

Disclaimer

This material is only for internal use and is given to the students to prepare the final evaluation for the course of Applied Genomics of the Master degree course of Bioinformatics.

This file and its content are confidential and intended solely for the use of the individuals to whom they are given. If you have received this file it means that you are a student of the course of Applied Genomics of the master degree course in Bioinformatics, regularly enrolled for the academic year 2018-2019. If you are not a student of this course you should not disseminate, distribute or copy this file. Please notify the professor immediately by e-mail if you have received this file.

In any case, also students of this course are notified that disclosing, copying, distributing or taking any action in reliance on the contents of this information is strictly prohibited.

For the students: please note that the content of this file is not enough to pass the exam. The content of this file could contain a few errors as it has not been peer reviewed or edited after its preparation.

72847 - APPLIED GENOMICS

Prof. Luca Fontanesi

Department of Agricultural and Food Sciences (DISTAL)

Division of Animal Sciences

Viale Fanin 46, 40127 Bologna

Tel: 051 2096535

E-mail: luca.fontanesi@unibo.it

<http://www.unibo.it/sitoweb/luca.fontanesi>

Who I am

Full Professor of Animal Breeding and Genetics

Department of Agricultural and Food Sciences
(Division of Animal Sciences)

Dipartimento di Scienze e Tecnologie Agro-Alimentari

Research activities in Animal Genomics

Where and when you can find me.

Where:

Faculty of Agriculture
Viale Fanin 46, Bologna

Building: North Block, second floor, former DIPROVAL administrative area
Office 2

When:

Appointment by e-mail: luca.fontanesi@unibo.it

Subject of your e-mails: please include always “Applied Genomics”



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DIPARTIMENTO

SCIENZE E TECNOLOGIE AGRO-ALIMENTARI - DISTAL

[Biblioteche](#) [Contatti](#)

Home



Dipartimento

La storia, l'organizzazione, le persone, gli organi



Ricerca

Gli ambiti, le pubblicazioni, i progetti, le attività



Didattica

Lauree e Lauree magistrali, i dottorati e gli altri corsi



Centri e Laboratori



Dipartimento di Scienze e Tecnologie Agro-Alimentari

L'organizzazione, le persone, l'amministrazione e le strutture che compongono il Dipartimento di Scienze e Tecnologie Agro-Alimentari (DISTAL)

In evidenza

on focus

Avvisi



[PROROGA ISCRIZIONI - Freschi e Trasformati
Ortofrutticoli: Innovazione & Internazionalizzazione](#)
15 ottobre 2013

[Studenti: Avviso bollettino MAV](#)
18 marzo 2013

[Chiusure programmate Distal - anno 2013](#)
14 marzo 2013

[Tutti gli avvisi](#)

Eventi



Non ci sono eventi da visualizzare

[Tutti gli eventi](#)

Course description

Prerequisites:

The knowledge of genetics and molecular biology are important prerequisites to follow the course. Who might have gaps in these fields is strongly encouraged to contact the professor at the beginning of the course.

Program of the course:

- 1) Foundational concepts in genetics (including population and quantitative genetics) and genomics.
- 2) Genome structure and variability in vertebrates
- 3) The transcriptional landscape of the mammalian genome
- 4) High throughput technologies for genotyping and next generation sequencing (NGS) platforms
- 5) Applications of NGS, array comparative genome hybridization
- 6) Linkage disequilibrium and linkage analysis, genetic mapping
- 7) QTL mapping, eQTL
- 8) Candidate gene analysis, genome wide association studies, selection signature
- 9) Relevant genomic projects: modENCODE, ENCODE, 1000 genome project, The Mammalian Genome Project, 10K Genome project.
- 10) Discussion of relevant scientific literature

FINAL EXAM

The final exam evaluates the students according to the following objectives:

- 1) knowledge of the foundational concepts in genetics and genomics;
- 2) knowledge of advanced technologies for genome analysis;
- 3) knowledge of advanced applications of genomic tools to answer biological questions in vertebrates;
- 4) capacity planning of genomic experiments.

The final exam has two levels:

- 1) Preparation of a genomic project: a text should be written including an appropriate introduction to the problem/question that the experiment or project would like to analyse or answer, aim of the project, a section with materials and methods, expected results and impact. The project should be submitted to the professor one week before the interview.
- 2) Interview based on the project submitted and other two questions according to the main objectives of the course. Only students that are positively evaluated at the first level are admitted at the second level.

The final score will be based on the performance of the students at the two levels.

References/text

Scientific articles, references and software are provided to the students during the course.

Lecture slides are given to the students usually at the end of the course.

To fill gaps: any text of Genetics could be useful

My research activities and topics of possible master theses in Applied Genomics

- 1) Genome wide association studies in livestock species for production traits
- 2) High throughput genotyping data for allocation of animal breeds
- 3) Copy number variation analysis in dairy cattle and pig and association with production traits
- 4) Resequencing animal genomes – and reassembling – including aquaculture species
- 5) Ancient DNA – *Prolagus sardus*
- 6) Next generation sequencing for authentication of food products
- 7) Metabolomics merged to genomics
- 8) Phenomics
- 9) Collaborations with TGAC (UK – Norwich) – 10K Genome Project – LaGomiCs – Pig genomics – vertebrate genomics –etc.

First exam session:

Oral sessions:

To be defined

Preliminary activities

- 1) Survey of your background and expertises
- 2) What do you expect from this course ?

Glossary of genetics terms

Why this course?

Genomics is the study of genome structure and function.

This is a new and exciting area that has recently witnessed many conceptual and technical advances.

This information is vital to our day-to-day living in this century.

Such a course would also make Bioinformatics students competitive in current job market that is seeking Bioinformaticians who can understand and handle genomics data

What is genomics?

Study of genomes

What is the genome?

Entire genetic
content/information of an
organism

...ACGTGTGCGTGAAAGGG...

Applied genomics

Use of technologies, tools
and experimental designs
to analyse genomes and
extract information

Living organisms

Current research theories support the division of living organisms into three domains

1. Bacteria
2. Eukaryota
3. Archaea living in the most inhospitable regions of the earth
 - Thermophiles tolerate extremely high temperatures
 - Halophiles tolerate very high salt concentrations
 - Methanogens produce methane as a by-product of metabolism

How many types of genomes are there?

- Virus genomes

- Prokaryotic genomes

Prokaryotes = Monera (simple bacteria) and Archaea

- Eukaryotic Genomes

 - Nuclear Genomes

 - Mitochondrial genomes

 - Choloroplast genomes

A complex organism (human) is a unique creation!

Life's little book of instructions

DNA blue print of life!

Human body has 10^{13} cells and each cell has 6 billion base pairs (A, C, G, T)

A hidden language/code determines which proteins should be made and when

This language is common to all organisms

What can genome sequence tell us?

- Almost everything about the organism's life
- Its developmental program
- Disease resistance or susceptibility
- (Its reactions against the environment)
- History
- Evolutionary perspectives (future)
- ...

How is human genome organized?

3% coding and rest of it junk (repetitive DNA).

Nuclear and mitochondrial

You are 99.99% similar to your neighbor

Why horse is a horse and duck is a duck?

It is in their genes!

DNA replicates by making a copy of itself and passes to next generation of cells or organisms

Purity of lineages maintained

Evolution

Now look at your neighbor

- What do you see?
- Someone is different than you!
- Could be that your friend differs in his/her sex, looks, nature, smartness, or simply the way he/she dresses and talks
- How much similarity you think you share with your friend at the gene level?
- 99.9%

Now look at your own body (hands and legs)

- Do they look similar?
- But they contain the same DNA in each of their cells
- DNA makes RNA makes proteins
- Different genes are expressed differently in different cells, tissues and organs of an organism
- Having a gene does not mean it will be expressed.

All people have cancer genes

- They are normal genes that got mutated or changed and do not perform the regular job
- But having a gene does not mean you will get cancer
- Because environment has a big role in turning a gene on or off
- Different genes and their products also interact: microecosystem
- Genes do not work alone (G X E)

Genomics is also the study
of all genes present in an
organism

Origin of terminology

- The term genome was used by German botanist Hans Winker in 1920
- Collection of genes in haploid set of chromosomes
- Now it encompasses all DNA in a cell
- In 1986 mouse geneticist Thomas Roderick used Genomics for “mapping, sequencing and characterizing genomes”
- New terms: Functional genomics, transcriptomics, proteomics, metabolomics, phenomics (Omics)

What is genomics?

- A marriage of genetics, molecular biology, robotics, and computing
- Tools and techniques of recombinant DNA technology
e.g., DNA sequencing, making libraries and PCRs
- High-throughput technologies
e.g., sequencing
- Computers are essential for processing and analyzing the large quantities of data generated

Origin of Genomics

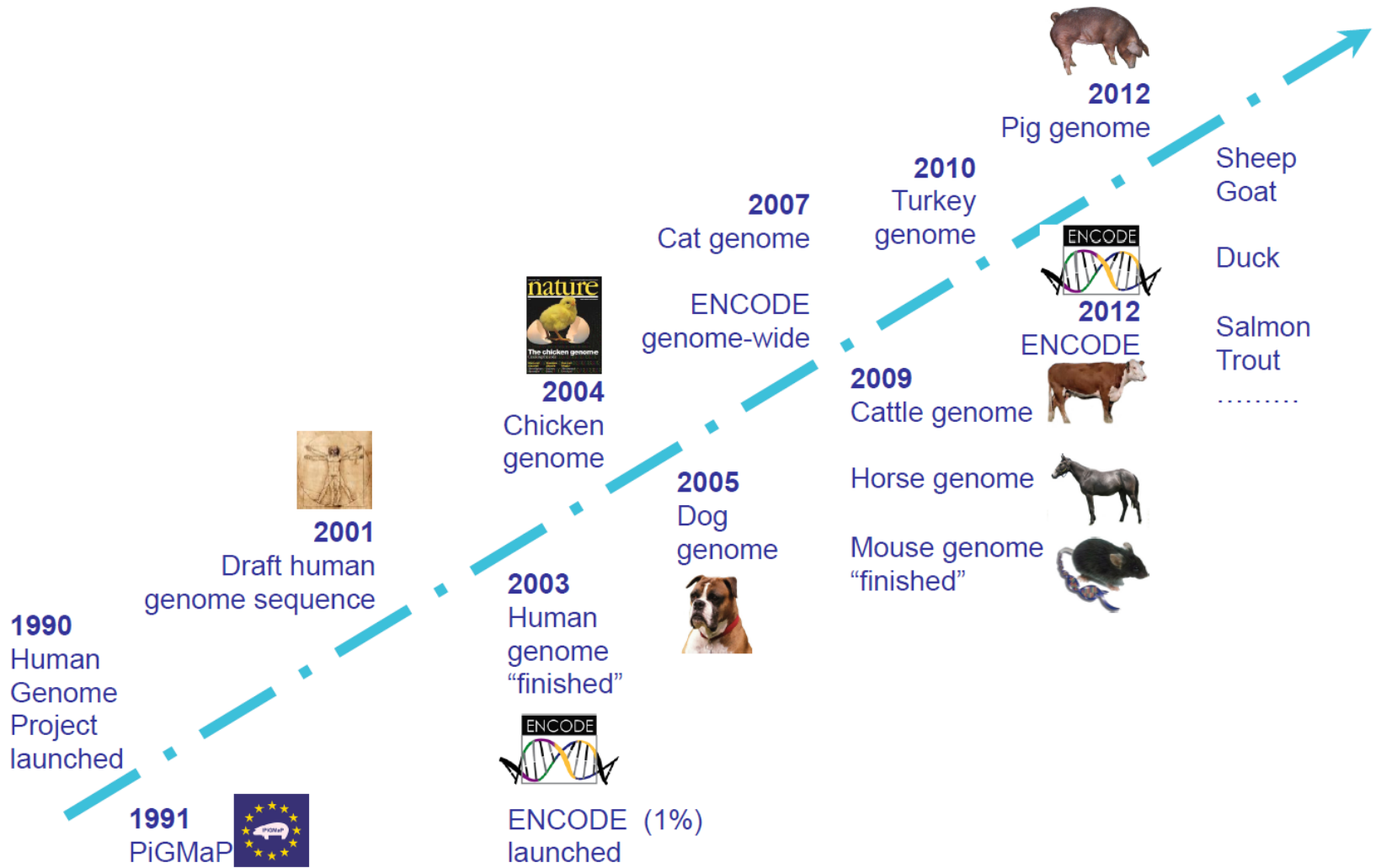
- Human Genome Project:

Goal: sequence 3 billion base pairs

- High-quality sequence (<1 error per 10 K bases) ACGT
- Immensity of task required new technologies

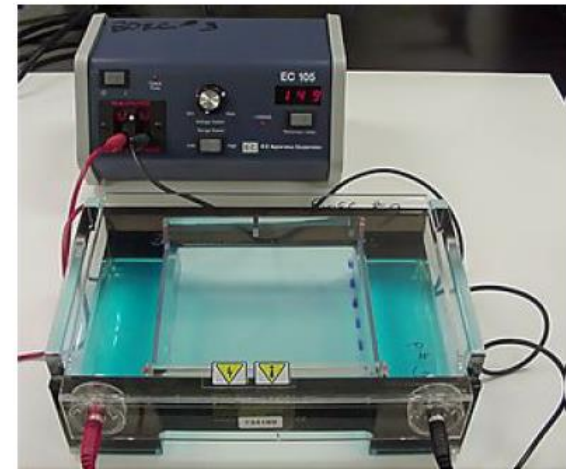
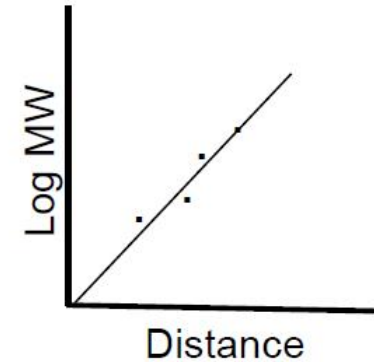
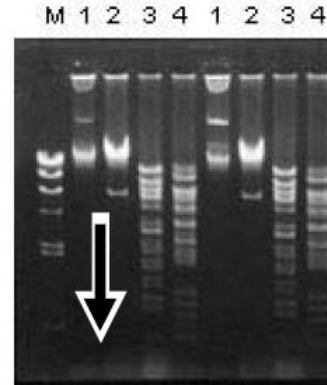
Automated sequencing

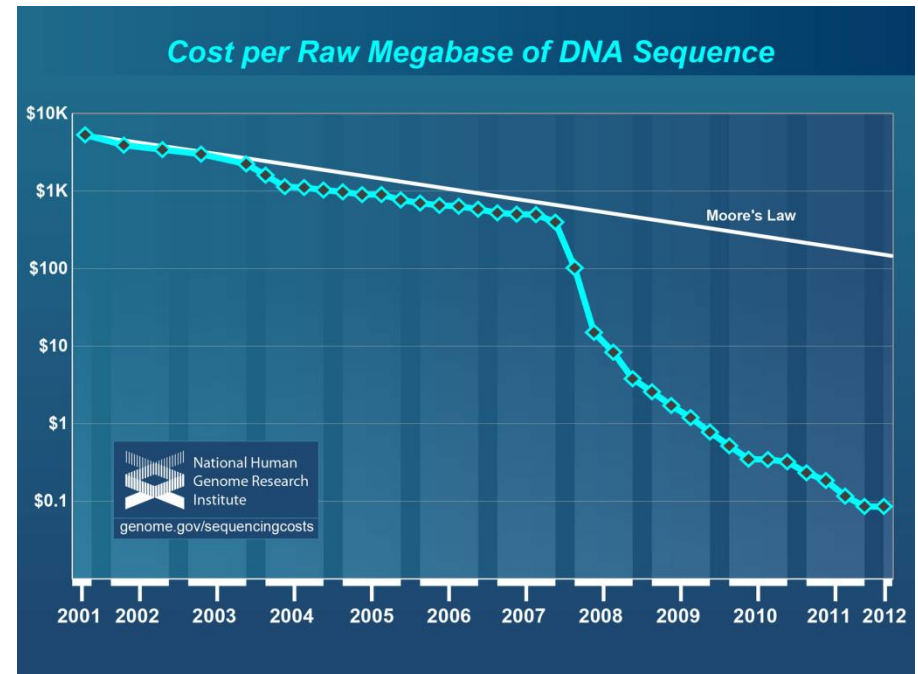
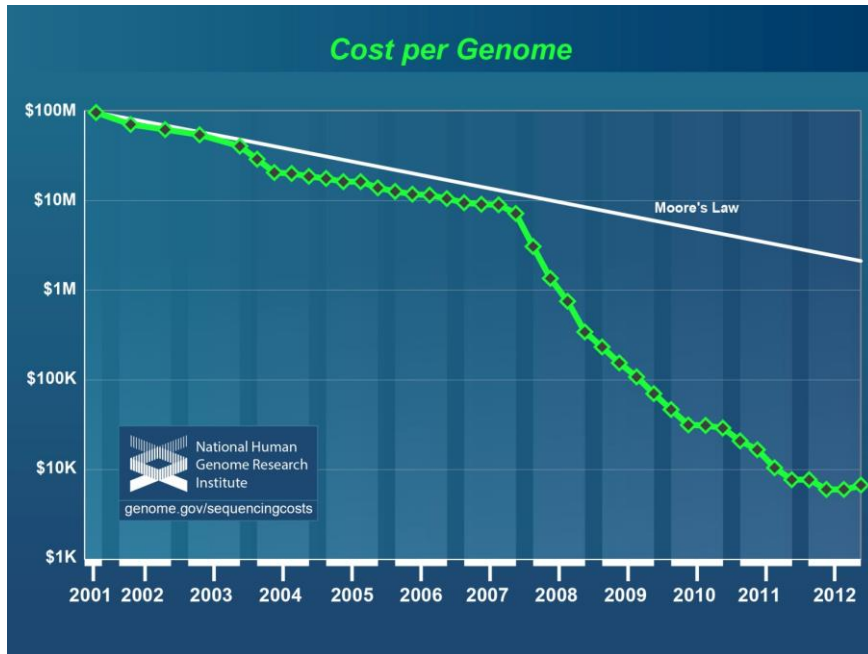
- Decision to sequence other genomes: yeast and bacteria
- Beginnings of comparative genomics



Technical foundations of genomics

- Molecular biology:
recombinant-DNA technology
- **DNA sequencing**
- Library construction
- PCR amplification
- Hybridization techniques





PERSPECTIVE

Big Data: Astronomical or Genomical?

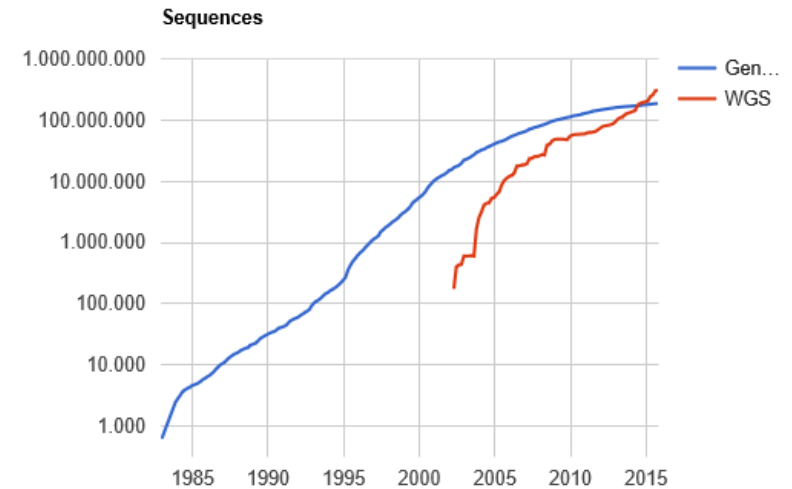
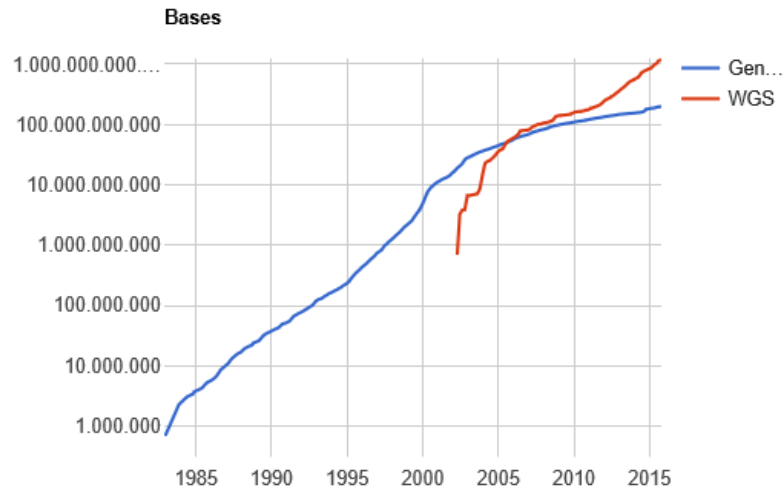
Zachary D. Stephens¹, Skylar Y. Lee¹, Faraz Faghri², Roy H. Campbell², Chengxiang Zhai³, Miles J. Efron⁴, Ravishankar Iyer¹, Michael C. Schatz^{5*}, Saurabh Sinha^{3*}, Gene E. Robinson^{6*}

1 Coordinated Science Laboratory and Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America, **2** Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America, **3** Carl R. Woese Institute for Genomic Biology & Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America, **4** School of Library and Information Science, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America, **5** Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, United States of America, **6** Carl R. Woese Institute for Genomic Biology, Department of Entomology, and Neuroscience Program, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America

* mschatz@cshl.edu (MCS); sinhas@illinois.edu (SS); generobi@illinois.edu (GER)



Growth of GenBank and WGS



Whole Genome Shotgun (WGS) projects are genome assemblies of incomplete genomes or incomplete chromosomes of prokaryotes or eukaryotes that are generally being sequenced by a whole genome shotgun strategy.

Genomic data

Zachary et al. 2015

<u>Data Phase</u>	<u>Astronomy</u>	<u>Twitter</u>	<u>YouTube</u>	<u>Genomics</u>
Acquisition	25 zetta-bytes/year	0.5–15 billion tweets/year	500–900 million hours/year	1 zetta-bases/year
Storage	1 EB/year	1–17 PB/year	1–2 EB/year	2–40 EB/year
Analysis	In situ data reduction	Topic and sentiment mining	Limited requirements	Heterogeneous data and analysis
	Real-time processing	Metadata analysis		Variant calling, ~2 trillion central processing unit (CPU) hours
	Massive volumes			All-pairs genome alignments, ~10,000 trillion CPU hours
Distribution	Dedicated lines from antennae to server (600 TB/s)	Small units of distribution	Major component of modern user's bandwidth (10 MB/s)	Many small (10 MB/s) and fewer massive (10 TB/s) data movement

doi:10.1371/journal.pbio.1002195.t001

Terabyte (1 000 000 000 000 bytes)

Petabyte (1 000 000 000 000 000 bytes)

Exabyte (1 000 000 000 000 000 000 bytes)

Zettabyte (1 000 000 000 000 000 000 000 bytes)

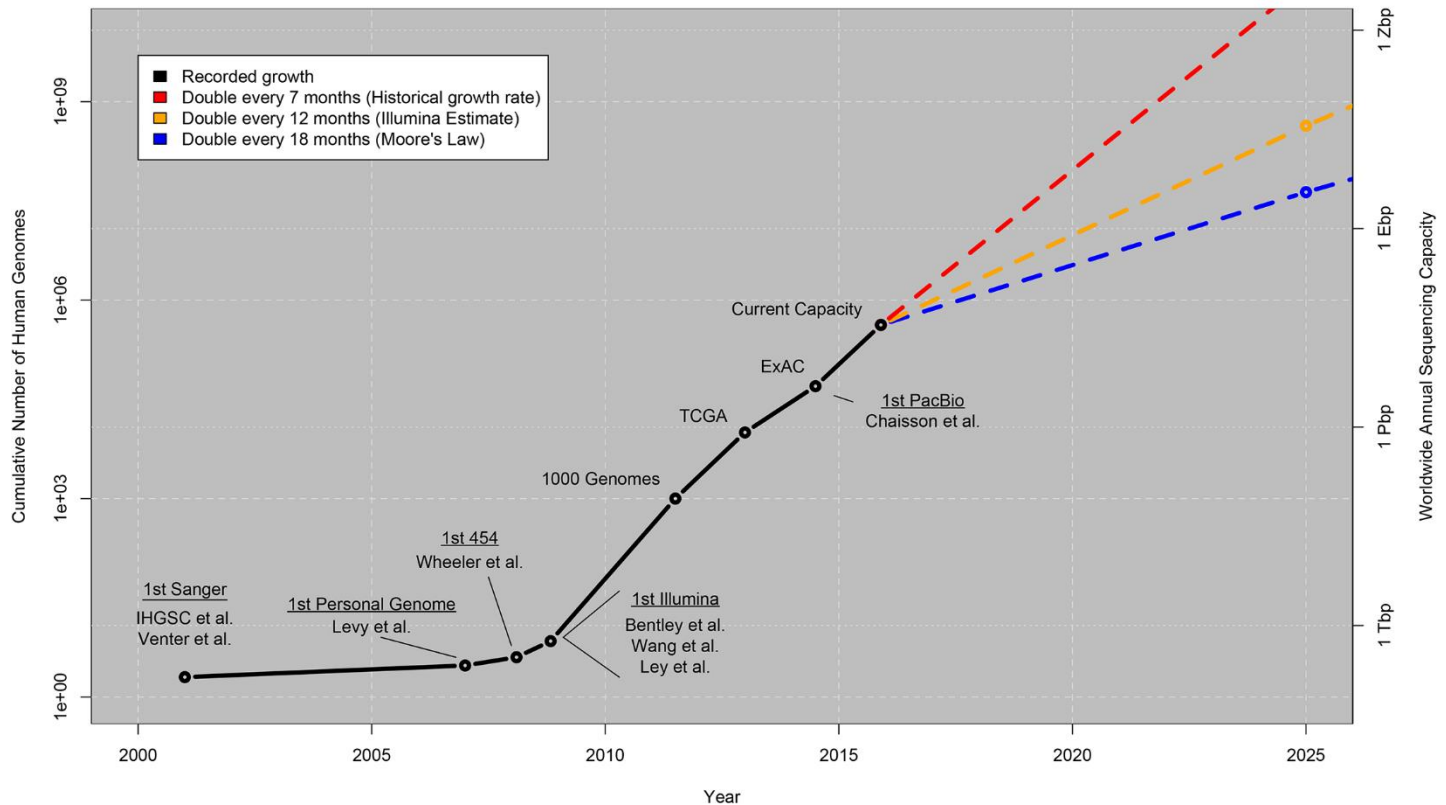
Genomic data

Zachary et al. 2015

Data Phase	Astronomy	Twitter	YouTube	Genomics
Acquisition	25 zetta-bytes/year	0.5–15 billion tweets/year	500–900 million hours/year	1 zetta-bases/year
Storage	1 EB/year	1–17 PB/year	1–2 EB/year	2–40 EB/year
Analysis	In situ data reduction	Topic and sentiment mining	Limited requirements	Heterogeneous data and analysis
	Real-time processing	Metadata analysis		Variant calling, ~2 trillion central processing unit (CPU) hours
	Massive volumes			All-pairs genome alignments, ~10,000 trillion CPU hours
Distribution	Dedicated lines from antennae to server (600 TB/s)	Small units of distribution	Major component of modern user's bandwidth (10 MB/s)	Many small (10 MB/s) and fewer massive (10 TB/s) data movement

doi:10.1371/journal.pbio.1002195.t001

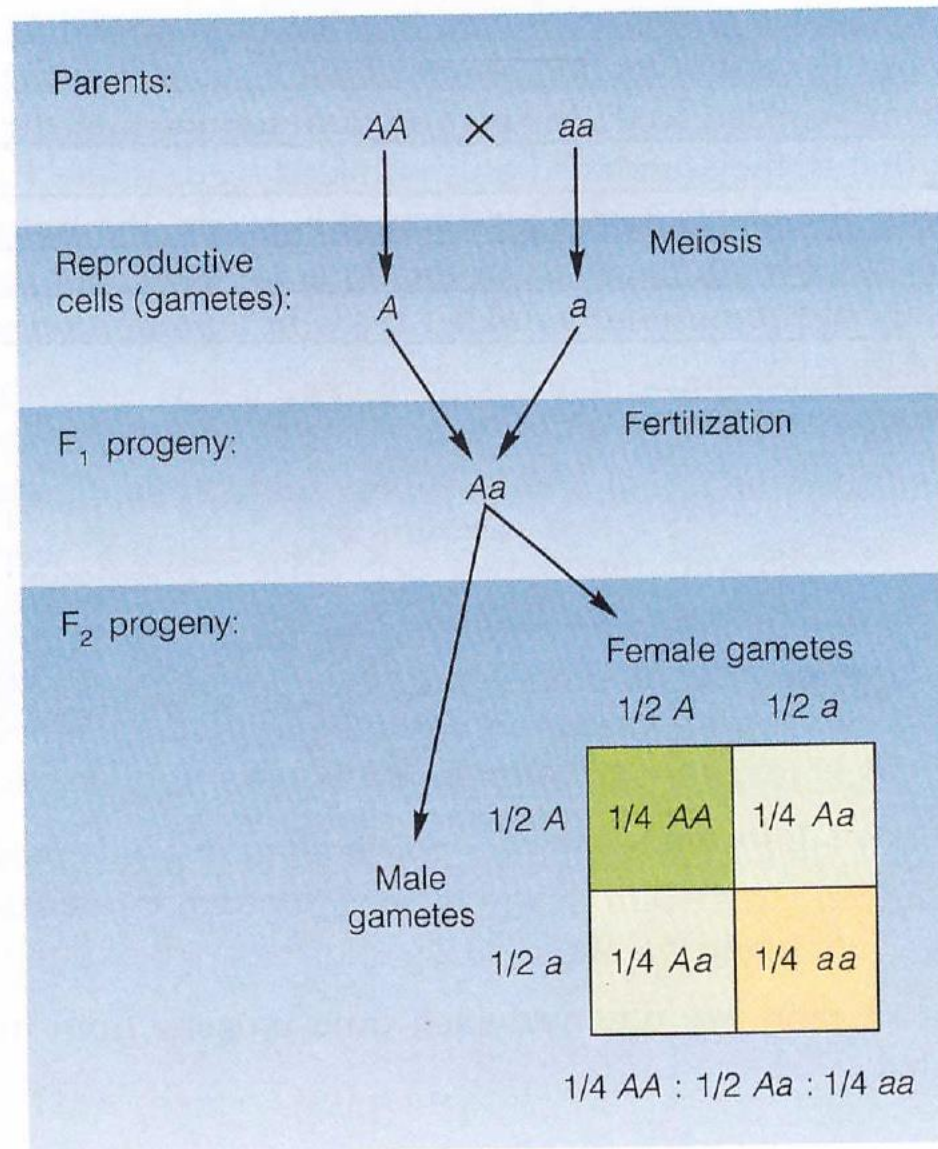
Growth of DNA Sequencing



Fundamental concepts in genetics

The Principle of Segregation (Mendel's first law)

Crossing individuals



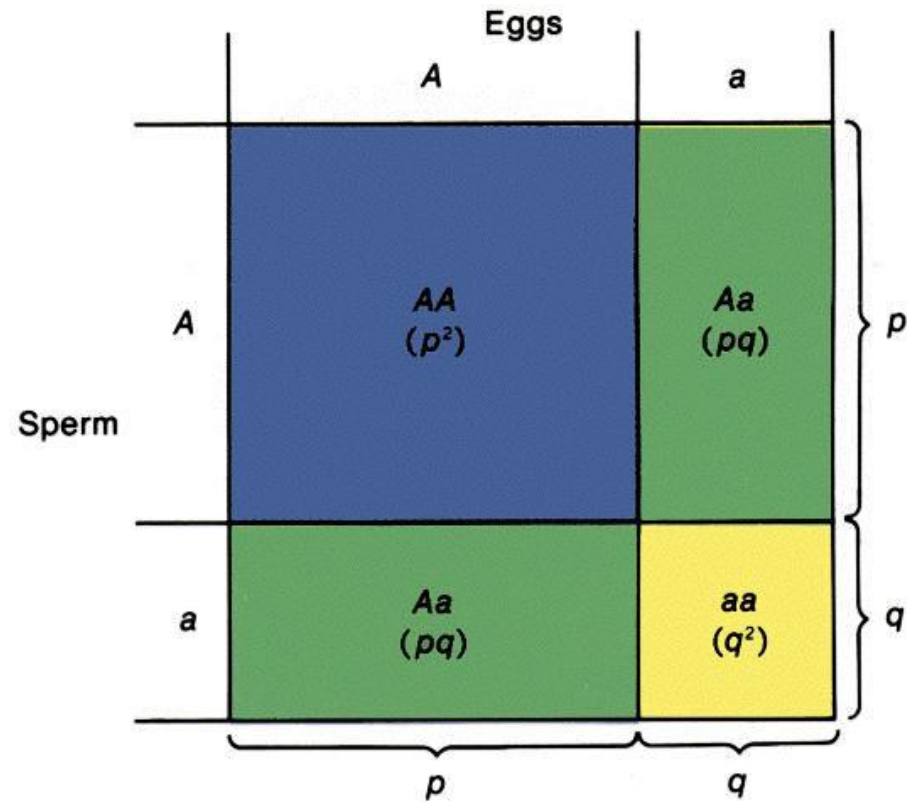
-Mendel's particulate heredity elements are called genes

-The various forms of a given gene are called alleles

-The genotype is the genetic constitution

- The phenotype is an observable property of an organism

In a population



If alleles are more than 2?

In a population

	p	q	r
p			
q			
r			

If alleles are more than 2?

In a population

	p	q	r
p			
q			
r			

Expected homozygous have the frequency that is = the square of the allele frequency

If alleles are more than 2?

In a population

	p	q	r
--	---	---	---

p			
q			
r			

The heterozygous have frequency = $2 \times$ (the product of the frequencies of the alleles)

The Principle of Independent Assortment (Mendel's second law)

Crossing individuals

