FORTH TO THE PAST!!

From Biotech to Cultural Heritage: A leap forward !

IMBB LABORATORY

Dimitris Kafetzopoulos

kafetzo@imbb.forth.gr

11η επιστημονική ΔΙΗΜΕΡΙΔΑ ITE 2017

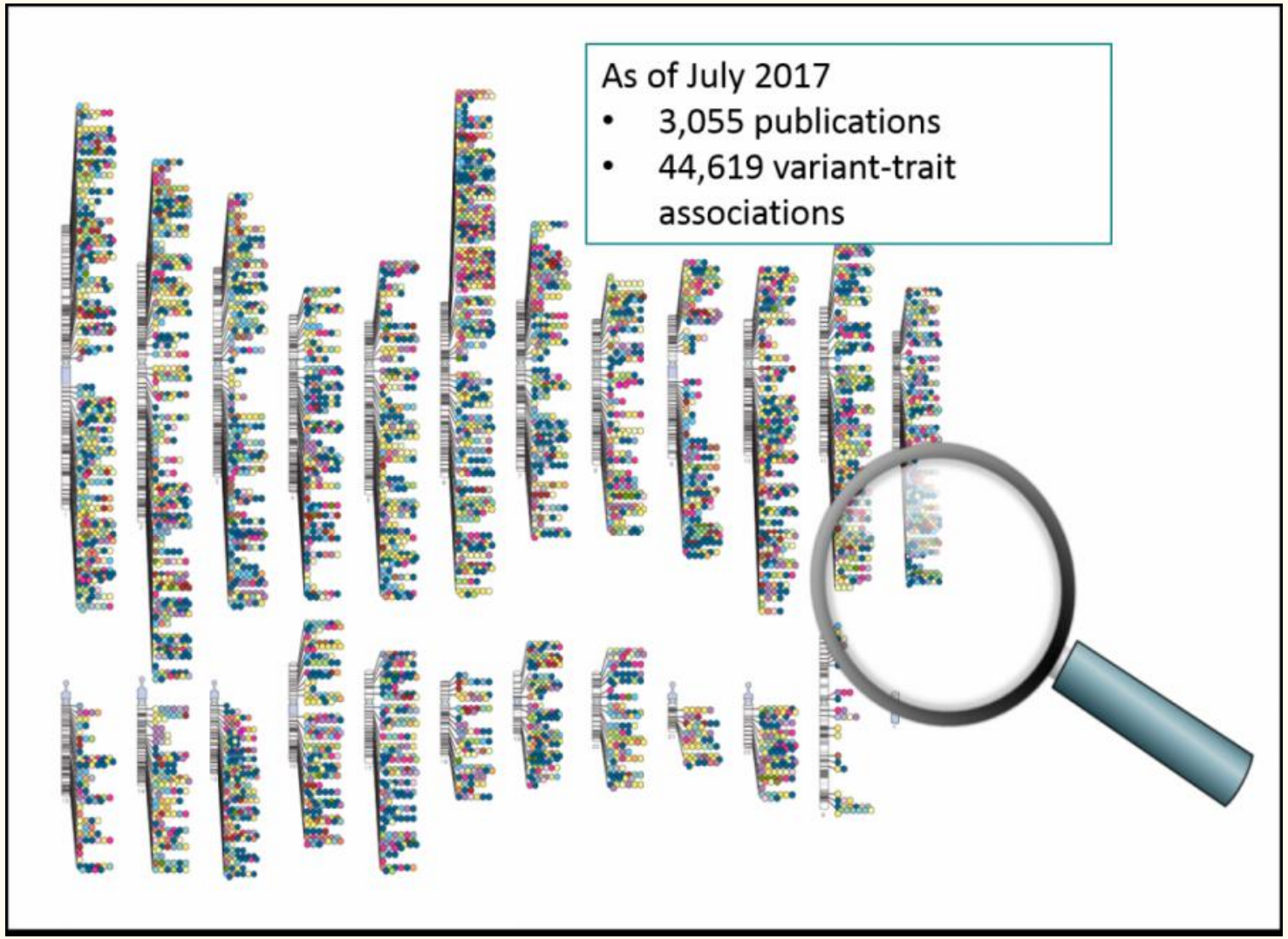
Molecular Biotechnology Break-through

Completion of the Human Genome Project (2003) and publication of the HapMap Project results (2005) initiated the development of a wide range of technologies and applications including genomics, (NGS, nanopore, single molecule), transcriptomics (microarrays, molecular signature assays), proteomics (Mass spec), metabolomics (GC-MS, NMR), eDNA profiling, genetic breeding, precision medicine, personal genomics etc.

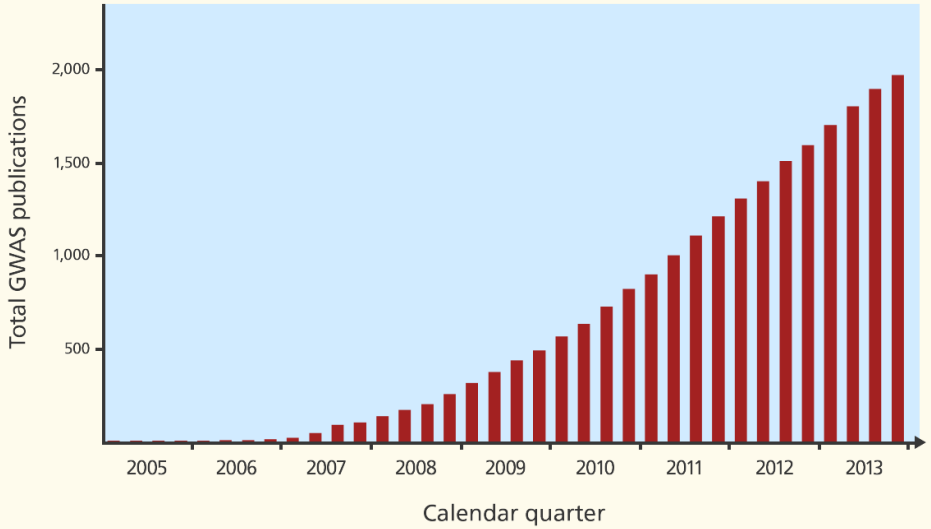
Obama, announcing his "Precision Medicine Initiative" Jan 30, 2015:

One study found that every dollar we spent to map the human genome has already returned $140 to our economy

GWAS & SNP maps

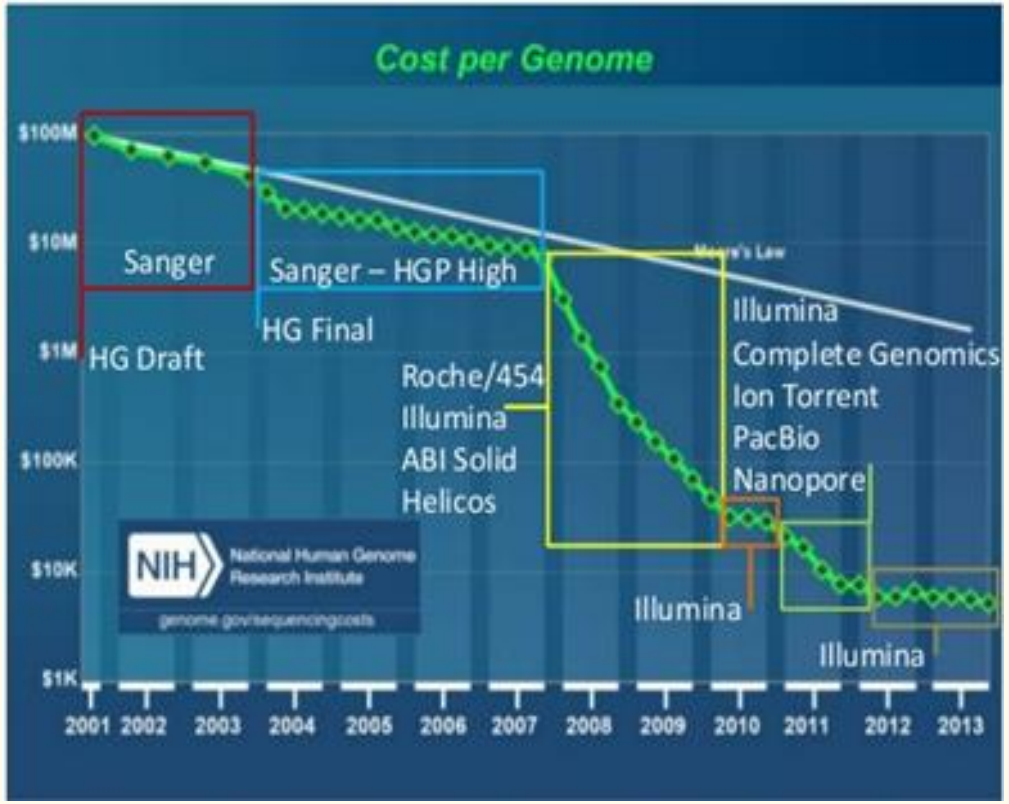


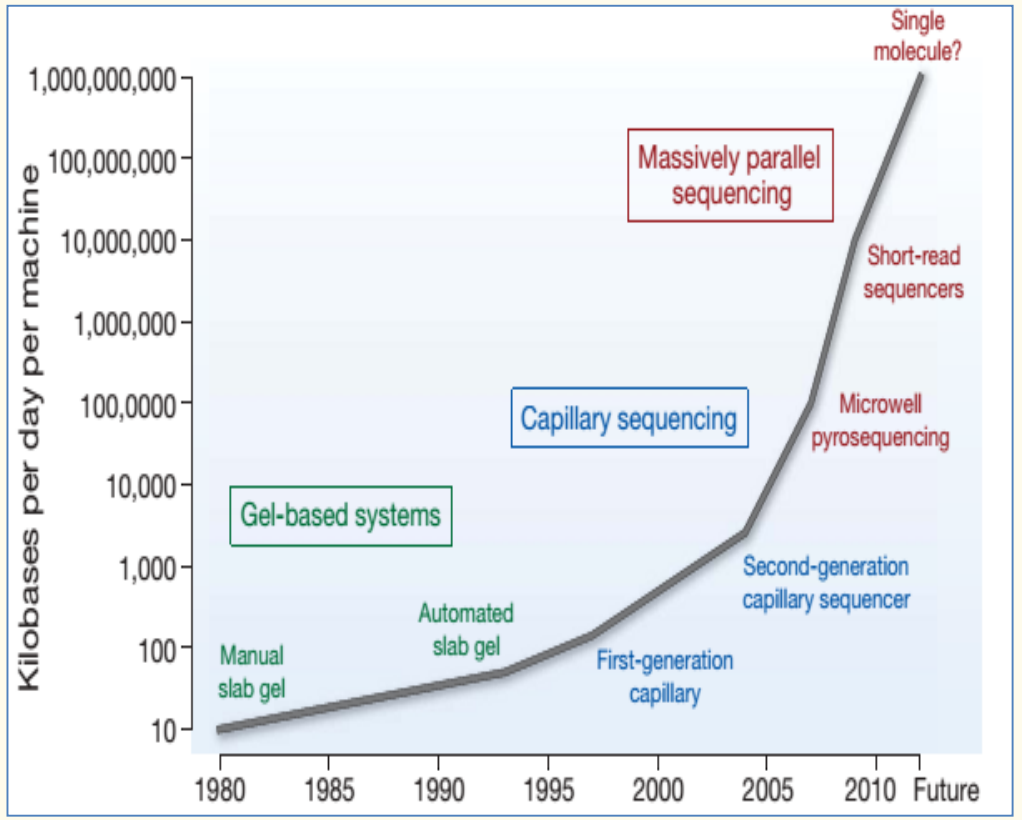
Number of GWAS since HapMap Project



DNA sequencing revolution

Low cost, huge output, broadly accessible and versatile





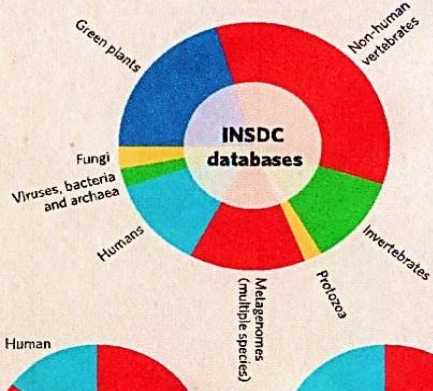
THE SEQUENCE EXPLOSION

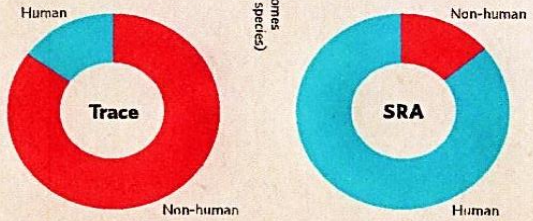
At the time of the announcement of the first drafts of the human genome in 2000, there were 8 billion base pairs of sequence in the three main databases for 'finished' sequence: GenBank, run by the US National Center for Biotechnology Information; the DNA Databank of Japan; and the European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database. The databases share their data regularly as part of the International Nucleotide Sequence Database Collaboration (INSDC). In the subsequent first post-genome decade, they have added another 270 billion bases to the collection of finished sequence, doubling the size of the database roughly every 18 months. But this number is dwarfed by the amount of raw sequence that has been created and stored by researchers around the world in the Trace archive and Sequence Read Archive (SRA).

See Editorial, page 649, and human genome special at www.nature.com/humangenome

DNA SEQUENCES BY TAXONOMY

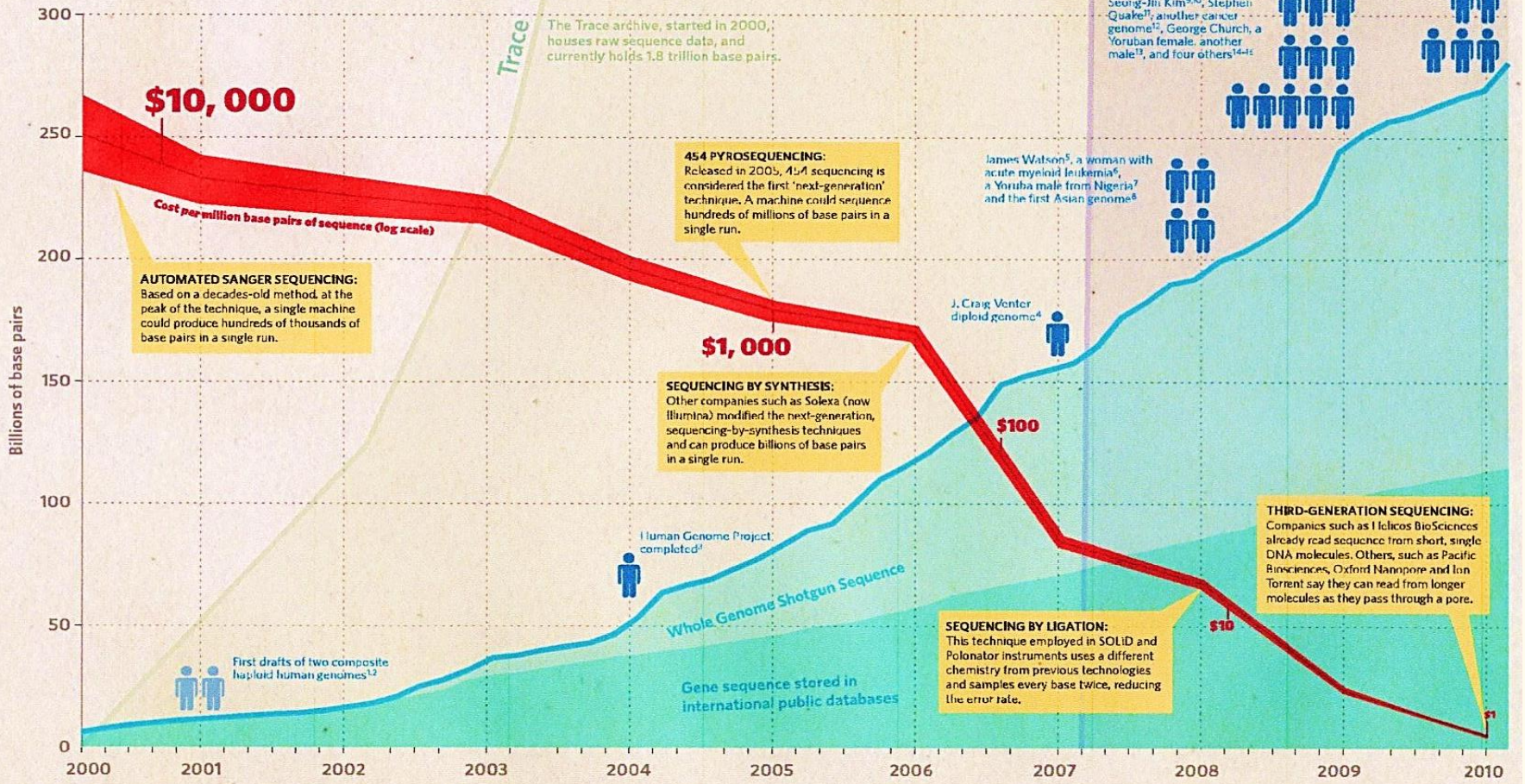
International Nucleotide Sequence Database Collaboration: The main repositories of 'finished' sequence span a wide range of organisms, representing the many priorities of scientists worldwide.





Trace Archive: Developed to house the raw output of high-throughput sequencers built in the late 1990s, the trace archive spans a wide range of taxa.

Sequence Read Archive: Houses raw data from next-generation sequencers. Dominated by human sequence, including multiple coverage for more than 170 people.



HOW MANY HUMAN GENOMES?

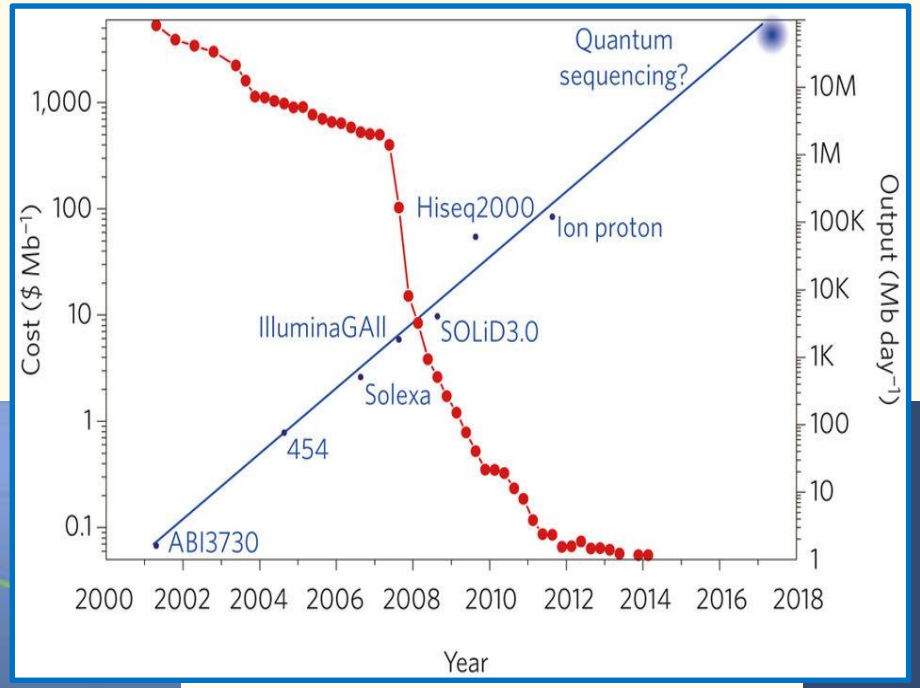
The graphic shows all published, fully sequenced human genomes since 2000, including nine from the first quarter of 2010. Some are resequencing efforts on the same person and the list does not include unpublished completed genomes.

1. Venter, J. C. *et al. Science* **291,** 1304–1351 (2001).
2. International Human Genome Sequencing Consortium *Nature* **409,** 860–921 (2001).
3. International Human Genome Sequencing Consortium *Nature* **431,** 931–945 (2004).
4. Levy, S. *et al. PLoS Biol.* **5,** e254 (2007).
5. Wheeler, D. A. *et al. Nature* **452,** 872–876 (2008).
6. Ley, T. J. *et al. Nature* **456,** 66–72 (2008).
7. Bentley, D. R. *et al. Nature* **456,** 53–59 (2008).
8. Wang, J. *et al. Nature* **456,** 60–65 (2008).
9. Ahn, S.-M. *et al. Genome Res.* **19,** 1622–1629 (2009).
10. Kim, J.-I. *et al. Nature* **460,** 1011–1015 (2009).
11. Pushkarev, D., Neff, N. F. & Quake, S. R. *Nature Biotechnol.* **27,** 847–850 (2009).
12. Mardis, E. R. *et al. N. Engl. J. Med.* **10,** 1058–1066 (2009).
13. Drmanac, R. *et al. Science* **327,** 78–81 (2009).
14. McKernan, K. J. *et al. Genome Res.* **19,** 1527–1541 (2009).
15. Pleasance, E. D. *et al. Nature* **463,** 191–196 (2010).
16. Pleasance, E. D. *et al. Nature* **463,** 184–190 (2010).
17. Clark, M. J. *et al. PLoS Genet.* **6,** e1000832 (2010).
18. Rasmussen, M. *et al. Nature* **463,** 757–762 (2010).
19. Schuster, S. C. *et al. Nature* **463,** 943–947 (2010).
20. Lupski, J. R. *et al. N. Engl. J. Med.* doi:10.1056/NEJMoa0908094 (2010).
21. Roach, J. C. *et al. Science* doi:10.1126/science.1186802 (2010).

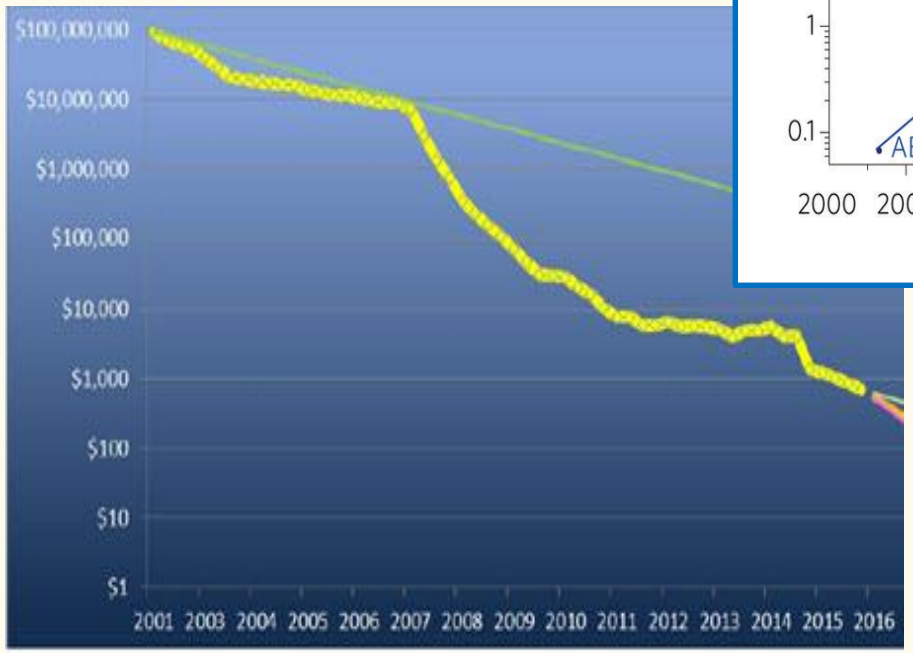
11η επιστημονική ΔΙΗΜΕΡΙΔΑ ITE 2017
Dimitris Kafetzopoulos
Heraklion, October 14

Trends in Human Genome Sequencing Costs

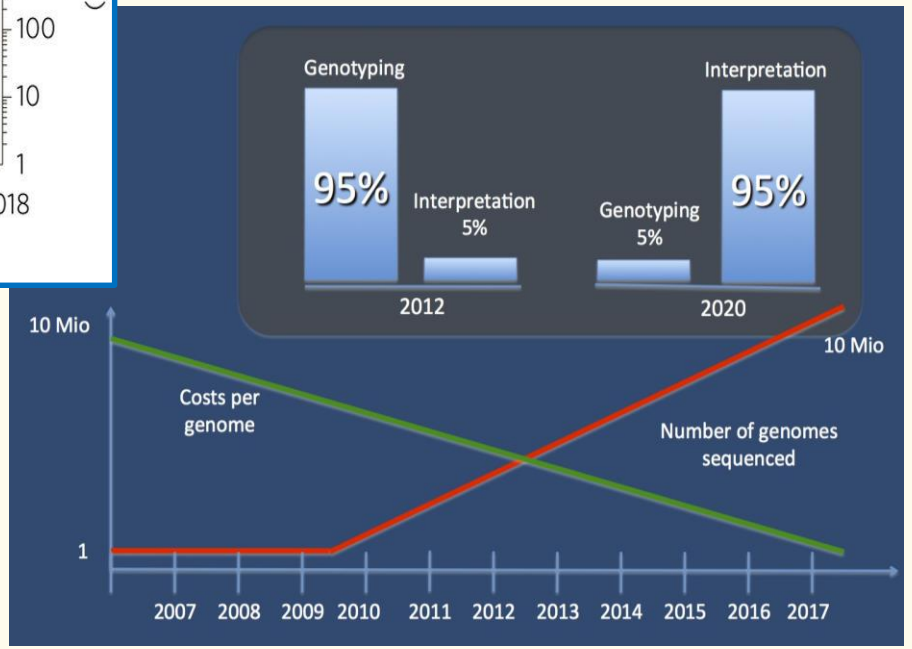
Cost per genome vs Platform output



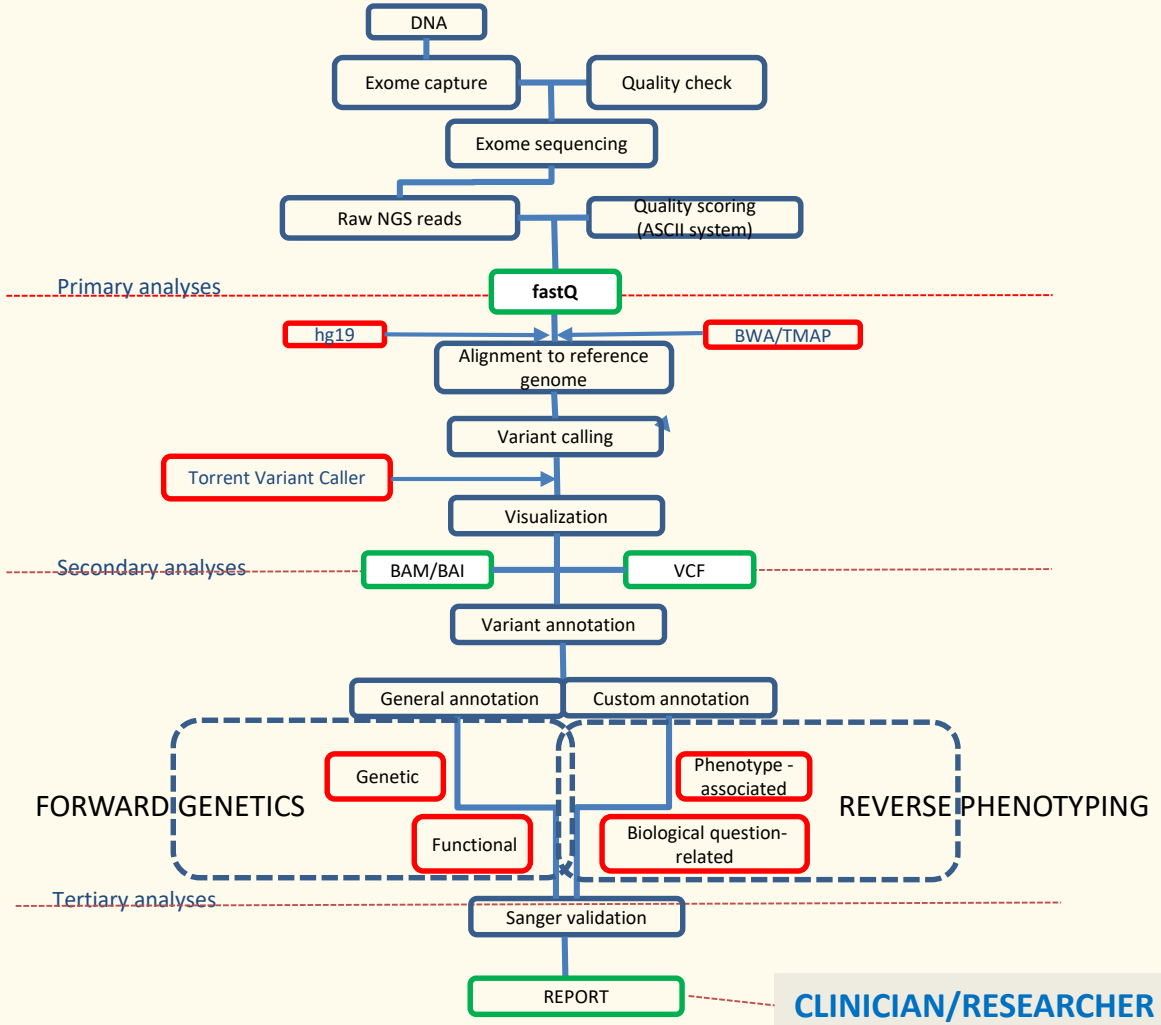
Cost per genome vs Moore



Genome vs Interpretation Costs



Whole Exome Sequencing Workflow

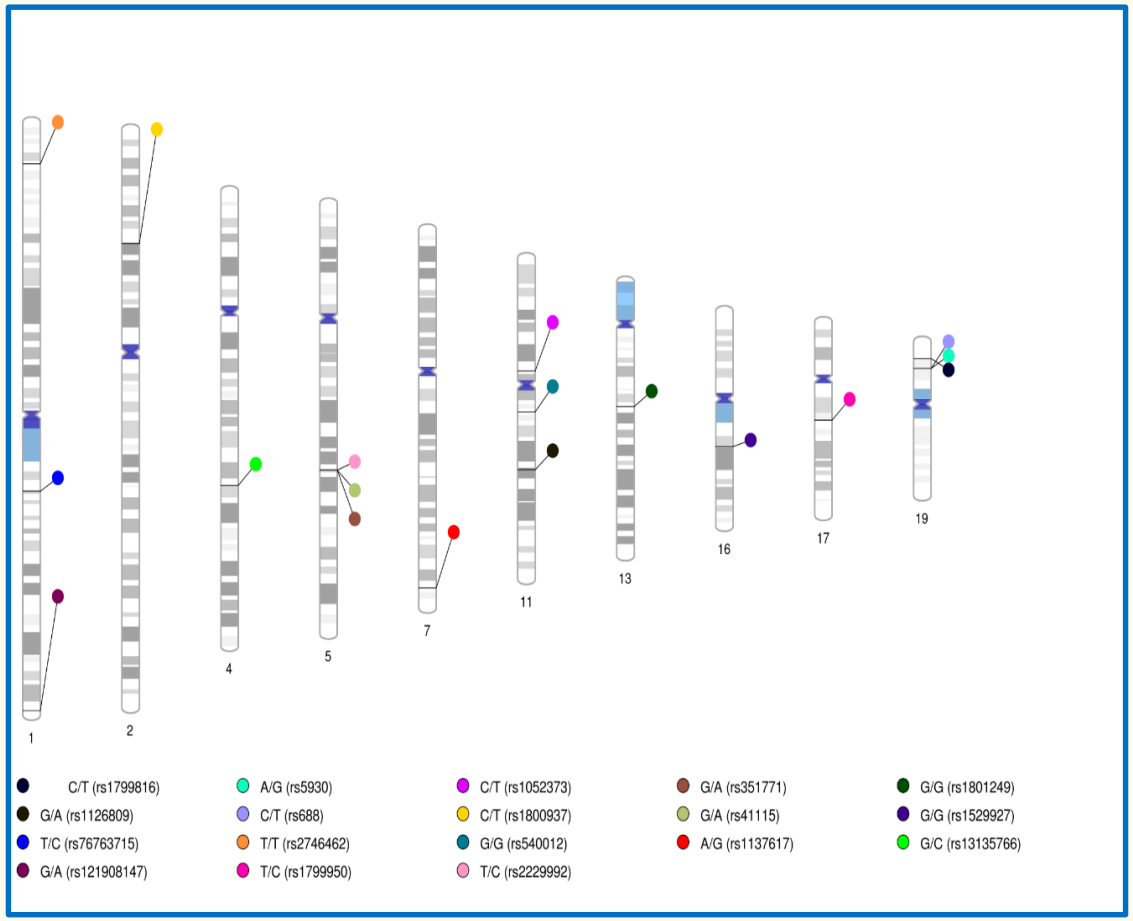


MinoTech Genomics Facility

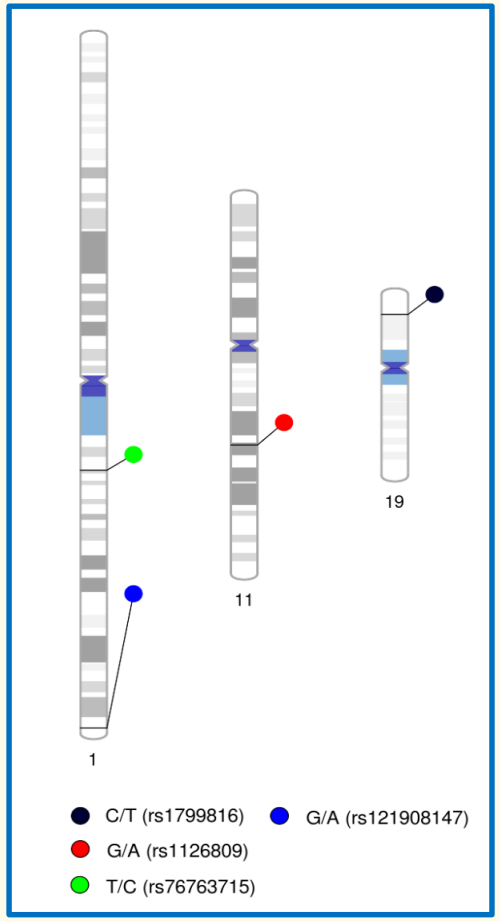
Geno-gram:

Reported Pathogenic Variants after manual annotation/curation

Initial 21 reported pathogenic variants



1 pathogenic + 3 conflicting



Human Genetic Morphometry

Human Diversity

- 3.000.000 polymorthisms (SNP)
- 2/3 transitions from C to T
- Occur in both coding and not coding regions
- Evolutionarily stable
- Show different frequencies
- Are useful for genome mapping
- Are inherited grouped in haplotypes
- Show different penetration in phenotype.

Genomic determinants of facial characteristics

A Practical Guide to the HIrisPlex System: Simultaneous Prediction of Eye and Hair Color from DNA

Susan Walsh and Manfred Kayser

Number	DNA variant	Gene	Blond(beta)	Blond(p)	Brown(beta)	Brown(p)	Black(beta)	Black(p)	Red(beta)	Red(p)
1	N29insA	MC1R	–	–	–	–	–	–	-21.9731	0.994026
2	rs11547464	MC1R	−0.947299	0.081175	−0.4007191	0.441688	−16.782634	0.995907	−2.8866	4.42E−08
3	rs885479	MC1R	0.272536	3.36E−01	0.1938828	0.460717	2.29E−01	0.575679	0.315529	0.707292
4	rs1805008	MC1R	−0.57034	0.003874	−0.3058868	0.097798	−5.66E−01	0.084668	−3.02472	2.20E−16
5	rs1805005	MC1R	0.20689	2.28E−01	0.2382036	0.128146	−1.57E−01	0.539306	−0.86742	0.025064
6	rs1805006	MC1R	1.718508	0.045418	2.1268136	0.009857	−1.70E+01	0.996356	−2.43626	0.001714
7	rs1805007	MC1R	−0.53542	0.030279	−0.1503278	0.508278	−1.32E+00	0.009567	−3.59956	2.20E−16
8	rs1805009	MC1R	0.550547	5.60E−01	0.5309897	0.49513	−4.70E−01	0.693758	−4.25774	4.14E−08
9	Y152OCH	MC1R	–	–	–	–	–	–	−19.3501	0.992969
10	rs2228479	MC1R	−0.025643	8.83E−01	−0.1128742	0.483857	1.98E−01	0.413966	−0.61967	0.110936
11	rs1110400	MC1R	−0.366071	0.338334	−0.5920858	0.123046	6.63E−01	0.21252	−1.67775	0.009302
12	rs28777	SLC45A2	0.566568	0.414238	0.3138274	0.561428	4.85E−01	0.468883	−0.41607	0.743869
13	rs16891982	SLC45A2	0.863795	0.194837	0.2562763	0.618846	6.29E−01	0.326034	0.891013	0.522114
14	rs12821256	KITLG	−0.434962	0.020898	−0.1743193	0.32142	−6.87E−01	0.056556	0.406751	0.312582
15	rs4959270	EXOC2	−0.251437	0.019073	−0.1555227	0.120958	−2.71E−01	0.104087	−0.34639	0.107774
16	rs12203592	IRF4	1.741377	2.20E−16	1.0810914	2.22E−16	8.80E−01	2.35E 06	0.071132	0.773323
17	rs1042602	TYR	0.125113	0.24551	0.141479	0.155781	−4.52E−02	0.779493	−0.3842	0.071464
18	rs1800407	OCA2	−0.204189	0.331948	−0.0048133	0.97935	−3.53E−01	0.202517	0.223931	0.580501
19	rs2402130	SLC24A4	0.354085	0.00797	0.2752735	0.023746	4.36E−02	0.820086	−0.08861	0.724429
20	rs12913832	HERC2	1.372353	2.20E−16	0.6797949	6.83E−10	1.19E+00	6.65E 13	0.754729	0.004319
21	rs2378249	PIGU/ASIP	0.088319	0.526489	0.1828612	0.154928	−1.64E−01	0.449722	−0.72184	0.002302

Genomic determinants of facial characteristics

OPEN ACCESS Freely available online

PLOS | GENETICS

A Genome-Wide Association Study Identifies Five Loci Influencing Facial Morphology in Europeans

Fan Liu[1], Fedde van der Lijn[1,2,3], Claudia Schurmann[4], Gu Zhu[5], M. Mallar Chakravarty[6,7], Pirro G. Hysi[8], Andreas Wollstein[1], Oscar Lao[1], Marleen de Bruijne[2,3], M. Arfan Ikram[3,9], Aad van der Lugt[3], Fernando Rivadeneira[9,10], André G. Uitterlinden[9,10], Albert Hofman[9], Wiro J. Niessen[2,3,11], Georg Homuth[4], Greig de Zubicaray[12], Katie L. McMahon[12], P… R… …nka Pauso… G… …icking[17], S… T… …Manfred K… G…

1 D… Era… Th… Me… To… Era… Ro… 12… Lo… Gr… Un… Bio…

…r Rotterdam, R… …artment of Radio… …Moritz-Arndt Un… …ronto, Toronto, …, King's College … …epartment of In… …aculty of Applie… …oratory of Neu… …ent of Prosthodo… …iversity Medicin… …rsity of Queens… …al Neurological…

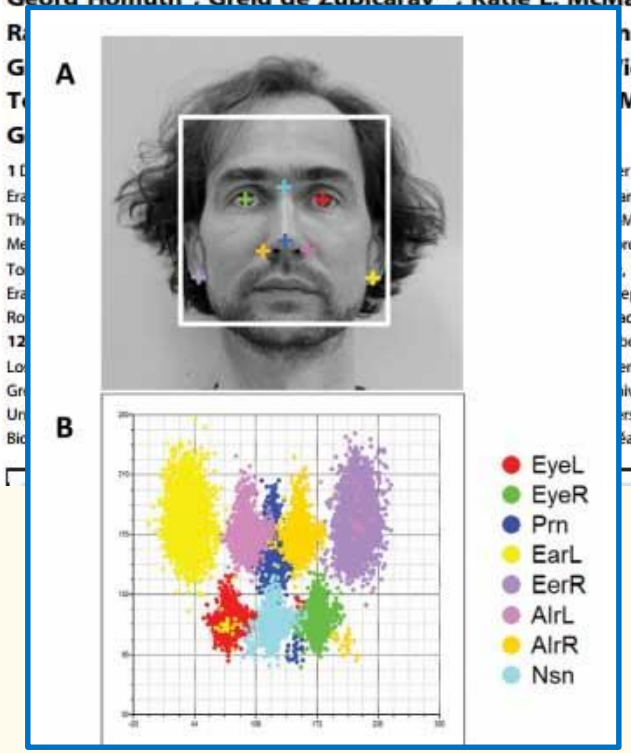
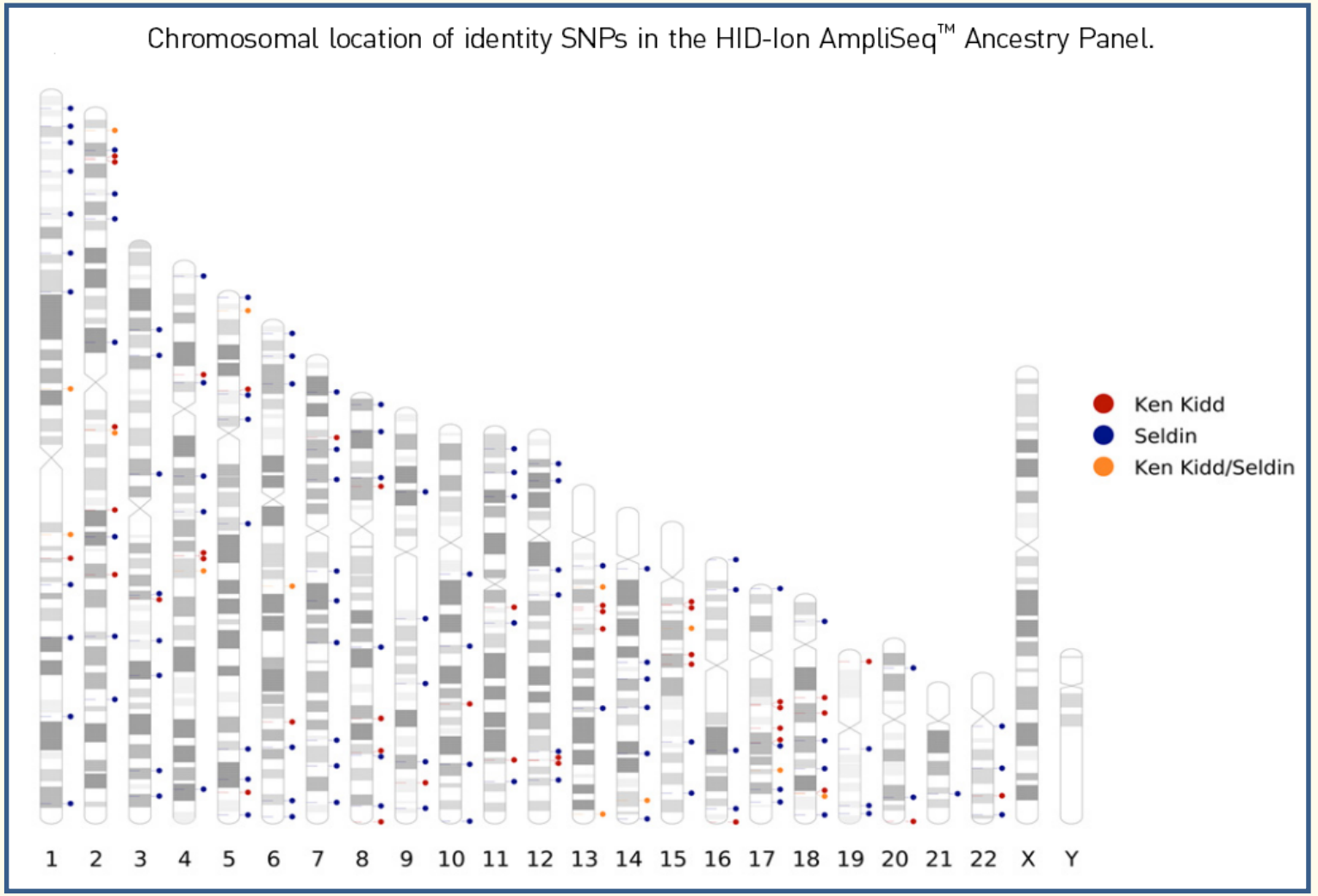


Table 2. SNPs associated with facial shape features from discovery GWAS and their replications.

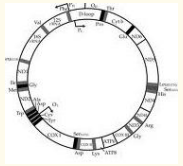
								Discovery (N = 5,388)			SYS (N = 568)				BLTS+TwinSUK (N = 3,867)			
Gene	SNP	Chr	BP	Eff	Alt	FreqEff	Trait*	Beta	Se	P	Beta	se	P	%	Beta	se	P	%
PRDM16	rs4648379	1p36.23-p33	3251376	T	C	0.28	AlrL-Prn	−0.26	0.05	1.13E-08	0.02	0.21	0.930	1.1	0.13	0.09	0.152	0.0
							AlrR-Prn	−0.24	0.05	2.50E-07	−0.04	0.22	0.841		0.15	0.09	0.096	
PAX3	rs974448	2q35	222713558	G	A	0.17	EyeR-Nsn	0.29	0.05	1.56E-08	−0.19	0.20	3.6E-01	1.0	0.10	0.13	0.438	3.6
							EyeL-Nsn	0.29	0.05	7.06E-08	0.06	0.14	6.6E-01		0.21	0.12	0.076	
TP63	rs17447439	3q28	191032117	G	A	0.04	EyeR-EyeL	-0.91	0.15	4.44E-08	−0.42	0.68	5.4E-01	6.4	−0.56	0.27	**0.043**	21.4
C5orf50	rs6555969	5q35.1	171061069	T	C	0.33	ZygR-Nsn	0.41	0.07	1.17E-09	0.31	0.14	**3.2E-02**	16.0	—	—	—	17.9
							ZygL-Nsn	0.35	0.07	5.80E-07	0.39	0.14	**5.6E-03**		—	—	—	
							EyeR-Nsn	0.24	0.04	2.05E-08	0.42	0.12	**3.7E-04**		0.06	0.10	0.590	
							EyeL-Nsn	0.26	0.04	2.28E-09	0.47	0.12	**7.5E-05**		0.21	0.10	**0.031**	
COL17A1	rs805722	10q24.3	105800390	T	C	0.19	EyeL-Nsn	0.29	0.05	3.97E-08	0.54	0.16	**5.9E-04**	18.1	0.08	0.10	0.510	0.0
							EyeR-Nsn	0.26	0.05	6.47E-07	0.51	0.15	**9.7E-04**		−0.23	0.13	0.074	

Chromosomal SNPs in Ancestry Panel

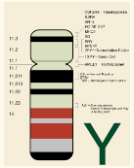


mtDNA & Chr-Y DNA

as species and ancestry identification markers

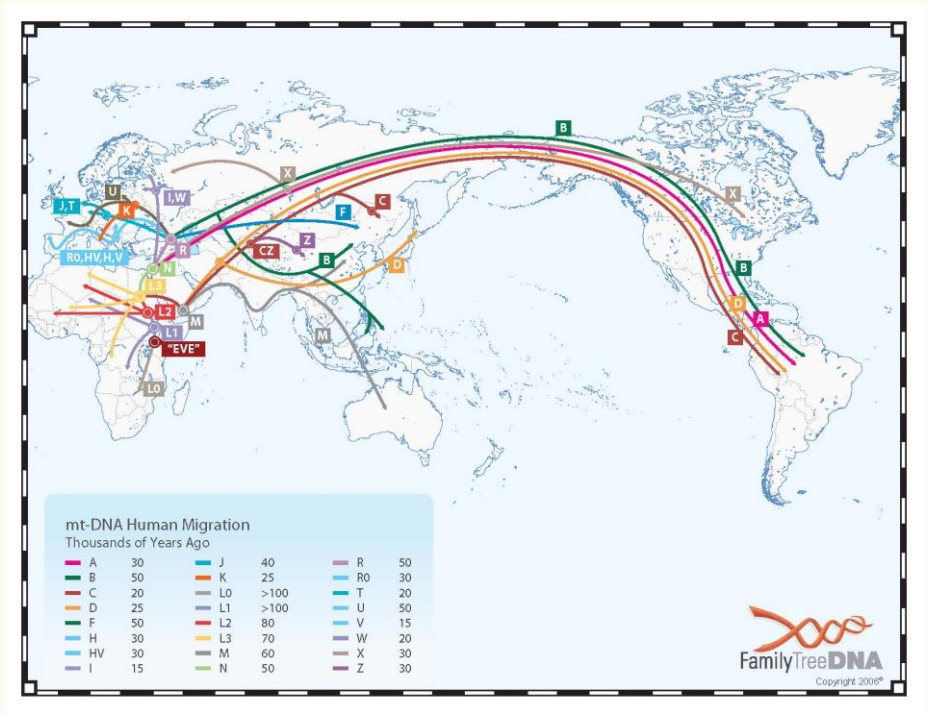


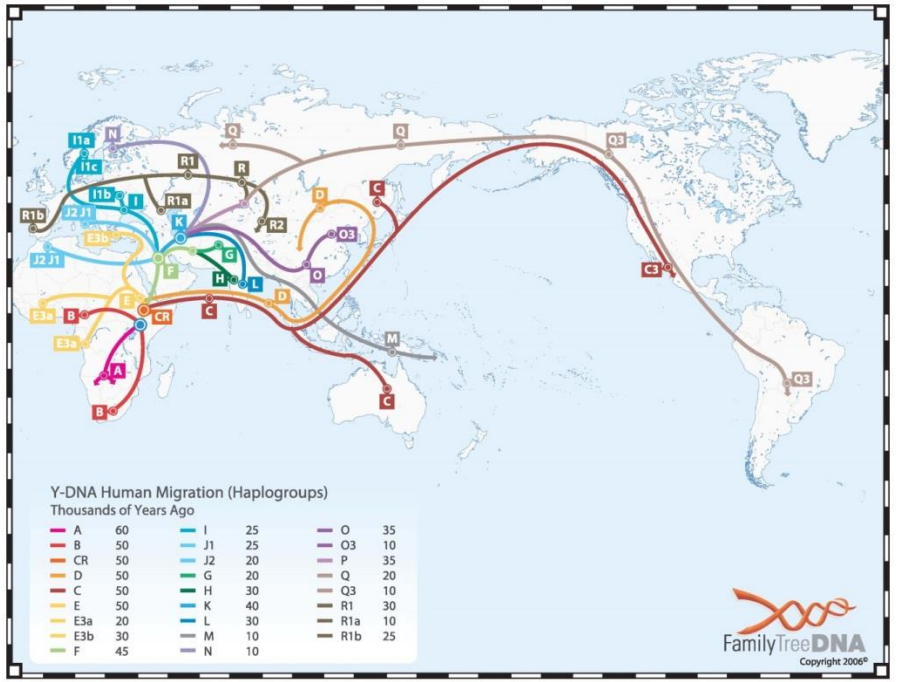
mtDNA in most species is maternally inherited, matching individuals with “recent” ancestry in haplogroups different in various populations.



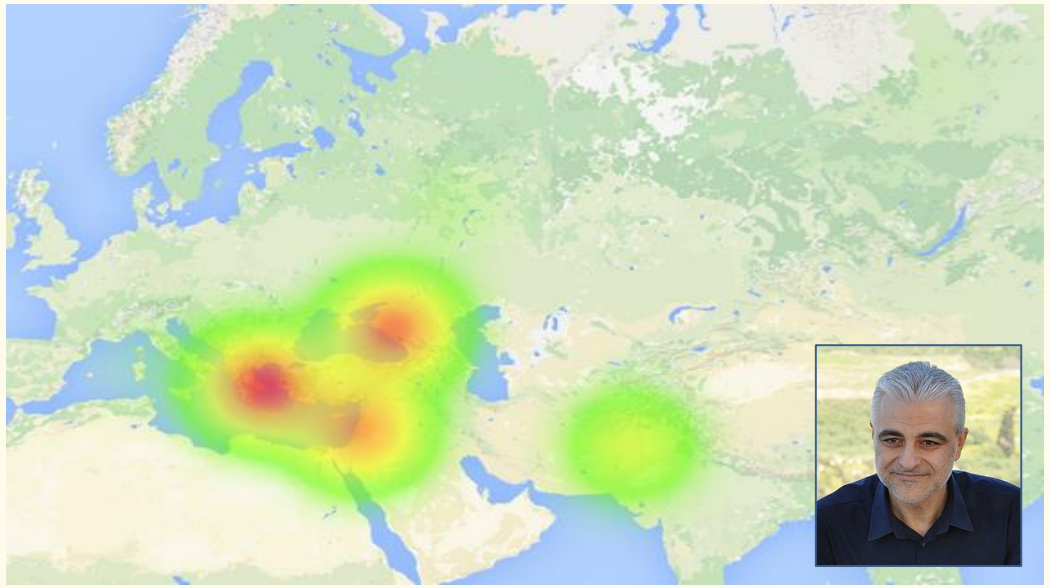
Chr-Y is a marker of patrilinear ancestry with Short Tandem Repeats (STRs) and Single Nucleotide Polymorphisms (SNPs) markers defining also haplogroups.

Deduced human migration map based on mtDNA & Chr -Y

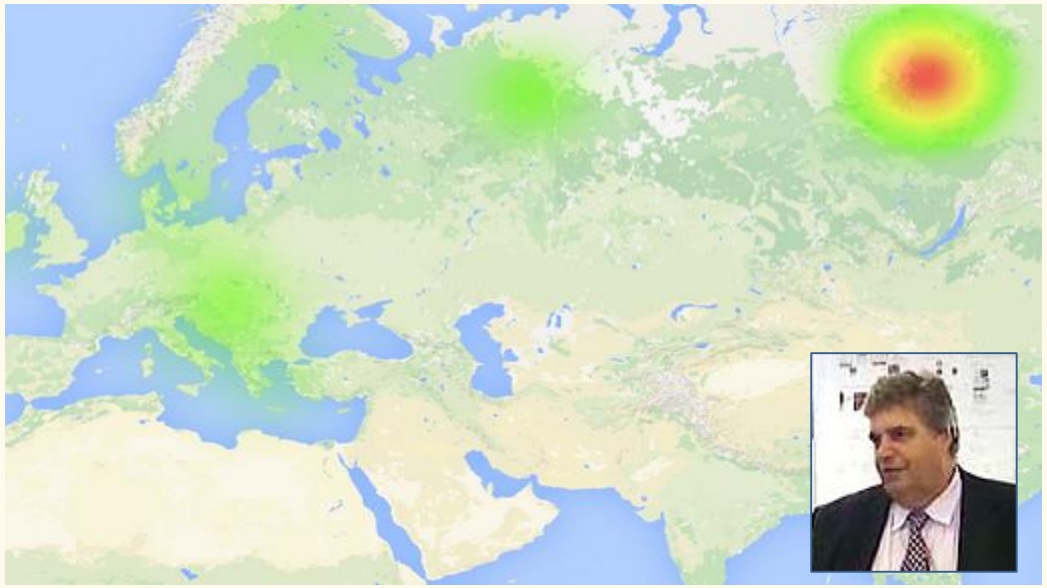




Ancestry tests

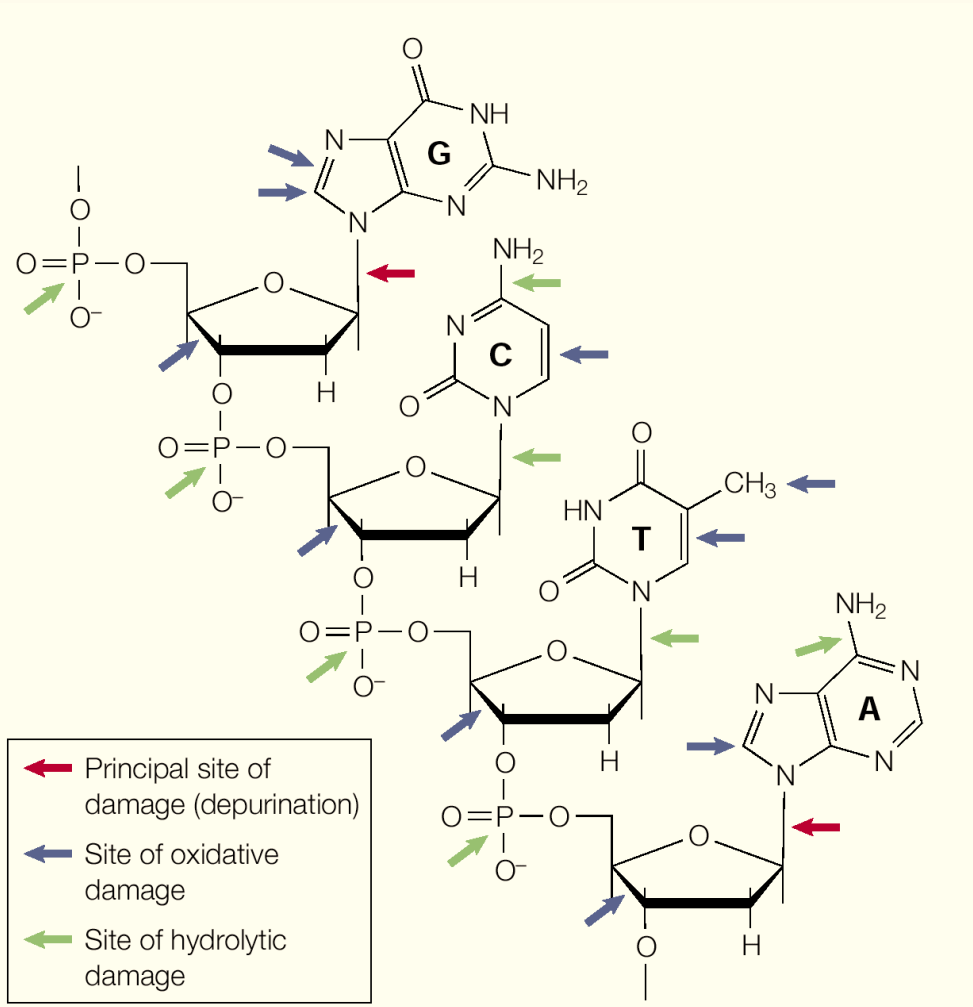


Population Name	Geographic Location	Sample Size	Likelihood
Greeks	Europe	104	1.65E-13
Adygei	Europe	108	1.45E-13
Druze	Asia	212	1.22E-13
Jews, Ashkenazi	Europe	166	7.76E-14
Pashtun	Asia	184	5.04E-14
Chuvash	Europe	84	6.38E-15
Sardinian	Europe	68	3.75E-15
Samaritans	Europe	82	3.00E-15
Italians	Europe	172	2.48E-15
Jews, Sephardic	Europe	54	1.21E-15



Population Name	Geographic Location	Sample Size	Likelihood
Russians	Europe	96	7.02E-14
Russians	Europe	68	1.15E-14
Komi-Zyrian	Asia	94	8.45E-15
Hungarian	Europe	184	6.83E-15
Greeks	Europe	104	3.07E-15
Danes	Europe	102	3.02E-15
Finns	Europe	72	2.87E-15
Irish	Europe	232	2.55E-15
Italians	Europe	172	1.92E-15

Ancient DNA Damages



Hofreiter *et al* *Nat Rev Genet*. 2001 **2**:353

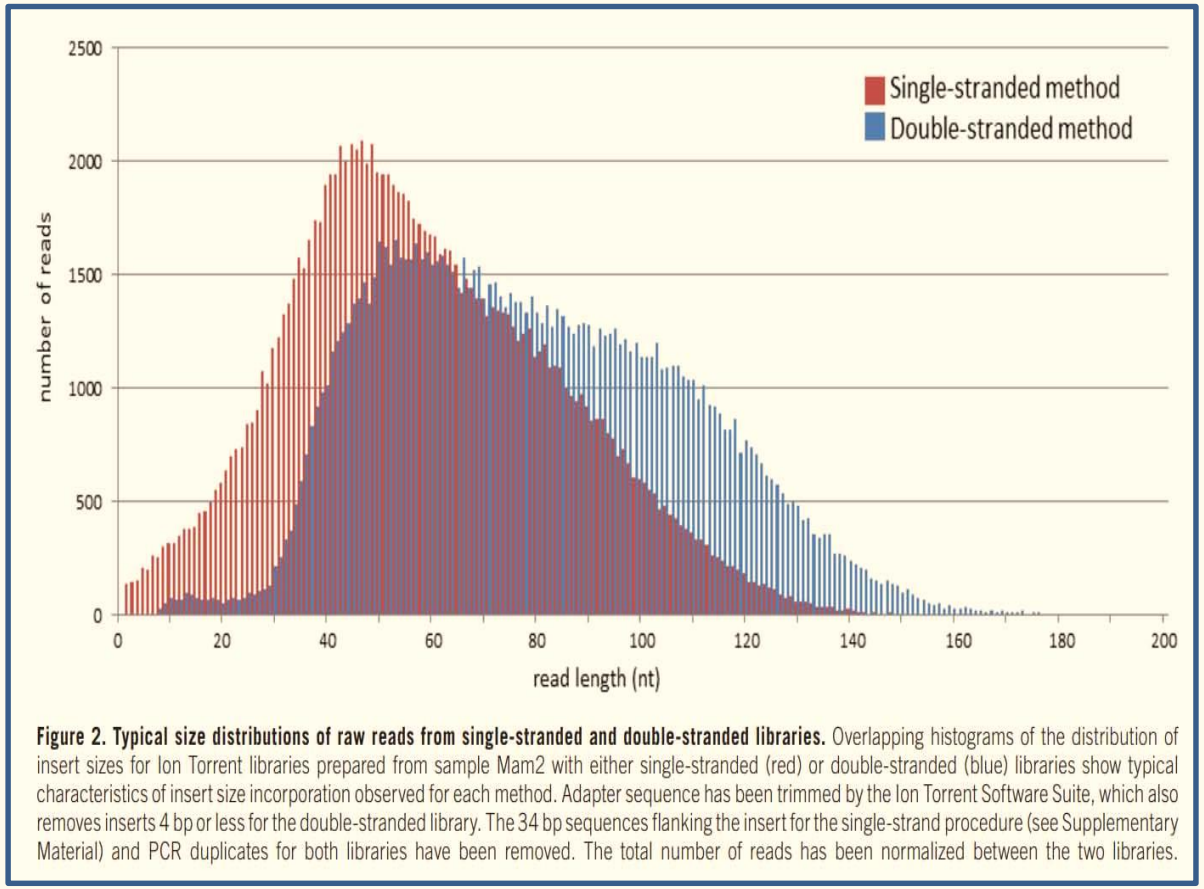


Figure 2. Typical size distributions of raw reads from single-stranded and double-stranded libraries. Overlapping histograms of the distribution of insert sizes for Ion Torrent libraries prepared from sample Mam2 with either single-stranded (red) or double-stranded (blue) libraries show typical characteristics of insert size incorporation observed for each method. Adapter sequence has been trimmed by the Ion Torrent Software Suite, which also removes inserts 4 bp or less for the double-stranded library. The 34 bp sequences flanking the insert for the single-strand procedure (see Supplementary Material) and PCR duplicates for both libraries have been removed. The total number of reads has been normalized between the two libraries.

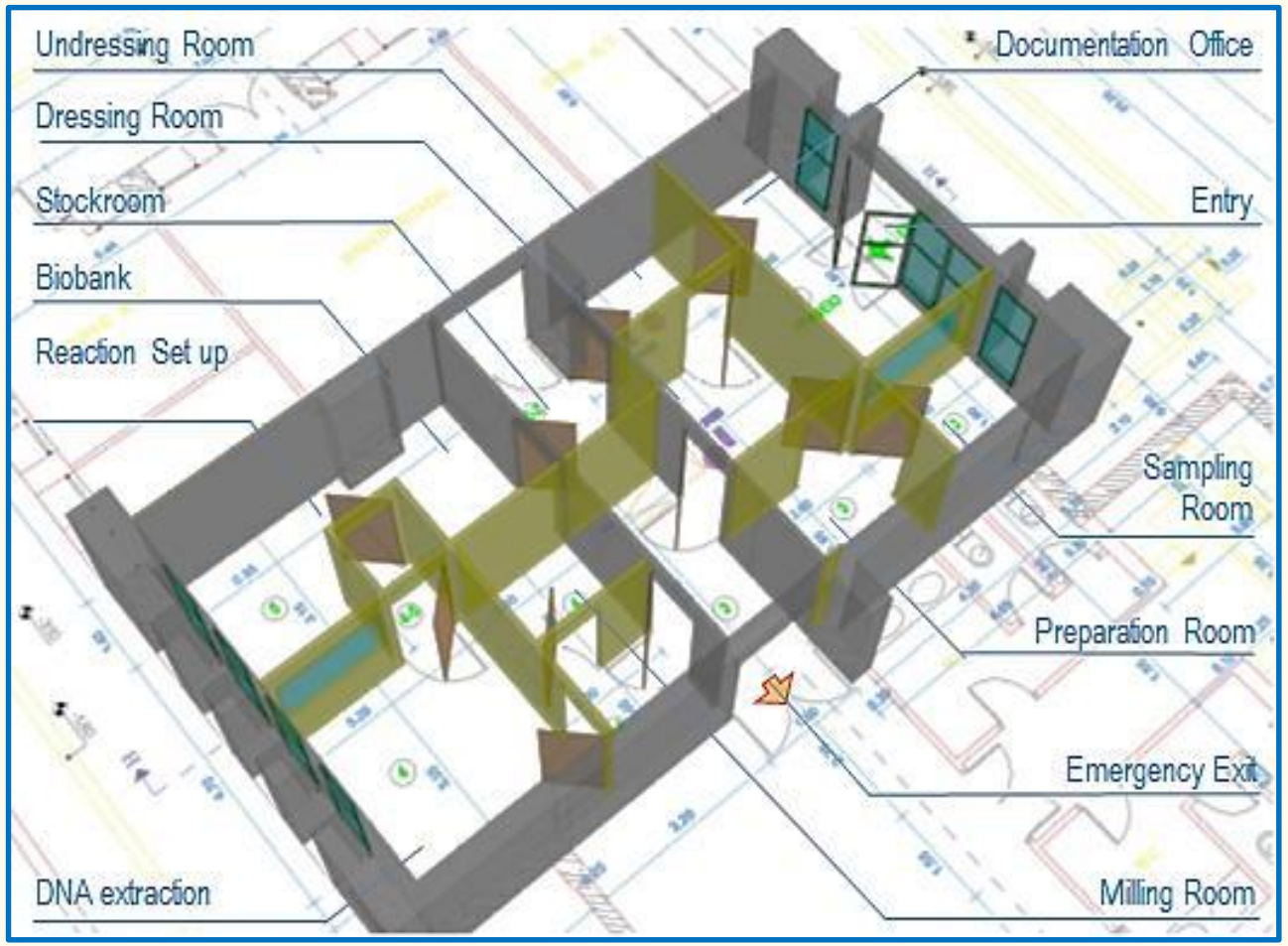
Bennett *et al* *BioTechniques 2014* ***56***:289

Ancient DNA sequencing challenges

The IMBB Ancient DNA Laboratory

ΠΕΡΙΦΕΡΕΙΑ ΚΡΗΤΗΣ
REGION OF CRETE

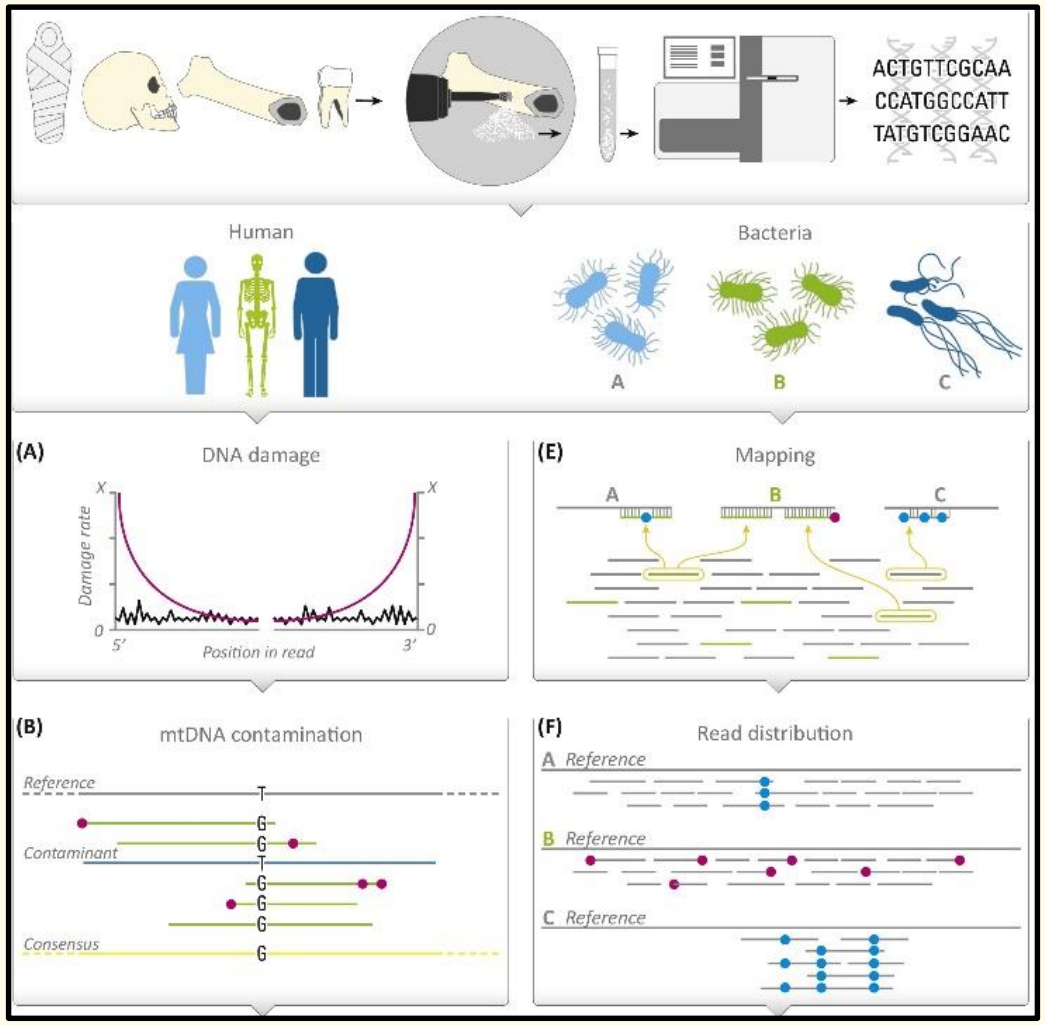
LAB ARRANGEMENT



LAB SPECIFICATIONS

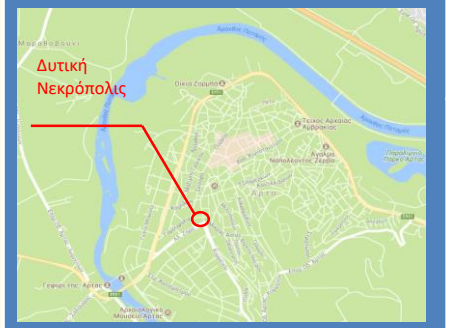
- Building distant from all DNA analysis *(DNA free)*.
- Insulated rooms with positive pressure gradient
- HEPA filter ventilation
- Central vacuum cleaning
- Furniture & clothing made from non-biological materials
- Exclusive equipment
- Bleach-peroxide & UV light decontamination
- Controlled & restricted access
- Keeping detailed protocol of interventions
- Following standard and certified procedures
- Use forensics grade consumables

aDNA NGS Workflow

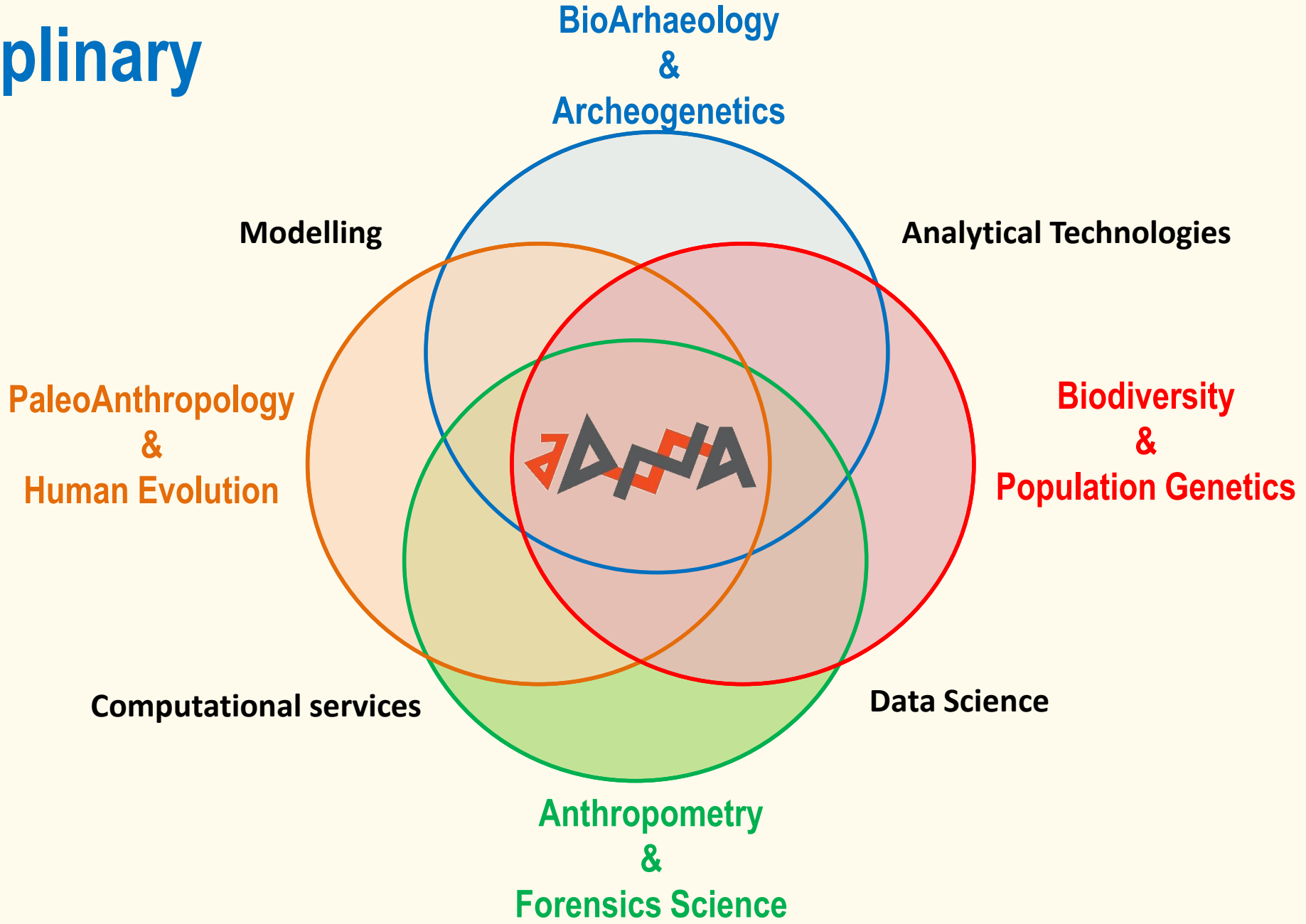


aDNA sequencing analysis

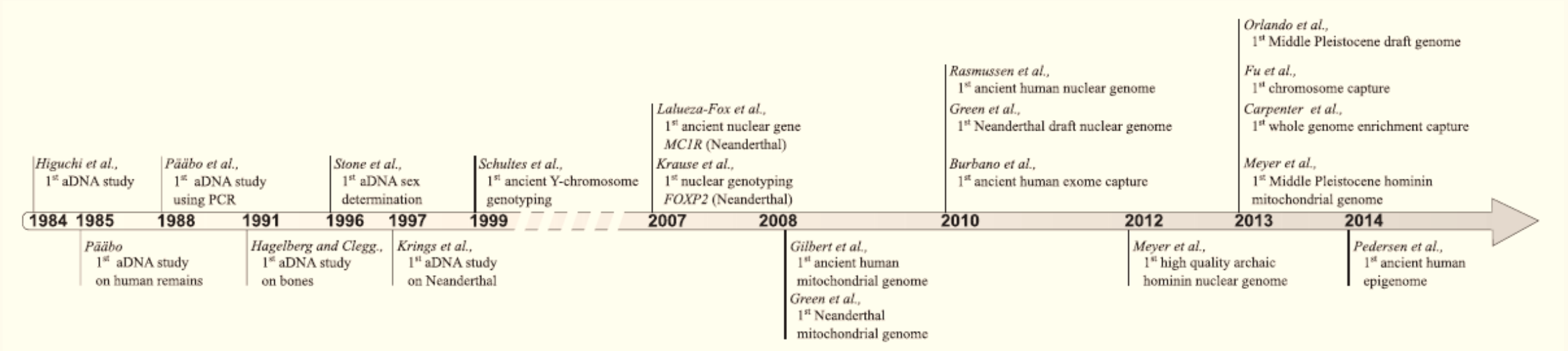
of archeological remains from Amvrakia

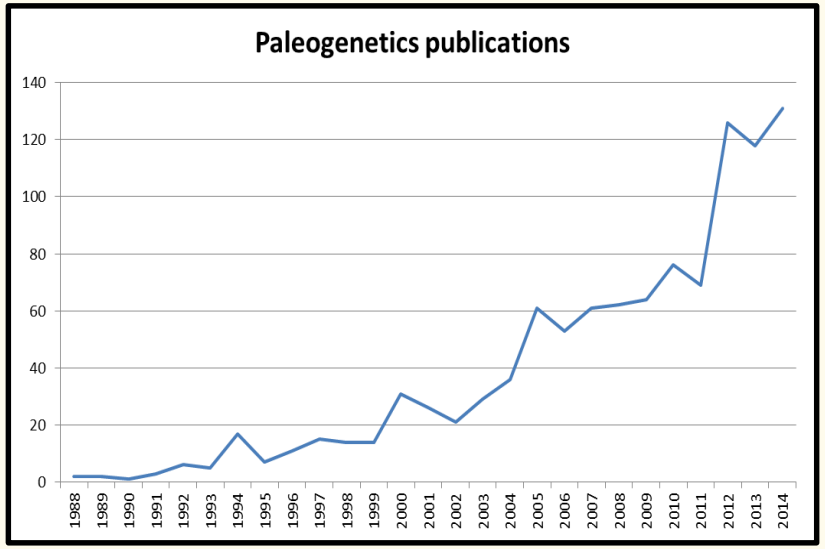
AMVRAKIA WEST NECROPOLIS	Tomb	Sample type	Average Read Length	Total No of Reads	Mapped Reads (hs 37)	% Human Reads	Gender
	CCLXXII, East Support Wall 15/11/12	Cuspid 27	65,7	24.254.094	437.598	1,92	Male
	CXXII, West Support Wall 11/10/12	Molar 16	57,6	46.713.285	984.268	2,23	Female
	CXXV, West Support Wall 24/10/12	Petrous Bone R	56,5	24.262.723	9.108.559	39,28	Female

Multidisciplinary Field of aDNA

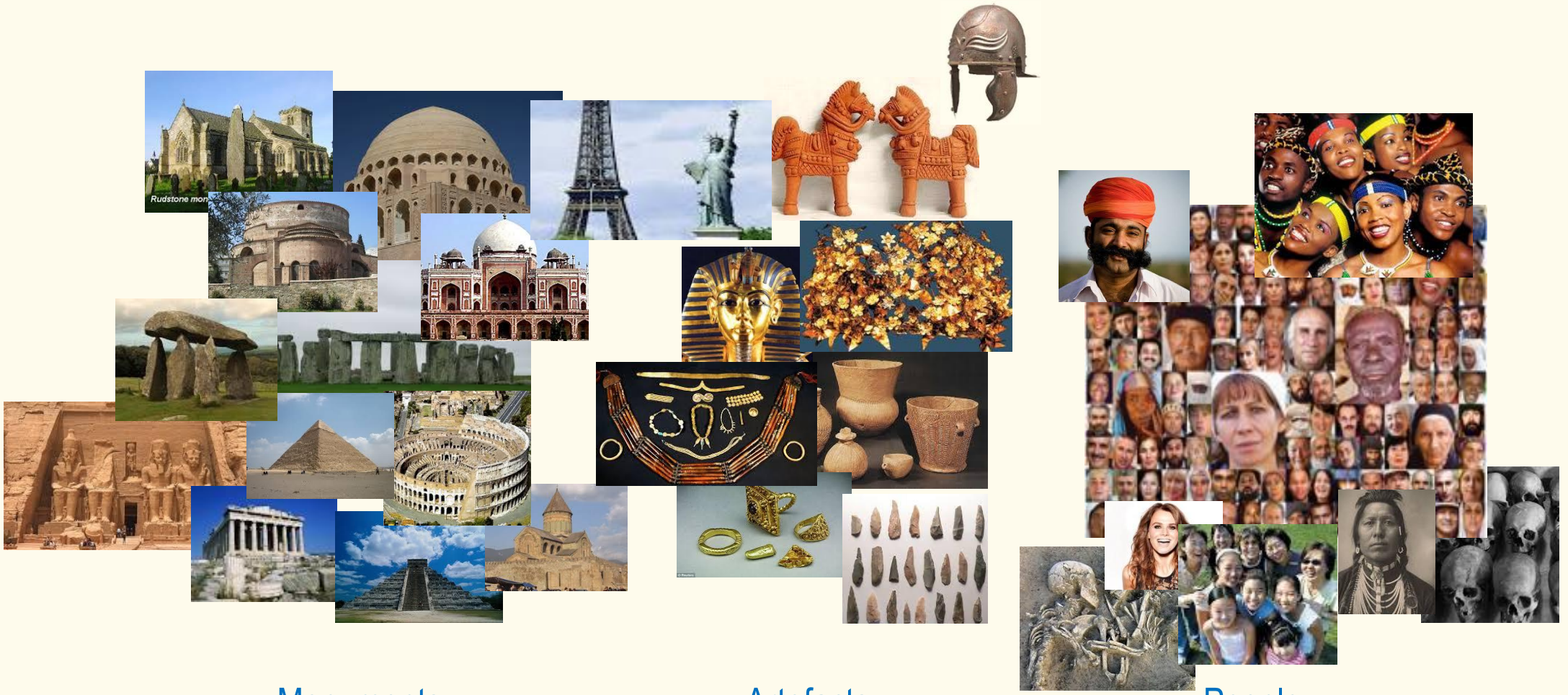


Ancient DNA and Palaeogenetics explosion





Tangible Cultural Heritage Assets



Monuments

Artefacts

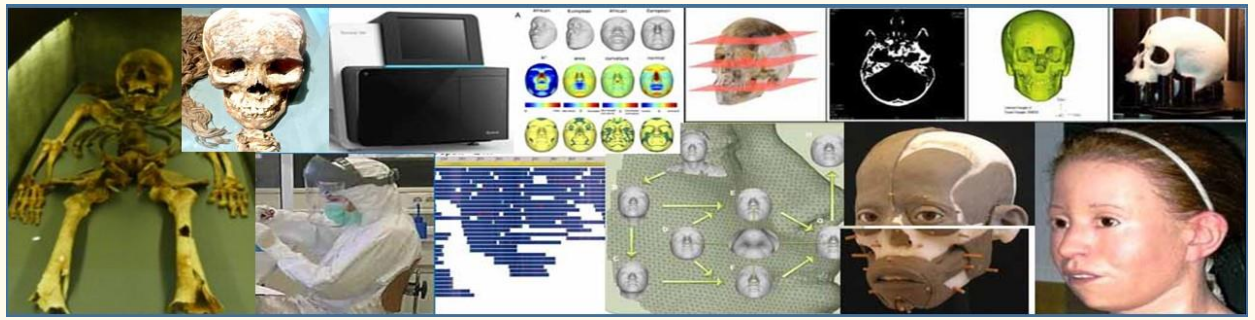
People

Research Plans

- Studies for elucidating origins, migrations, environment and relations of ancient populations: Focus on the Corinthian Colonization

(proposal submitted)

- Genetic characterization, Digitization & Reconstruction of ancient individuals



(proposal in preparation)

- Exploration & modelling of genetic markers as a means for spatiotemporal classification of human archeological remains

Perspectives

- Establish a state-of-the-art Archeogenetics facility within the framework of the National and European Research Infrastructures of Cultural Heritage
- Develop advanced analytical methods, handling protocols and interpretation services, addressing the needs of the Archeological community and unravelling the huge wealth of Greece in archeological findings
- Promote multidisciplinary approaches in Archeogenetics, Bioarcheology and Paleontology research, expand towards Archeoproteomics and provide access to an advanced, competence center facility
- Bridge the gap between humanities and biotechnologies and work towards the integration of highly heterogeneous and distributed Archeological research data

HELLAS-CH
National Research Infrastructure

E-RIHS
EUROPEAN RESEARCH INFRASTRUCTURE FOR HERITAGE SCIENCE

Acknowledgements

Members

- Eirini Stratidaki
- Despina Vassou
- Eugenia Tabakaki
- Nikos Psonis
- Christiana Choulaki

Collaborators

- Eva-Maria Geigl, JMI-CNRS, FR
- Morten Erik Allentoft, Geogenetics, DK
- Varvara Papadopoulou, Arch. Ephorate of Arta, GR
- Manolis Maravelakis, Tech. Un. of Crete, Chania GR
- Manolis Papagrigorakis, Sch. of Dentistry, Un. Athens, GR

Funding Agencies & Projects

ΠΕΡΙΦΕΡΕΙΑ ΚΡΗΤΗΣ
REGION OF CRETE

EMBO

PARTHENOS
Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies

ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη

ΚΡΗΠΙΣ
POLITEIA
ΠΟΛΙτισμός - ΤΕχνολογΙΑ

HELLAS-CH
National Research Infrastructure

FORTH TO THE PAST!!

From Biotech to Cultural Heritage:
A leap forward !

IMBB LABORATORY

Dimitris Kafetzopoulos
kafetzo@imbb.forth.gr

11η επιστημονική ΔΙΗΜΕΡΙΔΑ ITE 2017

-Coalescence-

Wright-Fisher Model for N = 6

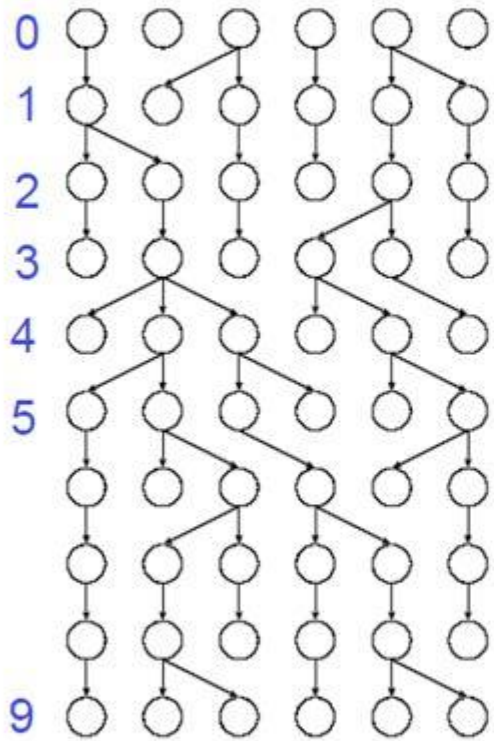


Figure 1: Forward process in the WFM.

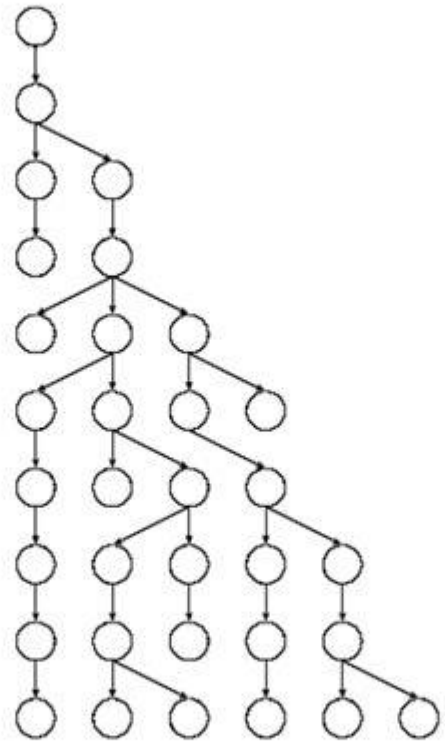


Figure 2: Pruned forward process in the WFM.

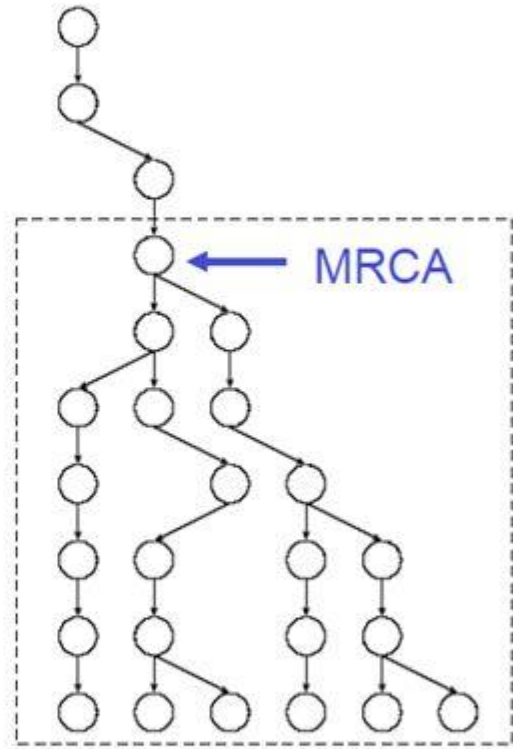
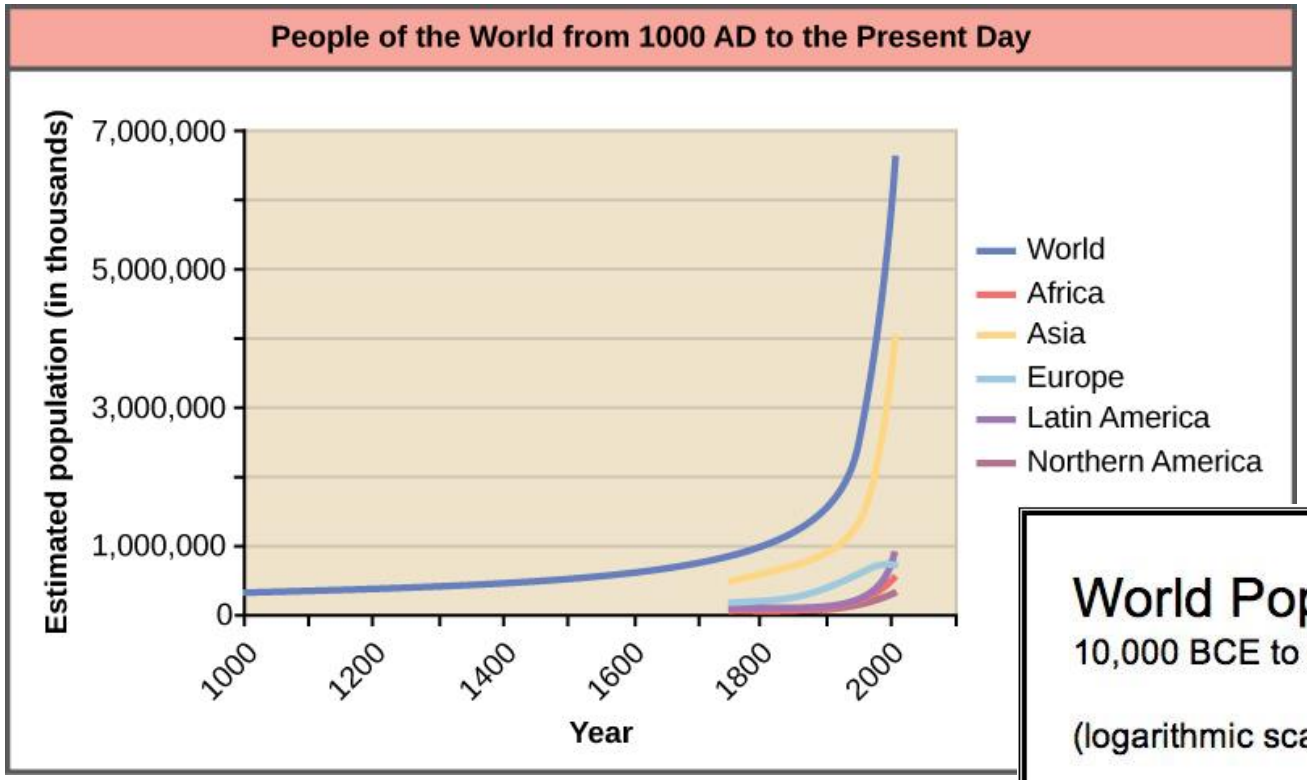
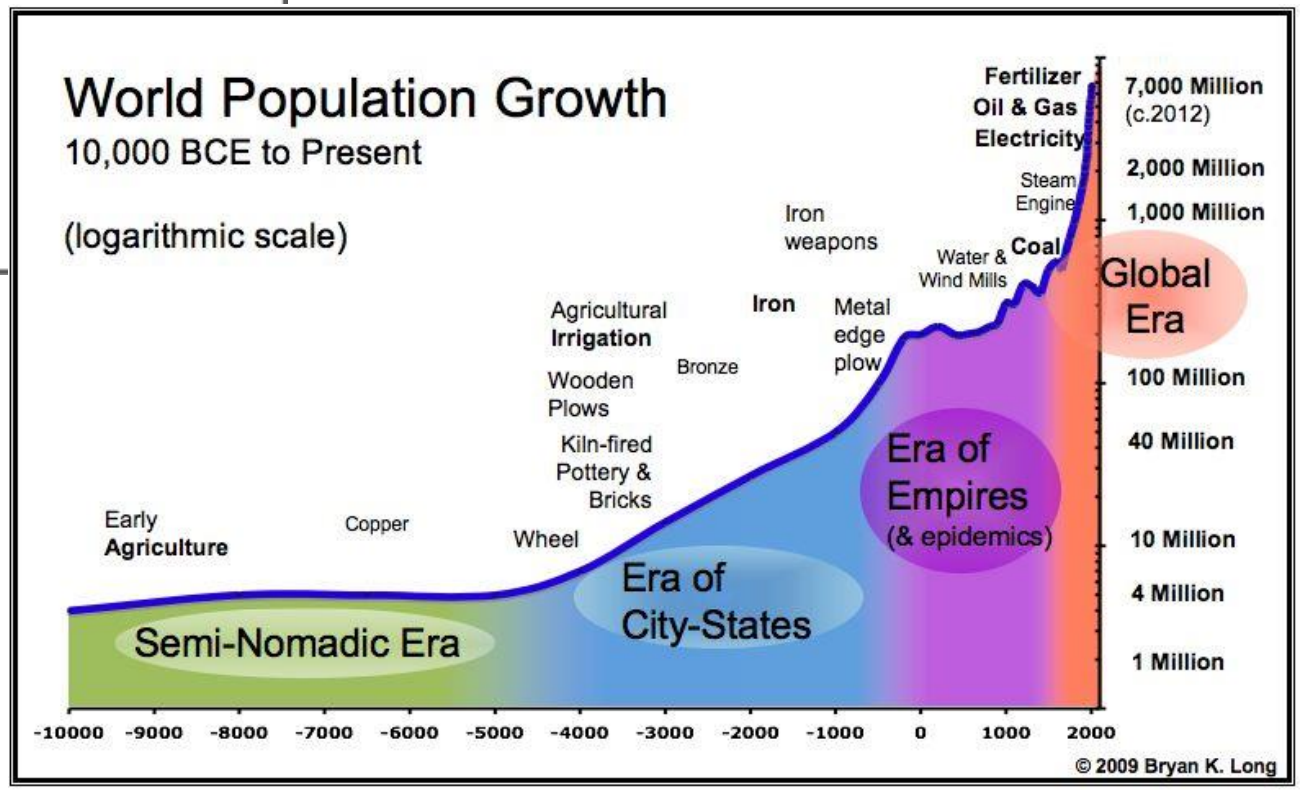


Figure 3: Relatives of alleles present in the final generation.

MRCA = Most Recent Common Ancestor


People of the World from 1000 AD to the Present Day
Estimated population (in thousands)
7,000,000
5,000,000
3,000,000
1,000,000
0
1000
1200
1400
1600
1800
2000
Year
World
Africa
Asia
Europe
Latin America
Northern America


World Population Growth
10,000 BCE to Present
(logarithmic scale)
Fertilizer
Oil & Gas
Electricity
Steam Engine
Coal
Water & Wind Mills
Iron weapons
Iron
Metal edge plow
Bronze
Agricultural Irrigation
Wooden Plows
Kiln-fired Pottery & Bricks
Wheel
Copper
Early Agriculture
Global Era
Era of Empires
(& epidemics)
Era of City-States
Semi-Nomadic Era
7,000 Million
(c.2012)
2,000 Million
1,000 Million
100 Million
40 Million
10 Million
4 Million
1 Million
-10000
-9000
-8000
-7000
-6000
-5000
-4000
-3000
-2000
-1000
0
1000
2000
© 2009 Bryan K. Long