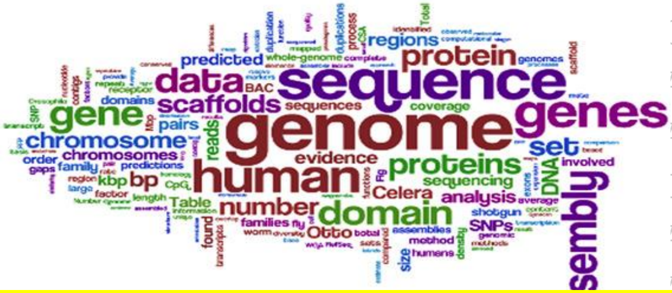
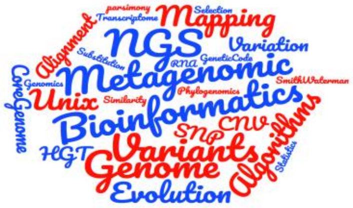
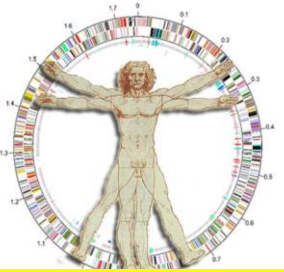






La Saga des Génomes

ou

L'Essor fulgurant de la Génomique et de la Bioinformatique

Fredj Tekaia, Institut Pasteur Paris

tekaia@pasteur.fr

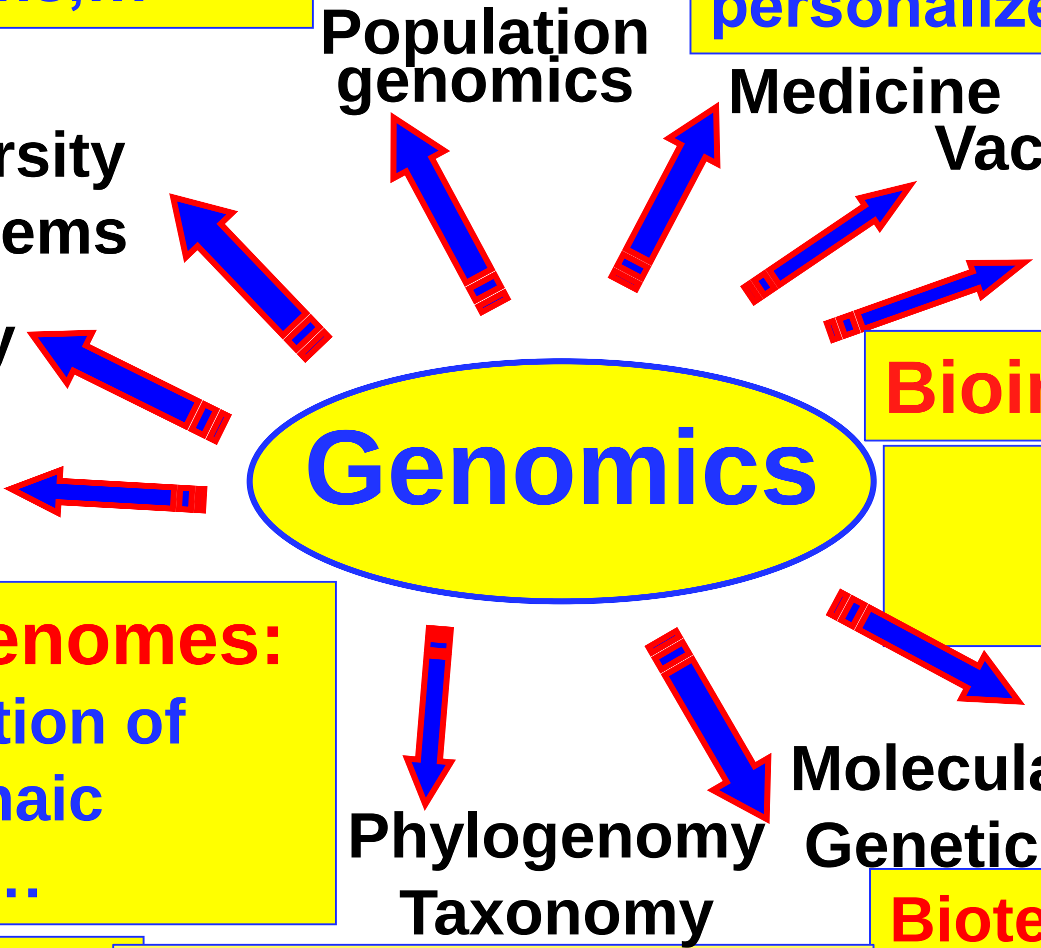
https://webext.pasteur.fr/tekaia/BGA_courses.html

Académie Tunisienne des Sciences des Lettres et des Arts, *Beit al-Hikma*

Tunis. 19-11-2020.

Environment:
Metagenomics, soil, water, oceans,…

Human health:
Cancer, Microbiome, personalized medicine,….



Genomics

Population genomics

Medicine

Vaccines

Pharmacology

Biodiversity
Ecosystems

Ecology

Evolution

Bioinformatics

Scientific Research

Microbiology

Molecular Virology
Genetics

Phylogenomy
Taxonomy

Ancient genomes:
Reconstruction of extinct, archaic organisms,…

Agronomy:
Animal, plant selection,…

New technologies development:
Sequencing, imaging,..

Biotechnology:
Genome editing, engineering, Synthetic Biology,…

Plan

I

- **Statistics on Genome Sequencing follow up;**
- **Bioinformatics in the genome era;**
- **Examples of large projects of Genome sequencing;**

II

- **Examples of genome structural variants studies in human;**
- **Examples of Microbiome studies in human;**

III

- **Examples of Ancient genome reconstruction of archaic human and organisms;**

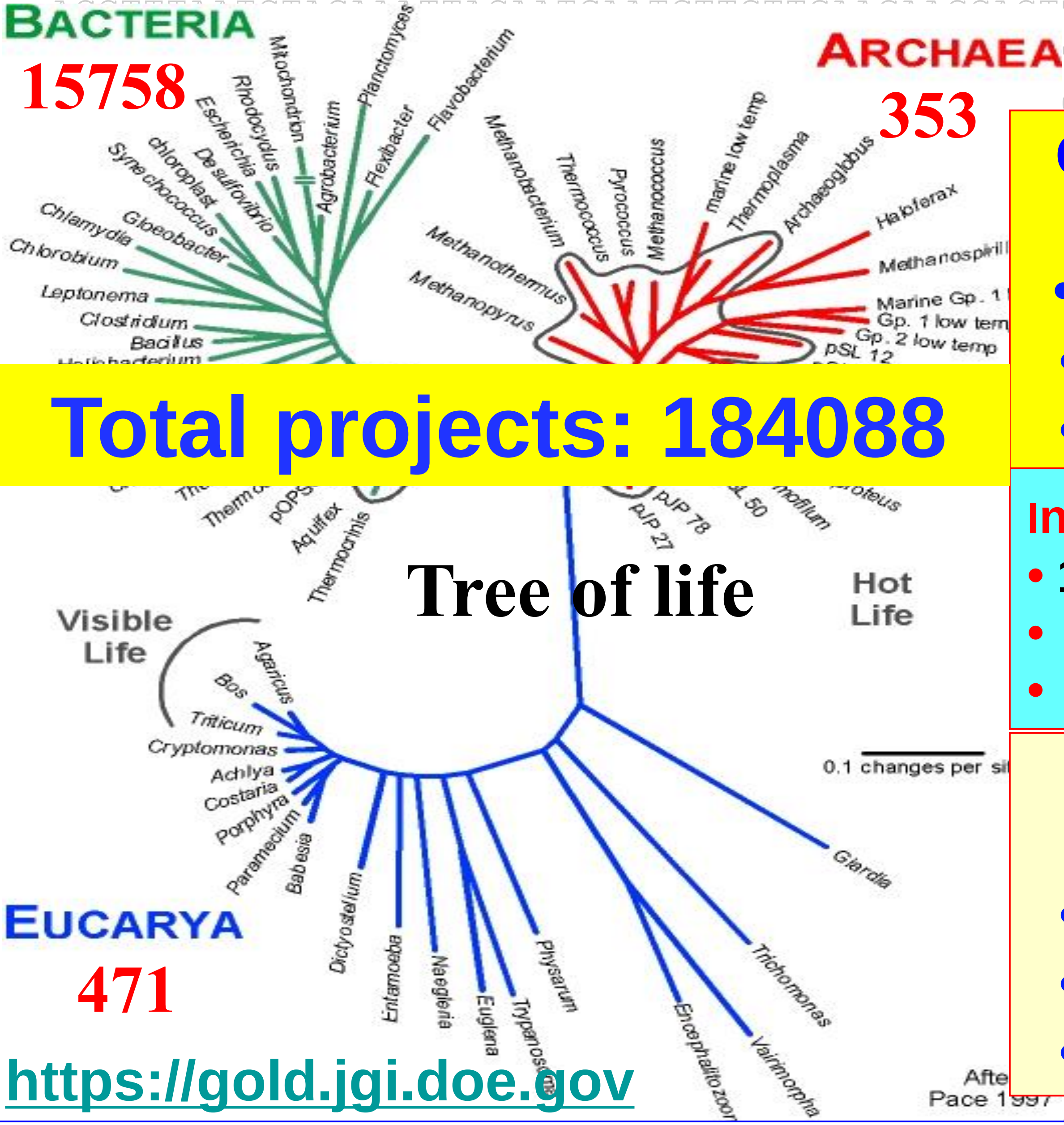
IV

- **Genome Editing: Designing minimal genomes;**

- **Conclusions – Perspectives (Propositions).**

Note: For more details, References are included in each slide

TATTTGATTTTTTGCTTAATTTTTATGTAATTACTTCCCCACAATTTTTGCCCACATATTGTGGATAAAT



Tree of life

15/10/2020

Complete sequenced genomes: 16582
- **15758 Bacteria**
- **353 Archaea**
- **471 Eukaryotes**

Total projects: 184088

Incomplete genomes: 23228
- **14584 Bacteria**
- **218 Archaea**
- **8166 Eukaryotes**

Permanent Draft genomes: 144278
- **138651 Bacteria**
- **1073 Archaea**
- **4554 Eukaryotes**

https://gold.jgi.doe.gov

ACAGACACTAACAATAACAAAGAAGA

Viruses: • Completed: 3502 • Permanent draft: 5083

```
>NC_010163.1 Acholeplasma laidlawii PG-8A, complete genome
TATTTGATTTTTTGCTTAATTTTTATGTAATTACTTCCCCACAATTTTTGCCCACATATTGTGGATAAAT
TTTCCACATTTTATTCACAATGTTGATAAGTTGTGTAAGTCGACTTGTTGGCTTTATAAAGCAAATGACA
```

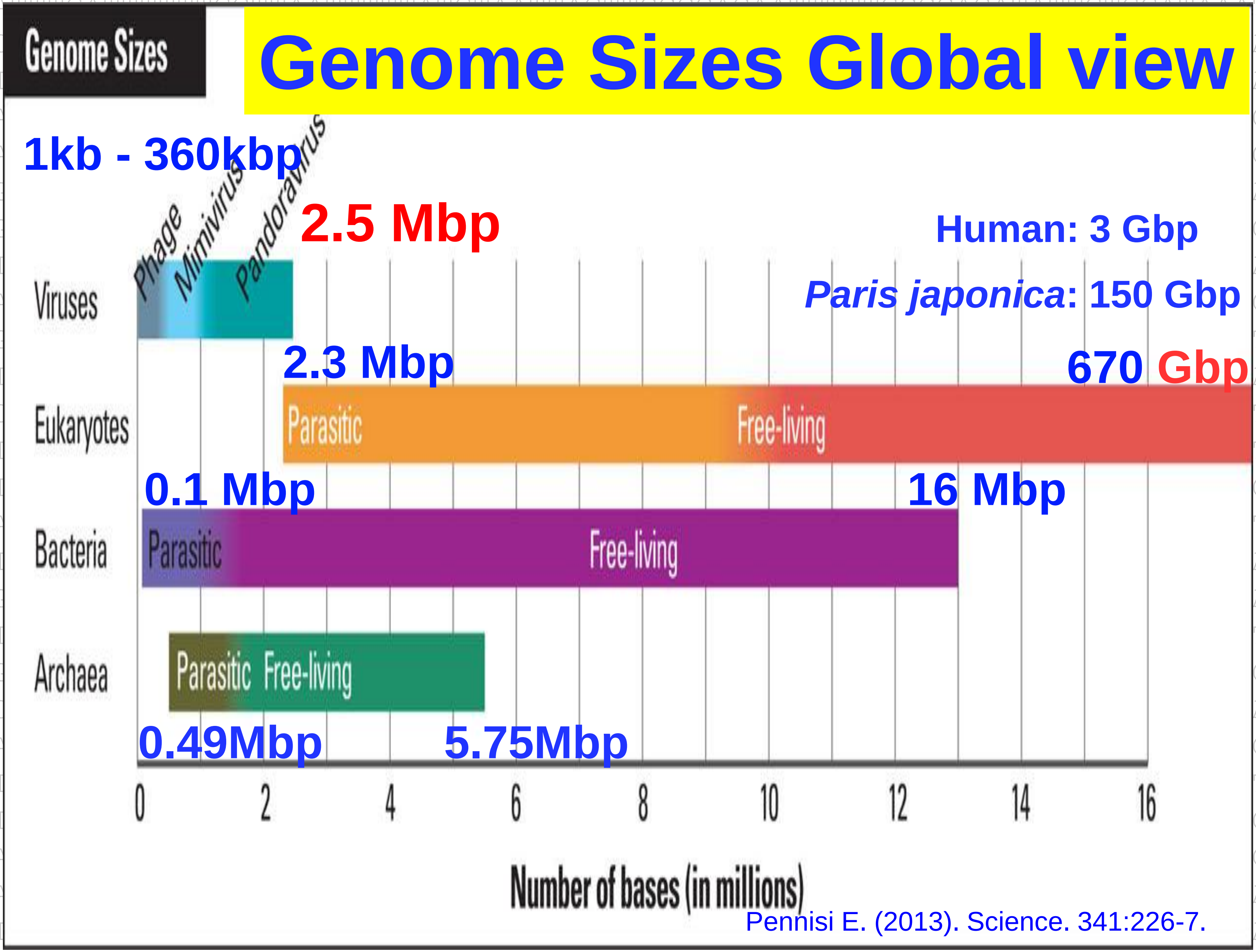
The Genome is an organism's complete set of DNA

```
GGAGACTTACAACGAGCTATTTCTACCAGTGACTTCTACTTTTAAAGATCAAAACGGATTACTTACAATG
GTTGT                                                           TTGCTA
CAAAA                                                           AGTAGC
```

Succession of A,T,C or G bases

```
TGATC                                                           TTTGTT
GTTGGAAAATCTAACATGTTTGCTTTTCGTATGGCGATGAAGGTTGCTGATCAACCTGGAGCAGTAGCAA
ACCCTTTCTACATATTTGGTGATGTAGGTTTAGGTAAAACCCATCTTATGCAAGCAATAGGTAACTATAT
ATTAGATAATGATGTCGAAAAACGTATCTTATATGTTAAAGCTGATAACTTTATTGAAGACTTTGTATCT
TTATTATCAAGAAACAAAAATAAGACTGAAGAATTCAATGCTAAATATAAAGATATTGACGTTATATTAG
TAGACGACATCCAAATTATGGCAAATGCTAGTAAAACTCAAATGGAATTCTTTAAGCTCTTTGACTACCT
ATATTTAAATAATAAACAAATCGTTATAACATCTGATAAACCAGCTTCACAATTAACAAATATCATGCCA
CGTTTAACGACACGTTTTGAAGCTGGTCTCTCTGTAGACATACAAATACCTGAATTAGAACATAGAATAA
GTATTTTAAAGAGAAAAACAGCTACATTAGATGCCAACTTAGAGGTAAGTGAGGATATTTTAACCTTTAT
AGCATCTCAATTTGCAGCAAACATTAGAGAAATGGAAGGTGCACTCATTCGTTTAATTAGTTATGCACAG
ACCTTTAATCTAGAAATTACAATGAATGTTGTTGAAGAAGCACTTGGTGCTGTATTAAAAACTAAGAAGA
AAACAAATGATTTAAACGAAAATAACTACGATAAGATCCAAAGTATTGTAGCAGATTACTTCCAAGTTTC
ATTACCAGACTTAATTGGTAAGAAAAGACATGCTAAATTCACATTACCTAGACATATAGCGATGTATCTT
ATTAAACTTAAATACAATATTCCTTATAAAACAATTGGATCTTTATTTAATGATAGAGACCACTCCACAG
TATTATCTGCTTGTGAAAAAGTAGAACGCGATATGAGGATGGATTCGAACTTAAAGTTTGCTGTTGACTC
GATTGTCAAAAAAATAGATTCACCATTATTAAAGTGATAAATGTTTATAAAAATGATTAAATGTGGTAAA
ATAAATGGTAGATGAAGCGATTATTGCTAGTTTCCCACTTTCCCACAGACACTAACAATAACAAAGAAGA
AATAATAATTAATAAAAGGGGTAAAATATGAATTTTACAATTGAAAGAGATAT...............
```

Genome Sizes Global view



1kb - 360kbp

2.5 Mbp

Human: 3 Gbp

Paris japonica: 150 Gbp

2.3 Mbp

670 Gbp

0.1 Mbp

16 Mbp

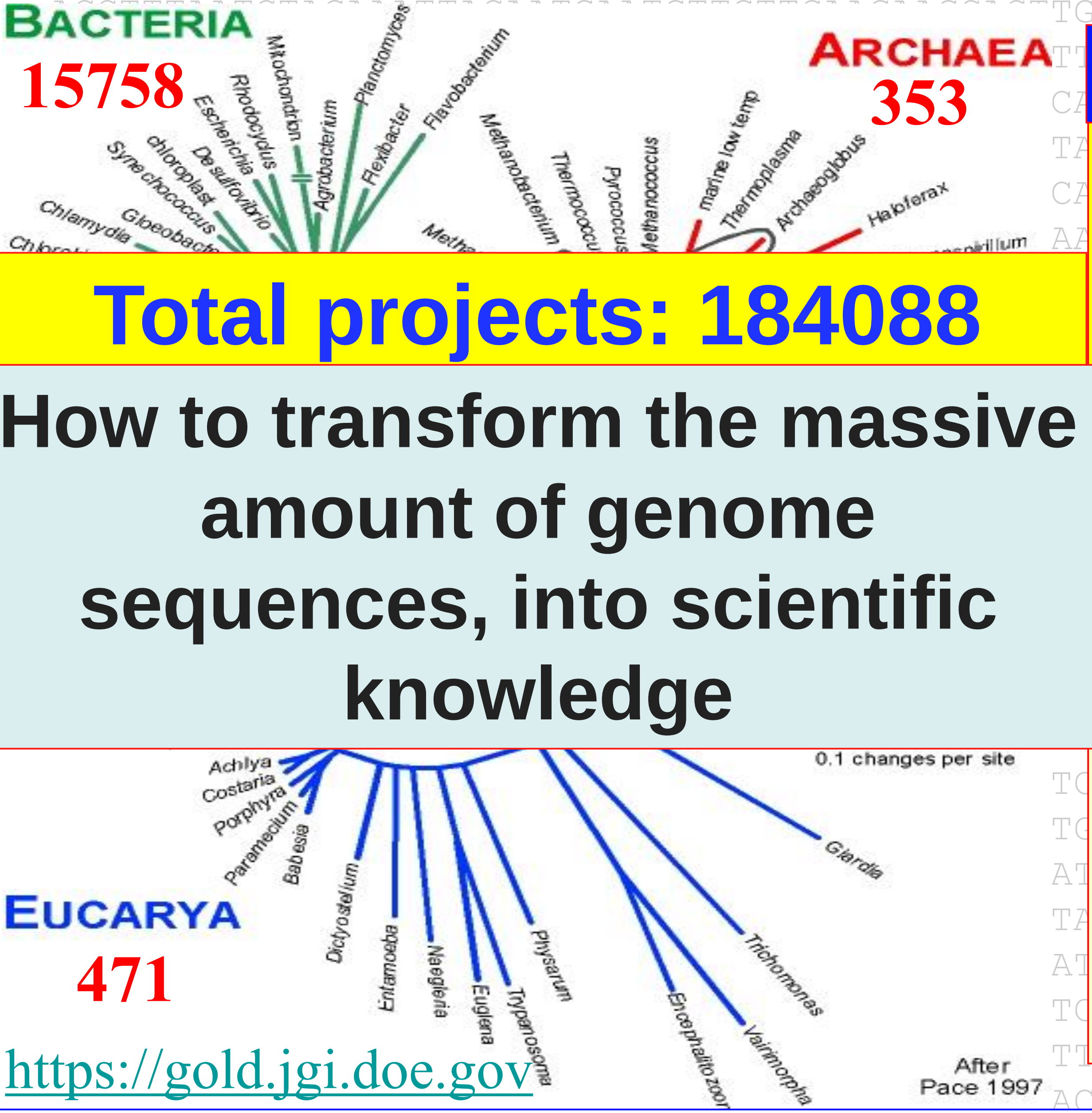
0.49Mbp

5.75Mbp

Pennisi E. (2013). Science. 341:226-7.

TATTTGATTTTTTGCTTAATTTTTATGTAATTACTTCCCCACAATTTTTGCCCACATATTGTGGATAAAT

TGGTGCTGTATTAAAAACTAAGAAGA



15758

353

471

15/10/2020

Complete sequenced genomes: 16582

- **15758 Bacteria**
- **353 Archaea**
- **471 Eukaryotes**

Incomplete genomes: 23228

- **14584 Bacteria**
- **218 Archaea**
- **8166 Eukaryotes**

Permanent Draft genomes: 144278

- **138651 Bacteria**
- **1073 Archaea**
- **4554 Eukaryotes**

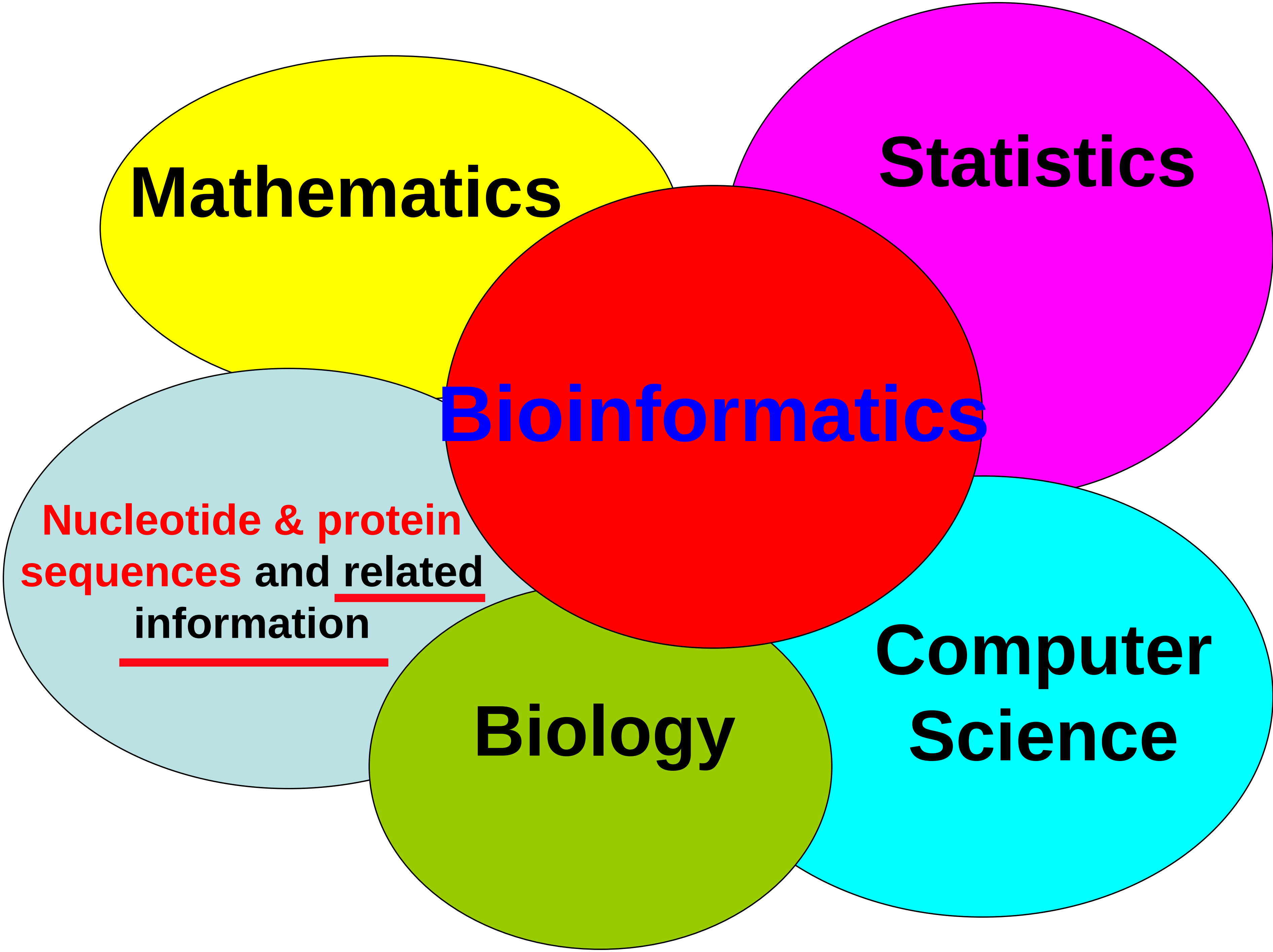
Total projects: 184088

How to transform the massive amount of genome sequences, into scientific knowledge

https://gold.jgi.doe.gov

ACAGACACTAACAATAACAAAGAAGA

Viruses: • Completed: 3502 • Permanent draft: 5083


Mathematics
Statistics
Bioinformatics
Nucleotide & protein sequences and related information
Biology
Computer Science

Examples of used Mathematics & Statistics in Bioinformatics

Biology Information creates interesting research questions

Algorithms – Data Structure – Theory - Model
-Probability; Information Theory,
-Most statistical models: Bayesian, Maximum Likelihood, Parsimony,
Distribution: Binomial, Multinomial, Normal,
Statistical tests: t-test, Chi square test, F-test,
Similarity (score, distance,..)
Graph Theory,
Matrices: diagonalization, Eigen values, Eigen vectors, Sum, Products,..
Multidimensional data Analysis Methods (Principal Component Analysis, Correspondence Analysis, Discrimination Analysis,...),
Clustering Methods (K-means, hierarchical clustering, Neighborhood,...),
Machine Learning,
Deep Learning Methods,
Support Vector Machines, Random Forests, Hidden Markov Models,
Bayesian Networks, Gaussian Networks
Neural Network,………

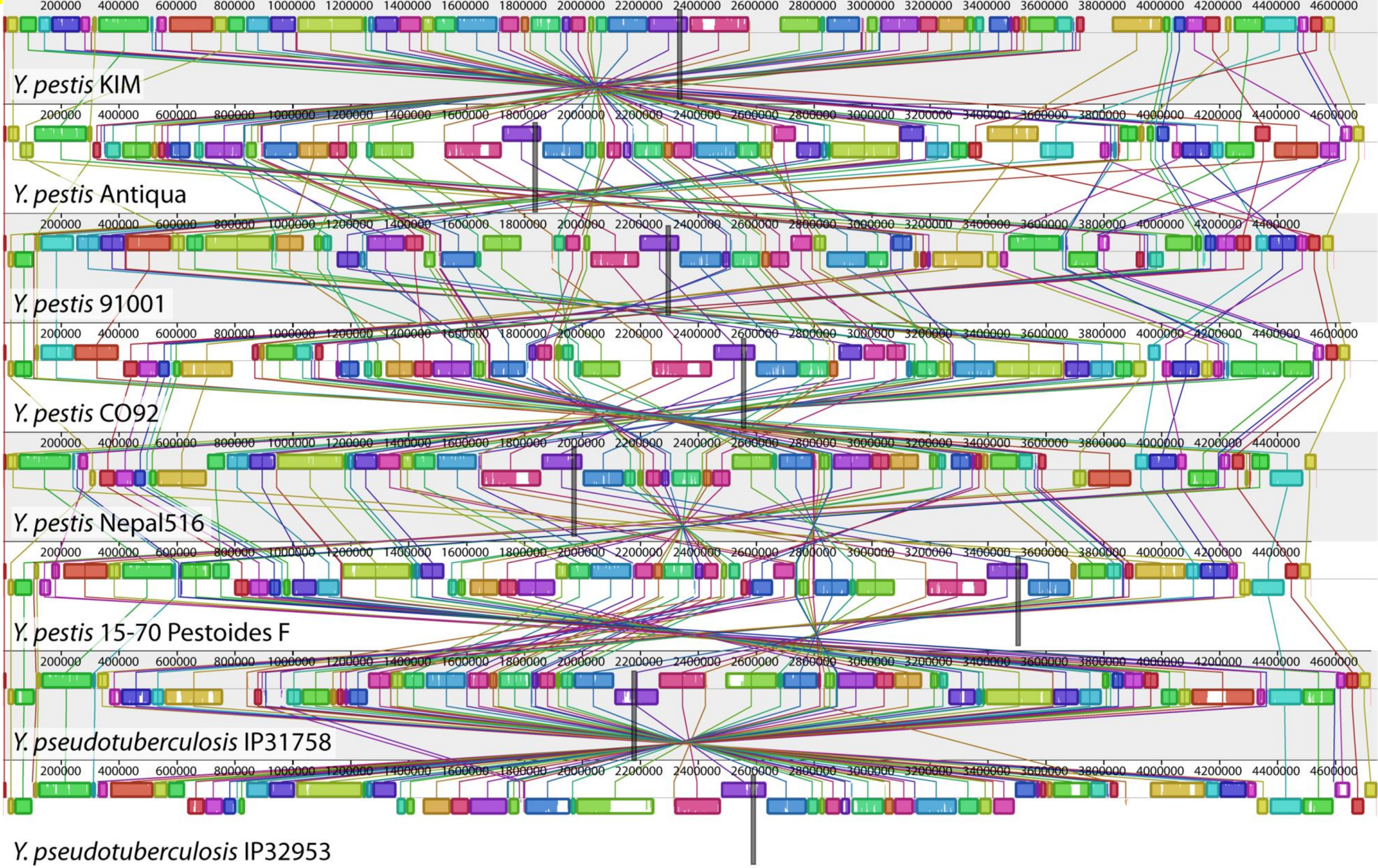
Bioinformatics in the genome era: Methods and Tools

- Genome Assembly (most important step)
- Gene prediction resources
- Large-scale genome comparisons
- Phylogenomics - Evolution
- Whole genome alignments
- Genome Structural Variants
- Relationships visualization

Large-scale genome comparisons

- Duplication (genes, chromosome segments, whole genome)
- Conservation (genes, chromosome segments);
- Specificity (species-specific genes);
- Inferring Paralogs, orthologs and clustering;
- Protein conservation profiles;
- Gene Transfer, introgression between species;
- Genome structural variant detection

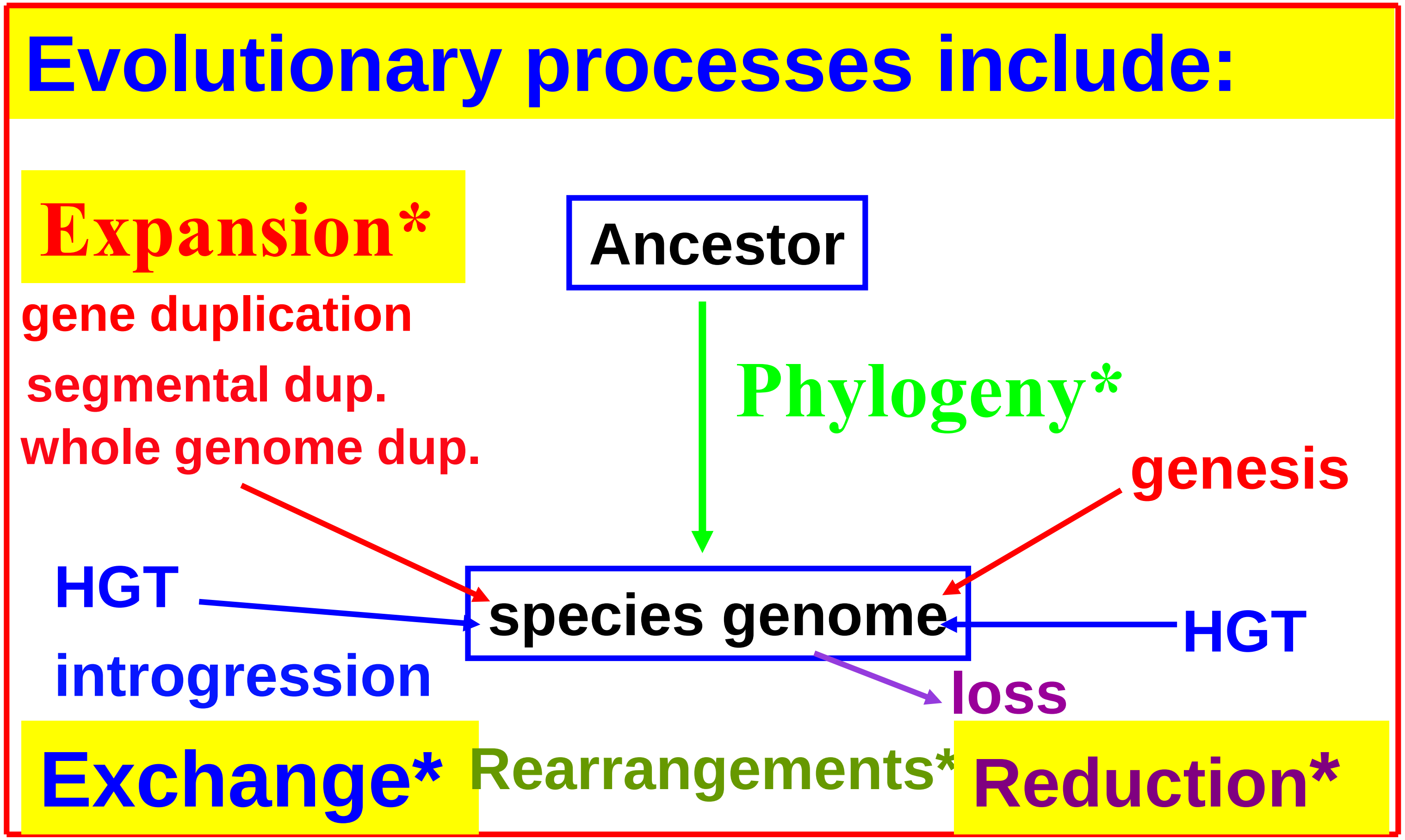
Multiple whole genome alignments



Aaron E. Darling, Istvan Miklos, Mark A. Ragan (2008). Dynamics of Genome Rearrangement in Bacterial Populations. PLoS Genet. 2008 July; 4(7): e1000128.

Large-scale comparative predicted proteomes revealed significant evolutionary processes:

Expansion, Exchange and Reduction



Evolutionary processes include:

Expansion*

gene duplication

segmental dup.

whole genome dup.

Ancestor

Phylogeny*

genesis

HGT

species genome

HGT

introgression

loss

Exchange*

Rearrangements*

Reduction*

HGT & introgression emerged as major evolutionary processes that shaped genomes

Not all our genes are from our ancestors - some are “foreign”!

Mušhuššu (Sirrush: serpent/dragon rouge). Ishtar gate. Pergamon Museum, Berlin

http://www.medicalnewstoday.com/articles/290835.php

Genomic Structural Variant Detection

Two types:

- **Single nucleotide variants (1, few bases)**

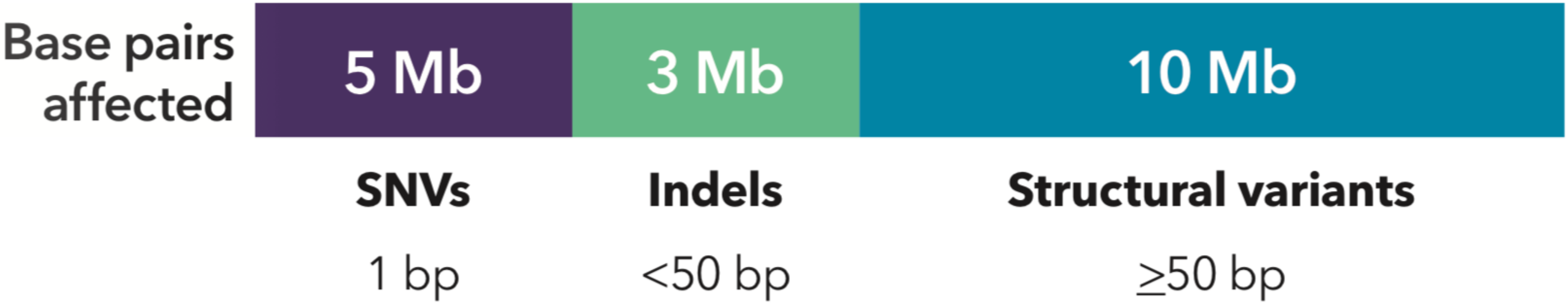
SNP, Insertion, Deletion, Indel, Substitution.

- **Large structural variants (>= 50bp)**

CNV, Inversion, Translocation.

Quantification of Human Genetic Variation?

- **Human genome has ~20,000 structural variants, spanning 10 million base pairs, more than twice the number of bases affected by SNVs.**



- **Structural variants of all types are known to cause Mendelian disease and contribute to complex disease**

https://www.pacb.com/wp-content/uploads/Whitepaper-Human-Structural-Variation.pdf

TATTTGATTTTTTGCTTAATTTTTATGTAATTACTTCCCCACAATTTTTGCCCACATATTGTGGATAAAT
ACCTTTAATCTAGAAATTACAATGAATGTTGTTGAAGAAGCACTTGGTGCTGTATTAAAAACTAAGAAGA
TTTCCACATTTTATTCACAATGTTGATAAGTTGTGTAAGTCGACTTGTTGGCTTTATAAAGCAAATGACA
CACTGAAAGTTATCCACAAACAAATTTTTCAAAAGTGTTTAAATCAAAAAGTTATCCACAAATAATGTGG
AAAACTTTTTAATAAATTTGTCGTTTCTTATGCTATCATAGTTTTACATAAATTATTAACTACTATAGGG
AGGCAGTCATGAGTCCAAACAGCACATTATGGCAGACAATATTACAGGATTTAGAAAACTATACAACGA
GGAGACTTACAACGAGCTATTTCTACCAGTGACTTCTACTTTTAAAGATCAAAACGGATTACTTACAATG
GTTGTAGCTAATGAGTTCTTAAAGAATCGTATCAATAAACTATACATTGCAAAAATTAACGAACTTGCTA
CAAAATATTCAAGTACTCCAGTTAGATTGAAATTCGTATCACAAGAAGAAGTTATTGAAGAACCAGTAGC
TGATCGTAAATTAACCATTGATTATCGTCAAGGGAACTTAAACTCTACCTATACCTTTGACTCTTTTGTT

Examples of Large Projects of Genome Sequencing

CGTTTAACGACACGTTTTGAAGCTGGTCTCTCTGTAGACATACAAATACCTGAATTAGAACATAGAATAA
GTATTTTAAAGAGAAAAACAGCTACATTAGATGCCAACTTAGAGGTAAGTGAGGATATTTTAACCTTTAT
AGCATCTCAATTTGCAGCAAACATTAGAGAAATGGAAGGTGCACTCATTCGTTTAATTAGTTATGCACAG
ACCTTTAATCTAGAAATTACAATGAATGTTGTTGAAGAAGCACTTGGTGCTGTATTAAAAACTAAGAAGA
AAACAAATGATTTAAACGAAAATAACTACGATAAGATCCAAAGTATTGTAGCAGATTACTTCCAAGTTTC
ATTACCAGACTTAATTGGTAAGAAAAGACATGCTAAATTCACATTACCTAGACATATAGCGATGTATCTT
ATTAAACTTAAATACAATATTCCTTATAAAACAATTGGATCTTTATTTAATGATAGAGACCACTCCACAG
TATTATCTGCTTGTGAAAAAGTAGAACGCGATATGAGGATGGATTCGAACTTAAAGTTTGCTGTTGACTC
GATTGTCAAAAAAATAGATTCACCATTATTAAAGTGATAAATGTTTATAAAAATGATTAAATGTGGTAAA
ATAAATGGTAGATGAAGCGATTATTGCTAGTTTCCCACTTTCCCACAGACACTAACAATAACAAAGAAGA
AATAATAATTAATAAAAGGGGTAAAATATGAATTTTACAATTGAAAGAGATATACGTTCAGTTAACCGGA

- Human 1000 genomes project (2504 individuals)

A global reference for human genetic variation

Nature 526, 68-74 (01 October 2015) doi:10.1038/nature15393

- Genomic England: The 100K Genomes Project

https://www.genomicsengland.co.uk/the-100000-genomes-project/

The project was established to sequence 100,000 genomes from National Health Service patients affected by rare diseases or cancer.

- 100K Pathogen Genome Project

https://100kgenomes.org

The 100K Pathogen Genome Project is producing genome sequences of diverse pathogens from food, animal disease, human disease, wildlife, and environmental reservoirs of those pathogens.

Genomic island

Nature Genetics 47,1221, 2015.https://doi.org/10.1038/ng.3436

- **Iceland** (https://www.nature.com/articles/ng.3247) (**2015**)

Whole genomes of 2,636 Icelanders. https://doi.org/10.1038/ng.3247

- **The UK10K (http://www.uk10k.org/)**

(***Nature* 526, 82-90, 2015. https://doi.org/10.1038/nature14962.**)

The UK10K project identifies rare variants in health and disease

- **The SardiNIA Project (https://sardinia.irp.nia.nih.gov/)**

Large project Including 3400 complete sequenced genomes

Identify genetic bases for prominent age-associated changes

A 1000 Arab genome project to study the Emirati citizens

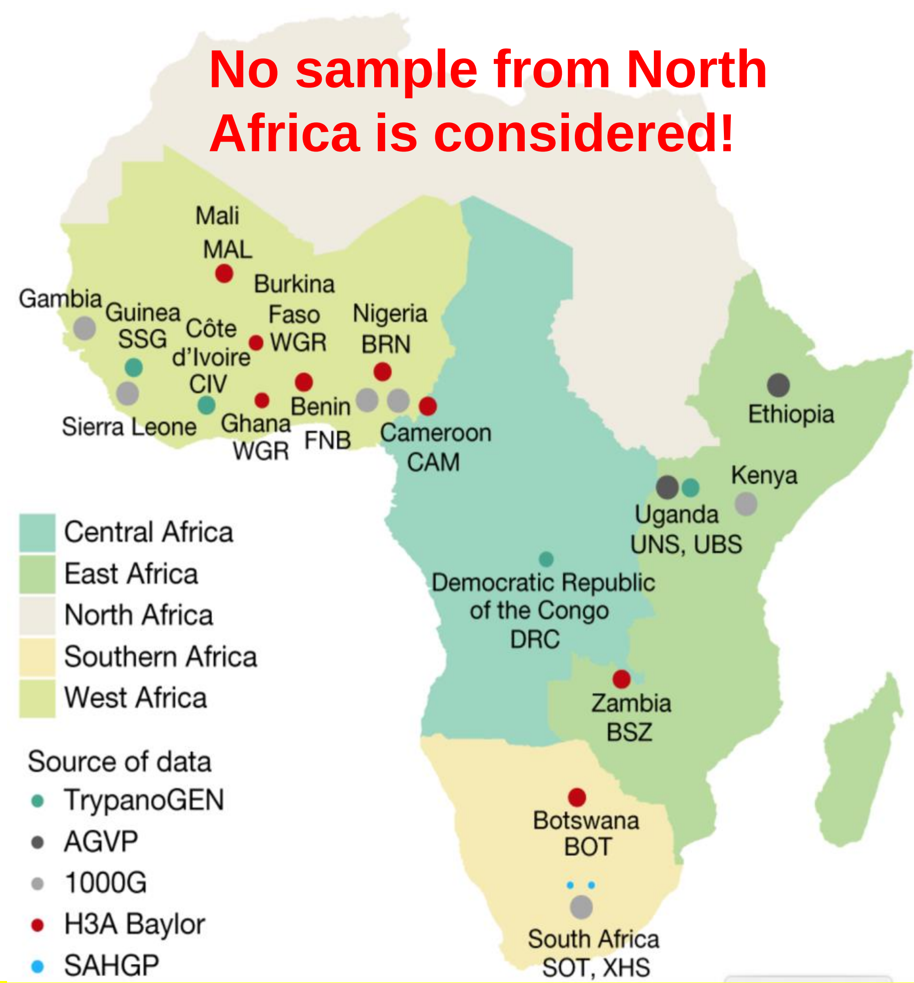
- The project aims to sequence the genome of Emirati citizens. (10% of 11 million residents in the UAE).

Goals include:

- The creation of an Emirati genome reference
- The identification of specific genetic variations
- Provide genome data for future healthcare applications in the UAE

Al-Ali, M. *et al.* A 1000 Arab genome project to study the Emirati population. *J Hum Genet* 63, 533–536 (2018). https://doi.org/10.1038/s10038-017-0402-y

Exploring the breadth of genomic diversity across Africa 2020



No sample from North Africa is considered!

- **Whole genome sequencing of 426 individuals, from 50 ethnolinguistic groups.**

- **3 million (31,160,639) previously undescribed variants**
- **62 previously unreported chromosomes locations that are under strong selection pressure, predominantly found in genes that are involved in viral immunity, DNA repair and metabolism.**

Choudhury, A. *et al.* High-depth African genomes inform human migration and health. *Nature* 586, 741–748 (2020). https://doi.org/10.1038/s41586-020-2859-7

Main ideas developed so far:

- Genomics and its implication in a variety of Domains
- Availability of Massive amounts of genome data
- Importance of Bioinformatics Methods & Tools
- Many large Projects of Genome Sequencing

II

Two examples of genome structural variants in human health

Microbiome

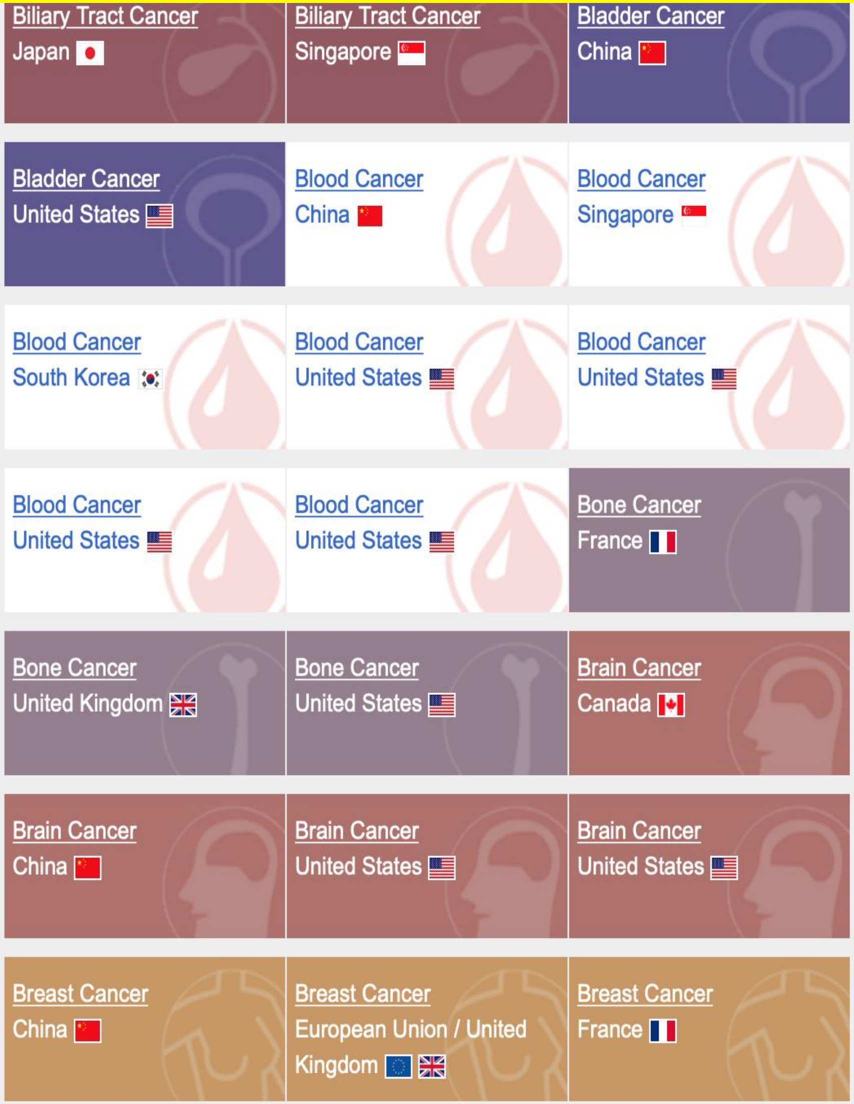
The Genome Aggregation Database (gnomAD)

https://gnomad.broadinstitute.org

- Genetic variants that inactivate protein-coding genes are a powerful source of information about the phenotypic consequences of gene disruption.
- To capture the extent of genome variation among a large group of individuals, the Genome Aggregation Database (gnomAD) has aggregated 15,708 whole genomes and 125,748 exomes (the protein-coding part of the genome).
- Analyses of these data provided a catalogue of the different types of variation,
revealing their potential functional impact and
how this information could help to identify disease-causing mutations and guide on the choice of therapeutic protocols.
- Also allowed to identify 443,769 high-confidence predicted Loss-Of-Function variants (https://doi.org/10.1038/s41586-020-2308-7)

International Cancer Genome Consortium
(https://icgc.org/)

Cancer Genome Project (90 project teams) 84 tumor types



Biliary Tract Cancer Japan	Biliary Tract Cancer Singapore	Bladder Cancer China
Bladder Cancer United States	Blood Cancer China	Blood Cancer Singapore
Blood Cancer South Korea	Blood Cancer United States	Blood Cancer United States
Blood Cancer United States	Blood Cancer United States	Bone Cancer France
Bone Cancer United Kingdom	Bone Cancer United States	Brain Cancer Canada
Brain Cancer China	Brain Cancer United States	Brain Cancer United States
Breast Cancer China	Breast Cancer European Union / United Kingdom	Breast Cancer France

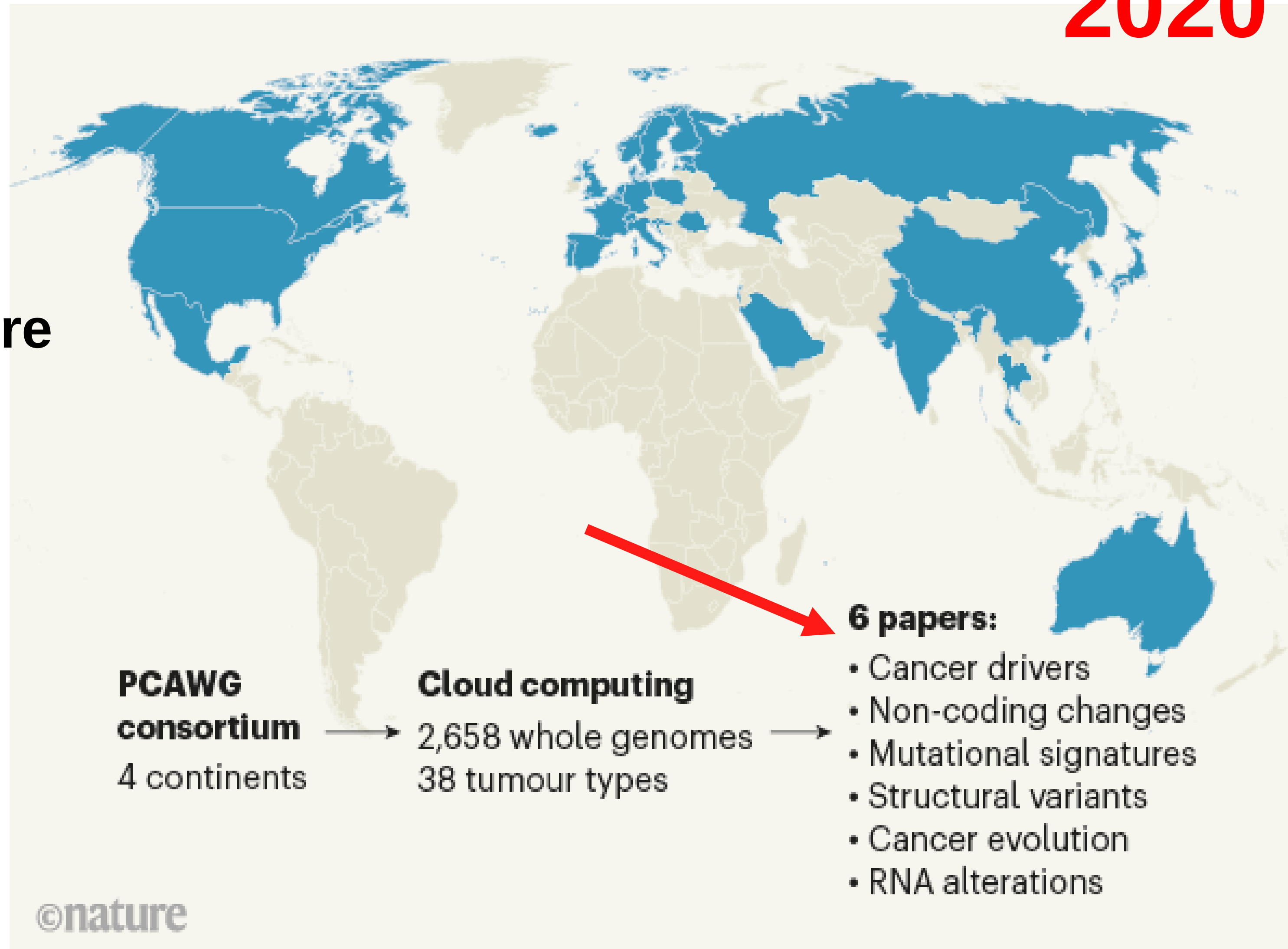
1. The ICGC aimed to define the **genomes of 25,000** primary **untreated cancers** (**the 25K Initiative**).

2. The Pan Cancer Analysis of Whole Genomes (**PCAWG**), known as the **Pan-Cancer Project**, **defined similarities and differences between cancer types**.

Pan-cancer Analysis of Whole Genomes

2020

A massive international effort has yielded multifaceted studies of more than 2,658 tumours from 38 tissues, generating a wealth of insights into the genetic basis of cancer.



Nature **578**, 39-40 (2020). https://doi.org/10.1038/d41586-020-00213-2

The repertoire of mutational signatures in human cancer

- **Characterization of mutational signatures using 84,729,690 somatic mutations from 4,645 whole-genome and 19,184 exome sequences that encompass most types of cancer.**

- **Identification of mutational signatures: 49 single-base-substitution (SBS), 11 doublet-base-substitution (DBS), 4 clustered-SBS and 17 small insertion-and-deletion.**

- **This analysis provided a systematic perspective on the repertoire of mutational processes that contribute to the development of human cancer.**

Alexandrov et al. Nature volume 578, pages94–101(2020). https://doi.org/10.1038/s41586-020-1943-3

Metagenomics

The field of metagenomics is rapidly evolving and becomes a topic of great scientific and public interest.

Human Microbiome Project (HMP)

A bag of surprises

Aims to characterize microbial communities found in human body including: nasal passages, oral cavities, skin, gastrointestinal tract, and uro-genital tract and to analyse the role of these microbes in human health and disease.

ATAAATGGTAGATGAAGCGATTATTGCTAGTTTCCCACT

http://commonfund.nih.gov/hmp/

Microbiome studies have already allowed significant advances in many diseases

- **Understanding the relationships between gut microbes and host phenotypes is critical for understanding wellness and disease.**

- **Recent studies showed well-established links between the diversity of for example a healthy gut microbiome and protection against diseases.**

Nussinov R, Papin JA (2017) How can computation advance microbiome research? PLoS Comput Biol13(9): e1005547.

…Microbiome studies have already allowed significant advances in many diseases

- There is emerging evidence for relationships between disruptions in the human microbiome and among many other diseases:

- Cancer,

Vogtmann E, Goedert JJ, (2016) Epidemiologic studies of the human microbiome and cancer. *British Journal of Cancer*, 114 (3) 237–42.

- Cardiovascular disease,

Wang Z, et al. (2011). Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472 (7341) 57–63 pmid:21475195.

- Obesity,

Cox LM, et al. (2014) Altering the Intestinal Microbiota during a Critical Developmental Window Has Lasting Metabolic Consequences. *Cell* 158 (4) 725–721.

Microbiome studies have already allowed significant advances in many diseases

…….continued

- Food allergies,

Stefka AT, et al. (2014) Commensal bacteria protect against food allergen sensitization. *PNAS* 111 (36) 13145–13150 pmid:25157157.

- Asthma

Huang YJ, Boushey HA, (2015) The microbiome in asthma. *The Journal of Allergy and Clinical Immunonlogy* 135 (1) 25–30.

- **Microbiota has important roles in cancer therapy.**

Roy S, Trinchieri G, (2017) Microbiota: a key orchestrator of cancer therapy. *Nature Reviews Cancer* 17, 271–285.

- **The microbiome plays a critical role in immune systems development.**

Zheng, D., Liwinski, T. & Elinav, E. *Cell Res* 30, 492–506 (2020). https://doi.org/10.1038/s41422-020-0332-7

Health and disease markers correlate with gut microbiome composition

2020

- Examination of the gut microbiota and ~150 host phenotypic features across ~3,400 individuals reveals associations between microbiome composition and host clinical markers and lifestyle factors.

- These results suggest, in order to improve host health, potential opportunities for targeted interventions that alter the composition of the microbiome.

Manor, O., Dai, C.L. *et al.* *Nat Commun* 11, 5206 (2020).
https://doi.org/10.1038/s41467-020-18871-1

Main ideas developed so far:

- Investigations on structural variants in cancer genomics allowed to determine:

-A catalogue of different types of variation
-A repertoire of Mutational signatures

- Investigations on gut microbiome and disease shows links between:

-Healthy gut microbiome and protection against diseases
-Disruption in human microbiome and diseases

III

Ancient Genome Sequencing

- Sequencing technologies developments have triggered considerable progress in the study of ancient DNA (aDNA), enabling complete genomes of past humans and organisms to be reconstructed.

- Providing new insights into human migration, evolution, adaptation and disease susceptibility through time.

- Also the reconstruction of Ancient pathogens

Fast evolving domain with continuous new discoveries

Late Pleistocene sites that have yielded Neanderthal DNA (red) and Denisovan DNA (blue)

2013

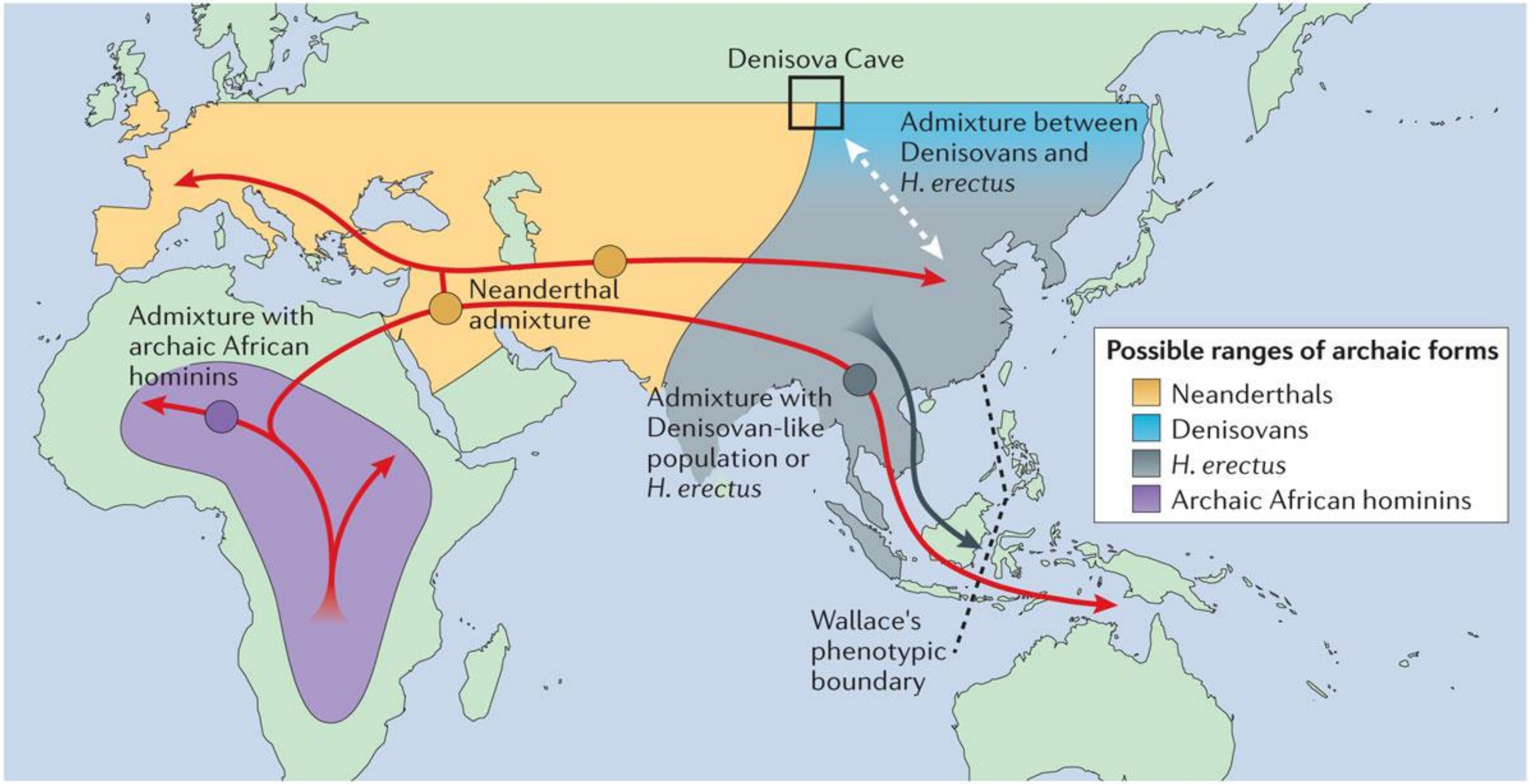
M Meyer *et al. Nature* 000, 1-4 (2013) doi:10.1038/nature12788

nature

A mitochondrial genome sequence of a hominin from Sima de los Huesos

The impact of whole-genome sequencing on the reconstruction of human population history

2014

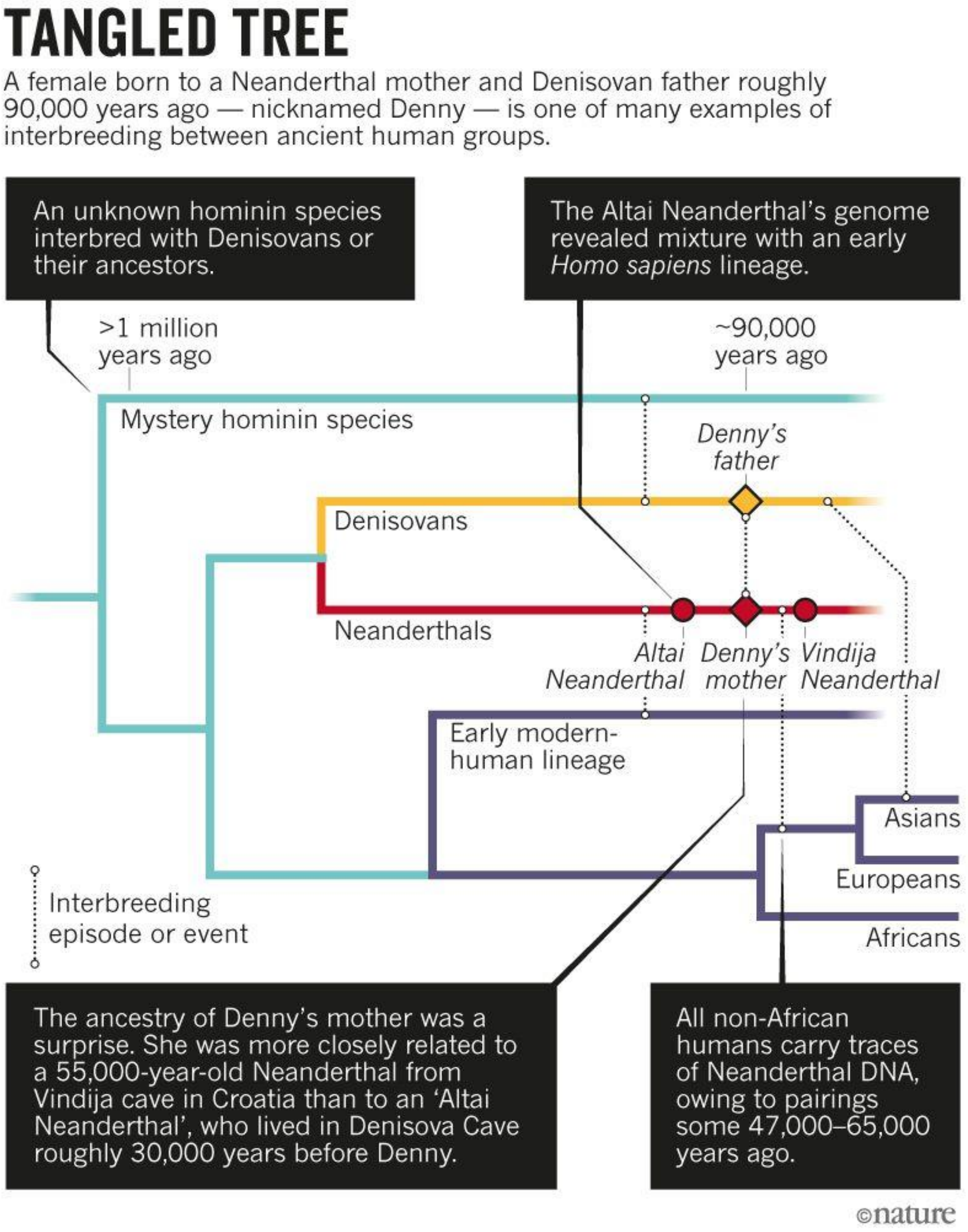


Nature Reviews | Genetics

Krishna R. Veeramah & Michael F. Hammer (2014). The impact of whole-genome sequencing on the reconstruction of human population history. *Nature Reviews Genetics* 15, 149-162 (2014)

Mum's a Neanderthal, Dad's a Denisovan: First discovery of an ancient-human hybrid dating 90 kyr ago in Siberia. 2018

Although the girl's remains were found in Siberia, her Neandertal DNA more closely matches a western European Neandertal from Vindija Cave in Croatia — thousands of kilometers to the west — than an older Neandertal from the same cave as the girl. That finding may mean that eastern Neandertals spread into western Europe sometime after 90,000 years ago, or that western Neandertals beat them to the punch, invading eastward into Siberia before 90,000 years ago and partially replacing the Neandertals living there. Researchers need to test more DNA from western European Neandertals to determine which scenario is correct.



TANGLED TREE

A female born to a Neanderthal mother and Denisovan father roughly 90,000 years ago — nicknamed Denny — is one of many examples of interbreeding between ancient human groups.

An unknown hominin species interbred with Denisovans or their ancestors.

The Altai Neanderthal's genome revealed mixture with an early *Homo sapiens* lineage.

The ancestry of Denny's mother was a surprise. She was more closely related to a 55,000-year-old Neanderthal from Vindija cave in Croatia than to an 'Altai Neanderthal', who lived in Denisova Cave roughly 30,000 years before Denny.

All non-African humans carry traces of Neanderthal DNA, owing to pairings some 47,000–65,000 years ago.

Nature 560, 417-418 (2018). *doi: 10.1038/d41586-018-06004-0*

The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes 2020

- Whole-genome sequences from 27,566 Icelanders (10% of pop) to 292 Sub-Saharan complete genomes detected 14.4 million putative archaic chromosome fragments i.e found in Icelanders and not in Africans. (Africans assumed to have no Neanderthal introgression)

- After combining identical and overlapping fragments, the authors identified more than 56,388 distinct archaic fragments covering 38-48% of the readable genome.

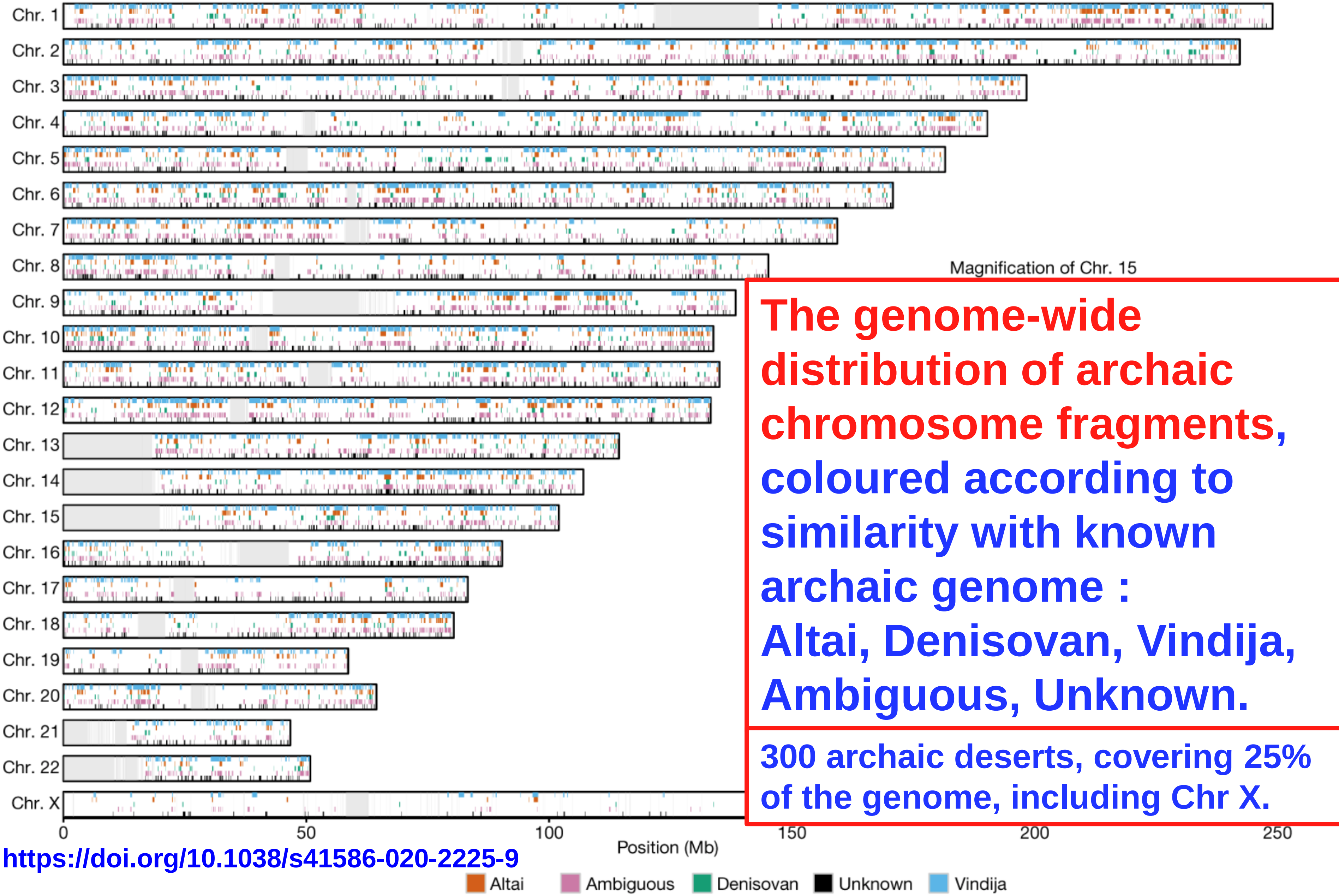
- These archaic fragments carried 395,304 SNPs that are absent in Africans. Of these 147,925 were found in Denisovan, Altai Neanderthal and Vindija Neanderthal genomes (see distribution next slide).

- Note that nearly 300 “archaic deserts” where there are no archaic fragments; these cover nearly 25% of the genome, including the entire X chromosome.

Skov, L., *et al.* The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. *Nature* 582, 78–83 (2020). https://doi.org/10.1038/s41586-020-2225-9

84.5% of archaic fragments assigned to Altai or Vindija Neanderthal origin, 3.3% to Denisovan origin and 12.2% unknown origin.

2020



The genome-wide distribution of archaic chromosome fragments, coloured according to similarity with known archaic genome : Altai, Denisovan, Vindija, Ambiguous, Unknown.

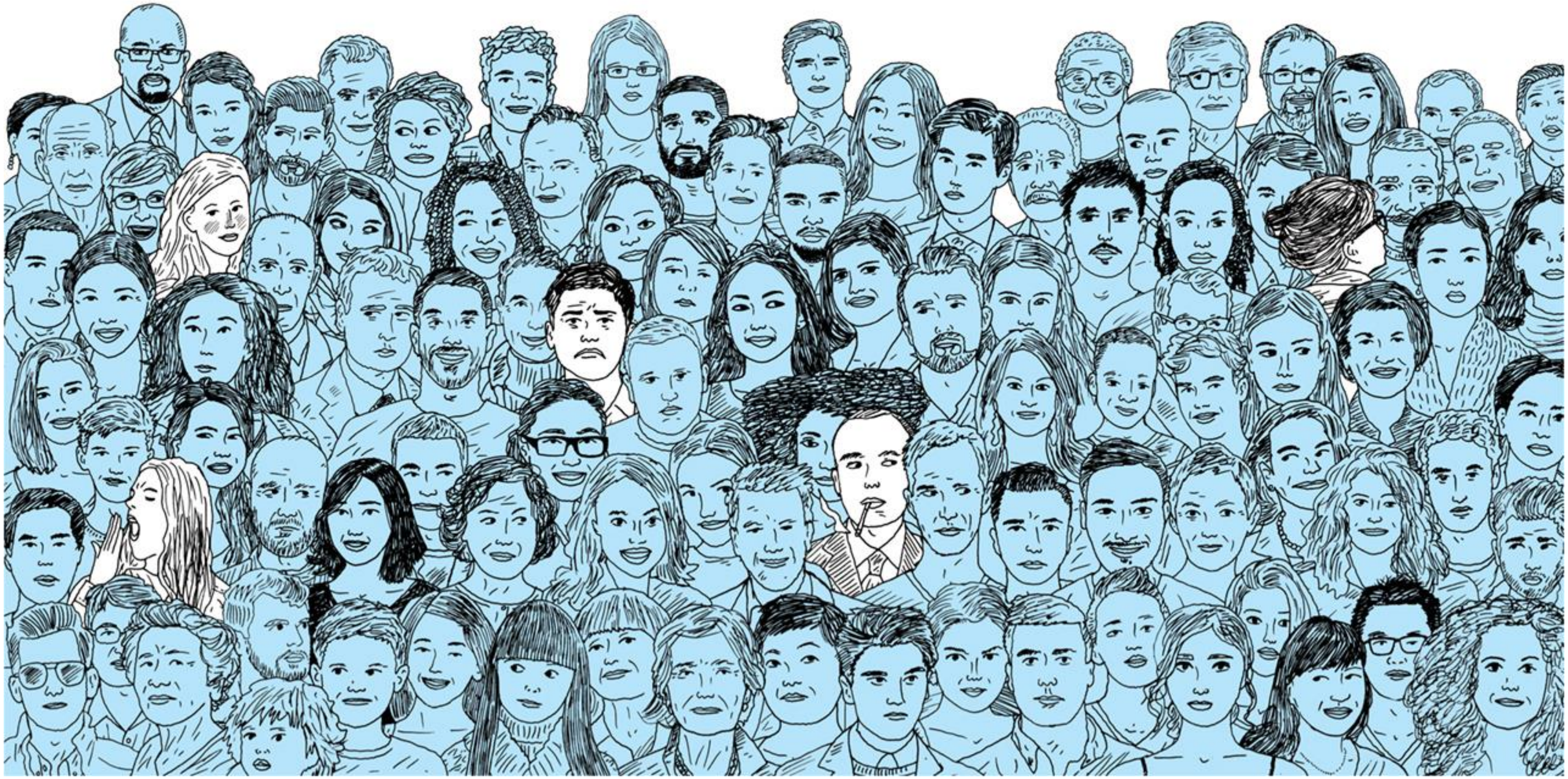
300 archaic deserts, covering 25% of the genome, including Chr X.

https://doi.org/10.1038/s41586-020-2225-9

The half-million people in the UK Biobank hold the genetic legacy of Neanderthals and clues to how we are still evolving.

UK Biobank (UKB), a large database that holds genetic and phenotypic health records for 500,000 British volunteers.

2019



Neanderthal DNA legs predispose to traits such as propensity to sunburn, staying up late, depression, smoking, and feeling lonely.

Ann Gibbons (2019). Spotting Evolution Among us. Science 363, 6422, 21-23.
https://doi.org/10.1126/science.363.6422.21

Neanderthal DNA highlights complexity of risk factors in modern humans

- Some 1–4% of the modern human genome comes from ancient relatives: Denisovans and Neanderthals (50-60 kyr ago) (Green et al., 2010).

- Many of the surviving archaic genes are harmful to modern humans, and are associated with infertility and an increased risk of disease.

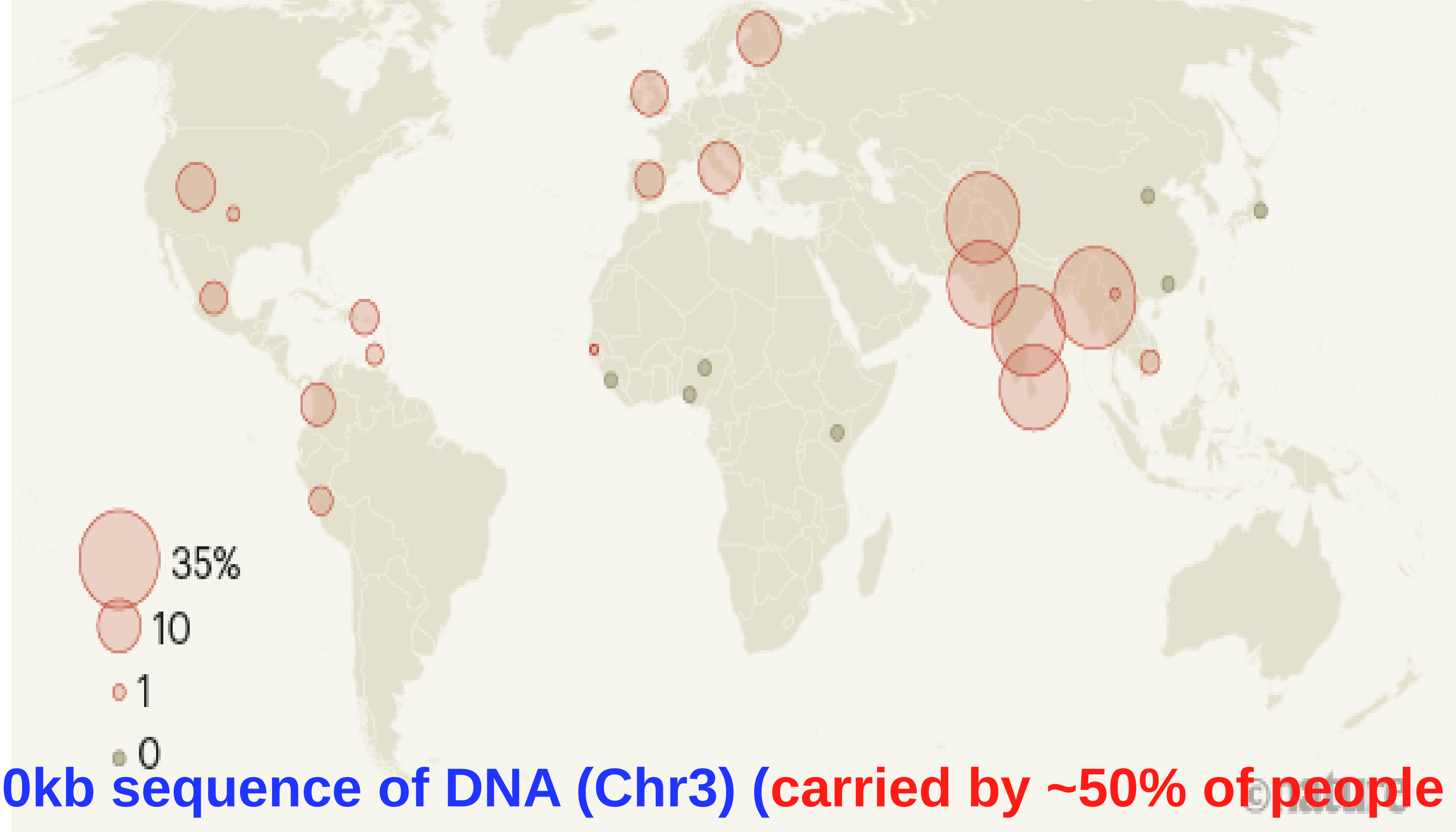
- But few archaic genes are beneficial:

\- the Denisovan-like version of a gene called *EPAS1* that helps modern Tibetans to cope with life at extremely high altitudes,

\- a Neanderthal gene that increases sensitivity to pain and others that help us fend off viruses.

•Hu H, et al. (2017) Evolutionary history of Tibetans inferred from whole-genome sequencing. PLoS Genet 13(4): e1006675. https:// doi.org/10.1371/journal.pgen.1006675
•https://doi.org/10.1038/d41586-020-02957-3

The major genetic risk factor for severe COVID-19 is inherited from Neanderthals



A 50kb sequence of DNA (Chr3) (carried by ~50% of people in south Asia and ~16% of people in Europe) is associated with severe respiratory failure upon COVID-19 infection.

This genomic segment is inherited from Neanderthals.

Zeberg, H. & Pääbo, S. *Nature* https://doi.org/10.1038/s41586-020-2818-3. (2020)

Ancient Pathogens: Collapse of Aztec society linked to catastrophic *salmonella enterica* outbreak

- Ancient bacterial DNA recovered from several of the victims matched that of *Salmonella*, based on comparisons with a database of more than 2,700 modern bacterial genomes.

- DNA of 500-year-old bacteria is first direct evidence of the worst epidemic in human history that devastated Mexico's native population by killing more than 80% of them.

2018

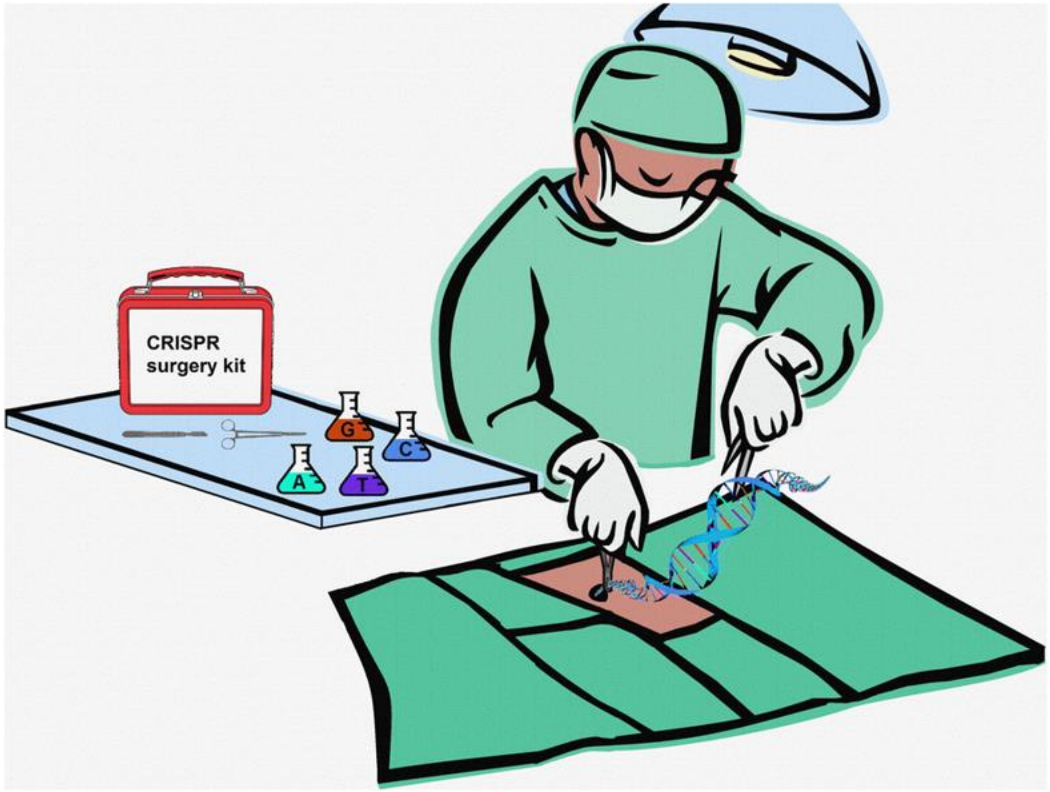
Vågene, Å.J., *et al. Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat Ecol Evol* 2, 520–528 (2018). https://doi.org/10.1038/s41559-017-0446-6

Main ideas developed so far:

- Reconstruction of ancient human genomes: Neanderthal and Denisovan
- DNA legs in modern human and consequences
- Ancient Pathogen genomes highlight on historical catastrophic epidemics

IV

Genome Editing



“Understanding by creating”

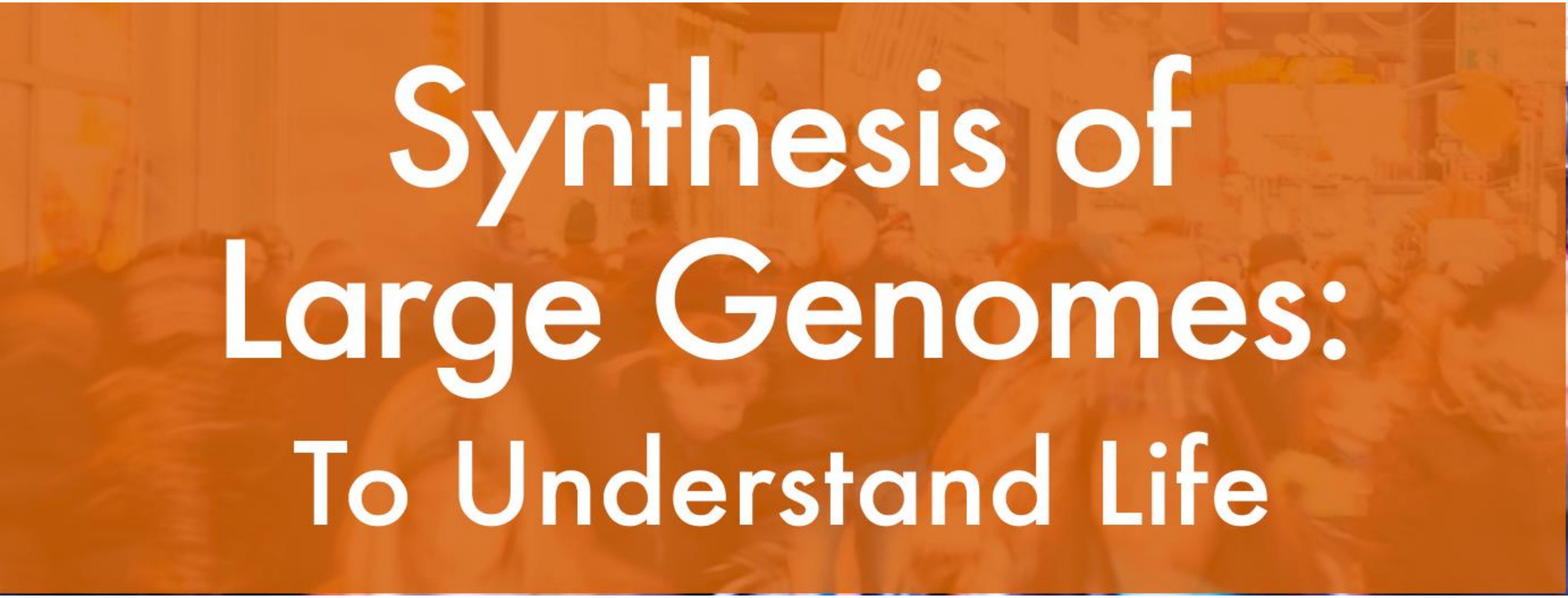
inspired the efforts to generate synthetic cells.

Special reviews on synthetic biology:
https://www.nature.com/nature/focus/synbio/index.html

Kannan K and Gibson DG. 2017. Yeast genome, by design. Science, 2017;35pp.1024-1025

Genome Project-write: GP-write

http://engineeringbiologycenter.org



GP-write will enable scientists to move beyond Observation to Action.

Minimal Genome
(Minimum genetic material needed for life)

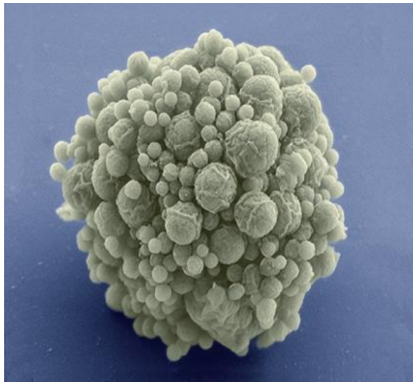
Designing and building a minimal bacterial genome

2016

it was possible to reduce the size of the *Micoplasma mycoides* genome by around 50% with retention of quasi-essential genes, producing JCVI-syn3.0 (531 kilobase pairs, 473 genes).

Procedure:

- **Redesigning genome segments by computer,**
- **Chemically synthesizing the fragments, then**
- **Assembling them.**



Design and synthesis of a minimal bacterial genome. Hutchison CA 3rd, et al. Science. 2016;351(6280). doi: 10.1126/science.aad6253

Another kind of minimization: 2019 codon reduction

Through design, synthesis and assembly, it was possible to construct an *E. coli* genome that uses only 61 of the 64 available codons in its protein-coding sequences, replacing 2 serine codons and one stop codon with synonyms.

Fredens, J. *et al. Nature* 569, 514–518 (2019)

The Synthetic Yeast Genome Project (Sc2.0)

Richardson SM, et al. Design of a synthetic yeast genome. Science. 2017;355(6329):1040–4.

2017

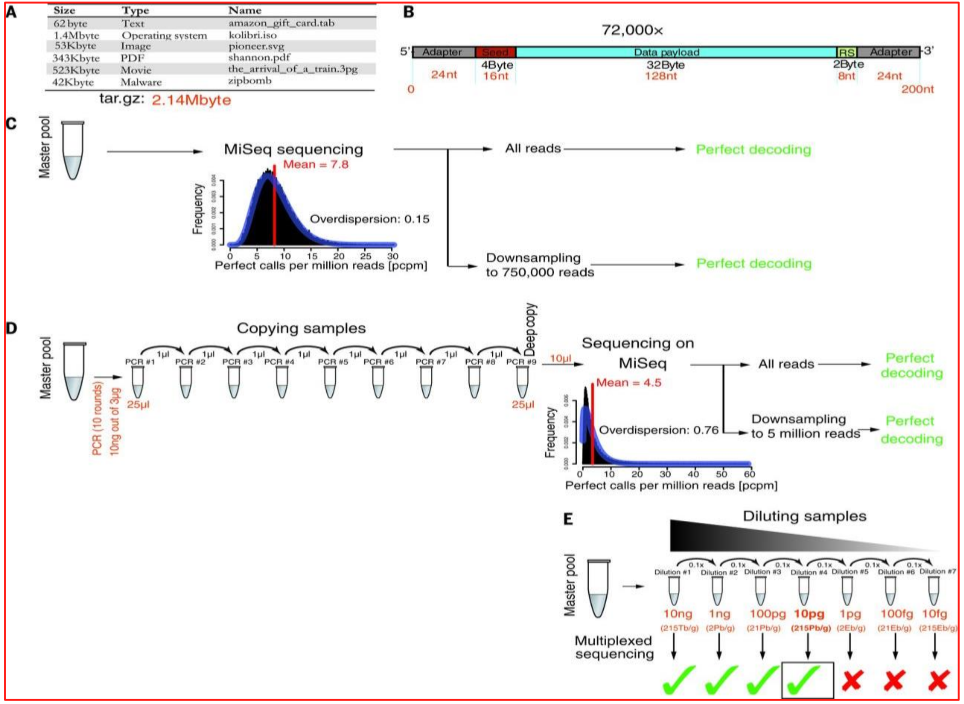
One-third of the 16 chromosomes in the *Saccharomyces cerevisiae* yeast genome have been designed and synthesized

Now fully completed and started a new SC3.0 project. 2020

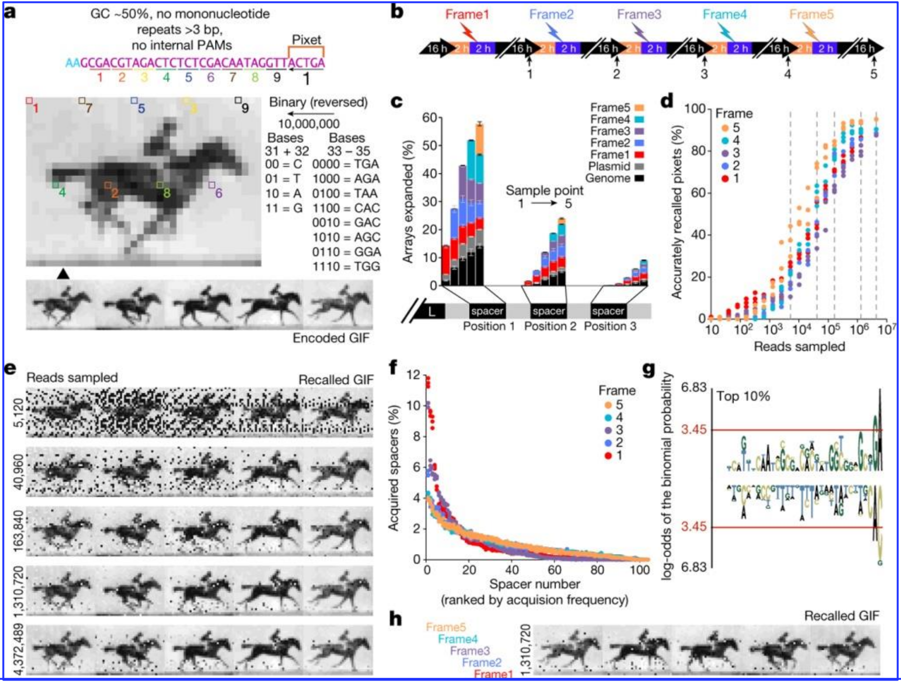
Illustration of a hypothetical yeast genome structure model encompassing all synthetic chromosomes (gold) completed to date (white, native chromosomes). **One-third of the chromosomes in the yeast genome have now been designed and synthesized** by the Synthetic Yeast Genome Project (Sc2.0). A three-dimensional model of the chromosomes was generated with the Hi-C method. An "envelope" (thick tube shapes) represents the population of chromosome-interacting molecules; translucent tips reveal a 30-nanometer fiber modeled inside.

Other applications involving DNA in storing data

Experimental setting for storing data on DNA



CRISPR–Cas encoding of a digital movie into the genomes of a population of living bacteria



Yaniv Erlich, and Dina Zielinski Science 2017;355:950-954
DOI: 10.1126/science.aaj2038

Shipman SL et al. *Nature* 547, 345–349 (2017).
http://dx.doi.org/10.1038/nature23017

Conclusions

- **Genomics and Bioinformatics full disciplines, self-improving with successive new advances.**

Thanks to the increasing complementarity between Mathematics, Statistics, Computer science and Biology.

Perspectives

=> Create opportunities for Young researchers to develop the necessary skills and excel in Genomics and Bioinformatics

- **Universities should fully integrate Bioinformatics and Genomics into their curriculum.**

- **Create a favourable environment to foster research development in these domains:**

Genome Research Institute!

Set up of a “1000 Tunisian Genome Project”!

- Tunisian population is known by its cultural diversity and origin.

Données cartographiques © 2020 Google

- Central position in the Mediterranean area.

- The structure of the Tunisian population genetic diversity has to be investigated!

Mediterranean Genome Project!

Given the historical depth of the Mediterranean people, involving many successive waves of exchange between European and African,
it would be an amazing project to consider whole genome sequencing of humans from all around the Mediterranean Sea.

It is up to motivated local researchers and Funding agencies to enrich scientific knowledge by enlightening on human genomes in this area.

Comment on : Choudhury, A., Aron, S., Botigué, L.R. *et al.* High-depth African genomes inform human migration and health. *Nature* **586,** 741–748 (2020). https://doi.org/10.1038/s41586-020-2859-7
Fredj Tekaia, Abdellatif Bodabous

Acknowledgements

Professor Habiba Bouhamed Chaabouni
Directrice du département des sciences de l’Académie Tunisienne des Sciences des Lettres et des Arts, Beit al-Hikma, Tunis.

Madame Thouraya Lasram
Directrice Administrative de l’Académie
Coordinatrice générale des affaires scientifiques et de la coopération internationale

Monsieur Nizar Yakoubi
Technicien Informatique à l’Académie
For his help in setting up the technical communication.

Professor Abdellatif Boudabous
Fac. Des Sciences, Université Tunis El Manar

Stay up to date with the scientific literature

➔ Young Researchers should keep up to date with the Scientific literature

➔ Follow up of the main scientific journals (Thanks to the Open Access Initiatives)

➔ Sign up for TOCs (Table Of Contents)

Follow up of the main scientific journals

- **Nature:** http://www.nature.com/
- **Nature Genetics** http://www.nature.com/nrg/index.html
- **Sciences:** http://www.sciencemag.org/magazine
- **Genome Research:** http://genome.cshlp.org/
- **Plos Biology:** https://journals.plos.org/plosbiology/
- **Plos Computational Biology:** https://journals.plos.org/ploscompbiol/
- **BMC Genomics:** https://bmcgenomics.biomedcentral.com
- **Current Biology:** http://www.cell.com/current-biology/home
- **Daily follow up of literature about microbiome:** https://microbiomedigest.com
- **PubMed:** http://www.ncbi.nlm.nih.gov/pubmed/

References

- **Large-scale Genome Analyses**

http://webext.pasteur.fr/tekaia/BCGAIPT2018/TALKS/Tekaia_BCGAIPT2018_LargeScaleGenomeAnal.pptx

- **Bioinformatics & Genome Analyses course**

https://webext.pasteur.fr/tekaia/BGA_courses.html
https://webext.pasteur.fr/tekaia/BCGA_WProgs.html

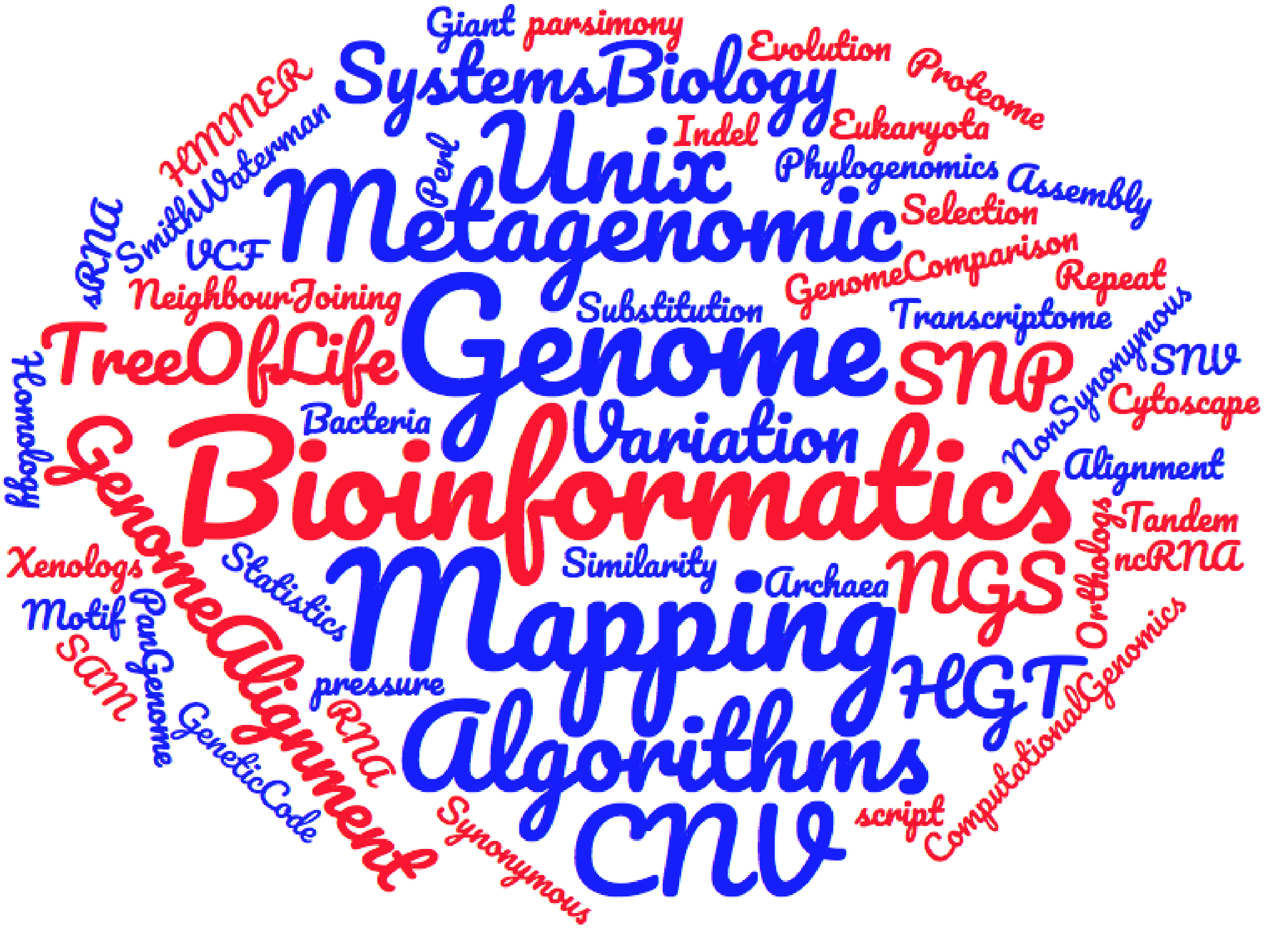
- **Designing and running an advanced Bioinformatics and genome analyses course in Tunisia.**

PLoS Comput Biol 15(1): e1006373.
https://doi.org/10.1371/journal.pcbi.1006373

- **Enhancing Bioinformatics and Genomics Courses: Building Capacity and Skills via Lab Meeting Activities.**

BioEssays. 2020. 42(10):e2000134. doi:10.1002/bies.202000134.

- **https://webext.pasteur.fr/tekaia/tekaia.publications.html**


Giant
parsimony
Evolution
Proteome
HMMER
SystemsBiology
SmithWaterman
Indel
Eukaryota
Perl
Unix
Phylogenomics
Assembly
sRNA
Metagenomic
Selection
VCF
GenomeComparison
Repeat
NeighbourJoining
Substitution
Transcriptome
NonSynonymous
TreeOfLife
Genome
SNP
SNV
Homology
Bacteria
Variation
Cytoscape
GenomeAlignment
Bioinformatics
Alignment
Tandem
Orthologs
Xenologs
Statistics
Similarity
Archaea
NGS
ncRNA
Motif
PanGenome
Mapping
SAM
pressure
HGT
ComputationalGenomics
GeneticCode
RNA
Algorithms
script
Synonymous
CNV

Thank You