

Lesson 2: Introducing the Course (and Haak et al. 2015)

Wednesday July 11: 9:00 – 11:30 am

What did you learn from Professor
Fu's lecture?

What are you hoping to learn from
this course?

July - August 2018

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
1	2	3	4	5	6	7
8	9	10	11	12	13	14
Preparation: Read two studies for Monday journal club.	9:00-11:30, QF: Introduction to ancient DNA research; 2:00-4:00 pm: QF+MY: Journal Club	Preparation: Follow instructions to install python/access server; Read Haak et al. 2015 paper.	9:00-11:30, MY: Introduction to Haak et al 2015/Data sets; 2:00-4:30 pm, MY, QF: Lab preparation and data processing		9:00-11:30, AK: Uniparental markers, mtDNA/chrY	
15	16	17	18	19	20	21
	9:00-11:30, MY: D-statistics (Test of Treeness) 2:00-4:00 pm: HW: Journal Club		9:00-11:30, HW: PCA		9:00-11:30, MY: Outgroup f3-statistic/Introduce Mini-Project	
22	23	24	25	26	27	28
	9:00-11:30, HS: D-statistics (Test of Admixture); 2:00-4:00 pm: AK: Journal Club		9:00-11:30, MY: qpWave		9:00-11:30, MY: qpAdm	
29	30	31	1	2	3	4
	9:00-11:30, HW: ADMIXTURE/PLINK; 2:00-4:00 pm: HS: Journal Club		9:00-11:30, MY: ADMIXTUREGRAPH/Modeling		9:00-11:30, MY: Overflow Time. Meetings with Mel?	
5	6	Notes:				
	9:00-11:30: HS/AK/HS/MY/QF: Final Presentation	Lecturer/Discussion Leaders are listed by initials, as follows: QF=Qiaomei Fu; MY=Melinda Yang; AK=Albert Ko; HW=Hongru Wang; HS=Hassan Shafiey				

July - August 2018

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
1	2	3	4	5	6	7
8	9	10	11	12	13	14
Preparation: Read two studies for Monday journal club.	9:00-11:30, QF: Introduction to ancient DNA research; 2:00-4:00 pm: QF+MY: Journal Club	Preparation: Follow instructions to install python/access server; Read Haak et al. 2015 paper.	9:00-11:30, MY: Introduction to Haak et al 2015/Data sets; 2:00-4:30 pm, MY, QF: Lab preparation and data processing		9:00-11:30, AK: Uniparental markers, mtDNA/chrY	
15	16	17	18	19	20	21
	9:00-11:30, MY: D-statistics (Test of Treeness) 2:00-4:00 pm: Journal Club HW:		9:00-11:30, HW: PCA		9:00-11:30, MY: Outgroup f3-statistic/Introduce Mini-Project	
22	23	24	25	26	27	28
	9:00-11:30, HS: D-statistics (Test of Admixture); 2:00-4:00 pm: AK: Journal Club		9:00-11:30, MY: qpWave		9:00-11:30, MY: qpAdm	
29	30	31	1	2	3	4
	9:00-11:30, HW: ADMIXTURE/PLINK; 2:00-4:00 pm: Journal Club HS:		9:00-11:30, MY: ADMIXTUREGRAPH/Modeling		9:00-11:30, MY: Overflow Time. Meetings with Mel?	
5	6	Notes:				
	9:00-11:30: HS/AK/HS/MY/QF: Final Presentation	Lecturer/Discussion Leaders are listed by initials, as follows: QF=Qiaomei Fu; MY=Melinda Yang; AK=Albert Ko; HW=Hongru Wang; HS=Hassan Shafiey				

Journal Clubs

- Two papers per week (Monday 2-4 pm)
- Short presentation in English within 15 minutes
 - Share background, main argument, opinion or questions
- Every student prepares ≥ 2 questions per paper.

Date	Guide	Paper	Student Leader
2018.07.09	QF+MY	Lipson,2017,Nature	白帆
		Skoglund,2016,Nature	郎曼
2018.07.16	HW	Lazaridis,2017,Nature	王恬怡
		Damgaard,2018,Nature	刘俊岑，苗波
2018.07.23	AK	Kivisild,2017,Human Genetic	王文君
		Rierux,2014,Mol Bio Evo	郑泽权
2018.07.30	HS	Damgaard,2018,Science	张明
		Ilardo,2018,Cell	丁曼雨

July - August 2018

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
1	2	3	4	5	6	7
8	9	10	11	12	13	14
Preparation: Read two studies for Monday journal club.	9:00-11:30, QF: Introduction to ancient DNA research; 2:00-4:00 pm: QF+MY: Journal Club	Preparation: Follow instructions to install python/access server; Read Haak et al. 2015 paper.	9:00-11:30, MY: Introduction to Haak et al 2015/Data sets; 2:00-4:30 pm, MY, QF: Lab preparation and data processing		9:00-11:30, AK: Uniparental markers, mtDNA/chrY	
15	16	17	18	19	20	21
	9:00-11:30, MY: D-statistics (Test of Treeness) 2:00-4:00 pm: HW: Journal Club		9:00-11:30, HW: PCA		9:00-11:30, MY: Outgroup f3-statistic/Introduce Mini-Project	
22	23	24	25	26	27	28
	9:00-11:30, HS: D-statistics (Test of Admixture); 2:00-4:00 pm: AK: Journal Club		9:00-11:30, MY: qpWave		9:00-11:30, MY: qpAdm	
29	30	31	1	2	3	4
	9:00-11:30, HW: ADMIXTURE/PLINK; 2:00-4:00 pm: Journal Club		9:00-11:30, MY: ADMIXTUREGRAPH/Modeling		9:00-11:30, MY: Overflow Time. Meetings with Mel?	
5	6	Notes:				
	9:00-11:30: HS/AK/HS/MY/QF: Final Presentation	Lecturer/Discussion Leaders are listed by initials, as follows: QF=Qiaomei Fu; MY=Melinda Yang; AK=Albert Ko; HW=Hongru Wang; HS=Hassan Shafiey				

Expectations

- Meet MWF 9-11:30 am, with one W 2-4:30 pm class
- Class is a mixture of lecture and discussions
- Speaking in class – invited and encouraged!
- Lecturers want you to learn and have fun
- Do the homework!
- Try your best – that's all we ask for

Expectations

- Meet MWF 9-11:30 am, with one W 2-4:30 pm class
- Class is a mixture of lecture and discussions
 - Homework Presentation
 - New material
 - Lecture
 - Discussion
 - Class Activities (often in small groups)
 - Review/Questions

Expectations

- Meet MWF 9-11:30 am, with one W 2-4:30 pm class
- Class is a mixture of lecture and discussions
- Speaking in class
 - Be ready to **TALK** and discuss previous material
 - Always **ASK** when you are confused
 - Don't be afraid to speak English – practice makes perfect!

Expectations

- Meet MWF 9-11:30 am, with one W 2-4:30 pm class
- Class is a mixture of lecture and discussions
- Speaking in class – invited and encouraged!
- Lecturers want you to learn and have fun
- Do the homework!
 - The best learning is here
 - Ask questions of each other and the lecturers
 - It will take time, but be patient

Expectations

- Meet MWF 9-11:30 am, with one W 2-4:30 pm class
- Class is a mixture of lecture and discussions
- Speaking in class – invited and encouraged!
- Lecturers want you to learn and have fun
- Do the homework!
- Try your best – that's all we ask for

Homework

- Trying to design homework to take at least 7-8 hours.
- Mix of questions
 - Short answer/searching primary literature
 - Running introduced software
 - Writing short scripts to retrieve information, prepare data for analysis
 - More advanced scripts and/or thinking creatively.
 - You will upload homework onto the server – more later!
- Be prepared to present some homework in future classes!

Preparing for the course

- Does everyone have internet access?
- Could everyone install jupyter notebook and the VPN?
- Can everyone connect to the VPN?
 - In a Terminal, try 'ssh [student@192.168.1.2](#)'
 - It will prompt you for a password (Ask me!).
 - This brings you to the ADMIN node, where you should work as little as possible.
 - Type ssh comput12
 - This brings you to computing node 12.
 - Type 'htop' and press enter.

Do's and Don'ts of Server Access

- htop – way to check CPU and memory usage
 - DON'T use too many CPUs (for class, use one/run)
 - DON'T use too much RAM (We will warn you about memory intense runs, but check in htop to be sure!)
- Hard Disk Storage – limited space, make sure to only write new files into the folder you make for yourself at `/public/adna/summer2018/`
 - DON'T copy large data files into the home directory!

Useful Unix commands

- less
- cat
- grep
- paste
- cut
- sort
- screen
- head
- tail
- vim
- mv
- cp
- rm
- mkdir
- pwd
- man
- wget
- |
- df -h

Make folder for your files

- The main directory we will work in is:
 - /public/adna/student/
 - In 2018class/, make a folder for yourself titled “lastname_firstname/”
 - This is where you will make new folders and files for the course on the server.
 - Make a folder for your homework exercises. You will upload your home problems to this folder.

Tunneling to the comput node

Type `vim ~/.ssh/config` onto a Terminal or Cygwin window on your computer.

In the text file, press 'i' and then copy the 'Host...' and lower lines.

Press 'esc' and then ':', and then 'w' to save.

Press ':' and then 'q' to quit.

Try `ssh student12` in your Terminal window – you should have to input the password twice, but then end up directly in the `comput12` node!

Host student12

User student

HostName comput12

ProxyCommand `ssh student@192.168.1.2 nc %h %p 2> /dev/null`

Accessing jupyter notebook

After saving the new config file, type
'ssh -NfL localhost:8880:localhost:8880
student@student12' in your Terminal window.

If no error happens, open a web browser and
type 'localhost:8880'

If they request a password, ask me!

Transferring files

- ssh – need Terminal/Cygwin to login to server
- jupyter notebook (and accessing remotely) – easy way of accessing server when writing scripts
- scp – how to transfer files to and from server

```
scp -r student12:/public/adna/student/usefulfiles/  
ssh_config_file.txt .
```

-This transfers the text file from the server onto your computer.

```
scp -r ssh_config_file.txt student12:/public/adna/  
student/2018class/lastname_firstname/
```

-This transfers the text file from your computer onto the server.

Let's begin!

Haak et al. (2015)

LETTER

doi:10.1038/nature14317

Massive migration from the steppe was a source for Indo-European languages in Europe

Wolfgang Haak^{1*}, Iosif Lazaridis^{2,3*}, Nick Patterson³, Nadin Rohland^{2,3}, Swapan Mallick^{2,3,4}, Bastien Llamas¹, Guido Brandt⁵, Susanne Nordenfelt^{2,3}, Eadaoin Harney^{2,3,4}, Kristin Stewardson^{2,3,4}, Qiaomei Fu^{2,3,6,7}, Alissa Mittnik⁸, Eszter Bánffy^{9,10}, Christos Economou¹¹, Michael Francken¹², Susanne Friederich¹³, Rafael Garrido Pena¹⁴, Fredrik Hallgren¹⁵, Valery Khartanovich¹⁶, Aleksandr Khokhlov¹⁷, Michael Kunst¹⁸, Pavel Kuznetsov¹⁷, Harald Meller¹³, Oleg Mochalov¹⁷, Vayacheslav Moiseyev¹⁶, Nicole Nicklisch^{5,13,19}, Sandra L. Pichler²⁰, Roberto Risch²¹, Manuel A. Rojo Guerra²², Christina Roth⁵, Anna Szécsényi-Nagy^{5,9}, Joachim Wahl²³, Matthias Meyer⁶, Johannes Krause^{8,12,24}, Dorcas Brown²⁵, David Anthony²⁵, Alan Cooper¹, Kurt Werner Alt^{5,13,19,20} & David Reich^{2,3,4}

Haak et al. (2015)

LETTER

doi:10.1038/nature14317

Massive migration from the steppe was a source for Indo-European languages in Europe

Wolfgang Haak^{1*}, Iosif Lazaridis^{2,3*}, Nick Patterson³, Nadin Rohland^{2,3}, Swapan Mallick^{2,3,4}, Bastien Llamas¹, Guido Brandt⁵, Susanne Nordenfelt^{2,3}, Eadaoin Harney^{2,3,4}, Kristin Stewardson^{2,3,4}, Qiaomei Fu^{2,3,6,7}, Alissa Mittnik⁸, Eszter Bánffy^{9,10}, Christos Economou¹¹, Michael Francken¹², Susanne Friederich¹³, Rafael Garrido Pena¹⁴, Fredrik Hallgren¹⁵, Valery Khartanovich¹⁶, Aleksandr Khokhlov¹⁷, Michael Kunst¹⁸, Pavel Kuznetsov¹⁷, Harald Meller¹³, Oleg Mochalov¹⁷, Vayacheslav Moiseyev¹⁶, Nicole Nicklisch^{5,13,19}, Sandra L. Pichler²⁰, Roberto Risch²¹, Manuel A. Rojo Guerra²², Christina Roth⁵, Anna Szécsényi-Nagy^{5,9}, Joachim Wahl²³, Matthias Meyer⁶, Johannes Krause^{8,12,24}, Dorcas Brown²⁵, David Anthony²⁵, Alan Cooper¹, Kurt Werner Alt^{5,13,19,20} & David Reich^{2,3,4}

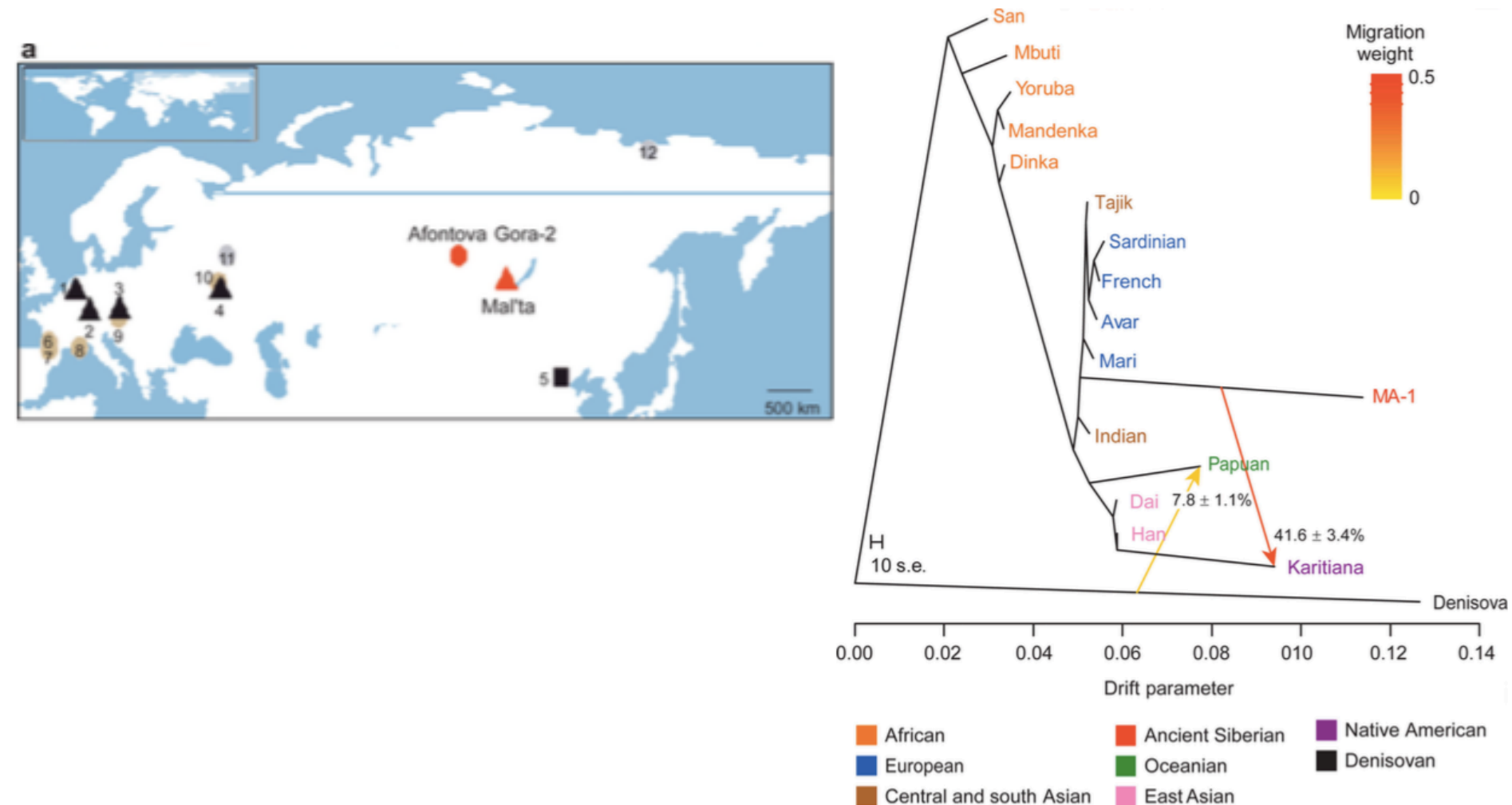
What can you tell me about this study?

References

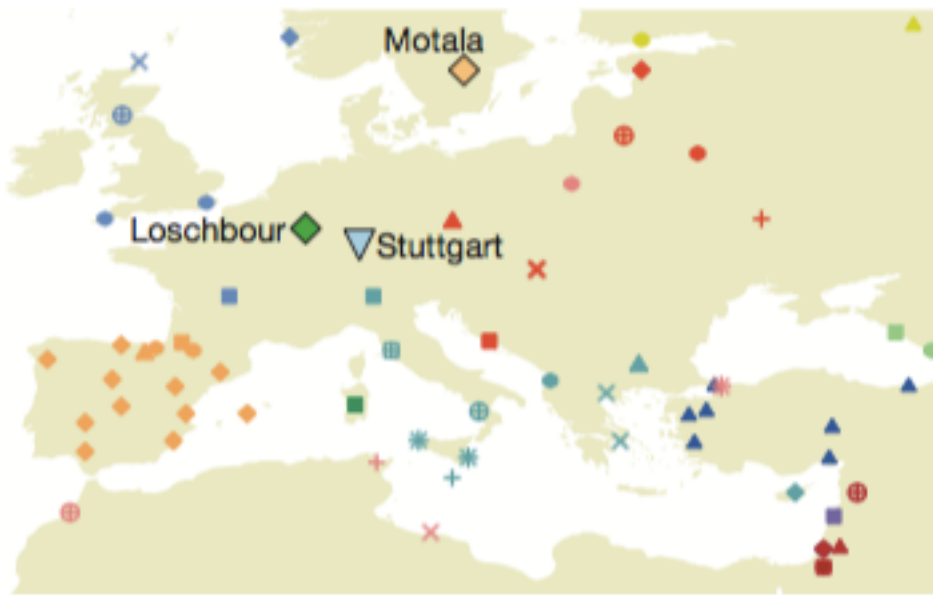
1. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014).
2. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nature Commun.* **5**, 5257 (2014).
3. Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nature Commun.* **3**, 698 (2012).
4. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
5. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225–228 (2014).
6. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
7. Seguin-Orlando, A. *et al.* Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113–1118 (2014).
8. Skoglund, P. *et al.* Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* **344**, 747–750 (2014).
9. Anthony, D. W. *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World* (Princeton Univ. Press, 2007).
10. Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).
11. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. Lond. B* **370**, 20130624 (2015).
12. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
13. Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).

Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans

Maanasa Raghavan^{1,*}, Pontus Skoglund^{2,*}, Kelly E. Graf³, Mait Metspalu^{4,5,6}, Anders

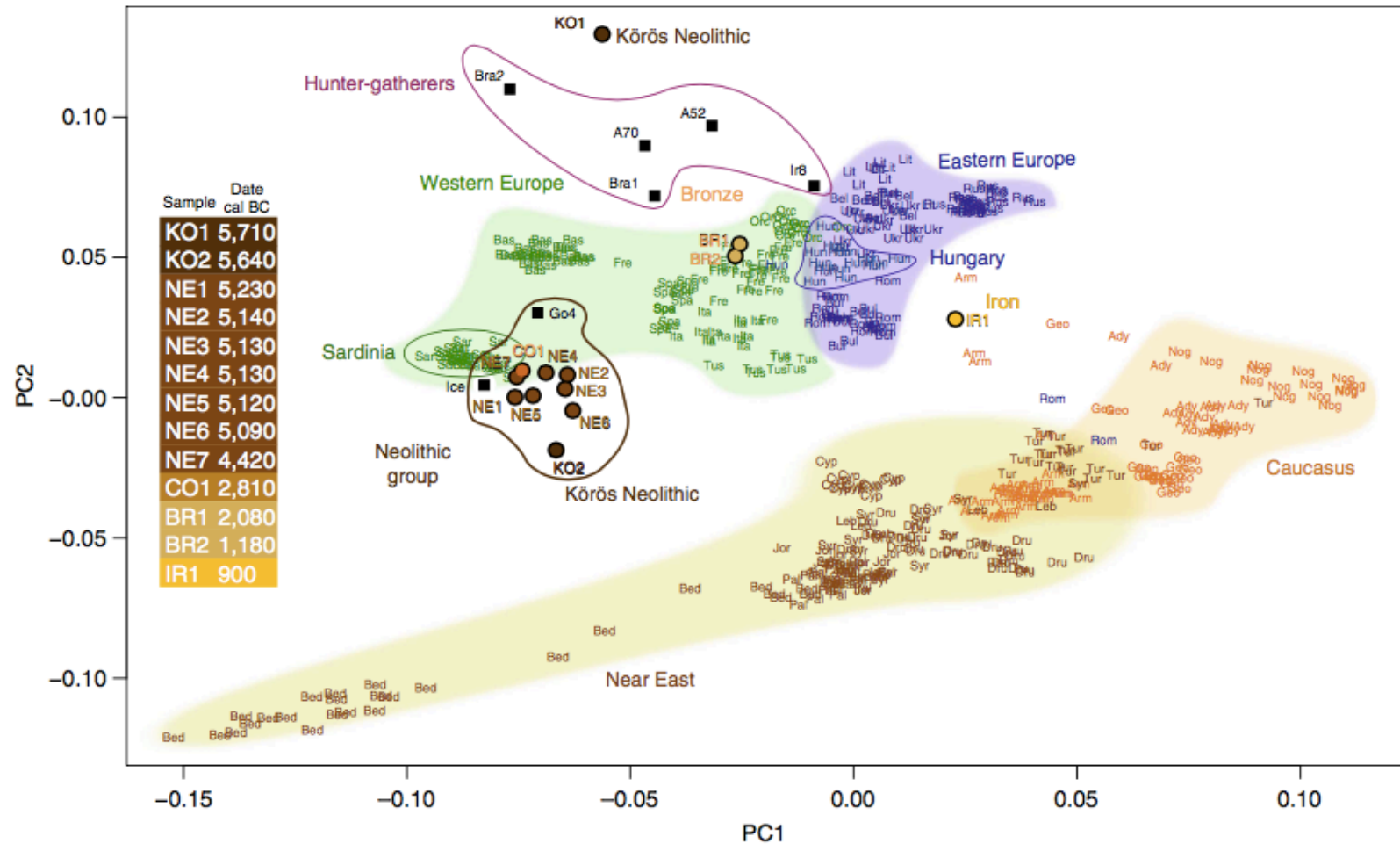


Ancient human genomes suggest three ancestral populations for present-day Europeans



Genome flux and stasis in a five millennium transect of European prehistory

Cristina Gamba^{1,2,3}, Eppie R. Jones³, Matthew D. Teasdale³, Russell L. McLaughlin³, Gloria Gonzalez-Fortes⁴,



Major Players in Haak et al.

- Western European H-G (WHG)
- Eastern European H-G (EHG)*
- Scandinavian H-G (SHG)
- Early Neolithic (EN)*
- Middle Neolithic (MN)*
- Late Neolithic (LN)*
- Bronze Age (BA)*
- Present-day

Ancient individuals available



Time (ky sc)	Group	West	Central	East
0.9	Iron Age			Hungary IA (1)
1.1	Late Bronze Age		Halberstadt (1)	Hungary BA (2)
2.2–1.6	Early Bronze Age		Unetice (8)	
2.5–2.2	Late Neolithic		Alberstedt (1) Benzigerode/Heimbürg (3) Bell Beaker (6) Karsdorf (1) Corded Ware (4)	
3.3–2.7	Late Copper Age (steppe)			Hungary CA (1) Sweden NHG (3) Yamnaya (9) Sweden MN (1)
4–3	Mid Neolithic	La Mina (4)	Iceman (1) Esperstedt (1) Baalberge (3)	
6–5.5	Early Neolithic	Eis Trocs (5)	Stuttgart (1) LBK (12)	LBKT (1) Hungary EN (8) Starcevo (1)
6–4.6	Holocene hunter-gatherer	La Brana1 (1)	Loschbour (1) Hungary HG (1)	Sweden MHG (1) Karelia (1) Motala (7) Samara (1)
43–22	Pleistocene hunter-gatherer			MA1 (1) Kostenki14 (1) Ust Ishim (1)

What questions are there?

What questions are there?

- If Europeans consist of three types of ancestry, when and how did each ancestry enter the European pool?
- What are differences between western and eastern Europeans? At different time periods?
- With many individuals from a small region (central Germany), what are local changes through time, and how does that compare to other regions?

Accessing the dataset

- <https://reich.hms.harvard.edu/datasets>
- AdmixTools:
<https://github.com/DReichLab/AdmixTools>
 - PACKEDANCESTRYMAP format

Accessing the dataset

- <https://reich.hms.harvard.edu/datasets>
- AdmixTools:
<https://github.com/DReichLab/AdmixTools>
 - PACKEDANCESTRYMAP format
 - CONVERTF helps us to convert to something readable – EIGENSTRAT format.

Accessing the dataset

- <https://reich.hms.harvard.edu/datasets>
- AdmixTools:
<https://github.com/DReichLab/AdmixTools>
 - PACKEDANCESTRYMAP format
 - CONVERTF helps us to convert to something readable – EIGENSTRAT format.
- The data can be found at /public/adna/student/data/
- Let's go to jupyter notebook.

Class Activity

- Can we identify the ancient individuals in the dataset, and develop a text file containing extra information?
 - What information do we want?
 - Where do we look for this information?
- As a class, we will fill out an Excel with this information, standardizing nomenclature.

Gathering info on our individuals.

- We will divide into four groups, and each group will be responsible retrieving information for one of the four groups. I'll do Motala.
- EHG: Karelia, Samara
- SHG: Motala
- EN: LBK, LBKT, Starcevo, Spain_EN
- MN/CA: Yamnaya, Esperstedt, Baalberge, Spain_MN
- LN: Alberstedt, BenzigerodeHeimburg, Bell Beaker, Karsdorf, Corded Ware
- BA/IA: Halberstadt, Unetice

Assigning Ancient Individuals/ Populations

ID	N	ID	N	ID	N
HungaryGamba_BA*	2	Corded_Ware_LN	4	Starcevo_EN	1
HungaryGamba_CA*	1	SwedenSkoglund_MHG*	1	LaBrana1*	1
HungaryGamba_IA*	1	SwedenSkoglund_NHG*	3	Loschbour*	1
HungaryGamba_HG*	1	SwedenSkoglund_MN*	1	Motala_HG	7
HungaryGamba_EN*	8	Yamnaya	9	Karelia_HG	1
Halberstadt_LBA	1	Iceman*	1	Samara_HG	1
Unetice_EBA	7	Esperstedt_MN	1	MA1*	1
Alberstedt_LN	1	Baalberge_MN	3	Kostenki14*	1
BenzigerodeHeimburg_LN	3	Stuttgart*	1	Ust_Ishim*	1
Bell_Beaker_LN	6	LBKT_EN	1	Spain_MN (La Mina)	4
Karsdorf_LN	1	LBK_EN	12	Spain_EN (Els Trocs)	4

Each student will be responsible for one of the starred individuals.

Two students together responsible for Sweden data.

Two students together responsible for HungaryGamba data.

Eleven all together.

Bolded=Haak et al. data

*=Previously published data used in Haak et al. (2015)