LEIDEN UNIVERSITY MEDICAL CENTER

# Single molecule sequencing in bacterial metagenomics

**Jeroen F. J. Laros**

**Leiden Genome Technology Center**

**Department of Human Genetics**

**Center for Human and Clinical Genetics**

# Introduction

***Metafor***

NWO Forensic Science grant.

*Metagenomic analysis of Forensic biological traces.*

- Analyse all present DNA in a forensic trace.
- Single molecule sequencing.

Pilot *ancient DNA* conducted at the LUMC.

**http://www.nwo.nl/forensicscience/**

# Introduction

## *Ancient DNA*

- Orlando L., Ginolhac A., Raghavan M., Vilstrup J., Rasmussen M., Magnussen K., Steinmann K.E., Kapranov P., Thompson J.F., Zazula G., Froese D., Moltke I., Shapiro B., Hofreiter M., Al-Rasheid K.A., Gilbert M.T., and Willerslev E. True single-molecule dna sequencing of a pleistocene horse bone. *Genome Research*, 21(10):1705–1719, October 2011.

# Introduction

## *Ancient DNA*

- Orlando L., Ginolhac A., Raghavan M., Vilstrup J., Rasmussen M., Magnussen K., Steinmann K.E., Kapranov P., Thompson J.F., Zazula G., Froese D., Moltke I., Shapiro B., Hofreiter M., Al-Rasheid K.A., Gilbert M.T., and Willerslev E. True single-molecule dna sequencing of a pleistocene horse bone. *Genome Research*, 21(10):1705–1719, October 2011.

Differences with modern DNA:

- Deamination (**C** $\Rightarrow$ **U**) in combination with PCR:
  - **C** $\Rightarrow$ **T**.
  - **G** $\Rightarrow$ **A**.
  - In short: the **GC** percentage drops.

# Introduction

## *Ancient DNA*

- Orlando L., Ginolhac A., Raghavan M., Vilstrup J., Rasmussen M., Magnussen K., Steinmann K.E., Kapranov P., Thompson J.F., Zazula G., Froese D., Moltke I., Shapiro B., Hofreiter M., Al-Rasheid K.A., Gilbert M.T., and Willerslev E. True single-molecule dna sequencing of a pleistocene horse bone. *Genome Research*, 21(10):1705–1719, October 2011.

Differences with modern DNA:

- Deamination (`C` $\Rightarrow$ `U`) in combination with PCR:
  - `C` $\Rightarrow$ `T`.
  - `G` $\Rightarrow$ `A`.
  - In short: the `GC` percentage drops.
- Short fragments.
  - Can not be amplified.

# Introduction

## *Ancient DNA*

- Orlando L., Ginolhac A., Raghavan M., Vilstrup J., Rasmussen M., Magnussen K., Steinmann K.E., Kapranov P., Thompson J.F., Zazula G., Froese D., Moltke I., Shapiro B., Hofreiter M., Al-Rasheid K.A., Gilbert M.T., and Willerslev E. True single-molecule dna sequencing of a pleistocene horse bone. *Genome Research*, 21(10):1705–1719, October 2011.

Differences with modern DNA:

- Deamination (**C** $\Rightarrow$ **U**) in combination with PCR:
  - **C** $\Rightarrow$ **T**.
  - **G** $\Rightarrow$ **A**.
  - In short: the **GC** percentage drops.
- Short fragments.
  - Can not be amplified.

Preferential amplification of modern DNA.

# Introduction

## *Ancient DNA*

Possible solutions:

- Moderate denaturation temperature (80 instead of 95 degrees Celsius).
- Single molecule sequencing.

# Introduction

## *Ancient DNA*

Possible solutions:

- Moderate denaturation temperature (80 instead of 95 degrees Celsius).
- Single molecule sequencing.

What applies for ancient DNA, probably also applies for mixed samples i.e., a metagenome:

- Human DNA has a **GC** percentage of around 40%.
- E.coli DNA has a **GC** percentage of around 50%.
- May vary greatly for the organisms of interest.

# Single molecule sequencing

## *Helicos true Single Molecule Sequencing*


Figure 1: HeliScope

Characteristics:

- No amplification.
- Short reads (±32bp).
- Relatively high error rate (±4%).
- Direct RNA sequencing.

# Single molecule sequencing

## *Pacific Biosciences Single Molecule, Real-Time*

Figure 2: PacBio RS

# Single molecule sequencing

## *Pacific Biosciences Single Molecule, Real-Time*

Characteristics:

- Long reads (several kilobases).
- High error rate (15-20%).
- Relatively high throughput (comparable with the Roche 454).

# Single molecule sequencing

***Pacific Biosciences Single Molecule, Real-Time***

Characteristics:

- Long reads (several kilobases).
- High error rate (15-20%).
- Relatively high throughput (comparable with the Roche 454).

Circular consensus sequencing.

- Sequence the same molecule several times.
- Extremely high accuracy.
- Acceptable read length ($\pm$250bp).

# Ancient DNA study

## *Sample retrieval*

Figure 3: Excavation in Eindhoven

# Ancient DNA study

***Sample description***

5 ancient DNA samples from teeth:

- Roche 454.
- Helicos.

4 modern DNA samples (mouth swabs):

- Helicos.

# Ancient DNA study

***Sample description***

5 ancient DNA samples from teeth:

- Roche 454.
- Helicos.

4 modern DNA samples (mouth swabs):

- Helicos.

The first objective was to find familial relationships.

A lot of unclassified DNA was left.

# Ancient DNA study

***Objective***

What else can we learn from these samples?

Method:

- Remove Human reads.

# Ancient DNA study

***Objective***

What else can we learn from these samples?

Method:

- Remove Human reads.
- Quality control.
  - Align data sets to each other.

# Ancient DNA study

***Objective***

What else can we learn from these samples?

Method:

- Remove Human reads.
- Quality control.
  - Align data sets to each other.
- Further analysis:
  - 16S rRNA database.
  - NCBI shotgun databases.
  - Full genomes.
  - De novo assembly.

# Ancient DNA study

***Objective***

What else can we learn from these samples?

Method:

- Remove Human reads.
- Quality control.
  - Align data sets to each other.
- Further analysis:
  - 16S rRNA database.
  - NCBI shotgun databases.
  - Full genomes.
  - De novo assembly.

# Ancient DNA study

## *Quality control*

| | | Roche 454 | | | | |
|---|---|---|---|---|---|---|
| | | A1 | A2 | A3 | A4 | A5 |
| Helicos | A1 | 0.200 | 0.011 | 0.050 | 0.360 | 0.600 |
| | A2 | 0.200 | 0.006 | 0.040 | 0.060 | 0.160 |
| | A3 | 0.400 | 0.013 | 0.110 | 0.290 | 1.370 |
| | A4 | 0.800 | 0.020 | 0.080 | 2.160 | 1.640 |
| | A5 | 0.300 | 0.001 | 0.010 | 2.040 | 10.320 |
| | M1 | 0.110 | 0.020 | 0.020 | 1.560 | 5.600 |
| | M2 | 0.170 | 0.030 | 0.020 | 2.060 | 7.100 |
| | M3 | 0.160 | 0.030 | 0.030 | 2.180 | 7.400 |
| | M4 | 0.180 | 0.040 | 0.040 | 2.560 | 7.800 |

Table 1: Percentage of Helicos reads mapped to Roche 454 reads

# Ancient DNA study

## *Full genome analysis*

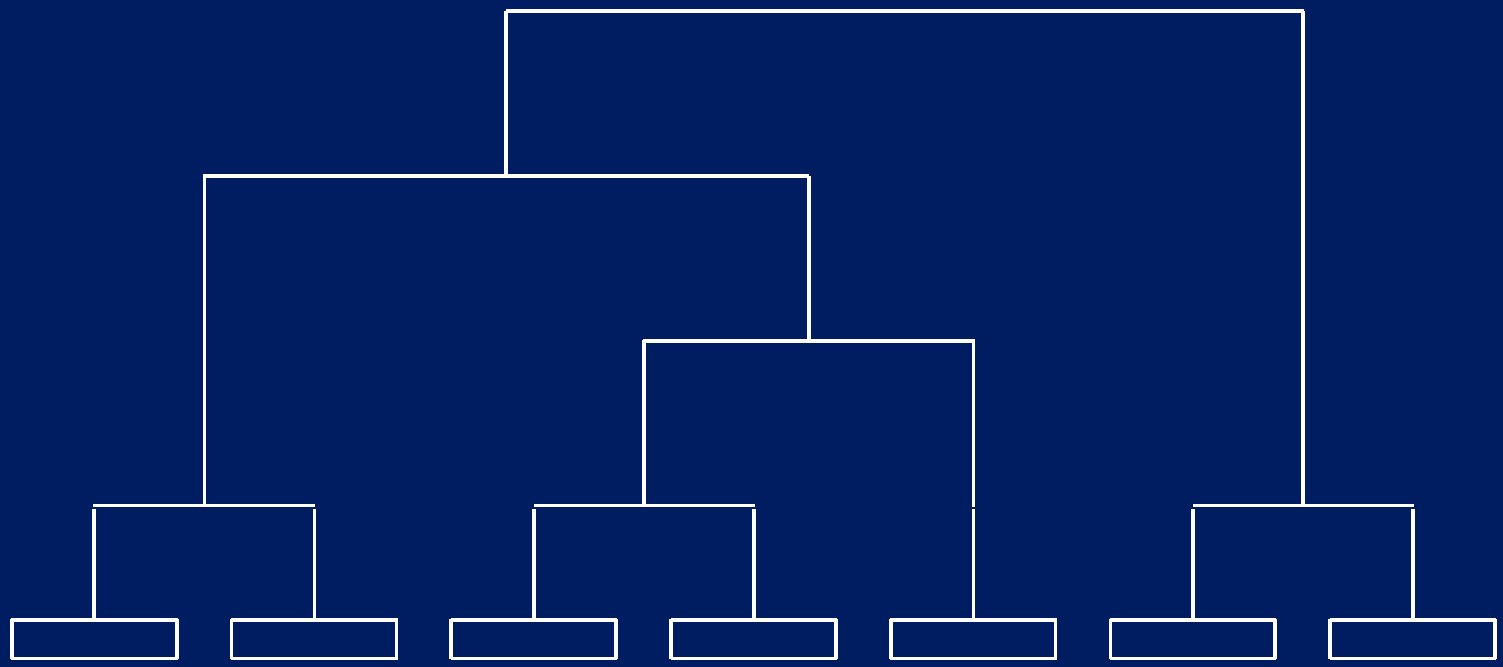

Figure 4: Horizontal coverage

# Ancient DNA study

## *Full genome analysis*

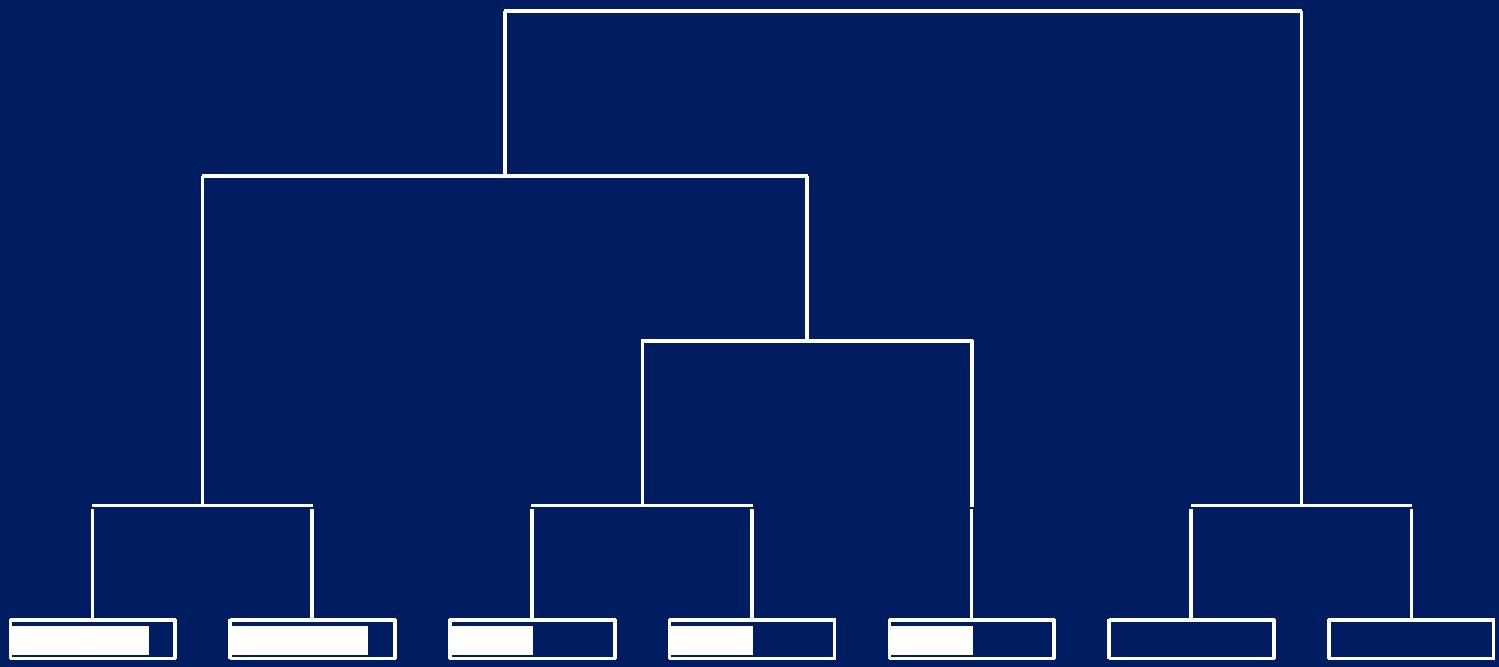

Figure 4: Horizontal coverage

# Ancient DNA study

## *Full genome analysis*

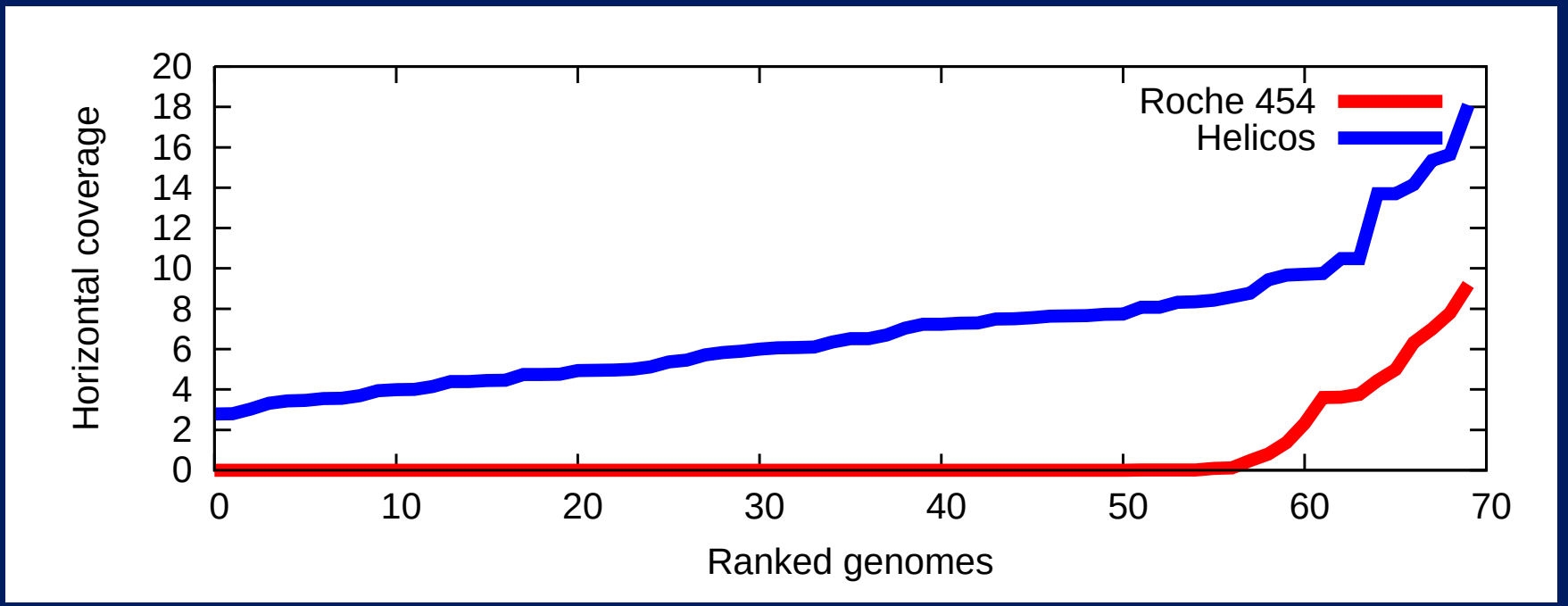


Figure 5: Horizontal coverage of ranked genomes

# Conclusions

***What we have learned so far***

Single molecule sequencing solves:

- PCR artefacts / biases.
- Fragmentation biases (Helicos).
    - Luckily we do not have this problem with forensic traces.

# Conclusions

***What we have learned so far***

Single molecule sequencing solves:

- PCR artefacts / biases.
- Fragmentation biases (Helicos).
    - Luckily we do not have this problem with forensic traces.

Has proven to be more sensitive for:

- Ancient DNA.
- Mixed samples.

# Further research

## *k-mer profiles*

Analysis of raw dataset:

- Make a profile of all $k$-mers of a certain length.
- Compare the profiles with a specialised distance measure.
- Possibly separate the dataset based on frequency distribution.

# Further research

## *k-mer profiles*

Analysis of raw dataset:

- Make a profile of all $k$-mers of a certain length.
- Compare the profiles with a specialised distance measure.
- Possibly separate the dataset based on frequency distribution.

Challenges:

- Different dataset sizes.
  - Dynamic $k$-mer profiles.
- Large choice of distance measures.

# Questions?

Acknowledgements:

Irina Pulyakhina
Eveline Altena
Michiel van Galen
Peter-Bram 't Hoen
Johan den Dunnen
Peter de Knijff