

```

```
72  # Read file PCA_2PCs.txt using the original directory variable as input
73  pca <- read.table(file.path(directory,"PCA_2PCs.txt"),
74                    header = TRUE,
75                    sep = "\t",
76                    stringsAsFactors = FALSE)
77
78  # Read file BreedComposition.txt using the original directory variable as input
79  breed <- read.table(file.path(directory,"breedComposition.txt"),
80                      header = TRUE,
81                      sep = "\t",
82                      stringsAsFactors = FALSE)
83
84  # Merge by animal ID
85  merged <- merge(pca,
86                  breed,
87                  by = "ID")
88
89  # Save merged file, PCA_Breed_Merged.txt, using the original directory variab
90  write.table(merged,
91              file = file.path(directory,"PCA_Breed_Merged.txt"),
92              sep = "\t",
93              quote = FALSE,
94              row.names = FALSE)
95
```

```r
# Read file PCA_2PCs.txt using the original directory variable as input
pca <- read.table(file.path(directory,"PCA_2PCs.txt"),
                  header = TRUE,
                  sep = "\t",
                  stringsAsFactors = FALSE)

# Read file BreedComposition.txt using the original directory variable as input
breed <- read.table(file.path(directory,"breedComposition.txt"),
                    header = TRUE,
                    sep = "\t",
                    stringsAsFactors = FALSE)

# Merge by animal ID
merged <- merge(pca,
                breed,
                by = "ID")

# Save merged file, PCA_Breed_Merged.txt, using the original directory variab
write.table(merged,
            file = file.path(directory,"PCA_Breed_Merged.txt"),
            sep = "\t",
            quote = FALSE,
            row.names = FALSE)
```

```
71
72  # Read file PCA_2PCs.txt using the original directory variable as input
73  pca <- read.table(file.path(directory,"PCA_2PCs.txt"),
74                    header = TRUE,
75                    sep = "\t",
76                    stringsAsFactors = FALSE)
77
78  # Read file BreedComposition.txt using the original directory variable as input
79  breed <- read.table(file.path(directory,"breedComposition.txt"),
80                      header = TRUE,
81                      sep = "\t",
82                      stringsAsFactors = FALSE)
83
84  # Merge by animal ID
85  merged <- merge(pca,
86                  breed,
87                  by = "ID")
88
89  # Save merged file, PCA_Breed_Merged.txt, using the original directory variab
90  write.table(merged,
91              file = file.path(directory,"PCA_Breed_Merged.txt"),
92              sep = "\t",
93              quote = FALSE,
94              row.names = FALSE)
95
```

```
71
72  # Read file PCA_2PCs.txt using the original directory variable as input
73  pca <- read.table(file.path(directory,"PCA_2PCs.txt"),
74                    header = TRUE,
75                    sep = "\t",
76                    stringsAsFactors = FALSE)
77
78  # Read file BreedComposition.txt using the original directory variable as input
79  breed <- read.table(file.path(directory,"breedComposition.txt"),
80                      header = TRUE,
81                      sep = "\t",
82                      stringsAsFactors = FALSE)
83
84  # Merge by animal ID
85  merged <- merge(pca,
86                  breed,
87                  by = "ID")
88
89  # Save merged file, PCA_Breed_Merged.txt, using the original directory variab
90  write.table(merged,
91              file = file.path(directory,"PCA_Breed_Merged.txt"),
92              sep = "\t",
93              quote = FALSE,
94              row.names = FALSE)
95
```

```r

# loading an already installed library to be used downstream, ggplot2
library(ggplot2)

# Use library ggplot to create a figure
ggplot(merged, aes(x = PC1, y = PC2, color = percentage)) +   scale_color_manual(values = c(
  "B_100-96"    = "blue4", "B_95-75" = "lightseagreen", "B_74-51" = "slateblue1", "B_50" = "purple",
  "B_49-25" = "hotpink4",  "B_24-5" = "hotpink1","B_4-0" = "magenta1")) +
  geom_point(size = 3) + theme_minimal()

```

```
95
96  # loading an already installed library to be used downstream, ggplot2
97  library(ggplot2)
98
99  # Use library ggplot to create a figure
100 ggplot(merged, aes(x = PC1, y = PC2, color = percentage)) +   scale_color_manual(values = c(
101   "B_100-96"    = "blue4", "B_95-75" = "lightseagreen", "B_74-51" = "slateblue1", "B_50" = "purple",
102   "B_49-25" = "hotpink4",  "B_24-5" = "hotpink1","B_4-0" = "magenta1")) +
103   geom_point(size = 3) + theme_minimal()
104
```

Loads a library named “ggplot”

```
95
96  # loading an already installed library to be used downstream, ggplot2
97  library(ggplot2)
98
99  # Use library ggplot to create a figure
100 ggplot(merged, aes(x = PC1, y = PC2, color = percentage)) +   scale_color_manual(values = c(
101   "B_100-96"    = "blue4", "B_95-75" = "lightseagreen", "B_74-51" = "slateblue1", "B_50" = "purple",
102   "B_49-25" = "hotpink4",  "B_24-5" = "hotpink1","B_4-0" = "magenta1")) +
103   geom_point(size = 3) + theme_minimal()
104
```

Loads a library named “ggplot”

Function to create a plot

```r

# loading an already installed library to be used downstream, ggplot2
library(ggplot2)

# Use library ggplot to create a figure
ggplot(merged, aes(x = PC1, y = PC2, color = percentage)) +   scale_color_manual(values = c(
  "B_100-96"    = "blue4", "B_95-75" = "lightseagreen", "B_74-51" = "slateblue1", "B_50" = "purple",
  "B_49-25" = "hotpink4",  "B_24-5" = "hotpink1","B_4-0" = "magenta1")) +
  geom_point(size = 3) + theme_minimal()

```

```
95
96  # loading an already installed library to be used downstream, ggplot2
97  library(ggplot2)
98
99  # Use library ggplot to create a figure
100 ggplot(merged, aes(x = PC1, y = PC2, color = percentage)) +   scale_color_manual(values = c(
101   "B_100-96"    = "blue4", "B_95-75" = "lightseagreen", "B_74-51" = "slateblue1", "B_50" = "purple",
102   "B_49-25" = "hotpink4",  "B_24-5" = "hotpink1","B_4-0" = "magenta1")) +
103   geom_point(size = 3) + theme_minimal()
104
```

It will plot variable PC1 versus variable PC2

```
95
96  # loading an already installed library to be used downstream, ggplot2
97  library(ggplot2)
98
99  # Use library ggplot to create a figure
100 ggplot(merged, aes(x = PC1, y = PC2, color = percentage)) +   scale_color_manual(values = c(
101   "B_100-96"    = "blue4", "B_95-75" = "lightseagreen", "B_74-51" = "slateblue1", "B_50" = "purple",
102   "B_49-25" = "hotpink4",  "B_24-5" = "hotpink1","B_4-0" = "magenta1")) +
103   geom_point(size = 3) + theme_minimal()
104
```

It will plot variable PC1 versus variable PC2

Each dot will be colored by the variable “percentage”

```
95
96  # loading an already installed library to be used downstream, ggplot2
97  library(ggplot2)
98
99  # Use library ggplot to create a figure
100 ggplot(merged, aes(x = PC1, y = PC2, color = percentage)) +  scale_color_manual(values = c(
101   "B_100-96"   = "blue4", "B_95-75" = "lightseagreen", "B_74-51" = "slateblue1", "B_50" = "purple",
102   "B_49-25" = "hotpink4",  "B_24-5" = "hotpink1","B_4-0" = "magenta1")) +
103   geom_point(size = 3) + theme_minimal()
104
```

It will plot variable PC1 versus variable PC2

Each dot will be colored by the variable “percentage”

Pallet of colors to be used

Run the code

populationStructureCode.R*

Source on Save | Run | Source

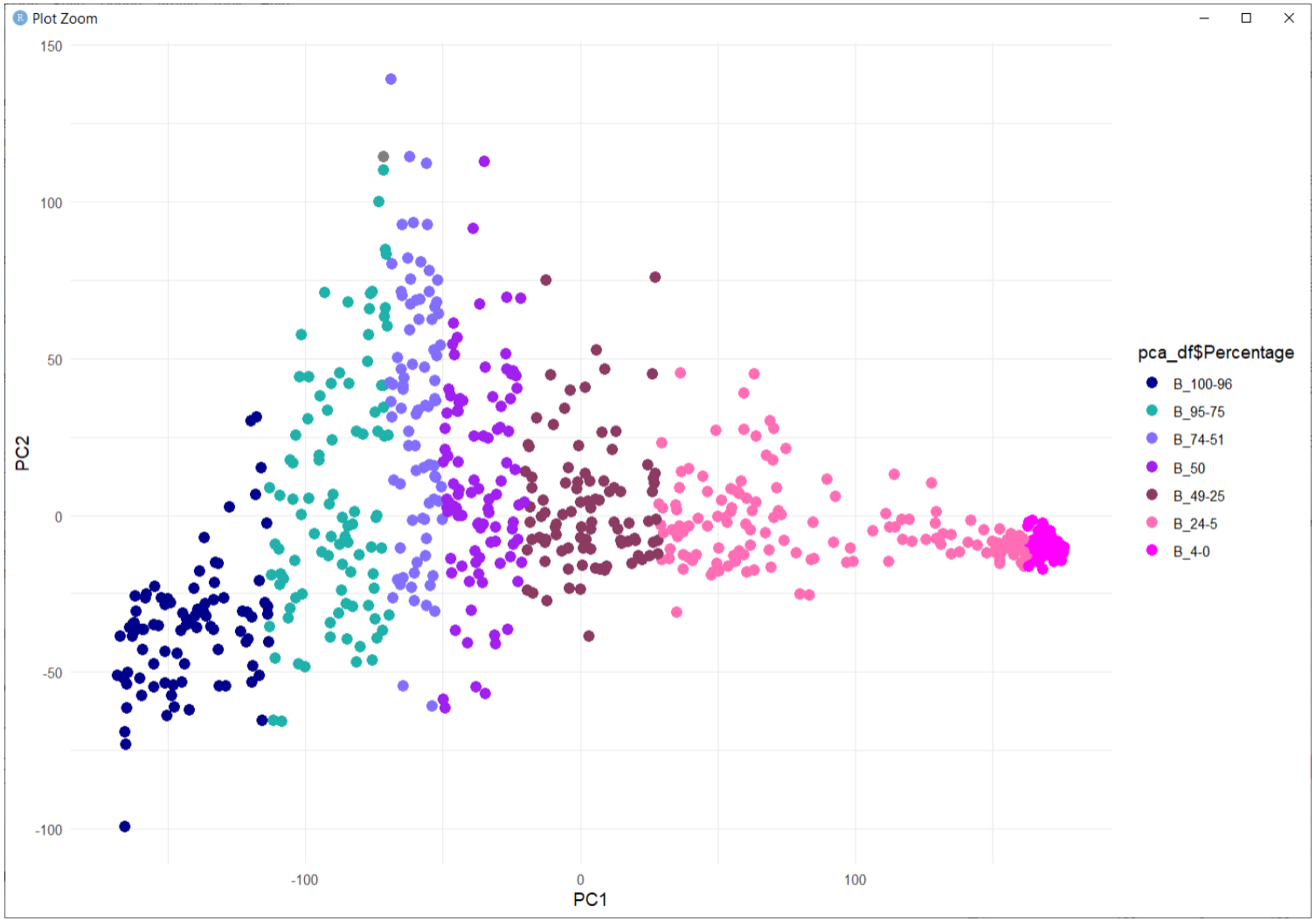
```
83
84  # Merge by animal ID
85  merged <- merge(pca,
86                  breed,
87                  by = "ID")
88
89  # Save merged file, PCA_Breed_Merged.txt, using the original directory variable as input
90  write.table(merged,
91              file = file.path(directory,"PCA_Breed_Merged.txt"),
92              sep = "\t",
93              quote = FALSE,
94              row.names = FALSE)
95
96  # loading an already installed library to be used downstream, ggplot2
97  library(ggplot2)
98
99  # Use library ggplot to create a figure
100 ggplot(merged, aes(x = PC1, y = PC2, color = percentage)) +   scale_color_manual(values = c(
101   "B_100-96"    = "blue4", "B_95-75" = "lightseagreen", "B_74-51" = "slateblue1", "B_50" = "purple",
102   "B_49-25" = "hotpink4",   "B_24-5" = "hotpink1","B_4-0" = "magenta1")) +
103   geom_point(size = 3) + theme_minimal()
104
```

populationStructureCode.R*

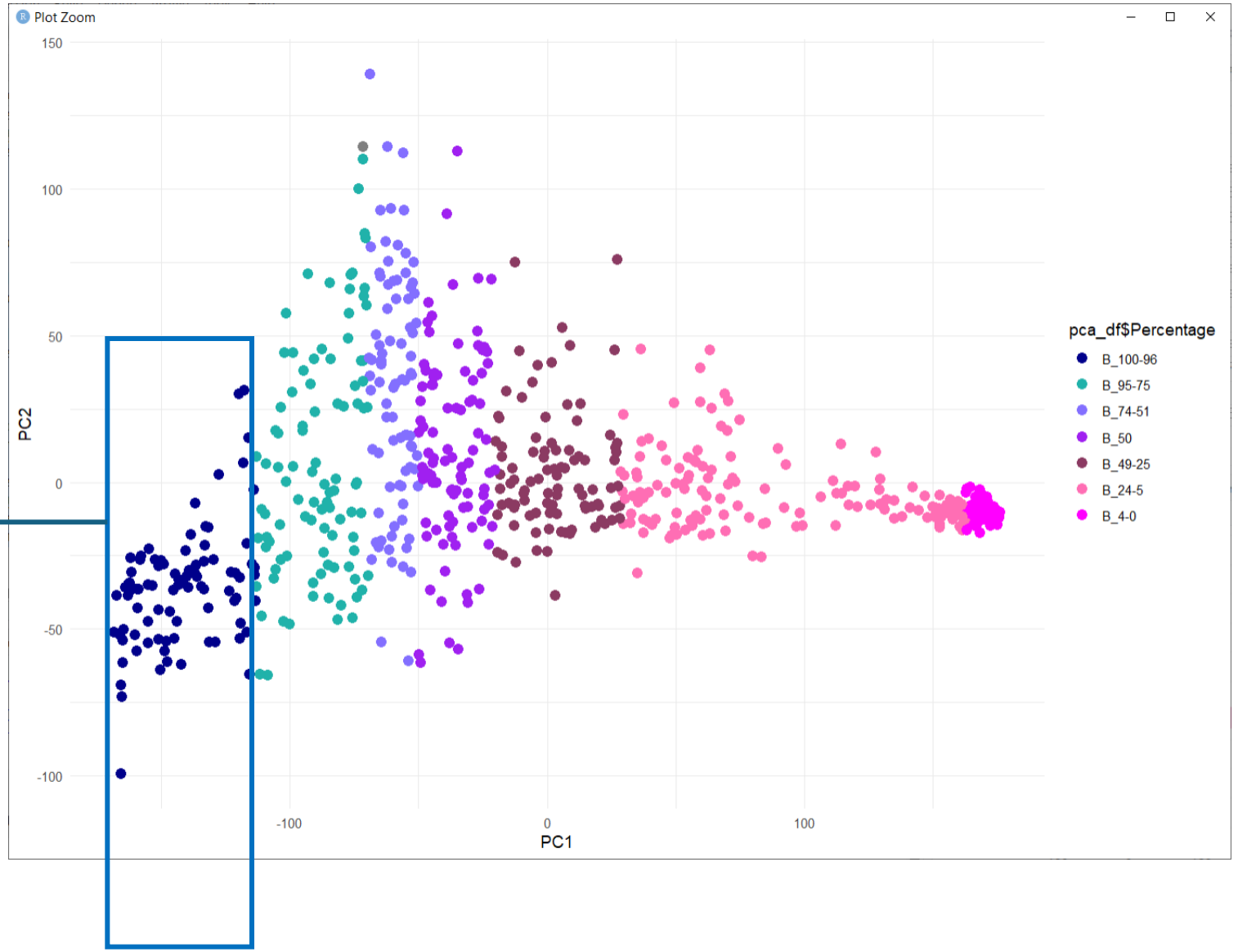
Source on Save | Run | Source

```
83
84  # Merge by animal ID
85  merged <- merge(pca,
86                  breed,
87                  by = "ID")
88
89  # Save merged file, PCA_Breed_Merged.txt, using the original directory variable as input
90  write.table(merged,
91              file = file.path(directory,"PCA_Breed_Merged.txt"),
92              sep = "\t",
93              quote = FALSE,
94              row.names = FALSE)
95
96  # loading an already installed library to be used downstream, ggplot2
97  library(ggplot2)
98
99  # Use library ggplot to create a figure
100 ggplot(merged, aes(x = PC1, y = PC2, color = percentage)) +   scale_color_manual(values = c(
101   "B_100-96"    = "blue4", "B_95-75" = "lightseagreen", "B_74-51" = "slateblue1", "B_50" = "purple",
102   "B_49-25" = "hotpink4",   "B_24-5" = "hotpink1","B_4-0" = "magenta1")) +
103   geom_point(size = 3) + theme_minimal()
```

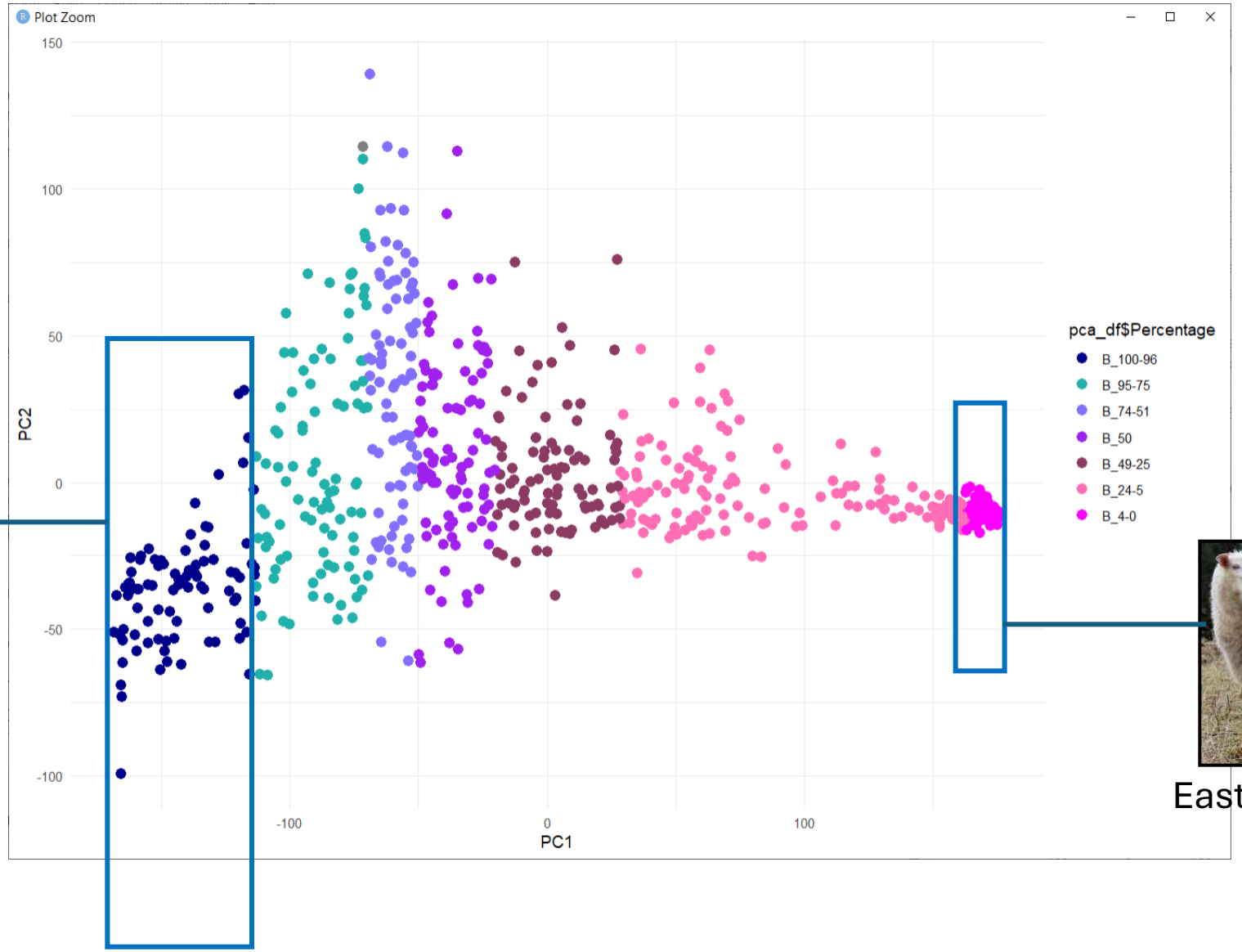
Results


Plot Zoom
150
100
50
0
-50
-100
PC2
-100
0
100
PC1
pca_df$Percentage
B_100-96
B_95-75
B_74-51
B_50
B_49-25
B_24-5
B_4-0

Suffolk sheep



Suffolk sheep



East Friesian sheep

Suffolk sheep

50:50

East Friesian sheep


Alleles influencing meat production
50:50
Suffolk sheep
East Friesian sheep
PC1
PC2
pca_df$Percentage
B_100-96
B_95-75
B_74-51
B_50
B_49-25
B_24-5
B_4-0

Suffolk sheep

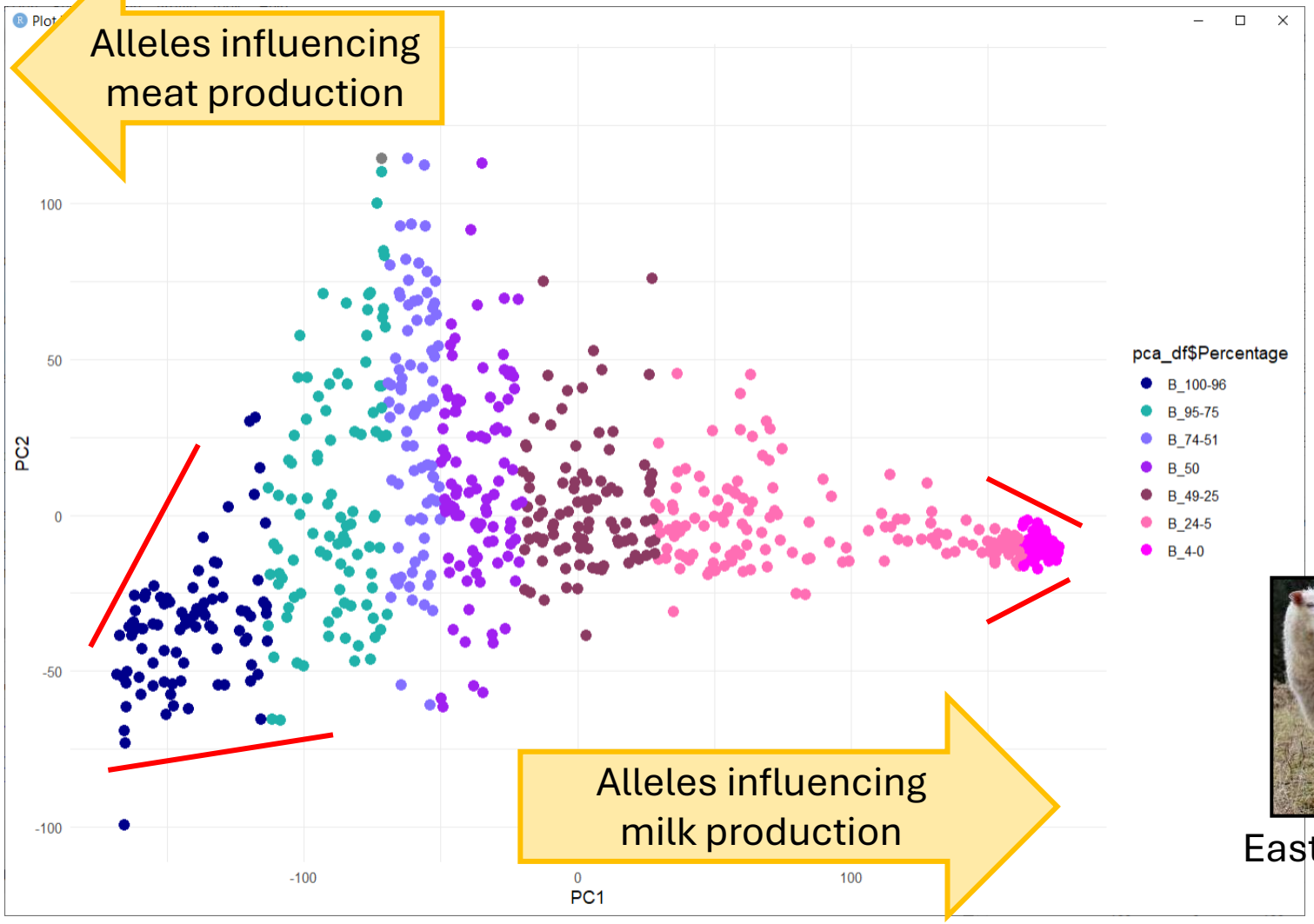
Alleles influencing meat production

50:50

Alleles influencing milk production

East Friesian sheep

Suffolk sheep



East Friesian sheep

50:50

Alleles influencing meat production
Alleles influencing milk production
PC1
PC2
100
50
0
-50
-100
-100
0
100
pca_df$Percentage
B_100-96
B_95-75
B_74-51
B_50
B_49-25
B_24-5
B_4-0

Activities:

1. Verify the breed composition of an animal for a producer interested in high-performance dairy females.

2. Create a grayscale figure for the original breed composition analysis using the provided code.

3. Create a plot using PC2 vs PC3 instead of the original PC1 vs PC2

Activities:

1. Verify the breed composition of an animal for a producer interested in high-performance dairy females.

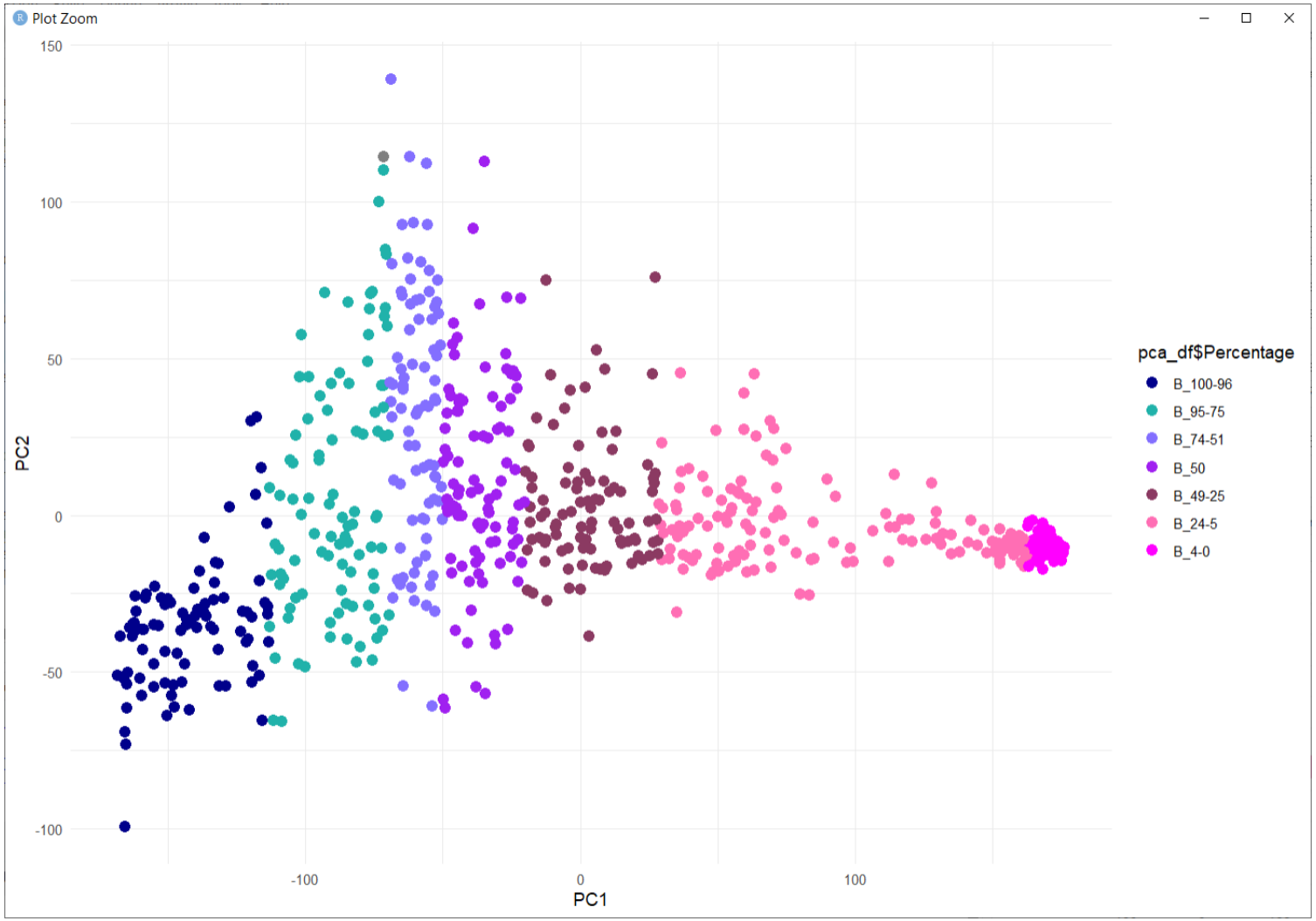
2. Create a grayscale figure for the original breed composition analysis using the provided code.

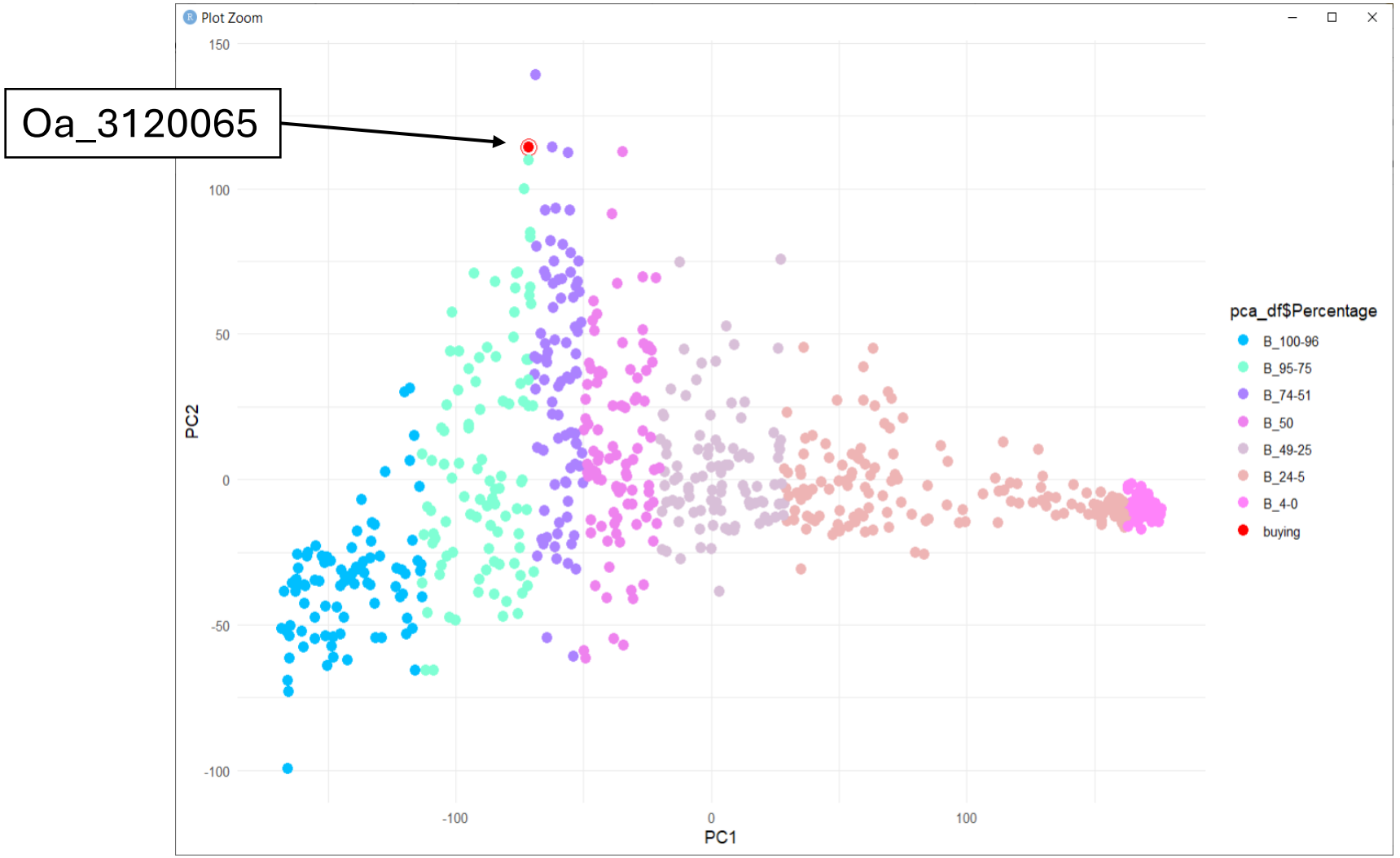
3. Create a plot using PC2 vs PC3 instead of the original PC1 vs PC2

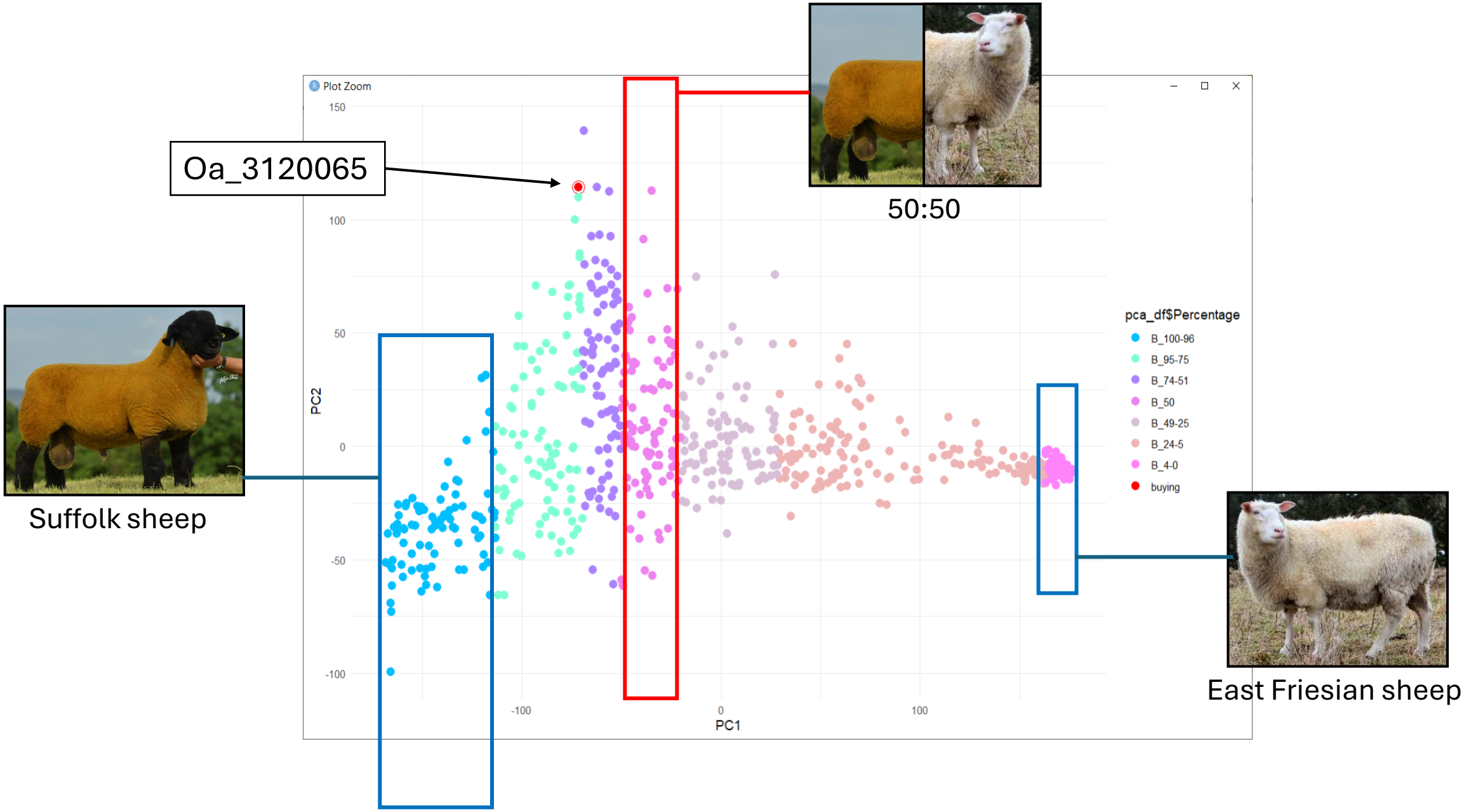
- A dairy producer wants to buy an animal from this crossbred/purebred population. They selected Oa_3120065 and were told she was a suitable purebred for their system.

- A dairy producer wants to buy an animal from this crossbred/purebred population. They selected Oa_3120065 and were told she was a suitable purebred for their system. Is this the case based on the breed composition analysis performed?

```
###################################
###### Activities ###############
###################################

######################
######    1     ######
######################

# Objective: Verify the breed composition of an animal for a producer interested in high performance dairy females

# Read file breedCompositionBuying.txt using the original directory variable as input
breed <- read.table(file.path(directory,"breedCompositionBuying.txt"), header = TRUE,
                    sep = "\t", stringsAsFactors = FALSE)

# Merge by animal ID objects pca and breed using animal ID
merged <- merge(pca, breed, by = "ID")

# Save merged file using the original directory variable as input
write.table(merged, file = file.path(directory,"PCA_Breed_MergedBuying.txt"), sep = "\t", quote = FALSE, row.names = FALSE)

# Use library ggplot to create a figure
ggplot(merged, aes(x = PC1, y = PC2, color = percentage)) +
  scale_color_manual(values = c("B_100-96"   = "deepskyblue", "B_95-75" = "aquamarine1", "B_74-51" = "mediumpurple1",
                                "B_50" = "violet", "B_49-25" = "thistle",  "B_24-5" = "rosybrown2","B_4-0" = "orchid1",
                                "buying" = "red")) +
  geom_point(data = subset(merged, ID == "Oa_3120065"), color = "red", shape = 21, size = 5) +
  geom_point(size = 3) + theme_minimal()
```


Plot Zoom
150
100
50
0
-50
-100
PC2
-100
0
100
PC1
pca_df$Percentage
B_100-96
B_95-75
B_74-51
B_50
B_49-25
B_24-5
B_4-0


Oa_3120065
Plot Zoom
150
100
50
0
-50
-100
PC2
-100
0
100
PC1
pca_df$Percentage
B_100-96
B_95-75
B_74-51
B_50
B_49-25
B_24-5
B_4-0
buying


Oa_3120065
50:50
Suffolk sheep
East Friesian sheep
Plot Zoom
PC2
PC1
150
100
50
0
-50
-100
-100
0
100
pca_df$Percentage
B_100-96
B_95-75
B_74-51
B_50
B_49-25
B_24-5
B_4-0
buying

Is this animal suitable for a dairy production system?

50:50

Oa_3120065

Suffolk sheep

East Friesian sheep

Activities:

1. Verify the breed composition of an animal for a producer interested in high-performance dairy females.

2. Create a grayscale figure for the original breed composition analysis using the provided code.

3. Create a plot using PC2 versus PC3

```r

######################
######    2     ######
######################

# Objective: Create a figure in gray scale using the following code

# Use library ggplot to create a figure
ggplot(merged, aes(x = PC1, y = PC2, color = percentage)) +
  scale_color_manual(values = c("B_100-96"   = "deepskyblue", "B_95-75" = "aquamarine1", "B_74-51" = "mediumpurple1",
                                "B_50" = "violet", "B_49-25" = "thistle",  "B_24-5" = "rosybrown2","B_4-0" = "orchid1",
                                "buying" = "red")) +
  geom_point(data = subset(merged, ID == "Oa_3120065"), color = "red", shape = 21, size = 5) +
  geom_point(size = 3) + theme_minimal()

```

Activities:

1. Verify the breed composition of an animal for a producer interested in high-performance dairy females.

2. Create a grayscale figure for the original breed composition analysis using the provided code.

3. Create a plot using PC2 versus PC3

```
146
147 ######################
148 ######    3     ######
149 ######################
150
151 # Objective: Create a plot using PC2 versus PC3
152
```

Lincoln University of Missouri | My Interactive Sessions - RSS C | RStudio Server | Ask Gemini

ondemand.rnet.missouri.edu/rnode/c159.mgmt.hellbender/57973/

File Edit Code View Plots Session Build Debug Profile Tools Help

lincolnu-lealgutierrezj

Go to file/function | Addins | Project: (None)

Console Terminal Jobs

R 4.2.0 · ~/

```
R version 4.2.0 (2022-04-22) -- "Vigorous Calisthenics"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.


> getwd
function ()
.Internal(getwd())
<bytecode: 0x564cfc2cc6a0>
<environment: namespace:base>
> getwd()
[1] "/home/lincolnu-lealgutierrezj"
Connected to your session in progress, last started 2026-Jul-02 13:28:38 UTC (37 minutes ago)
>
```

Environment History Connections Tutorial

Import Dataset | 178 MiB | List

R | Global Environment

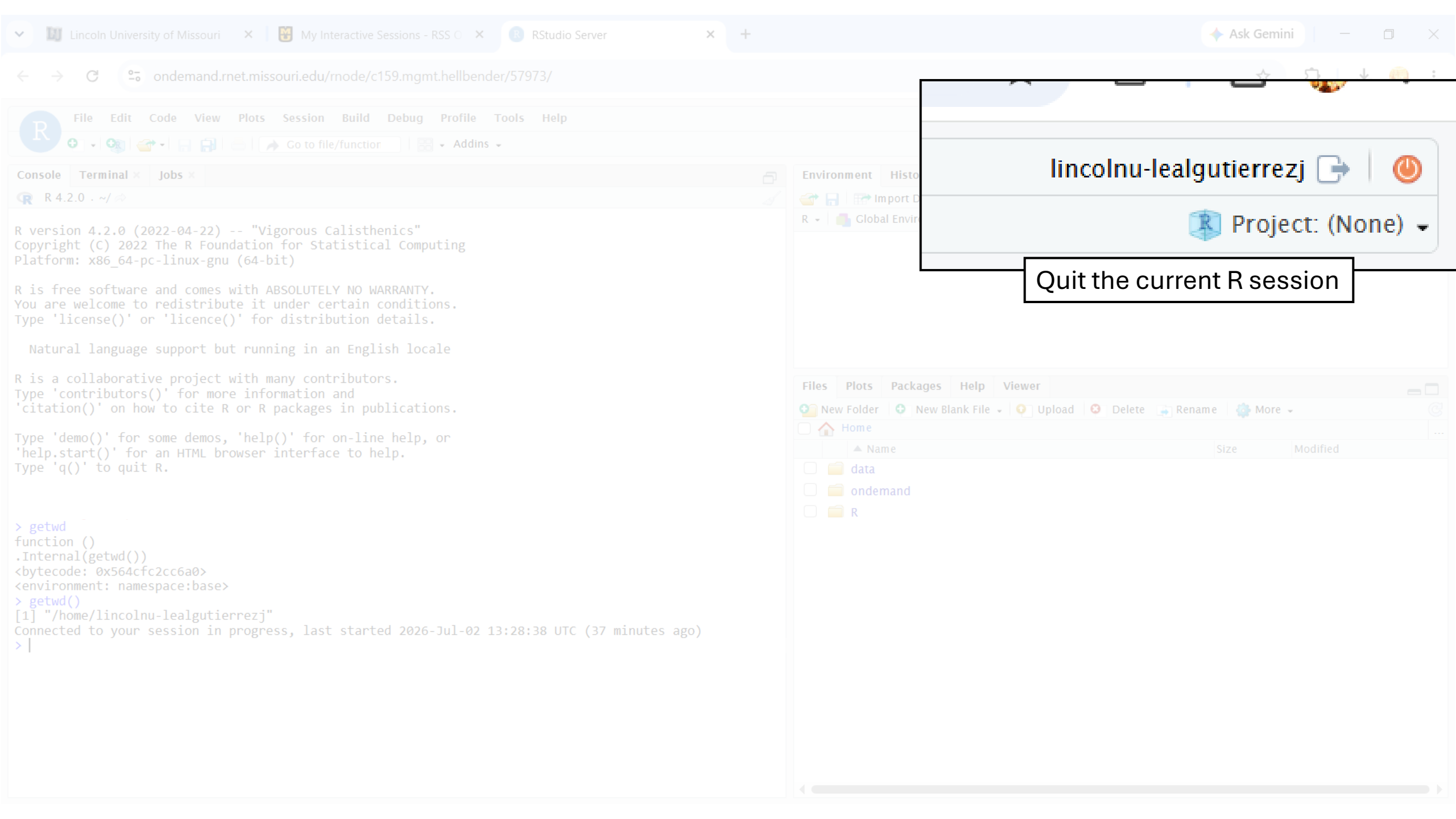
Environment is empty

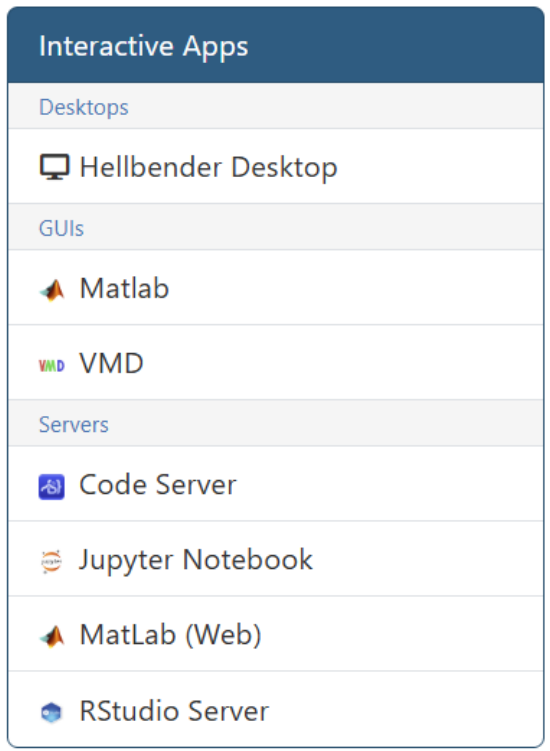
Files Plots Packages Help Viewer

New Folder | New Blank File | Upload | Delete | Rename | More

Home

Name	Size	Modified
data		
ondemand		
R		


lincolnu-lealgutierrezj
Project: (None)
Quit the current R session



Hellbender Desktop (15283304)

Completed | | |

Created at: 2026-07-23 16:08:33 CDT

Delete

Session ID: 7662a6b1-f81c-465d-9652-c6e9606845fc

For debugging purposes, this card will be retained for 6 more days

RStudio Server (15283292)

Completed | | |

Created at: 2026-07-23 16:06:53 CDT

Delete

Session ID: 4aa70e52-4190-49d0-9e63-73e3b0c2e0b7

Conclusions

We presented an example of how HPC resources can be applied to address complex genomic questions.

www.modernfarmer.com

We presented an example of how HPC resources can be applied to address complex genomic questions. Can I determine the breed composition of individuals in a crossbred sheep population?

www.modernfarmer.com

We presented an example of how HPC resources can be applied to address complex genomic questions. Can I determine the breed composition of individuals in a crossbred sheep population? Can I use this information to draw productive recommendations to farmers?

www.modernfarmer.com

Acknowledgment

This work is supported by the Research Capacity Fund (Evans-Allen) project award no. 7009120, from the U.S. Department of Agriculture's National Institute of Food and Agriculture.

USDA National Institute of Food and Agriculture
U.S. DEPARTMENT OF AGRICULTURE

LINCOLN UNIVERSITY MISSOURI
College of Agriculture

Laboratory of Genomic Dissection
Lincoln University of Missouri

www.labofgenomicdissection.com

Thanks!!
Questions??