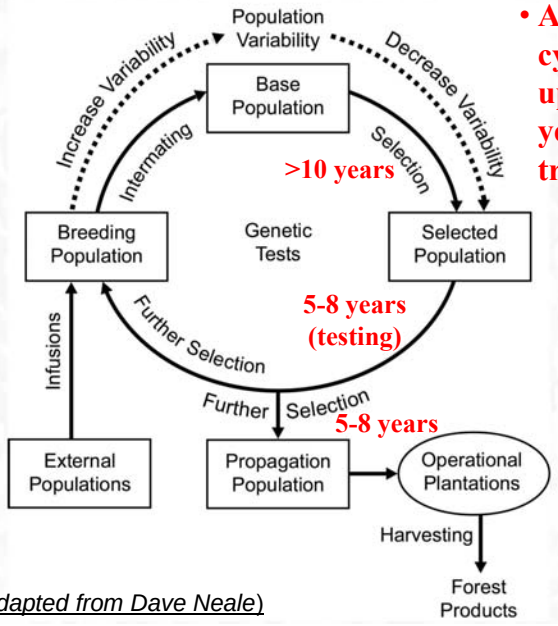
Приложения в геномике

- Genomic selection
- Metagenomics
- Nutrigenomics
- Phylogenomics

Traditional forest tree breeding



- **A full breeding cycle may take up to 20-25 years in forest trees!**

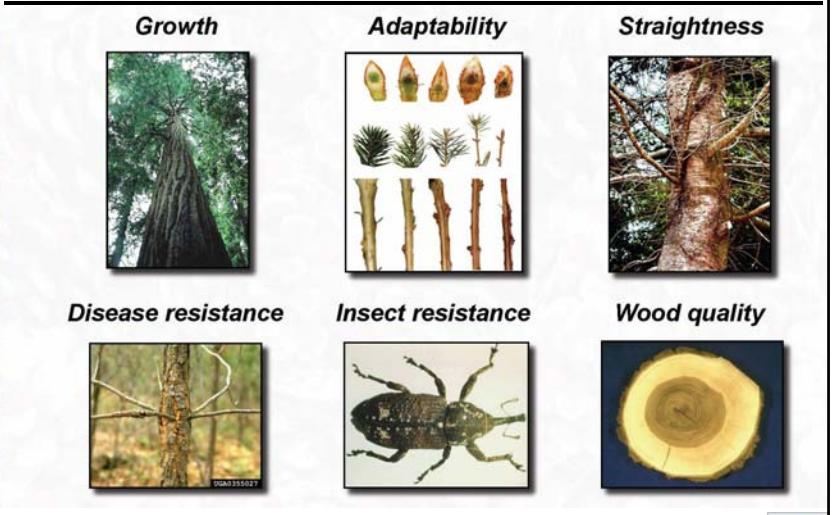
(*Adapted from Dave Neale*)

Traditional forest tree breeding

What we have learned from traditional forest tree breeding:

- Most breeding and adaptive traits are complex quantitative traits controlled by environment and multiple genes of small effect

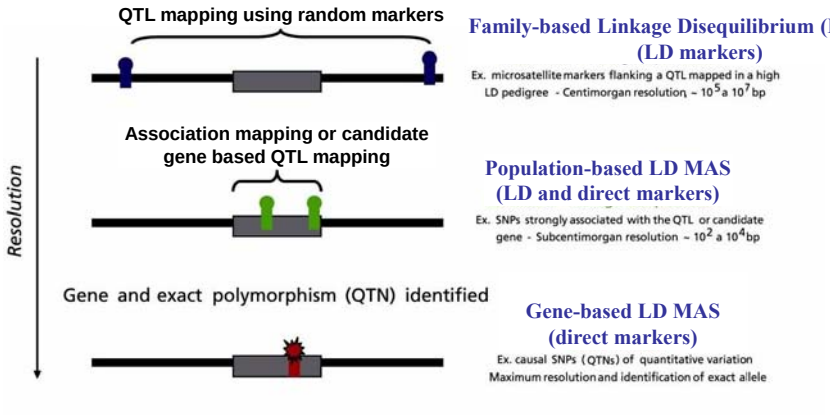
Traditional molecular breeding and Marker-Aided Selection (MAS)



(*Adapted from Dave Neale*)

Traditional molecular breeding and Marker-Aided Selection (MAS)

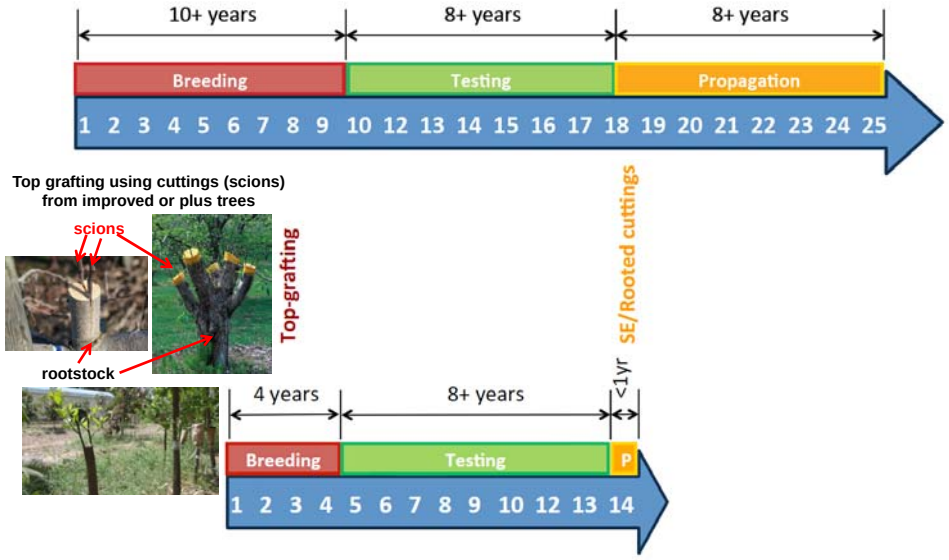
Classification of three different types and resolutions of marker-trait associations:



Modified from Grattapaglia (2007)

5

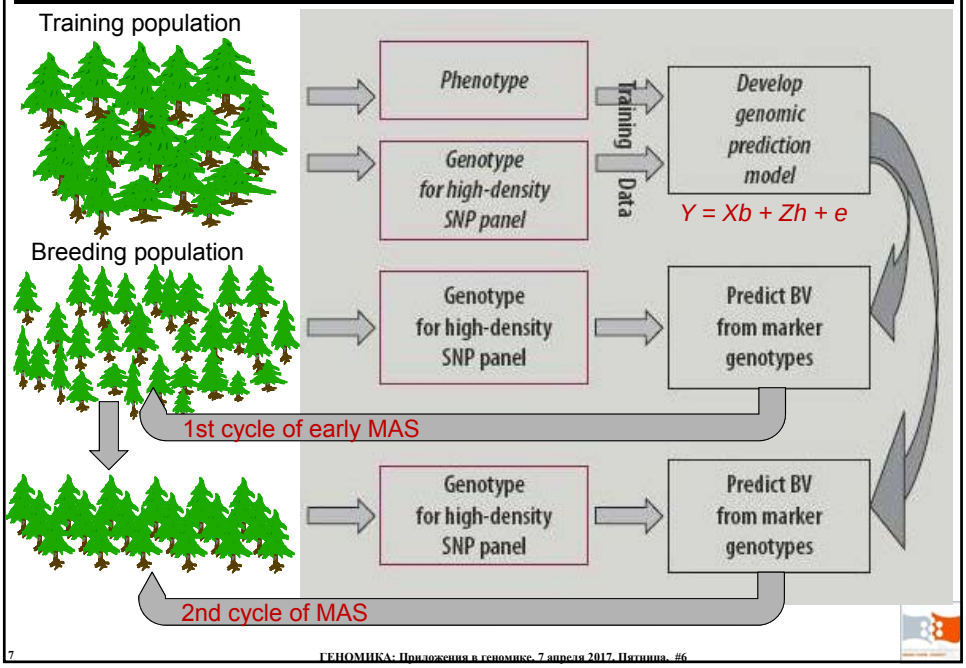
Traditional pine breeding



(*Adapted from Matias Kirst*)

6

Genomic selection

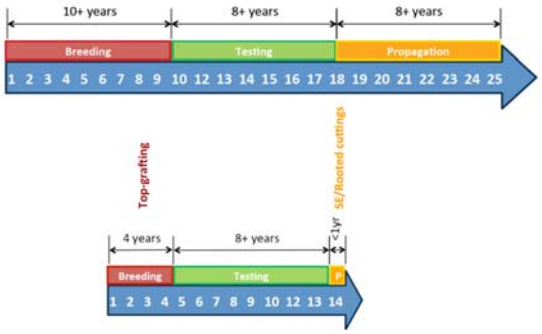


New type of Marker-Aided Selection (MAS):

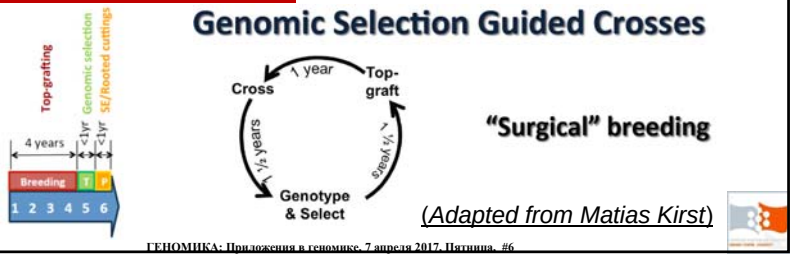
Genome-wide based selection or Genomic selection

- needs genome-wide comprehensive number of markers
- needs efficient high-throughput genotyping
- needs complex regression models to predict phenotypes and breeding values (e.g., GBLUP, Bayes A/B)
- needs high-quality phenotyping
- depends on Linkage Disequilibrium (LD) (ideally – genotyping-by-sequencing – GBS):
  - low LD – more markers are needed;
  - high LD – less number of markers needed

Traditional pine breeding:



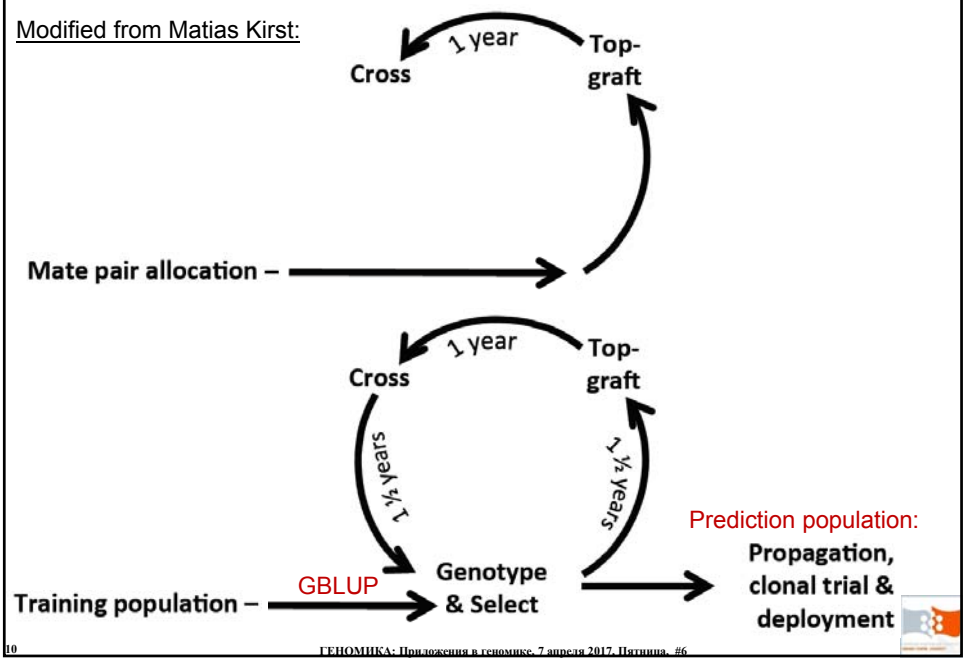
Genomic selection:



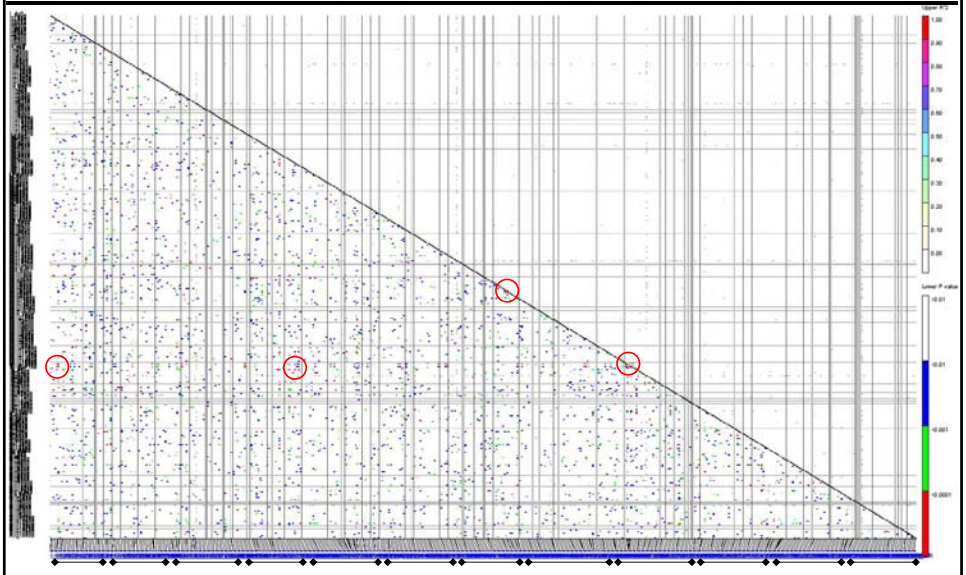
(*Adapted from Matias Kirst*)

Genomic selection incorporated into pine breeding

Modified from Matias Kirst:



Genome-wide linkage disequilibrium (LD)

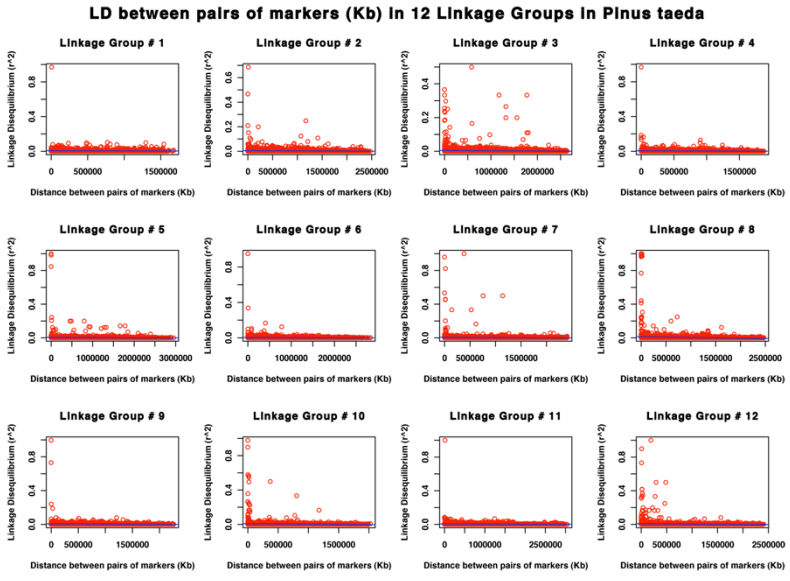


LGs: I II III IV V VI VII VIII IX X XI XII

502 mapped SNPs in 12 loblolly pine linkage groups (I-XII)

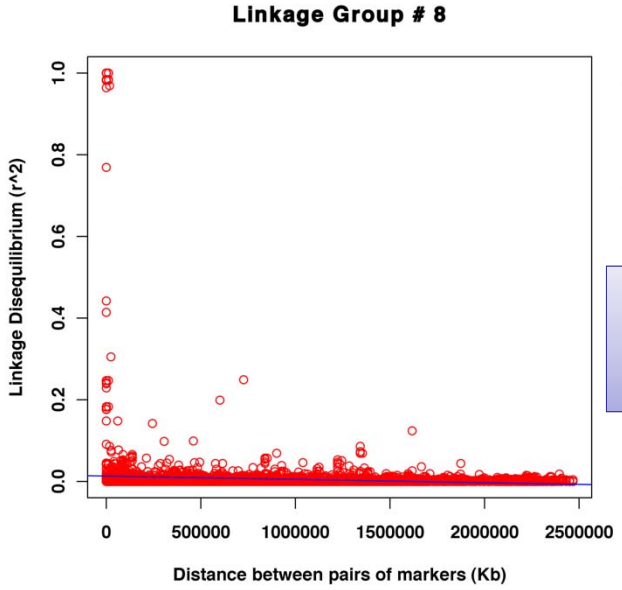
Chhatre, V., T. Byram, D. B. Neale, J. L. Wegrzyn, and K. V. Krutovsky, 2013 Genetic structure and association mapping of adaptive and selective traits in the East Texas loblolly pine (*Pinus taeda* L.) breeding populations. *Tree Genetics and Genomes* 9(5): 1161-1178.

Pairwise LDs for SNPs mapped in 12 loblolly pine linkage groups



Based on Linkage Map by Eckert *et al* (2010)
Mapping Distance to Physical Distance Conversion Based on O'Brien *et al* (1996)

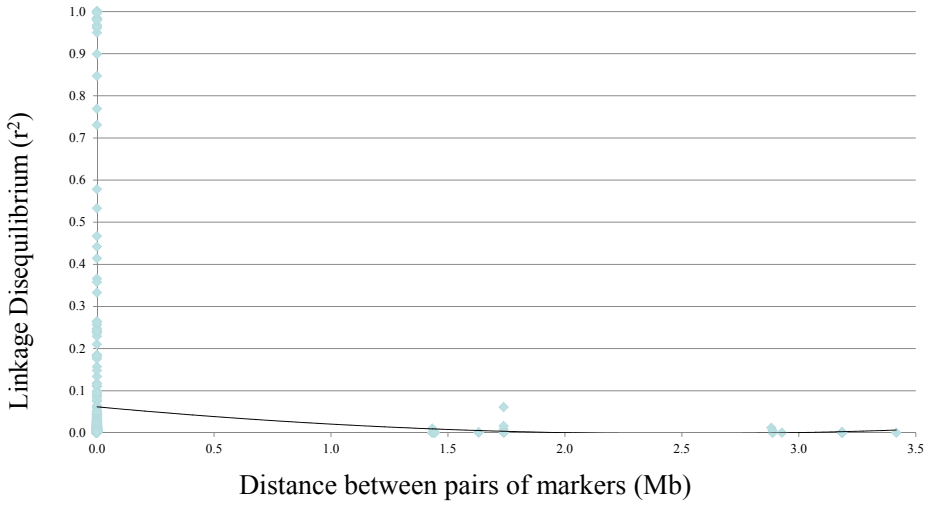
LD in E. Texas Natural Populations



124 Markers
463 Individuals

Low map resolution and SNP density to trace LD decay

LD in the first generation selection trees in Texas (natural populations) Zoomed-in region: 0 – 3.5Mb



Low map resolution and SNP density to trace LD decay

How much LD do we need for genomic selection?

Simulation of probability for detecting association between marker and QTL at different LD levels

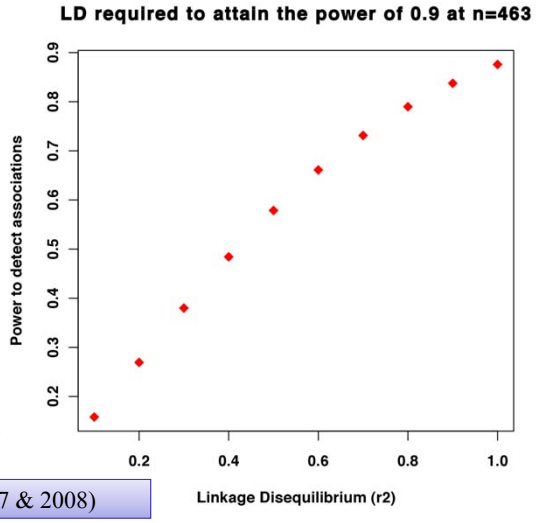
Given conditions:

- Sample size = 463
- Marker minimal allele frequency (MAF) = 0.2
- QTL MAF= 0.2
- Additive effects
- Statistical significance = 95%

Goal:

Power to detect LD between marker & QTL = 0.9

ldDesign R Package (Ball 2007 & 2008)



How many samples do we need for genomic selection?

Simulation of **LD and sample size needed for genome-wide association mapping**

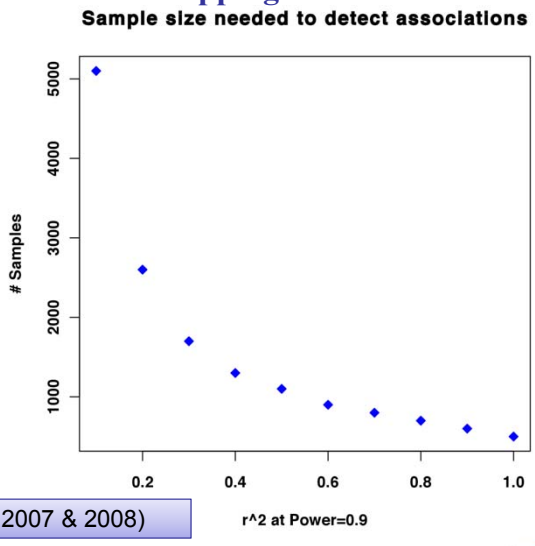
Given conditions:

- Power = 0.9
- Marker MAF = 0.2
- QTL MAF = 0.2
- Additive Effects
- Significance = 95%
- R^2 = 0.1 to 1.0

Goal:

Estimate sample size needed at a given level of LD

ldDesign R Package (Ball 2007 & 2008)



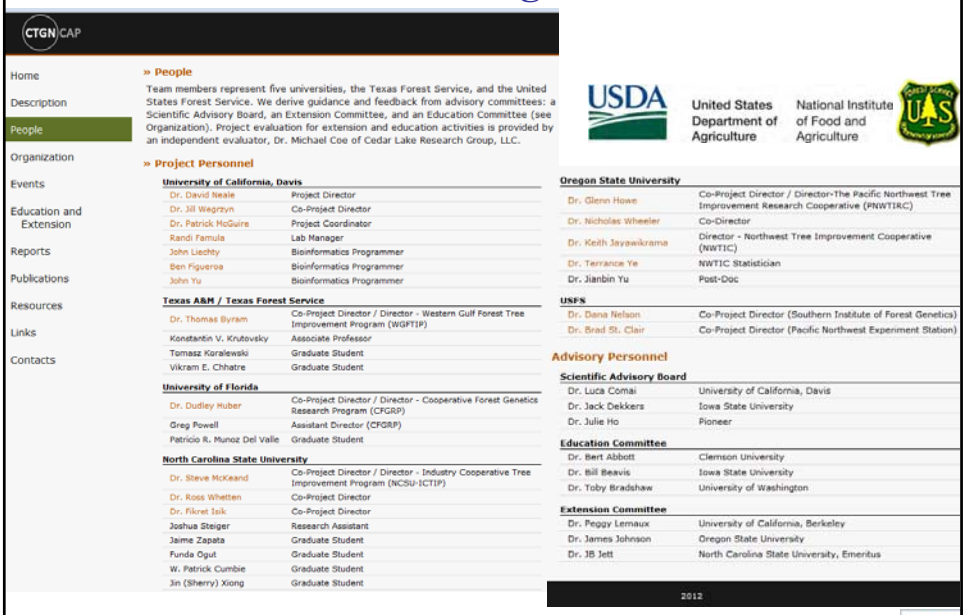
Pairwise relationship (kinship) coefficients for all 1,548 trees in the first (G1) and second (G2) generation selections based on 4,187 SNP markers (partial 1548 x 1548 matrix)

ALL LOCI	T0062	T0159	T0160	T0162	T0163	T0164	T0166	T0169	T0170	T0171	T0172	T0173	T0174	T0175	T0177	T0178	T0179	T0180	T0182	T0184	T0185	T0187	T0188	T0330	T0397	T0419	T0419
T0062		0.004	-0.134	-0.178	-0.082	-0.066	0.005	-0.038	-0.153	-0.096	-0.095	-0.036	-0.143	-0.090	-0.084	-0.199	-0.169	-0.166	-0.058	-0.074	-0.108	-0.108	-0.148	-0.081	-0.037	-0.033	-0.027
T0159	0.004		0.022	-0.029	0.022	0.048	-0.012	0.033	-0.004	0.051	0.007	0.066	-0.041	0.079	0.025	-0.098	-0.008	-0.029	0.036	-0.032	0.017	-0.002	-0.089	-0.073	0.013	-0.092	-0.089
T0160	-0.134	0.022		-0.054	-0.028	-0.023	-0.084	-0.066	-0.042	-0.046	-0.083	0.037	-0.064	0.109	-0.048	-0.076	-0.081	-0.093	0.009	-0.019	-0.016	0.049	-0.051	-0.172	-0.040	-0.163	-0.155
T0162	-0.178	-0.029	-0.054		-0.024	-0.057	-0.115	-0.037	0.464	0.454	0.404	0.438	-0.073	-0.026	0.523	0.131	0.487	-0.017	0.016	0.021	0.028	0.202	0.237	-0.162	-0.138	-0.172	-0.184
T0163	-0.082	0.022	-0.028	-0.024		-0.022	-0.046	0.530	-0.027	-0.041	-0.054	0.057	-0.083	0.112	-0.001	-0.122	-0.072	-0.091	0.051	0.011	-0.027	-0.056	-0.090	-0.133	-0.040	-0.102	-0.098
T0164	-0.066	0.048	-0.023	-0.057	-0.022		-0.009	-0.022	-0.079	-0.021	0.009	0.041	-0.065	0.038	0.007	-0.061	-0.064	-0.040	0.100	-0.007	-0.042	-0.021	-0.093	-0.104	-0.042	-0.115	-0.105
T0166	0.005	-0.012	-0.084	-0.115	-0.046	-0.009		-0.039	-0.122	-0.080	-0.106	-0.013	-0.171	-0.020	-0.059	-0.175	-0.168	-0.145	-0.051	-0.047	-0.058	-0.108	-0.180	-0.131	-0.045	0.018	0.020
T0169	-0.038	0.033	-0.066	-0.037	0.530	-0.022	-0.039		-0.029	-0.032	-0.030	0.077	-0.073	0.108	-0.006	-0.099	-0.032	-0.034	0.103	0.031	0.011	-0.040	-0.053	-0.136	-0.029	-0.085	-0.073
T0170	-0.153	-0.004	-0.042	0.464	-0.027	-0.079	-0.122	-0.029		0.460	0.441	0.464	-0.045	-0.012	0.478	0.161	0.460	-0.044	-0.020	0.003	0.039	0.187	0.180	-0.137	-0.141	-0.200	-0.203
T0171	-0.096	0.051	-0.046	0.454	-0.041	-0.021	-0.080	-0.032	0.460		0.448	0.533	-0.074	0.010	0.504	0.056	0.457	-0.046	0.006	0.023	0.017	0.214	0.161	-0.149	-0.107	-0.156	-0.153
T0172	-0.095	0.007	-0.083	0.404	-0.054	0.009	-0.106	-0.030	0.441	0.448		0.552	-0.080	0.001	0.492	0.117	0.457	-0.034	-0.004	-0.003	0.004	0.210	0.154	-0.148	-0.158	-0.183	-0.186
T0173	-0.036	0.066	0.037	0.438	0.057	0.041	-0.013	0.077	0.464	0.533	0.552		-0.042	0.064	0.549	0.212	0.506	0.020	0.089	0.069	0.044	0.237	0.214	-0.047	-0.033	-0.098	-0.104
T0174	-0.143	-0.041	-0.064	-0.073	-0.083	-0.065	-0.171	-0.073	-0.045	-0.074	-0.080	-0.042		-0.051	-0.061	-0.140	-0.072	-0.062	-0.012	-0.089	-0.060	0.014	-0.084	-0.209	-0.152	-0.208	-0.204
T0175	-0.090	0.079	0.109	-0.026	0.112	0.038	-0.020	0.108	-0.012	0.010	0.001	0.064	-0.051		0.039	-0.050	-0.033	-0.039	0.098	0.025	0.038	0.071	0.043	-0.065	0.045	-0.071	-0.064
T0177	-0.084	0.025	-0.048	0.523	-0.001	0.007	-0.059	-0.006	0.478	0.504	0.492	0.549	-0.061	0.039		-0.090	-0.011	0.026	0.068	0.035	0.025	-0.006	-0.019	-0.097	-0.075	-0.123	-0.122
T0178	-0.199	-0.098	-0.076	0.131	-0.122	-0.061	-0.175	-0.099	0.161	0.056	0.117	0.212	-0.140	-0.050	-0.090		0.444	-0.065	-0.032	-0.056	0.159	0.479	0.451	-0.169	-0.182	-0.234	-0.255
T0179	-0.169	-0.008	-0.081	0.487	-0.072	-0.064	-0.168	-0.032	0.460	0.457	0.457	0.506	-0.072	-0.033	-0.011	0.444		-0.065	-0.036	-0.006	0.040	0.426	0.443	-0.158	-0.172	-0.218	-0.236
T0180	-0.166	-0.029	-0.093	-0.017	-0.091	-0.040	-0.145	-0.034	-0.044	-0.046	-0.034	0.020	-0.062	-0.039	0.026	-0.065	-0.065		0.004	-0.023	-0.043	-0.044	-0.110	-0.163	-0.156	-0.170	-0.178
T0182	-0.058	0.036	0.009	0.016	0.051	0.100	-0.051	0.103	-0.020	0.006	-0.004	0.089	-0.012	0.098	0.068	-0.032	-0.036	0.004		0.086	0.047	0.019	-0.055	-0.099	0.001	-0.091	-0.092
T0184	-0.074	-0.032	-0.019	0.021	0.011	-0.007	-0.047	0.031	0.003	0.023	-0.003	0.069	-0.089	0.025	0.035	-0.056	-0.006	-0.023	0.086		-0.034	0.020	-0.061	-0.140	-0.046	-0.076	-0.088
T0185	-0.108	0.017	-0.016	0.028	-0.027	-0.042	-0.058	0.011	0.039	0.017	0.004	0.044	-0.060	0.038	0.025	0.159	0.040	-0.043	0.047	-0.034		0.172	0.066	-0.105	-0.104	-0.162	-0.170
T0187	-0.108	-0.002	0.049	0.202	-0.056	-0.021	-0.108	-0.040	0.187	0.214	0.210	0.237	0.014	0.071	-0.006	0.479	0.426	-0.044	0.019	0.020	0.172		0.486	-0.122	-0.088	-0.167	-0.165
T0188	-0.148	-0.089	-0.051	0.237	-0.090	-0.093	-0.180	-0.053	0.180	0.161	0.154	0.214	-0.084	0.043	-0.019	0.451	0.443	-0.110	-0.055	-0.061	0.066	0.486		-0.140	-0.180	-0.230	-0.233
T0330	-0.081	-0.073	-0.172	-0.162	-0.133	-0.104	-0.131	-0.136	-0.137	-0.149	-0.148	-0.047	-0.209	-0.065	-0.097	-0.169	-0.158	-0.163	-0.099	-0.140	-0.105	-0.122	-0.140		-0.148	-0.081	-0.082
T0397	-0.037	0.013	-0.040	-0.138	-0.040	-0.042	-0.045	-0.029	-0.141	-0.107	-0.158	-0.033	-0.152	0.045	-0.075	-0.182	-0.172	-0.156	0.001	-0.046	-0.104	-0.088	-0.180	-0.148		-0.069	-0.065
T0419	-0.033	-0.092	-0.163	-0.172	-0.102	-0.115	0.018	-0.085	-0.200	-0.156	-0.183	-0.098	-0.208	-0.071	-0.123	-0.234	-0.218	-0.170	-0.091	-0.076	-0.162	-0.167	-0.230	-0.081	-0.069		0.999
T0419	-0.027	-0.089	-0.155	-0.184	-0.098	-0.105	0.020	-0.073	-0.203	-0.153	-0.186	-0.104	-0.204	-0.064	-0.122	-0.255	-0.236	-0.178	-0.092	-0.088	-0.170	-0.165	-0.233	-0.082	-0.065	0.999	
T0420	-0.033	-0.102	-0.139	-0.216	-0.128	-0.082	-0.007	-0.123	-0.208	-0.166	-0.145	-0.100	-0.223	-0.067	-0.176	-0.207	-0.196	-0.225	-0.079	-0.105	-0.085	-0.129	-0.189	-0.111	-0.118	-0.078	-0.083
T0421	-0.026	-0.083	-0.110	-0.178	-0.101	-0.154	0.002	-0.134	-0.180	-0.162	-0.165	-0.106	-0.200	-0.049	-0.128	-0.240	-0.220	-0.204	-0.060	-0.071	-0.137	-0.153	-0.212	-0.084	-0.103	0.182	0.187
T0422	-0.083	-0.156	-0.155	-0.222	-0.144	-0.151	-0.048	-0.148	-0.268	-0.182	-0.184	-0.110	-0.209	-0.131	-0.176	-0.277	-0.262	-0.227	-0.140	-0.112	-0.193	-0.185	-0.247	-0.157	-0.179	0.237	0.233
T0423	0.011	-0.115	-0.105	-0.186	-0.144	-0.064	-0.006	-0.113	-0.194	-0.123	-0.141	-0.070	-0.161	-0.050	-0.137	-0.269	-0.212	-0.245	-0.124	-0.113	-0.148	-0.162	-0.214	-0.176	-0.052	-0.023	-0.009
T0424	-0.058	-0.147	-0.130	-0.219	-0.108	-0.106	0.017	-0.120	-0.242	-0.152	-0.186	-0.118	-0.198	-0.106	-0.172	-0.235	-0.241	-0.211	-0.090	-0.158	-0.161	-0.176	-0.242	-0.128	-0.093	0.249	0.251
T0425	-0.053	-0.073	-0.094	-0.130	-0.086	-0.046	-0.091	-0.092	-0.116	-0.049	-0.074	0.005	-0.104	-0.074	-0.047	-0.138	-0.147	-0.120	-0.047	-0.096	-0.056	-0.032	-0.101	-0.100	-0.101	-0.089	-0.074
T0426	-0.035	-0.116	-0.125	-0.224	-0.103	-0.078	-0.032	-0.122	-0.266	-0.184	-0.242	-0.142	-0.205	-0.119	-0.219	-0.249	-0.253	-0.258	-0.108	-0.133	-0.157	-0.168	-0.243	-0.124	-0.083	0.241	0.243
T0427	0.013	-0.088	-0.150	-0.237	-0.146	-0.065	0.033	-0.127	-0.231	-0.187	-0.188	-0.094	-0.255	-0.083	-0.188	-0.251	-0.234	-0.246	-0.097	-0.109	-0.176	-0.138	-0.207	-0.134	-0.052	-0.016	-0.013
T0428	-0.090	-0.048	-0.111	-0.104	-0.086	-0.081	-0.098	-0.088	-0.126	-0.071	-0.093	-0.004	-0.157	-0.055	-0.053	-0.146	-0.121	-0.132	-0.040	-0.076	-0.055	-0.065	-0.155	-0.122	-0.109	0.161	0.165

Genomic selection – Conclusions

- Accuracy of prediction will increase with:
  - more markers
  - more individuals
  - higher heritability
  - higher LD
- It can be done, but most likely in the family based breeding

Acknowledgements



http://dendrome.ucdavis.edu/ctgn/people/

Western Gulf Forest Tree Improvement Program

Texas Forest Service Gene Conservation Program
Forest Science Laboratory, Texas A&M University,
College Station, TX, USA

http://www.ars-grin.gov/misc/wgftip/about.html

- The WGFTIP is a cooperative tree breeding project founded in 1969 with the objective of providing the best genetic quality seed for use in forest regeneration programs in the Western Gulf Region of the United States.
- Base Population: 3300 loblolly & 1000 slash pines.
- Progeny Tests: > 1500, 3 mln trees, 4,000 ac
- Current members include 5 states and 8 industrial members collectively responsible for planting 300,000,000 seedlings per year.
- The cooperative is preserving and improving populations of five southern pine species and several hardwood species

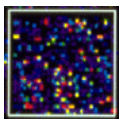
Medical genomics

- can discover disease associated genes
- can discover disease causing genes.
- provides understanding of disease
- provides basis for novel drug development
- provides basis for novel genetic and stem cell therapies
- provides the basis for preventive medicine

Use of genomic information

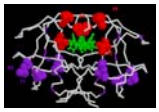
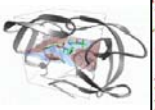
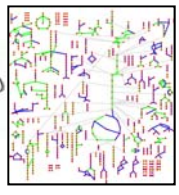
Novel Diagnostics

- Microchips & Microarrays - DNA
- Gene Expression - RNA
- Proteomics - Protein


Novel Therapeutics

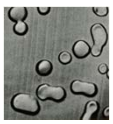
- Drug Target Discovery
- Rational Drug Design
- Molecular Docking
- Gene Therapy
- Stem Cell Therapy

Understanding Metabolism

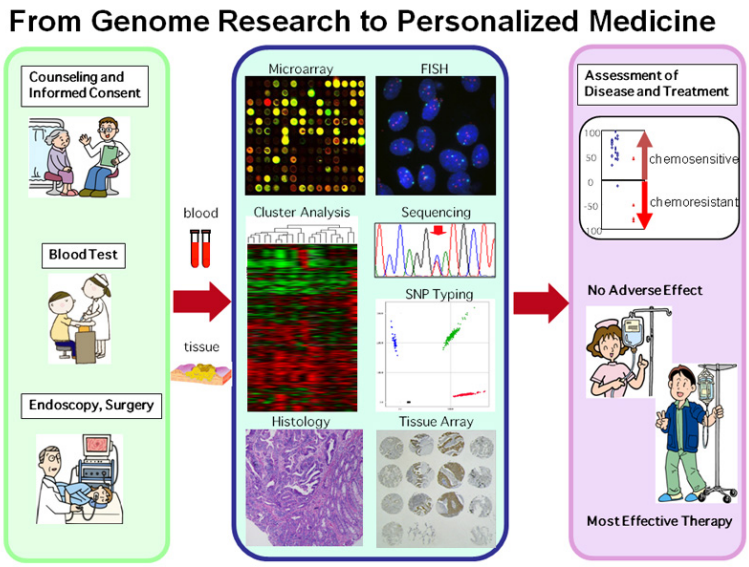
Understanding Disease

- Inherited Diseases - OMIM
- Infectious Diseases
- Pathogenic Bacteria
- Viruses




Personalized genomic medicine

The right treatment, for the right patient, at the right



Examples for complex polygenic diseases & responses

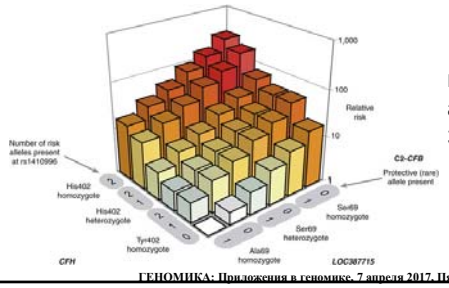
Medullary thyroid cancer & *RET* mutation testing: Multiple Endocrine Neoplasia 2 (MEN2) (If RET +, prophylactic thyroidectomy is offerred)

Predicting toxicity from chemotherapy based on retrospective analysis of clinical trial data. Toxicity and sensitivity depend on thiopurine methyltransferase (TPMT) activity. There is individual genetic polymorphisms that affect this enzymatic activity.

Multiple contributors to **asthma**: *Genetics (*beta-adrenergic receptor, GSTM1, GSTT1, IL-4, IL-4RA, IL-13, TNF-alpha, and 30-50 other genes) + *Environment* (mites, cockroaches, pollens, animal danders, cigarette smoke, diesel fuel)

Estimate of lifetime diabetes risk based on presence/absence of disease-associated mutations

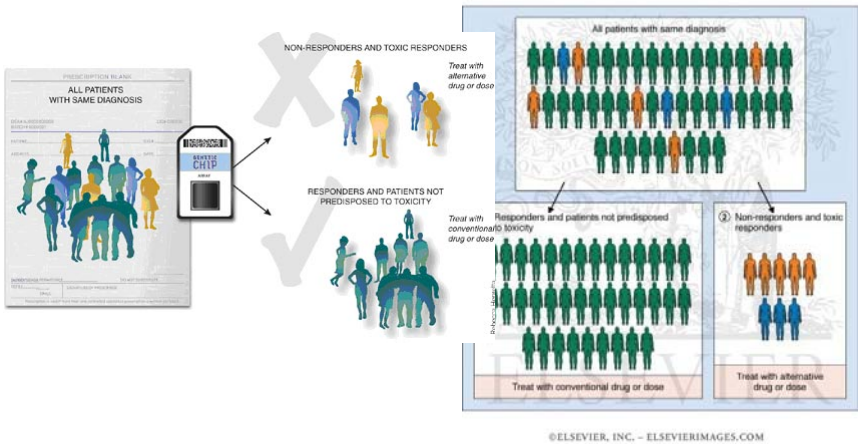
Risk of age-related macular degeneration (AMD) depends on variation in 3 genes



1% have > 50% risk of AMD most have risk close to average (Nat Genet 2006; 38:1055-9)

Personalized genomic medicine

The right treatment, for the right patient, at the right time

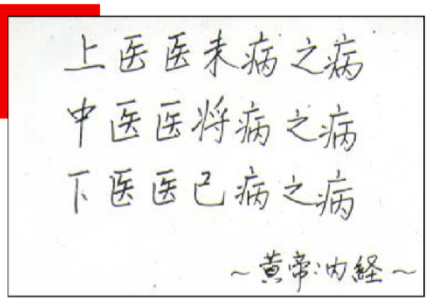


Personalized genomic medicine

Article in the „New York Times“ June 4, 2014: «In a First, Test of DNA Finds Root of Illness» - tells about a miraculous cure of a young boy due to the Next-Generation Sequencing (NGS), described in The New England Journal of Medicine (Wilson et al. 2014)

- Joshua Osborn, 14, laid in a coma at American Family Children's Hospital in Madison, Wis. For weeks his brain had been swelling with fluid, and a battery of tests had failed to reveal the cause.
- DNA-based test for diagnosing elusive pathogens
- DNA was isolated from different tissues, sequenced and compared with database within 48 hours
- Joshua's cerebrospinal fluid contained DNA from a potentially lethal type of bacteria called Leptospira
- Leptospira was readily treated with penicillin.

Preventive medicine

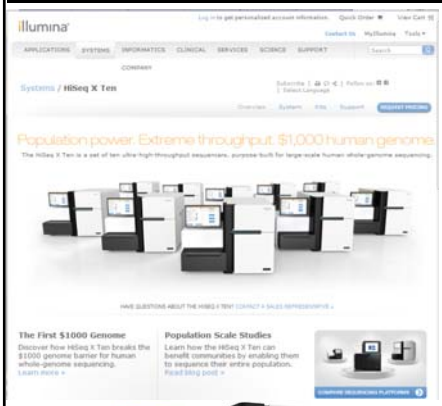


"Superior Doctors Prevent the Disease.
Mediocre Doctors Treat the Disease Before Evident.
Inferior Doctors Treat the Full Blown Disease."
-Huang Dee: Nai - Ching (2600 B.C. 1st Chinese Medical Text

Genomics allows to predict deseases, establish their relations with particular genes and genotypes, and therefore creates a foundation to prevent them

When thinking about diseases, I never think about how to cure them, but instead I think about how to prevent them.
-Louis Pasteur (1822-1895)

Preventive medicine



- Preventive medicine based on the whole genome sequencing is becoming a reality**!**
- Illumina presented a new and the most powerfull sequencer **HiSeq X** at the Plant and Animal Genome conference in San-Diego in January, 2014
- In his presentation, Illumina's chief executive Jay Flatley said the **HiSeq X** would be able to deliver **a human genome for just under $1,000**
- He said the world is "entering the **supersonic age of genomics**".
- **1.6-1.8 Tb for 3 days = >500 human genomes!**
- **Qatar's human genome project** (http://www.qatartodayonline.com/qatar-genome-launched-at-wish)

Paleogenomics and sequencing of ancient DNA

Scientists create complete genetic map of a Neanderthal from a TOE - and put it online for free

- Scientists from Germany's Max Planck Institute sequenced genome from toe bone found in southern Siberia
- New techniques allowed them to sequence every position in the genome 50 times over for greater accuracy
- They hope it will help answer questions about our own genetic history and how we're related to Neanderthals

By DAMIEN GAYLE

PUBLISHED: 14:52 GMT, 22 March 2013 | UPDATED: 17:28 GMT, 20 March 2013

Share Tweet Share 39 View comments

The first complete Neanderthal genome sequence has been completed and made available free-of-charge to researchers across the world.

Scientists from the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, have made the data available as a free download from their website.

The group will present a paper describing the genome later this year.

'But we make the genome sequence freely available now to allow other scientists to profit from it even before it is published' said Dr Svante Pääbo, who led the project.

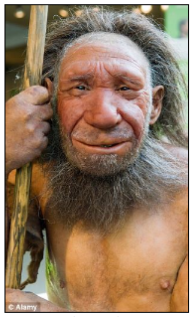
Dr Pääbo and his colleagues in 2010 presented the first draft of the Neanderthal genome from data collected from three bones found in a cave in Croatia.

They have now used a toe bone excavated in 2010 in Denisova Cave in southern Siberia to generate a high-quality genome from a single Neandertal individual.

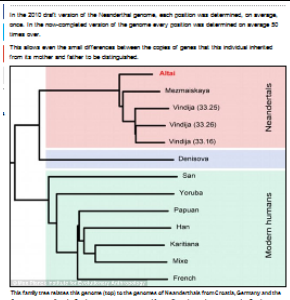
The Leipzig team used sensitive techniques developed there over the past two years to sequence every position in the genome about 50 times over, using DNA extracted from 0.038 grams of the bone.

The analysis of the genome together with partial genome sequences from other Neanderthals, and the genome from a small finger bone discovered in the same cave, shows that the individual is closely related to other Neanderthals in Europe and western Russia.

Remarkably, Neanderthals and their relatives, Denisovans, were both present in this unique cave in the Altai Mountains on the border between Russia, China, Mongolia and Kazakhstan.

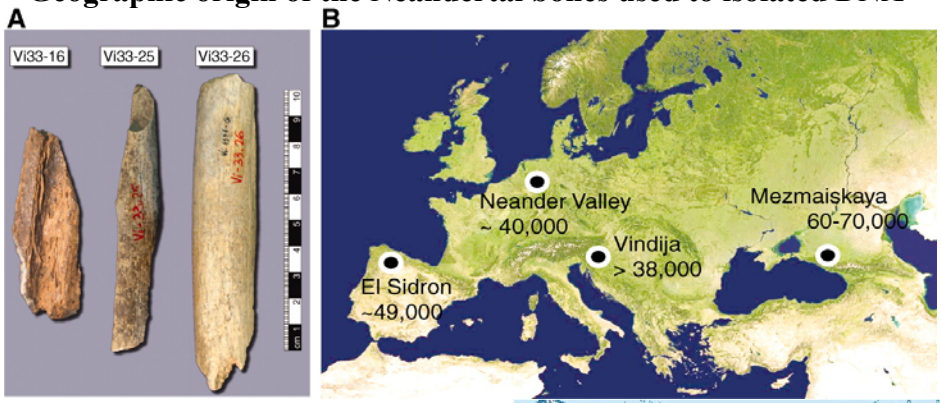


Sequenced: The first full Neanderthal genome has been sequenced and made available free-of-charge by the Max Planck Institute



Paleogenomics and sequencing of ancient DNA

Geographic origin of the Neandertal bones used to isolated DNA

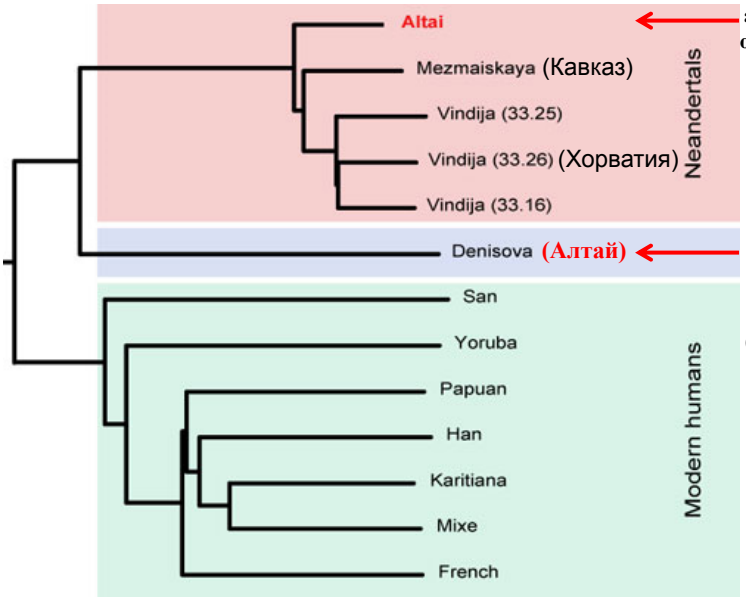


(A) The three bones from Vindija from which Neandertal DNA was sequenced.

(B) Map showing the four archaeological sites from which bones were used and their approximate dates (years B.P.)

Green et al. Science 2010; 328: 710-722

Paleogenomics and hominid paleophylogenomics

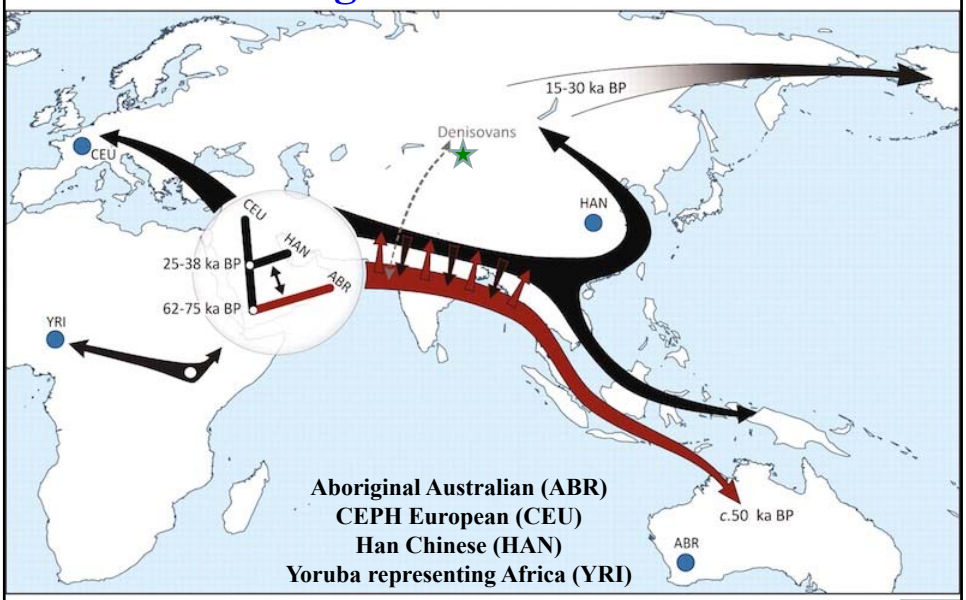


Neandertal genome assembled from DNA of a tooth found in the Denisova Cave

Genome assembled from DNA of a pedal phalanx found in the Denisova Cave in 2010 (Meyer et al. Pääbo 2012 Science 338(6104): 222-226)

http://www.eva.mpg.de/neandertal/index.html

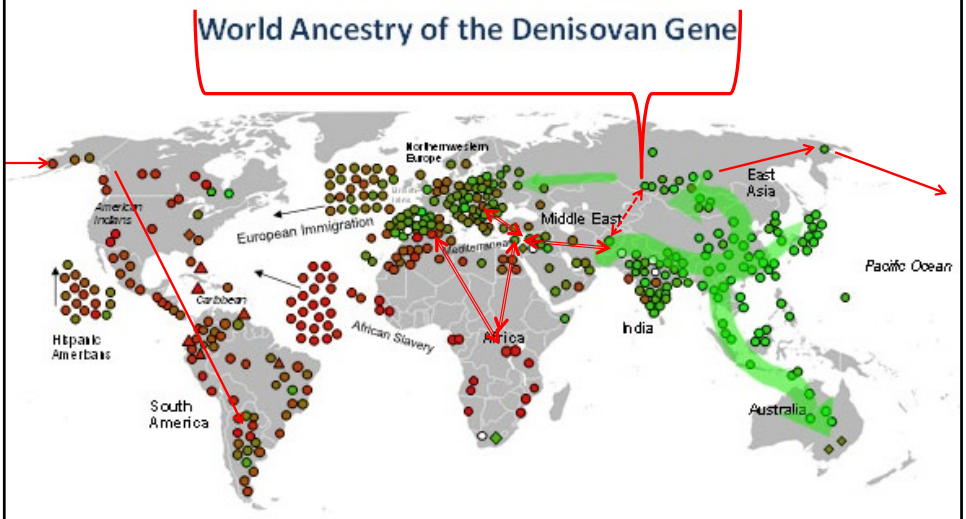
Historic migration of modern human



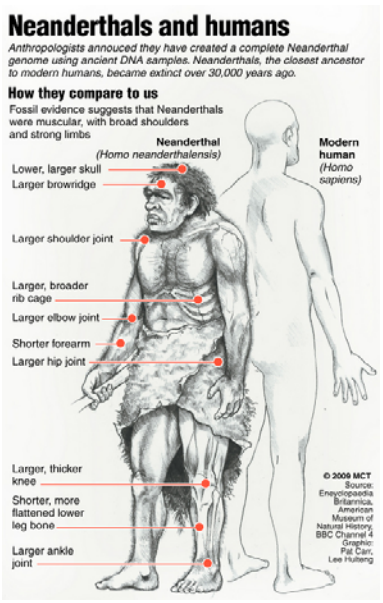
Aboriginal Australian (ABR)
CEPH European (CEU)
Han Chinese (HAN)
Yoruba representing Africa (YRI)

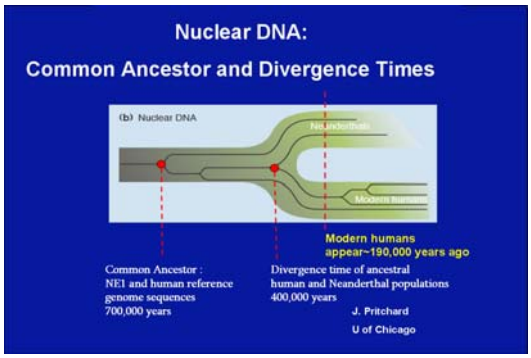
Morten Rasmussen et al. **An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia** *Science 2011: Vol. 334 no. 6052 pp. 94-98.*

Paleogenomics and hominid ancestry



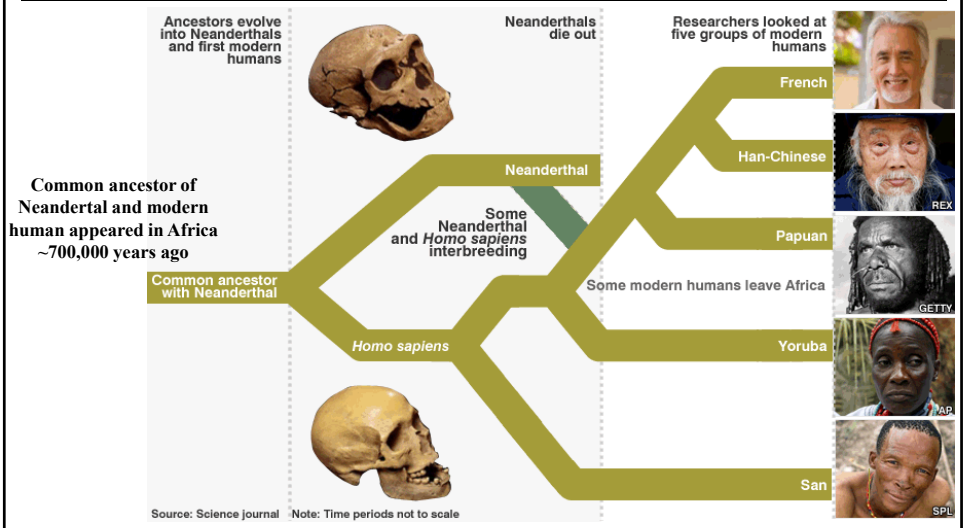
Paleogenomics and hominid paleophylogenomics





Analysis of genomic DNA from fossilized Neanderthal bones indicated that *Homo sapiens* and *Homo neanderthalensis* last shared a common ancestor approximately 700,000 years ago. The two hominids split into separate species approximately 400,000 years ago, with no evidence of any significant crossbreeding between the two after that time.

Paleogenomics and hominid paleophylogenomics



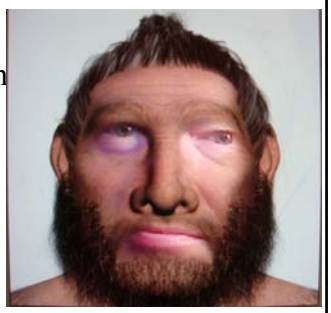
Split between Neandertal and modern human occured ~400,000 years ago

Dispersal of modern human from Africa to Eurasia began ~40,000-70,000 years ago

Paleogenomics and sequencing of ancient DNA

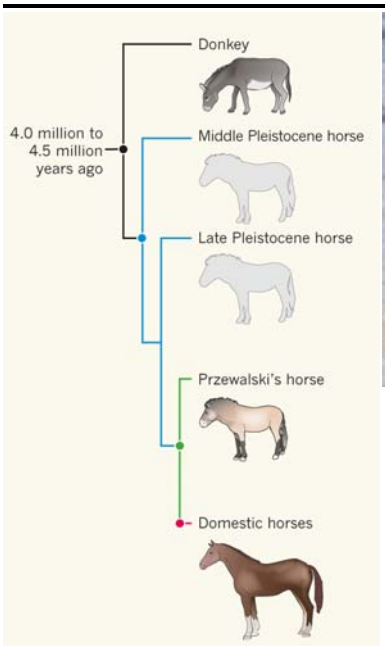
Special challenges:

- Ancient DNA is degraded by nucleases
- The majority of DNA in samples derives from unrelated organisms such as bacteria that invaded after death
- The majority of DNA in samples is contaminated by human DNA
- Determination of authenticity requires special controls, and analysis of multiple independent extracts



Green, R. E. *et al.* A draft sequence of the Neandertal genome. Science 328, 710–722 (2010)

Paleogenomics and paleophylogenomics



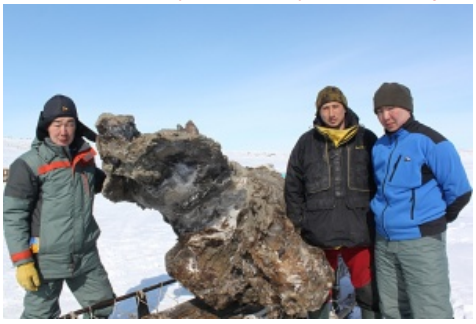
- The most ancient, 700,000 year old DNA was isolated from the remnants of the ancient horse found in the permafrost in Canada and was used to assemble a whole genome.
- Phylogenomic analysis demonstrated that the common ancestor of domestic horses, zebras and their relatives lived ~4 mln years ago (Orlando *et al*. Nature 2013: http://dx.doi.org/10.1038/nature12323).

Paleogenomics and sequencing of ancient DNA

Genome of wool mammoth (***Mammuthus primigenius***)was partially sequenced in 2008 using hairs of two females found in permafrost in Siberia and dated as ~20,000 and 60,000 year old (Miller et al. 2008 Nature 456: 387-390).

The best preserved wool mammoth was found in **2013 in Maly Lyakhovsky Island in the far north of Siberia**

Scientists from the Siberian Northeastern Federal University in Yakutsk and the Siberian Federal University in Krasnoyarsk have a joint project for the whole genome sequencing



Paleogenomics and sequencing of ancient DNA

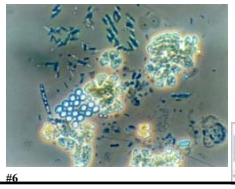
Dissection and sampling of the Maly Lyakhovsky mammoth by the scientists from the Siberian Northeastern Federal University and the Siberian Federal University in 2014 . The genome sequencing will be done at the Genome Research and Education Center of the Siberian Federal University (http://genome.sfu-kras.ru/en)



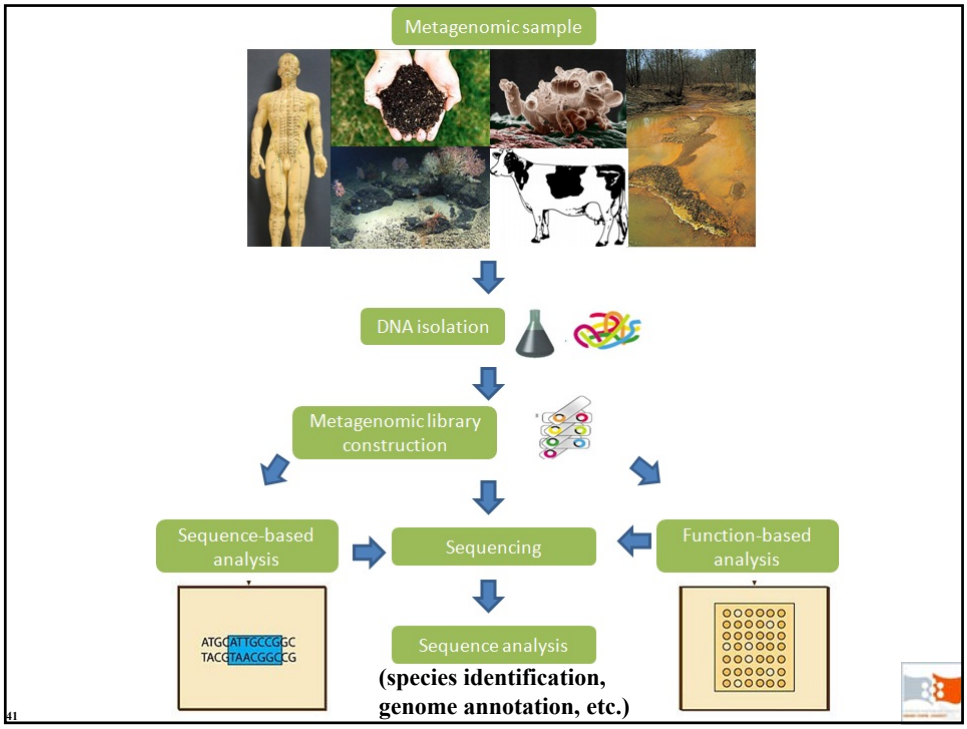


Metagenomics and sequencing of complex communities

Metagenomics (also **Environmental Genomics**, **Ecogenomics** or **Community Genomics**) is the study of genetic material recovered directly from environmental samples:



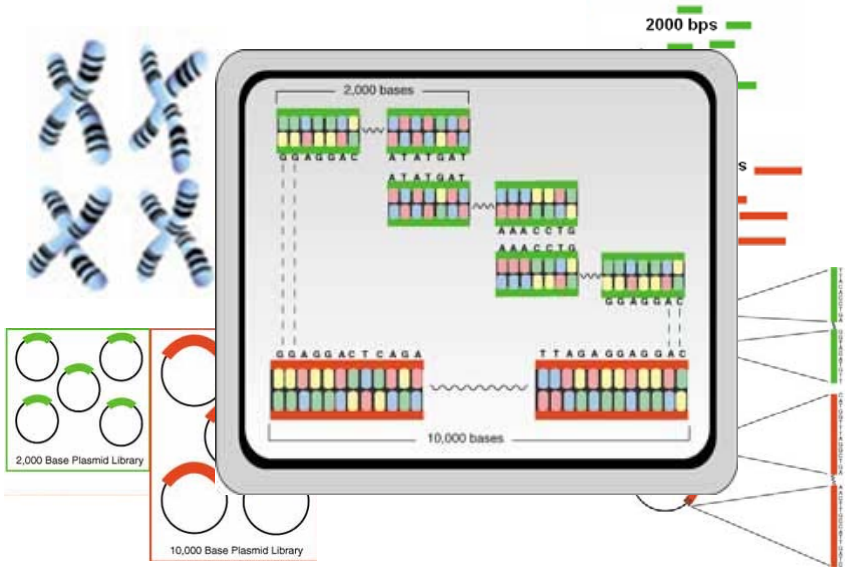
- external environments (ecological)
  hot spring, ocean, sludge, soil, etc.
- internal environments (organismal)
  guts, saliva, feces, lung, etc.



Sampling in Metagenomics

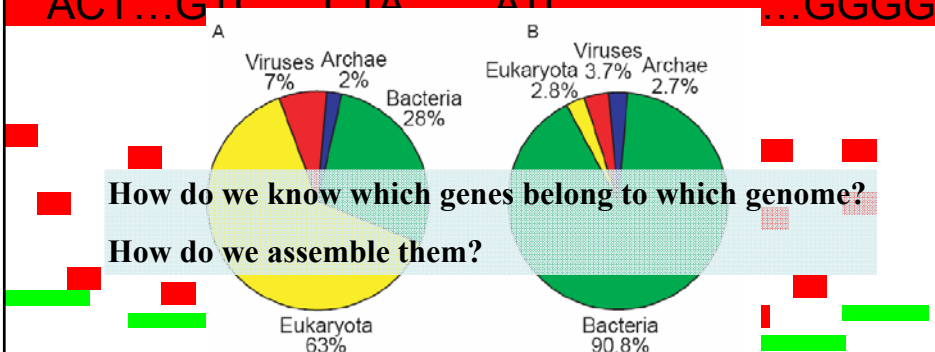
- Take a sample off of the environment
- Isolate and amplify DNA/mRNA
- Sequence it

Shotgun Sequencing



Computer assembly

ACT…GTC CTA ATC …GGGG



How do we know which genes belong to which genome?

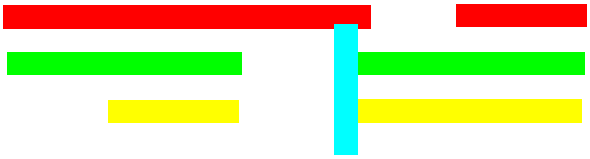
How do we assemble them?

Global Ocean Sampling (GOS) metagenomic data

The Best Case Scenario

Coverage is enough to assemble independent genomes

What normally happens

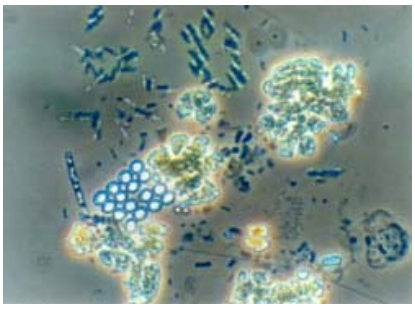


Coverage is not enough and assembly is fragmentary

Worst Case Scenario: Some fragments can not be assigned

Down Side of Metagenomics

- often fragmentary
- often highly divergent
- rarely any known activity
- no chromosomal placement
- no organism of origin
- *ab initio* ORF predictions
- huge data

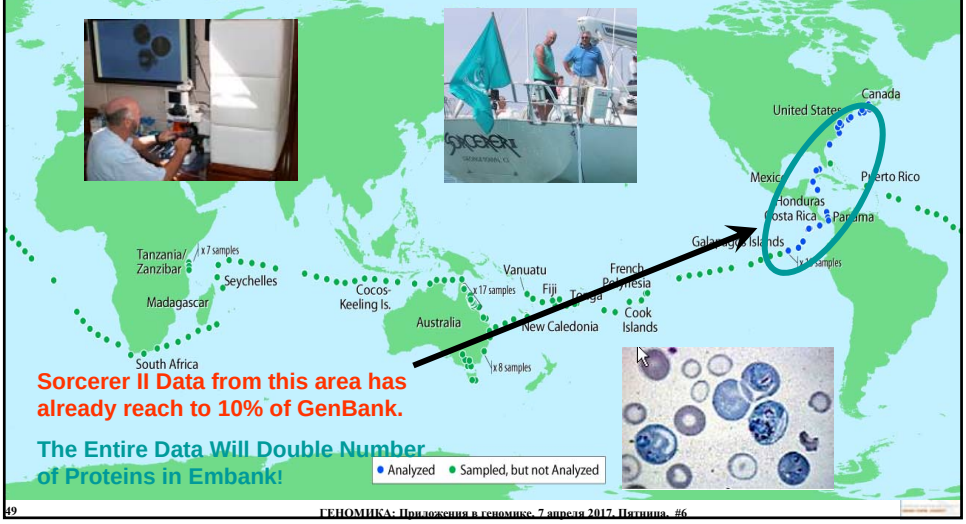


Marine Metagenomics

- Microbes account for more than **90% of ocean biomass**, mediate all biochemical cycles in the oceans and are responsible for **98% of primary production** in the sea.

- Metagenomics is a breakthrough sequencing approach to examine the open-space microbial species **without the need for isolation and lab cultivation of individual species**.

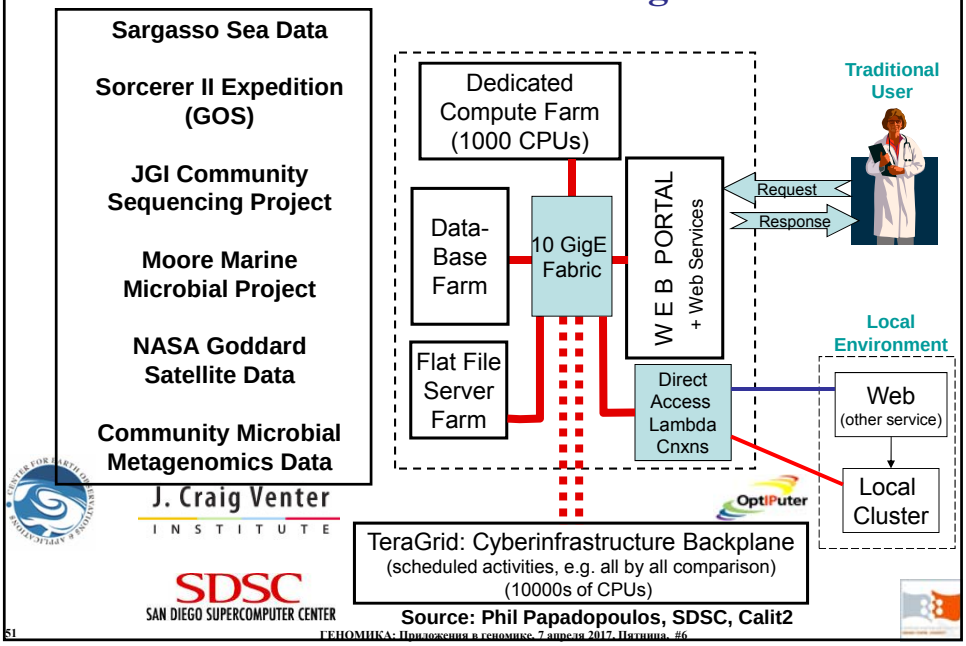
Marine Genome Sequencing Project Measuring the Genetic Diversity of Ocean Microbes



Sample Metadata from GOS (Global Ocean Sampling)

- **Site Metadata**
  - Location (lat/long, water depth)
  - Site characterization (finite list of types plus "other")
  - Site description (free text)
  - Country
- **Sampling Metadata**
  - Sample collection date/time
  - Sampling depth
  - Conditions at time of sampling (e.g., stormy, surface temperature)
  - Sample physical/chemical measurements
  - "author"
- **Experimental Parameters**
  - Filter size
  - Insert size

Calit2's Direct Access Core Architecture Will Create Next Generation Metagenomics Server



Source: Phil Papadopoulos, SDSC, Calit2

Marine Metagenomics

Drug discovery

Metabolic pathway discovery

Microbial genetic survey

Environmental survey

Symbiosis

Who is there?

Evolution study

Endosymbiosis

Organism discovery

Microbial genomic survey

Bioenergy discovery

Marine conservation

Biogeochemistry mapping

Ecological restoration

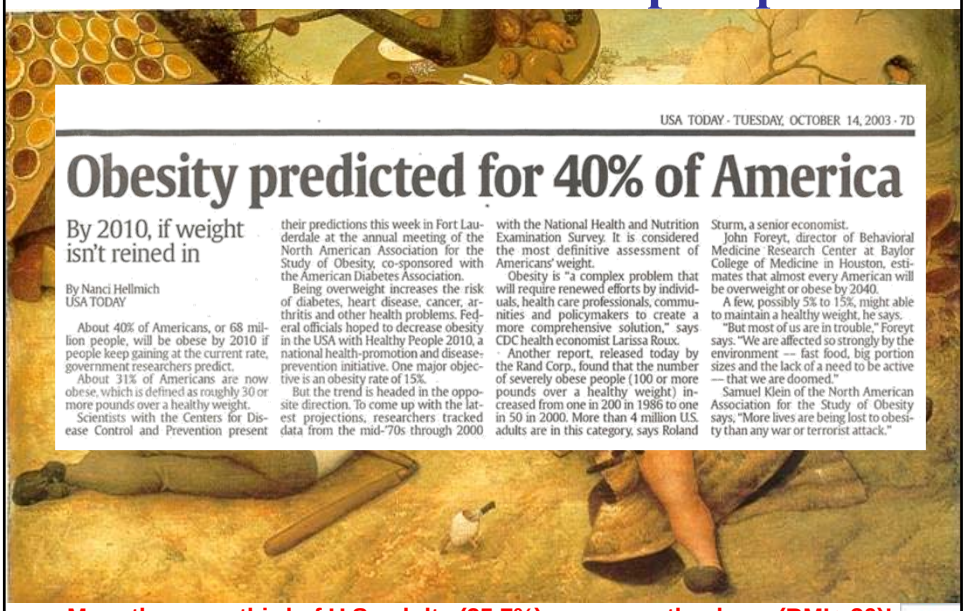
What is Nutrigenomics?

- Nutrigenomics is the science that examines the response of individuals to food compounds using post-genomic and related technologies.
- The long-term aim of nutrigenomics is to understand how the whole body responds to real foods using an **integrated system biology approach**.
- Studies using this approach can examine people (i.e. populations, subpopulations - based on genes or disease - and individuals), food, life-stage and life-style without preconceived ideas.

Why is Nutrigenomics important?

- Most non-genetic diseases are **nutrition** related.
- **Diabetes, obesity and other nutrition related diseases are growing!!!** Of course genes are a factor.
- **Finding the right combination of nutrients for each genotype** can help in changing behavior and preventing many of these diseases.
- This combination may change with age, sex!

Problem 1: Nutrition – complex problem



USA TODAY · TUESDAY, OCTOBER 14, 2003 · 7D

Obesity predicted for 40% of America

By 2010, if weight isn't reined in

By Nanci Hellmich
USA TODAY

About 40% of Americans, or 68 million people, will be obese by 2010 if people keep gaining at the current rate, government researchers predict.

About 31% of Americans are now obese, which is defined as roughly 30 or more pounds over a healthy weight.

Scientists with the Centers for Disease Control and Prevention present their predictions this week in Fort Lauderdale at the annual meeting of the North American Association for the Study of Obesity, co-sponsored with the American Diabetes Association.

Being overweight increases the risk of diabetes, heart disease, cancer, arthritis and other health problems. Federal officials hoped to decrease obesity in the USA with Healthy People 2010, a national health-promotion and disease-prevention initiative. One major objective is an obesity rate of 15%.

But the trend is headed in the opposite direction. To come up with the latest projections, researchers tracked data from the mid-'70s through 2000 with the National Health and Nutrition Examination Survey. It is considered the most definitive assessment of Americans' weight.

Obesity is "a complex problem that will require renewed efforts by individuals, health care professionals, communities and policymakers to create a more comprehensive solution," says CDC health economist Larissa Roux.

Another report, released today by the Rand Corp., found that the number of severely obese people (100 or more pounds over a healthy weight) increased from one in 200 in 1986 to one in 50 in 2000. More than 4 million U.S. adults are in this category, says Roland Sturm, a senior economist.

John Foreyt, director of Behavioral Medicine Research Center at Baylor College of Medicine in Houston, estimates that almost every American will be overweight or obese by 2040.

A few, possibly 5% to 15%, might able to maintain a healthy weight, he says.

"But most of us are in trouble," Foreyt says. "We are affected so strongly by the environment — fast food, big portion sizes and the lack of a need to be active — that we are doomed."

Samuel Klein of the North American Association for the Study of Obesity says, "More lives are being lost to obesity than any war or terrorist attack."

More than one-third of U.S. adults (35.7%) are currently obese (BMI >30)!
http://www.cdc.gov/obesity/data/adult.html

Genes – Lifestyle – Calories

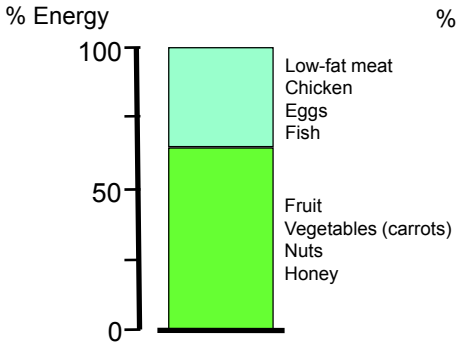
The same genes – The changed diet

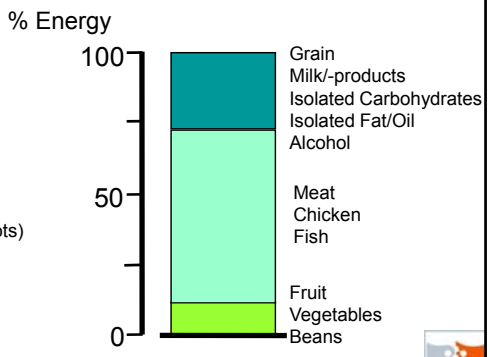
Paleolithic era

1.200.000 Generations between feast and famine

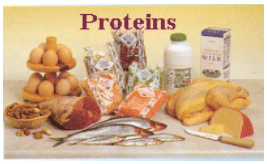
Modern Times

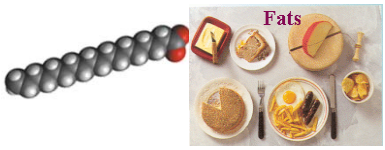
2-3 Generations in energy abundance

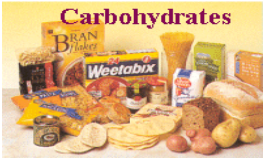


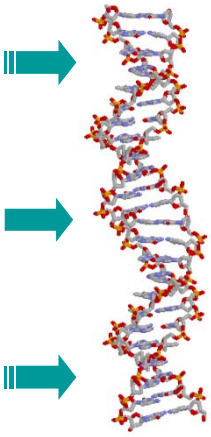


Molecular nutrition





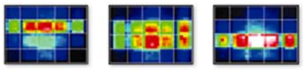




Problem 2:
Our “gene passports” and nutrition

Individual genotype
Functional phenotype

AA AB BB



Optimal Nutrition

Lifestyle

Improvement
Maintenance
of Health

“Eat right for your genotype??”

Personalized diets?

genelex

Nutritional Genetic Profile
Request Form

Client Information

To order testing, either contact Genelex directly or complete this form and return it either by fax at (425) 825-1870 or mail to Genelex Corporation, 12277 134th Ct NE, Ste. 130 Redmond, WA 98052.

Name: _____ Phone: _____ E-mail: _____

Address: _____

City: _____ State: _____ Zip: _____

Nutritional Genetic Profile Requested

Item	Number ordered	Cost (per item)	Total
Nutritional Genetic Panel		$445.00	
Nutritional Genetic Collection Kit (Additional $410 due with samples)		$35.00	
International Shipping		$50.00	
Amount Due			

Payment: Prepayment is required. Send Cash, Check, or Money Order to the address shown above.

Cash ☐ Check or Money Order ☐ Credit Card (all major cards) ☐

Type of credit card: _____

Print cardholder's name: _____

Card number: _____ Expiration date: _____

For immediate consultation Call 800-TEST-DNA (800-837-8362)
Hours 7:00 AM to 6:00 PM PST, 10:00 AM to 9:00 PM EST, fax 425-825-1870,
e-mail: info@genelex.com
www.genelex.com
©2002 Genelex Corporation

Consumers warned that time is not yet ripe for nutrition profiling

Erika Check

Looks good, tastes good, and one day individuals may know exactly how much good it does them.

One day, information about your genome may well help you decide what breakfast cereal to eat. But that day's a long way off, the second International Nutrigenomics Conference in Amsterdam was told last week. In the meantime, researchers at the meeting heard, the emerging field badly needs a regulatory framework that will stop its first customers from being scared off.

Nutrigenomics researchers aim to learn how nutrients interact with genes to lead to health or disease. But people eat wildly different levels of nutrients over their lifetimes, and teasing apart the precise interactions is notoriously difficult.

The researchers who gathered in Amsterdam on 6–7 November were in optimistic mood, however. Their science is progressing quickly, and food industry executives have expressed interest in the idea of using genetic information to customize their products.

In January, the US National Institutes of Health used a 5-year, $6.5-million grant to create a National Center of Excellence for Nutritional Genomics at the University of California, Davis, and the Children's Hospital Oakland Research Institute (CHORI) in Oakland. In July, the European Commission set up the European Nutrigenomics Organisation to coordinate work. Now the Netherlands looks set to embark on a $20-million nutrigenomics project, jointly funded by the government and the food industry.

But some researchers warn that the field is in danger of developing too quickly. They want experts to back off from the sometimes-extravagant claims for the field's potential, and instead to sit down and patiently work out a scientific vision and ethical framework for the discipline.

"Our aim is to bring the field a little bit back down to Earth, because people tend to start with a lot of science fiction," says Michael Müller, a gerontologist at Wageningen University in the Netherlands who helped to organize the meeting.

The main fruits of this field are still years away, researchers say. So far, most of the studies on profiling gene expression — measuring genome-wide responses to nutrients — have been done in mice. And much more work is needed on the basic mechanisms by which nutrients turn genes on or off. But that hasn't stopped a handful of companies from selling nutritional profiles directly to consumers over the Internet.

The companies test a tissue sample — such as a cheek swab — from a "patient". The patient can choose which genetic profile he or she wants to learn about, for example skin ageing or susceptibility to osteoporosis. The company then gives the patient a "personalized profile" based on its tests for single nucleotide polymorphisms (SNPs): genetic variants that have been linked to disease. For instance, one company, GeneLink of Margate, New Jersey, tells people what vitamins they should take, based on SNPs involved in cellular responses to certain toxins. GeneLink declined to comment on its products.

But many scientists argue that it's far too early for most of these tests to be useful. "The idea of marketing any individual genetic test at this point assumes there is information to justify the use of that test, and we really don't have evidence that any single genetic marker carries enough information to guide dietary treatments," says Ronald Krauss, director of atherosclerosis research at CHORI.

The direct-to-consumer tests also raise ethical issues that affect the whole field. For instance, some companies sell the results of their genetic profiles to other firms, which use the information for research on genes and disease. Although consumers must give their consent, they may not necessarily understand what they're agreeing to, says ethicist David Castle of the University of Guelph. Castle is collaborating with the University of Toronto Joint Center for Bioethics in soliciting comments on a joint working paper on ethics and nutrigenomics.

At the nutrigenomics meeting, Castle argued that even though the field is very young, scientists must begin talking to the public about such issues.

"This technology could end up affecting something that every person does every day, which is eat," Castle says. "It's not a situation where you want to roll out the science and the products and then go back and ask people how they feel about it."

NATURE | VOL 426 | 13 NOVEMBER 2003 |

Nutrients acts as dietary signals

Nutritional factors

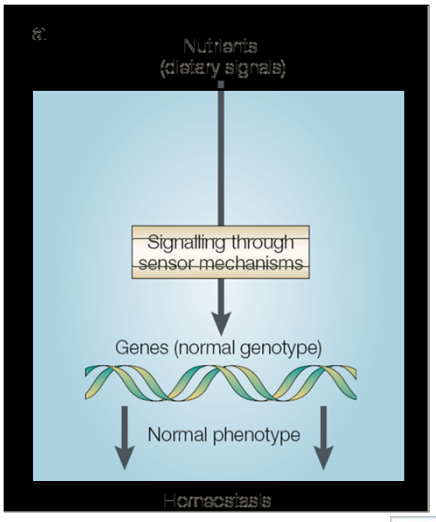
↓

Transcription factors

↓

Gene transcription

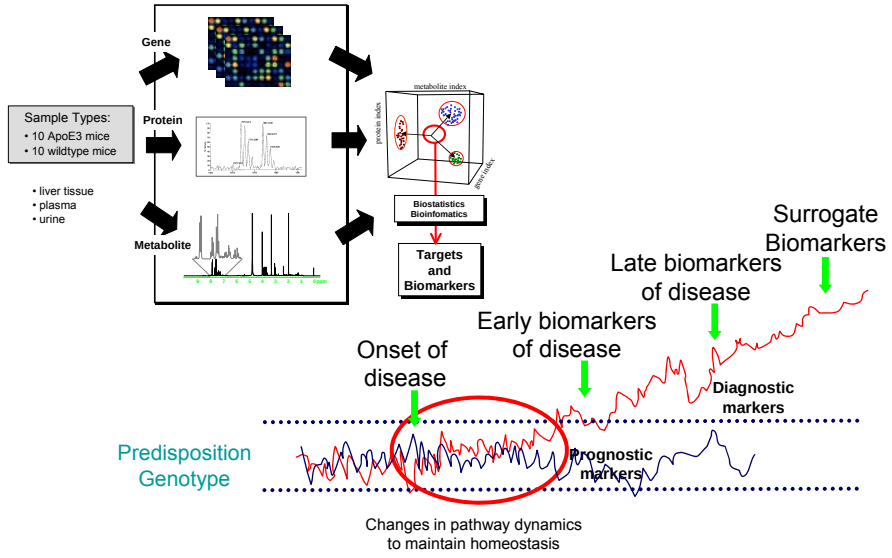
Energy homeostasis | Cell proliferation | Nutrient absorption



Transcription-factor pathways mediating nutrient-gene interaction

Nutrient	Compound	Transcription factor
Macronutrients		
Fats	Fatty acids Cholesterol	PPARs, SREBPs, LXR, HNF4, ChREBP SREBPs, LXRs, FXR
Carbohydrates	Glucose	USFs, SREBPs, ChREBP
Proteins	Amino acids	C/EBPs
Micronutrients		
Vitamins	Vitamin A Vitamin D Vitamin E	RAR, RXR VDR PXR
Minerals	Calcium Iron Zinc	Calcineurin/NF-ATs IRP1, IRP2 MTF1
Other food components		
	Flavonoids Xenobiotics	ER, NFκB, AP1 CAR, PXR

Nutritional Systems Biology



Nutrigenomics

Foods
Nutrition

Target Genes
Mechanisms
Pathways

Signatures
Profiles
Biomarkers

Molecular Nutrition
& Genomics

Nutritional
Systems Biology

•Identification of dietary signals
•Identification of dietary sensors
•Identification of target genes
•Reconstruction of signaling pathways

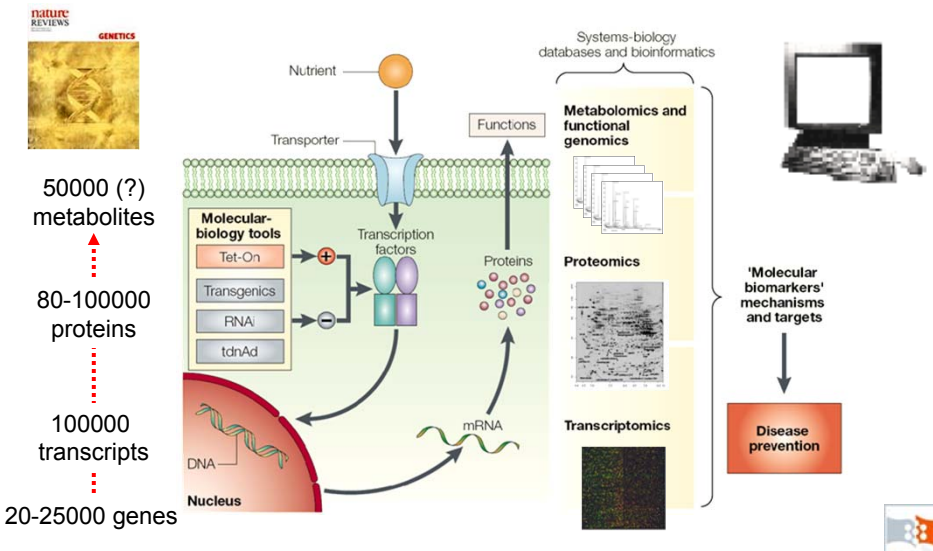
•Measurement of stress signatures
•Identification of early biomarkers

Small research groups
Small budgets

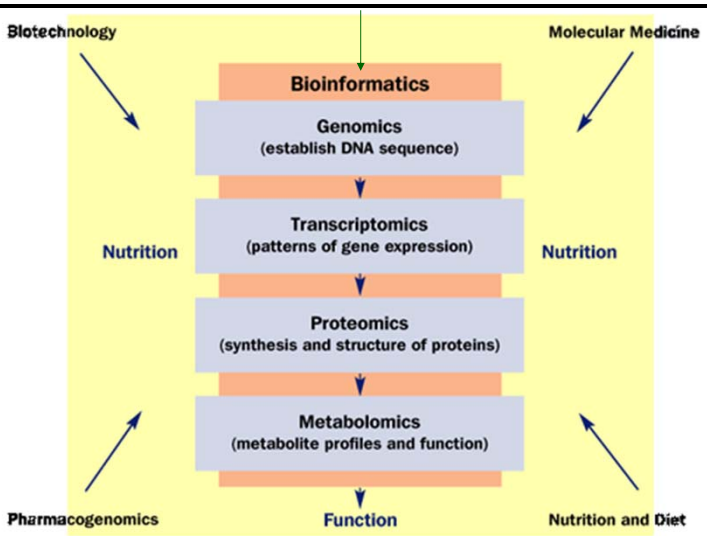
Large research consortia
Big money

Complexity

“Molecular Nutrition & Genomics”
The strategy of Nutrigenomics



Integration of enabling technologies in nutrigenomics



Two Strategies

(1) The traditional hypothesis-driven approach: specific genes and proteins, the expression of which is influenced by nutrients, are identified using genomics tools — such as transcriptomics, proteomics and metabolomics — which subsequently allows the regulatory pathways through which diet influences homeostasis to be identified. Transgenic mouse models and cellular models are essential tools.

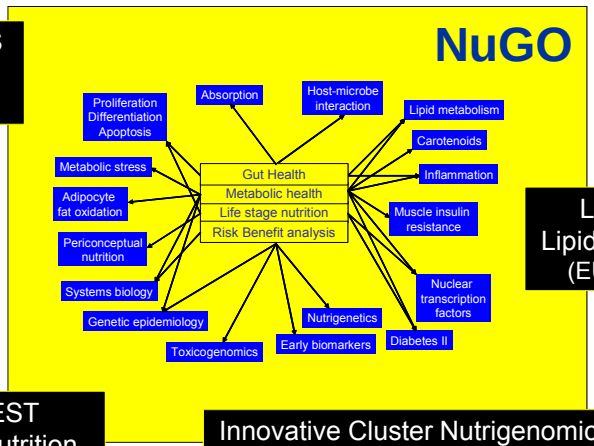
provide us with detailed molecular data on the interaction between nutrition and the genome.

(2) The SYSTEMS BIOLOGY approach: gene, protein and metabolite signatures that are associated with specific nutrients, or nutritional regimes, are catalogued, and might provide 'early warning'molecular biomarkers for nutrient-induced changes to homeostasis.

Be more important for human nutrition, given the difficulty of collecting tissue samples from 'healthy' individuals.

EU programs

DIOGENES
obesity
(EU, 12M€)



LIPGEN
Lipids & genes
(EU, 14M€)

EARNEST
early life nutrition
(EU, 14M€)

Innovative Cluster Nutrigenomics
Chronic metabolic stress
(Dutch, 21M€)

Conclusion and future perspective

(1) Nutrigenomics researchers must know the challenge of understanding polygenic diet related diseases.

(2) Short-term goals:

1. to identify the dietary signals.

2. to elucidate the dietary sensor mechanisms.

3. to characterize the target genes of these sensors.

4. to understand the interaction between these signalling pathways and pro-inflammatory signalling to search for sensitizing genotypes.

5. to find '**signatures**' (gene/protein expression and metabolite profiles).

(3) Long-term goals:

Nutrigenomics is to help to understand how we can use nutrition to prevent many of the same diseases for which pharmacogenomics is attempting to identify cures.

Future → **personalized diets**

Gerontogenomics

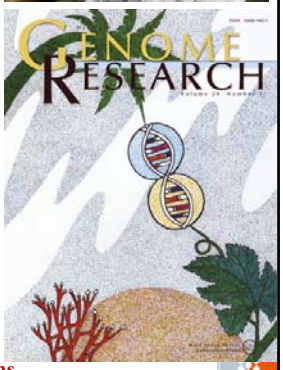
GerontoGenomics is the genomics of aging and senescence

Research

Somatic mutations found in the healthy blood compartment of a 115-yr-old woman demonstrate oligoclonal hematopoiesis

Henne Holstege,[1,10] Wayne Pfeiffer,[2] Daoud Sie,[3] Marc Hulsman,[4] Thomas J. Nicholas,[5] Clarence C. Lee,[6] Tristen Ross,[6] Jue Lin,[7] Mark A. Miller,[2] Bauke Ylstra,[3] Hanne Meijers-Heijboer,[1] Martijn H. Brugman,[8] Frank J.T. Staal,[8] Gert Holstege,[9] Marcel J.T. Reinders,[4] Timothy T. Harkins,[6] Samuel Levy,[5] and Erik A. Sistermans[1]

- Individual genome in the multiple blood cells of **Hendrikje van Andel-Schipper (1890-2005)**, at one point the oldest woman in the world, were sequenced and compared (Holstege et al. 2014 *Genome Res*. 24(5): 733-742)
- She was remarkably healthy until her death
- **450 mutations were found in her cells, but none of them was detrimental**



- **genomes of 17 of the world's oldest living people (110-116 year old) have been sequenced and published recently** (Gierman et al. 2014 PLoS ONE 9(11): e112430 http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0112430)
- **Japanese project to sequence genome and metagenome of all centenarians**