

Welcome to Population Genetics!



Todays schedule

Introduction to the course (HRS)

- Who is this person teaching?
- What is population genetics about?
- Why is it so important today?
- Practical aspects about the course

History of population genetics (HRS)

- From Darwin to the 1000 genomes project

The Hardy-Weinberg law (HRS)

Introduction to Linux (RH)

Who is this person teaching?

Hans R. Siegismund

Associate professor

Section for Computational
and RNA Biology

Department of Biology

Biocenter

Room 1.2.16

Population genetics

African ungulates

Muskox

Great Apes

(Foot- and mouth disease virus)

Population genetics of African ungulates



Great Apes



Chimpanzee



Bonobo



Lowland gorilla



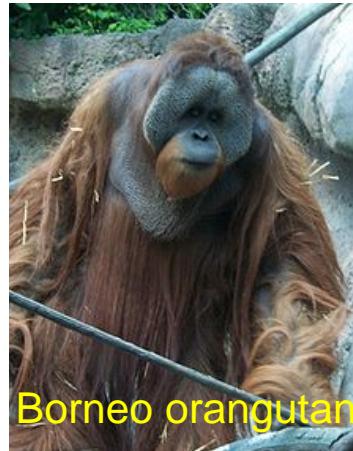
Mountain gorilla



Christina Hvilsom



Peter Frandsen



Borneo orangutan



Sumatra orangutan

Transboundary Animal Diseases in East Africa

Infrastructure

2 molecular genetic laboratories

Personnel

3 project coordinators

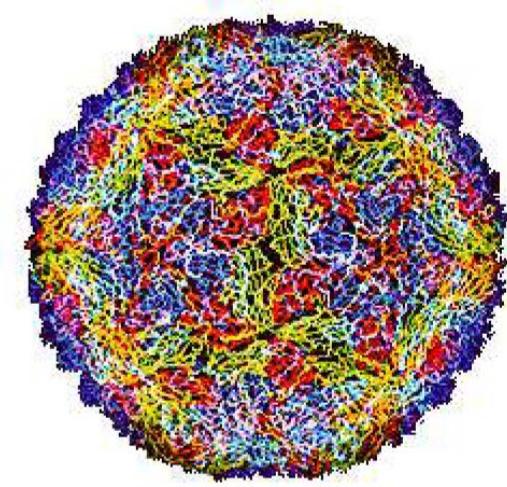
2 post doc

7 PhD students

3 technicians

Funding

Danida: DKK 19,300,000
(2006-2016)



Foot-and mouth-disease virus

RNA-virus

~8300 bp long

Other people teaching

Anders Albrechtsen



Emil Jørsboe



Ida Moltke



Kristoffer Vitting-Seerup



Peter Frandsen



Rasmus Heller



Ryan Waples



Fernando Racimo



most are from the

[popgen group](#) at the Section for Computational and
RNA Biology

(plus Copenhagen Zoo and the Natural History
Museum)

Introduction to the course

Main focus

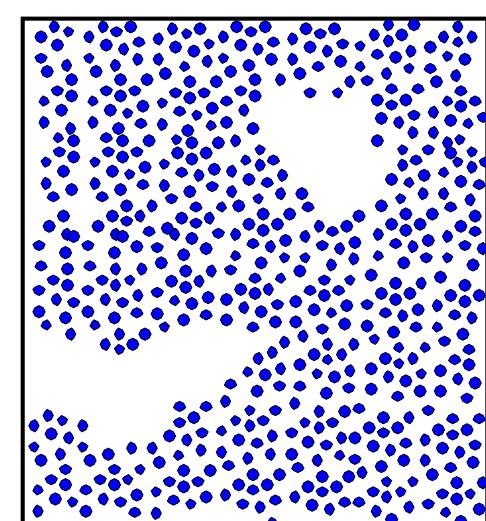
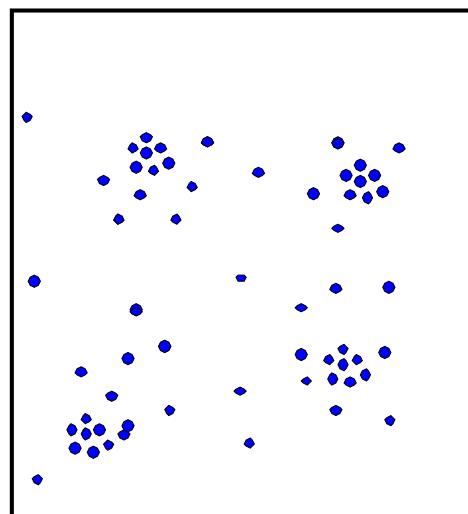
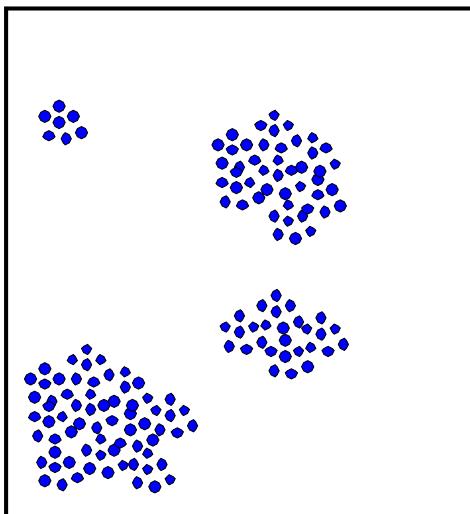
- Understand concepts of population genetics
- Analyse current data sets (genomic data) with up-to-date analytical tools

Introduction to the course

- What is population genetics?

What is a population?

Most species are geographically structured.



Introduction to the course

- What is population genetics?

What is a population?

Most species are geographically structured.



Introduction to the course

- What is population genetics?

What is a population?

Most species are geographically structured.

We focus on local **interbreeding** populations.

(More genetic interactions within than between populations)

Other names:

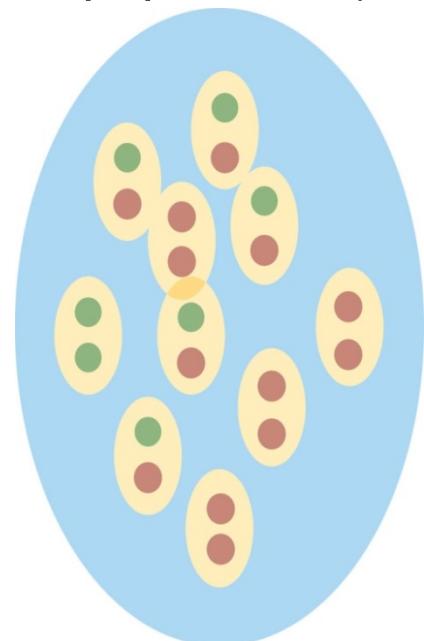
Local population

Subpopulation

Deme

Mendelian population

Population



Introduction to the course

Population structure in humans

Can we distinguish
Norwegians from Swedes?
Norwegians and Swedes from Italians?



Introduction to the course

Population structure in humans

Data

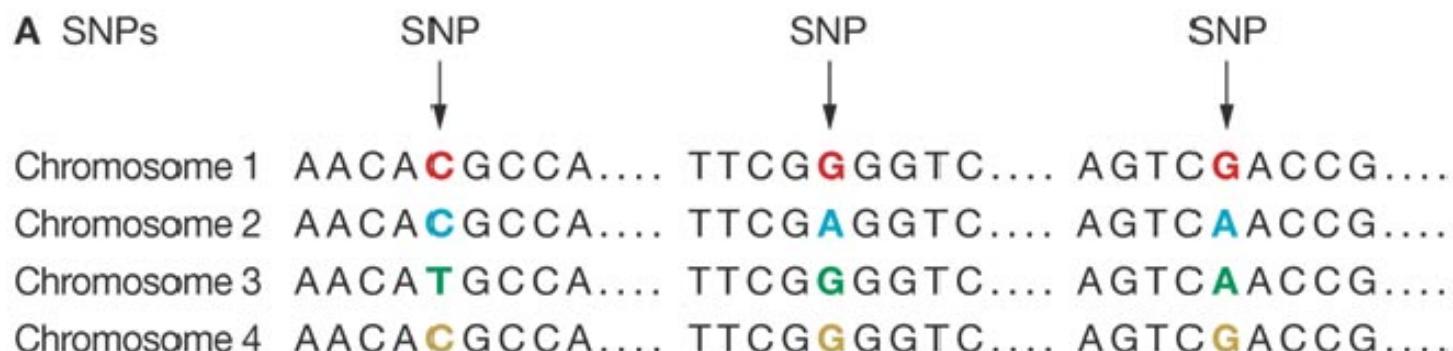
$N = 3,000$

500,000 SNP

(Single nucleotide polymorphism)

Analysis

Principle
component
analysis



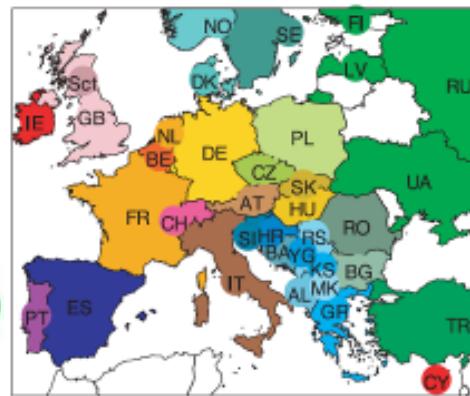
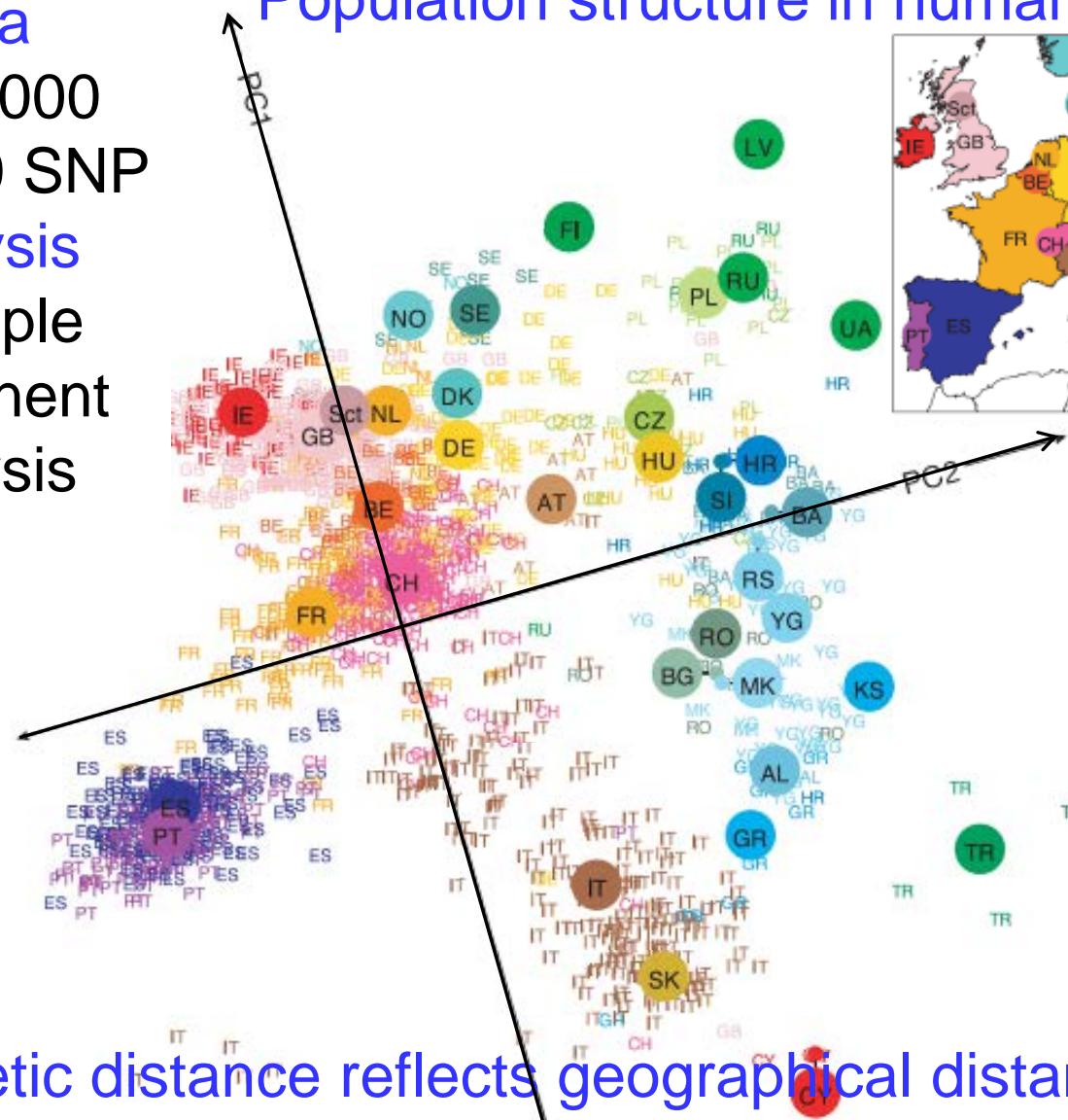
Introduction to the course

Data

$N = 3,000$

500,000 SNP

Analysis Principle component analysis



Novembre et al. (2008)

Genetic distance reflects geographical distance. Clinal variation.

Introduction to the course

Data

$N = 36$

300,000 SNP

Analysis

Principle
component
analysis

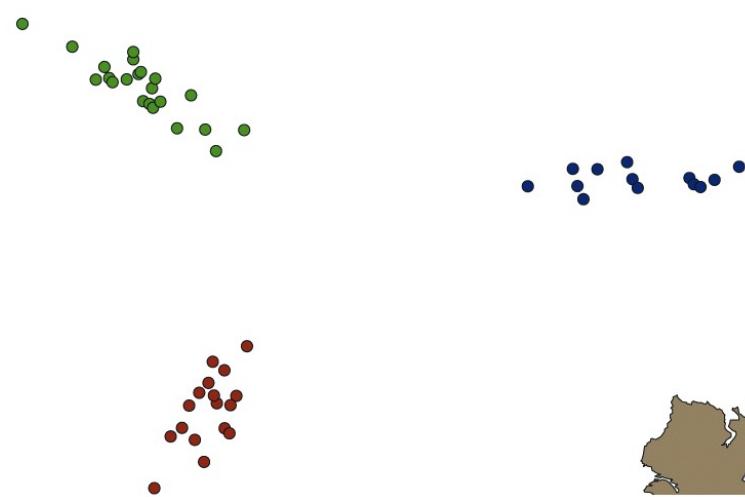
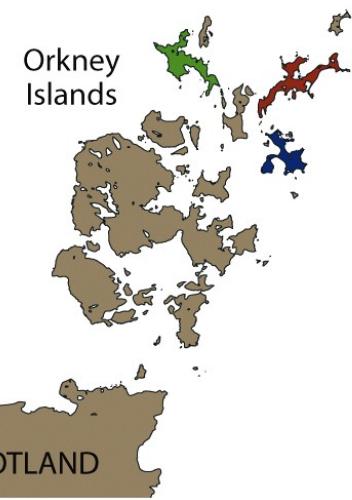


Figure 10.6b Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)



Discrete populations
(sampled individuals had all
four grandparents from the same isle)

How did the populations differentiate from each other?

Introduction to the course

- What is population genetics?

Population genetics studies the distribution of genetic variation in time and space

and

analyses the impact of the evolutionary forces

mutation

recombination

genetic drift

gene flow/migration

natural selection

Introduction to the course

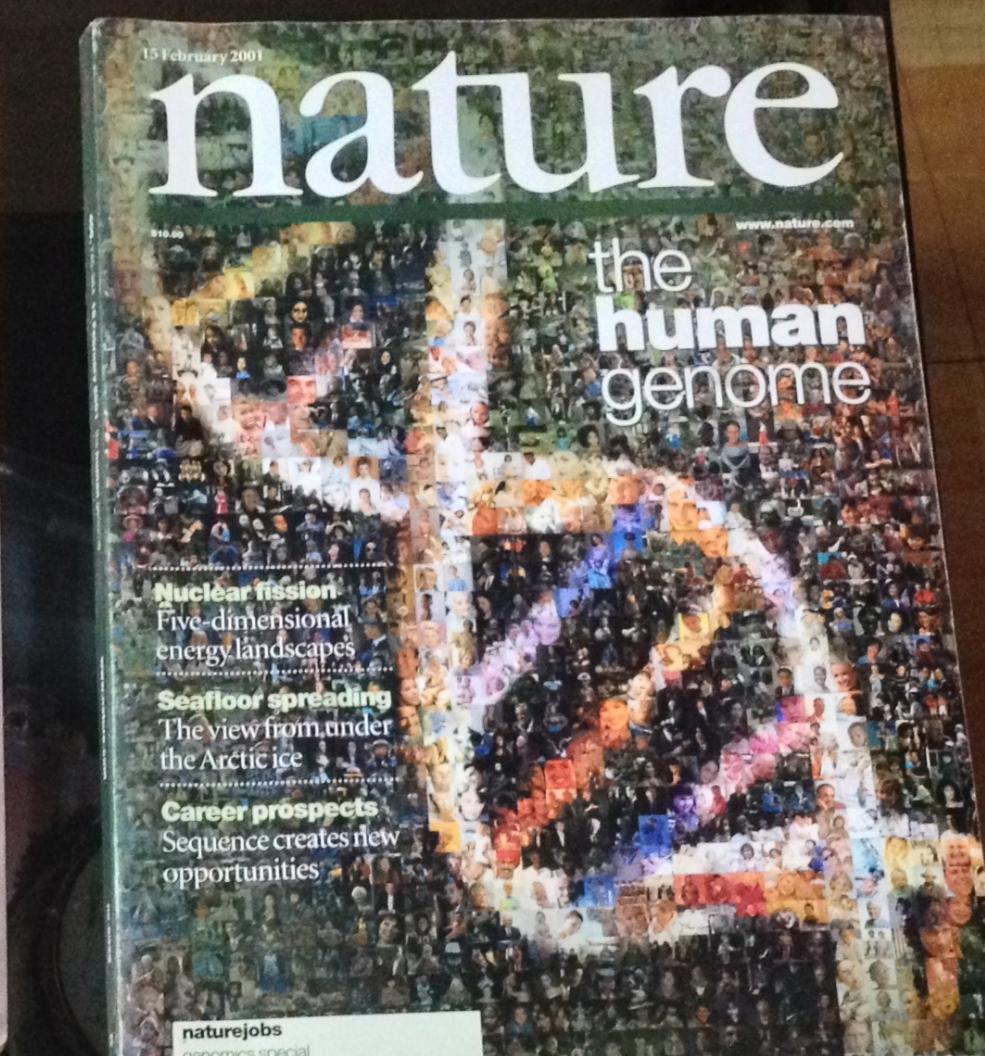
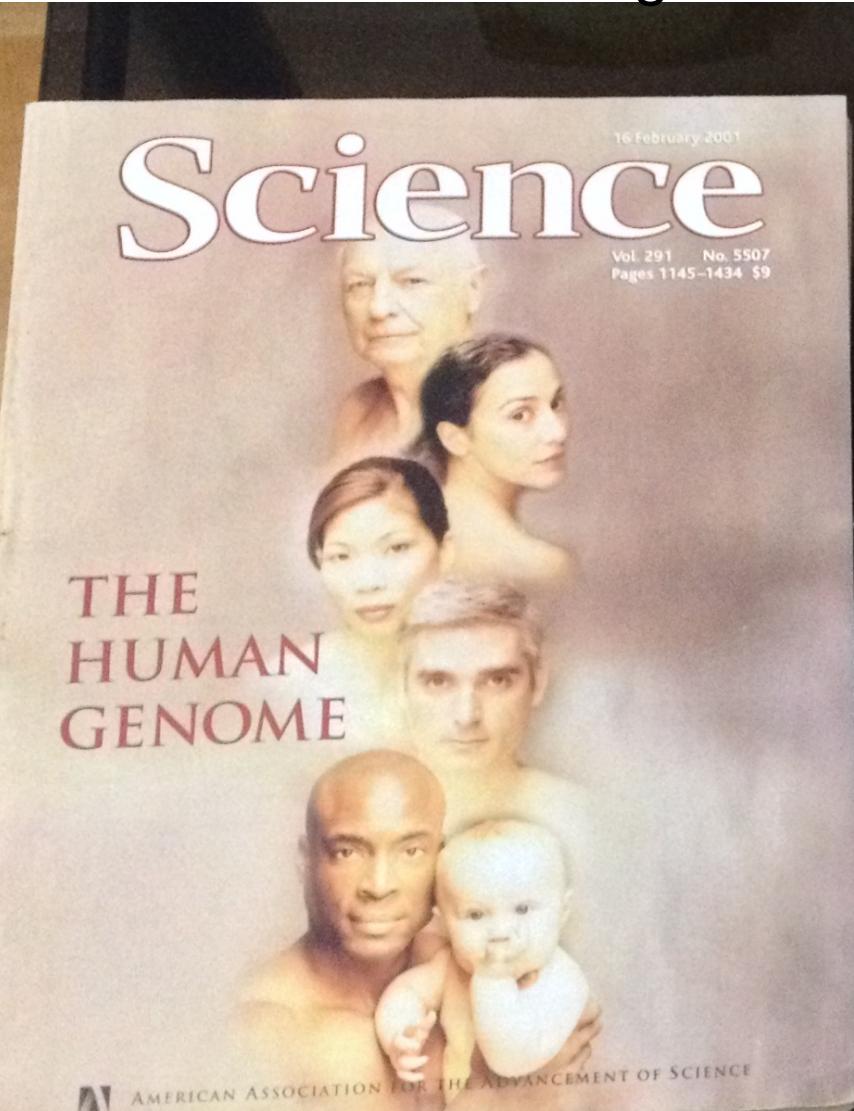
- Population genetics has important implications in a number of biological disciplines
 - Genome research
 - Evolutionary biology
 - Conservation genetics
 - Animal and plant breeding
 - Molecular ecology
 - Molecular medicine

Introduction to the course

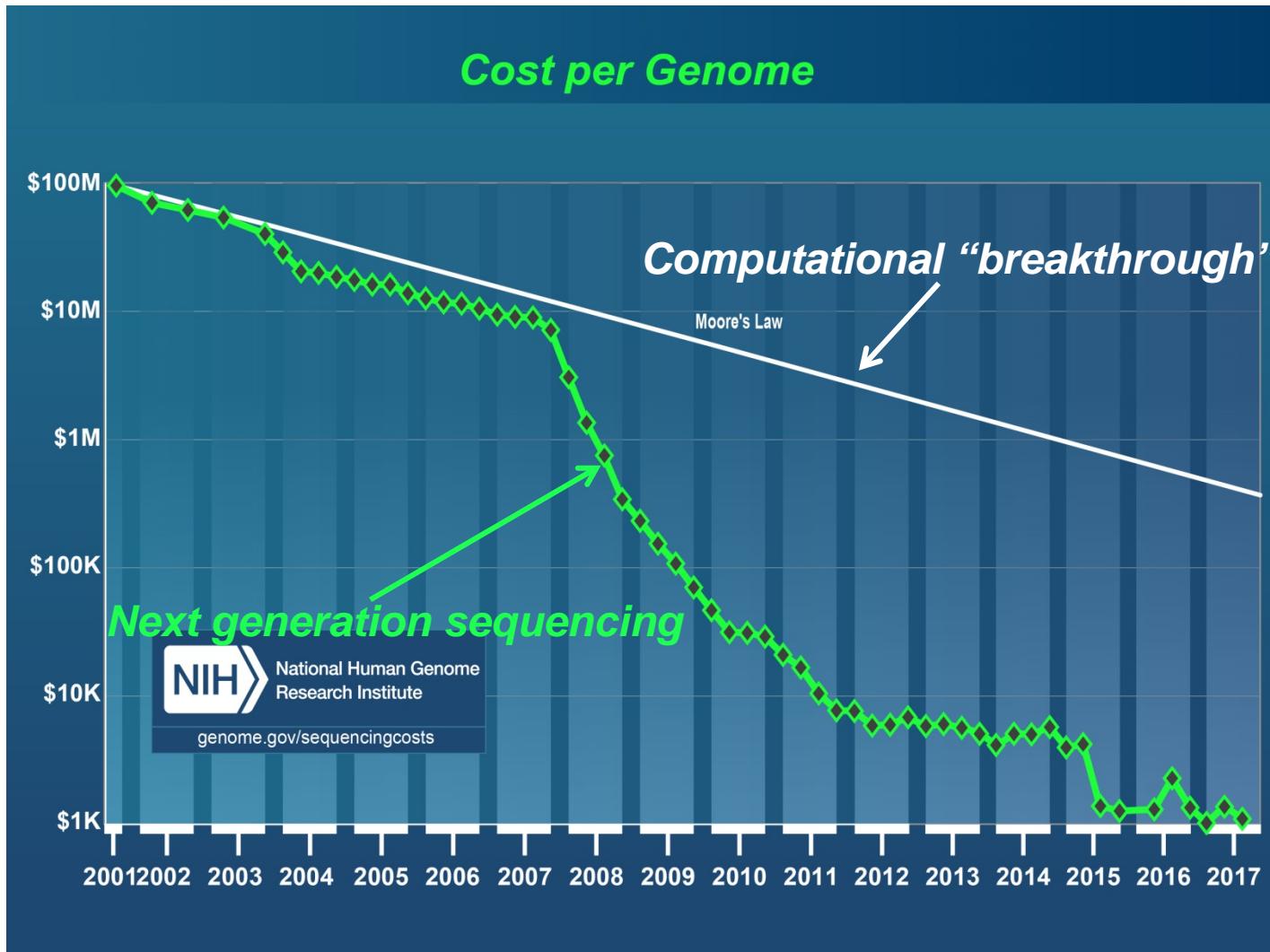
- Why is population genetics especially important today?
- Population genetics has been revolutionized in three important areas
 - Sequencing technology
 - Sequencing costs are drastically reduced
 - Computational breakthrough
 - Conceptually with Kingman's coalescence

Sequencing technology

The human genome (2001): 100,000,000\$



Sequencing technology



Computational breakthrough

The IBM Model 350 disk file with a storage space of 5MB from 1956 and a Micro SD Card



100,000 kg/Gb



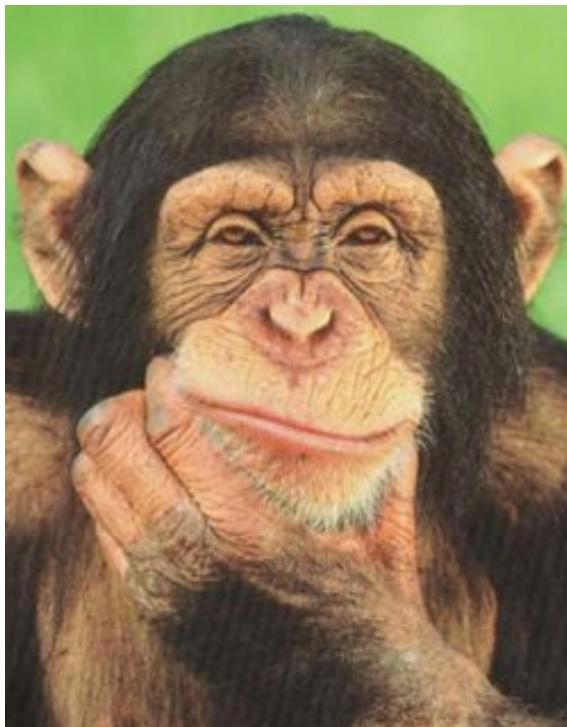
5 g

0.000,04 kg/Gb

Sequencing technology

Population genetics 2012

Chimpanzee History and Evolution



Chimpanzee (*Pan troglodytes*)



Christina Hvilsom

Sequencing technology

Population genetics 2012

Christina's PhD study:

2008: 25 microsatellites, 250 individuals

2010: Complete mitochondrial genomes, 90 individuals
(16,500 bp)

2011: Exome sequencing, 29 individuals

(\approx 23,000 protein coding genes

\approx 180,000 exons

\approx 1% of genome \approx 30 megabases)

2012: Complete genomes, 30 chimpanzees + 80 other primates
(\approx 3.1 gigabases)

2013

Great ape genetic diversity and population history

Javier Prado-Martinez^{1*}, Peter H. Sudmant^{2*}, Jeffrey M. Kidd^{3,4}, Heng Li⁵, Joanna L. Kelley⁴, Belen Lorente-Galdos¹, Krishna R. Veeramah⁶, August E. Woerner⁶, Timothy D. O'Connor², Gabriel Santpere¹, Alexander Cagan⁷, Christoph Theunert⁷, Ferran Casals¹, Hafid Laayouni¹, Kasper Munch⁸, Asger Hobolth⁸, Anders E. Halager⁸, Maika Malig², Jessica Hernandez-Rodriguez¹, Irene Hernando-Herraez¹, Kay Prüfer⁷, Marc Pybus¹, Laurel Johnstone⁶, Michael Lachmann⁷, Can Alkan⁹, Dorina Twigg³, Natalia Petit¹, Carl Baker², Fereydoun Hormozdiari², Marcos Fernandez-Callejo¹, Marc Dabad¹, Michael L. Wilson¹⁰, Laurie Stevenson¹¹, Cristina Camprubí¹², Tiago Carvalho¹, Aurora Ruiz-Herrera^{12,13}, Laura Vives², Marta Mele^{1†}, Teresa Abello¹⁴, Ivanela Kondova¹⁵, Ronald E. Bontrop¹⁵, Anne Pusey¹⁶, Felix Lankester^{17,18}, John A. Kiyang¹⁷, Richard A. Berg¹⁹, Elizabeth Lonsdorf²⁰, Simon Myers²¹, Mario Ventura²², Pascal Gagneux²³, David Comas¹, Hans Siegismund²⁴, Julie Blanc²⁵, Lidia Agueda-Calpena²⁵, Marta Gut²⁵, Lucinda Fulton²⁶, Sarah A. Tishkoff²⁷, James C. Mullikin²⁸, Richard K. Wilson²⁶, Ivo G. Gut²⁵, Mary Katherine Gonder²⁹, Oliver A. Ryder³⁰, Beatrice H. Hahn³¹, Arcadi Navarro^{1,32,33}, Joshua M. Akey², Jaume Bertranpetti¹, David Reich⁵, Thomas Mailund⁸, Mikkel H. Schierup^{8,34}, Christina Hvilsom^{24,35}, Aida M. Andrés⁷, Jeffrey D. Wall¹¹, Carlos D. Bustamante⁴, Michael F. Hammer⁶, Evan E. Eichler^{2,36} & Tomas Marques-Bonet^{1,33}

Population genetics course 2019

Course project:

Group work

≈ 4-5 students

at least 2 bioinformatics students plus

at least 2 biology students

Introduction to the course

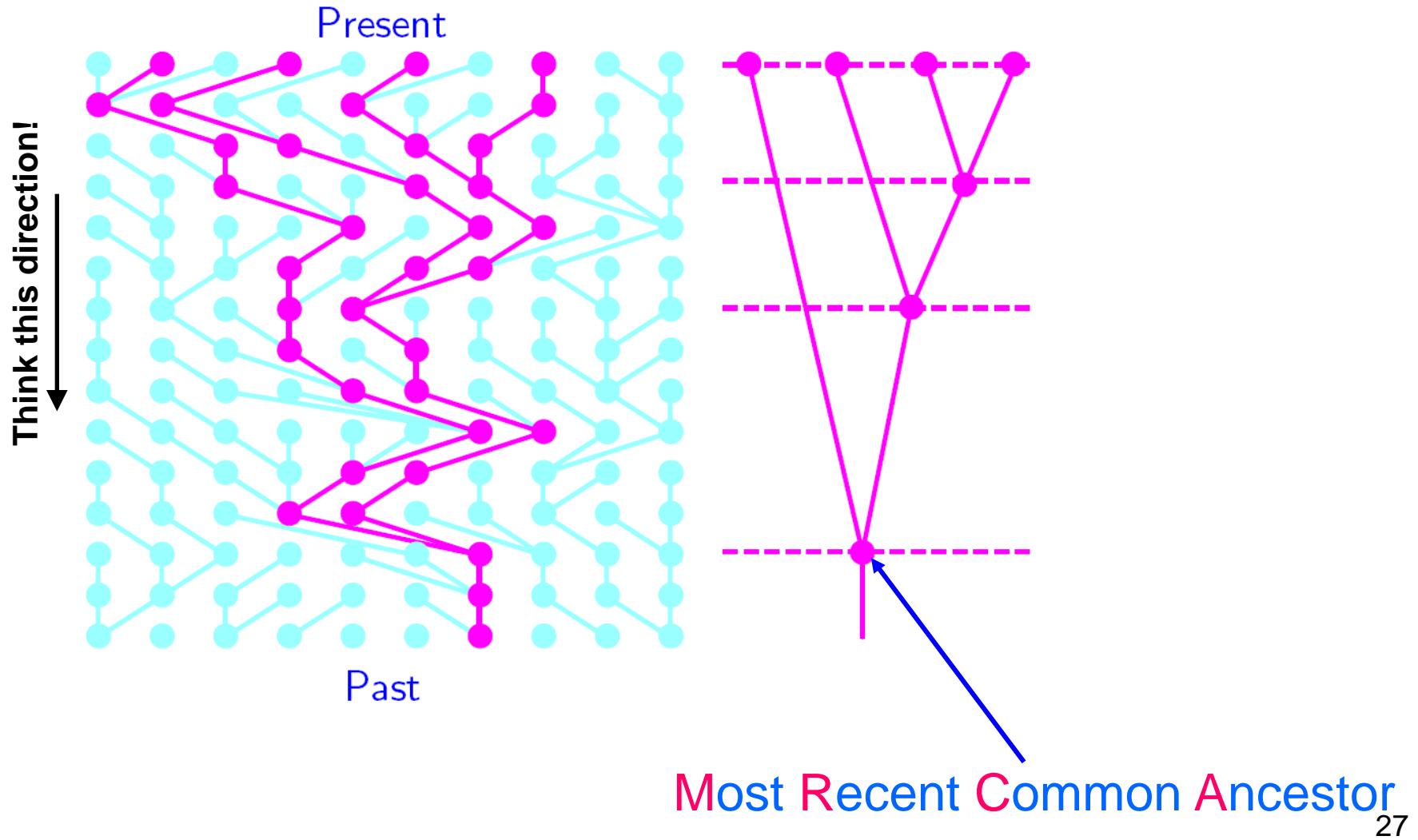
- Why is population genetics important today?
- Population genetics has been revolutionized in three important areas
 - Sequencing technology
 - Computational breakthrough
 - Conceptually with Kingman's coalescence



(1982)

John Kingman (1939-)

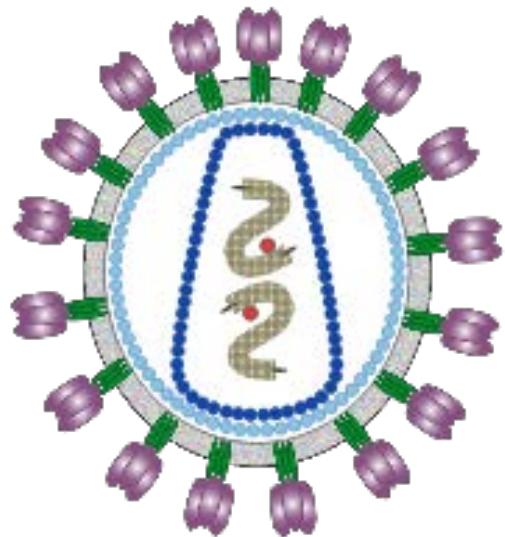
Coalescence process



Coalescence process

Evolution of HIV-1

Human immunodeficiency virus



Virus identified in 1983

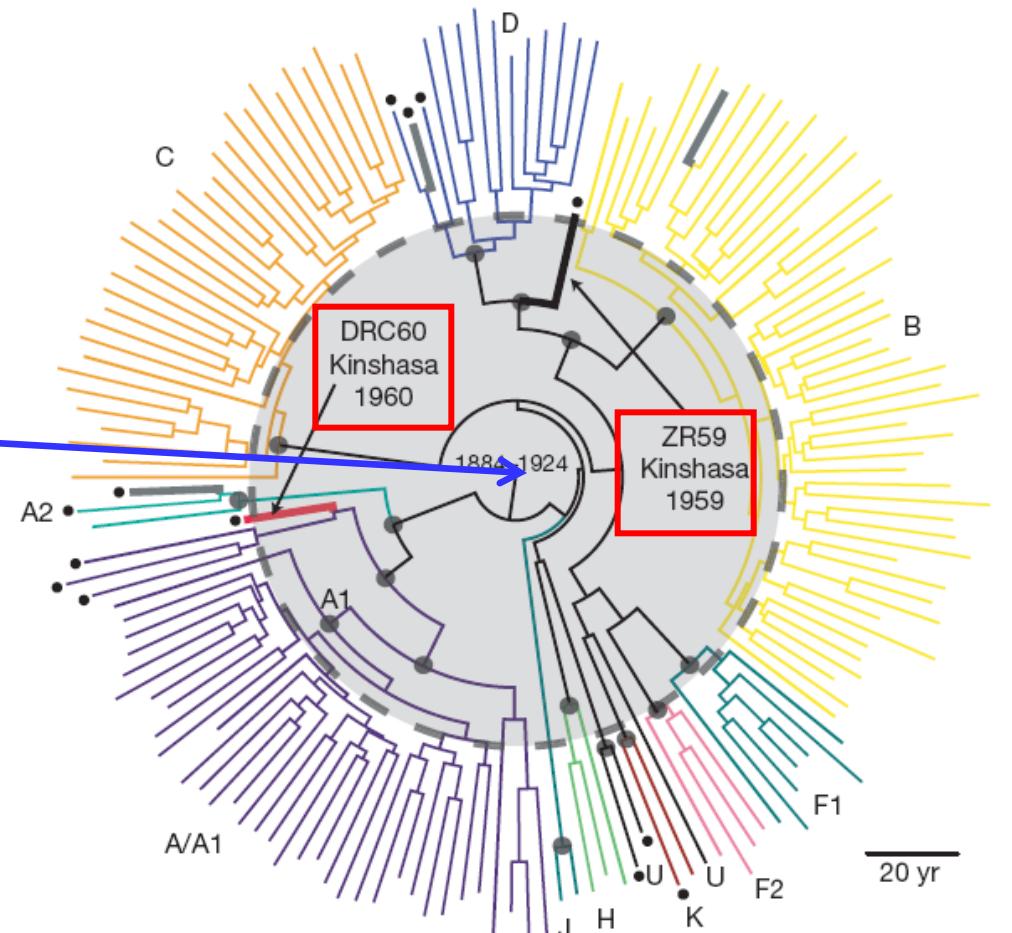
First cases of HIV-1 in 1959

Coalescence process

Evolution of HIV-1

How long has HIV-1
been among humans?

TMRCA: 1908



Time to
Most
Recent
Common
Ancestor

Worobey et al. 2008 Nature 455:661-665

Content of the course

- How do we quantify genetic variation in populations?
- Natural selection
- Genetic drift and inbreeding
- Coalescent: theory and applications
- The neutral theory and molecular population genetics
- Population structure
- Evolutionary quantitative genetics
- Population genomics
- Human population genetics

Content of the course

First six weeks

Lectures

Exercises

Paper & pen

Statistical analyses of data

Introduction to programs

Research lectures

Last two weeks

Project work with chosen subject or data set

Additional lectures/exercises for those that are inexperienced in Linux and R

Content of the course

Exercises

Based on up-to-date data sets
and analytical tools (many of them are Linux based)

Virtual Linux Server

You should have MobaXterm installed on your machine
(If you use Windows)

How to acces the Linux Server

Home

View progress

+ Module

Modules

Announcements

People

Assignments

Quizzes

Discussions

Files

The screenshot shows a course management interface. On the left is a sidebar with links: Home, View progress, + Module, Modules, Announcements, People, Assignments, Quizzes, Discussions, and Files. The Home link is highlighted with a red box. The main area displays a module structure. At the top is a section titled "Welcome to the population genetics course!" with three icons: a green checkmark, a plus sign, and a gear. Below it is a "Course Information" section with the same three icons. At the bottom is a section titled "A brief introduction to Linux and how to access the server" which is highlighted with a red border. This section also has the three standard icons.

⋮ Welcome to the population genetics course!

⋮ Course Information

⋮ A brief introduction to Linux and how to access the server

How to acces the Linux Server

Open a terminal and ssh (secure shell)

Login-server:

ssh -X USERNAME@ssh-bio-stud.science.ku.dk

Popgen server:

ssh -X USERNAME@popgen-bio.science.ku.dk

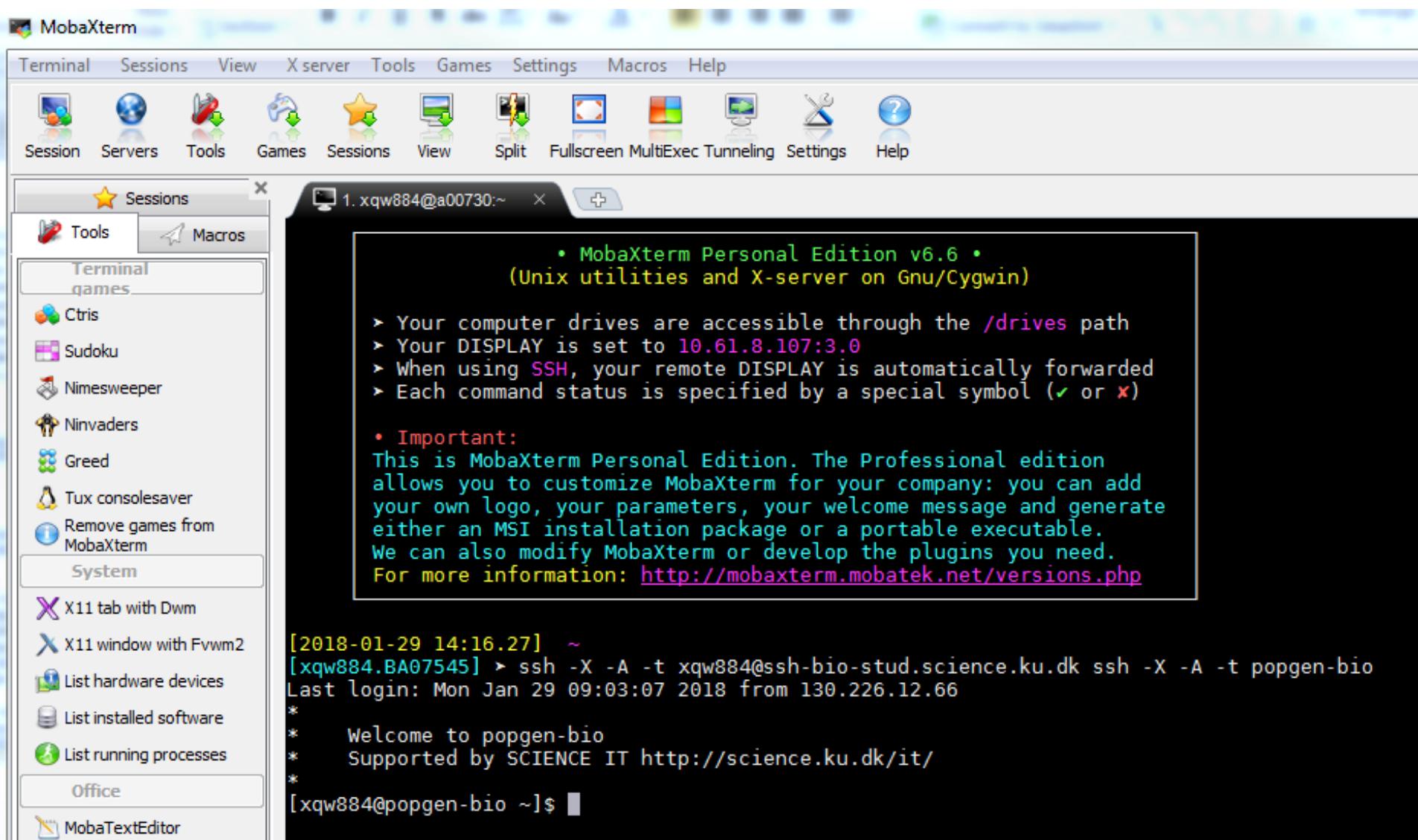
Combining the two steps

ssh -X -A -t USERNAME@ssh-bio-stud.science.ku.dk ssh -X -A -t popgen-bio

USERNAME is your KU-id , like abc123

The password is your KU password

How to acces the Linux Server



How to access the Linux Server

```
[2018-01-29 14:16.27] ~
[xqw884.BA07545] > ssh -X -A -t xqw884@ssh-bio-stud.science.ku.dk ssh -X -A -t popgen-bio
Last login: Mon Jan 29 09:03:07 2018 from 130.226.12.66
*
*   Welcome to popgen-bio
*   Supported by SCIENCE IT http://science.ku.dk/it/
*
[xqw884@popgen-bio ~]$ R    Starting R

R version 3.4.3 (2017-11-30) -- "Kite-Eating Tree"
Copyright (C) 2017 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.

> 2 + 2    Wow!!
[1] 4
> q()  #exit R again■  Exit R Everything to the right of # is a comment
```

Course schedule

Week	4 February	6 February
6	Monday	Wednesday
		8-12 (KV-S) Introduction to Linux and R (R part)
	13-17 Introduction to the course (HRS) History of population genetics (HRS) The Hardy-Weinberg law (L/E) (HRS) Introduction to Linux and R (RH) (Linux part)	13-16 (FR) Genetic drift and basic coalescence theory (L/E)

How to get most out of the course

Read the textbook and background literature

Download (and read) the exercises

(paper and pen should be solved)

Download the data

before you come!

Workload

Category	Hours
Exam	16
Preparation	81 
Lectures	27
Project work	55
Theory exercises	27
Total	206

It is a course about population genetics, not about R and Linux

Exercise in estimating nucleotide diversity

Estimating the nucleotide diversity along the chromosome

```
## Function for generating sliding windows
slidingwindowplot <- function(mainv, xlabv, ylabv, ylimv, window.size,
step.size,input_x_data,input_y_data)
{
if (window.size > step.size)
  step.positions <- seq(window.size/2 + 1, length(input_x_data)- window.size/2,
  by=step.size)
else
  step.positions <- seq(step.size/2 + 1, length(input_x_data)- step.size,
  by=step.size)
n <- length(step.positions)
means_x <- numeric(n)
means_y <- numeric(n)
for (i in 1:n) {
  chunk_x <- input_x_data[(step.positions[i]-
  window.size/2):(step.positions[i]+window.size-1)]
  means_x[i] <- mean(chunk_x,na.rm=TRUE)
  chunk_y <- input_y_data[(step.positions[i]-
  window.size/2):(step.positions[i]+window.size-1)]
  means_y[i] <- mean(chunk_y,na.rm=TRUE)
}
plot(means_x,means_y,type="b",main=mainv,xlab=xlabv,ylab=ylabv,ylim=ylimv,cex=0.25,
pch=20, cex.main=0.75)
vec <- c(0.025,0.5,0.975)
zz <- means_y[!is.na(means_y)]
abline(h=quantile(zz,0.025,na.rm=TRUE),col="blue")
abline(h=quantile(zz,0.925,na.rm=TRUE),col="blue")
abline(h=mean(input_y_data))
}
```



But you are welcome
to explore the code...

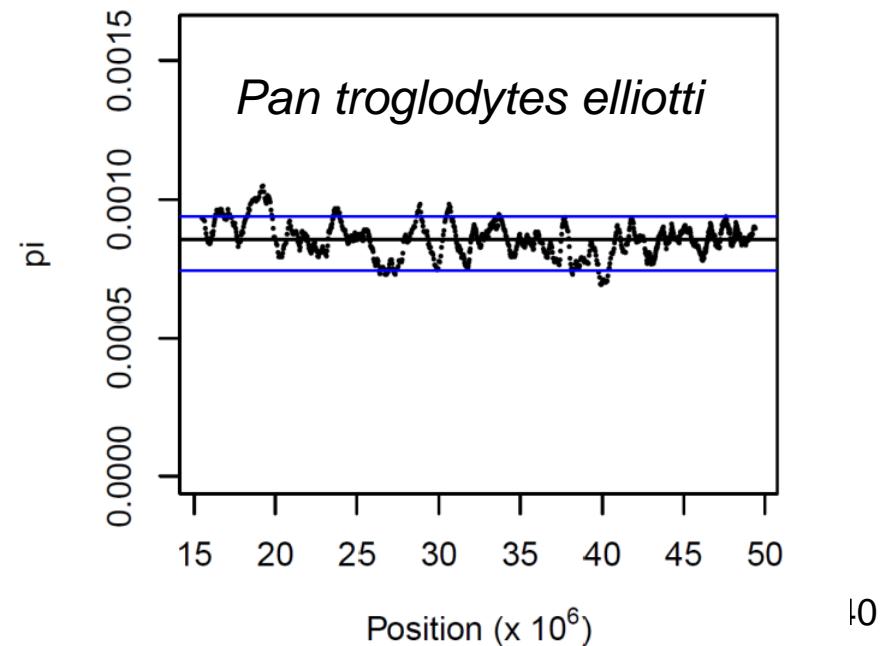
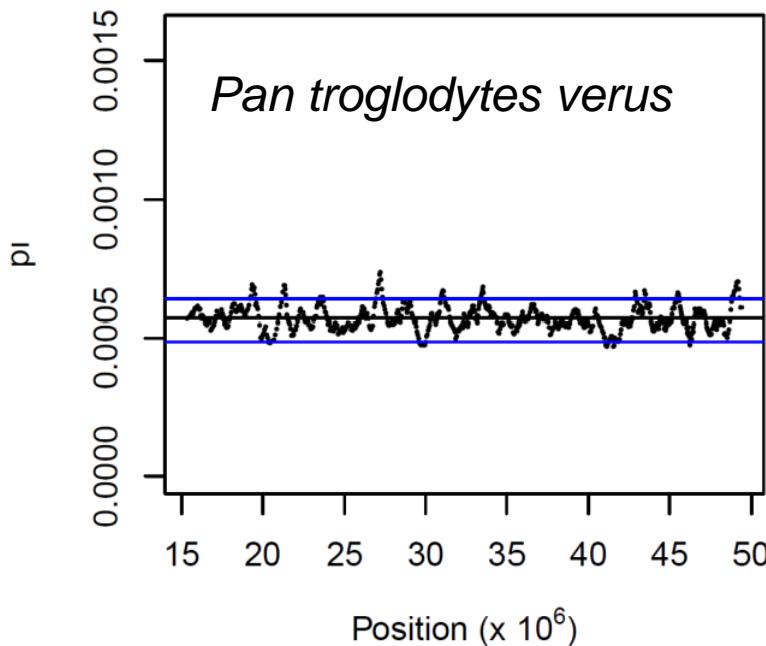
It is a course about population genetics, not about R and Linux

Exercise in estimating nucleotide diversity

Estimating the nucleotide diversity along the chromosome

Is the diversity evenly distributed along the chromosomes?

Do subspecies differ among each other?

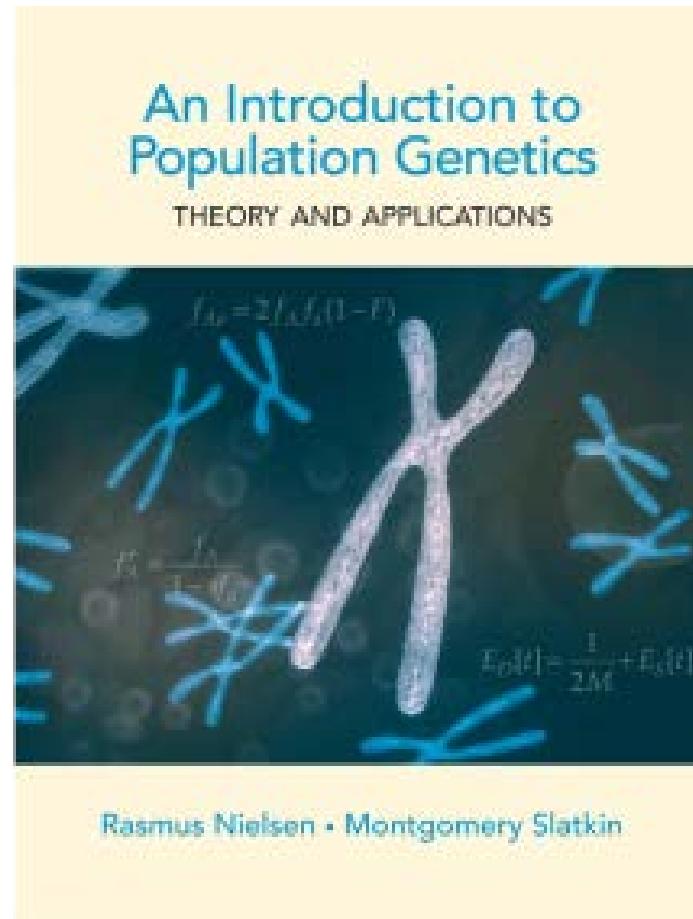


Recommended reading

Nielsen, R. & M. Slatkin 2013.

An Introduction to Population Genetics: Theory and Applications.
Sinauer Associates.

Supplemented with other
material (free).



Rasmus Nielsen



Montgomery
Slatkin



Center for Theoretical
Evolutionary Genomics
University of California, Berkeley

Other good books

Hahn, M. 2018

Molecular Population Genetics, Sinauer Associates.

Joe Felsenstein 2016.

Theoretical Evolutionary Genetics.

Free pdf available at

<http://evolution.gs.washington.edu/pgbook/pgbook.html>

Hedrick, P. W. 2011.

Genetics of Populations, 4. ed., Jones and Bartlett Publishers

Hartl, D.L. & A.G. Clark 2007

Principles of Population Genetics, Fourth edition, Sinauer
Associates

Exam

Paper written about population
genetic analyses of real data sets.
(Group work, 4-5 students)

Title

Authors

Abstract

Introduction

Material & Methods

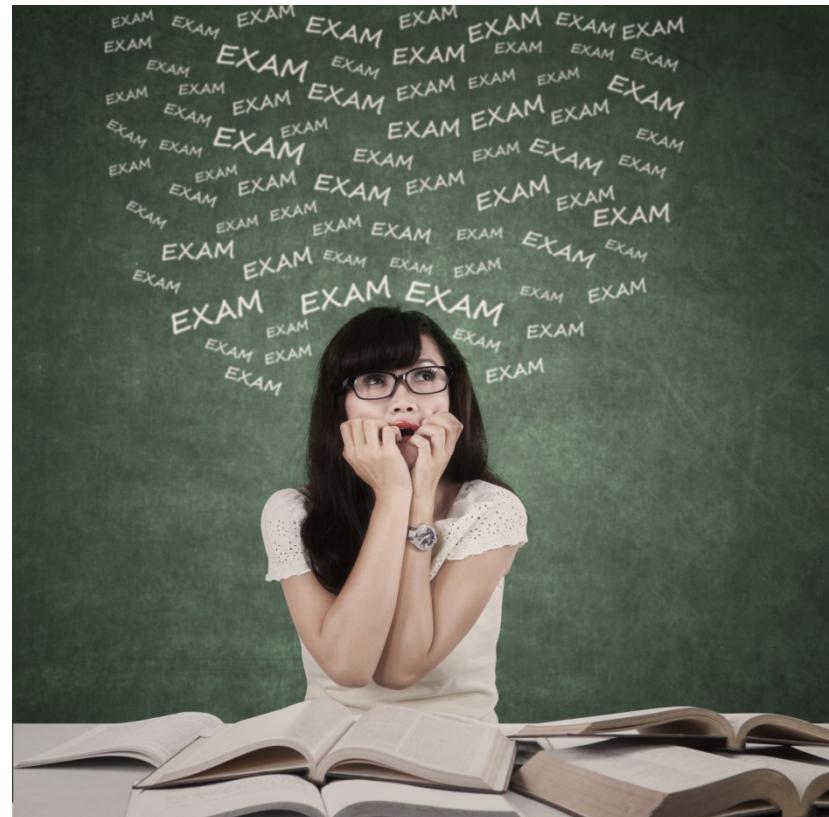
Results

Discussion

References

Uploaded on Absalon

(with automatic check for plagiarism)



Paper

The population genetic structure and variability of wild and captive common chimpanzee (*Pan troglodytes*) populations

Darren C. B. Hassan¹, Michael Gervais¹, Michael Gillingswater¹, Eric Lai¹, Kristina Krouse¹, Alexander Ring¹

The remaining categories (i.e., unadjusted) populations are ones, resulting from the removal of all bias through Categorical variables, or the higher level variable of *Group*. The last two categories are the adjusted populations, which are the same as the unadjusted populations, except that they have been adjusted for the categorical variables in the model (Cox & Wermuth, 2013). The different categories are difficult to distinguish when examining the results, as they are often used when referring to categories of population in the literature. In this study, the adjusted populations were used, as the unadjusted populations were not clearly defined. There are three following that I will elaborate designated *C* in *Table 1* as *C* (unadjusted), *C* (adjusted), *C* (adjusted and *G*), and *C* (adjusted and *G* + *M*) (Cox & Wermuth, 2013).

i) Explore the chimpanzee non-human primate and human origins, genetic diversity, compare it to the wild populations

References

- similar diversity, especially for highly polydispersive microsatellite markers, and in cases where assumptions for traditional calculations are violated (Just 2006).
- For genetic population assignment analysis, we used GENALICA2 (Peyr et al. 2014), implementing the Bayesian assignment criteria of Falush and Stephens (2003), which was recommended as the best

act of assignment interaction by Condit et al. (2008). The population assignments, individuals were highlighted as potential migrants which either had a higher posterior probability of originating in the 'resident' population, or any individual that had a 'less posterior probability' of originating in its own population. We present the assignment results probabilistically illustrating the methodology of Leiserson and 2010. For each reference population, the values between the 1st and 99th percentile of the posterior distribution of language probabilities. These values display the language posterior distributions for the 'W' as a consequence of an individual drawn from the population, and should be regarded as a 'mimic' application of both the overall spread of the population, and of a resulting community.

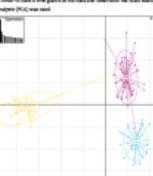
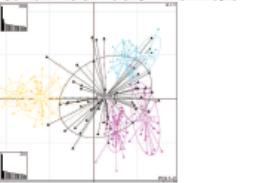


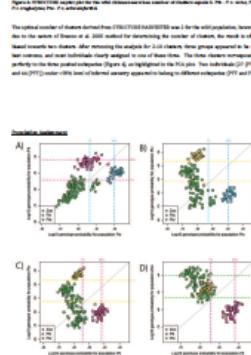
Figure 5 Principal Component Analysis of genetic differentiation of wild *Oligosoma*. Mean values of the multilocus genet identity populations PPT1 = P. punctatum, PPT2 = P. maccanni, PPT3 = P. laeve and PPT4 = P. laeve var. *laeve*

From the PCA it can be seen that the wild-type samples are distributed into three distinct populations. These three clusters match well with the previously defined subtypes, with *A. megasporus*(PM) and *A. schoenherri*(PH) samples being most closely related, while the third cluster (*A. fumigatus*, *A. flavipes*(TF)) is more distant. Two individuals that are not clustering with their subtypes were identified, *A. fumigatus*(TF) that was grouping with *A. schoenherri* samples and *A. flavipes*(TF), falling in the *A. fumigatus* cluster (Figure 7).



The Principal Component analysis of the wild and captive populations demonstrates there being a distinctness of the population individuals in those belonging to each of the three original wild subspecies (*P. t. tenuis*, *P. t. agassizii* and *P. t. macrourus*), as well as those being a heterogeneous mixture of hybrids in the one population (Figure 2). A continuous shift of the one population individuals to the 3 subspecies in relation to the wild species can be observed. This might also due to a relatively large number of mixing events in the one population as well as people performing PCR, thus making some individuals to have the 3 subspecies.

In Figure 1, we present a typical example of a group difference of standard T1 hybrids for potassium uptake and translocation in the presence of different concentrations of PTK. It is evident that PTK is incorporated in all genotypes.



With measures and $\text{Jaccard } D_{ij}$ values for the distances between the wild subspecies correspond to the D_{ij} values, however, additional level of differentiation was reported, showing the two populations to be nearly divergent from the $P. t. antillensis$ and most closely related to the $P. t. mearnsi$ (Table 1). The Hausdorff distance also agrees with the patterns.

respectively) than the one they were assigned to (corresponding to $\text{X}(\text{A})$ target). Dispersion of scores (0 to 6) on PPI and 40% on PI target points, suggesting that it's a mixed anxiety test.

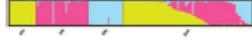


Figure 5. The 1000X genome project for the wild rice *Oryza sativa* (cv. Nipponbare) has been used to test the *h* allele at the *H* locus. A *h* allele was found.

- Individuals with lower than 90% inferred ancestry in one of the three clusters were then eliminated. All 1000X accessions were then re-phenotyped with the 104 "young" accessions, together with the original 1000X accessions. The results are shown in Table 1. The proportion of individuals with a *h* allele was 0.15 in PFTT, 0.16 in PFTL, and 0.17 in PFTC (not significantly different).
- The remaining 1000X accessions had a hybrid ancestry. The mean inferred ancestry of these 46 accessions was 0.91, 0.91 in PFTT, 0.91 in PFTL and 0.92 in PFTC.
- The 1000X genome project for the wild rice species population (101 individuals with only) suggested there is no linkage between *H* and *h*.
- Our findings are consistent with an additional possible hybrid status.

Table 3: Mean allele patterns across populations in the different demographic periods			
Population	n. samples	n. individuals	n. individuals
No. of loci (nucleotide)	10.24	11.02	11.02
No. of alleles (nucleotide)	4.06	4.03	4.06
No. of different alleles	4.06	4.04	4.06
Standard deviation (nucleotide)	1.77	2.05	2.05
No. Private alleles	0.76	1.09	1.06
Expected heterozygosity	0.71	0.65	0.62
Variance in expected heterozygosity	0.75	0.66	0.62

There are a number of different ways of measuring genetic diversity. In the table above (Table 1), the three subspecies are described in terms of the number of different alleles, number of private alleles and Shannon's information per locus. Most of these measures indicate that the P. t. mitchillae subspecies appears to have a higher genetic diversity than P. t. tenebrosa (it is unshaded).

Individuals can also potentially be assigned to the PHS and PTY groups, however the highest fraction of the negative chimpanzees are unlikely to belong to any cluster and must be recognized as hybrids (Jones et al.).

Discussion
 We found that in general, the catchback times the wild were used to reoccupy soil in the widely recognized 4-step model is 0.1, 0.4, 0.8 and 1.2 years. The time-limited step-backs were and reported by the Principle Component analysis method, as well as by the clustering method in PROSTIM. The theoretical model predicted slightly exaggerated time-limits being the best scenario for the wild populations. There are also some differences between the results of the cluster analysis in the higher range (1-1.5) as compared to a limited range of 0.1-0.8. As the cluster analysis corresponds particularly to the initial step-back scenario, it is expected that the results will be more accurate.

and genetic gains from clearly being a substantial fraction of individuals that manage to immigrate to any of the wild populations. The arrangement was also stable stable after applying the STRUCTURE algorithm to the first 1000 time steps. These individuals were not of mixed descent and became the result of interbreeding between subpopulations in time.

using equally distant in all three subspecies. However, the F_{ST} value and FST's distances supported the subspecies, which was also clearly visible in the population assignment analysis.

For a detailed analysis of the information, a number of potential methods have been tested. The results of these analyses are presented below. A detailed description of the methods used can be found in a long report of the study (Bergman et al., 1992). In this paper we will focus on the results. The following reports are very new. Compared with our earlier data (involving older or experienced patients), the present data are more complete and more likely to reflect medical reality about the effectiveness of the treatment. Furthermore, an attempt has been made to compare the results with those from other studies. The results from these studies have been analysed by means of grouping them according to different variables that may influence the outcome. These were time since stroke and whether or not they were still working at the time of the study.

In addition to the important material, educational and scientific missions, some areas as measures and assessment of cultural diversity, in terms of individuals, and their greater material. It is important that the new mission

Exam

Oral presentation (individually)

20 minutes:

10 minutes

Presentation of project

10 minutes

Discussion of project/presentation

Population genetic questions

(Which cover the whole textbook
except chapter 10)

2, maybe 3 days in week 15



Presentation

Population genetics

UNIVERSITY OF COPENHAGEN

A population genetic approach to the investigation of the structure and diversity of plains zebra

Ricard Argelaguet
April 2015

Introduction

Morphology-based classification into six subspecies:

- *E. burchelli*
- *E. boehmi*
- *E. quagga*
- *E. Chapmani*
- *E. burchellii*
- *E. quagga* (extinct)

Figures: Leppla et al. (2008); Lovejoy et al. (2008)

Aim of the project:

Use genetic data to answer:

- (1) Does the taxonomic classification reflect the current population structure of the plains zebra?
- (2) Is the quagga a plains zebra?

Materials and methods

174,226 SNP data (20,501 after filtering) from 59 individuals

- *E. q. boehmi* (30)
- *E. q. boernensis* (3)
- *E. q. burchelli* (5)
- *E. q. chapmani* (8)
- *E. q. crassicaudata* (14)
- *E. q. quagga* (1)

Results: principal component analysis

The individuals belonging to the same subspecies tend to cluster together, but there is no clear distinction between subspecies.

Climat trend that matches with the geographical distribution

The single quagga specimen falls within the burchelli group.

Results: unrooted phylogenetic tree

Some subspecies define deep monophyletic groups (Burchelli, Boehmi, Chapmani), whereas other subspecies form polyphyletic clusters (Burchelli, Chapmani).

The single quagga specimen falls within the burchelli group.

Results: summary genetic statistics

Subspecies	N	H _e	H _W	F _{is}	
Jackson	30	0.00303	0.00313	0.00177	
Boehmi	3	0.00296	0.00250	0.00541	
Chapmani	8	0.00418	0.00457	1.0	-0.07704
Burchelli	5	0.00375	0.00325	1.0	0.03215
Crassicaudata	14	0.00476	0.00466	1.0	0.03768
TOTAL	60	0.00305	0.00319	0.00262	

Table 1: Summary genetic statistics for each subspecies and for the total plain zebra population: number of individuals (N), expected heterozygosity under Hardy-Weinberg proportions (H_e), nucleotide diversity per site (*r*), fraction of SNPs in Hardy-Weinberg proportions (H_W), inbreeding coefficient (F_{is})

Results: pairwise Fst

Subspecies	Jackson	Boehmi	Chapmani	Burchelli	Crassicaudata
Jackson	-	0.06034	0.05891	0.04330	0.0211773
Boehmi	-	0.11447	0.10056	0.097193	
Chapmani	-	-	0.0900833	0.0699572	
Burchelli	-	-	-	0.0485734	
Crassicaudata	-	-	-	-	

Table 2: Fst coefficients between each pair of subspecies

Results: admixture analysis

K=5 we don't observe clear separation in taxonomic subspecies

K=2 we observe a clear climat trend that matches geography

K=3 boernensis appear as a different subpopulation

Conclusion

Was the extinct quagga a plains zebra?

According to genetic data, the quagga falls within the diversity of the plains zebra, in particular among the burchelli individuals.

E. q. boernensis, *E. q. chapmani*, *E. q. quagga*

Conclusion

Does the taxonomic classification reflect the current population structure of the plains zebra?

The morpho-geographic classification is only partially consistent with genetic data.

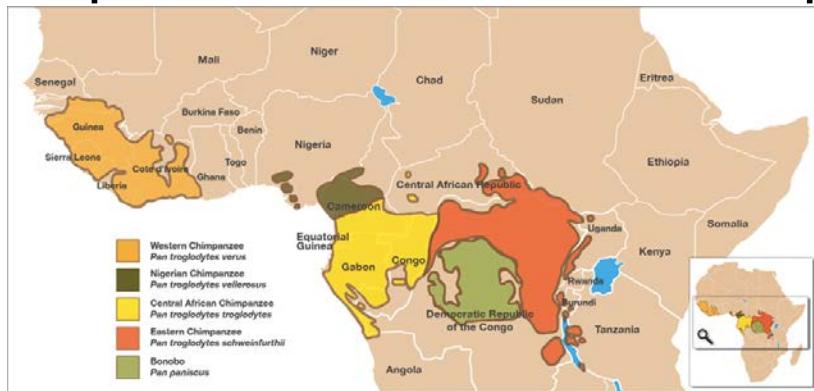
There is a continuous genetic change from North-East to South and we don't distinguish clear discrete genetic clusters, except for boernensis.

Limitations

- In our initial hypothesis we considered the subspecies to be isolated populations, thus the genetic statistics calculated here assume the existence of such groups.
- Sample size: boernensis (n=3) and burchelli (n=5)
- Locations of sampling of the individuals
- Human influence

Projects

- Selection in the exomes of chimpanzee Subspecies
- Population structure of chimpanzees



- Mountain gorilla
Effect of bottleneck on genomic structure
- Population structure in gorillas



Challenges and opportunities in the course

Challenge in this course:

Students background:

Bioinformatics

Biology

plus other biological disciplines

Opportunities in this course:

Course project:

Group work

≈ 4-5 students

at least 2 bioinformatics students plus

at least 2 biology students

Improvement of course

Regular (few short!) meetings with ≈ 4 students
With different backgrounds
Volunteers?

External lecturers

Bo T. Simonsen

Head of Section of Forensic Genetics

Modern crime-buster: The population geneticist

DNA Testing Has Exonerated Several Hundred Wrongfully Convicted People



The Innocents
The Innocents Project, New York

The Innocents

U.S. Supreme Court to rule on DNA 'fingerprinting'

The U.S. Supreme Court is going to determine whether Maryland's decision to collect DNA samples from people arrested for serious crimes represents an unconstitutional invasion of privacy or a crime-solving breakthrough with the potential to be the "fingerprinting of the 21st century."

Survey based on Socrative

A screenshot of a web browser window titled "Socrative". The address bar shows the URL "b.socrative.com/login/student/". The main content area displays the Socrative logo and the text "Enter room number 265011 and Join room". A red arrow points from the text "Please go to http://b.socrative.com/login/student/" to the address bar. Another red arrow points from the text "Enter room number 265011 and Join room" to the "Join Room" button on the right side of the screen.

Hans R. X

b.socrative.com/login/student/ Hans R. X

Please go to <http://b.socrative.com/login/student/>

Welcome to Socrative!

Enter your teacher's room below

Room Name

Join Room

Enter room number 265011 and Join room