SBS 800

Supervisor: Prof V. PERUMAL

By AKHIL KUMAR

About the Project

Temporal evolution of mono and di -nucleotides in human mtDNA hasn't been studied yet. We found an interesting database AmtDB(Jan 2019) having the ancient mitochondrial sequences dated back to ~50k BC

Ancient Mitochondrial Database (amtDB)

Ancient DNA

Ancient specimens

Undergoes fragmentation, post-mortem damages caused mainly by environmental factors

Ancient mtDNA

For tracing human past demographic events

Population variability
Maternal inheritance
High mutation rate
Absence of recombination

The Database

Released earlier this year

1300+ entries
889 with FASTA files
Metadata with geographical locators, archaeological culture affiliation, sex, epoch

Data Analysis

Data collection

FASTA and metadata files from amtDB

Data cleaning

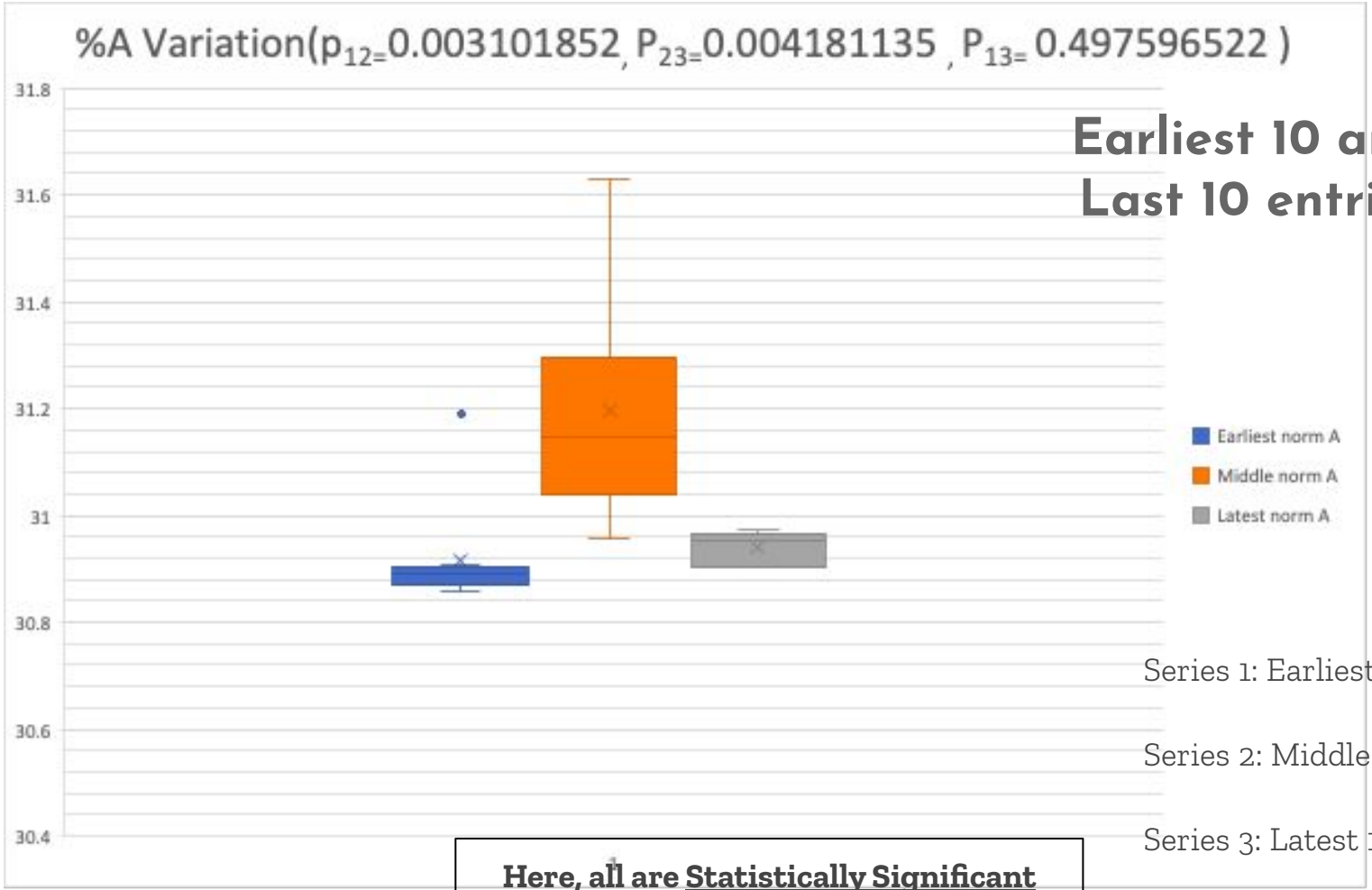
Filtered the data and created a merged file

Data visualization

Produced some plots, performed statistical tests

Inferences

Reached on some interesting conclusions based on the evidence and reasoning



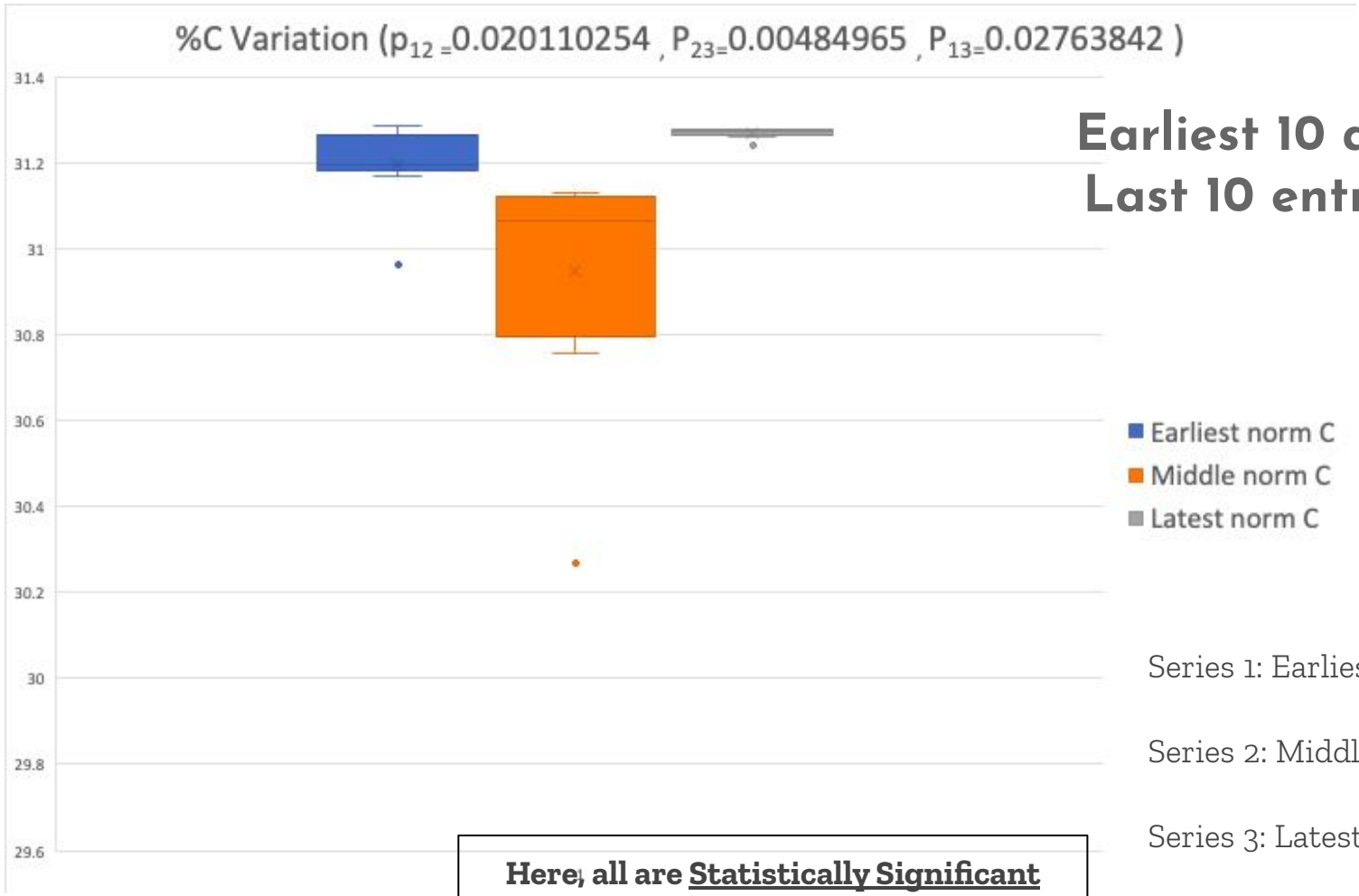
Earliest 10 and Last 10 entries

Series 1: Earliest 10

Series 2: Middle 10

Series 3: Latest 10

Here, all are <u>Statistically Significant</u>



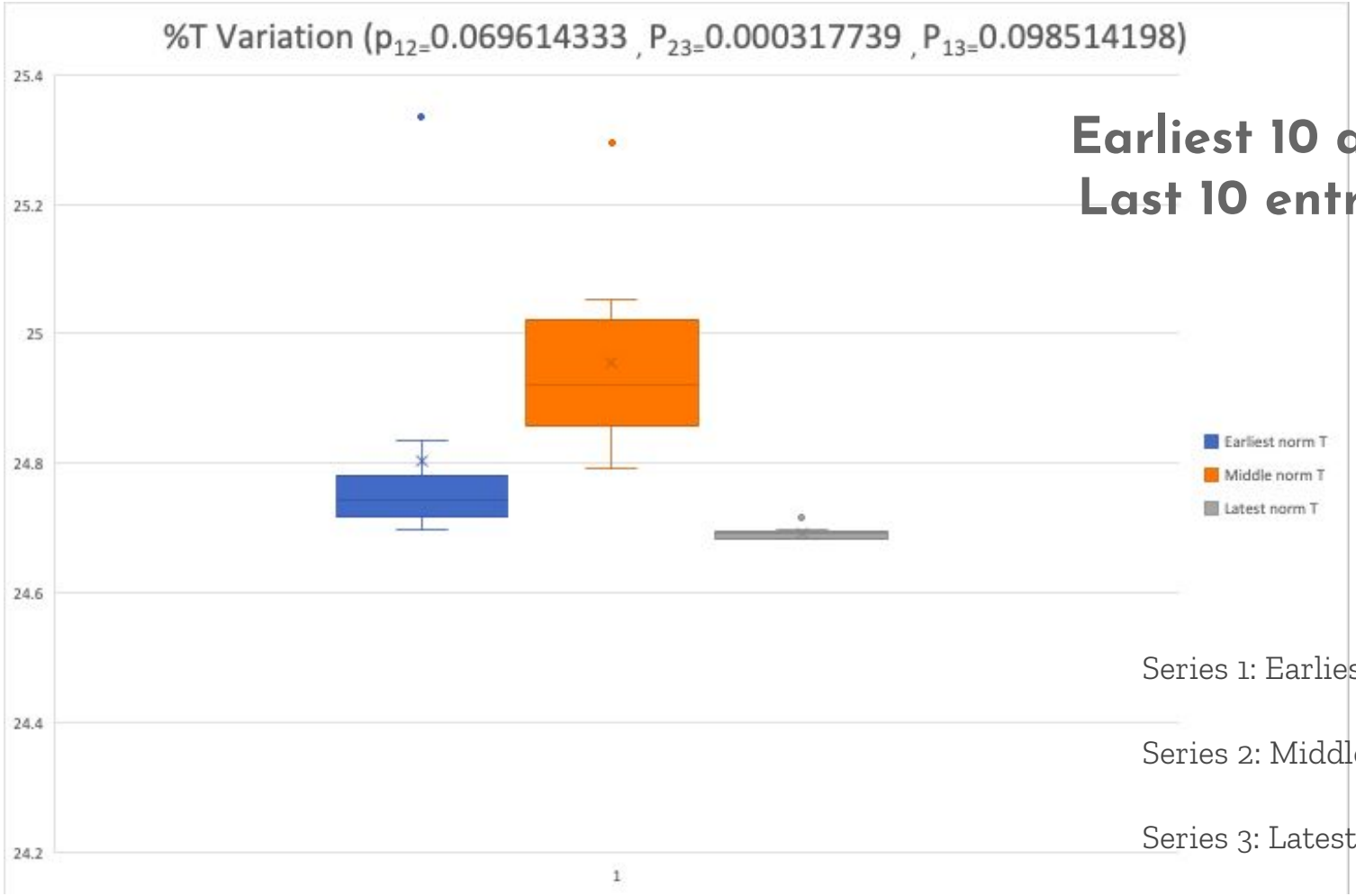
Earliest 10 and Last 10 entries

Series 1: Earliest 10

Series 2: Middle 10

Series 3: Latest 10

Here, all are <u>Statistically Significant</u>

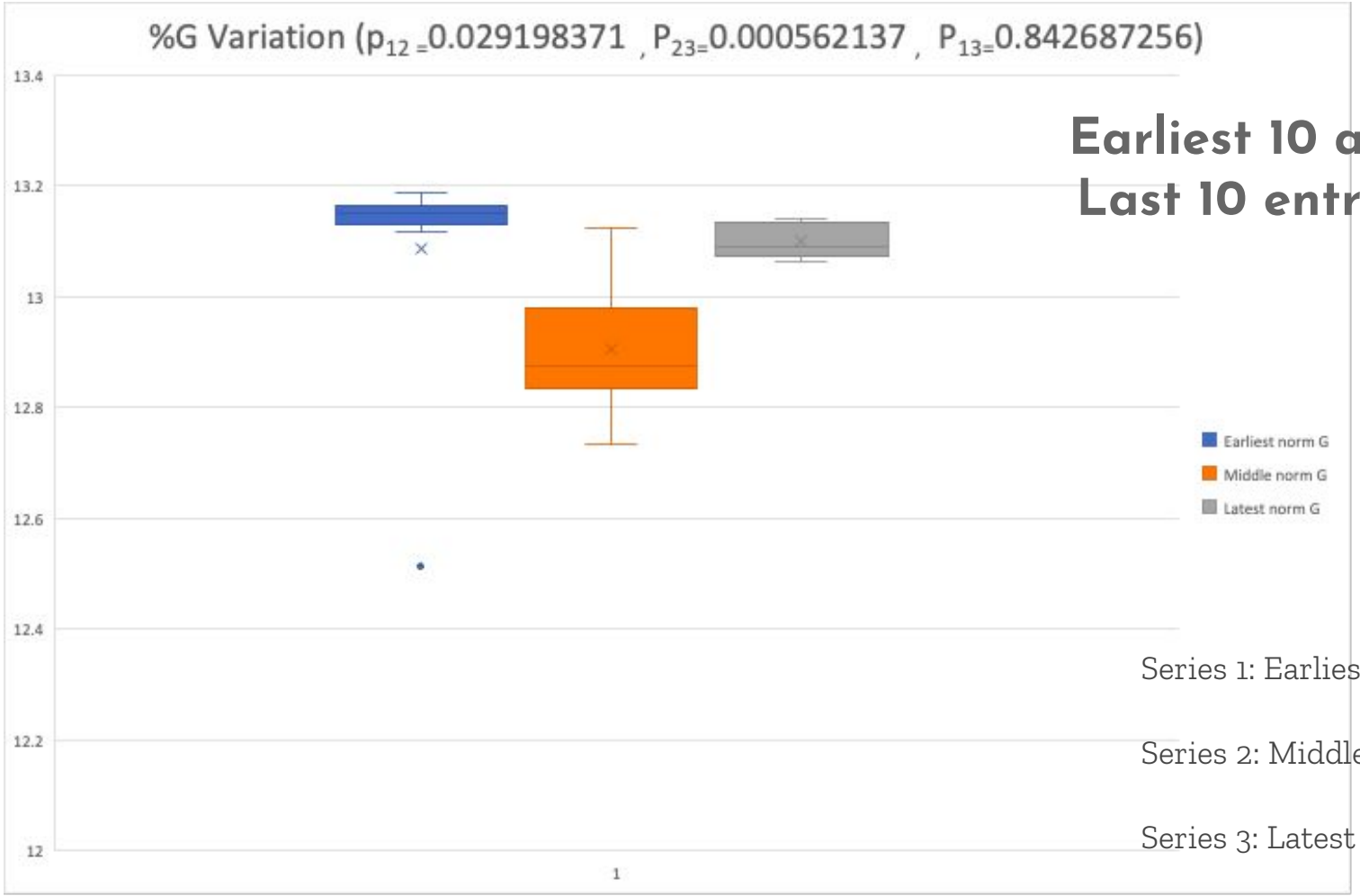


Earliest 10 and Last 10 entries

Series 1: Earliest 10

Series 2: Middle 10

Series 3: Latest 10

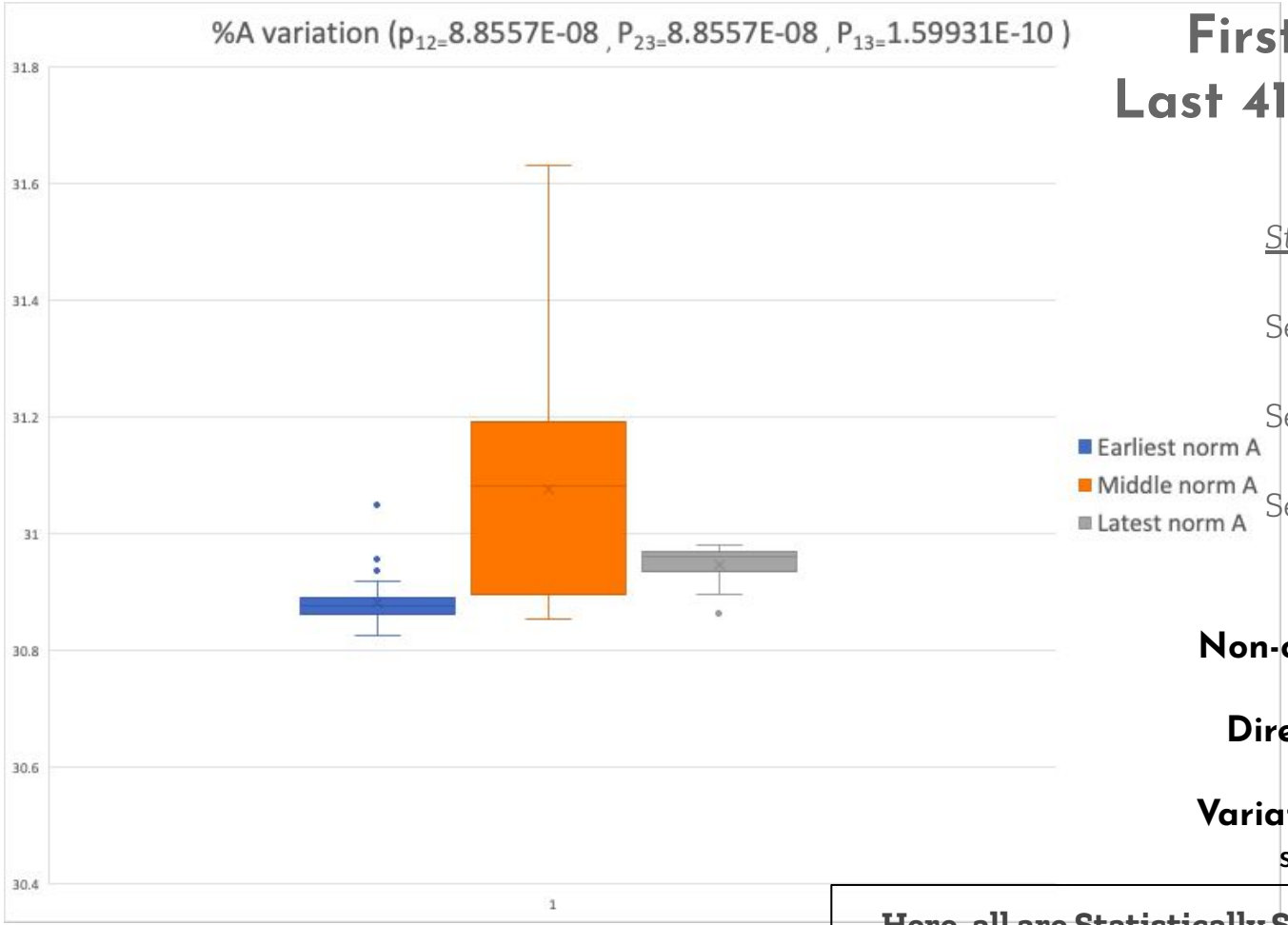


Earliest 10 and Last 10 entries

Series 1: Earliest 10

Series 2: Middle 10

Series 3: Latest 10



First 41 and Last 41 entries

Starting from -10k BC-

Series 1: Earliest 41

Series 2: Middle 41

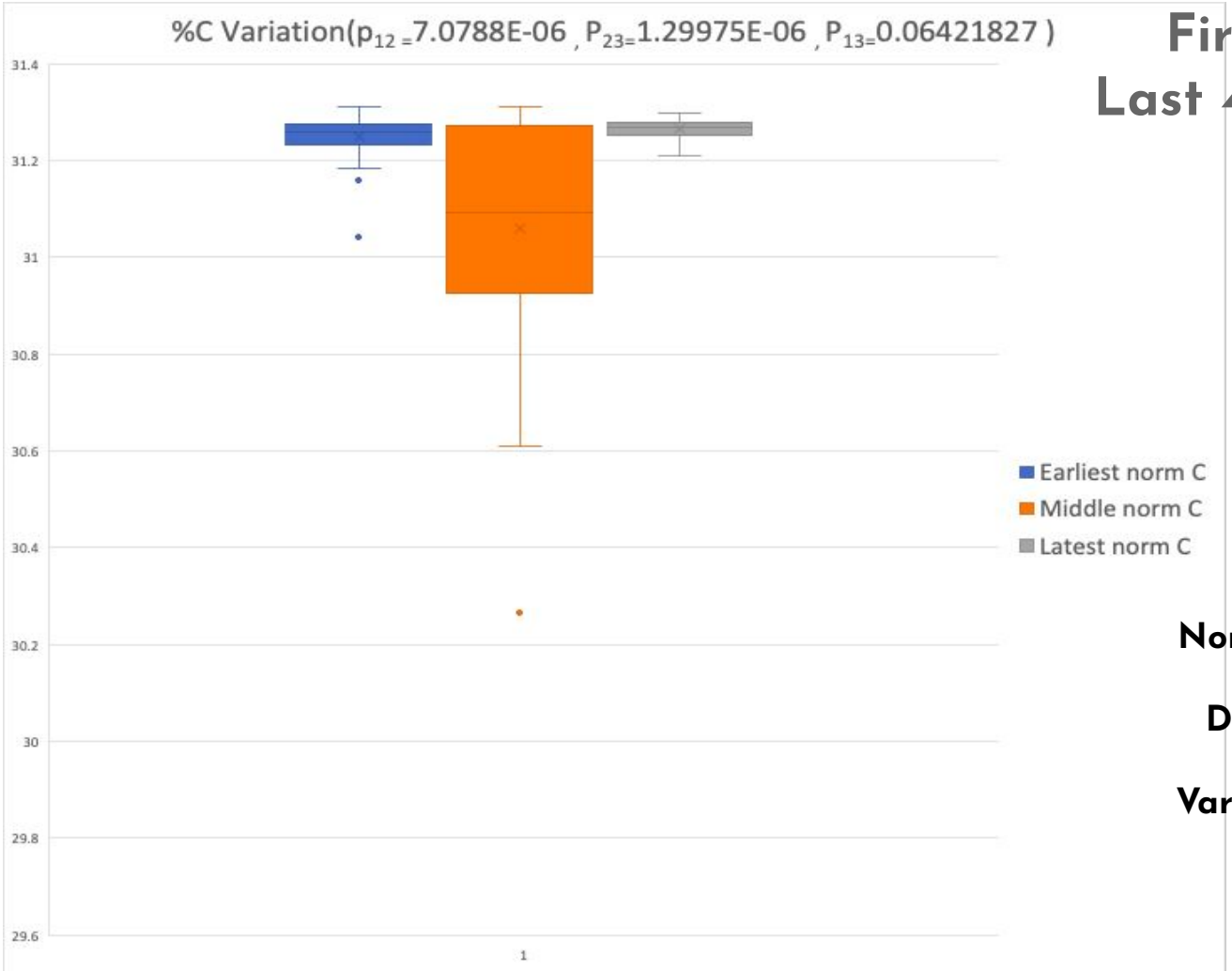
Series 3: Latest 41

Non-overlapping

Directionality

Variation (tighter spread)

Here, all are Statistically Significant



First 41 and Last 41 entries

Starting from -10k BC-

Series 1: Earliest 41

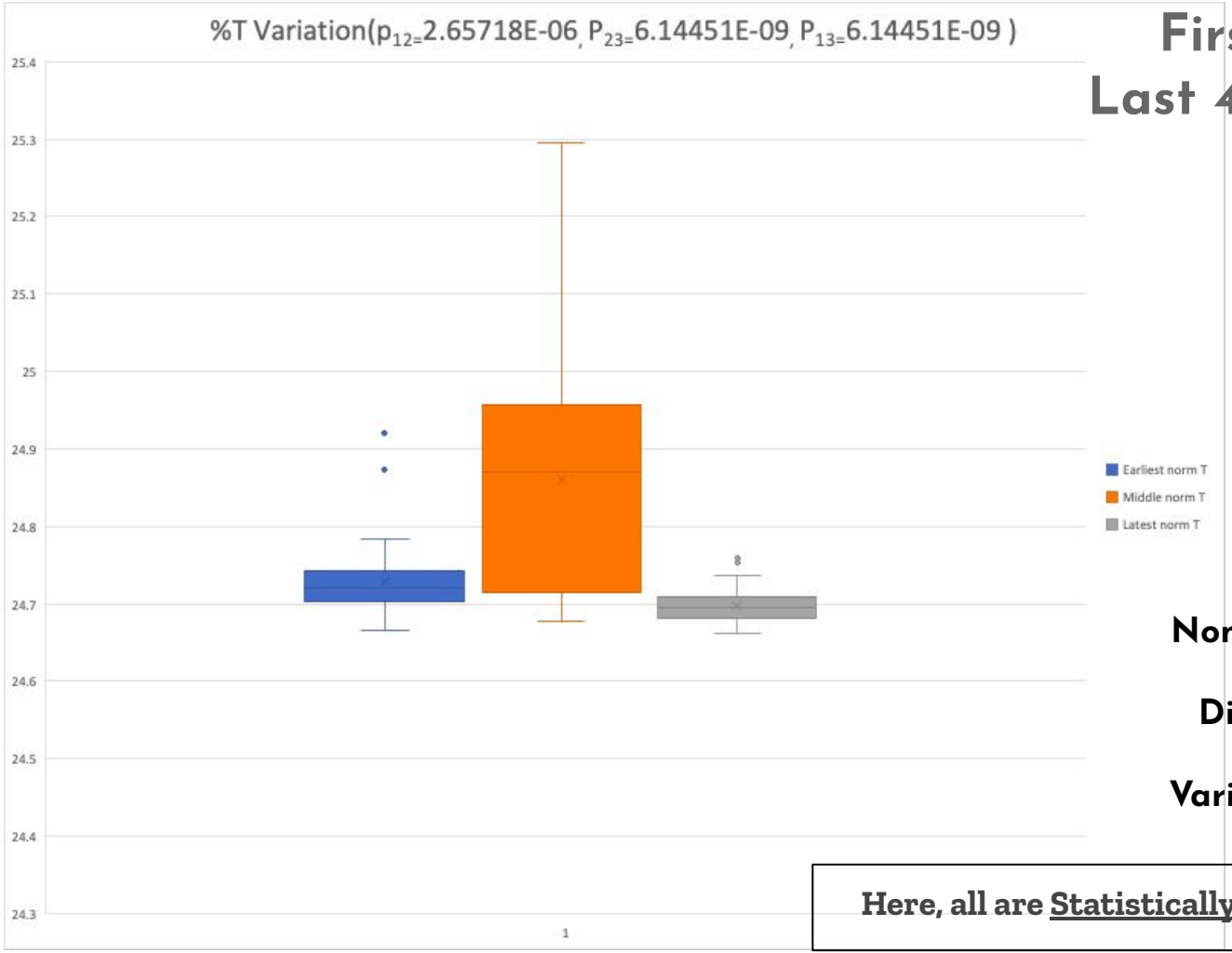
Series 2: Middle 41

Series 3: Latest 41

Non-overlapping

Directionality

Variation (tighter spread)



First 41 and Last 41 entries

Starting from -10k BC-

Series 1: Earliest 41

Series 2: Middle 41

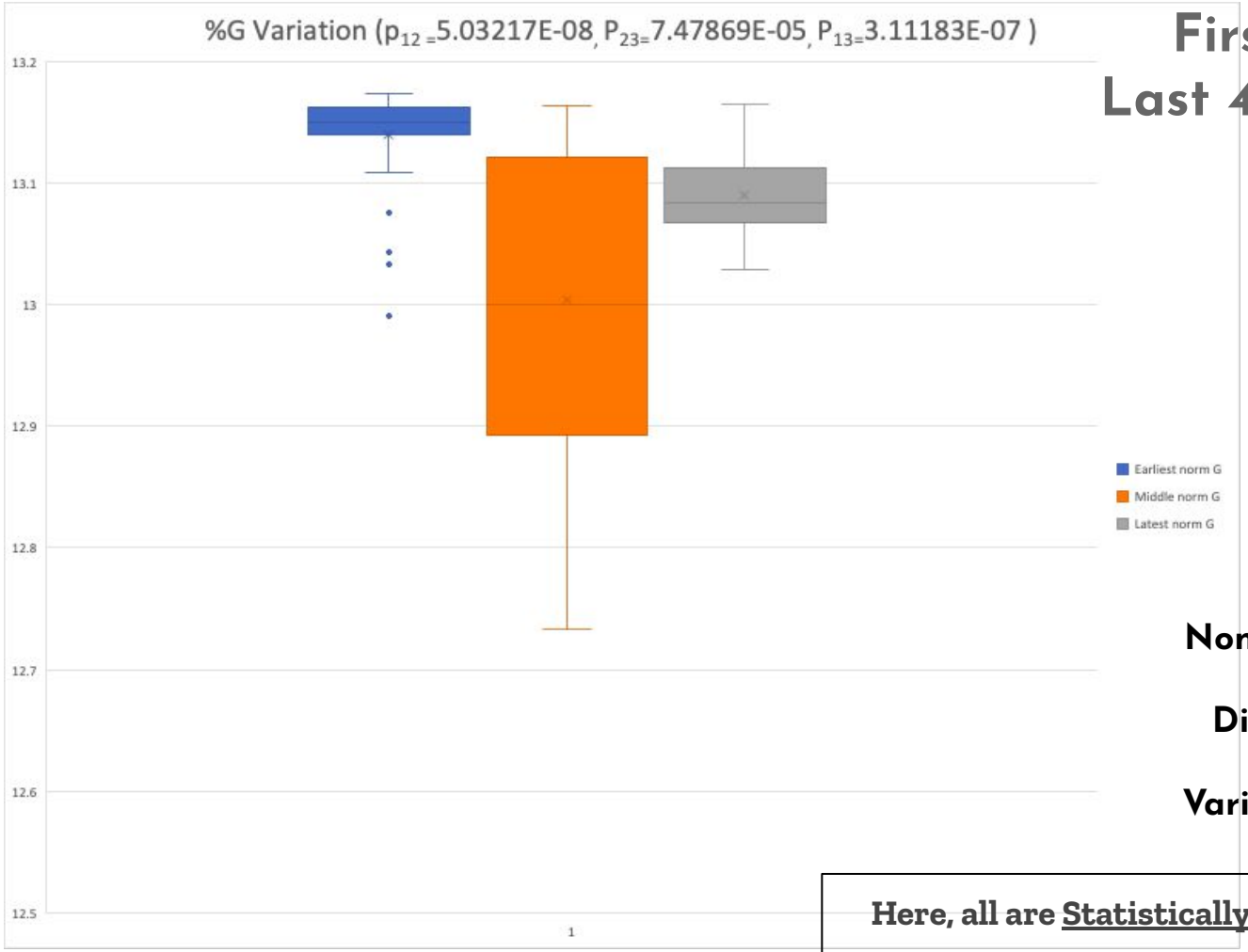
Series 3: Latest 41

Non-overlapping

Directionality

Variation (tighter spread)

Here, all are Statistically Significant



First 41 and Last 41 entries

Starting from -10k BC-

Series 1: Earliest 41

Series 2: Middle 41

Series 3: Latest 41

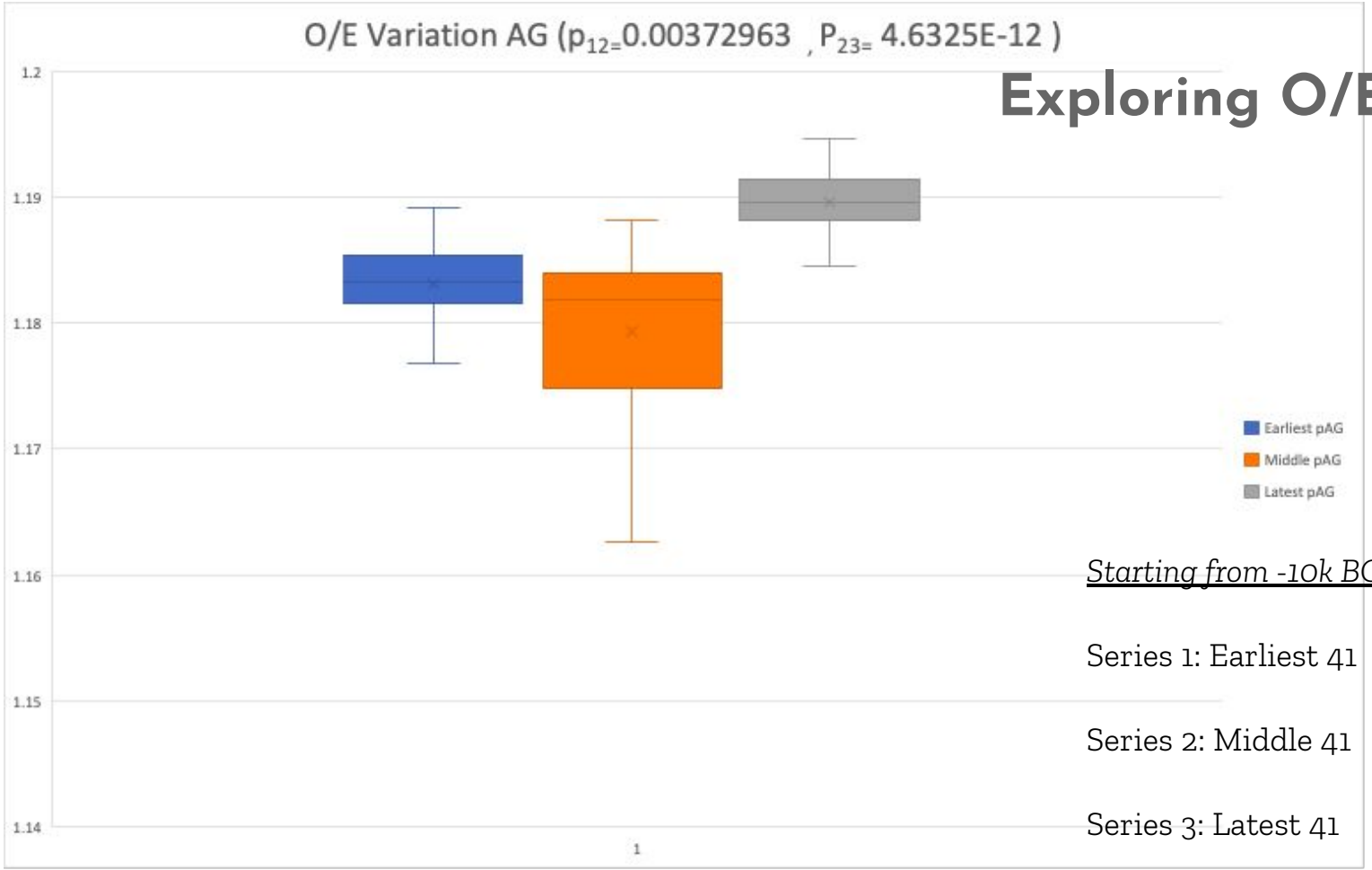
Non-overlapping

Directionality

Variation (tighter spread)

Here, all are Statistically Significant

Exploring O/Es



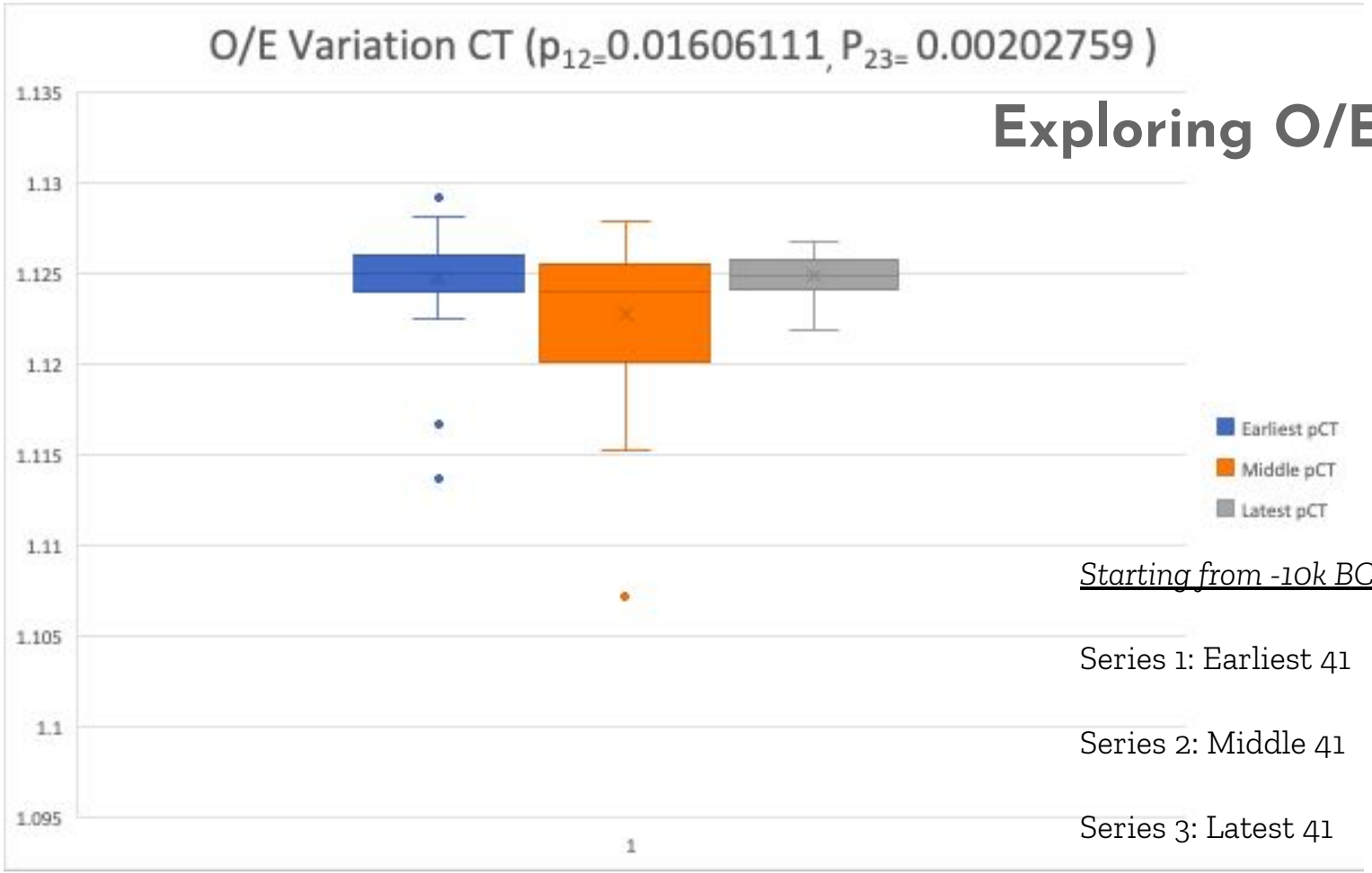
Starting from -10k BC-

Series 1: Earliest 41

Series 2: Middle 41

Series 3: Latest 41

Exploring O/Es



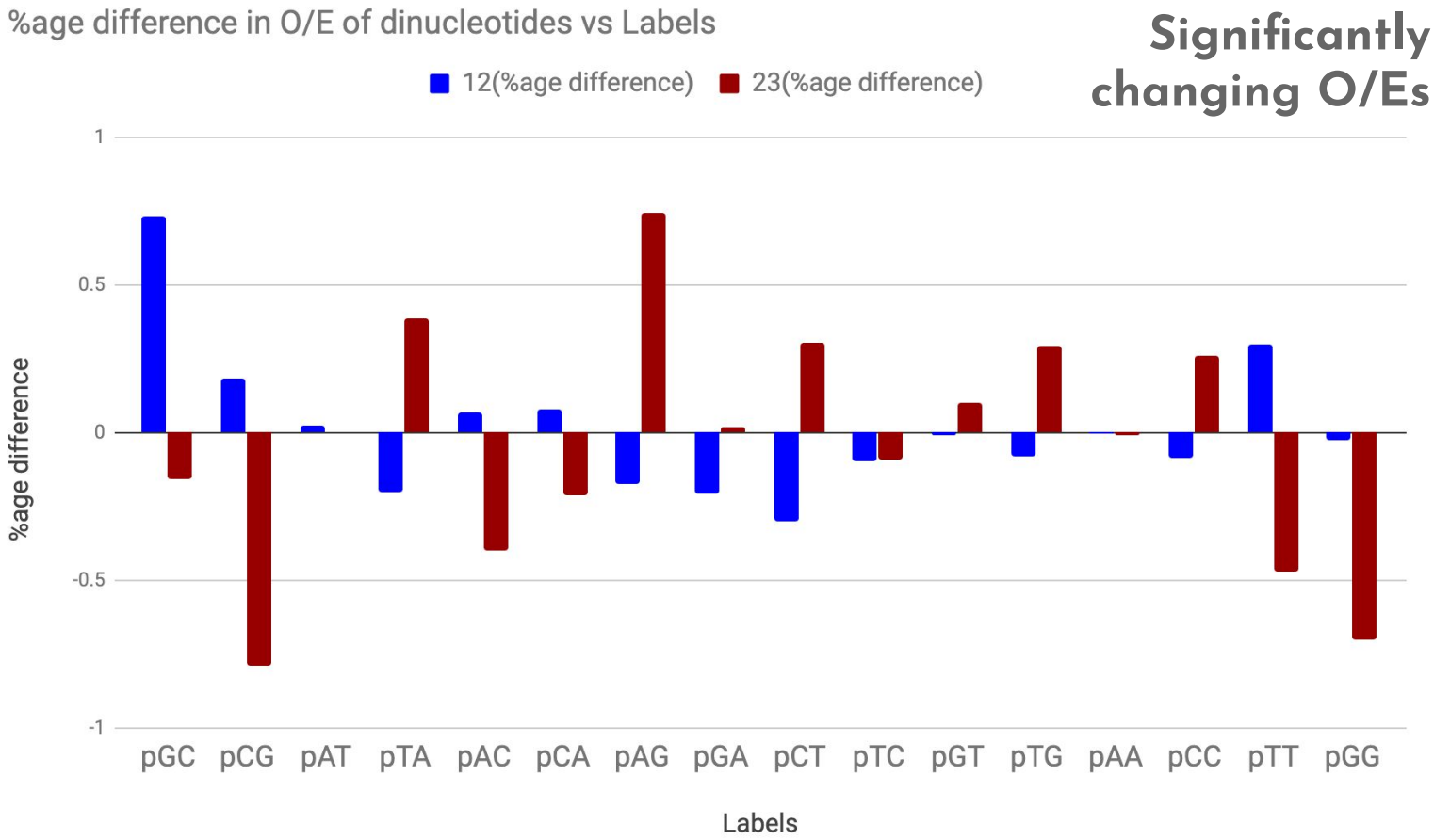
Starting from -10k BC-

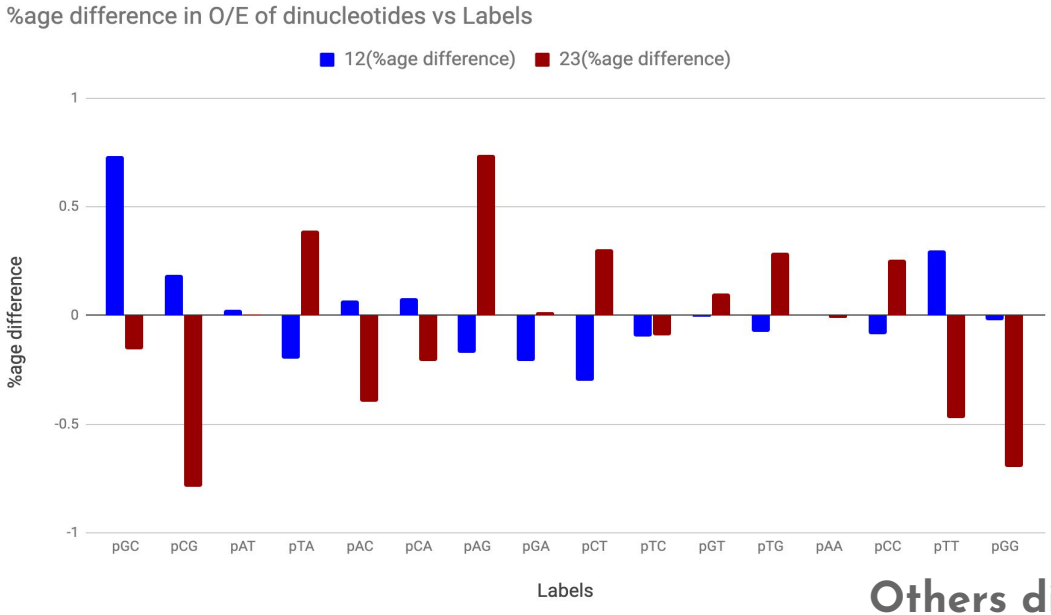
Series 1: Earliest 41

Series 2: Middle 41

Series 3: Latest 41

Significantly changing O/Es





Significantly changing O/Es

For O/E(AG)

p_{12}-value= 0.00372963

p_{23}-value= 4.6325E-12

For O/E(CT)

p_{12}-value= 0.01606111

p_{23}-value= 0.00202759

For O/E(TT)

p_{12}-value= 0.00561418

p_{23}-value= 1.1669E-06

For O/E(GG)

p_{12}-value= 0.05881906

p_{23}-value= 2.9238E-08

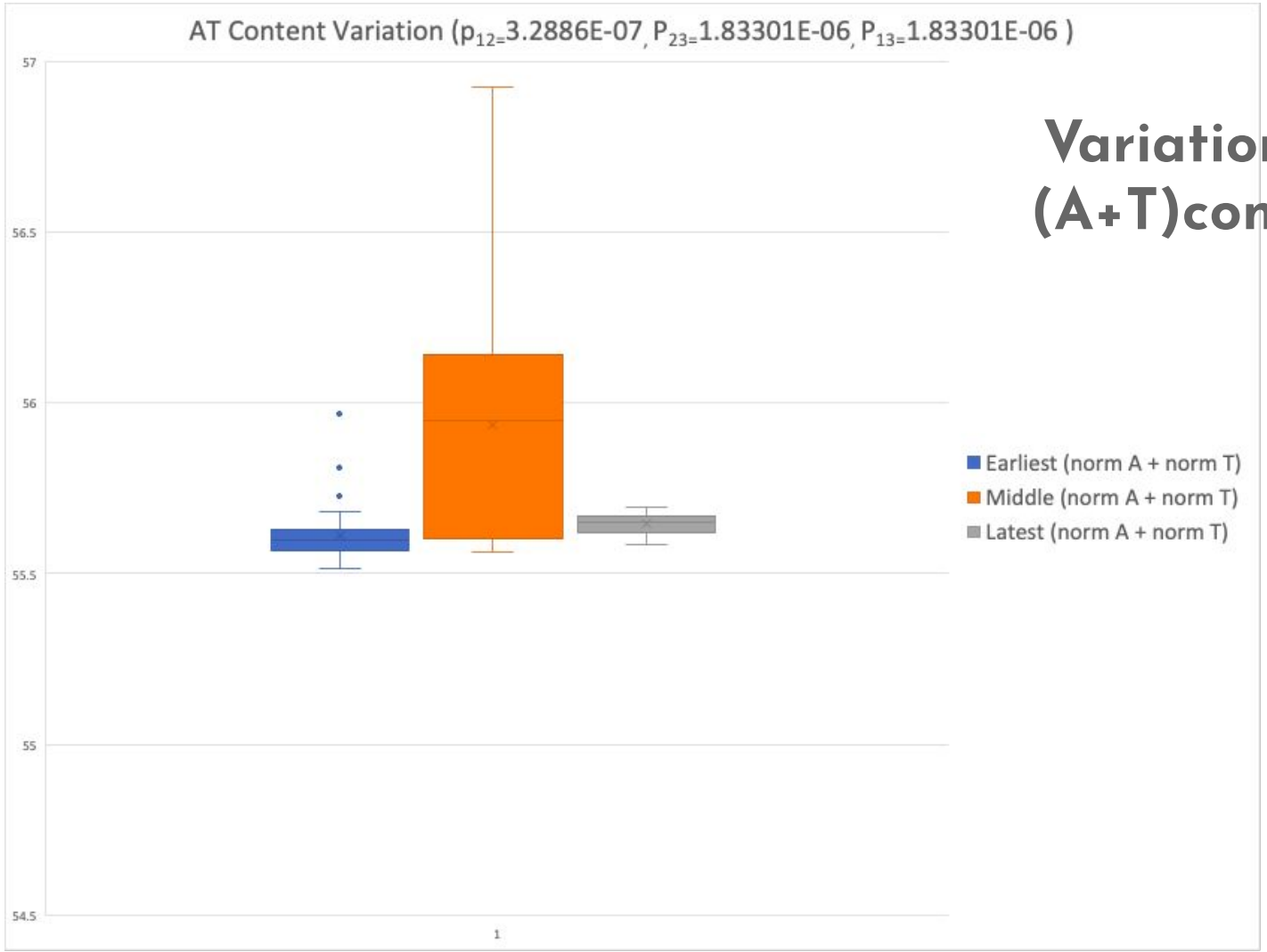
Others didn't have any <u>common</u> significant statistical backing for all the cases!

For O/E(CG)

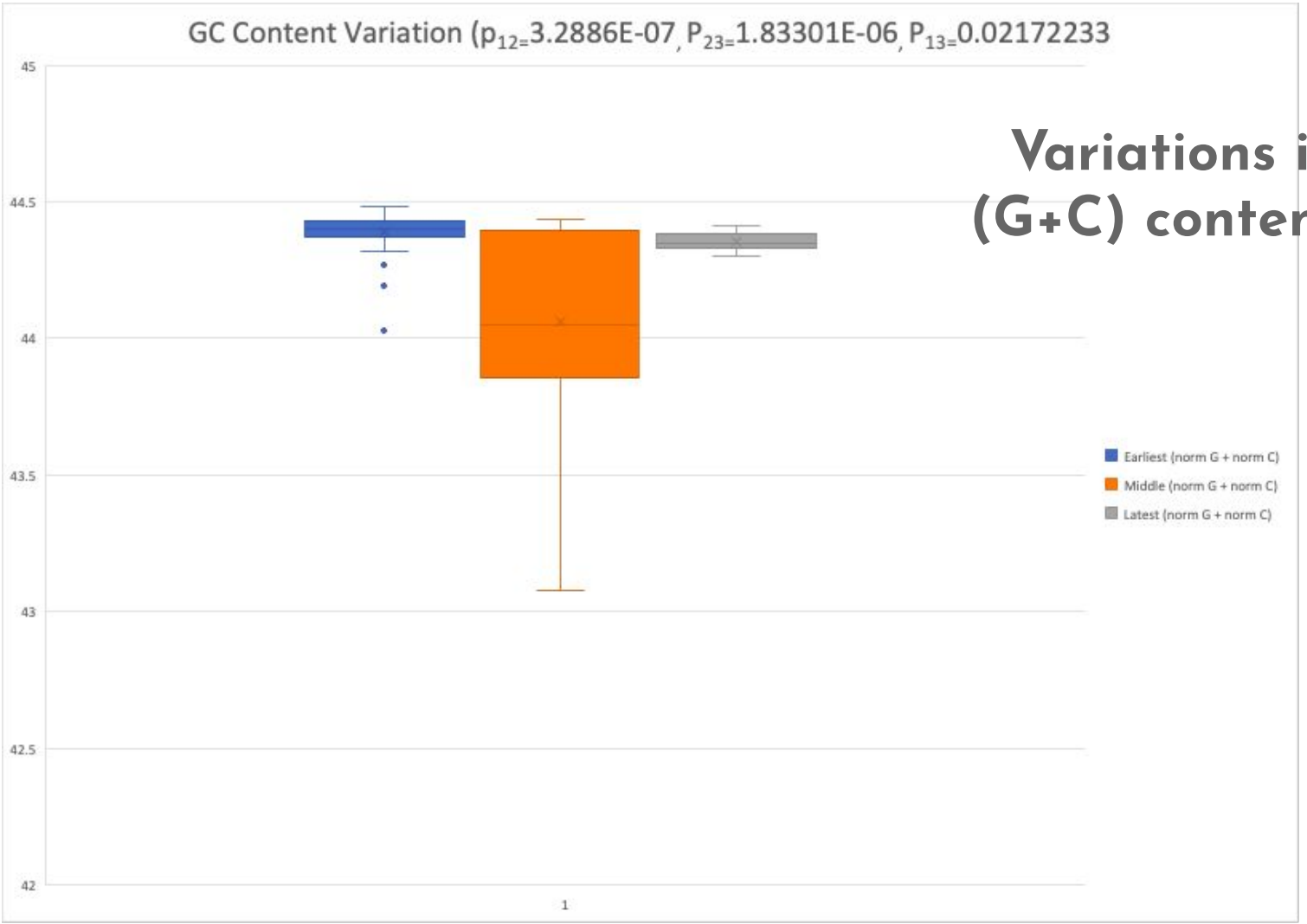
p_{12}-value= 0.34278979 *(not significant)*

p_{23}-value= 6.2134E-05

Variations in (A+T)content



Variations in (G+C) content



Because the average proportion of G and C in the human genome is 20.5% each (15), and this is reflected in the contemporary human DNA sequenced by 454 [supporting information (SI) Fig. 5], this suggests a slight overall bias toward GC-rich sequences in the ancient reads. Strikingly, at the -1 position of the 5′-ends, i.e., the first position upstream of the 5′-most base sequenced, the frequency of G is elevated from ≈22% seen across all Neandertal reads analyzed to 29% (Fisher's exact test, $P < 2.2 \times 10^{-16}$), and the frequency of A is elevated from ≈28% to 31% ($P = 3.5 \times 10^{-10}$), whereas C and T are depressed. Conversely, at the position + 1 downstream of 3′-ends, the frequency of C ($P < 2.2 \times 10^{-16}$) as well as T ($P = 1.32 \times 10^{-5}$) is elevated to ≈30%, whereas G and A are depressed. At the 5′-most sequenced positions, A is depressed to 23% ($P < 2.2 \times 10^{-16}$), whereas T is elevated to 31% ($P = 4.7 \times 10^{-13}$), whereas at the 3′-most sequenced position, A is elevated to 32% ($P = 2.8 \times 10^{-12}$) and T is depressed to 23% ($P < 2.2 \times 10^{-16}$).

Literature findings

Systemic biases in stored samples

Ancient sequences GC rich

PNAS

Patterns of damage in genomic DNA sequences from a Neandertal

Adrian W. Briggs*†, Udo Stenzel*, Philip L. F. Johnson‡, Richard E. Green*, Janet Kelso*, Kay Prüfer*, Matthias Meyer*, Johannes Krause*, Michael T. Ronan§, Michael Lachmann*, and Svante Pääbo*†

*Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany; ‡Biophysics Graduate Group, University of California, Berkeley, CA 94720; and §454 Life Sciences, Branford, CT 06405

Contributed by Svante Pääbo, May 25, 2007 (sent for review April 25, 2007)

High-throughput direct sequencing techniques have recently ... of their position in ancient DNA fragments. Although there is

Variations in (A+T), (G+C) content

%A was increasing, %T was increasing
~0.19% **~0.13%**

%C was decreasing, %G was decreasing
~0.18% **~0.13%**

Over the course of time, mtDNA is

losing GC content

Becoming rich in AT content

%A is decreasing, %T is decreasing
~0.13% **~0.16%**

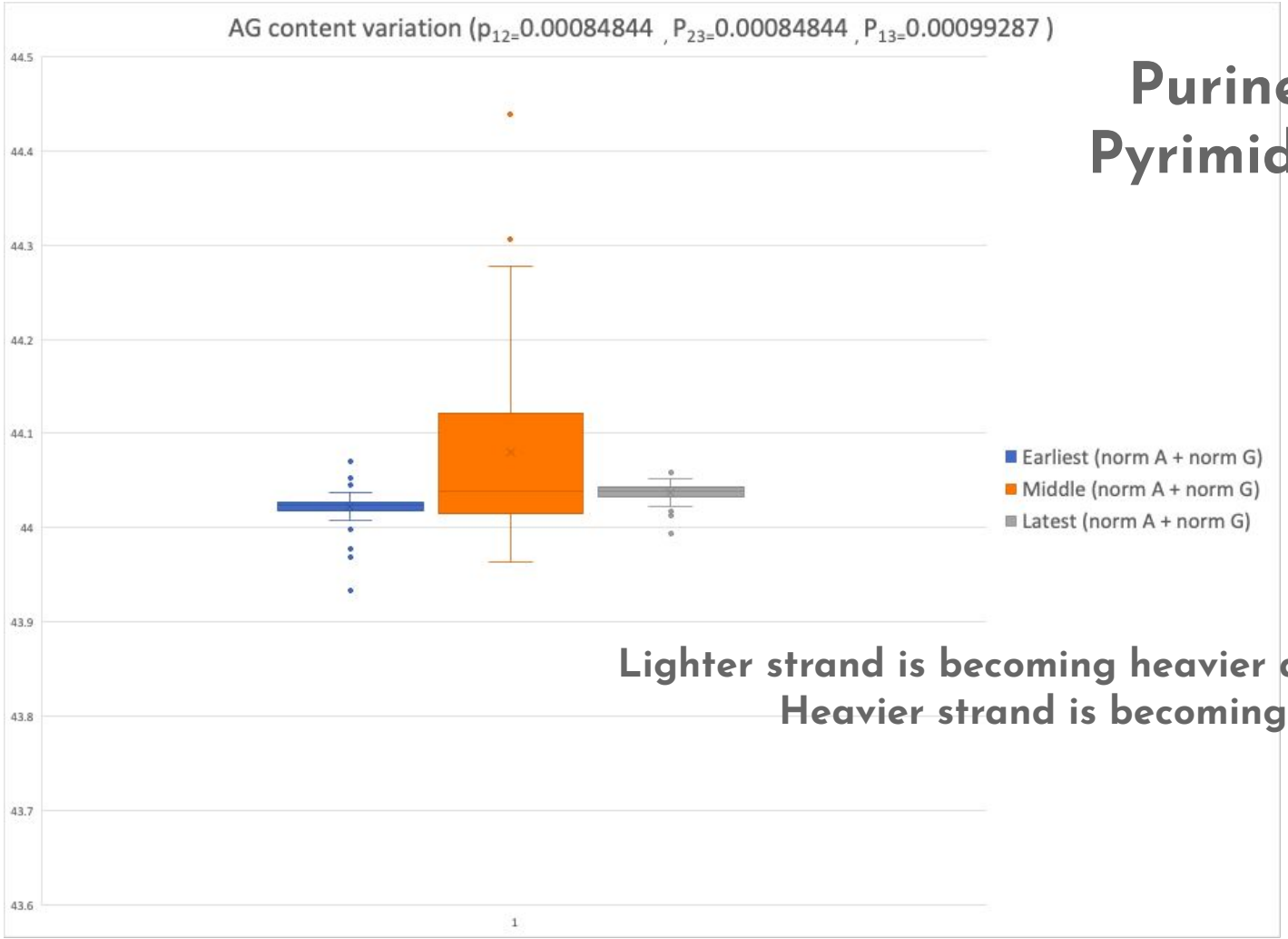
%C was increasing, %G was increasing
~0.20% **~0.08%**

For the new period, mtDNA is

gaining GC content

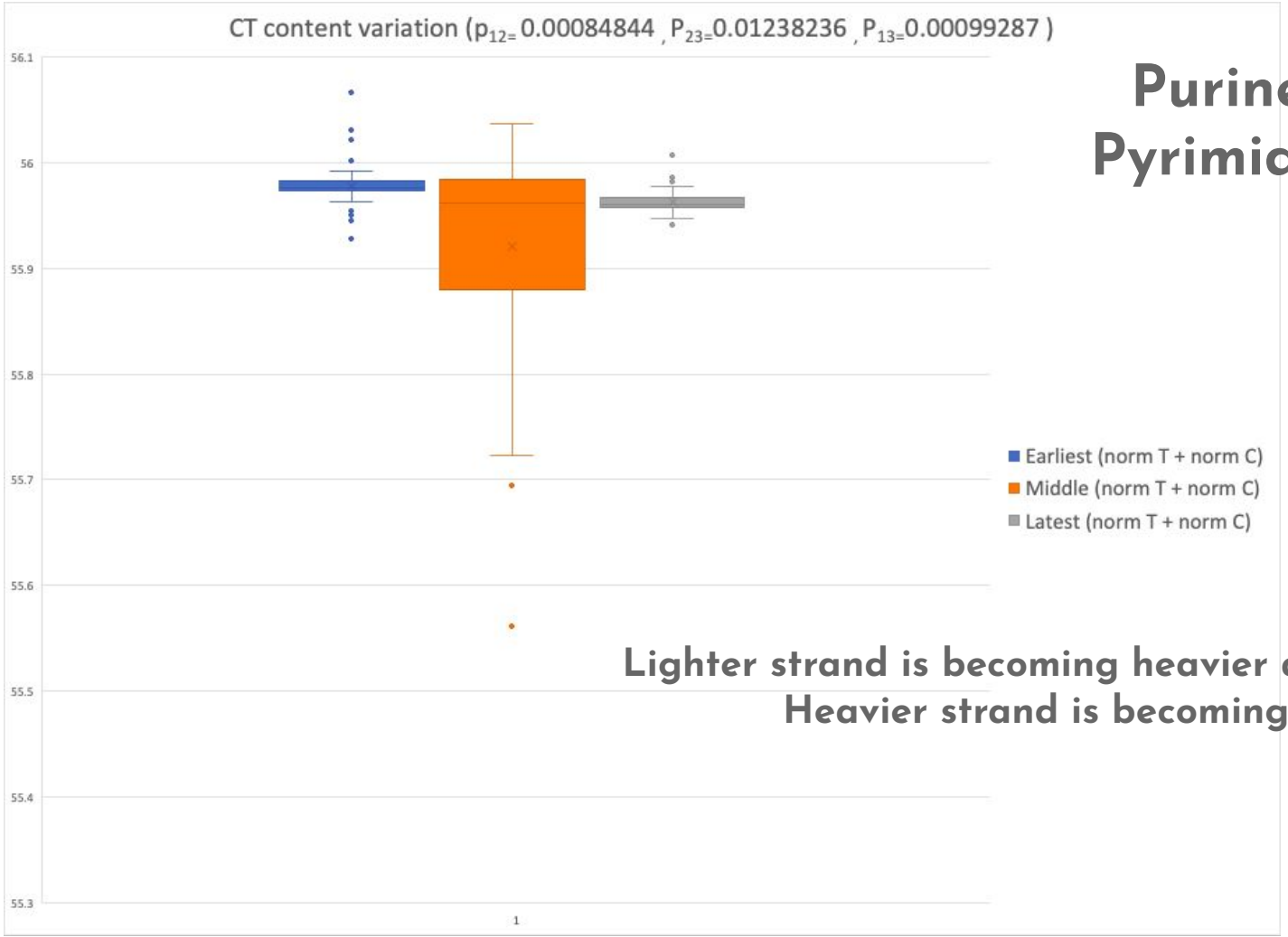
becoming poor in AT content

Purines vs Pyrimidines



Lighter strand is becoming heavier and the Heavier strand is becoming lighter

Purines vs Pyrimidines



Lighter strand is becoming heavier and the Heavier strand is becoming lighter

Future Work

The rate of adaptive evolution in animal mitochondria

JENNIFER E. JAMES,* GWENAEL PIGANEAU†‡ and ADAM EYRE-WALKER*
*School of Life Sciences, University of Sussex, Brighton BN1 9QG, UK, †UPMC Univ Paris 06, UMR 7232, Observatoire Oceanologique, Avenue de Fontaulé, BP 44, 66651 Banyuls-sur-Mer, France, ‡CNRS, UMR 7232, Observatoire Oceanologique, Avenue de Fontaulé, BP 44, 66651 Banyuls-sur-Mer, France

Investigation of **Hypervariable regions** the mtDNA sequence

MITOMAP
A human mitochondrial genome database

A compendium of polymorphisms and mutations in human mitochondrial DNA

Map Locus	Starting	Ending	Shorthand	Description
MT-HV1	16024	16383	CR:HVS1/HV1	Hypervariable segment 1 [classic:16024-16365]
MT-HV2	57	372	CR:HVS2	Hypervariable segment 2 [classic:73-340]
MT-HV3	438	574	CR:HVS3	Hypervariable segment 3

Thank You!

Does anyone have any questions?