

Parallel Programming in the Classroom

Analysis of Genome Data



Karl Frinkle - Mike Morris
Parallel Programming Seminar CS4973
Spring 2015



OU Supercomputer Symposium – September 23, 2015

We Started Here . . .



OU Supercomputer Symposium – September 23, 2015

I thought it might be appropriate to translate the title of our paper and explain what the whole exercise is about.

1. That's why we're here!
2. We researched the latest ACM accreditation guidelines and found that parallel computing is suggested in multiple areas of the CS curricula.
3. That is what we meant by interdisciplinary.
4. Small liberal arts colleges are notorious for not showering money on the sciences.

We Started Here . . .



OU Supercomputer Symposium – September 23, 2015

I thought it might be appropriate to translate the title of our paper and explain what the whole exercise is about.

1. That's why we're here!
2. We researched the latest ACM accreditation guidelines and found that parallel computing is suggested in multiple areas of the CS curricula.
3. That is what we meant by interdisciplinary.
4. Small liberal arts colleges are notorious for not showering money on the sciences.

We Progressed To . . .



**Our first cluster
– made from
junk – but it
worked!**



OU Supercomputer Symposium – September 23, 2015

This is where the 3-semester part came in.

Course 1: We needed a crash course in C that was tailored to parallel thinking, so we tweaked a standard introductory C course.

Course 2: After some MPI success it seemed like a good idea to introduce CUDA.

Course 3: After a year of fairly intense programming, we felt like students could use some validation as to why we do this in the first place. We just happened to have a real research-level math problem at hand.

Then Found LittleFe . . .



OU Supercomputer Symposium – September 23, 2015

Next We Used Sooner . . .



OU Supercomputer Symposium – September 23, 2015

This is the Oklahoma Supercomputing Center for Research and Education's old supercomputer, Sooner. It's been replaced by Boomer . . .

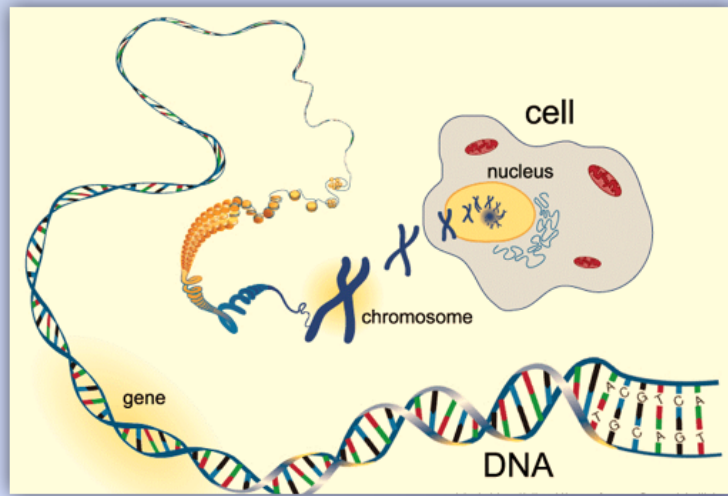
Now We Have . . .



OU Supercomputer Symposium – September 23, 2015

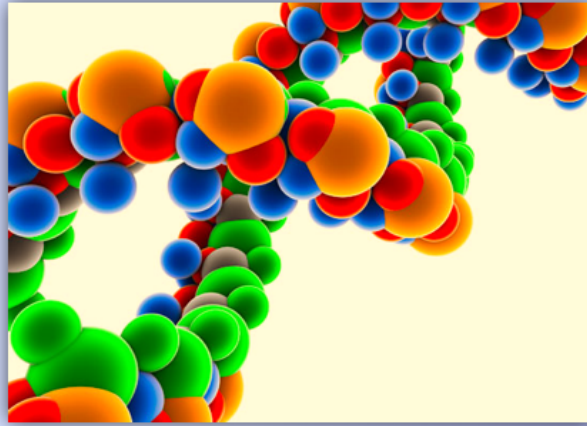
This is the Oklahoma Supercomputing Center for Research and Education's old supercomputer, Sooner. It's been replaced by Boomer . . .

Analyzing Genome Data



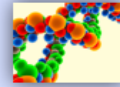
OU Supercomputer Symposium – September 23, 2015

PHASE 1 – Write Code



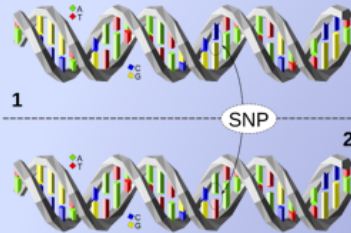
OU Supercomputer Symposium – September 23, 2015

PHASE 1 – Write Code



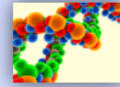
Definitions:

SNP: single-nucleotide polymorphism
pronounced “snip” is a DNA sequence
commonly varying within a population



OU Supercomputer Symposium – September 23, 2015

PHASE 1 – Write Code



Definitions:

SNP: single-nucleotide polymorphism

pronounced “snip” is a DNA sequence
commonly varying within a population

rsid: Reference SNP cluster ID

access number used to refer
to specific SNPs



OU Supercomputer Symposium – September 23, 2015

PHASE 1 – Write Code

- Harvard PGP* Database
- 23andME

# rsid	chromosome	position	genotype
...			
rs12564807	1	734462	AA
rs3131972	1	752721	GG
rs148828841	1	760998	CC
rs12124819	1	776546	AA ...

* PGP - Personal Genome Project



OU Supercomputer Symposium – September 23, 2015

PHASE 1 – Write Code

Harvard PGP Database

Per person, there were about 1,000,000 snips.

# rsid	chromosome	position	genotype
...			
rs12564807	1	734462	AA
rs3131972	1	752721	GG
rs148828841	1	760998	CC
rs12124819	1	776546	AA ...



OU Supercomputer Symposium – September 23, 2015

PHASE 1 – Write Code

- We started with 200 profiles
- Then gave 'em names
- That was about 5G of data

Per person, there
were about
1,000,000 snips.

# rsid	chromosome	position	genotype
...			
rs12564807	1	734462	AA
rs3131972	1	752721	GG
rs148828841	1	760998	CC
rs12124819	1	776546	AA ...



OU Supercomputer Symposium – September 23, 2015

!! Important Info for HPC Instructors !!

- **5G of data isn't necessarily "big"**
- **Our clusters are significantly small**



!! Important Info for HPC Instructors !!

- **5G of data isn't necessarily "big"**
- **Our clusters are significantly small**
 - **We're teaching concept and techniques**
 - **We can easily scale up to Boomer**



PHASE 1 – Write Code

Several programs begged to be written, and all were great candidates for parallelization.

- Search for a particular *rsid* for a given person
- Ditto for many persons
- Both of the above for a collection of *rsids*
- Compare 2 persons' makeup
 - we used a sliding window algorithm
- Compare many persons' makeup



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

PENELOPE KARDASHIAN

i6019305	1	891343	GG
rs13303106	1	891945	AG
i6019306	1	894379	GG
rs13303010	1	894573	AG
i6019308	1	897792	CC
i6019309	1	898082	AA
rs6696281	1	903104	CT
i6019310	1	905681	CC
i6019311	1	906114	CC
i6019312	1	907666	AA
i6060381	1	909238	CG
rs2340592	1	910935	--
rs13303118	1	918384	GT
rs78164078	1	921071	GG
rs6665000	1	924898	AC
rs2341362	1	927309	CC
rs9777703	1	928836	CT

STONEY BURKE

i6019305	1	891343	GG
rs13303106	1	891945	AG
i6019306	1	894379	GG
rs13303010	1	894573	AA
i6019308	1	897792	CC
i6019309	1	898082	AA
rs6696281	1	903104	CT
i6019310	1	905681	CC
i6019311	1	906114	CC
i6019312	1	907666	AA
i6060381	1	909238	CG
rs2340592	1	910935	GG
rs13303118	1	918384	GT
rs78164078	1	921071	GG
rs6665000	1	924898	AC
rs2341362	1	927309	CC
rs9777703	1	928836	TT



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

PENELOPE KARDASHIAN

```
i6019305 1 891343 GG
rs13303106 1 891945 AG
i6019306 1 894379 GG
rs13303010 1
i6019308 1
i6019309 1
rs6696281 1
i6019310 1
i6019311 1 906114 CC
i6019312 1 907666 AA
i6060381 1 909238 CG
rs2340592 1 910935 --
rs13303118 1 918384 GT
rs78164078 1 921071 GG
rs6665000 1 924898 AC
rs2341362 1 927309 CC
rs9777703 1 928836 CT
```

STONE BURKE

```
i6019305 1 891343 GG
rs13303106 1 891945 AG
i6019306 1 894379 GG
894573 AA
897792 CC
898082 AA
903104 CT
905681 CC
906114 CC
i6019311 1 907666 AA
i6019312 1 907666 AA
i6060381 1 909238 CG
rs2340592 1 910935 GG
rs13303118 1 918384 GT
rs78164078 1 921071 GG
rs6665000 1 924898 AC
rs2341362 1 927309 CC
rs9777703 1 928836 TT
```

If one person's rsid was unrecorded, both were tossed out.



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

PENELOPE KARDASHIAN

i6019305	1	891343	GG
rs13303106	1	891945	AG
i6019306	1	894379	GG
rs13303010	1	894573	AG
i6019308	1	897792	CC
i6019309	1	898082	AA
rs6696281	1	903104	CT
i6019310	1	905681	CC
i6019311	1	906114	CC
i6019312	1	907666	AA
i6060381	1	909238	CG
rs13303118	1	918384	GT
rs78164078	1	921071	GG
rs6665000	1	924898	AC
rs2341362	1	927309	CC
rs9777703	1	928836	CT

STONE BURKE

i6019305	1	891343	GG
rs13303106	1	891945	AG
i6019306	1	894379	GG
rs13303010	1	894573	AA
i6019308	1	897792	CC
i6019309	1	898082	AA
rs6696281	1	903104	CT
i6019310	1	905681	CC
i6019311	1	906114	CC
i6019312	1	907666	AA
i6060381	1	909238	CG
rs13303118	1	918384	GT
rs78164078	1	921071	GG
rs6665000	1	924898	AC
rs2341362	1	927309	CC
rs9777703	1	928836	TT

FIRST
VARIANCE



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

PENELOPE KARDASHIAN

i6019305	1	891343	GG
rs13303106	1	891945	AG
i6019306	1	894379	GG
rs13303010	1	894573	AG
i6019308	1	897792	CC
i6019309	1	898082	AA
rs6696281	1	903104	CT
i6019310	1	905681	CC
i6019311	1	906114	CC
i6019312	1	907666	AA
i6060381	1	909238	CG
rs13303118	1	918384	GT
rs78164078	1	921071	GG
rs6665000	1	924898	AC
rs2341362	1	927309	CC
rs9777703	1	928836	CT

STONEY BURKE

i6019305	1	891343	GG
rs13303106	1	891945	AG
i6019306	1	894379	GG
rs13303010	1	894573	AA
i6019308	1	897792	CC
i6019309	1	898082	AA
rs6696281	1	903104	CT
i6019310	1	905681	CC
i6019311	1	906114	CC
i6019312	1	907666	AA
i6060381	1	909238	CG
rs13303118	1	918384	GT
rs78164078	1	921071	GG
rs6665000	1	924898	AC
rs2341362	1	927309	CC
rs9777703	1	928836	TT

VARIANCES



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

PENELOPE KARDASHIAN

i6019305	1	891343	GG
rs13303106	1	891945	AG
i6019306	1	894379	GG
rs13303010	1	894573	AG
i6019308	1	897792	CC
i6019309	1		
rs6696281	1		
i6019310	1		
i6019311	1		
i6019312	1		
i6060381	1		
rs13303118	1		
rs78164078	1	921071	GG
rs6665000	1	924898	AC
rs2341362	1	927309	CC
rs9777703	1	928836	CT

STONEY BURKE

i6019305	1	891343	GG
rs13303106	1	891945	AG
i6019306	1	894379	GG
rs13303010	1	894573	AA
i6019308	1	897792	CC
		898082	AA
		903104	CT
		905681	CC
		906114	CC
		907666	AA
		909238	CG
		918384	GT
		921071	GG
rs78164078	1	921071	GG
rs6665000	1	924898	AC
rs2341362	1	927309	CC
rs9777703	1	928836	TT

A set number of variances were allowed in a given "block" size w/o causing a "difference".

VARIANCES



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

--
--
@@
--
--
--
--
--
--
--
--
@@
@@
--
@@
@@
--
--
--
--
--



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

This represents *rsid* locations. “@@” means difference. “—” means agreement.

--
--
@@
--
--
--
--
--
--
--
@@
@@
--
--
@@
@@
--
--
--
--
--



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

This represents *rsid* locations. “@@” means difference. “—” means agreement.

--
--
@@
--
--
--
--
--
--
--
@@
@@
--
@@
--
--
--
--
--

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

This represents *rsid* locations. “@@” means difference. “—” means agreement.

--
--
@@
--
--
--
--
--
--
@@
@@
--
--

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.

It's a “best case” algorithm. If an *rsid* is ever in a match box of 8, then it is forever a match.

--
--
--



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example,
if 2 or more do
not agree in a
group of 8, then
record a
difference.
Otherwise it is
recorded as a
match.

--
--
@@
--
--
--
--
--
--
--
@@
@@
--
@@
@@
--
--
--
--
--



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.

@@

@@

@@

@@

@@

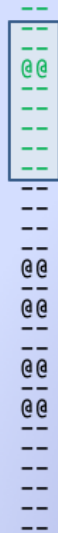
Green = Match
Red = Difference



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.



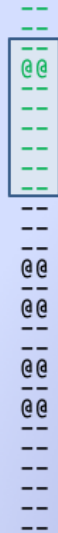
Green = Match
Red = Difference



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.



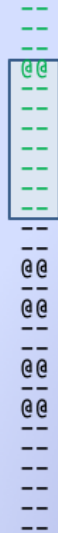
Green = Match
Red = Difference



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.

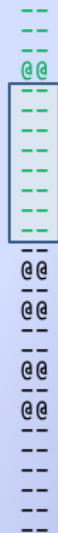


Green = Match
Red = Difference



Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.



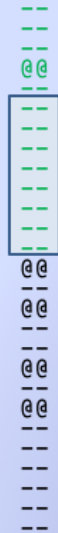
Green = Match
Red = Difference



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.



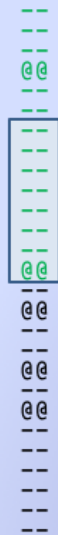
Green = Match
Red = Difference



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.



Green = Match
Red = Difference



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.

--
--
@@
--
--
--
--
--
@@
@@
--
@@
--
@@
--
--
--
--
--

Green = Match
Red = Difference



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.



Green = Match
Red = Difference

It's a "best case" algorithm. If an *rsid* is ever in a match box of 8, then it is forever a match.



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.



Green = Match
Red = Difference

It's a "best case" algorithm. If an *rsid* is ever in a match box of 8, then it is forever a match.



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example,
if 2 or more do
not agree in a
group of 8, then
record a
difference.
Otherwise it is
recorded as a
match.



Green = Match
Red = Difference



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.



Green = Match
Red = Difference



Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.

--
--
@@
--
--
--
--
--
--
--
@@
@@
@@
@@
@@
--
--
--
--
--

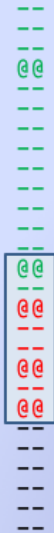
Green = Match
Red = Difference



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.



Green = Match
Red = Difference



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.

--
--
@@
--
--
--
--
--
--
--
@@
@@
--
@@
--
@@
--
--
--
--

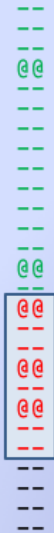
Green = Match
Red = Difference



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.



Green = Match
Red = Difference



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.



Green = Match
Red = Difference



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.

--
--
@@
--
--
--
--
--
--
--
@@
@@
--
@@
--
@@
--
--
--
--
--
--

Green = Match
Red = Difference

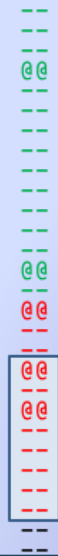


OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.

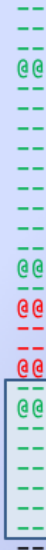
Green = Match
Red = Difference



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.

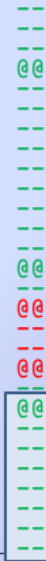


Green = Match
Red = Difference



Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.



Green = Match
Red = Difference



OU Supercomputer Symposium – September 23, 2015

Sliding Window Technique . . .

For this example, if 2 or more do not agree in a group of 8, then record a difference. Otherwise it is recorded as a match.

--
--
@@
--
--
--
--
--
--
--
--
@@
@@
--
--
@@
--
--
--
--
--
--
--

Green = Match
Red = Difference



OU Supercomputer Symposium – September 23, 2015

PHASE 2 – Web GUI



OU Supercomputer Symposium – September 23, 2015

PHASE 2 – Web GUI



Genome Data Analyzer

Single Person Multiple People Group Allele Search Two Person Comparison System Monitor

Select a person:

LASANDRA SIPPAL ▾

Count No Calls

Enter RSID(s) for detailed information:

(Type rsid and press enter to enter more)

Here are some for easy testing: rs586265 rs1676428 rs1253105 rs522929 rs17841327 rs2857887

Submit Query

PHASE 2 – Web GUI



Genome Data Analyzer

Single Person Multiple People Group Allele Search Two Person Comparison System Monitor

Select a person:

LASANDRA SIPPAL

Count No Calls

Enter RSID(s) for detailed information:

(Type rsid and press enter to enter more)

Here are some for easy testing: rs586265 rs1676428 rs1253105 rs522929 rs178413

Submit Query

Currently
we have 5
options.

Web GUI – Single Person



Genome Data Analyzer

Single Person Multiple People Group Allele Search Two Person Comparison System Monitor

Select a person:

IVY CONTRERES ▾

Count No Calls

Enter RSID(s) for detailed information:

(Type rsid and press enter to enter more)
Here are some for easy testing: rs586265 rs1676428 rs1253105 rs522929 rs17841327 rs2857887

rs1253105 x

Submit Query

Web GUI – Single Person



Genome Data
Analyzer

Output

CONTRERES, IVY

RSID

Allele

rs1253105

AC

[Lookup rs1253105 on snpedia.com](#)

No Calls

All no-call variant types indicate that the sequence could not be fully resolved, either because of limited or no information, or because of contradictory information.

3828

Green is Low | Yellow is Average | Red is High

Web GUI – Multiple Person



Genome Data Analyzer

Single Person Multiple People **Group Allele Search** Two Person Comparison System Monitor

Select some people:

DOREEN ZAGACKI x

JESUSITA SPEVACEK x

CREOLA ESPINA x

Count No Calls

Enter RSID(s) for detailed information:

(Type rsid and press enter to enter more)

Here are some for easy testing: rs586265 rs1676428 rs1253105 rs522929 rs17841327 rs2857887

rs522929 x

Submit Query

Web GUI – Multiple Person



Genome Data Analyzer

Output

RSID	ZAGACKI, D.	ESPINA, C.	SPEVACEK, J.	
rs2857887	AA	AA	AG	Lookup rs2857887 on snpedia.com
No Calls Count	6956	3624	3124	

[New Search](#)

Web GUI – Group Allele



Genome Data Analyzer

[Single Person](#) [Multiple People](#) [Group Allele Search](#) [Two Person Comparison](#) [System Monitor](#)

Enter RSID(s) for detailed information*:

*This may take a few seconds as it compares all data on all nodes.

(Type rsid and press enter to enter more)

Here are some for easy testing: rs586265 rs1676428 rs1253105 rs522929 rs17841327 rs2857887

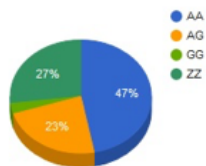
rs586265 x rs6019305 x
rs6090381 x rs9777703 x
rs6665000 x

Submit Query

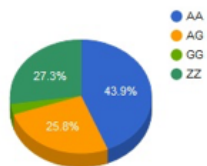
Web GUI – Group Allele

rs586265

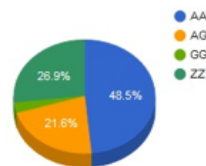
Total People: 200



Total Male: 66

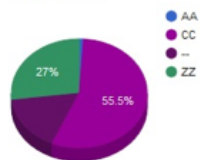


Total Female: 134

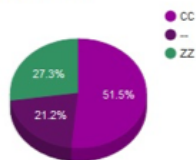


rs522929

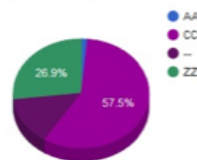
Total People: 200



Total Male: 66



Total Female: 134



Web GUI – Two Person Compare

Single Person Multiple People Group Allele Search **Two Person Comparison** System Monitor

Select two people:

DOREEN ZAGACKI

IVY CONTRERES

Two Allele Comparison

Ignore No Calls

Window Size: 76

Max mismatches per window: 5

Mbp tolerance: 1.72

Chromosomes to compare:

1 x	2 x	3 x	4 x	5 x	6 x	7 x	8 x
9 x	10 x	11 x	12 x	13 x	14 x	15 x	
16 x	17 x	18 x	19 x	20 x	21 x	22 x	

Submit Query

Web GUI – Two Person Compare



Genome Data
Analyzer

Output

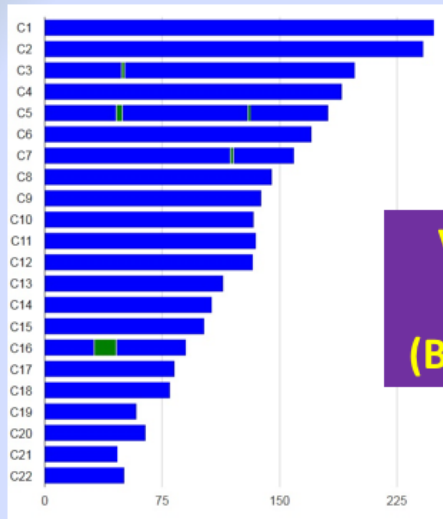
DOREEN ZAGACKI

IVY CONTRERES

Two Allele Comparison
5 Mismatches allowed per window size of 76
only showing results with matches greater than 1.72 mbp in length.
Ignore No Calls

Selected Chromosome Output:

Web GUI – Two Person Compare

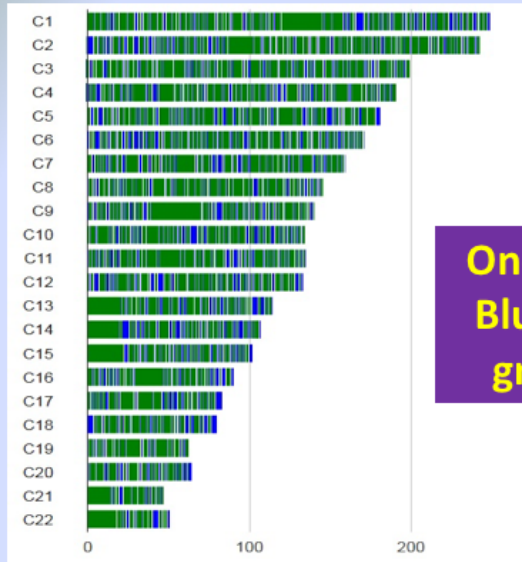


**Very different –
no connection.
(Blue is no match.)**



OU Supercomputer Symposium – September 23, 2015

Web GUI – Two Person Compare

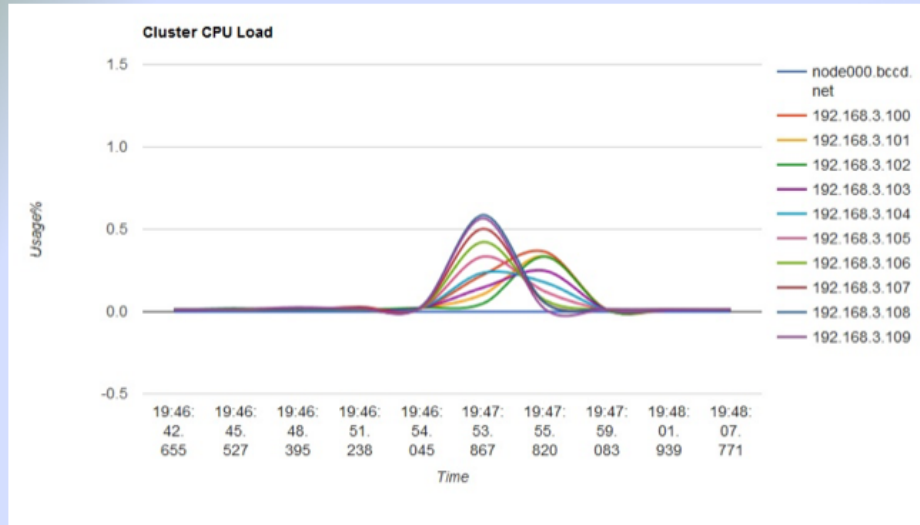


One-allele search.
Blue is no match,
green is match.



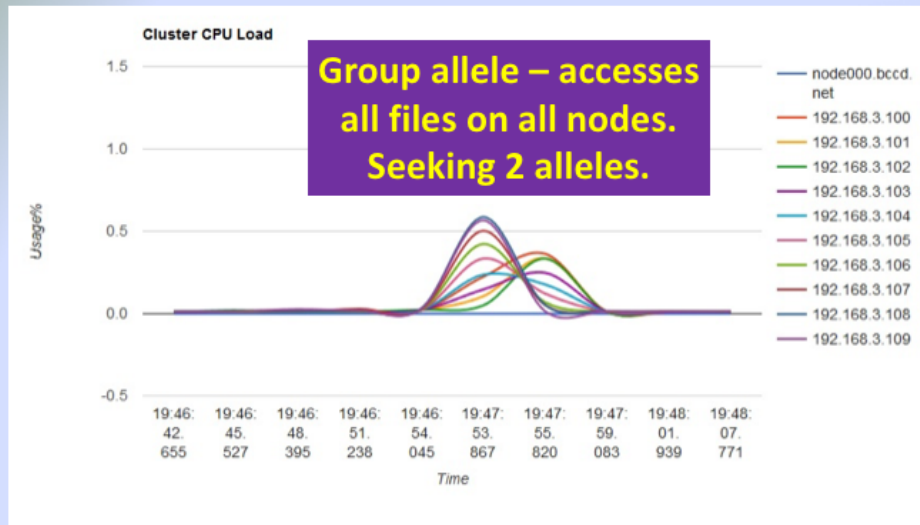
OU Supercomputer Symposium – September 23, 2015

Web GUI – System Monitor



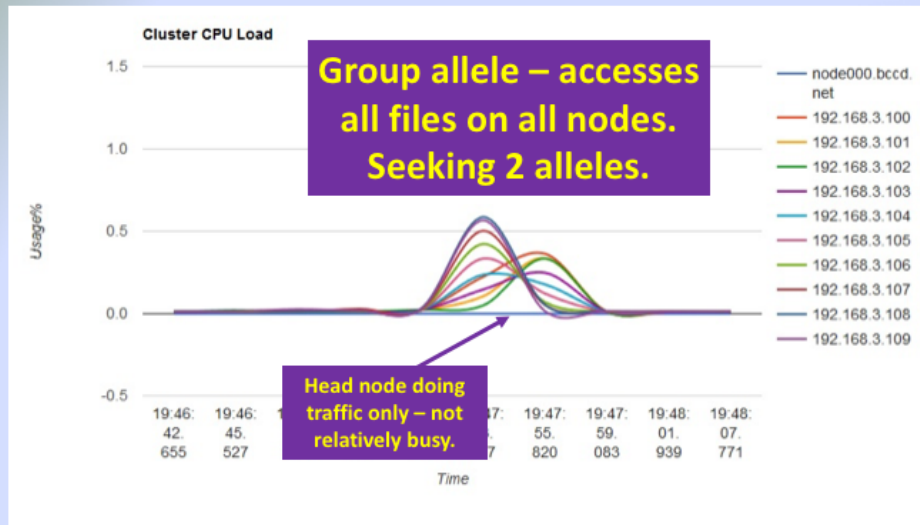
OU Supercomputer Symposium – September 23, 2015

Web GUI – System Monitor



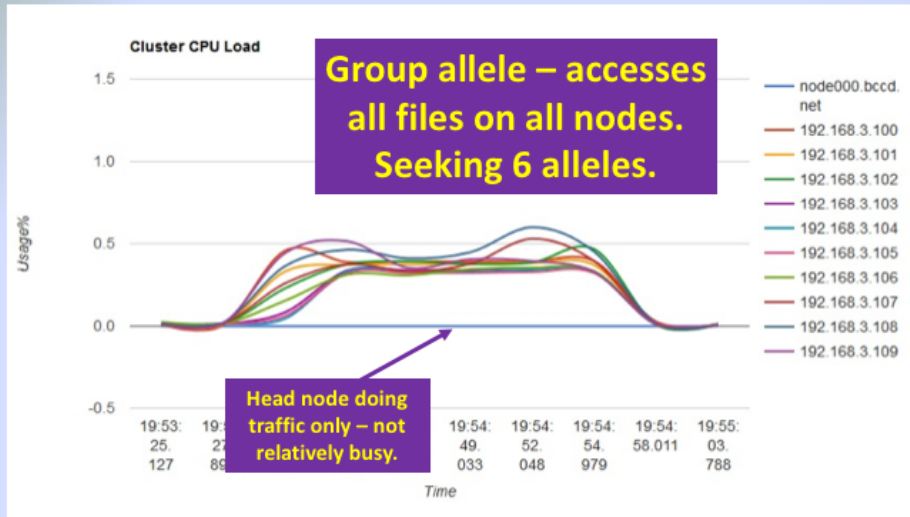
OU Supercomputer Symposium – September 23, 2015

Web GUI – System Monitor



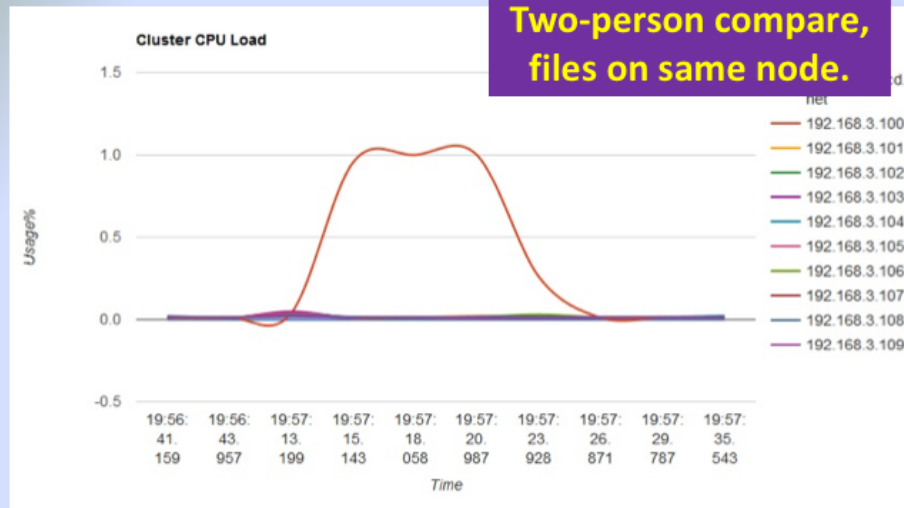
OU Supercomputer Symposium – September 23, 2015

Web GUI – System Monitor



OU Supercomputer Symposium – September 23, 2015

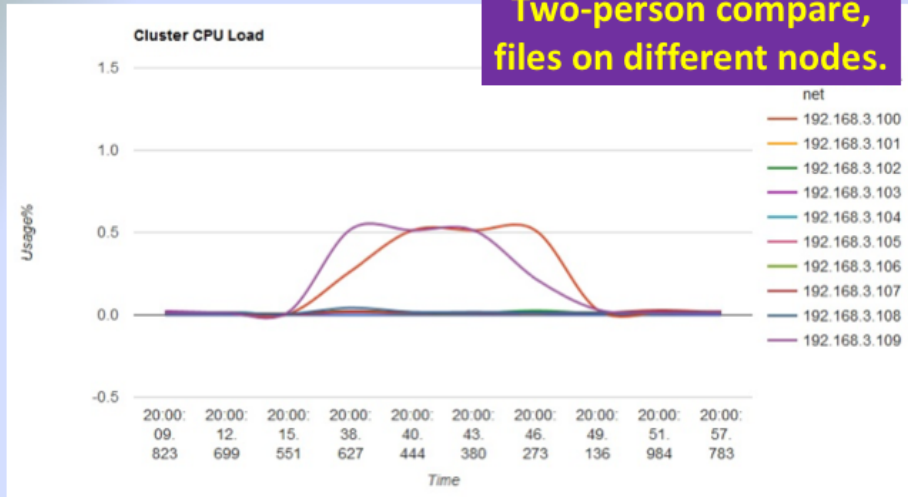
Web GUI – System Monitor



OU Supercomputer Symposium – September 23, 2015

Web GUI – System Monitor

Two-person compare, files on different nodes.



OU Supercomputer Symposium – September 23, 2015

Thank You!

We especially thank Henry Neeman, Charlie Peck, Tom Murphy, and everyone else associated with OU IT, the LittleFe project, the SOSU IT guys and all of our colleagues and friends in the educational community involved with HPC, for all the help we have received.

Karl Frinkle
Mike Morris



OU Supercomputer Symposium – September 23, 2015