Ongoing human chromosome end extension revealed by analysis of BioNano and nanopore data

Haojing Shao
Chromosome end’s past, present and future

https://www.yourgenome.org/facts/what-is-a-telomere
Subtelomeric paralogy map

Chromosome end is repetitive.

Elena V. 2005
Average read length 280kb
> Maximum subtelomeric repetitive 156kb

https://bionanogenomics.com/technology/platform-technology/
Bionano: Two trios and two Chinese
Summary of chromosome end

Missing 11%
Extending 22%
Polymorphism
Heritable in trios 94%

Extension preferred
some sequences
Chromosome end polymorphism and extension

Population Sample

Label | Ref | Unknown | Unaligned | 19p | 16q | 1p | famA | famB
--- | --- | --- | --- | --- | --- | --- | --- | ---
Reference GRCh37
CEPH Father
CEPH Daughter
CEPH Mother
Ashkenazi Father
Ashkenazi Son
Ashkenazi Mother
Chinese Han
Chinese YH

Population Sample

Label | Ref | Unknown | Unaligned | 19p | 16q | 1p | famA | famB
--- | --- | --- | --- | --- | --- | --- | --- | ---
Reference GRCh37
CEPH Father
CEPH Daughter
CEPH Mother
Ashkenazi Father
Ashkenazi Son
Ashkenazi Mother
Chinese Han
Chinese YH

Chromosome 9 q arm terminal region (500kb)

Chromosome 15 q arm terminal region (199kb)
Validating and filling by nanopore

Label 9q Unknown Unaligned 1p

Reference GRCh37

Bionano assembly

Predicted sequence(9q+1p)

Nanopore read(184 kb)

140900000 141000000 141100000 141200000 141300000

chromosome 9 q arm terminal region (500kb)

Label 15q Unknown Unaligned 15qCN1 CN2 CN3 CN4 CN5

Reference GRCh37

Bionano assembly

Predicted sequence(15q+4*tip)

Nanopore read(52 kb)

102400000 102450000 102500000 102550000

chromosome 15 q arm terminal region (199kb)
15q nanopore reads show telomeric like tandem repeat

Is sequence directly added to proper capping telomere?
Some sequences are favored in the extensions.
Phylogenetic tree

Non-terminal: 14 myr
Terminal: 0.07~2.6 myr

They are new sequence after humans diverged from chimps
Extension’s impact on 15 species terminal

Terminal missing sequence=extension
Extension’s impact on 15 species terminal

dotplot and violin plot for 4 terminals in 15 species

- Terminals are share (ancient terminal)

H: Uniq + telo + duplication + telo

Horse, Dog: Unique

- Duplications are new.
Conclusion: an extension model

- Past
- Current
- Future
• Lachlan Coin’s group
• IMB Postgraduate Office
• Nectar service
• Public data
  The 1000 Genome Project
  Bionano data (Zook et al. 2016)
  Nanopore data (Jain et al. 2017)

• Australian Research Council Grant
  DP140103164
• Scholarship
  IPRS
  UQ Centenial
Question?