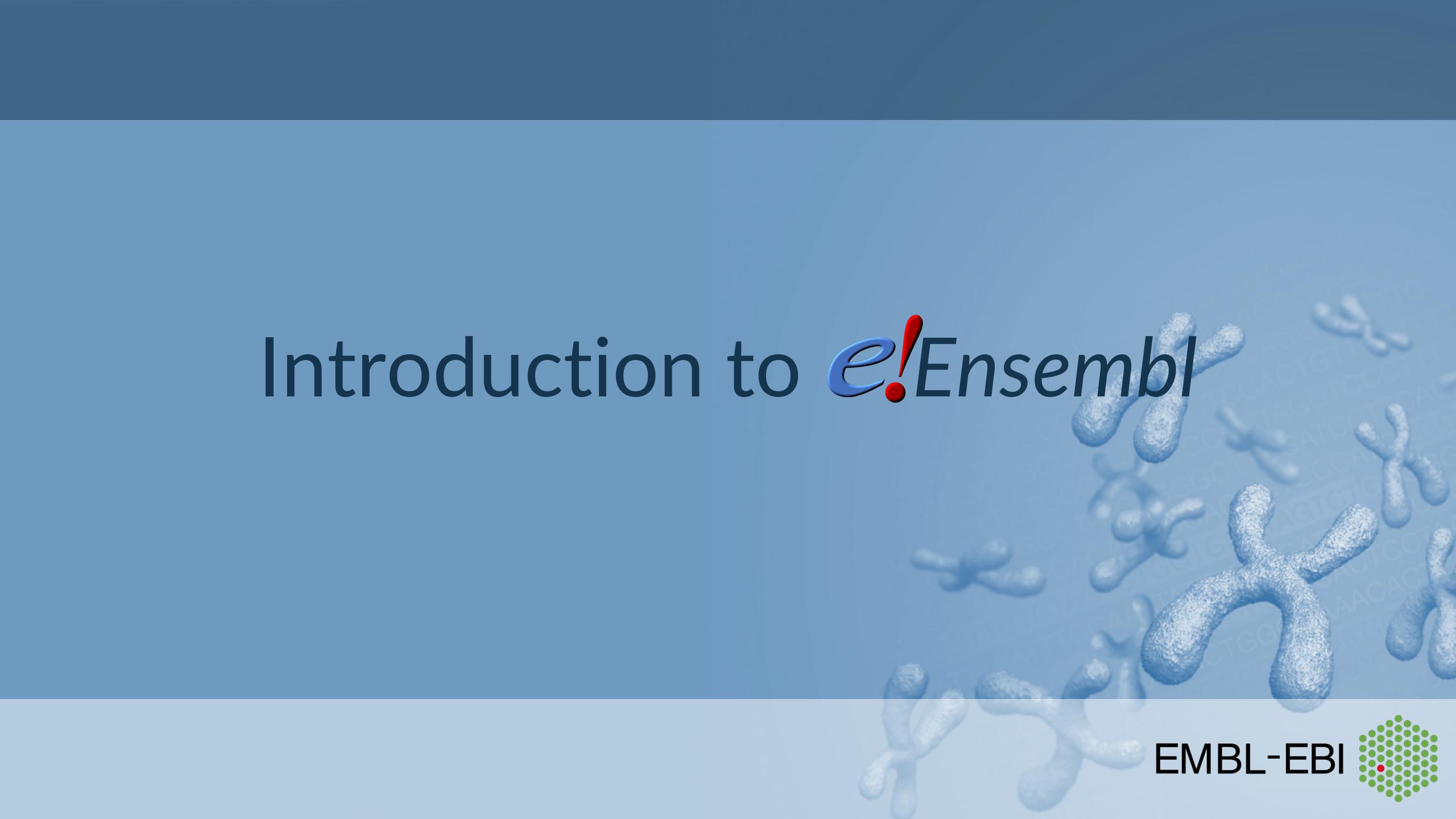


# Introduction to *e!Ensembl*



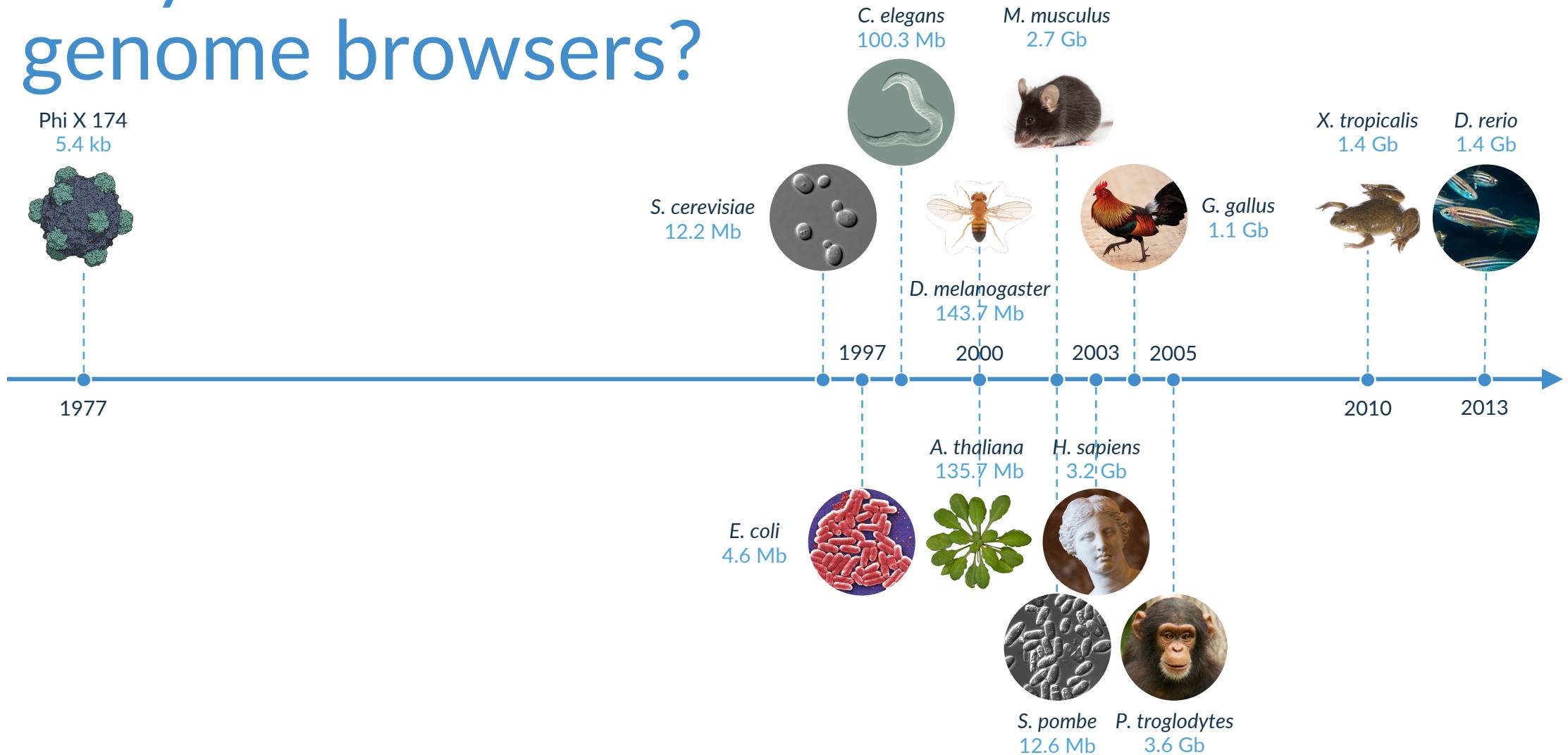
# Training materials



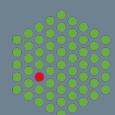
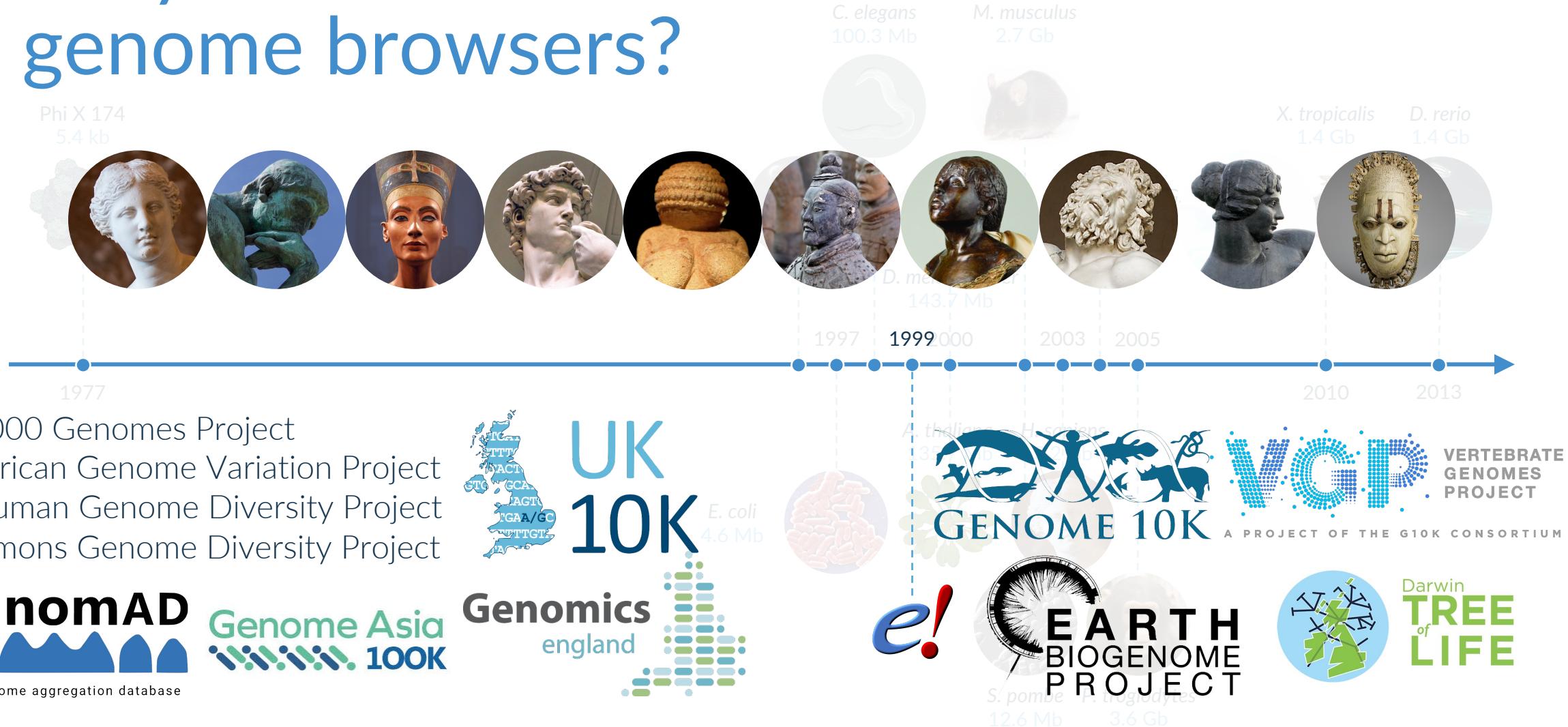
- Ensembl training materials are protected by a CC BY license:  
[creativecommons.org/licenses/by/4.0/](http://creativecommons.org/licenses/by/4.0/)
- If you wish to re-use these materials, please credit Ensembl for their creation
- If you use Ensembl for your work, please cite our papers:  
[ensembl.org/info/about/publications.html](http://ensembl.org/info/about/publications.html)

# Why do we need genome browsers?

# Why do we need genome browsers?



# Why do we need genome browsers?



# What is Ensembl?

- Ensembl maps genomic features to genome sequences and performs analysis on these data
- Ensembl is an added value resource that brings together information from wide range of other databases in a single site

# What is Ensembl?



- Ensembl maps genomic features to genome sequences and performs analysis on these data
- Ensembl is an added value resource that brings together information from wide range of other databases in a single site



Genome Browser

[genome.ucsc.edu](http://genome.ucsc.edu)



Genome Data Viewer

NCBI

[ncbi.nlm.nih.gov/genome/gdv](http://ncbi.nlm.nih.gov/genome/gdv)



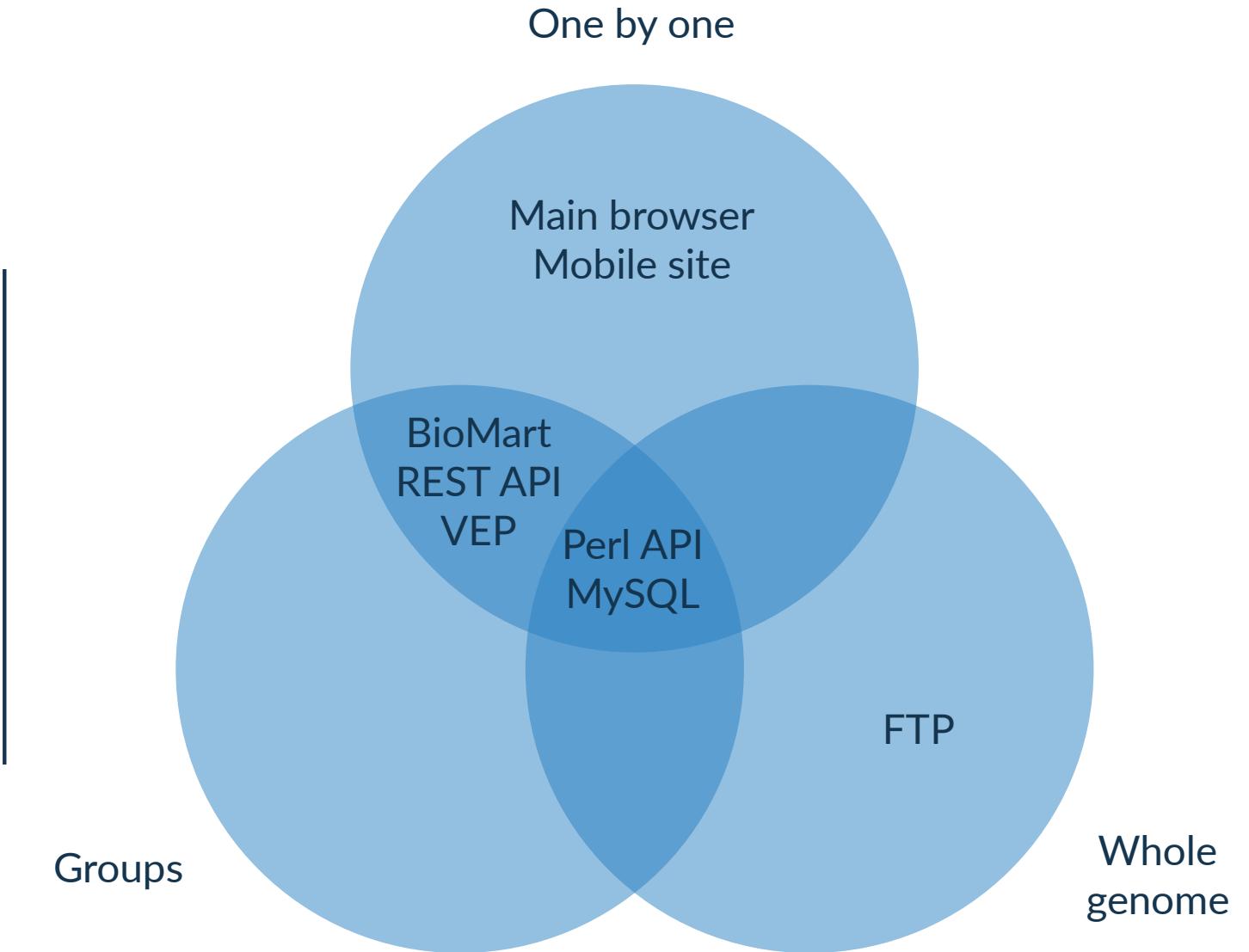
EMBL-EBI



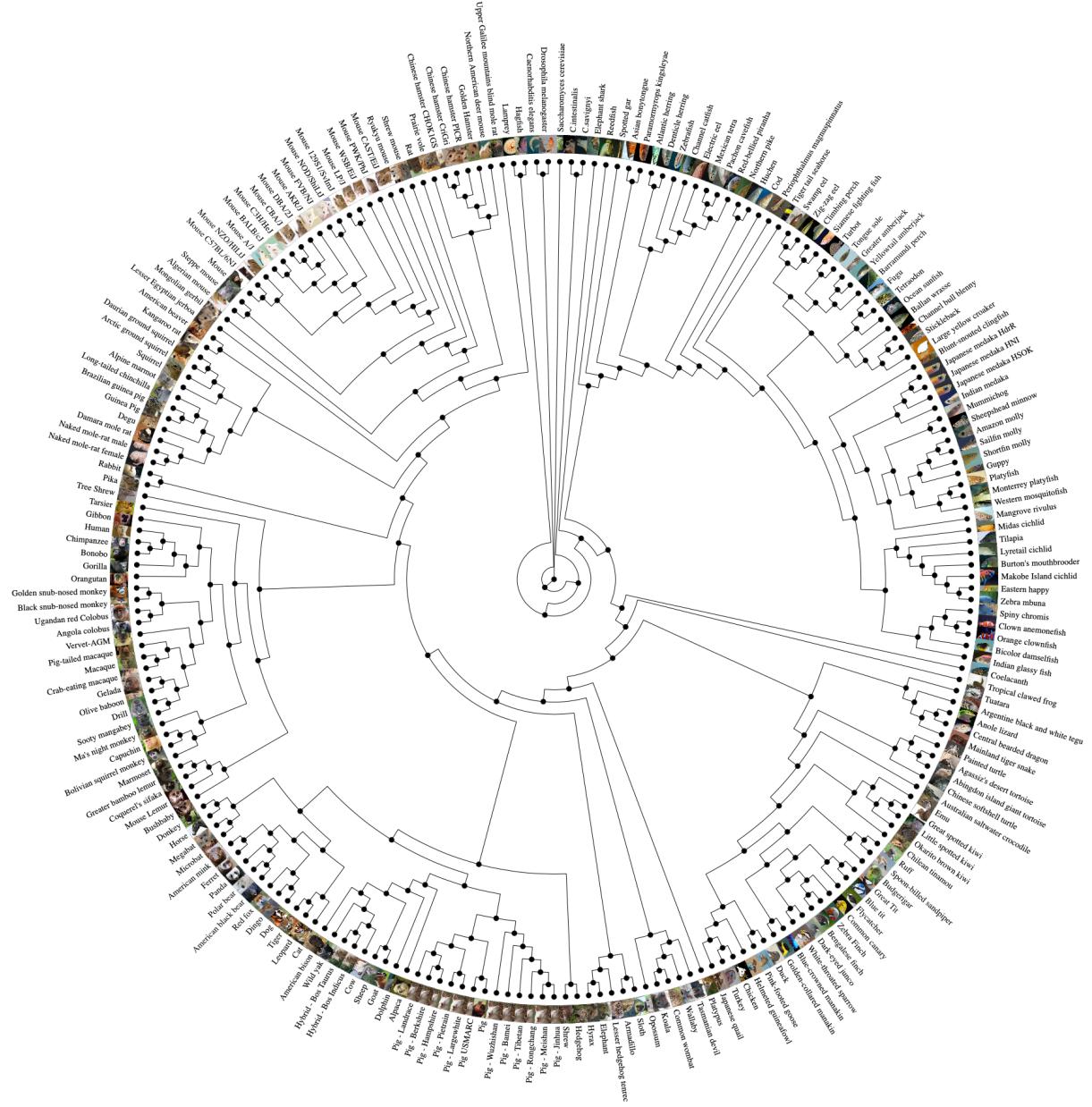
# Ensembl features

- Genomes and gene builds for >300 species
- Variation data
- Compara (alignments, gene trees, homologues)
- Regulatory build
- BioMart (data export)
- Tools for data processing, e.g. VEP
- Display your own data
- Programmatic access via APIs
- Completely Open Source (FTP, GitHub)

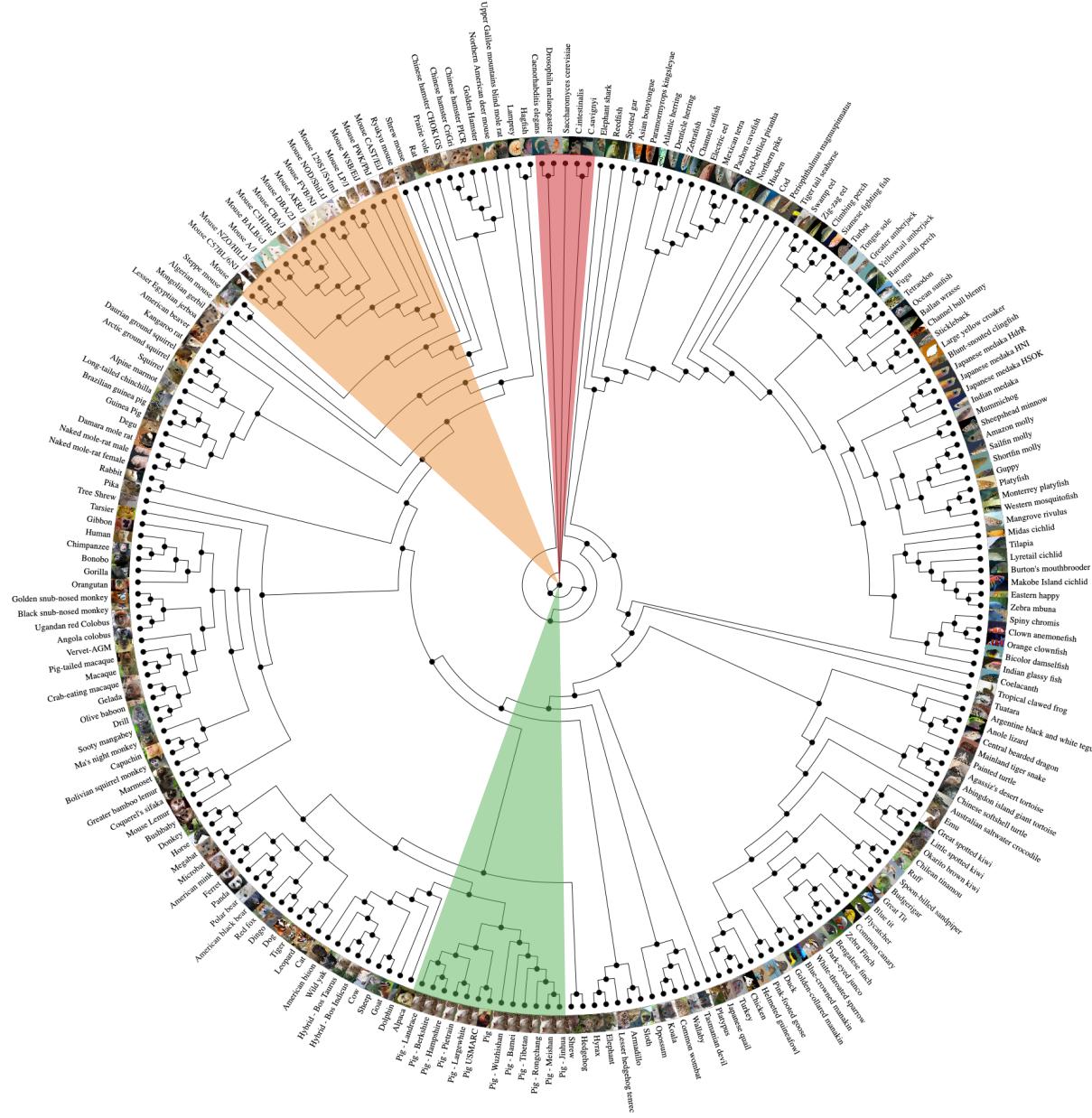
# Access scales



# Ensembl: vertebrates and model organisms



# Ensembl: vertebrates and model organisms





# Ensembl Genomes

[ensemblgenomes.org](http://ensemblgenomes.org)

- **e!EnsemblBacteria**  
31,332 genomes
- **e!EnsemblFungi**  
1,014 genomes
- **e!EnsemblMetazoa**  
123 genomes
- **e!EnsemblPlants**  
110 genomes
- **e!EnsemblProtists**  
237 genomes



# Ensembl and Ensembl Genomes

	Ensembl <a href="http://ensembl.org">ensembl.org</a>	Ensembl Genomes <a href="http://ensemblgenomes.org">ensemblgenomes.org</a>
Launched	2000	2009
Species	Vertebrates (and outgroups)	Non-vertebrates (bacteria, fungi, protists, plants, metazoa)
Annotation	By Ensembl	In collaboration with the scientific communities

# Genome assembly

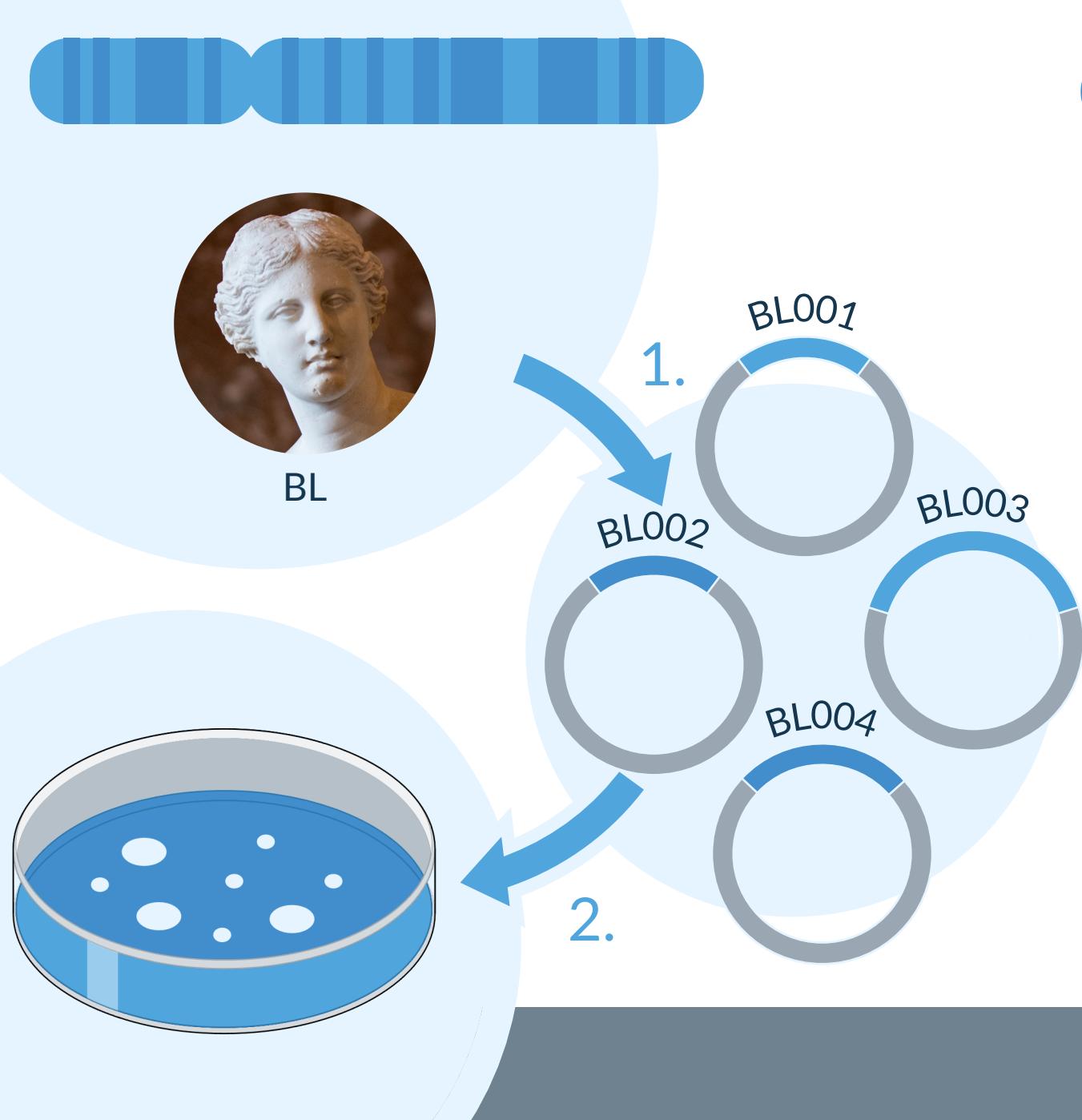


Genome: "DNA within a cell"



Genome assembly: representation of a genome

# Genome assembly: cloning into BACs



# Genome assembly: making a contig

- Sequence reads:

CAGCTGTCCCAGATGAC AACTTCCCTCCCAGCT TCCGCCCTTCAGCTCAAGACTTAACCTC  
CGGCCTTTGGGCTCC ACTTAACCTTCCCTCCCAGCTGTCC TCCCAGCTGTCCCAGATGACGCCATC  
CAGATGACGCC CGGCCTTTGGGCTCCGCCCTTCAGCTCAAGA GGGCTCCGCCCTTCAGCTC

- Match up overlaps:

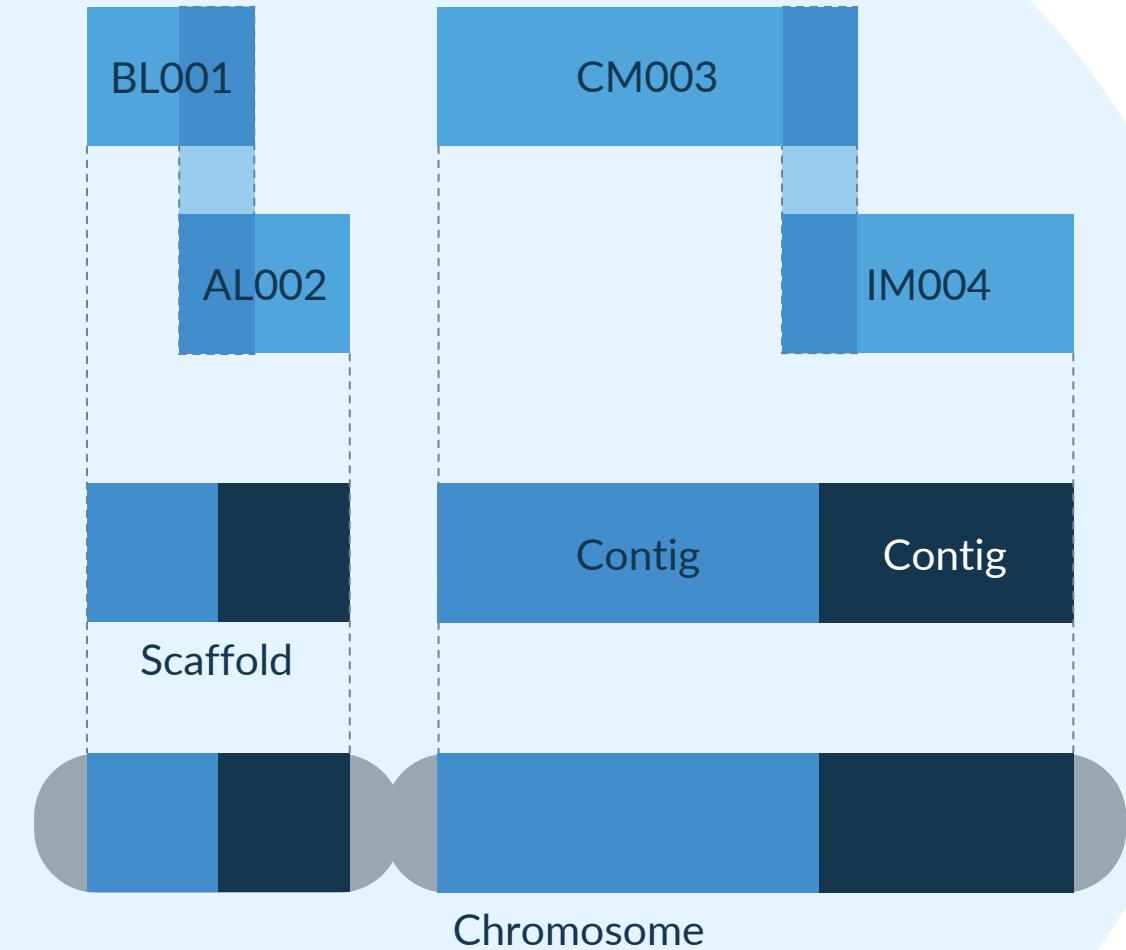
CGGCCTTTGGGCTCCGCCCTTCAGCTCAAGA AACTTCCCTCCCAGCT CAGATGACGCC  
TCCGCCCTTCAGCTCAAGACTTAACCTC TCCCAGCTGTCCCAGATGACGCCATC  
GGGCTCCGCCCTTCAGCTC ACTTAACCTTCCCTCCCAGCTGTCC  
CGGCCTTTGGGCTCC CAGCTCCCAGATGAC

- Contig:

CGGCCTTTGGGCTCCGCCCTTCAGCTCAAGACTTAACCTTCCCTCCCAGCTGTCCCAGATGACGCCATC

# Genome assembly: contig to scaffold

1. **Tilepath:** BACs from different individuals assembled together with overlaps
2. Overlaps trimmed to give **contigs**. A run of contigs with no gaps is a **scaffold**
3. Genetic maps used to assemble scaffolds into a **chromosome**



# nature

the  
**human**  
genome

£5.45 \$8.29 Fr54 DM15 Lire15000

www.nature.com

## Nuclear fission

Five-dimensional energy landscapes

## Seafloor spreading

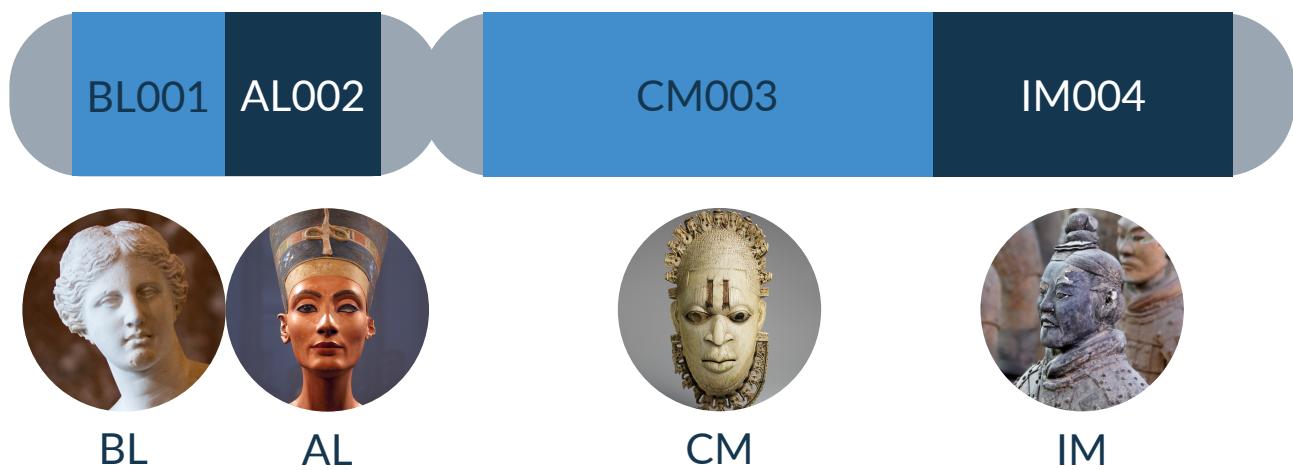
The view from under the Arctic ice

## Career prospects

Sequence creates new opportunities

## Genome assembly

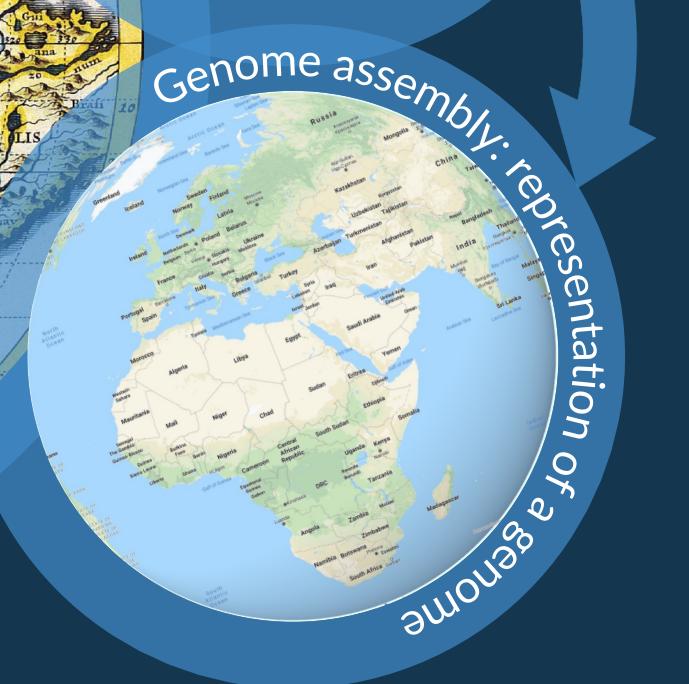
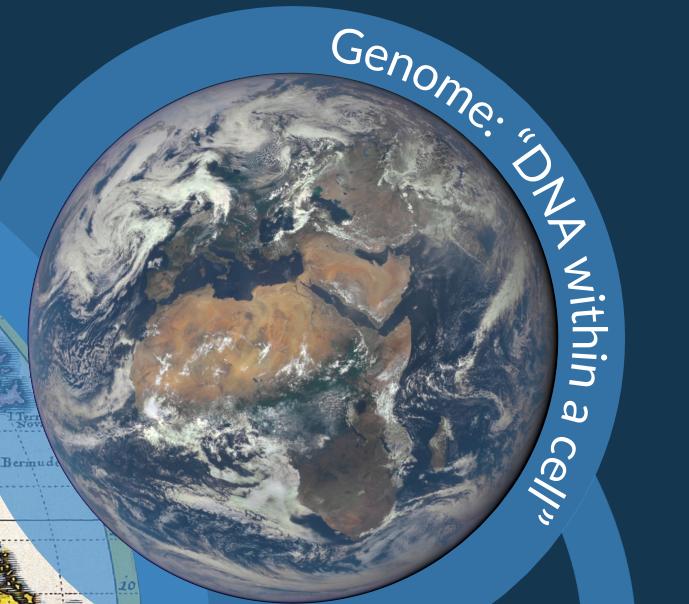
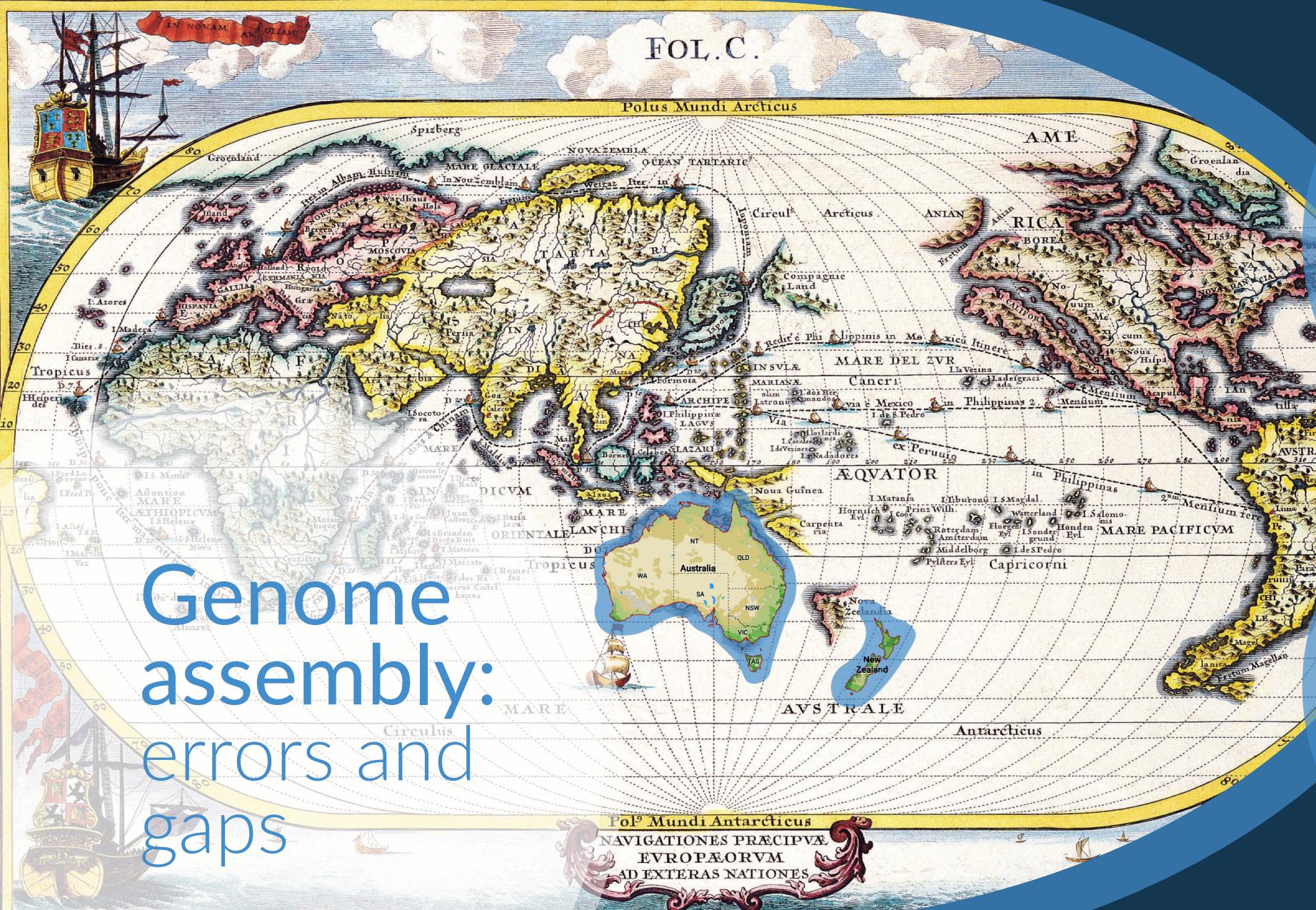
- The human **reference genome** is not a genome of a single person



# Genome assembly: errors and gaps



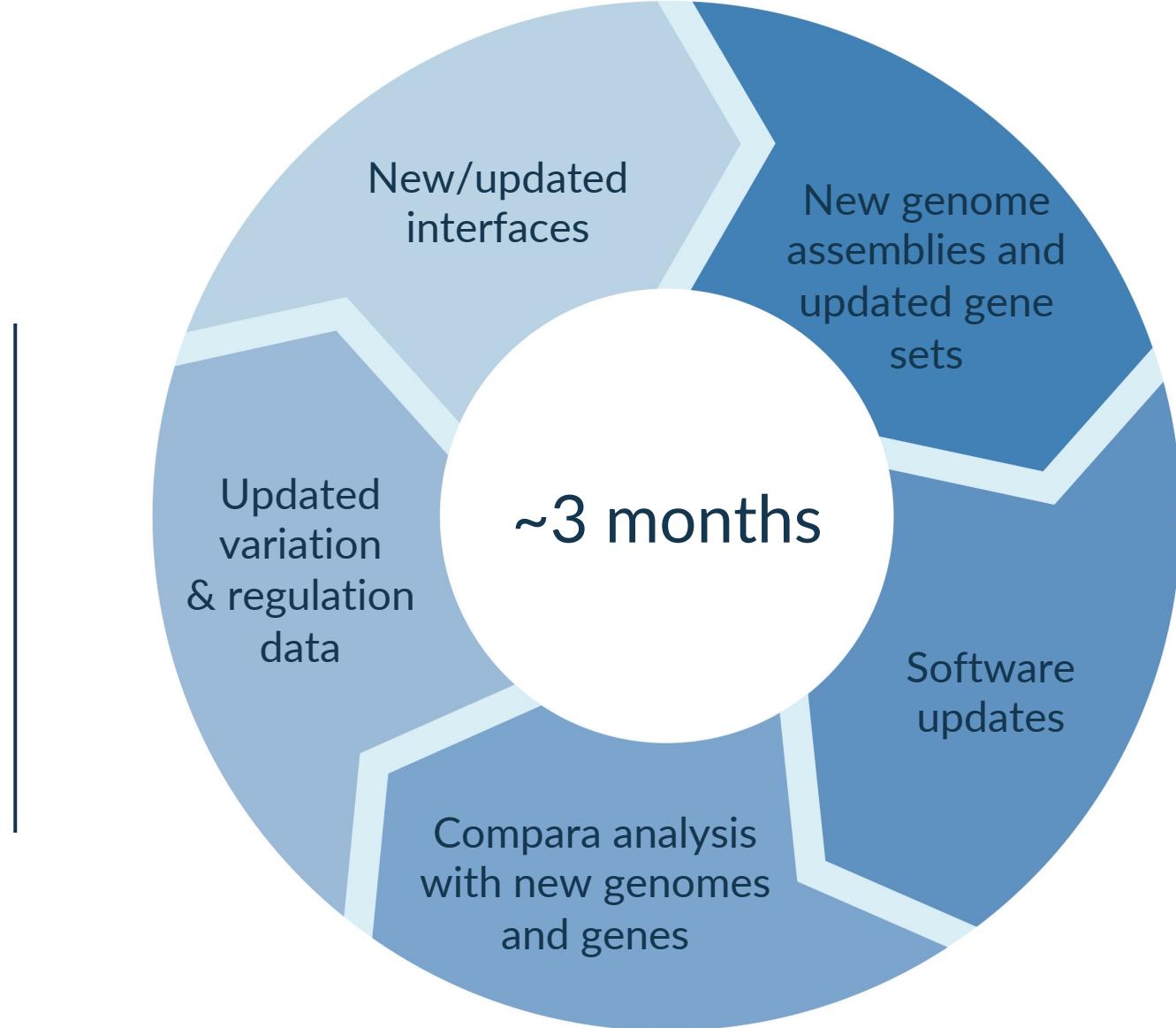
# Genome assembly: errors and gaps



# Human genome assemblies

- GRCh38 (aka hg38) 
  - Shorter gap length. Many rare/private alleles replaced
  - [ensembl.org](http://ensembl.org)
  - Most up-to-date and supported
- GRCh37 (aka hg19) 
  - Some large gaps
  - [grch37.ensembl.org](http://grch37.ensembl.org)
  - Limited data and software updates
- NCBI36 (aka hg18) 
  - Many gaps
  - [ncbi36.ensembl.org](http://ncbi36.ensembl.org)
  - No longer updated

# Release cycle: **e!Ensembl 104** **e!EnsemblGenomes 51** May 2021



# Ensembl Rapid Release

[rapid.ensembl.org](http://rapid.ensembl.org)

- New genomes & gene annotation ✓
- Released every two weeks ✓
- >200 species ✓
- BLAST ✓
- FTP ✓
- No BioMart ✗
- No Compara ✗
- No variation data ✗