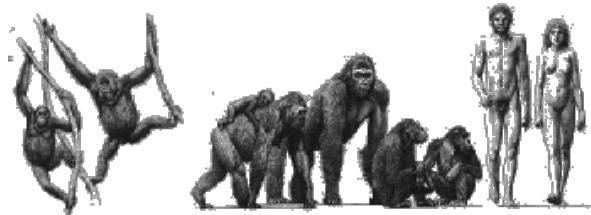


Great ape genomics, evolution of gene regulation and new applications of long read technologies.

## CSH, Programming for Biology 2018

Tomàs Marquès-Bonet

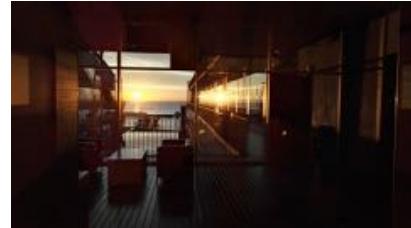
Institut de Biologia Evolutiva (UPF-CSIC, Barcelona)



## Humans as apes

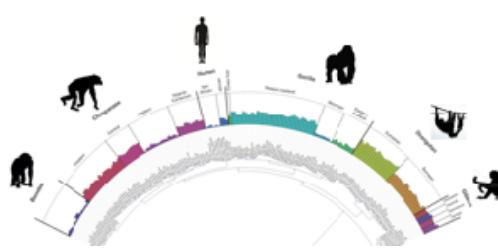


# Our group... PRBB, Barcelona

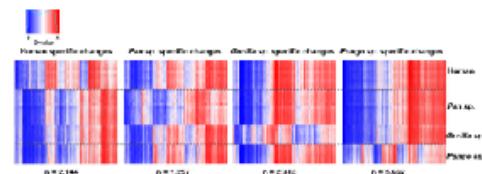


## What do we do?

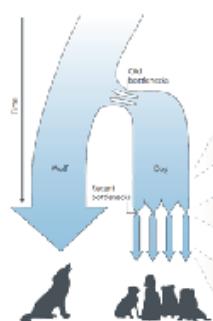
- Genomic variation and human evolution



- Transcriptome and SV and Epigenetics in humans and primates



- Canid evolution (CNVs)

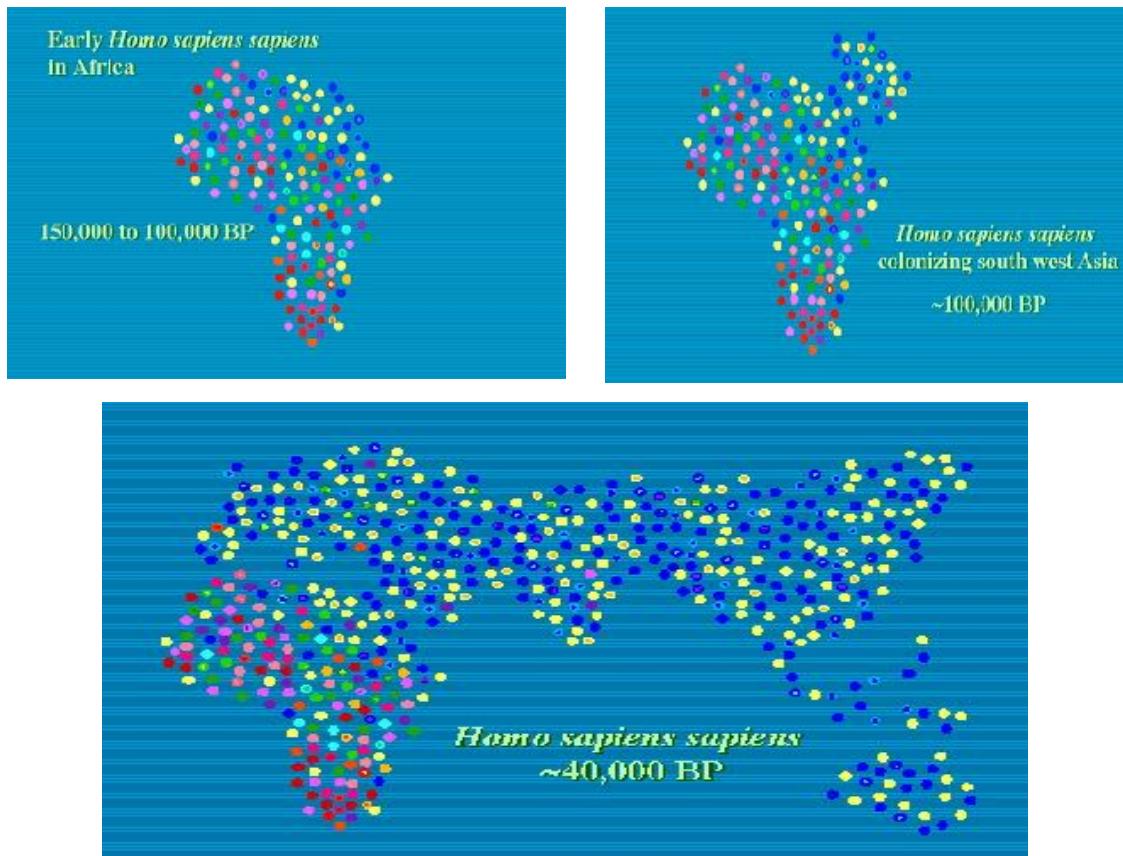


## Population genomics in apes

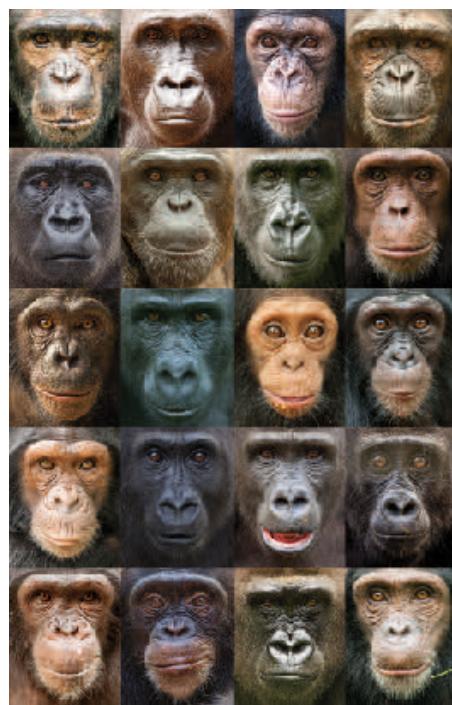


## Human evolutionary genetics

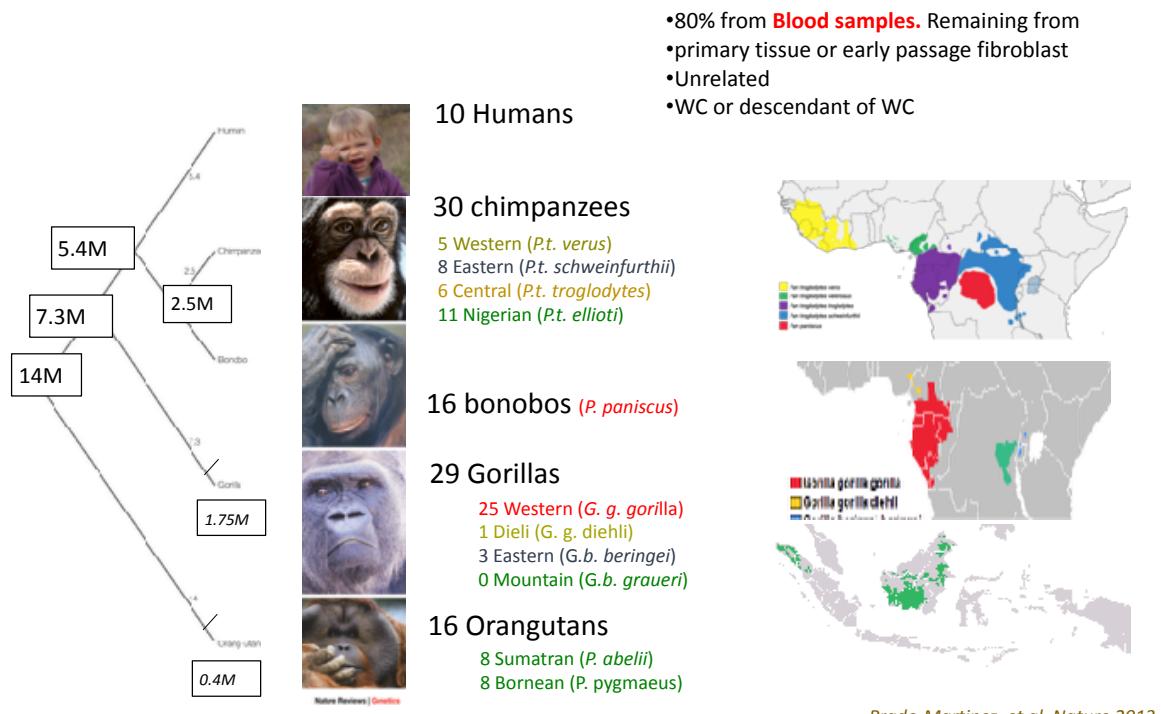




## What about apes?

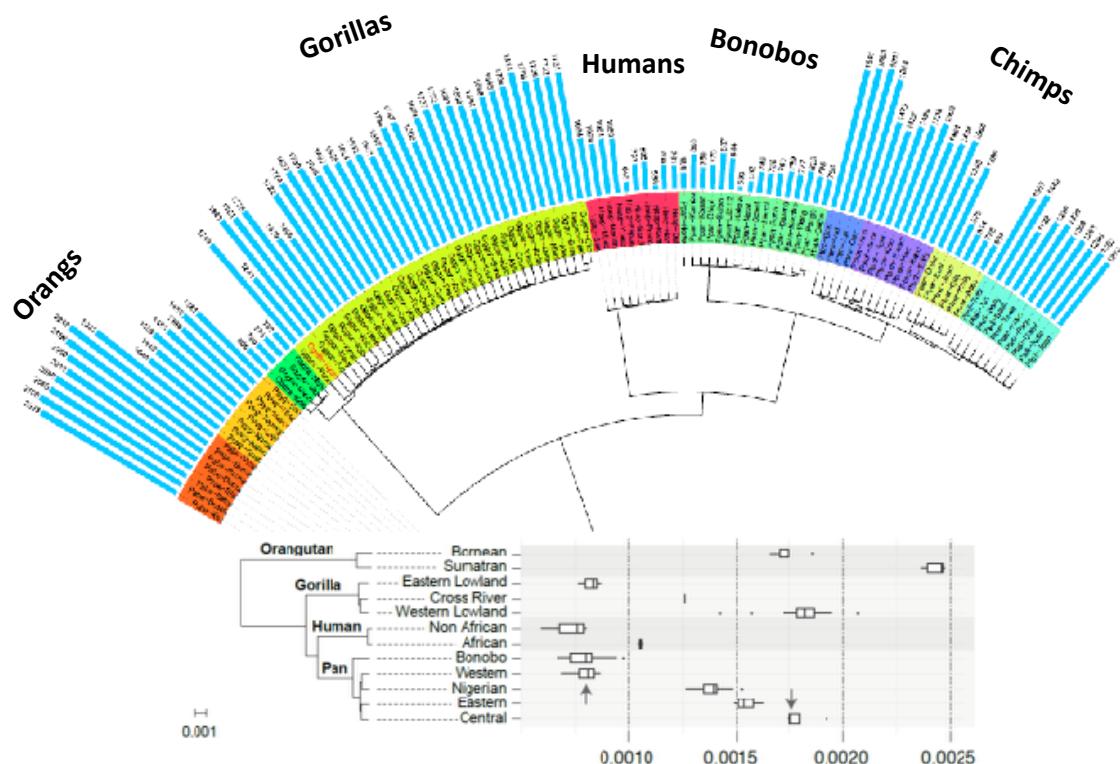


# Great Ape Genome diversity project

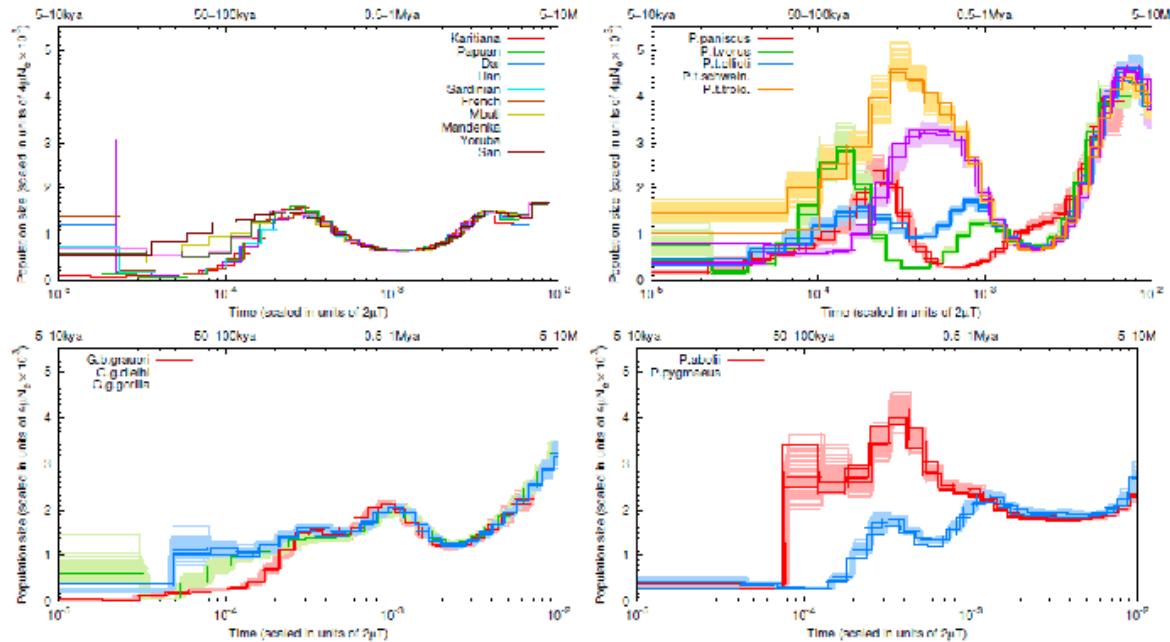


Prado-Martinez et al. *Nature* 2013

## Heterozygosity



# Demography (PSMC)

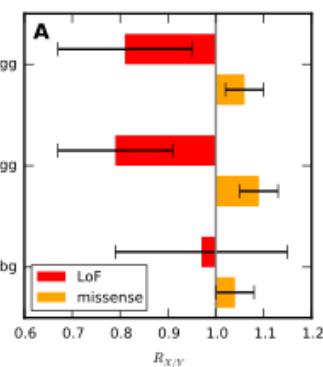
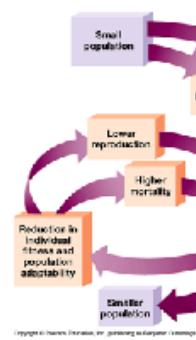
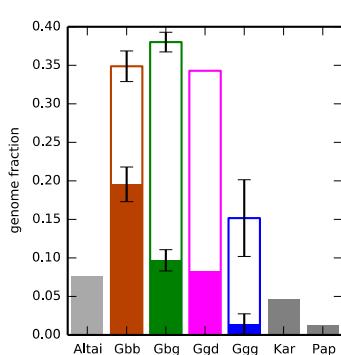


## Mountain gorilla genomes



PRIMATE GENOMICS

Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding

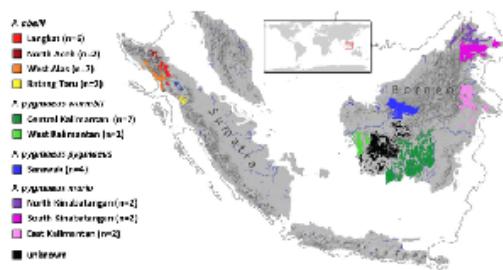


Xue, Prado-Martinez et al. Science 2015

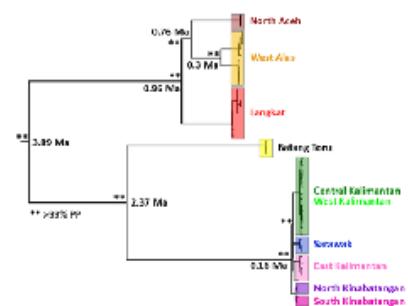
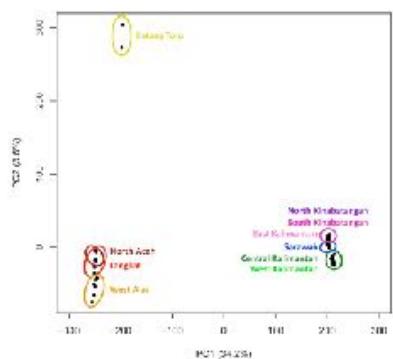
# Orangutans



*Science top 10 breakthrough of the year 2017*

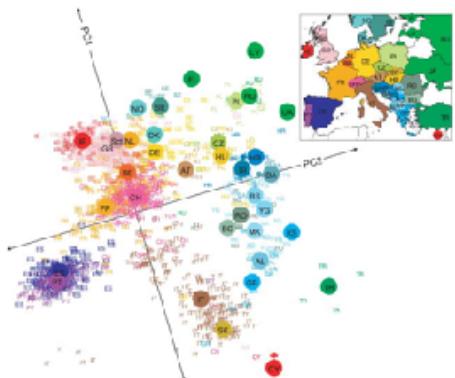


Total of 35 genomes covering the entire genus



Nater et al. Current Biology 2017

# Genes mirror geography

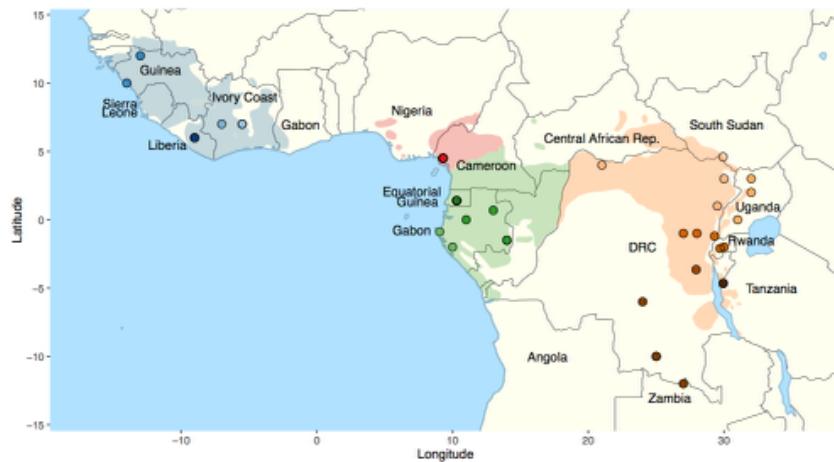


*Novembre et al. Nature 2008*  
*Lao et al. Current Biology 2008*

## Sample distribution and coverage

70 full genome of the *chimp and bonobo* lineage from **Blood samples**

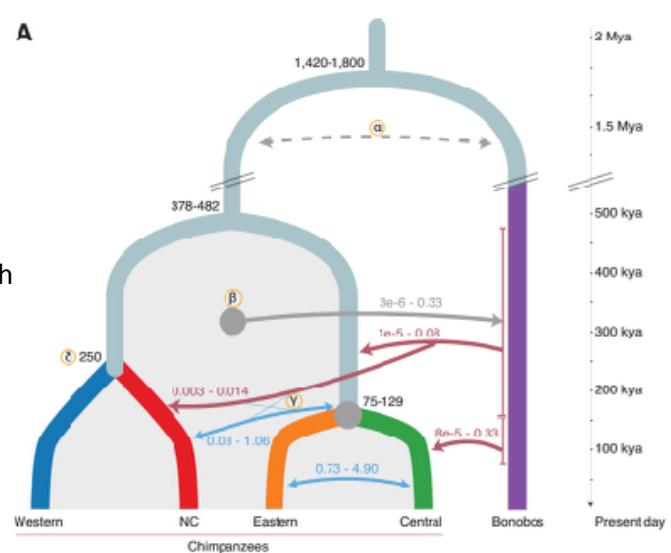
58 out of the 60 chimp genomes are georeferenced wild-born individuals spanning **9 countries**:



## Ancient gene flow between chimps and bonobos

### 5 lines of evidences:

- Excess of allele sharing (D stats)
- Clustered pattern of sites at high frequency in bonobos and shared with non-Western chimpanzee individuals.
- Differential haplotype divergences to bonobo
- Haplotype ages and variation

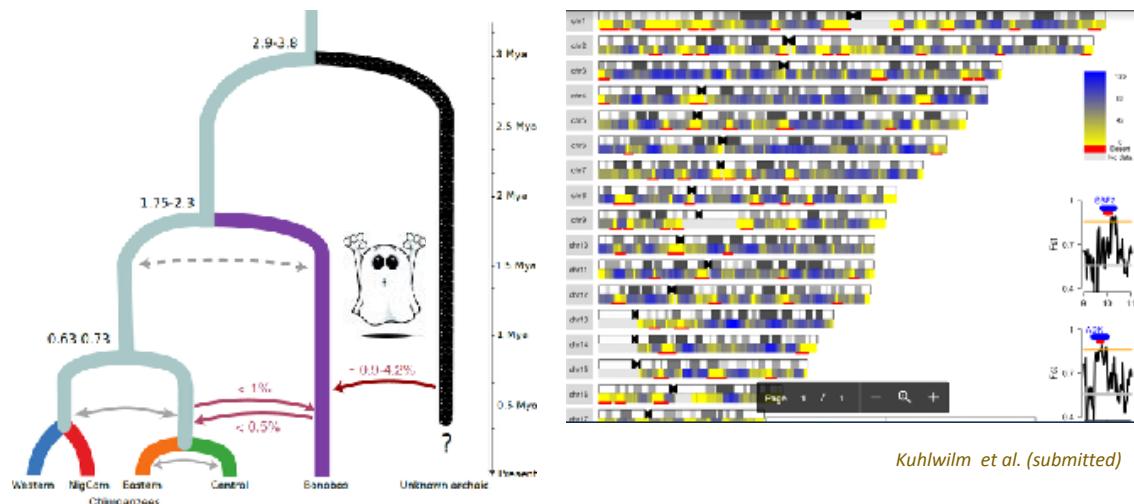


Possibly two gene flow events from bonobos into chimpanzees:

- 200-500kya
- <200kya

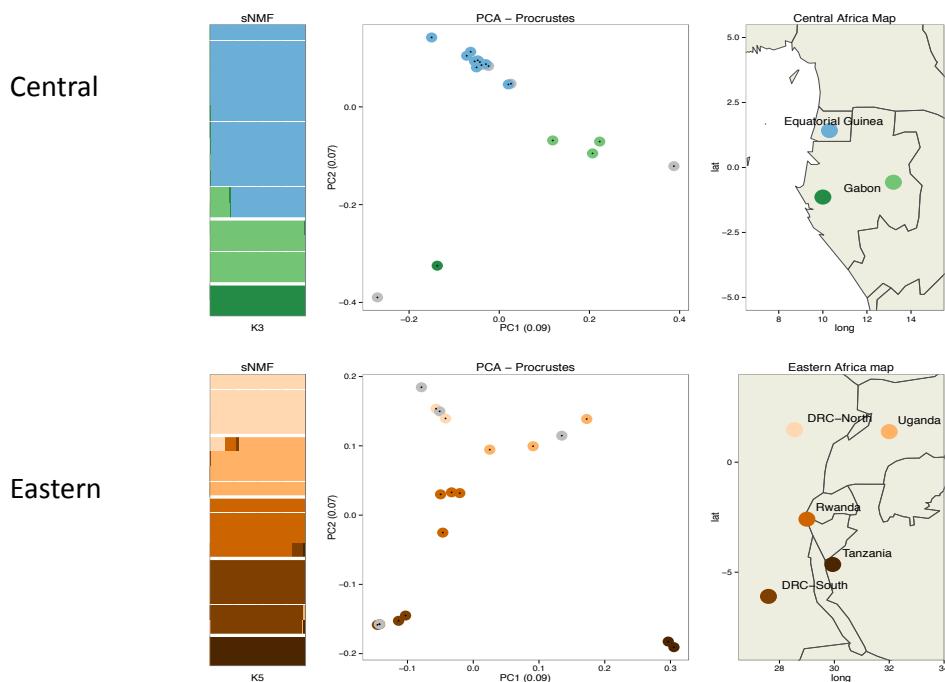
*deManuel et al. Science 2016*

## Excavating archaic ape genomes



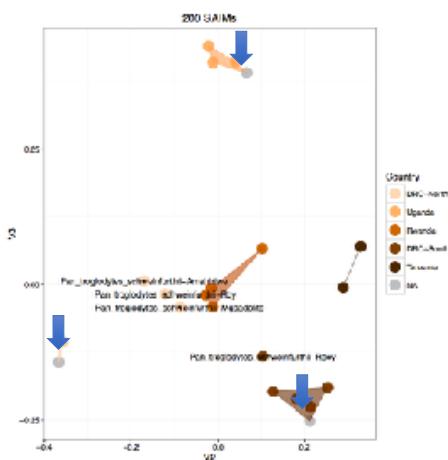
- No known fossils from Congo basin related to *Pan*
- Signature of introgression from “ghost” population found in bonobo genomes
- ~3% of genome from archaic lineage across 10 individuals

## Strong population stratification by country



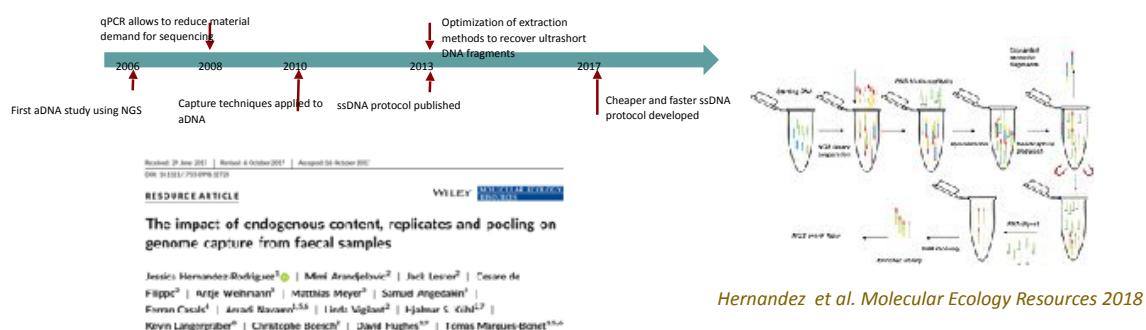
## How to validate information?

## Validation 1: Sequencing de novo samples with known origin

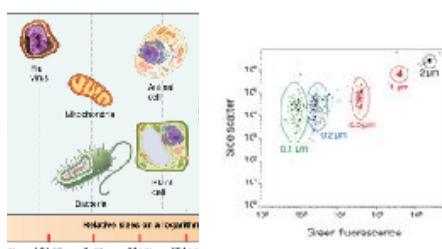


## Working with fecal samples

- ▶ Non-invasive samples are common (most of them geolocated) but little genome information can be retrieved (**degraded, usually < 2%**).



## A new method: CELL SORTING!

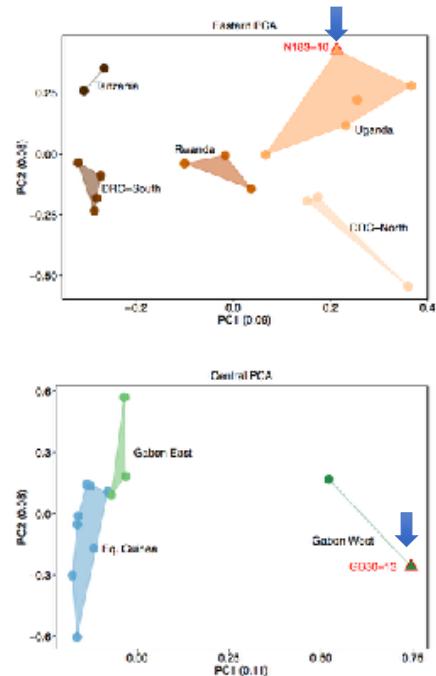


**Hundreds of cells**  
**DNA low input protocols**

Whole genome!

## How to validate information?

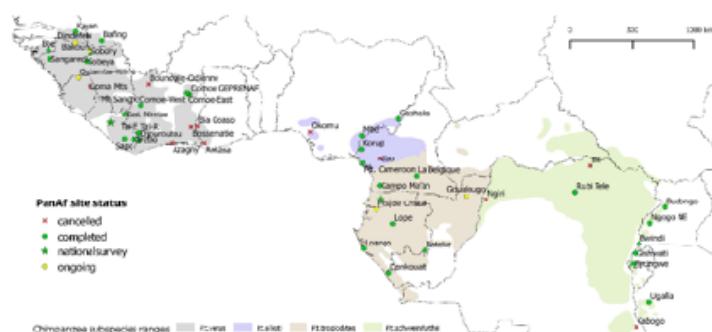
## Validation 2: Fecal samples



# A new framework for geolocalization (fecal samples from panAfrican)



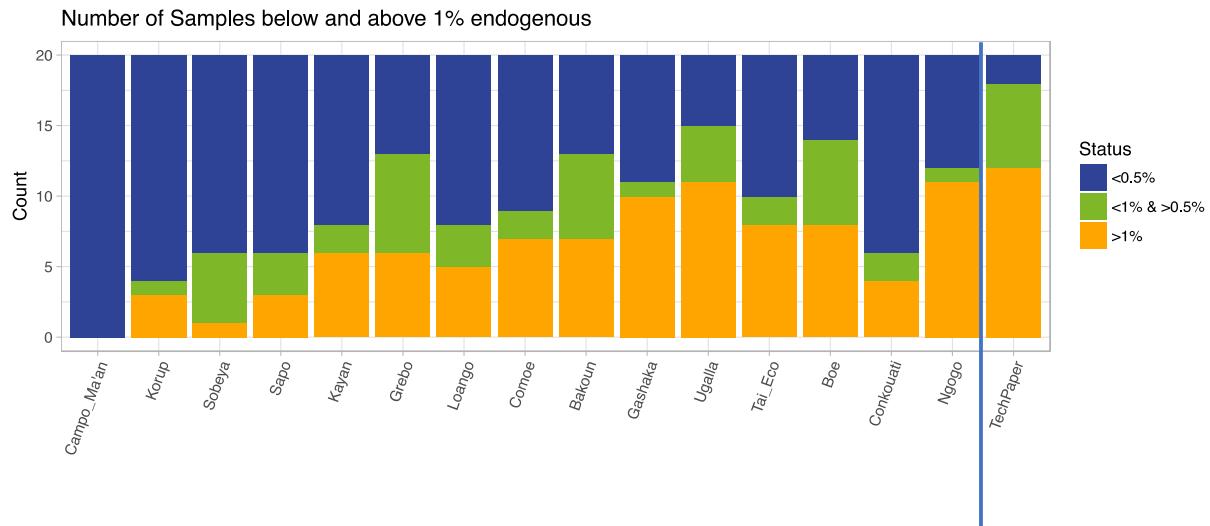
- 40 sites and 20 fecal samples/site



## Data and Methods

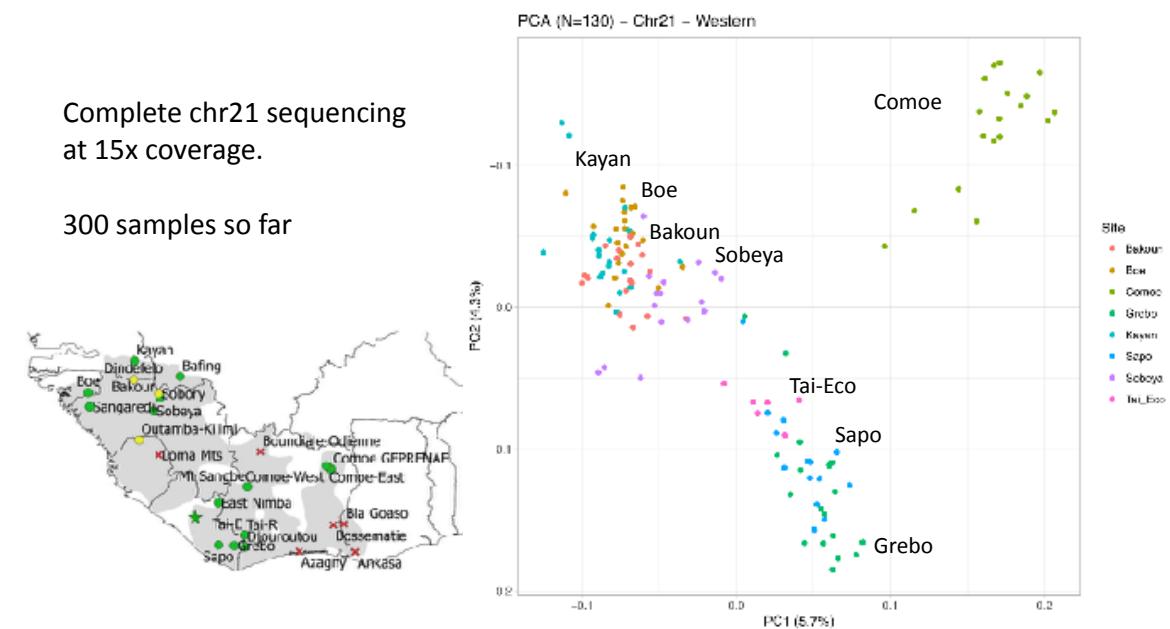
- 800 fecal DNA chimpanzee samples
  - Applied in house library preparation method ([Caroe et al](#)) and in-solution capture methods.
    - Agilent Exome V6
    - Whole chr21

## PanAf Dataset: Endogenous content



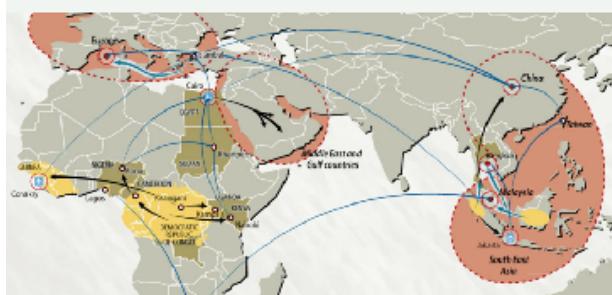
Most of the samples are < 0.5% endogenous!

## Within Subspecies: Western



# Application to conservation genetics

> Graphic > Stolen Ape - The... > Main International routes for the illegal trafficking of great apes



Number of great apes counted, presumed dead and lost, 2005–2011

	Confiscated/Confirmed	Presumed Dead	Undeclared Losses (k)	Total Loss
Gorillas	47	0	223	234
Bornean Orangutans	48	480	323	1,052
Gorilla	98	1,172	3,479	2,940
Orangutan	1,018	1,019	2,038	4,076
Total	1,363	5,171	11,109	22,218



## International Initiatives Battling the Illegal Trade in Great Apes

### Great Apes Survival Partnership (GASP)

GASP was established 2000 to battle powerful criminal organisations that threaten the survival of the world's three remaining great apes. The Coalition is made up of 17 UN member states, the World Bank, the European Commission, the International Fund for Animal Welfare, the World Wildlife Fund, the World Bank, the IUCN, and the World Conservation Strategy.

### International Union for the Conservation of Nature (IUCN)

IUCN is an international agreement between governments to protect nature and its resources. It is the world's largest environmental network, with over 1,000 member organisations in 160 countries and territories. IUCN's Species Survival Commission (SSC) is the world's oldest and largest global partnership dedicated to the long-term survival of wild species and their habitats.

### Great Apes and Forests Partnership (GAP)

GAP is a unique alliance launched to curb poaching, illegal trade and habitat destruction. It is not a fund to combat the conservation of apes. GAP is a political alliance of NGOs, foundations and governments that have joined together to combat the international illegal trade in great apes.

### MAPFIC

MAPFIC is a global wildlife crime fighting research network. MAPFIC is not a fund to combat the illegal trade in great apes. MAPFIC is a political alliance of NGOs, foundations and governments that have joined together to combat the international illegal trade in great apes.

### MAPFIC

MAPFIC is a global wildlife crime fighting research network. MAPFIC is not a fund to combat the illegal trade in great apes. MAPFIC is a political alliance of NGOs, foundations and governments that have joined together to combat the international illegal trade in great apes.

### GRASP (Global Rapid Assessment of Species Potential Crisis)

GRASP is a coalition of NGOs working to highlight a wide range of species that are at risk of extinction due to habitat loss and degradation, climate change, overexploitation, and other threats.

### UNESCO Man and Biosphere Programme

UNESCO Man and Biosphere Programme is a global network of protected areas that are managed to ensure the protection of biodiversity and the sustainable development of local communities.

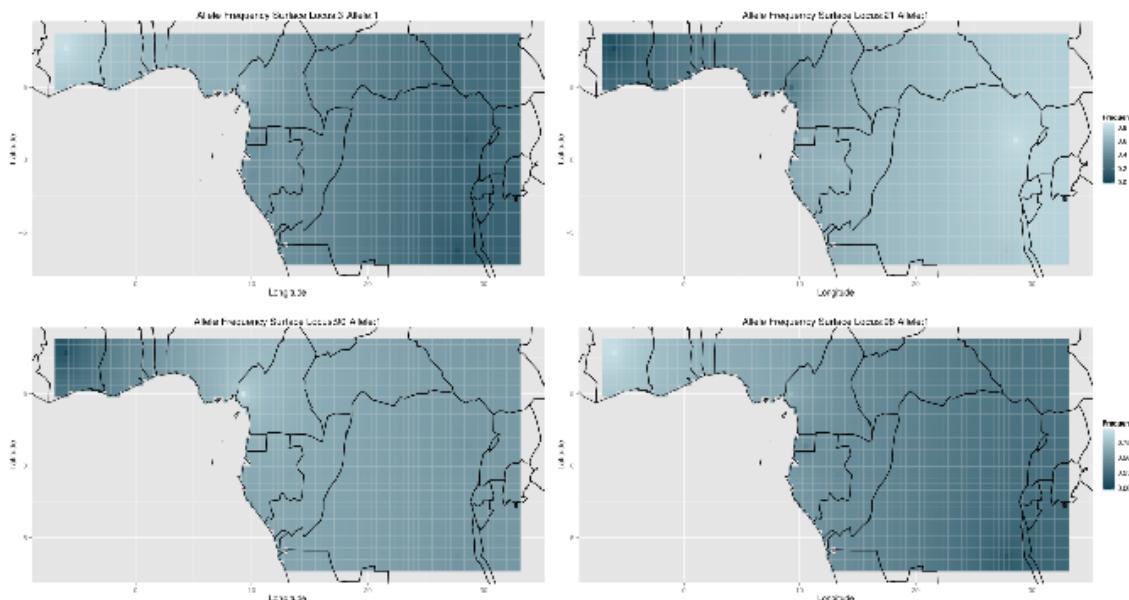
### Law Enforcement Agencies (LEAs)

LEAs are law enforcement agencies that are responsible for investigating and prosecuting crimes against the environment. LEAs include national parks, forest departments, and environmental protection agencies.



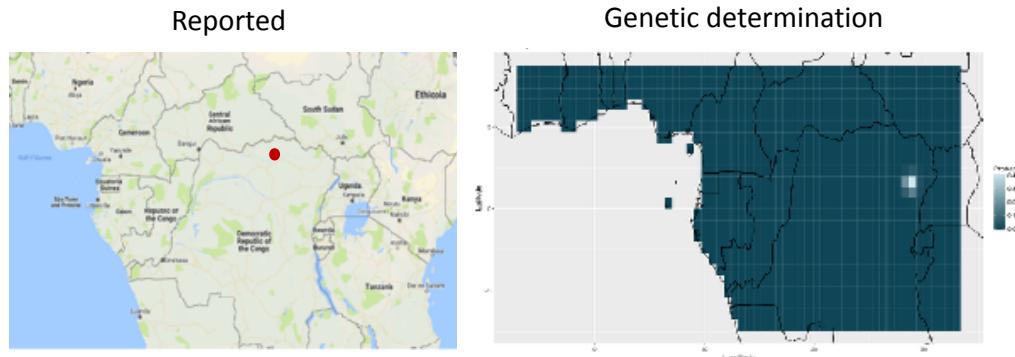
INTERPOL

# Application to conservation genetics



## Fighting illegal trafficking with genetics

- Bili was supposedly captured and smuggled from the Bili forest in the DRC.
  - Confirmed Eastern Chimpanzee.

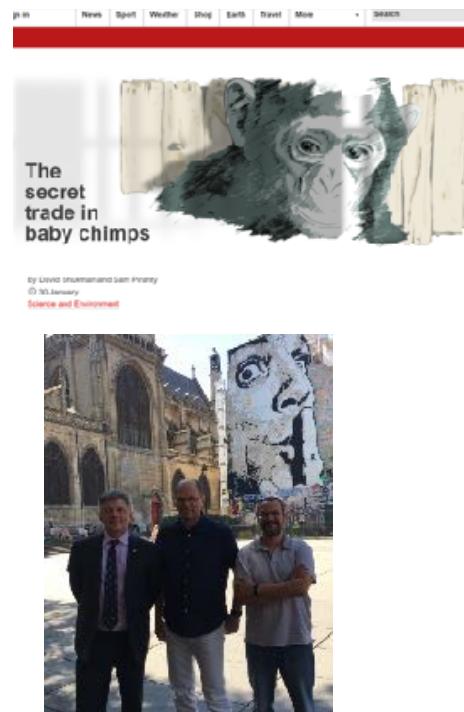


## Fighting illegal trafficking with genetics

- Jac was confiscated from laboratory and comes from Ruashi (DRC)

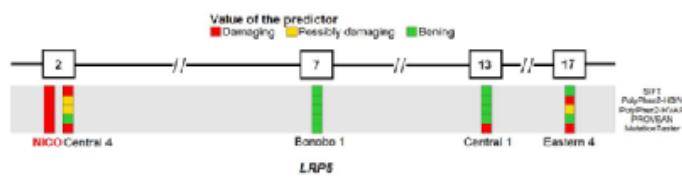


## UNESCO- Real example from this week



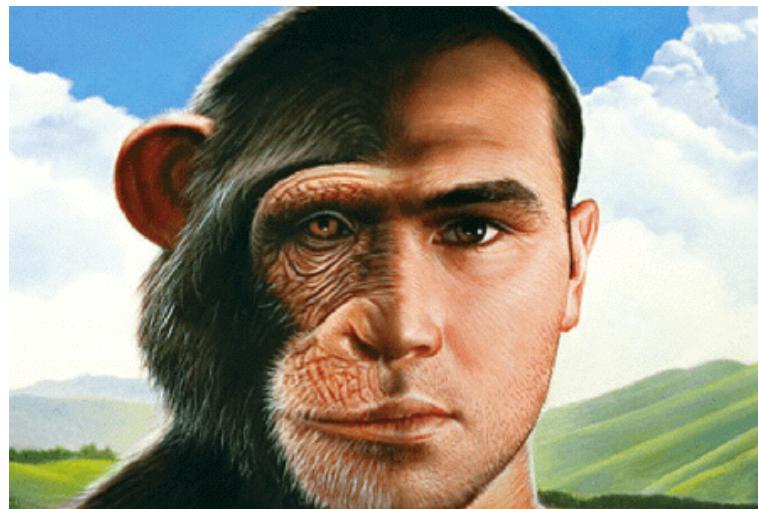
A good partnership,  
WAZA, Illumina and us

## Do all chimpanzees look the same?



Solis et al. Sci Reports 2017

# Evolution of gene regulation



MC King.....



31 April 1995, Volume 268, Number 5184 SCIENCE

## Evolution at Two Levels in Humans and Chimpanzees

These macromolecules are so alike that regulatory mutation may account for their biological differences.

Mary-Claire King and S. C. Williams

Editorial concerning the molecular basis of evolution in humans and chimpanzees. The authors find evolutionary changes in human DNA may be more subtle than expected. They argue that the sequence of genes that are present change little, while the way they are controlled by regulatory molecules account for the major biological differences between humans and chimpanzees.

Society of Human and Chimpanzee Genes

To compare human and chimpanzee genes, one compares other homologous proteins or nucleic acids. All the proteins

## Genome variation and gene expression

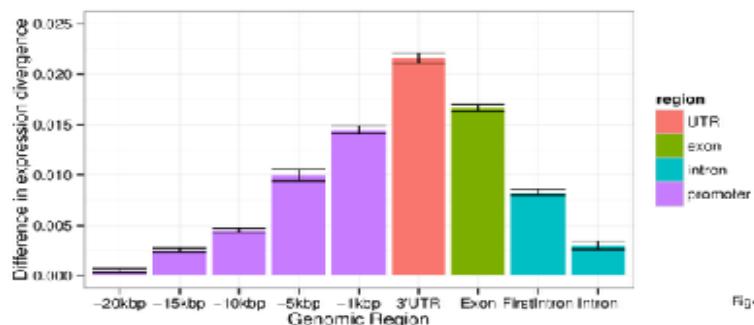
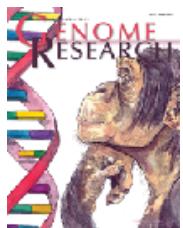
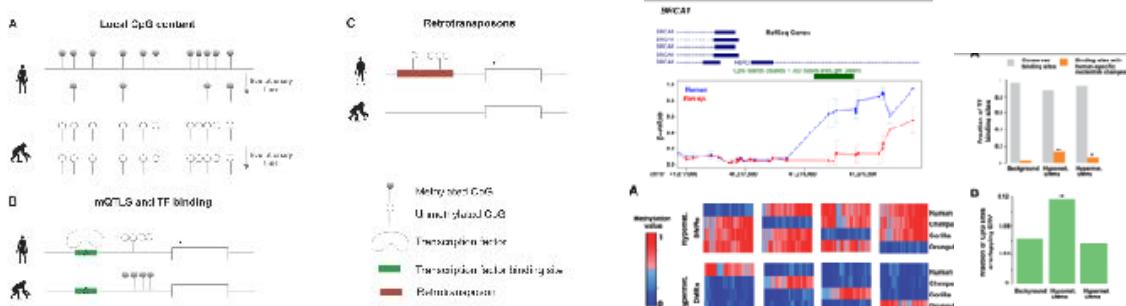


Fig 4

Expression divergence: 3' UTR >> repeats in exons >> repeats in promoter >> repeats in 1<sup>st</sup> intron >> repeats in any intron

Bilkin, Carvalho et al. *Genome Research* 2015

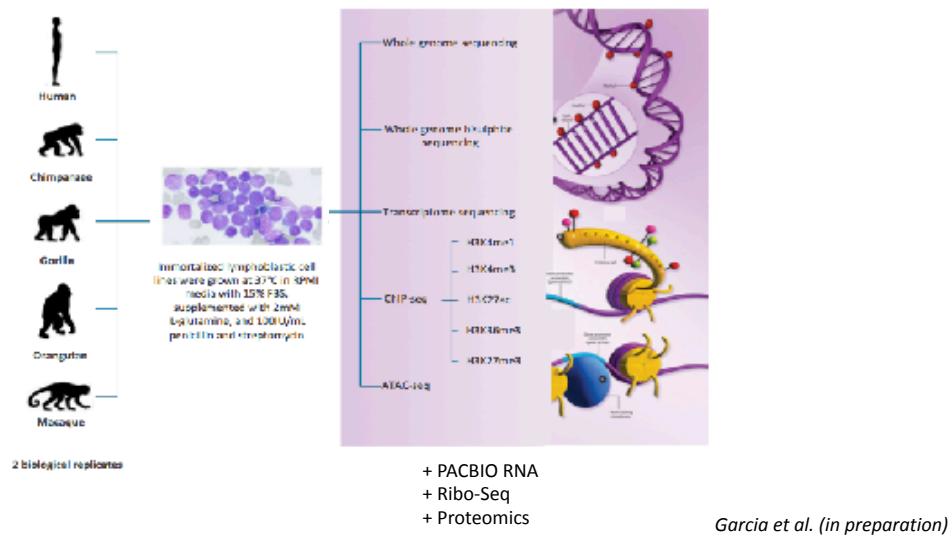
## Comparative Epigenomics



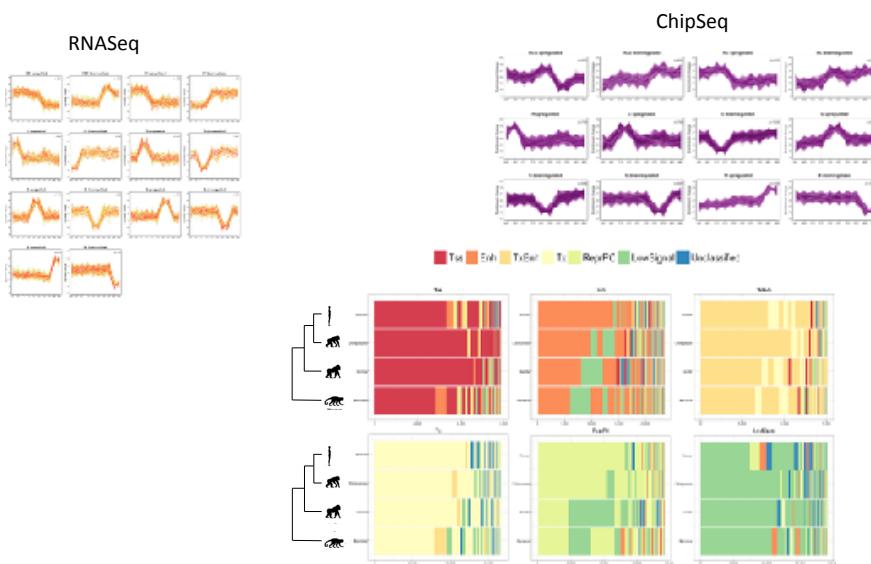
Hernando et al. *Plos genetics* 2015

Hernando et al. *Plos genetics* 2013

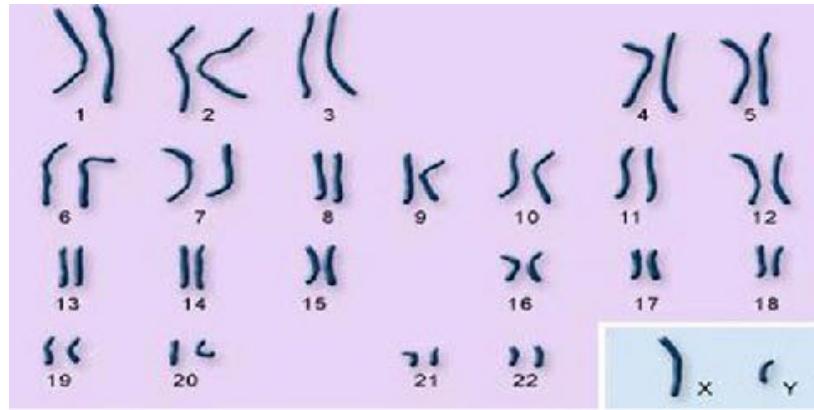
## Ok, but then what? Cell lines....



## Profiles for human derived molecular phenotypes



# The complicated Y chr....



Y chromosomes are notoriously hard to assemble!

## articles

### The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes

Helen Skaletsky<sup>1</sup>, Tomoko Kuroda-Kawaguchi<sup>2</sup>, Patrick J. Minx<sup>1</sup>, Holland S. Cerdan<sup>2</sup>, LaDesha Hillier<sup>1</sup>, Laura G. Brown<sup>1</sup>, Sjoerd Repping<sup>2</sup>, Tatyana Pyntikova<sup>1</sup>, Jelkar Ali<sup>1</sup>, Tamara L. Bleit<sup>1</sup>, Asif Chinnaiyan<sup>1</sup>, Lucinda Fulton<sup>1</sup>, Robert Fulton<sup>1</sup>, Tina Graves<sup>1</sup>, Shun-Fang Hsu<sup>1</sup>, John McPherson<sup>1</sup>, Tracie Miner<sup>1</sup>, William Nash<sup>1</sup>, Christine Nguyen<sup>1</sup>, Keisi Scott<sup>1</sup>, Brian Schultz<sup>1</sup>, Cindy Strong<sup>1</sup>, Aye Tin-Welham<sup>1</sup>, Shu & David C. Page<sup>1</sup>

Vol 463 | 28 January 2010 doi:10.1038/nature08700

## LETTERS

### Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content

Jennifer F. Hughes<sup>1</sup>, Helen Skaletsky<sup>1</sup>, Tatyana Pyntikova<sup>1</sup>, Tina A. Graves<sup>2</sup>, Sesilia K. M. van Daalen<sup>3</sup>, Sean D. McGrath<sup>4</sup>, Devir P. Locke<sup>5</sup>, Cynthia Friedman<sup>6</sup>, Barbara J. Trask<sup>4</sup>, Ir<sup>2</sup>, Sjoerd Repping<sup>2</sup>, Steve Rezen<sup>1</sup>, Richard K. Wilson<sup>2</sup> & David C. Page<sup>1</sup>

doi:10.1038/nature10843

## LETTER

### Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes

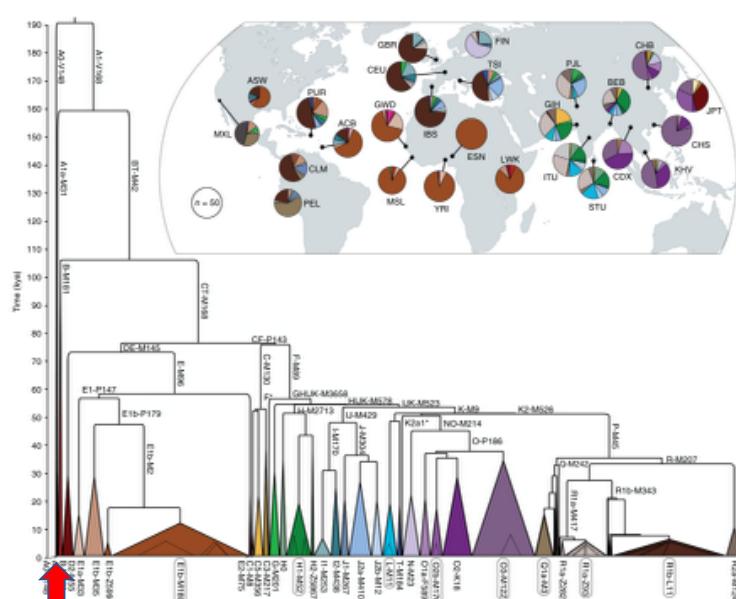
Jennifer F. Hughes<sup>1</sup>, Helen Skaletsky<sup>1</sup>, Laura G. Brown<sup>1</sup>, Tatyana Pyntikova<sup>1</sup>, Tina Graves<sup>2</sup>, Robert S. Fulton<sup>2</sup>, Shannon Dugay<sup>3</sup>, Yan Ding<sup>3</sup>, Christian J. Bailey<sup>3</sup>, Odile Krenitzin<sup>2</sup>, Qiaoyun Wang<sup>3</sup>, Hua Shen<sup>3</sup>, Michael Holder<sup>1</sup>, Donna Villasana<sup>3</sup>, Lynne V. Nazareth<sup>1</sup>, Andrew Cox<sup>1</sup>, Laura Courtney<sup>2</sup>, Joelle Vekter<sup>1</sup>, Holland Kotkiewicz<sup>2</sup>, Ting-Jan Cho<sup>1</sup>, Natalia Kourtseva<sup>1</sup>, Steve Rozen<sup>1</sup>, Donna M. Muzny<sup>1</sup>, Wesley C. Warren<sup>1</sup>, Richard K. Wilson<sup>2</sup> & David C. Page<sup>1</sup>

**There are some shortcuts...**

# A time- and cost-effective strategy to sequence mammalian Y Chromosomes: an application to the de novo assembly of gorilla Y

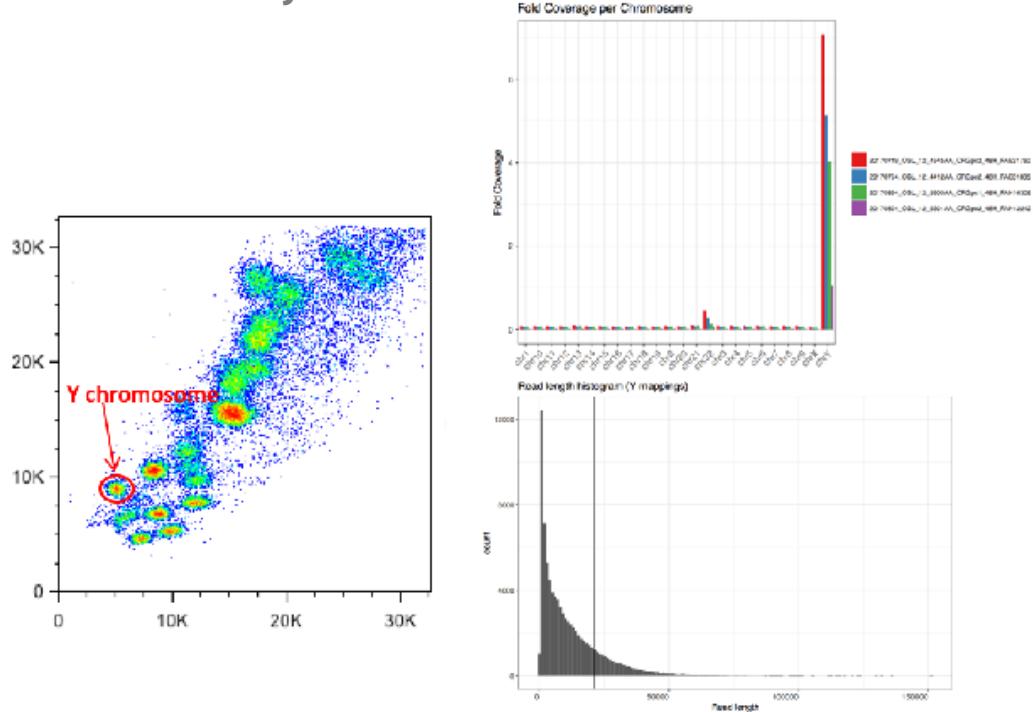
Marta Tomaszkiewicz,<sup>1,9</sup> Samarth Rangavittal,<sup>1,9</sup> Monika Cechova,<sup>1,9</sup> Rebeca Campos Sanchez,<sup>2</sup> Howard W. Fescemyer,<sup>1</sup> Robert Harris,<sup>1</sup> Danling Ye,<sup>1</sup> Patricia C.M. O'Brien,<sup>3</sup> Rayan Chikhi,<sup>4,5,6</sup> Oliver A. Ryder,<sup>7</sup> Malcolm A. Ferguson-Smith,<sup>3</sup> Paul Medvedev,<sup>5,6,8</sup> and Kateryna D. Makova<sup>1</sup>

## A proof of concept...



**Figure 2** Y-chromosome phylogeny and haplogroup distribution. Branch lengths are drawn proportional to the estimated times between successive splits, with the most ancient division occurring ~190 ka. Colored triangles represent the major clades, and the width of each base is proportional to one less than the corresponding sample size. We modeled expansions within eight of the major haplogroups (circled) (Fig. 4); dotted triangles represent the ages and sample sizes of the expanding lineages. Inset, world map indicating, for each of the 26 populations, the geographic source, sample size, and haplogroup distribution.

## Enrichment by flow sorting and some magic...



## Comparison of contiguity to gorilla Y

- Run Non-filtered data through Canu
  - ~25X of self-corrected reads as input for assembly
- Remap Nanopore reads, correct assembly with Nanopolish
- Use Illumina data to correct remaining errors where possible with Pilon

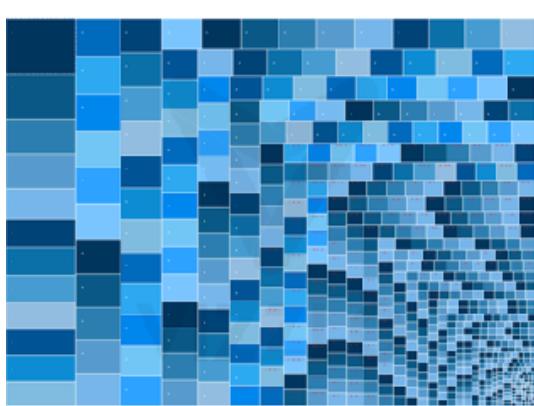
Results: 35 Contigs

N50 : 1,463,983

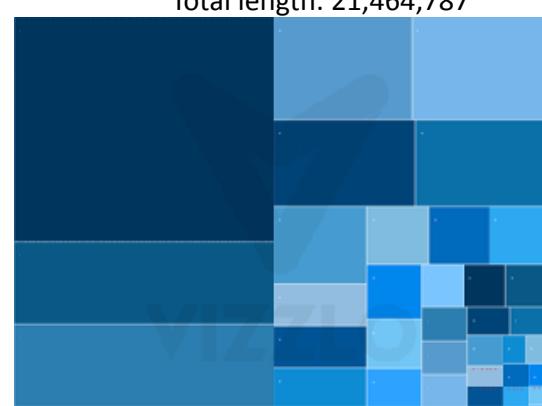
N90: 278,531

Total length: 21,464,787

*Shortcut: Gorilla Y chr*

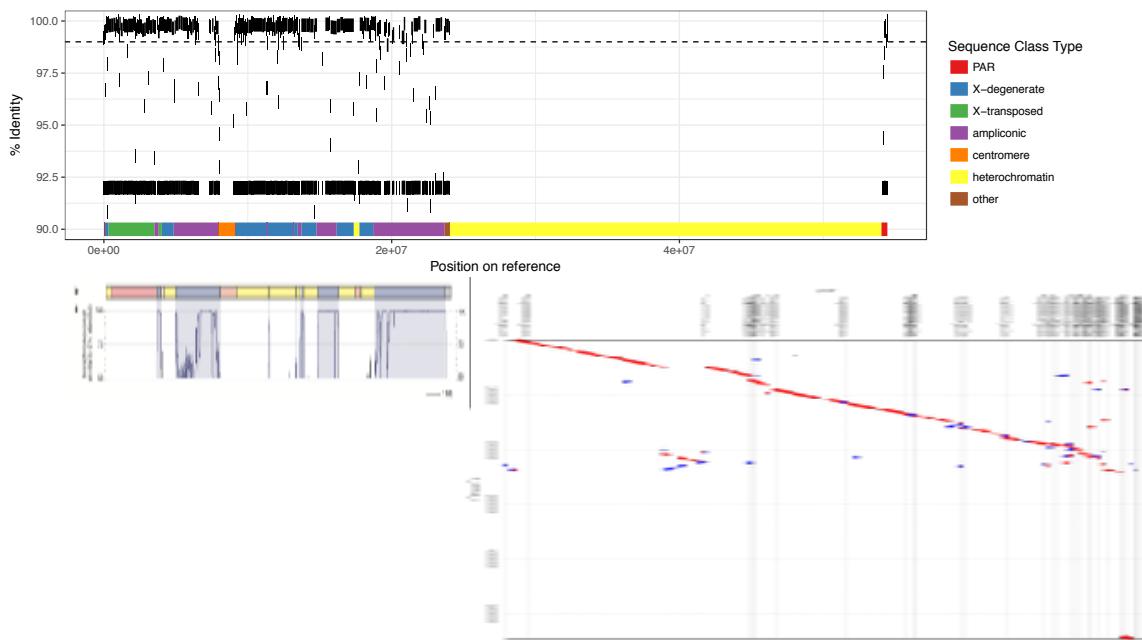


Contig N50: 17,836 bp

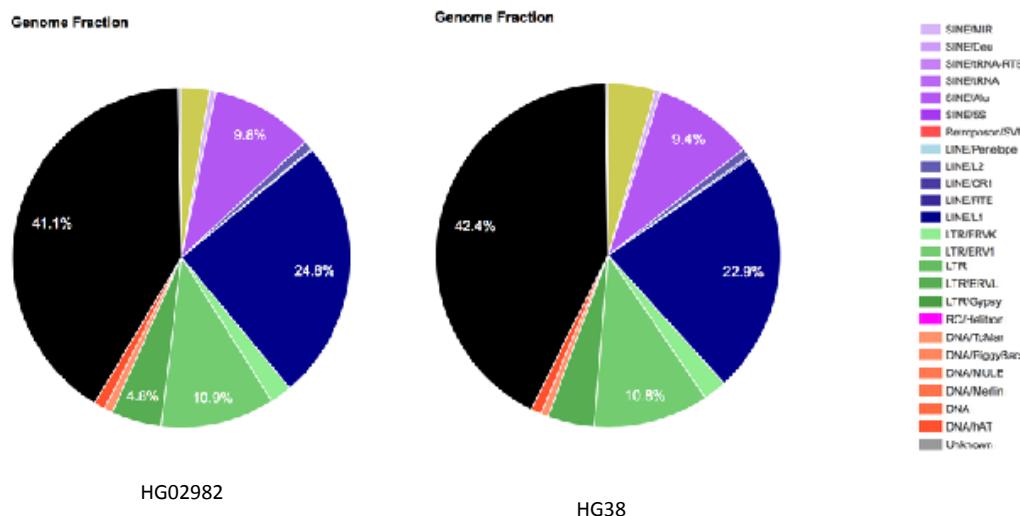


Contig N50: 1,463,983

## Limits of our Y chr: where do we fail?



## Landscape of interspersed repeats



## Some messages...

- Population genomics in apes and their applications to conservation
- Evolution of gene regulation and a new panel with excellent genomic characterizations
- Novel method to assemble Y chromosomes with Nanopore

## Acknowledgments

Copenhagen Zoo,

Christina Hvilsom

University of Copenhagen

Hans. R. Siegismund  
Peter Frandsen

University of Zurich

Michael Krutzen

Aarhus University

Mikkel Schierup

### Institut Biologia Evolutiva UPF/CSIC

Lukas Kuderna      Jessica Hernandez  
Claudia Fontserè      Irene Lobón  
Marc de Manuel      Aitor Serres  
Martin Kuhlwilm      Sojung Han  
Esther Lizano      Luis Fernandez  
Raquel Garcia      David deJuan  
David Hughes      Manuel Solis  
                            Laure Batlle  
                            Paula Esteller

Sanger, Wellcome Trust

Chris Tyler-Smith  
Yali Xue

Max Planck

Sergi Castellano  
Aida Andres  
Mimi Arandjelovic  
Linda Vigilant  
Hjalmar Kühl  
Christophe Burke

