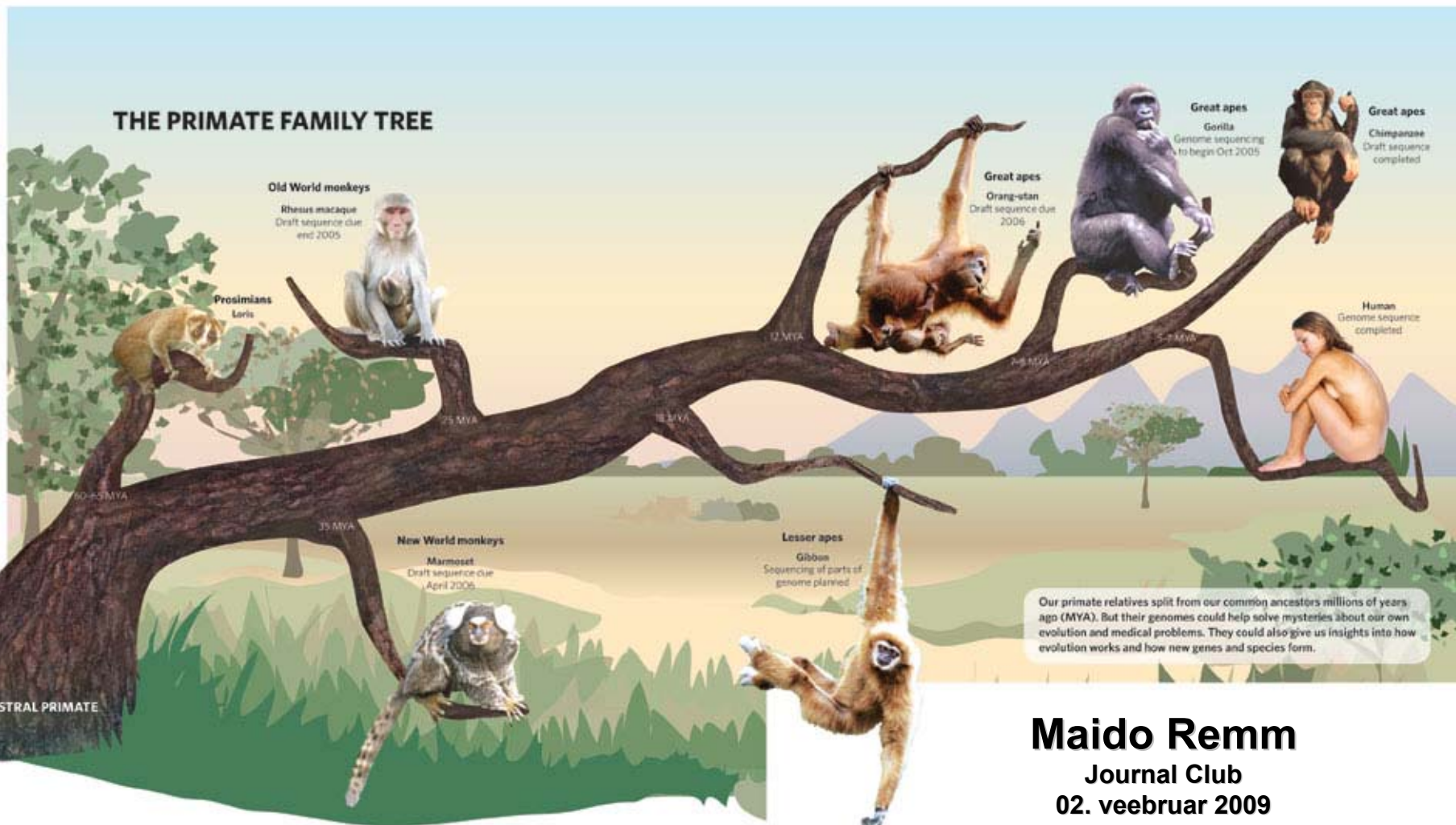


Evolutionary and Biomedical Insights from the Rhesus Macaque Genome

Rhesus Macaque Genome Sequencing and Analysis Consortium

SCIENCE VOL 316 222-234



Maido Remm
Journal Club
02. veebruar 2009

Mis teeb inimese eriliseks? Mis on genoomis “inimlikku”?

Mis annab inimesele oskuse lahendada Rubiku kuubiku sarnaseid abstraktseid ülesandeid, samas kui kaheksajalgadele (ja šimpansidele) on see liialt raske?

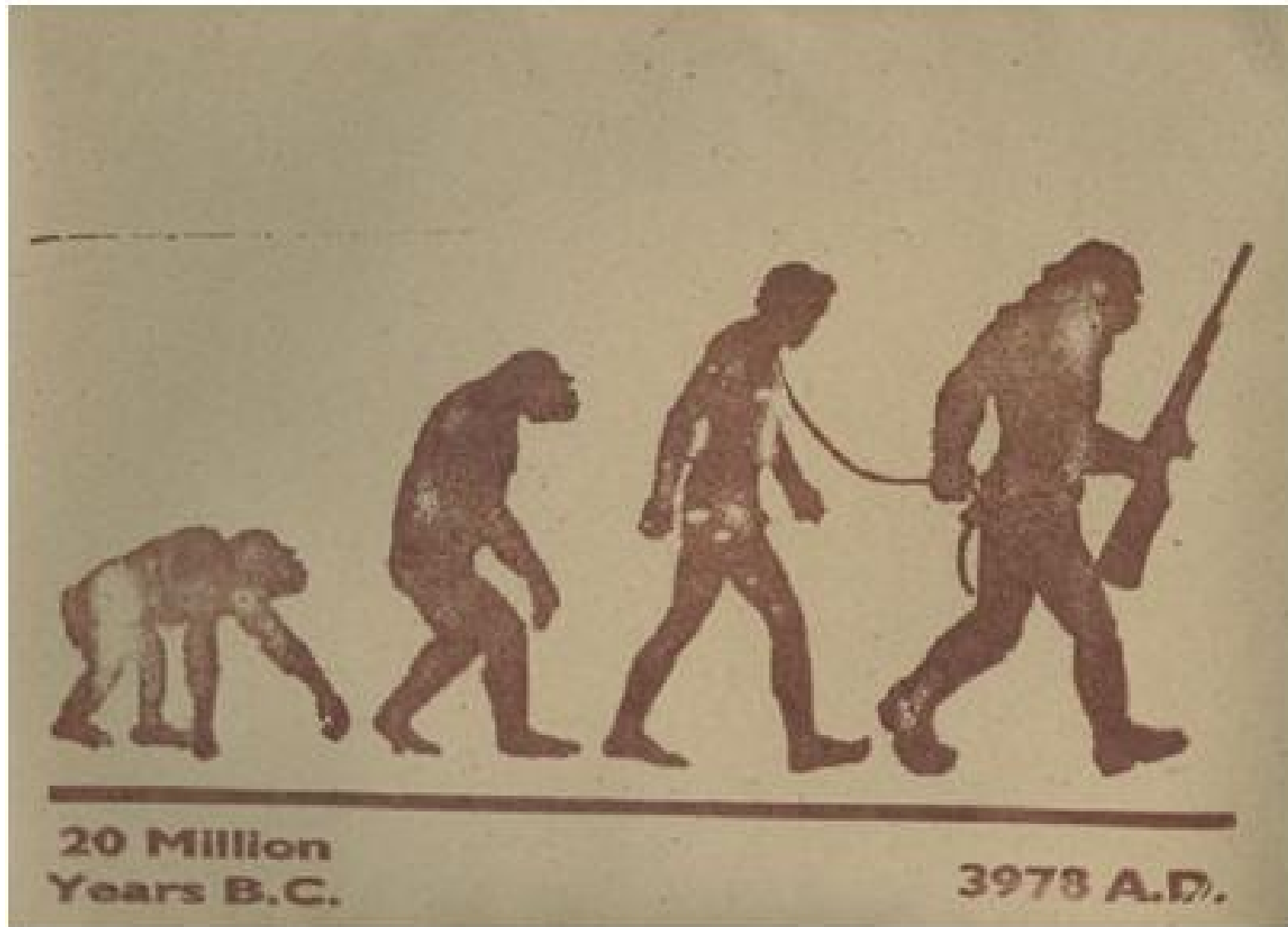


Mis teeb inimese eriliseks? Mis on genoomis “inimlikku”?

Miks annab inimestele võime, oskuse ja tahtmise vangistada teisi loomi samas kui loomad teise liigi esindajaid vangistuses ei hoia?



Kas evolutsioonis on võimalikud ka teistsugused stsenaariumid?



Kas EBK direktor on tulevikus selline?



... ja TÜMRI nõukogu selline?



Inimese ja šimpansi genoomi võrdlemine

Üht-teist saab inimese genoomi unikaalsuse kohta teada juba inimese genoomi võrdlemisel šimpansi genoomiga (Nature, Sep 2005, 437:69).

Samas on erinevusi liiga palju et ilma täiendava analüüsita saaks midagi järeldada.

Kokku on kahe genoomi vahel ca 5% nukleotiidide tasandil erinevusi

- 35 miljonit ühenukleotiidset asendust (kokku ca 35 Mb)
- 5 miljonit insertiooni/deletsiooni (kokku ca 90 Mb)

Nende hulgast oluliste leidmine ei ole lihtne.

Inimese ja šimpansi genoomi võrdlemine

Varki and Altheide “Comparing the human and chimpanzee genomes: Searching for needles in a haystack.”

Genome Research 156:1746-1758 (2005)

What makes humans different from their closest evolutionary relatives and how, why and when did these changes occur?

Ülevaade võimalikest strateegiatest “nõela” otsimisel “heinakuhjast”.

Inimese ja šimpansi genoomi võrdlemine

Table 3. Some candidate genes and gene families that may contribute to phenotypic differences between humans and apes^a

Gene(s)	Gene product(s)	Unusual hominid or human-specific features	Potential relevance to the human condition
Individual genes			
<i>FOXP2</i>	Putative transcription factor with polyglutamine tract and forkhead DNA binding domain	Two human-specific amino acid changes	Mutant humans have motoric speech disorder (developmental verbal dyspraxia). Region positively selected and fixed in humans <200,000 years ago (Enard et al. 2002b; Zhang et al. 2002)
<i>MYH16</i>	Myosin heavy chain 16	Human-specific 2-bp deletion causing frameshift—predicted 76-kD unstable head domain	Claimed to be cause of reduction in the type II fibres of human jaw muscle. (Stedman et al. 2004; Perry et al. 2005)
<i>CMAH</i>	CMP-Neu5Ac hydroxylase	92-bp deletion of exon 6 causing frameshift and inactive enzyme. Fixed in modern humans	Absence of sialic acid Neu5Gc. Change in resistance or susceptibility to pathogens. Loss of ligand for some Siglecs. Dated -2.5–3 Mya. Dietary Neu5Gc in meat became foreign antigen (Chou et al. 1998, 2002; Irie et al. 1998; Hayakawa et al. 2001)
<i>MAOA</i>	Monoamine oxidase A	Human-specific nonconservative change Glu151Lys in active site	Substitution affects protein dimerization according to a 3D structural model and predicts functional change (Andres et al. 2004)
<i>ASPM</i>	Modulator of mitotic spindle in neural progenitors?	Accelerated evolution in ape and human lineages	Deletions in <i>ASPM</i> lead to microcephaly. Presumed to be related to increased brain size and/or other features of human brain (Zhang 2003; Dorus et al. 2004; Evans et al. 2004; Kouprina et al. 2004b; Mekel-Bobrov et al. 2005)
<i>MCPH1</i>	Microcephalin	As above	As above (Dorus et al. 2004; Evans et al. 2004, 2005)
<i>TTR</i>	Transthyretin	Decreased expression in humans in blood and brain	May be related to altered thyroid hormone metabolism in humans versus chimpanzees (Gagneux et al. 2001)
<i>ST6GAL1</i>	Alpha 2–6 sialyltransferase	Apparent human-specific up-regulation on epithelia	Can explain relative resistance of chimpanzees to human influenza A virus. Other consequences unknown. (Gagneux et al. 2003)
<i>EMR4</i>	EGF-TM7 receptor family	Human-specific deletion in exon 8. Frameshift	Predominantly expressed by immune system cells. Functional significance unknown (Hamann et al. 2003)
<i>PCDH11Y</i>	Protocadherin XY	Duplicated onto Y in Yp11.2/Xq21.3 pseudoautosomal region only in humans	Expressed from Y and escapes X-inactivation? Y copy has undergone structural changes. Selectively expressed in brain. Probable adhesion molecule. Significance unknown, hypothesized to be involved in brain development, lateralization and schizophrenia risk (Ross et al. 2003; Blanco-Arias et al. 2004)
<i>IL9R (Y)</i>	Interleukin-9 receptor	As above	Expressed from Y and escapes X-inactivation? Growth factor for T cells, mast cells, and macrophages. Significance unknown. Related to asthma? (Vermeesch et al. 1997)
<i>SPRY3</i>	Sprouty 3	As above	Expressed from Y and escapes X-inactivation? Cysteine-rich protein—Homolog of <i>Drosophila</i> antagonist of FGF signaling that patterns apical airways branching. Significance unknown (Vermeesch et al. 1997)

Täiendavaid vastuseid aitavad leida reesusmakaagid (*Macaca mulatta*)

Kuuluvad gruppi, mida nimetatakse “Old World Monkeys”

Kingdom: [Animalia](#)

Phylum: [Chordata](#)

Class: [Mammalia](#)

Order: [Primates](#)

Family: [Cercopithecidae](#)

Genus: [Macaca](#)

Species: ***M. mulatta***

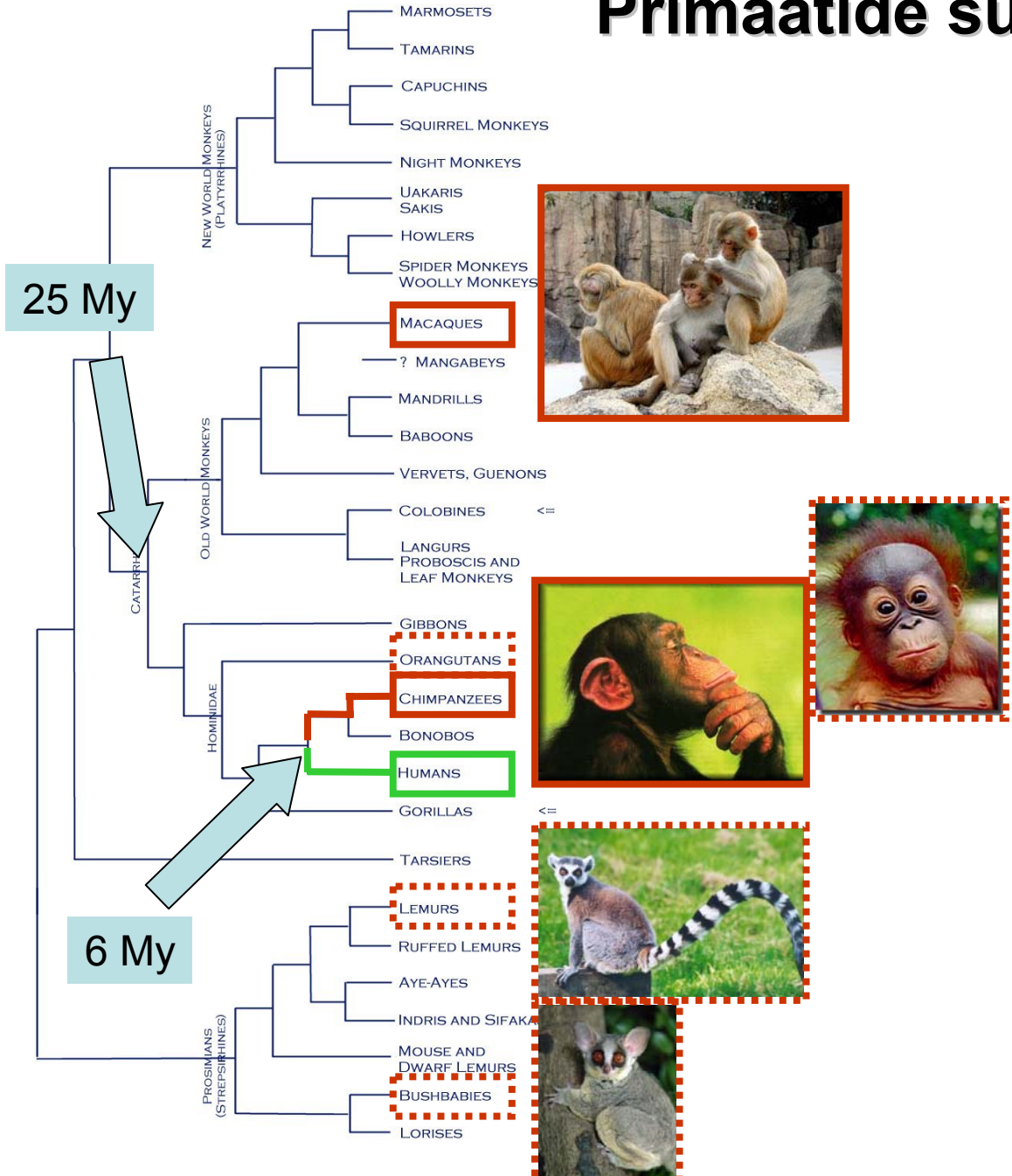


Adult body weight: Male 5.6-10.9 kg, Female 4.4-10.9 kg

Lifespan: 29 years

Female sexual maturity approximately 54 months

Primaatide sugupuu



Primate genomes that have been sequenced

Primate genomes for which sequencing is in progress

Reesusmakaakide leviala



Reesusmakaagi genoom

- Joondatavad geenijärjestused **inimene-makaak** võrdluses erinevad ca **7%**
- Joondatavad geenijärjestused **inimene-shimpans** võrdluses erinevad ca **1.2%**
- Joondatavad geenijärjestused **inimene-inimene** võrdluses erinevad ca **0.1%**



Paar fakti genoomi järjestamise kohta

- Sekveneriti üks emane. Kohati täiendati järjestust ühe isase genoomiga.
- Sekvenerimisstrateegiaks oli WGS (Whole Genome Shotgun), mille lähtematerjaliks kasutati 3, 10, 35 ja 180 kb pikkuste insertidega kloonide raamatukogusid.
- Kokku sekveneriti 18 Gb järjestust kasutades traditsioonilist Sangeri tehnoloogiat.
- Assambleerimisel kasutati kolme programmi: Atlas, PCAP ja Celera Assembler. Viimane andis pikimaid kontiige, kuid üldiselt oli kokkulangevus hea ja kasutati kõigi kolme programmi tulemusi. Inimese genoomi kasutati ainult ilmsete vigade leidmiseks, kuid mitte otseselt assambleerimiseks.
- Lõpptulemusena saadud mustandgenoomis oli 301 000 kontiigi N50 pikkusega >25kb ja 123 000 scaffoldi N50 pikkusega >24Mb

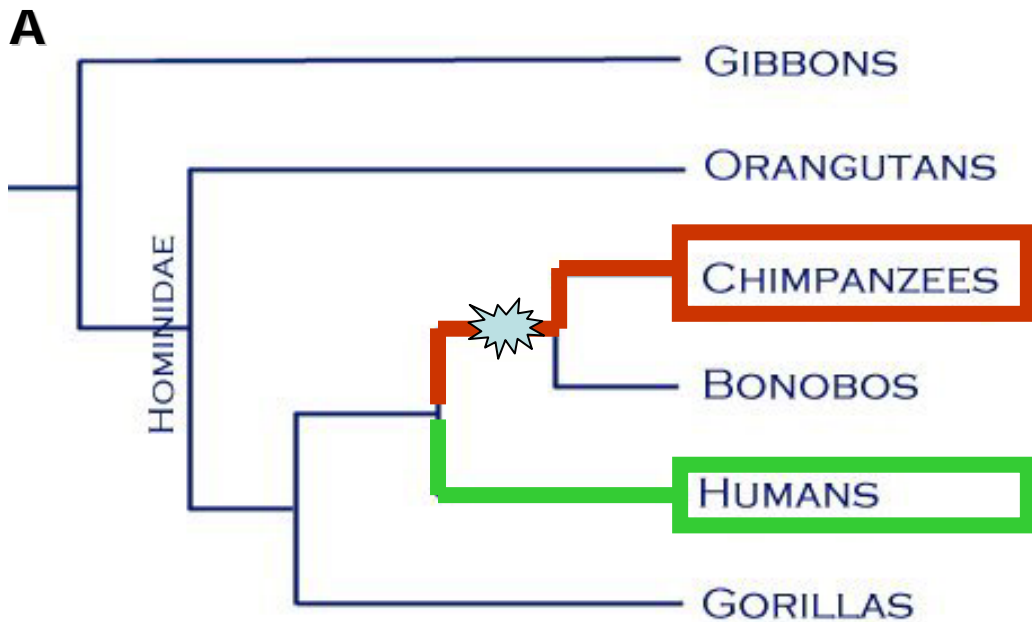
Paar fakti genoomi järjestuse kohta

- 20 autosoomi + XY kromosoomid
- GC% sarnane inimese genoomiga
- Korduste arv sarnane inimese ja shimpansi genoomiga (kokku ca 50% genoomist on kordused).
Ca 20 000 L1 ja ca 110 000 Alu kordust on tekkinud makaagi liinis.
- Geenide leidmiseks kasutati:
 - ENSEMBL tarkvara (ennustus toimub peamiselt inimese genoomi alusel)
 - Gnomon tarkvara (kasutab ka saadaolevat EST järjestuste infot)
 - Nscan tarkvara (kasutab mitmete liikide järjestuste joondusi ja cDNA järjestuste infot)

Kokku määrati ca 22 000 geeni asukoht genoomis. Neist hilisemas analüüsis kasutati 10 376 geeni (1:1:1 ortoloogid inimene-shimpans-reesusmakaak) .

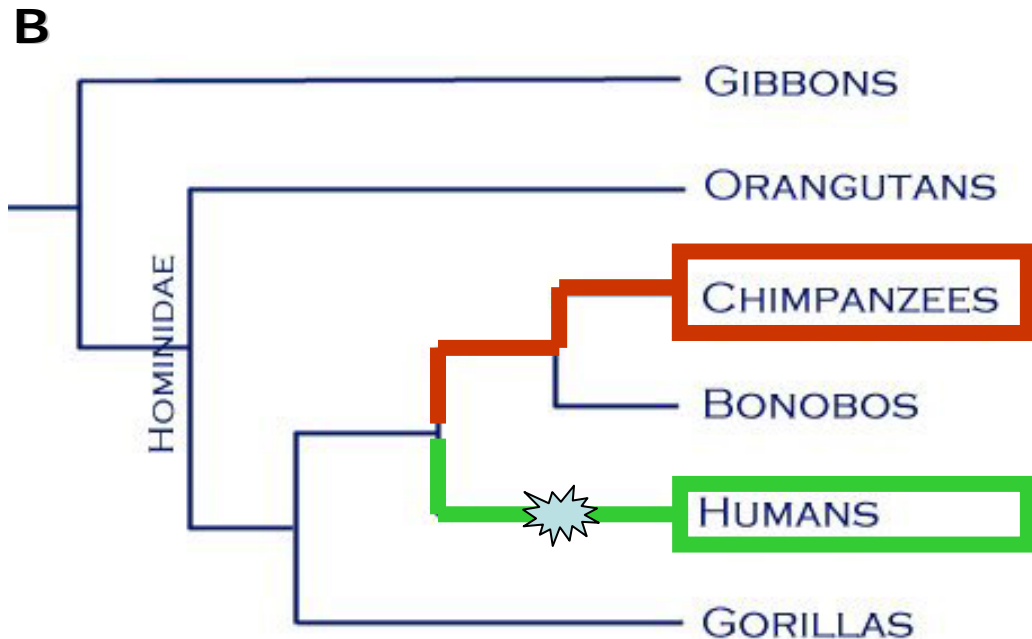
Mida saame reesusmakaagi genoomist kasulikku teada?

- Saame küsida millised uued geenid on tekkinud inimese genoomis inimese arengu jooksul.
- Ainult kahe genoomi (šimpans-inimene) võrdlusest me vastust ei saa kuna me ei tea kummas harus muutus on toimunud.



Problem #1:

It is difficult to tell whether changes have occurred in chimp lineage (A) or human lineage (B)



We would be more interested in studying changes that occurred in human lineage (variant B)

Mida saame reesusmakaagi genoomist kasulikku teada?

- Saame teada millised inimese geenid on MUUTUMAS uue funktsiooniga geenideks
- Sellist kiire muutumise protsessi nimetatakse POSITIIVSEKS SELEKTSIOONIKS või ADAPTIIVSEKS EVOLUTSIOONIKS

Positiivne selektsioon

Liikide võrdlusel on peamiseks loodusliku valiku surve mõõtühikuks oomega:

$$\omega = dN / dS$$

dN on mittesünonüümsete (aminohapet muutvate nukleotiidasenduste hulk geenis). Tuntud ka nime all K_a

dS sünonüümsete (aminohappe muutust mitte tekitavate) nukleotiidasenduste arv. Tuntud ka nime all K_s

Oomega väärtuse iga geeni kohta saab leida ortoloogsete geenide joondusest.

Positiivne selektsioon

$\omega = dN/dS$ väärtuseks on <1 kui geen on negatiivse valiku all – nõ. konserveerunud järjestus.

$\omega = dN/dS$ väärtuseks on ≈ 1 kui geen ei ole loodusliku valiku mõju all – neutraalselt evolutsioneeruv järjestus.

$\omega = dN/dS$ väärtuseks on >1 kui geen on positiivse valiku mõju all – adaptiivselt evolutsioneeruv järjestus.

Positiivne selektsioon

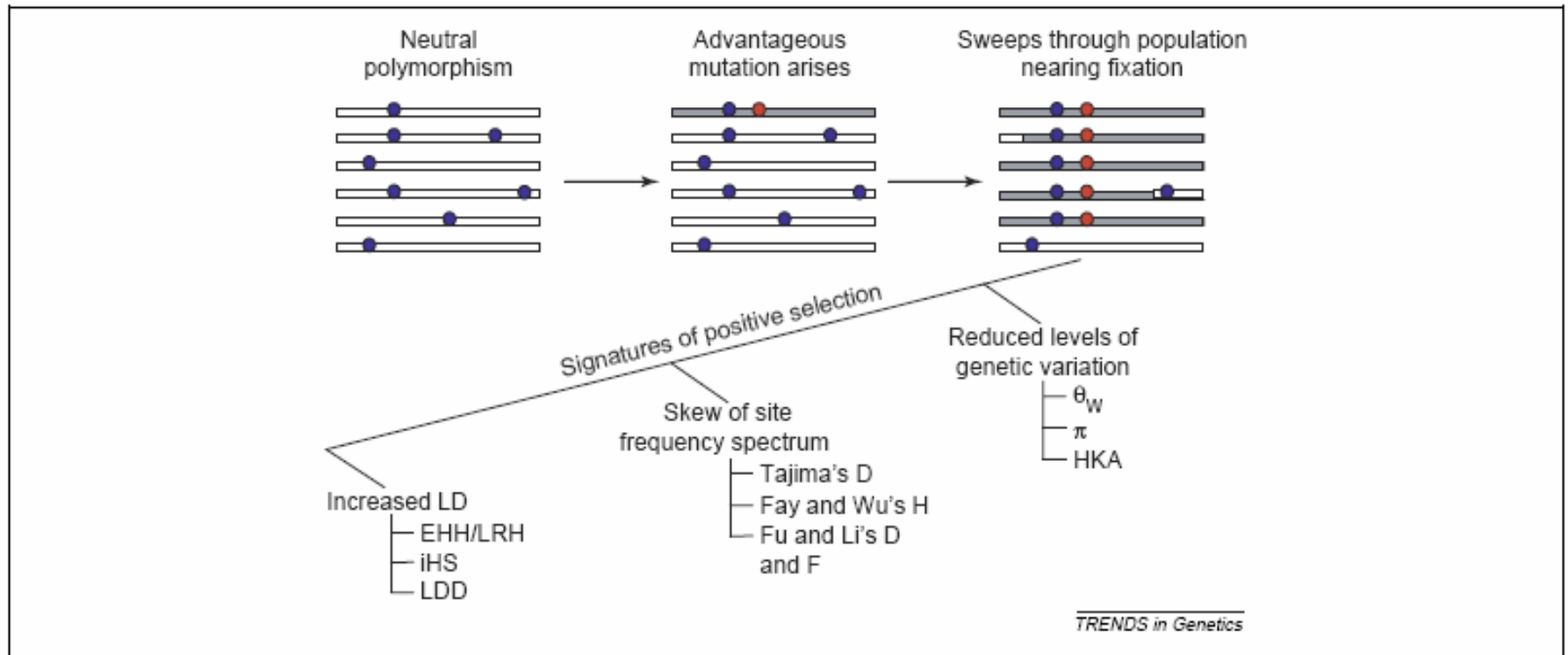
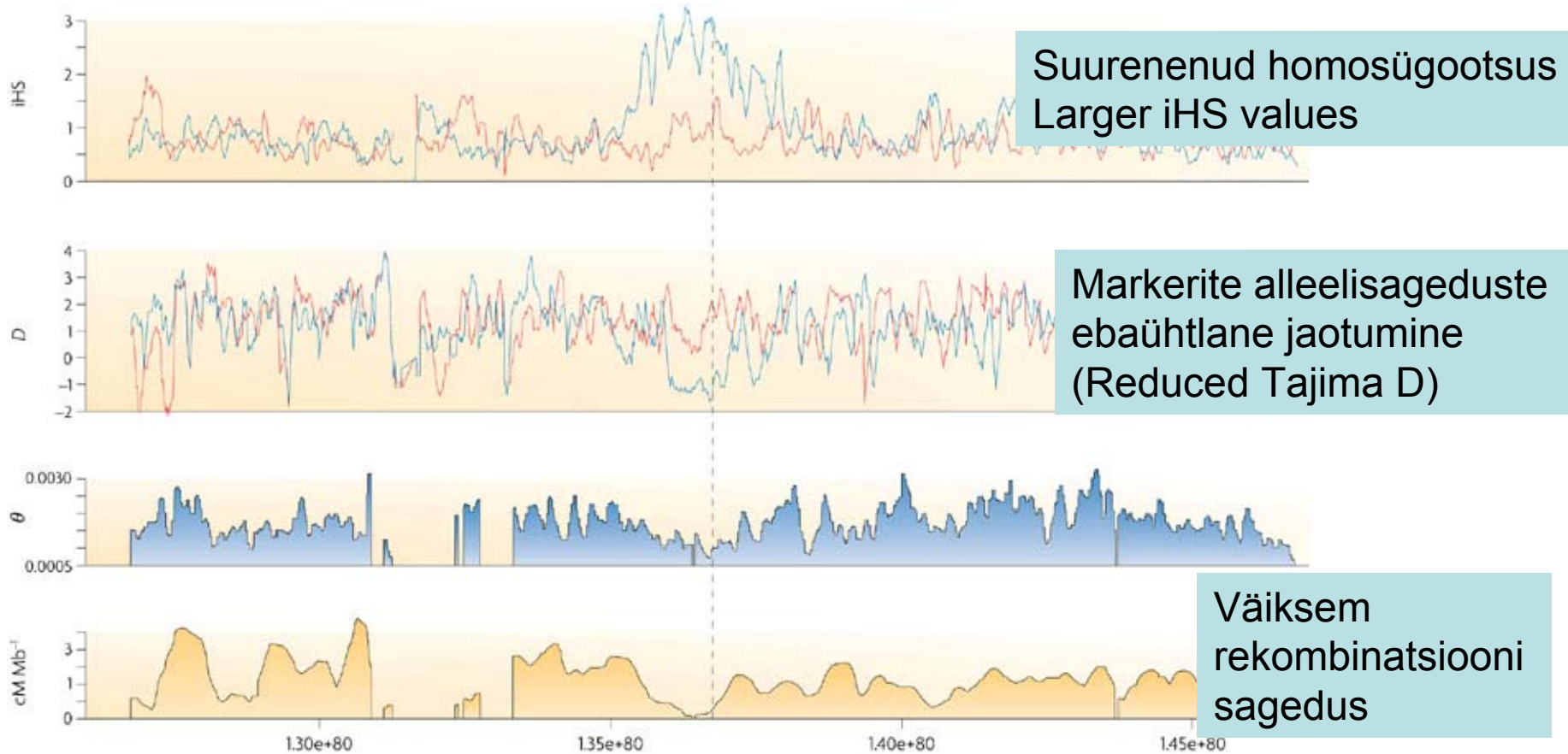


Figure 1. Signatures of positive selection. On the left, patterns of neutral polymorphism (denoted as blue circles) are shown for a sample of six haplotypes. A new advantageous mutation (indicated by the red circle) arises on a specific haplotype (middle panel highlighted in gray). As the advantageous allele increases in frequency it drags along linked neutral polymorphisms. On the right, an incomplete selective sweep is shown such that the advantageous allele has not yet reached fixation. This process perturbs patterns of genetic variation relative to neutral expectations and imparts signatures such as reduced levels of genetic variation, a skew in the site frequency spectrum (also referred to as allele frequency distribution), and increased levels of LD. Recombination between haplotypes carrying and not carrying the advantageous allele delimit the region over which the signature of selection extends. Commonly used summary statistics that have been proposed to test for these signatures are also indicated and described in more detail in Box 1. Note that the relative magnitude of these signatures of positive selection depend on many parameters such as when the advantageous allele arose, the strength of selection, whether the sweep is ongoing or has reached fixation, the amount of time that has elapsed since fixation, and local rates of recombination and mutation.

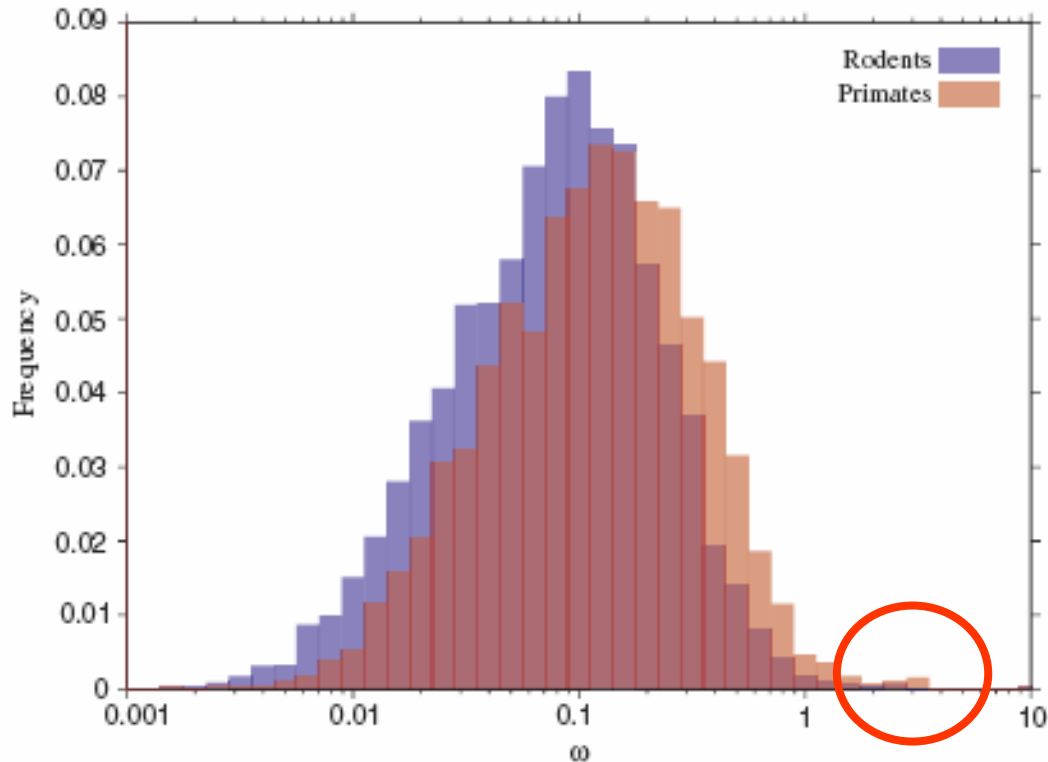
Positiivse selektsiooni detekteerimine inimeste sub-populatsioonis

The signature of an incomplete selective sweep in the region containing the lactase (*LCT*) gene.



Nature Reviews | Genetics

Mida saame reesusmakaagi genoomist kasulikku teada?



- Antud artiklis leiti ω väärtus kõigi 10 376 human-chimp-macaque (HCM) trio jaoks ja 5461 ortoloogse kvinteti HCM-mouse-rat jaoks.
- Leiti et primaatide geenides on loodusliku valiku toime vähem negatiivne (nõrgem valiku surve) kui näriliste geenides.
- Suuremat huvi pakuvad geenid, mille ω väärtus on märkimisväärselt suurem kui 1.

On võimalik testida millised geenid muutuvad kiiresti just inimese liinis

Kolme või rohkema genoomi olemasolul on võimalik eraldi testida iga evolutsioonipuu haru kohta, millised geenid muutuvad kiiresti just selles evolutsioonipuu harus.

Antud artiklis testiti seda ainult kolme primaadi omavahelises puus.

Kasutati R.Nielsen'i väljatöötatud positiivse seleksiooni testi (MBE 2002), mis võtab arvesse geenide GC-sisaldust, koodonite evolutsiooni mudelit jpm.

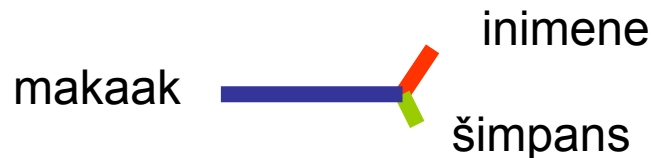
Tehti nelja sorti teste:

TA – testiti kas on toimunud positiivne valik mõnes puu harus

TH – **pos. valik inimese harus**

TC – **pos. valik šimpansi harus**

TM – **pos. valik makaagi harus**



Üksikutes primaatides kõige kiiremini evolutsioneeruvad geenid

Table 5. Selected genes from top 40 showing evidence of positive selection in primates. Accession, the number of the reference transcript for each gene (human). Chr, human chromosome on which reference gene resides. *P* value, nominal *P* value for test TA (7). Genes shown have FDR < 0.04. Test, the test (other than test TA) that detected the given gene. The Dup column has a checkmark if a gene overlaps a segmental duplication preceding the human/macaque divergence.

Accession	Gene name	Chr	Description	<i>P</i> value	Test	Dup
AB126077	<i>KRTAP5-8</i>	11	Keratin-associated protein 5-8	6.20×10^{-16}	TM	✓
NM_006669	<i>LILRB1</i>	19	Leukocyte immunoglobulin-like receptor	7.20×10^{-14}	TH, TC	✓
NM_001942	<i>DSG1</i>	18	Desmoglein 1 preproprotein	1.10×10^{-10}	–	
NM_173523	<i>MAGEB6</i>	X	Melanoma antigen family B, 6	5.30×10^{-8}	TC	✓
NM_054032	<i>MRGPRX4</i>	11	G protein-coupled receptor MRGX4	5.60×10^{-8}	TM	✓
NM_000397	<i>CYBB</i>	X	Cytochrome b-245, beta polypeptide	1.50×10^{-7}	TM	
NM_001911	<i>CTSG</i>	14	Cathepsin G preproprotein	1.50×10^{-7}	TM	
NM_000735	<i>CGA</i>	6	Glycoprotein hormones, alpha polypeptide	1.20×10^{-6}	TM	
NM_001012709	<i>KRTAP5-4</i>	11	Keratin-associated protein 5-4	2.70×10^{-6}	TM	✓
NM_000201	<i>ICAM1</i>	19	Intercellular adhesion molecule 1 precursor	2.70×10^{-6}	TM	
NM_001131	<i>CRISP1</i>	6	Acidic epididymal glycoprotein-like 1 isoform 1	1.60×10^{-5}	TM	
NM_002287	<i>LAIR1</i>	19	Leukocyte-associated immunoglobulin-like	3.10×10^{-5}	TM	✓
NM_153368	<i>CX40.1</i>	10	Connexin40.1	4.90×10^{-5}	–	
NM_018643	<i>TREM1</i>	6	Triggering receptor expressed on myeloid cells	6.30×10^{-5}	TM	
NM_000300	<i>PLA2G2A</i>	1	Phospholipase A2, group IIA	1.30×10^{-4}	–	
BC020840	<i>TCRA</i>	14	T cell receptor alpha chain C region	1.50×10^{-4}	–	
NM_000733	<i>CD3E</i>	11	CD3E antigen, epsilon polypeptide	1.50×10^{-4}	TM	
NM_001014975	<i>CFH</i>	1	Complement factor H isoform b precursor	1.50×10^{-4}	–	
NM_001423	<i>EMP1</i>	12	Epithelial membrane protein 1	1.50×10^{-4}	TM	
NM_001424	<i>EMP2</i>	16	Epithelial membrane protein 2	1.50×10^{-4}	TM	
NM_002170	<i>IFNA8</i>	9	Interferon, alpha 8	1.50×10^{-4}	–	
NM_030766	<i>BCL2L14</i>	12	BCL2-like 14 isoform 2	1.50×10^{-4}	–	
NM_006464	<i>TGOLN2</i>	2	Trans-golgi network protein 2	1.80×10^{-4}	TM	
NM_014317	<i>PDSS1</i>	10	Prenyl diphosphate synthase, subunit 1	1.80×10^{-4}	–	
NM_000518	<i>HBB</i>	11	Beta globin	2.00×10^{-4}	TM	

NB! Kiiresti evolutsioneeruvad hiljuti duplitseerunud geenid

Inimese liinist leiti selle lähenemisega kaks kiiresti evolutsioneeruvat geeni

Need geenid jäid oluliseks ka pärast mitmese testimise korrigeerimist.

LILRB1 - the leukocyte immunoglobulin-like receptor
On kiiresti muutuv ka šimpansi liinis

LOC399947 – hypothetical protein, unknown function, ca 200 aminohapet

Šimpansi liinis seevastu leiti 14 geeni mis olid samade kriteeriumide järgi positiivse valiku all.

Keskmisest kiiremini arenevad funktsioonid:

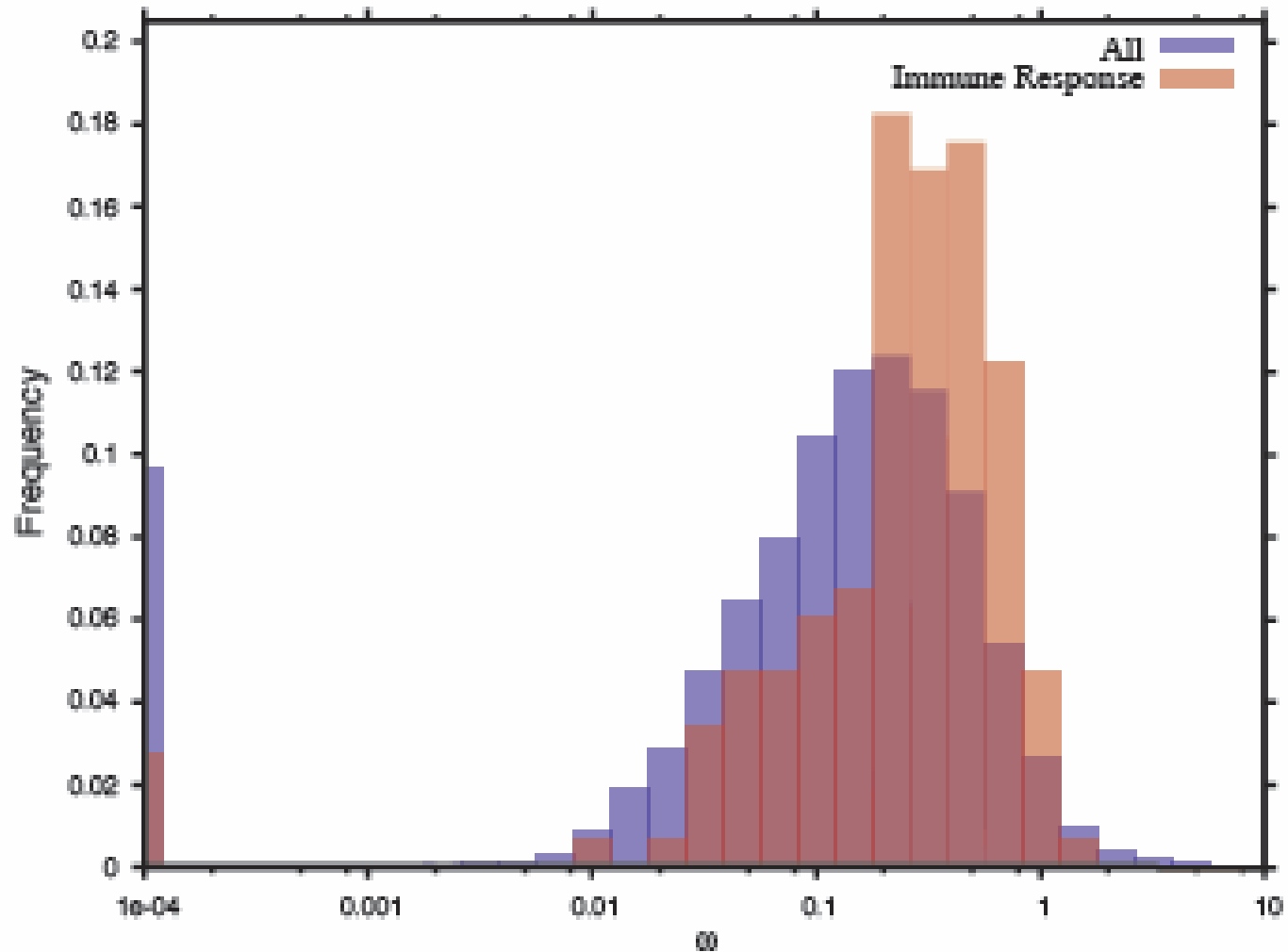
Table S6.6: GO Categories Showing an Excess of High Likelihood Ratios

Category	Description	N^a	P_A^b	P_H^c	P_C^d	P_M^e
GO:0051869	physiological response to stimulus	677	7.98e-13	2.94e-02	3.03e-03	2.66e-09
GO:0006955	immune response	321	1.42e-08	9.80e-02	8.61e-03	9.32e-08
GO:0002376	immune system process	378	5.84e-08	6.59e-02	1.85e-02	8.40e-08
GO:0002245	physiological response to wounding	185	2.27e-07	7.62e-01	1.72e-01	4.03e-10
GO:0007606	sensory perception of chemical stimulus	52	3.18e-07	7.53e-04	1.85e-02	4.68e-04
GO:0006952	defense response	256	3.57e-07	6.45e-01	3.07e-01	8.05e-08
GO:0050874	organismal physiological process	1002	3.89e-07	2.93e-02	1.92e-02	5.99e-07
GO:0009611	response to wounding	194	8.21e-07	7.49e-01	2.34e-01	3.20e-09
GO:0050909	sensory perception of taste	10	9.18e-07	5.95e-01	5.55e-01	5.09e-06
GO:0002217	physiological defense response	214	3.79e-06	5.66e-01	3.69e-01	2.08e-07
GO:0005576	extracellular region	567	4.82e-06	1.28e-01	8.13e-02	9.38e-11
GO:0048730	epidermis morphogenesis	13	2.90e-05	3.14e-03	6.56e-03	1.64e-01
GO:0009913	epidermal cell differentiation	13	2.90e-05	3.14e-03	6.56e-03	1.64e-01
GO:0007596	blood coagulation	40	8.97e-05	4.22e-01	8.78e-03	1.29e-05
GO:0009605	response to external stimulus	264	1.06e-04	6.98e-01	3.80e-01	1.00e-06
GO:0006954	inflammatory response	145	1.06e-04	7.85e-01	6.14e-01	6.82e-07
GO:0042060	wound healing	43	1.09e-04	5.64e-01	9.88e-03	2.34e-05
GO:0004984	olfactory receptor activity	40	2.28e-04	1.20e-05	6.89e-03	3.79e-02
GO:0005882	intermediate filament	56	1.69e-03	2.53e-02	8.18e-03	3.96e-05
GO:0045111	intermediate filament cytoskeleton	56	1.69e-03	2.53e-02	8.18e-03	3.96e-05
GO:0005615	extracellular space	241	1.88e-03	7.23e-02	9.53e-02	6.35e-06
GO:0045087	innate immune response	38	3.37e-03	6.73e-01	1.23e-01	2.93e-05
GO:0002541	activation of plasma proteins during acute inflammatory response	19	5.71e-03	7.32e-01	2.29e-01	1.37e-05
GO:0006956	complement activation	19	5.71e-03	7.32e-01	2.29e-01	1.37e-05
GO:0002526	acute inflammatory response	30	6.66e-03	6.77e-01	8.23e-01	3.32e-05
GO:0006118	electron transport	167	8.36e-03	4.72e-01	5.88e-01	3.49e-05
GO:0044421	extracellular region part	369	1.24e-02	3.40e-01	2.49e-01	6.78e-07

^aNumber of genes from set of 10,376 that were classified in each category or one of its descendant categories. Categories with $N < 10$ are excluded.

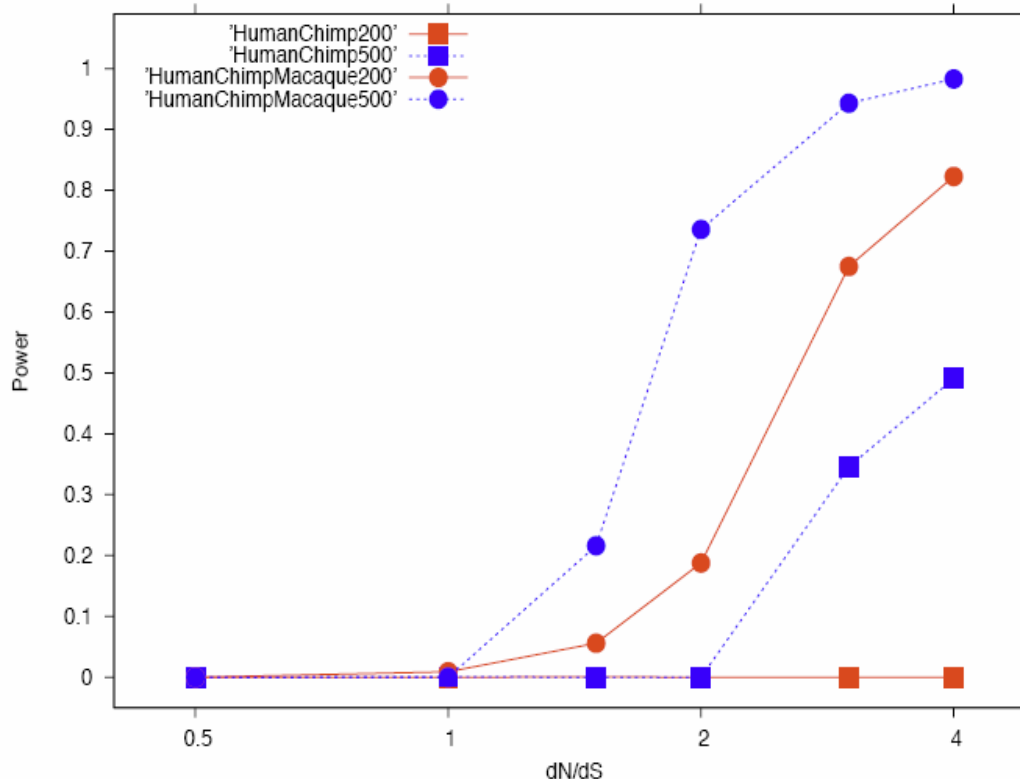
^bNominal one-sided P -value from MWU test of log likelihood ratios from test T_A (any branch). A small P -value indicates a significant shift toward larger T_A -based log likelihood ratios among genes within a category relative to genes not in the category. Bold indicates significance at 0.1 level after a conservative correction for FWER (Holm). Only categories significant in at least one test are shown.

Immuunsüsteemi geenid muutuvad keskmisest kiiremini



Mida andis meile makaagi genoomi järjestus?

Võimaluse analüüsida täpsemalt primaatide liinis toimuvaid evolutsioonilisi protsesse. Peamine eelis on analüüsi võimsuse suurenemine (muutuvate nukleotiidide arv on geenides üldiselt väike).



Eelnevates töödes (Nielsen 2005, PLoS Biology) kui kasutati ainult šimpansi ja inimese genoomi ei suudetud sama meetodikaga inimese liinis avastada ühtegi statistilise olulisuse nivood ületavat kiiresti evolutsioneeruvat geeni.

Huvitavaid näiteid üksikute mutatsioonide tasemel

A Ancestral alleles are now mutations in human

	<i>ABCA4</i> Lys223Gln							<i>CFTR</i> Phe87Leu						
boreoeutherian	R	V	T	Q	A	G	R	Y	G	I	L	L	Y	L
primate
catarrhine
hominid	F	.	.	.
human mutation	F	.	.	.
human normal	.	.	.	K	F	.	.	.
chimp	F	.	.	.
macaque
rat	.	.	A	.	.	.	W	.	.	V
mouse
rodent	W
cow	A	I	.	.	.
dog
elephant
opossum	.	.	G	H	L	.	M	H	.	.	I	.	.	.

Näide geenidest kus inimeses tekkinud variatsioon on peaaegu kinnistunud ja haruldane antsestraalne järjestust taastav mutatsioon tekitab haigusi.

ABCA4: Stargardt disease – sensitivity to light, vision loss

CFTR: tsüstiline fibroos

Kõikide liikide muutumisel mängib olulist rolli immuunsüsteemi kiire muutumine

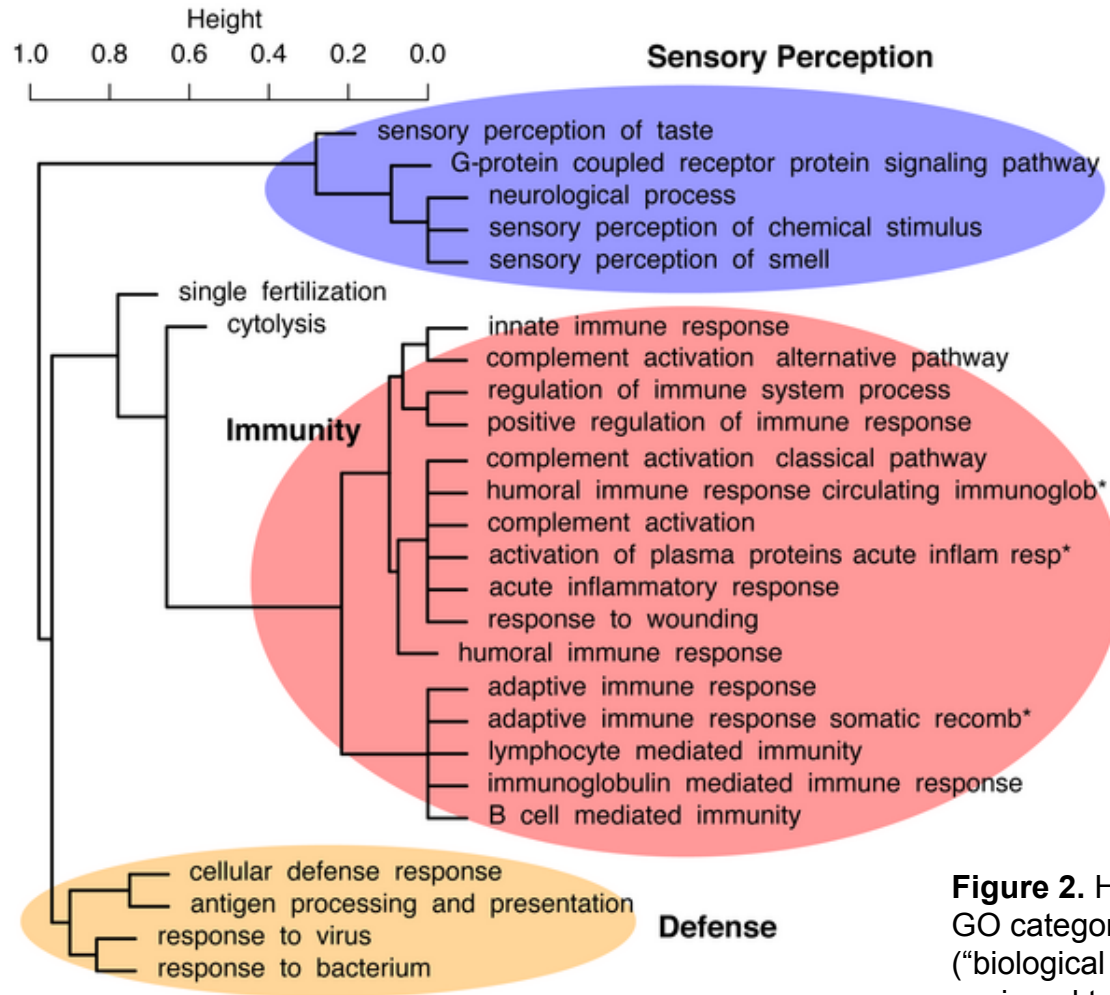


Figure 2. Hierarchical clustering of 27 over-represented GO categories identified by the Mann-Whitney *U* test (“biological process” group only), based on the genes assigned to each category.

Subjekttiivne kokkuvõte

- Immuunsüsteem, maitse- ja lõhnaretseptorid ning viljastamisega seotud geenid on olulised liikide tekkel.
- Inimesele kui liigile omaste geenide leidmine on raske, vähese võimsusega ja väga tundlik parameetrite valikule.
- Suurem osa infot positiivse valiku kohta valke kodeerivatest geenidest. Ülegenoomseid analüüse on tehtud suhteliselt vähe, kasutades näiteks HapMap projekti andmeid.
Need kirjeldavad vaid suhteliselt hiliseid evolutsioonilisi sündmusi.
- Puudub süstemaatiline info valiku all olevate geenide ja genoomi piirkondade kohta.