Comparative Genomics and Visualisation
BS32010
1.Introduction

The James Hutton Institute

Leighton Pritchard[1,2,3]
[1]Information and Computational Sciences,
[2]Centre for Human and Animal Pathogens in the Environment,
[3]Dundee Effector Consortium,
The James Hutton Institute, Invergowrie, Dundee, Scotland, DD2 5DA

Acceptable Use Policy

The James Hutton Institute

Recording of this talk, taking photos, discussing the content using email, Twitter, blogs, etc. is permitted (and encouraged), providing distraction to others is minimised.

These slides will be made available on SlideShare.

These slides, and supporting material including exercises, are available at https://github.com/widdowquinn/Teaching-2015-03-17-UoD_compgenvis

Table of Contents

The James Hutton Institute

What Is Comparative Genomics?

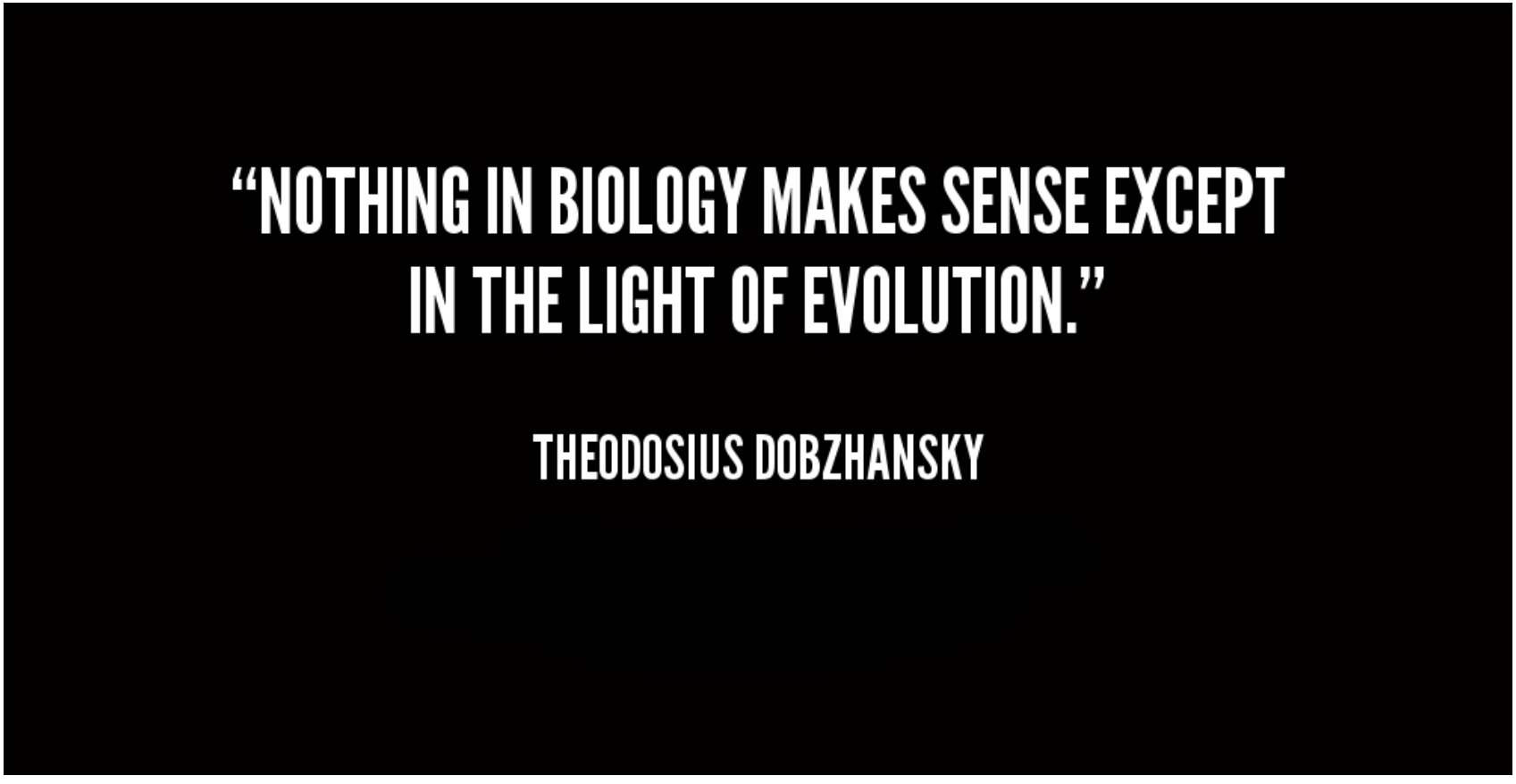
The James Hutton Institute

Comparative genomics is...

The combination of genomic data, and comparative and evolutionary biology, to address questions of genome structure, evolution, and function.

Evolution is the central concept

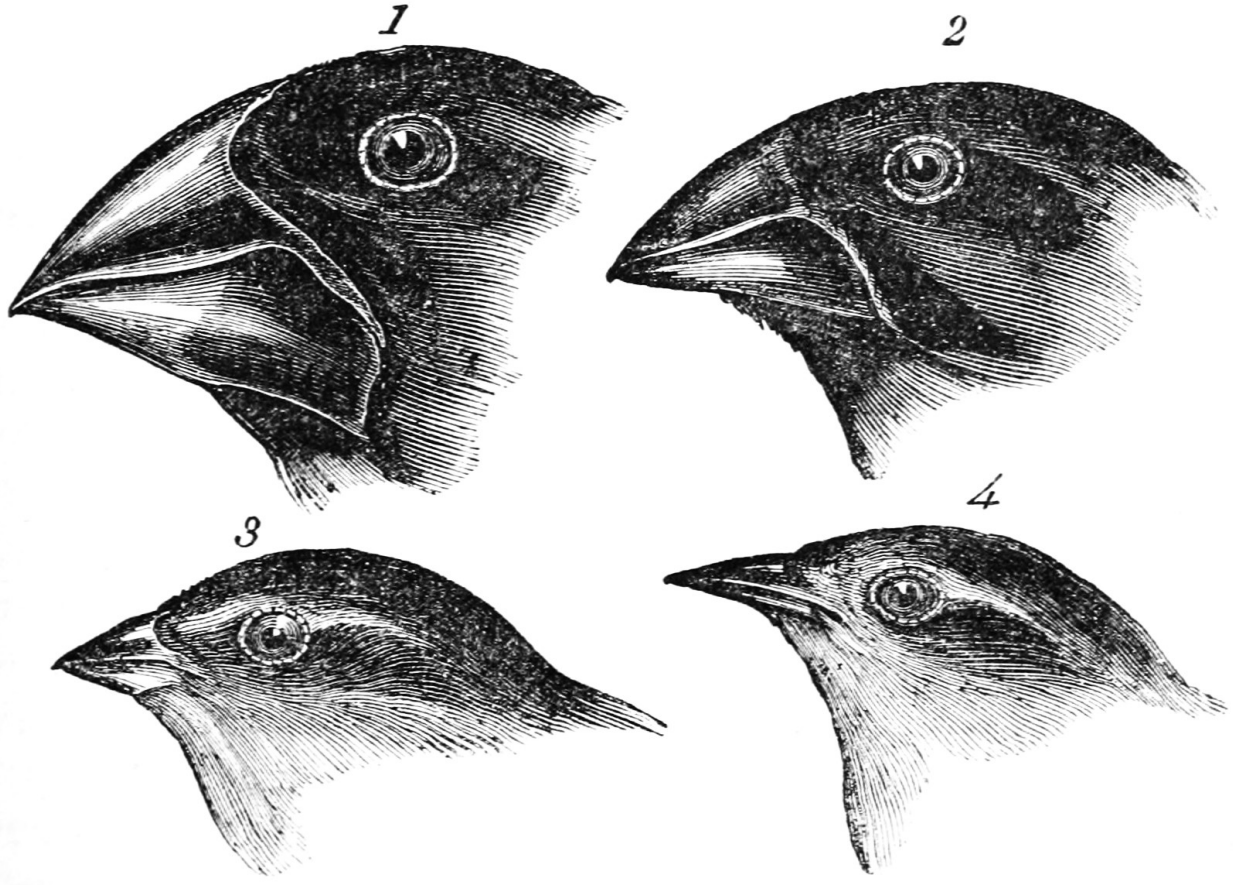
The James Hutton Institute



Comparison of physical features

The James Hutton Institute

How do we determine that features are related, and evolved?

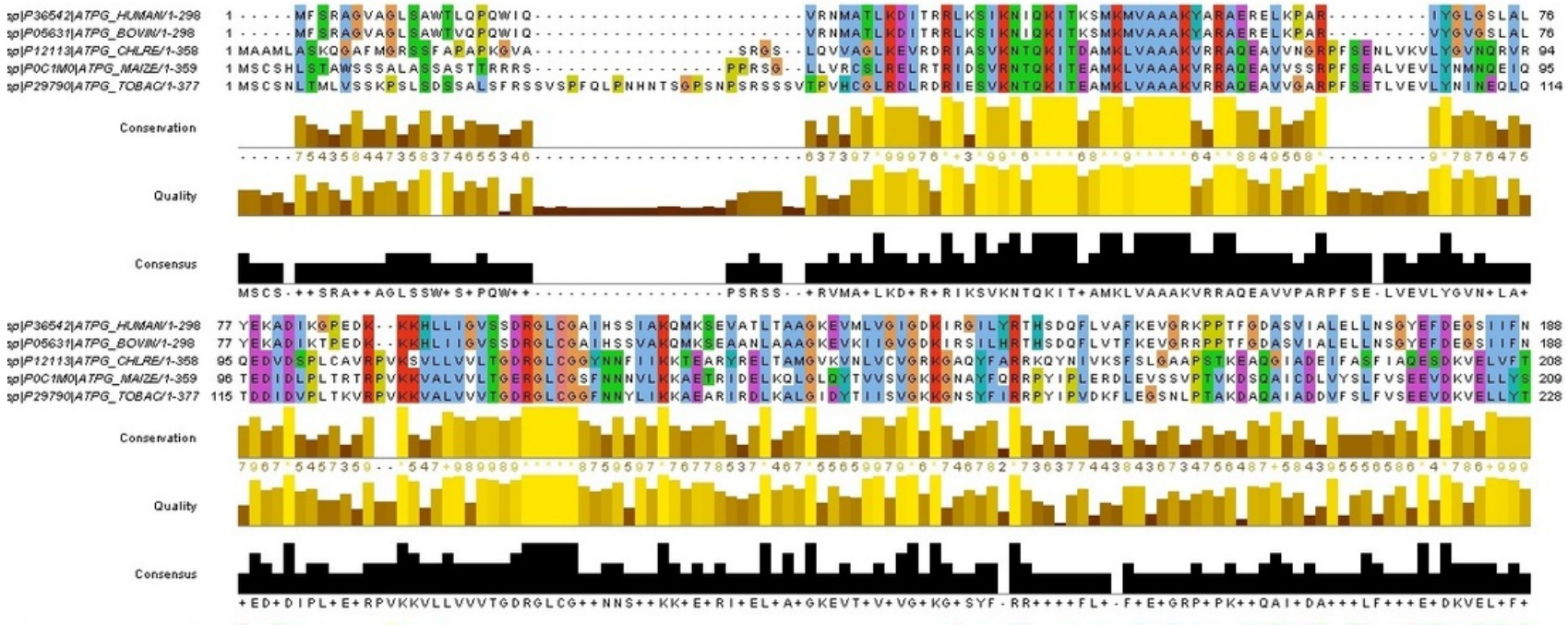


1. Geospiza magnirostris.
2. Geospiza fortis.
3. Geospiza parvula.
4. Certhidea olivasea.

Comparison of sequence features

The James Hutton Institute

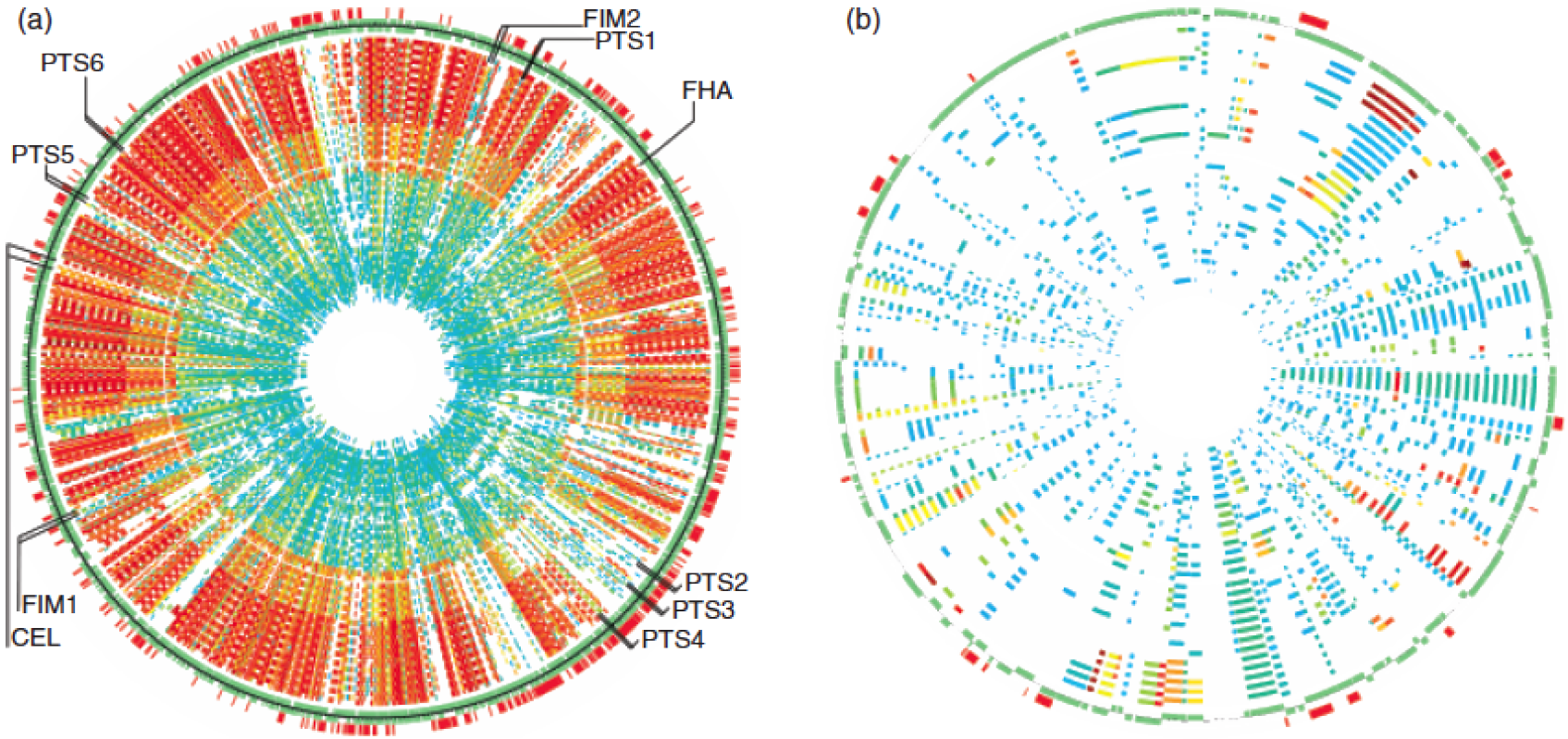
Multiple alignment of ATP synthase



Comparison of genome features

The James Hutton Institute

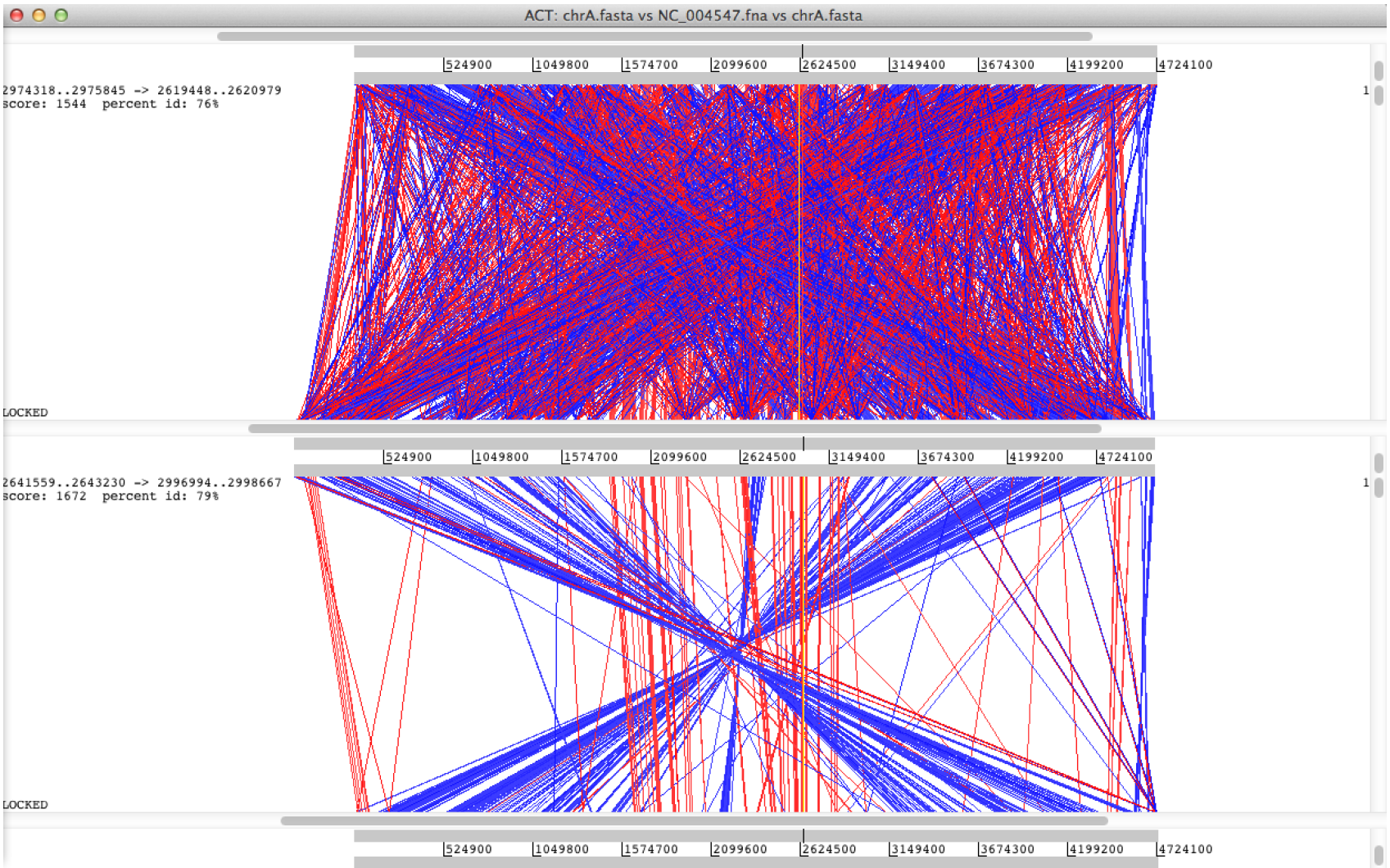
Sequence similarity of individual features



Comparison of genome features

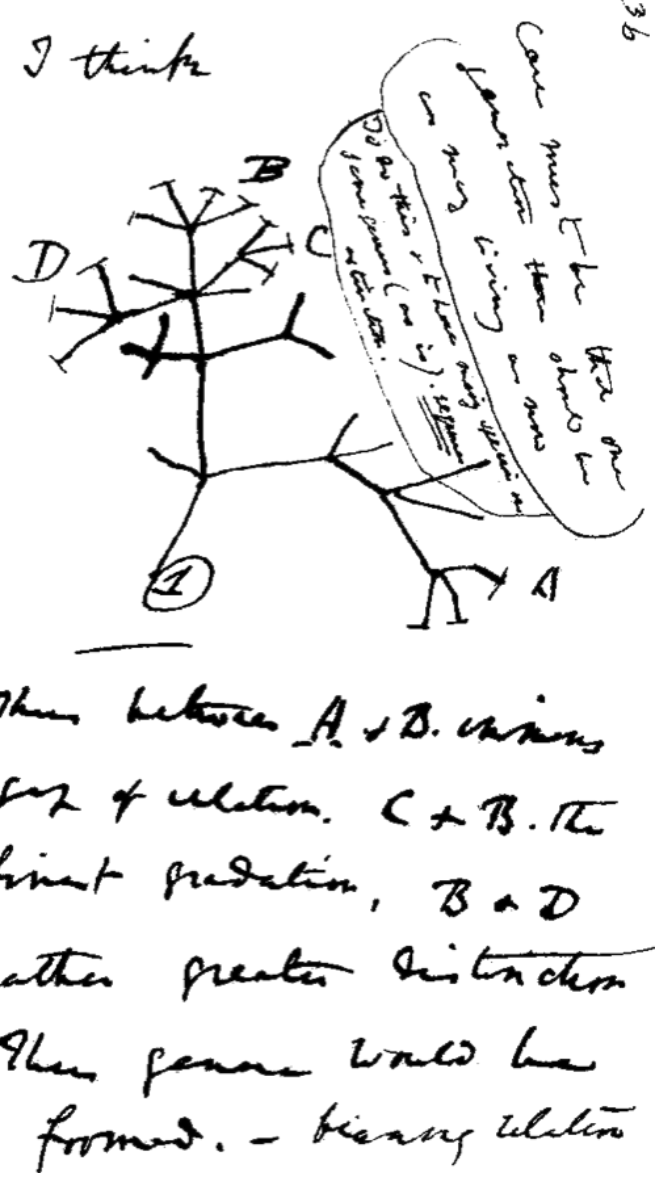
The James Hutton Institute

Genome structural rearrangements



Why comparative genomics?

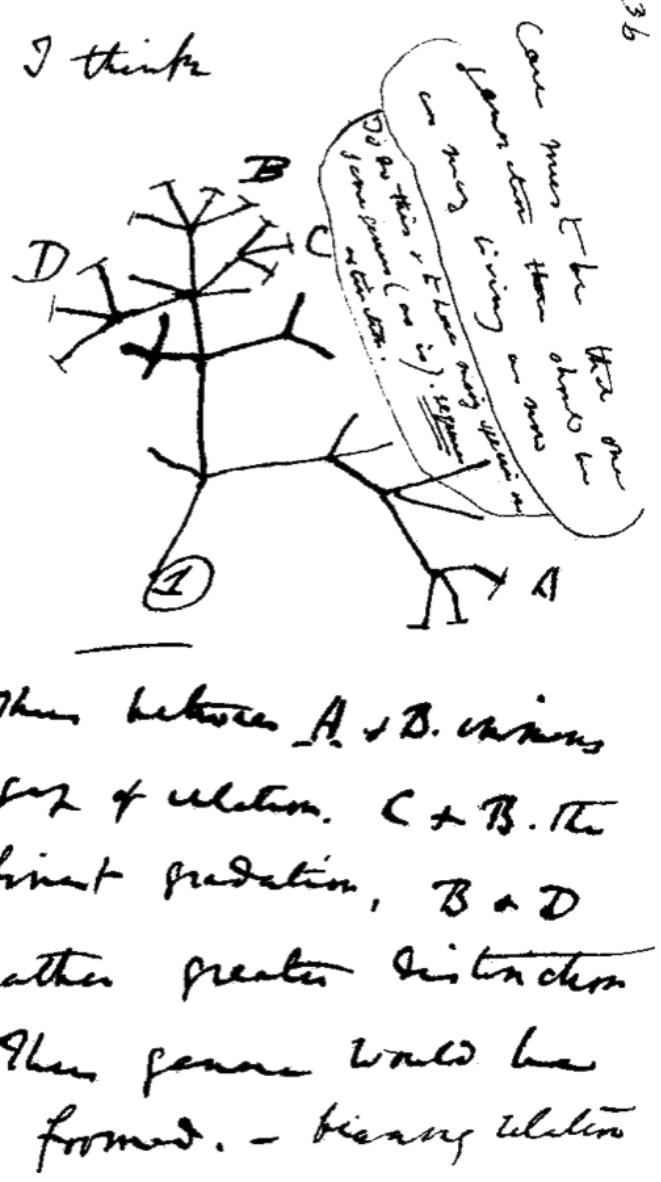
- Genomes describe heritable characteristics
- Related organisms share ancestral genomes
- Functional elements encoded in genomes are common to related organisms



Why comparative genomics?

The James Hutton Institute

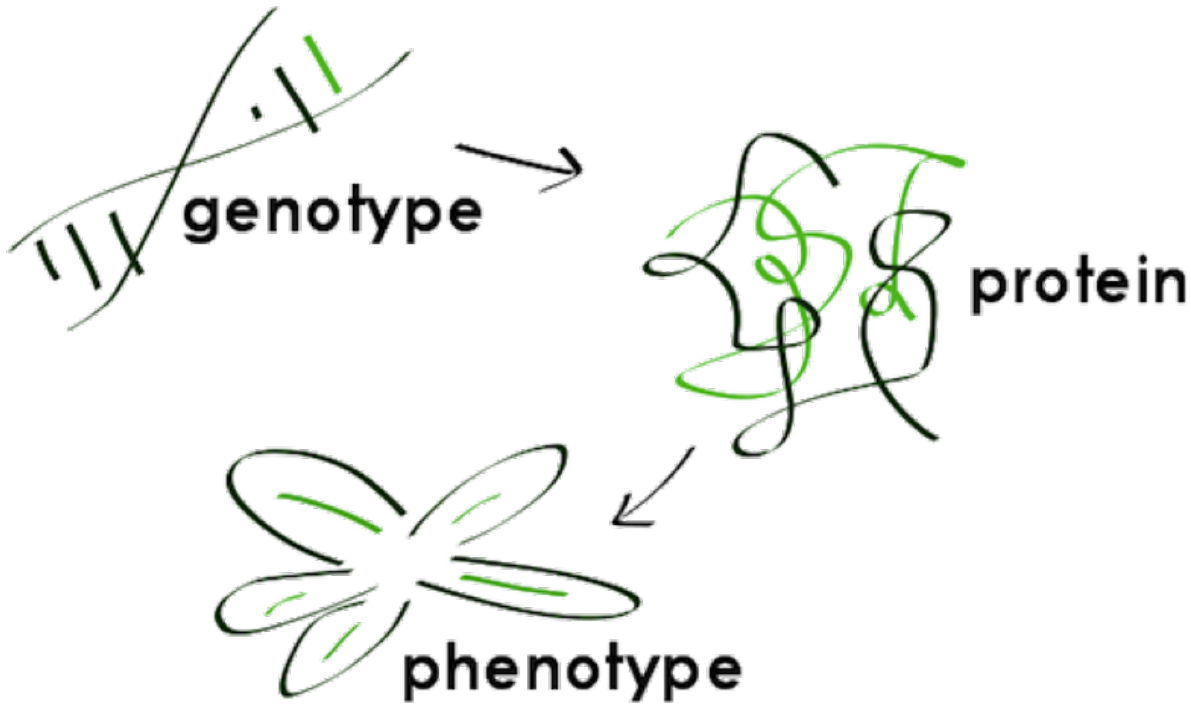
- Transfer of functional understanding from model systems (*E. coli*, *A. thaliana*, *D. melanogaster*) to non-model systems
- Genome comparisons can be informative, even for distantly-related organisms



Genomes are informative, but...

The James Hutton Institute

CONTEXT: epigenetics, tissue differentiation, mesoscale systems, etc.



PHENOTYPIC PLASTICITY: responses to temperature, stress, environment, etc.

Genomes to systems

The James Hutton Institute

Functional Genomics

- Genomic differences can underpin phenotypic (morphological or physiological) differences.
- Where phenotypes/other organism-level properties are known, comparison of genomes can give mechanistic or functional insight into differences (e.g. GWAS).
- Genomic changes reveal evolutionary processes and constraints.

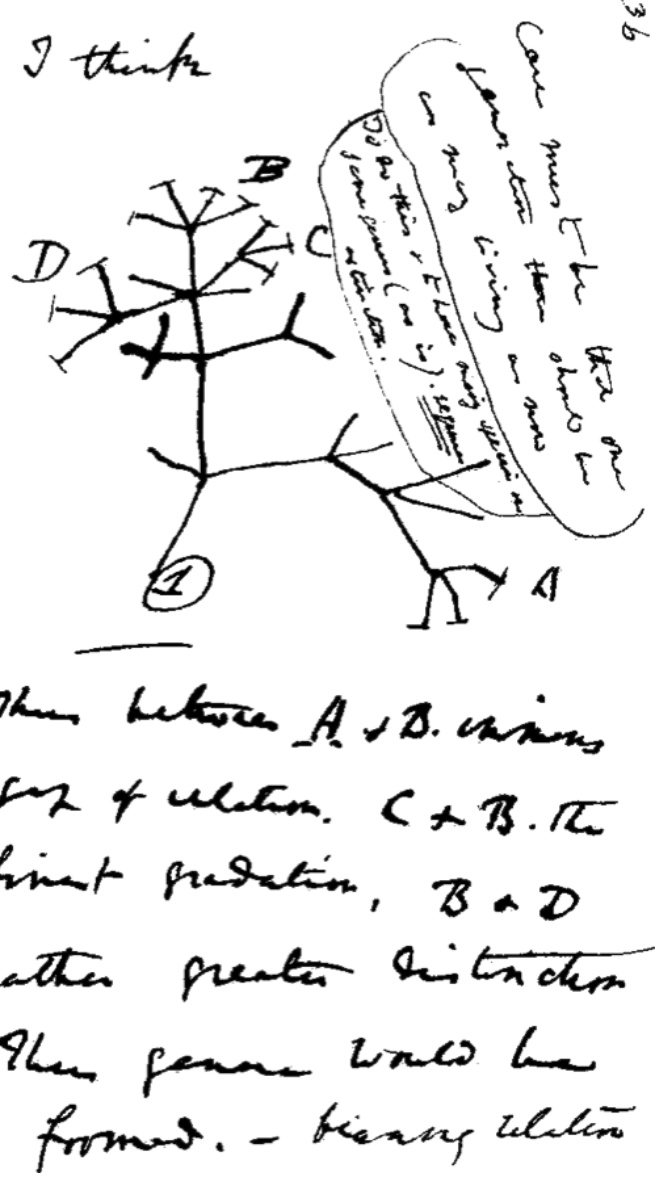


Table of Contents

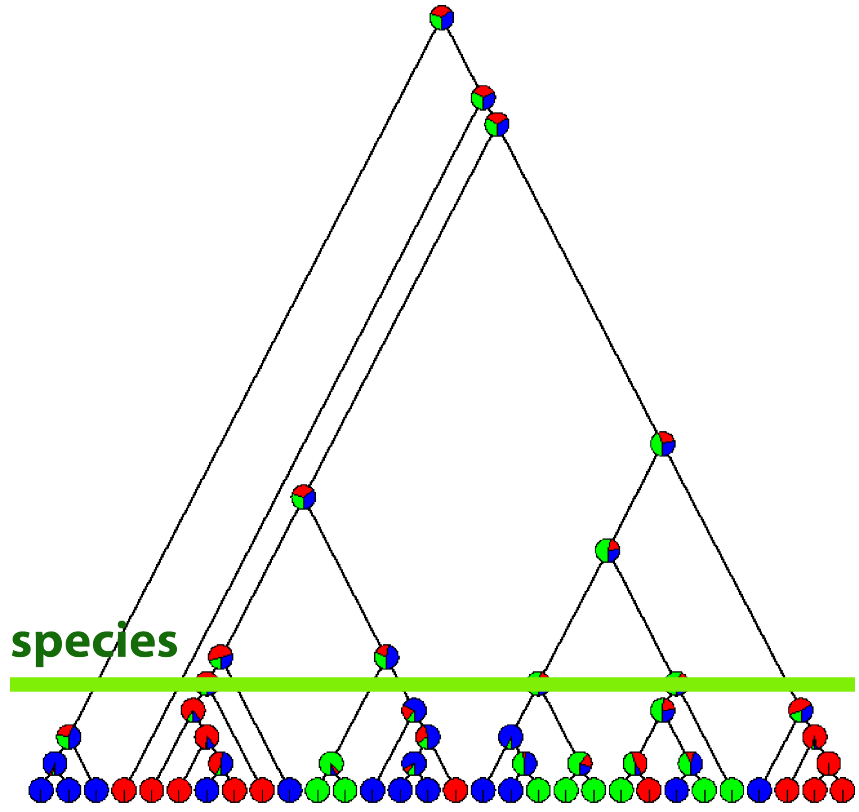
The James Hutton Institute

Types of comparison

The James Hutton Institute

Within species

- e.g. isolate-level (and even within individuals)
- which genome features may account for unique characteristics of organisms/cell-types (e.g. tumours)?
- what epigenetic changes occur in an individual?

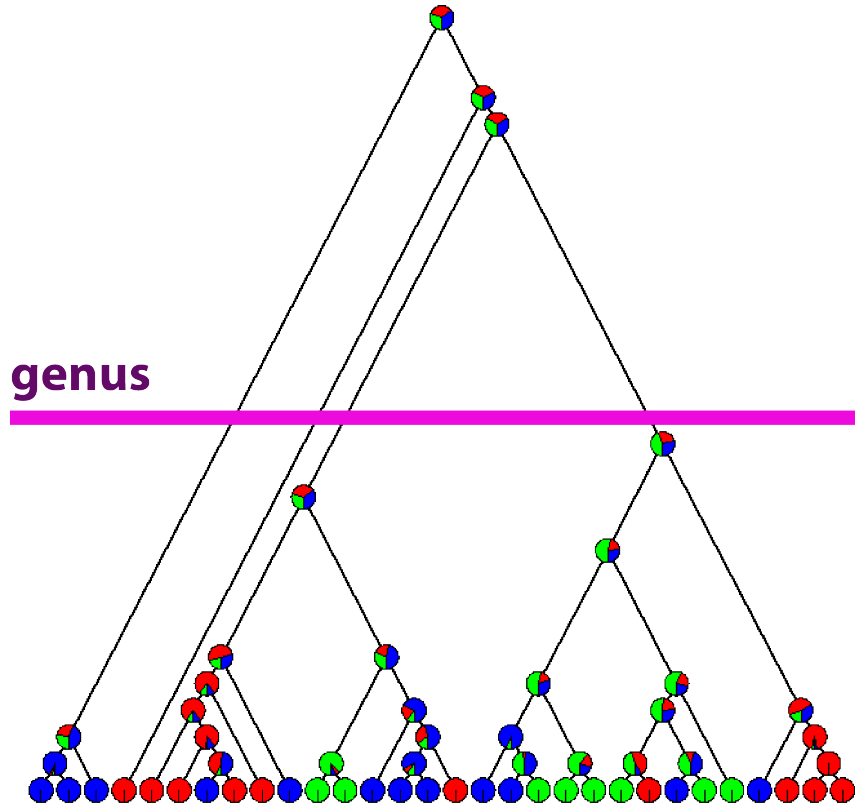


Types of comparison

The James Hutton Institute

Within genera/between species

- what genome features show evidence of selective pressure?
- which species are under selective pressure for which functions?

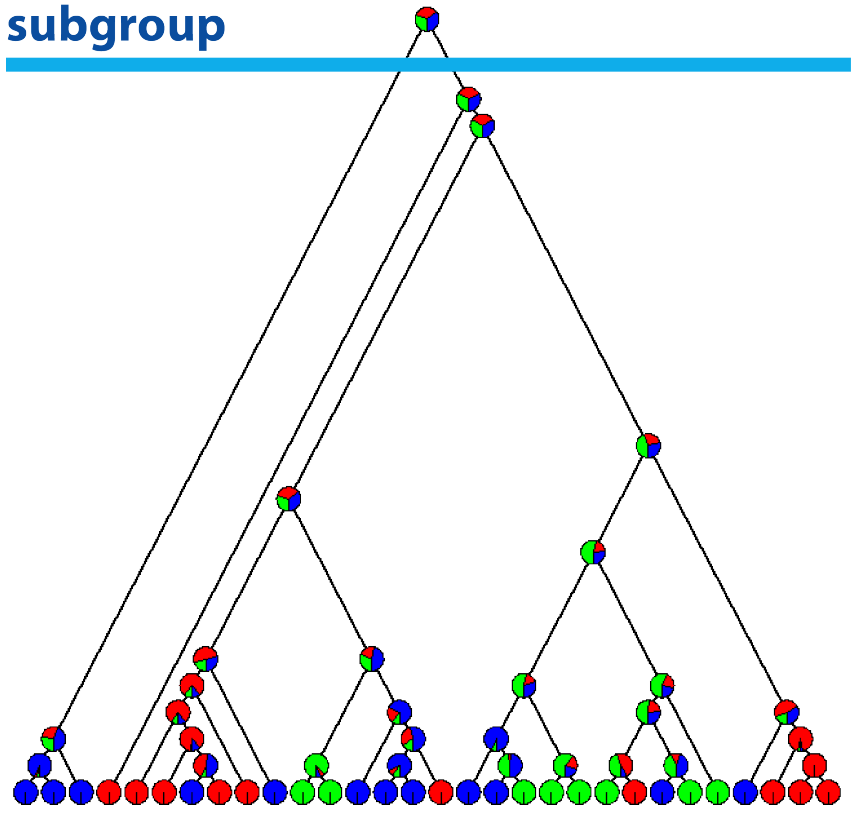


Types of comparison

The James Hutton Institute

Between subgroups

- what are the core set of genome features that define a subgroup or genus?
- what functions are present/absent between groups?



E. coli LTEE [a] [b] [c]

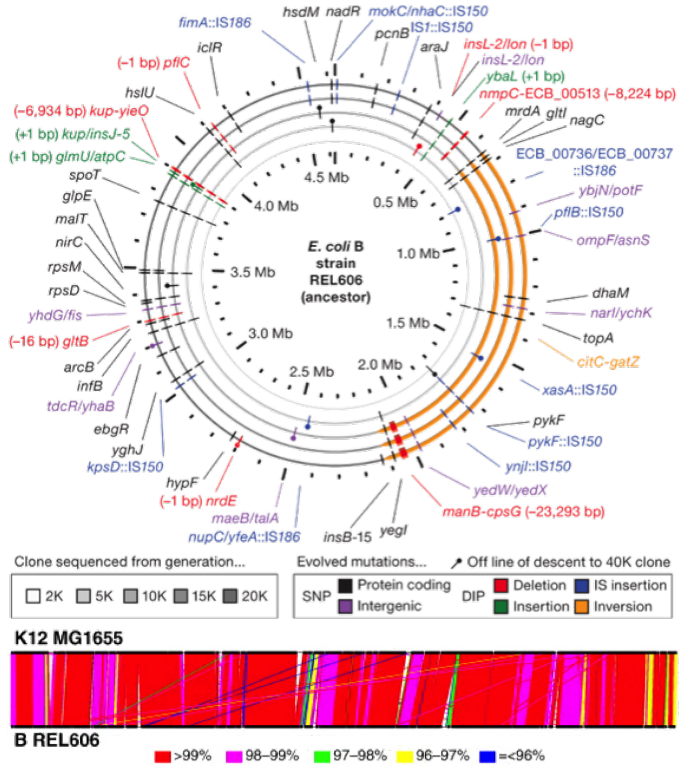
[a] Jeong *et al.* (2009) *J. Mol. Biol.* doi:10.1016/j.jmb.2009.09.052

[b] Barrick *et al.* (2009) *Nature* doi:10.1038/nature08480

[c] Wiser *et al.* (2013) *Science* doi:10.1126/science.1243357

The James Hutton Institute

- Run by the Lenski lab, Michigan State University since 1988 (http://myxo.css.msu.edu/ecoli/)
- 12 flasks, citrate usage selection
- >50,000 generations of *E coli*!
  - Cultures propagated every day
  - Every 500 generations (75 days), mixed-population samples stored
  - Mean fitness estimated at 500 generation intervals



Comparative genomics in the news [a] [b] [c]

[a] Washington Post 23/2/2015
[b] Ibarra-Laclette *et al.* (2013) *Nature* doi:10.1038/nature12132
[c] Carretero-Paulet *et al.* (2015) *Mol. Biol. Evol.* doi:10.1093/molbev/msv020

- *Utricularia gibba*: carnivorous bladderwort
- Genome: 82Mbp, 17,324 genes (wheat: 17bn bases, ≈94-96k genes)
- Intergenic region contraction (3% repeat elements; most plants: 10-60% repeat elements)
- Genomic context for flowering plants does not require "hidden regulators" (cf. ENCODE)

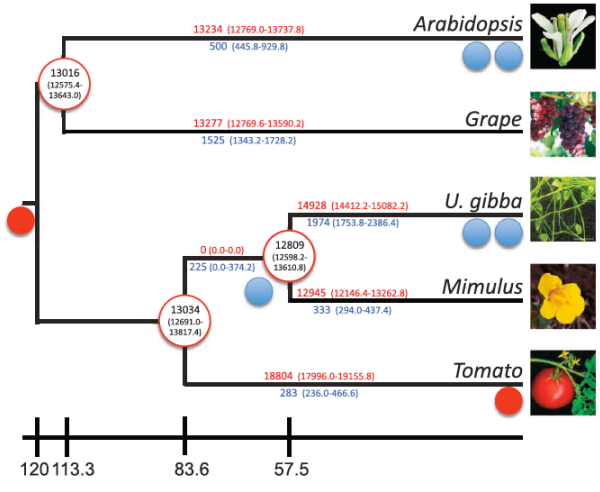


Table of Contents

The James Hutton Institute

Levels of comparison

The James Hutton Institute

Bulk Properties

- chromosome/plasmid counts and sizes, nucleotide content, etc.

Whole Genome Sequence

- sequence similarity
- organisation of genomic regions (synteny), etc.

Genome Features/Functional Components

- numbers and types of features (genes, ncRNA, regulatory elements, etc.)
- organisation of features (synteny, operons, regulons, etc.)
- complements of features
- selection pressure, etc.

Licence: CC-BY-SA

The James Hutton Institute

By: Leighton Pritchard