http://bioinfo.cipf.es

Genómica Comparativa y Biomedicina:

Predicción de Mutantes Deletéreos en el Genoma Humano

Hernán J. Dopazo

e-mail: hdopazo@cipf.es *http://bioinfo.cipf.es/hdopazo*

Departamento de Bioinformática
Unidad de Genómica Comparativa
Centro de Investigación Príncipe Felipe (CIPF)
Valencia, España

Sociedad Española de Genética
3er Curso de Genética Humana
Valencia
Enero de 2008

Bioinformatics

PRINCIPE FELIPE
CENTRO DE INVESTIGACION

http://bioinfo.cipf.es

3 Principal Investigators
2 Researchers
1 Postdoctral Fellow
2 Staticians
5 Bioinformaticians
8 Predoctoral Fellows
1 System Manager
1 Administrative Assistant

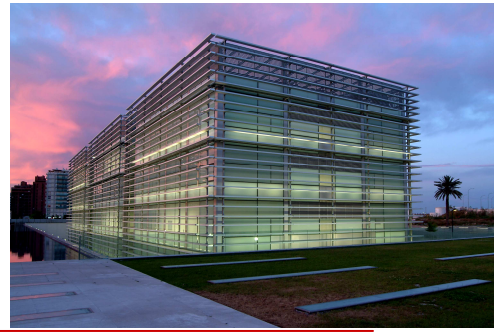
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

IBM, SGI and Bull Clusters
~**300 cores** (64 bits)

BSC, Barcelona
Mare Nostrum
10000 CPUs

Bioinformatics

PRINCIPE FELIPE
CENTRO DE INVESTIGACION



Head of Department

No-related

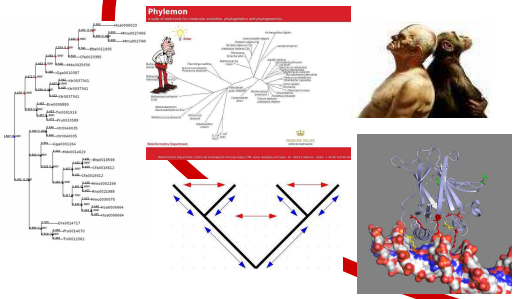
Hernán Dopazo
Comparative Genomics Unit

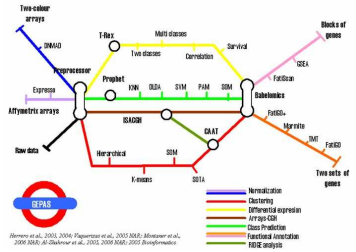
Ximo Dopazo
Functional Genomics Unit

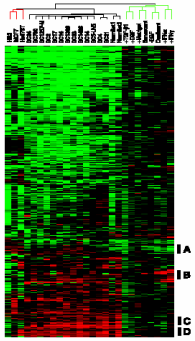
Marc A. Marti-Renom
Structural Genomics Unit

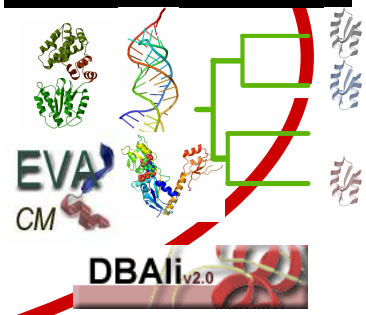
BABELOMICS

GEPAS









Pupa Suite

SIDE

Ensembl

INB INSTITUTO NACIONAL DE BIOINFORMATICA

CeGen Centro Nacional de Genotipado

Ricccc Red Temática de Investigación Cooperativa de Centros de Cáncer

MINISTERIO DE EDUCACIÓN Y CIENCIA

Genoma España

BBVA

GENERALITAT VALENCIANA

Tool usage around the World

You can see in this map the visits we had during the last day. It is automatically refreshed every five minutes.

Visits from the Antarctica are not really penguins. They are the visits we couldn't locate.

Hourly report	Daily report	Weekly report

Research Overview

Adaptive Evolution of the Human Genome

Positive selection / relaxation- *PLoS Comp. Biol., 2006*

Tissue-specific genes evolution- *working on*

Human Disease and Natural Selection

nsSNPs functional prediction- *JMB, 2006; HM 2007*

Pupas Web server- *NAR, 2006*

Phylogenetics/omics

Phylemon Web server- *NAR, 2007*

The Human Phylome- *Genome Biol. 2007*

Research Overview

Adaptive Evolution of the Human Genome

Positive selection / weak selection- *PLoS Comp. Biol., 2006*

Tissue-specific genes evolution- *working on*

Human Disease and Natural Selection

nsSNPs functional prediction- *JMB, 2006; HM 2007*

Pupas Web server- *NAR, 2006*

Phylogenetics/omics

Phylemon Web server- *NAR, 2007*

The Human Phylome- *Genome Biol. 2007*

Research Overview

Adaptive Evolution of the Human Genome

Positive selection / weak selection- *PLoS Comp. Biol., 2006*

Tissue-specific genes evolution- *working on*

Human Disease and Natural Selection

nsSNPs functional prediction- *JMB, 2006; HM 2007*

Pupas Web server- *NAR, 2006*

Phylogenetics/omics

Phylemon Web server- *NAR, 2007*

The Human Phylome- *Genome Biol. 2007*

Research Overview

Adaptive Evolution of the Human Genome

Positive selection / weak selection- *PLoS Comp. Biol., 2006*

Tissue-specific genes evolution- *working on*

Human Disease and Natural Selection

nsSNPs functional prediction- *JMB, 2006; HM 2007*

Pupas Web server- *NAR, 2006*

Phylogenetics/omics

Phylemon Web server- *NAR, 2007*

The Human Phylome- *Genome Biol. 2007*

Ancient Events of Positive Selection on Human-Chimp Genomes

Main Publications from 2003 to 2007

Inferring Nonneutral Evolution from Human-Chimp-Mouse Orthologous Gene Trios

Andrew G. Clark,[1] Stephen Glanowski,[3] Rasmus Nielsen,[2] Paul D. Thomas,[4] Anish Kejariwal,[4] Melissa A. Todd,[2] David M. Tanenbaum,[5] Daniel Civello,[6] Fu Lu,[5] Brian Murphy,[3] Steve Ferriera,[3] Gary Wang,[3] Xianqgun Zheng,[5] Thomas J. White,[6] John J. Sninsky,[6] Mark D. Adams,[5]* Michele Cargill[6]†

Open access, freely available online PLoS BIOL

A Scan for Positively Selected Genes in the Genomes of Humans and Chimpanzees

Rasmus Nielsen[1,2*], Carlos Bustamante[1], Andrew G. Clark[3], Stephen Glanowski[4], Timothy B. Sackton[3], Melissa J. Hubisz[1], Adi Fledel-Alon[1], David M. Tanenbaum[5], Daniel Civello[6], Thomas J. White[6], John J. Sninsky[6], Mark D. Adams[5¤], Michele Cargill[6]

1 Biological Statistics and Computational Biology, Cornell University, Ithaca, New York, United States of America, 2 Center for Bioinformatics, University of Copenhagen, Denmark, 3 Molecular Biology and Genetics, Cornell University, Ithaca, New York, United States of America, 4 Applied Biosystems, Rockville, Maryland, United States of America, 5 Celera Genomics, Rockville, Maryland, United States of America, 6 Celera Diagnostics, Alameda, California, United States of America

OPEN ACCESS Freely available online PLoS COMPUTATIONAL BIOLOGY

Positive Selection, Relaxation, and Acceleration in the Evolution of the Human and Chimp Genome

Leonardo Arbiza[1], Joaquín Dopazo[2], Hernán Dopazo[1*]

1 Pharmacogenomics and Comparative Genomics Unit, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain, 2 Functional Genomics Unit, Bioinformatics Department, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain

Initial sequence of the chimpanzee genome and comparison with the human genome

The Chimpanzee Sequencing and Analysis Consortium*

Natural selection on protein-coding genes in the human genome

Carlos D. Bustamante[1], Adi Fledel-Alon[1], Scott Williamson[1], Rasmus Nielsen[1,2], Melissa Todd Hubisz[1], Stephen Glanowski[3], David M. Tanenbaum[3], Thomas J. White[4], John J. Sninsky[4], Ryan D. Hernandez[1], Daniel Civello[4], Mark D. Adams[5], Michele Cargill[4]* & Andrew G. Clark[6]*

PNAS

More genes underwent positive selection in chimpanzee evolution than in human evolution

Margaret A. Bakewell, Peng Shi, and Jianzhi Zhang*

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109

Communicated by Morris Goodman, Wayne State University School of Medicine, Detroit, MI, February 26, 2007 (received for review Dece

Observations of numerous dramatic and presumably adaptive phenotypic modifications during human evolution prompt the common belief that more genes have undergone positive Darwin- number of deficiencies. First, both studies used outgroup, to distinguish between human-spec specific nucleotide substitutions, because of the

Cell, Vol. 119, 1027–1040, December 29, 2004, Copyright ©2004 by Cell Press

Accelerated Evolution of Nervous System Genes in the Origin of *Homo sapiens*

Steve Dorus,[1,2,4] Eric J. Vallender,[1,2,4] Patrick D. Evans,[1,2] Jeffrey R. Anderson,[1] Sandra L. Gilbert,[1] Michael Mahowald,[1] Gerald J. Wyckoff,[1,3] Christine M. Malcom,[1,3] and Bruce T. Lahn[1,*]

(Jerison, 1973; Byrne and Wh 1990; Matsuzawa, 2001). Mo sis of brain evolution has eme able discussion. Of particula garding what genes underlie

Comparing Human-Chimp PSGs studies

	#PSGs	GO terms (Biol. Process)	GO-PSGs diffs	Mult. Test Corr	Diff. PS-RX	H-C diff. analy	ML or Ka/Ks
Clark, et al. Science 2003 Celera	7,645 1,547H 1,534C	**Olfactation, Sensory perception' G-PCR**	YES	NO	NO	YES	ML
Nielsen, et al. PLoS Biol 2005 Celera	8,079 733 (H-C) 35 p<0.05	**Immune response,Sensory perception, Spermatogenesis, Apoptosis, Cell cyle**	---	YES	NO	NO	ML
The Chimp Seq Anal Consortium Nature 2005	13,454: 585 (H-C)	**Spermatogenesis, Perception of sound, Reproduction, Olfactation, Immune response**	YES NO	YES	YES	YES	Ka/Ks
Bustamante, et al. Nature 2005 Celera	10,767: 304H	**Apoptosis, Gametogenesis, Immune response, Sensory perception, mRNA transcription, Transcription factor**	---	---	NO	NO	MK
Arbiza, et al. PLoS CB2006 Ensembl	13,198 108-H 577-C	**Transcription factor, Cell. Prot. Metab. Olfactation, Immune response, G-PCR**	NO	YES	YES	YES	ML
Zhang, J. et al 2007 PNAS 2007 Ensembl	13,188Ñ 154H, 233C 2H, 21C	**Protein metabolism, mRNA transcrip., Ectoderm development, stress response, G PCR signaling**	YES	YES NO	YES	YES	ML

Comparing Human-Chimp PSGs studies

	#PSGs	GO terms (Biol. Process)	GO-PSGs diffs	Mult. Test Corr	Diff. PS-RX	H-C diff. analy	ML or Ka/Ks
Clark, et al. Science 2003 Celera	7,645 1,547H 1,534C	**Olfactation, Sensory perception' G-PCR**	YES	NO	NO	YES	ML
Nielsen, et al. PLoS Biol 2005 Celera	8,079 733 (H-C) 35 p<0.05	**Immune response,Sensory perception, Spermatogenesis, Apoptosis, Cell cyle**	---	YES	NO	NO	ML
The Chimp Seq Anal Consortium Nature 2005	13,454: 585 (H-C)	**Spermatogenesis, Perception of sound, Reproduction, Olfactation, Immune response**	YES NO	YES	YES	YES	Ka/Ks
Bustamante, et al. Nature 2005 Celera	10,767: 304H	**Apoptosis, Gametogenesis, Immune response, Sensory perception, mRNA transcription, Transcription factor**	---	---	NO	NO	MK
Arbiza, et al. PLoS CB2006 Ensembl	13,198 108-H 577-C	**Transcription factor, Cell. Prot. Metab. Olfactation, Immune response, G-PCR**	NO	YES	YES	YES	ML
Zhang, J. et al 2007 PNAS 2007 Ensembl	13,188Ñ 154H, 233C 2H, 21C	**Protein metabolism, mRNA transcrip., Ectoderm development, stress response, G PCR signaling**	YES	YES NO	YES	YES	ML

Comparing Human-Chimp PSGs studies

	#PSGs	GO terms (Biol. Process)	GO-PSGs diffs	Mult. Test Corr	Diff. PS-RX	H-C diff. analy	ML or Ka/Ks
Clark, et al. Science 2003 Celera	7,645 1,547H 1,534C	**Olfactation, Sensory perception' G-PCR**	YES	NO	NO	YES	ML
Nielsen, et al. PLoS Biol 2005 Celera	8,079 733 (H-C) 35 p<0.05	**Immune response,Sensory perception, Spermatogenesis, Apoptosis, Cell cyle**	---	YES	NO	NO	ML
The Chimp Seq Anal Consortium Nature 2005	13,454: 585 (H-C)	**Spermatogenesis, Perception of sound, Reproduction, Olfactation, Immune response**	YES NO	YES	YES	YES	Ka/Ks
Bustamante, et al. Nature 2005 Celera	10,767: 304H	**Apoptosis, Gametogenesis, Immune response, Sensory perception, mRNA transcription, Transcription factor**	---	---	NO	NO	MK
Arbiza, et al. PLoS CB2006 Ensembl	13,198 108-H 577-C	**Transcription factor, Cell. Prot. Metab. Olfactation, Immune response, G-PCR**	NO	YES	YES	YES	ML
Zhang, J. et al 2007 PNAS 2007 Ensembl	13,188Ñ 154H, 233C 2H, 21C	**Protein metabolism, mRNA transcrip., Ectoderm development, stress response, G PCR signaling**	YES	YES NO	YES	YES	ML

Comparing Human-Chimp PSGs studies

	#PSGs	GO terms (Biol. Process)	GO-PSGs diffs	Mult. Test Corr	Diff. PS-RX	H-C diff. analy	ML or Ka/Ks
Clark, et al. Science 2003 Celera	7,645 1,547H 1,534C	**Olfactation, Sensory perception' G-PCR**	YES	NO	NO	YES	ML
Nielsen, et al. PLoS Biol 2005 Celera	8,079 733 (H-C) 35 p<0.05	**Immune response,Sensory perception, Spermatogenesis, Apoptosis, Cell cyle**	---	YES	NO	NO	ML
The Chimp Seq Anal Consortium Nature 2005	13,454: 585 (H-C)	**Spermatogenesis, Perception of sound, Reproduction, Olfactation, Immune response**	YES NO	YES	YES	YES	Ka/Ks
Bustamante, et al. Nature 2005 Celera	10,767: 304H	**Apoptosis, Gametogenesis, Immune response, Sensory perception, mRNA transcription, Transcription factor**	---	---	NO	NO	MK
Arbiza, et al. PLoS CB2006 Ensembl	13,198 108-H 577-C	**Transcription factor, Cell. Prot. Metab. Olfactation, Immune response, G-PCR**	NO	YES	YES	YES	ML
Zhang, J. et al 2007 PNAS 2007 Ensembl	13,188Ñ 154H, 233C 2H, 21C	**Protein metabolism, mRNA transcrip., Ectoderm development, stress response, G PCR signaling**	YES	YES NO	YES	YES	ML

Comparing Human-Chimp PSGs studies

	#PSGs	GO terms (Biol. Process)	GO-PSGs diffs	Mult. Test Corr	Diff. PS-RX	H-C diff. analy	ML or Ka/Ks
Clark, et al. Science 2003 Celera	7,645 1,547H 1,534C	**Olfactation, Sensory perception' G-PCR**	YES	NO	NO	YES	ML
Nielsen, et al. PLoS Biol 2005 Celera	8,079 733 (H-C) 35 p<0.05	**Immune response,Sensory perception, Spermatogenesis, Apoptosis, Cell cyle**	---	YES	NO	NO	ML
The Chimp Seq Anal Consortium Nature 2005	13,454: 585 (H-C)	**Spermatogenesis, Perception of sound, Reproduction, Olfactation, Immune response**	YES NO	YES	YES	YES	Ka/Ks
Bustamante, et al. Nature 2005 Celera	10,767: 304H	**Apoptosis, Gametogenesis, Immune response, Sensory perception, mRNA transcription, Transcription factor**	---	---	NO	NO	MK
Arbiza, et al. PLoS CB2006 Ensembl	13,198 108-H 577-C	**Transcription factor, Cell. Prot. Metab. Olfactation, Immune response, G-PCR**	NO	YES	YES	YES	ML
Zhang, J. et al 2007 PNAS 2007 Ensembl	13,188Ñ 154H, 233C 2H, 21C	**Protein metabolism, mRNA transcrip., Ectoderm development, stress response, G PCR signaling**	YES	YES NO	YES	YES	ML

Comparing Human-Chimp PSGs studies

	#PSGs	GO terms (Biol. Process)	GO-PSGs diffs	Mult. Test Corr	Diff. PS-RX	H-C diff. analy	ML or Ka/Ks
Clark, et al. Science 2003 Celera	7,645 1,547H 1,534C	**Olfactation, Sensory perception' G-PCR**	YES	NO	NO	YES	ML
Nielsen, et al. PLoS Biol 2005 Celera	8,079 733 (H-C) 35 p<0.05	**Immune response,Sensory perception, Spermatogenesis, Apoptosis, Cell cyle**	---	YES	NO	NO	ML
The Chimp Seq Anal Consortium Nature 2005	13,454: 585 (H-C)	**Spermatogenesis, Perception of sound, Reproduction, Olfactation, Immune response**	YES NO	YES	YES	YES	Ka/Ks
Bustamante, et al. Nature 2005 Celera	10,767: 304H	**Apoptosis, Gametogenesis, Immune response, Sensory perception, mRNA transcription, Transcription factor**	---	---	NO	NO	MK
Arbiza, et al. PLoS CB2006 Ensembl	13,198 108-H 577-C	**Transcription factor, Cell. Prot. Metab. Olfactation, Immune response, G-PCR**	NO	YES	YES	YES	ML
Zhang, J. et al 2007 PNAS 2007 Ensembl	13,188Ñ 154H, 233C 2H, 21C	**Protein metabolism, mRNA transcrip., Ectoderm development, stress response, G PCR signaling**	YES	YES NO	YES	YES	ML

Our work…why?

- **Main Questions**

Positive Selection, Relaxation, and Acceleration in the Evolution of the Human and Chimp Genome

Leonardo Arbiza[1], Joaquín Dopazo[2], Hernán Dopazo[1*]

1 Pharmacogenomics and Comparative Genomics Unit, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain, 2 Functional Genomics Unit, Bioinformatics Department, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain

- ✓ Which are the full set of genes and functions that evolve outside of the molecular clock hypothesis?
- ✓ Which are the full set of genes and functions that were positively selected during evolution of each species, and which show evidence of weak selection/relaxation?
- ✓ How do these sets of genes compare amongst themselves and in between derived and ancestral lineages at a functional level?

Positive selection, relaxation and clock

Derived and ancestral lineages

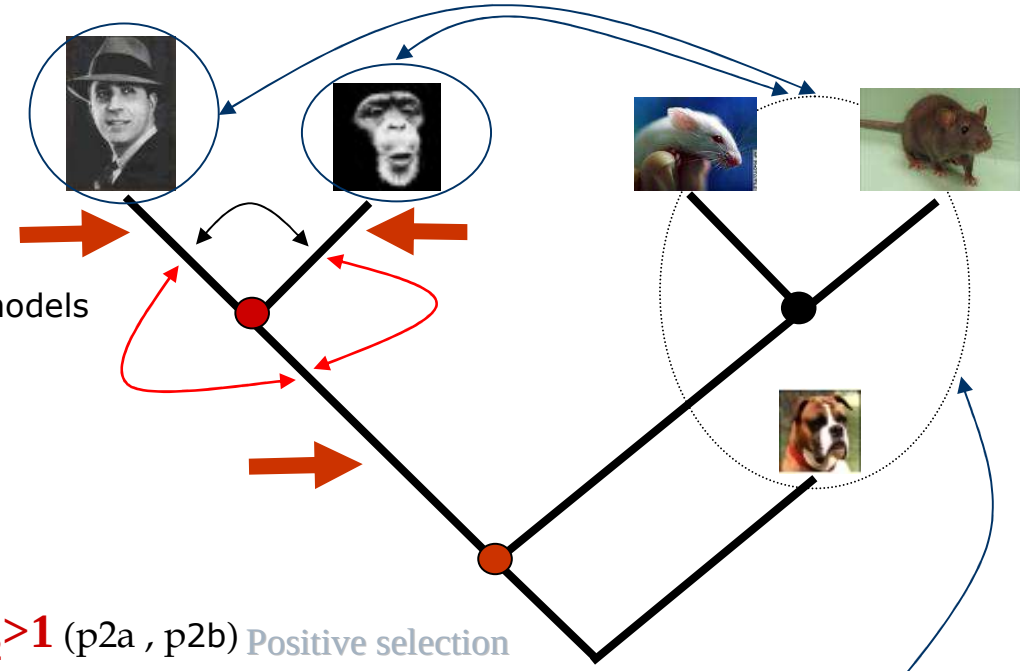
Testing PS and RX

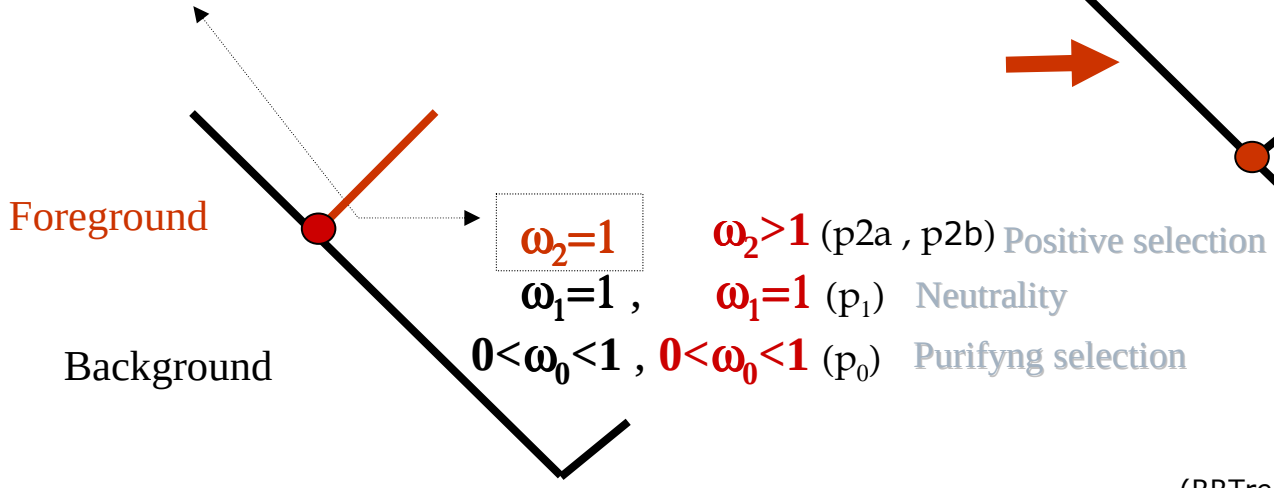
(CodeML, PAML, Zhang, Nielsen, Yang, Mol Biol Evol 2005)

Branch-site models: site classes (0, 1, 2)

Test I: M1a (neutral) vs **A** (Positive selection) models

Test II: A1 (neutral) vs **A** (Positive selection) models

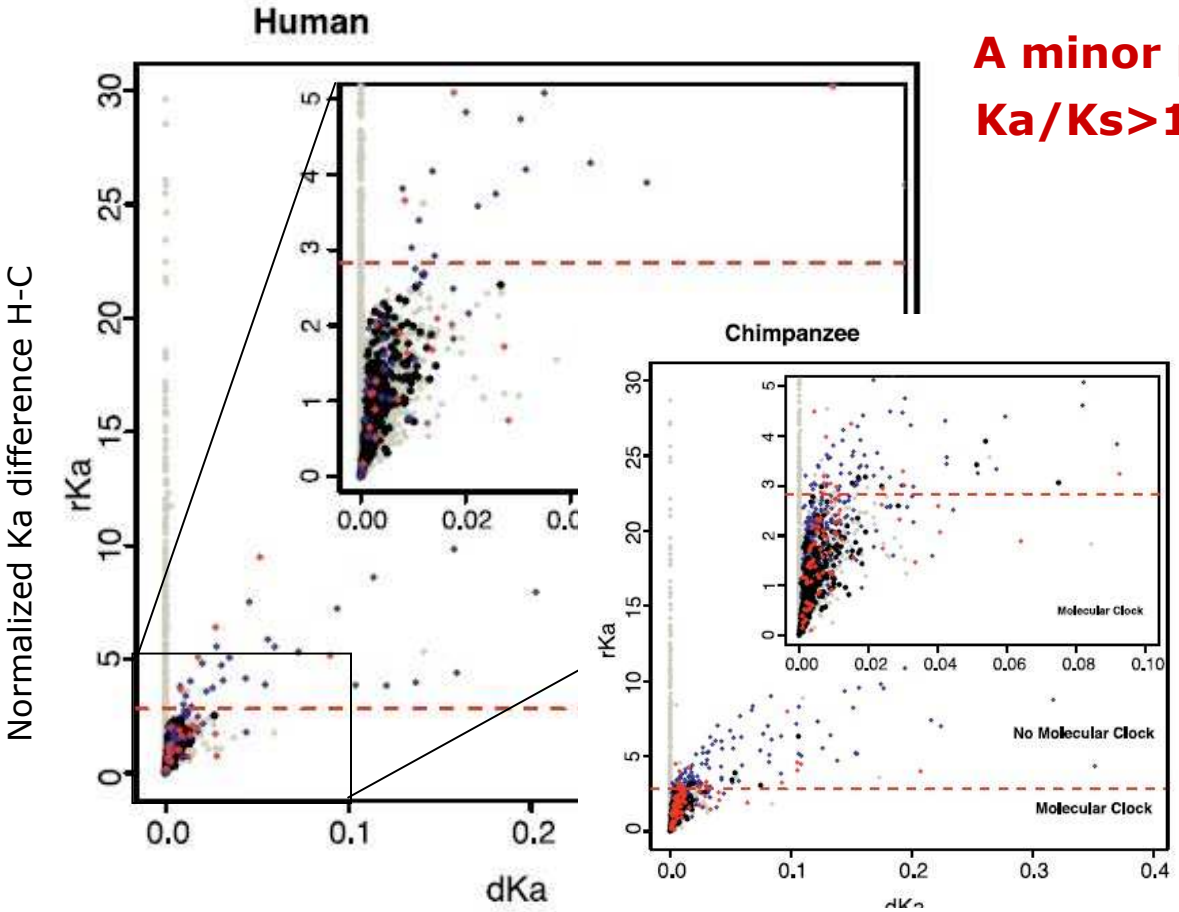




Testing Clock

(RRTree, Robinson-Rechavi, Bioinformatics 2000)

Clock assumption - Ka/Ks ratio - ML branch-site model

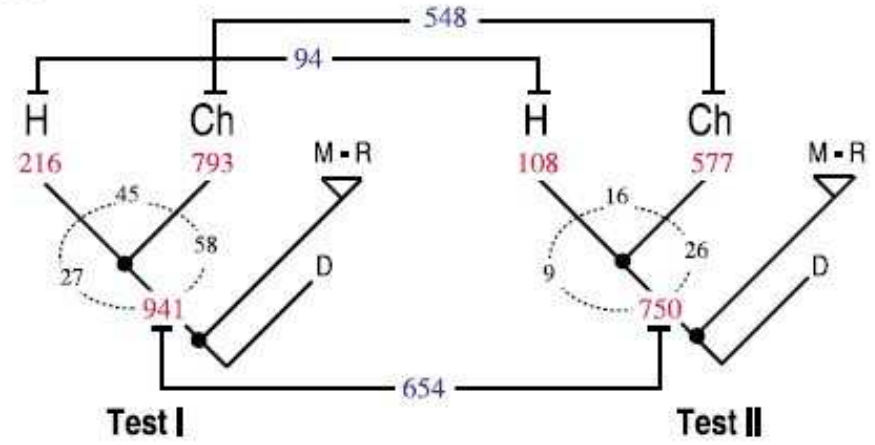


Normalized Ka difference H-C

A minor proportion of genes with Ka/Ks>1 match events of ML- PS

- **Ka-Ks > 1 and ML-PS**
- **Ka-Ks < 1 and ML-PS**
- **Ka-Ks > 1 no ML-PS**
- **Ka-Ks < 1 no ML-PS**

Distribution PSGs and RXGs



Derived – Ancestral change

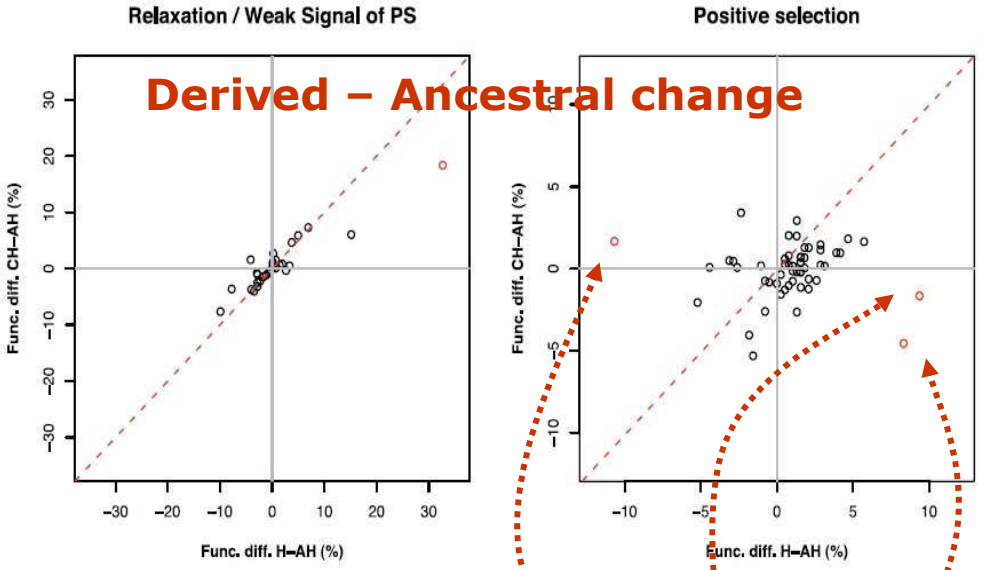


Figure 3. Ancestral and Derived Trends in Adaptation and RSC

Differences in GO term representation between the sets of the derived and the ancestral lineages (H-AH, human versus ancestral lineage; CH-AH, chimp versus ancestral lineage) are plotted against each other using genes exclusively observed in Test I (RSC) and Test II (PS). Each quadrant represents a particular evolutionary scenario increasing or decreasing in GO representation for each of the lineages after speciation. Terms showing a difference in representation between H-AH and CH-AH >10% were labeled in red: *G-coupled protein receptor* was found in both Test I (14.32%) and Test II (12.89%), and *sensory perception* (11.03%) and *cellular protein metabolism* (–12.34%) in Test II. Only the terms common to all lineages are shown.

DOI: 10.1371/journal.pcbi.0020038.g003

Functional Analysis PSGs and RXGs

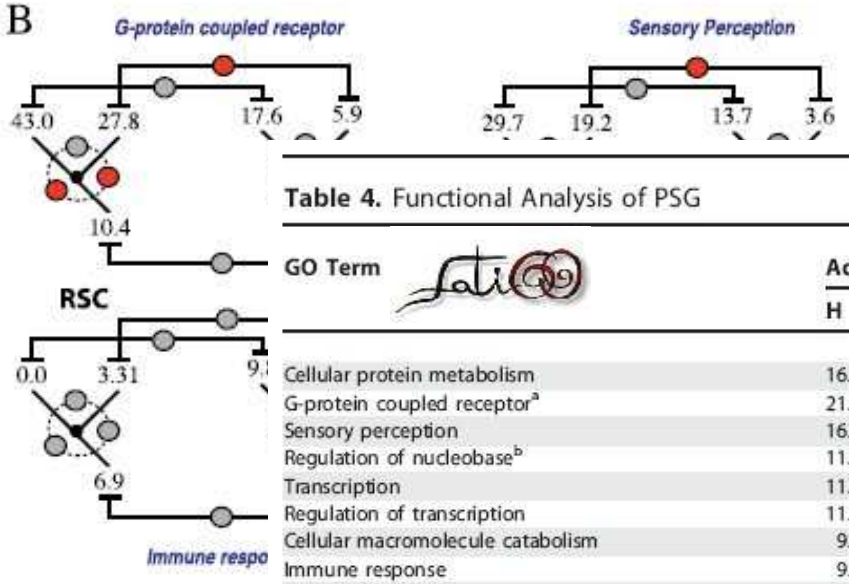


Table 4. Functional Analysis of PSG

GO Term	Adaptive Evolution				
	H	Ch	N_o	H-AH	Ch-AH
Cellular protein metabolism	16.67 (7)	31.00 (102)	3	−10.68	+1.66
G-protein coupled receptor[a]	21.43 (9)	6.08 (20)	0	+8.33	−4.56
Sensory perception	16.67 (7)	3.65 (12)	0	+9.37	−1.66
Regulation of nucleobase[b]	11.90 (5)	14.29 (47)	0	−3.12	+0.49
Transcription	11.90 (5)	15.20 (50)	0	−4.42	+0.08
Regulation of transcription	11.90 (5)	13.98 (46)	0	−2.86	+0.45
Cellular macromolecule catabolism	9.52 (4)	9.42 (31)	1	+2.09	+1.27
Immune response	9.52 (4)	4.86 (16)	0	−1.57	−5.31
Protein transport	7.14 (3)	4.86 (16)	1	+4.16	+0.96
Protein catabolism	7.14 (3)	8.81 (29)	1	+1.30	+1.99
Intracellular protein transport	7.14 (3)	3.95 (13)	1	+5.73	+1.64
Cytoskeleton organization and biogenesis	4.76 (2)	4.86 (16)	2	+4.68	+1.81
Phosphate metabolism	7.14 (3)	7.90 (26)	0	−5.21	−2.06
Cellular carbohydrate metabolism	4.76 (2)	2.74 (9)	1	+2.08	−1.24
Response to pest, pathogen, or parasite	7.14 (3)	2.43 (8)	0	−1.82	−4.05
DNA metabolism	4.76 (2)	6.08 (20)	1	+0.52	+0.59

Figure 2. Phylogeneti

(A) The differential dis
different pattern of oc
and ancestral lineage
genes detected in e
Numbers in black are
lineages. Numbers in
(B) The phylogenetic c
shown in human, in chimp, and in the ancestral lineage as depicted in

Table 5. A Small Sample of the Human and the Chimp Genes Deduced under Tests I and II

GO Description	Adaptive Evolution			Relaxation of Selective Constraints/Weak Signal of PS		
	Human	Chimpanzee	Ancestral	Human	Chimpanzee	Ancestral
Sensory perception	*EDN3 GRM6 HKR3*	*ABCA4 COL1A1 ERCC8*	*CCL1 CCL3 COL11A1*	*DFNB31 O2AG1 OR1072*	*CNGA2 GUCY2D O10D4*	*BBS2 CNGA1 COL1A2*
	OR2A14 OR51D1	*GJA3 MYH9 MYO9A*	*CRB2 DSPP GPRC5D*	*OR10K1 OR10Q1 OR2B6*	*OR13G1 OR2A12 OR2B2*	*GUCY2D OPA3 OPN1SW*
	OR5D18 TS1R1	*OR52N1OR5I1 ROBO1*	*IL8RA KPTN MYH14*	*OR3A3 OR4C16 OR51G1*	*OR2F2 OR2L2 OR4C13*	*OR11L1 OR2F2 OR4A16*
		USH1C	*O10D4 OR10T2OR1B6*	*OR52E5 OR52N1 OR5J2*	*OR4FE OR4K13 OR51B4*	*OR4C16 OR4E2 OR51B4*
			OR5A1 OR5P3 PROM1 RP1	*OR5T1 OR6K2 OR8J3*	*OR52E2 OR6P1 OR8I2*	*OR51I1 OR51V1 OR571*
			TAS1R3 TAS2R38 TAS2R41		*OR9A4 PCDH15 TAS2R60*	*OR5AS1 OR8J1 OR8J3*
			TRPA1		*TECTA*	*TNFRSF11A TRPM8*
						TULP2
G-PCR signalling pathway[a]	*GPR111 HKR3 OR52W1*	*ADRA1B ADRA2A SORCS1*	*AAR5 ADRA1A AS1R3*	*CAP1 IMPG2 OR10T2*	*ADRA1D CALM1ECE2*	*OR4C15 OR4E2 OR5T2*
	PTGER4 TS1R1	*GRPR AKAP12 TAAR1*	*CCL3 CCR2 CD3 EDRD2*	*OR3A3 Q8NG2 Q8NGU0*	*ELSR2 OR2B2 OR2T4*	*PYY TSHR*
		TAAR6 PARD3 HTR5A	*ENPP2 GABBR1 GALR2*	*Q8NH71 Q8NH88 RBP3*	*OR4C11 OR52M1 OR6K6*	
		EDG8 OR8D2	*GAP43 GLP1R GPCR116*	*TSHR*	*PLCE1 R4C13 RAMP3*	
			GPR154 GPR43 HTR1D			
			IL8RA MRGPRD OR13A1			
			PLCE1 PTHR1 RAI3			
Immune response[b]	*CCL4 ITGAL ITGB1*	*AFP AMBP CSF1R CSF2RB*	*AHSG ARTS1 AZGP1 CCL1*		CRIP1 ELF4 IVNS1ABP	CAMP CFH C1QG GSR
		GABBR1 HFE HLA-G HLA-J	*CCL3 CCR2 CD72 CD80*		ODZ1 PARP4 STAB1	INHA PRF1 PTGS2
		IGSF2 IKBKE IL1F10 IL1R1	*CRISP3 CSF3 D3E EXOSC9*			
		KLF6 OTUB1 SEMA7A	*FCGR2B FCN2 FTH1 GBP1*			
		STAT5A TCF7 UBD	*GBP2 HLA-DOB HLA-F*			
			HLA-G HLA-H ICOS IL18			
			ITGAL LIRA4 LTB4R LY75			
			NFX1 S100A9 STAT3 TLR1			
			TREM1 TRIM22			
DNA/RNA[c]	*CHTF18 NASP*	*ARID1A ERCC8 FANCG*	*ARSL CASKIN1 CIDEA*	*EPRS SUPT6H*	*ADRA1D CHD5 CHTF18*	*ADARB2 MCM3AP*
		LIG1 MSH4 MUS81 MYST3	*DCP2DHX15 ELAVL4*		*DARS IVNS1ABP NAP1L5*	*MRE11A MSH2 POLQ*
		POLD3 POLI RAD23B RFC1	*EXOSC9 HILS1 NEIL2*		*ORCL3 PAPR4 POLRMT*	*POLRMUT RPUSD4 SMN1*
		SUPT6H TOPBP1 TP73	*NF1B NFIC NFX1 NOLC1*		*SHPRH SUPT5H*	*CYCS SUPT6H SYV*
		UBE1 UV39H2 XRCC4	*OGG1 PARP2 POLE2 POLG*			*XRCC5*
			POLM POLN RAD51L3			
Transcription[d]	*ARNT2 KLF14 NFKB2*	*ASCL1 CDC5L CEPPZ*	*ATF4 BGALP BLZF1 CD80*	*AGGF1 CREM FOXI1*	*CHD5 E2F1 ELF4 GMEB2*	*CEBPZ ETS2 PER2*
		COA5 EDF1 ERCC8 FLI1	*CNOT4 DMRTC2 EGR4*	*GLIS3 GRIP1 LHX1 MYEF2*	*HNF4A HOXA1 HOXA3*	*PMFBP1 POLRMT SOX30*
		GLI1 HUWE1 KCNH5	*ERG ETV2 MEF2B MXD4*	*POLR3K SUPT6H TITF1*	*KLF3 NOC3L PHF19*	*SP110 SUPT6H TERF2IP*

Conclusions I

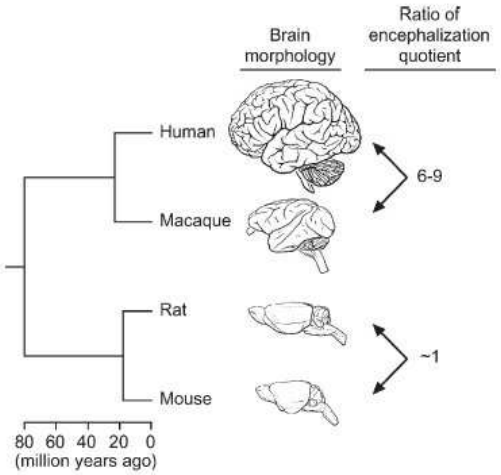
- Adaptive evolution is an infrequent process shaping the pattern of divergence between human and chimp genomes.
- Use of rate approaches (Ka/Ks) for concentrating cases of positive selection should be discarded in favor of more sensitive methodologies.
- Functional classes encompassed by the sets of genes evolving without clock and positive selection were the same and in similar proportions.
- Our analysis using the first release of the Chimp Genome from Ensembl database probable collects false positive cases of PSGs
- We are running a new analysis using an updated Chimp and Macaca genome

Positive selection: Nervous System

Cell, Vol. 119, 1027–1040, December 29, 2004, Copyright ©2004 by Cell Press

Accelerated Evolution of Nervous System Genes in the Origin of *Homo sapiens*

Steve Dorus,[1,2,4] Eric J. Vallender,[1,2,4] Patrick D. Evans,[1,2] Jeffrey R. Anderson,[1] Sandra L. Gilbert,[1] Michael Mahowald,[1] Gerald J. Wyckoff,[1,5] Christine M. Malcom,[1,3] and Bruce T. Lahn[1,*]



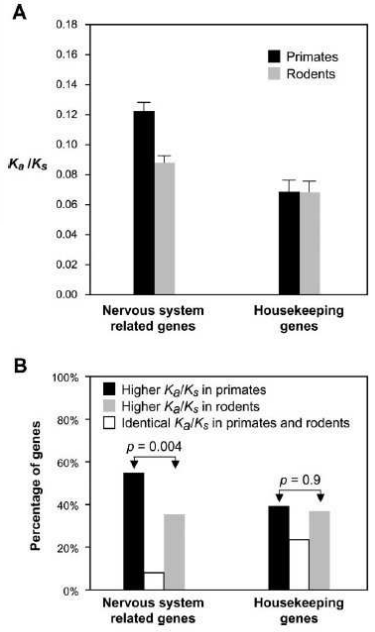
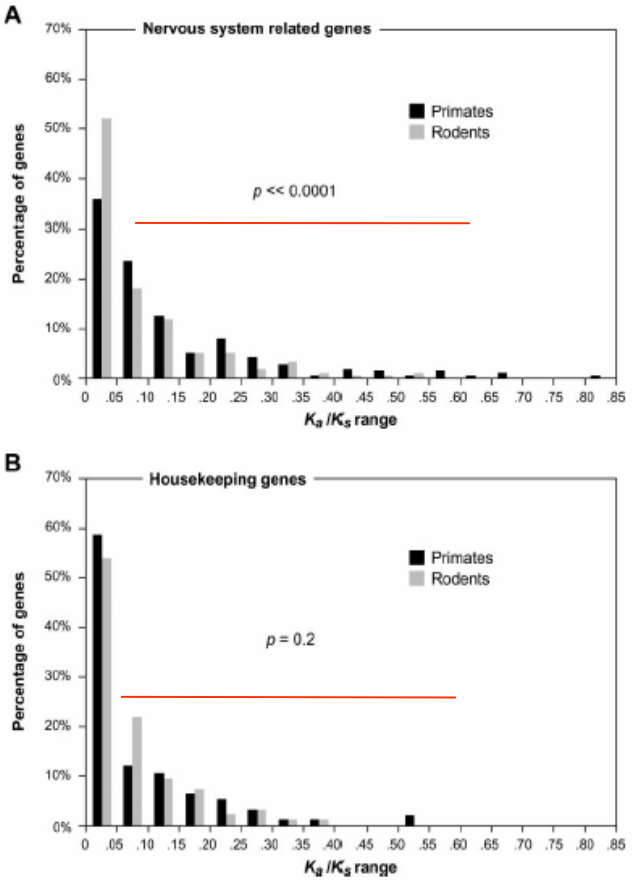


Figure 2. Evolution of Nervous System Genes and Housekeeping Genes in Primates and Rodents

(A) Evolutionary rates in primates and rodents.

(B) Percentage of genes that evolved with higher K_a/K_s in one or the other mammalian order.

The p values indicate the statistical significance of primate-rodent disparities.



Brain genes are between the most conserved genes analysed in PS studies

Open access, freely available online PLOS BIOLOGY

A Scan for Positively Selected Genes in the Genomes of Humans and Chimpanzees

Rasmus Nielsen[1,2], Carlos Bustamante[1], Andrew G. Clark[3], Stephen Glanowski[4], Timothy B. Sackton[3], Melissa J. Hubisz[1], Adi Fledel-Alon[1], David M. Tanenbaum[5], Daniel Civello[6], Thomas J. White[6], John J. Sninsky[6], Mark D. Adams[5¤], Michele Cargill[6]

1 Biological Statistics and Computational Biology, Cornell University, Ithaca, New York, United States of America, 2 Center for Bioinformatics, University of Copenhagen, Denmark, 3 Molecular Biology and Genetics, Cornell University, Ithaca, New York, United States of America, 4 Applied Biosystems, Rockville, Maryland, United States of America, 5 Celera Genomics, Rockville, Maryland, United States of America, 6 Celera Diagnostics, Alameda, California, United States of America

Since the divergence of humans and chimpanzees about 5 million years ago, these species have undergone a remarkable evolution with drastic divergence in anatomy and cognitive abilities. At the molecular level, despite the small overall magnitude of DNA sequence divergence, we might expect such evolutionary changes to leave a noticeable signature throughout the genome. We here compare 13,731 annotated genes from humans to their

Table 1. Biological Process Categories with an Excess of Putatively Positively Selected Genes (Nominal *p* less than 0.05; MWU) among a Total of 133 Biological Process Categories

Biological Process	Number of Genes	*p*-Value
Immunity and defense	417	0.0000
T-cell-mediated immunity	82	0.0000
Chemosensory perception	45	0.0000
Biological process unclassified	3,069	0.0000
Olfaction	28	0.0004
Gametogenesis	51	0.0005
Natural killer-cell-mediated immunity	30	0.0018
Spermatogenesis and motility	20	0.0037
Inhibition of apoptosis	40	0.0047
Interferon-mediated immunity	23	0.0080
Sensory perception	133	0.0160
B-cell- and antibody-mediated immunity	57	0.0298

Note that the categories overlap; e.g., "T-cell-mediated immunity" is entirely nested within "Immunity and defense."

DOI: 10.1371/journal.pbio.0030170.t001

Table 2. Test for an Excess of Putatively Positively Selected Genes by Tissue Type

Tissue of Maximal Expression	Number of Genes	*p*-Value
Testis	247	0.0002
Thyroid	66	0.0287
Thymus	82	0.0599
Prostate	76	0.0902
Fetal_liver	114	0.1668
Salivary_gland	195	0.1696
Whole_blood	405	0.239
Heart	120	0.2906
Lung	64	0.3381
Trachea	47	0.3976
Liver	244	0.4468
Uterus	51	0.493
Adrenal_gland	70	0.5434
Spleen	134	0.5582
Pancreas	358	0.6063
Pituitary_gland	60	0.6493
Placenta	179	0.7566
Cortex	36	0.7696
Kidney	179	0.801
Amygdala	43	0.8398
Corpus_callosum	101	0.8909
Caudate_nucleus	36	0.8945
Thalamus	33	0.9018
Fetal_brain	201	0.912
Ovary	133	0.9255
Whole_brain	83	0.965
Cerebellum	93	0.9903
Spinal_cord	14	1

Main Questions

How different is the evolution of human brain-specific genes from:

- ***others human T-SG's?***
- ***and in between lineages?***

Are there any effects of using alternative statistical methods

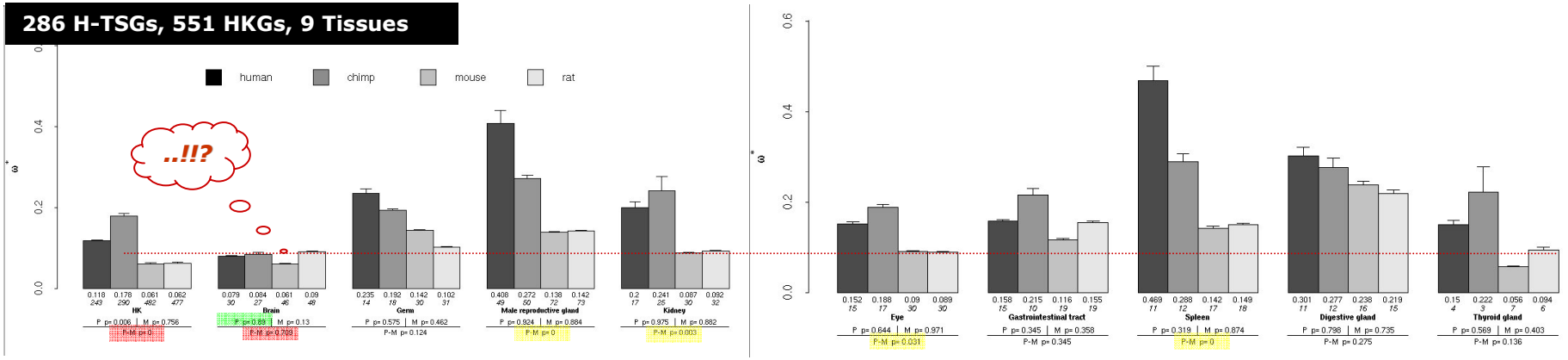
- ***Rate estimation?***
- ***Mean estimation?***

Table 1. Evolution of tissue-specific genes along primate and murid lineages using pairwise computing methods

Tissue-specific Genes		SM-Li			SM-ML			SEWM-ML			
		ω	N	p	ω	N	p	ω*	SE	N	p
Housekeeping	P	0.273	547	0.000*	0.251	547	0.000*	0.121	3.6e-03	419	0.000*
	M	0.138	550		0.114	550		0.060	1.7e-03	513	
Brain	P	0.454	51	0.006*	0.386	51	0.015*	0.096	2.2e-03	44	0.611
	M	0.223	51		0.190	51		0.084	8.0e-04	50	
Germ	P	0.501	32	0.053	0.496	31	0.061	0.261	5.1e-03	28	0.075
	M	0.332	33		0.286	33		0.141	1.5e-03	32	
Male reproductive gland	P	0.699	75	0.000*	0.686	73	0.000*	0.273	3.6e-03	71	0.000*
	M	0.295	74		0.245	75		0.140	7.0e-04	73	
Kidney	P	0.485	32	0.004*	0.392	32	0.025*	0.199	6.0e-03	29	0.028*
	M	0.218	33		0.186	33		0.109	1.5e-03	32	
Eye	P	0.400	32	0.058	0.318	31	0.197	0.164	2.9e-03	27	0.101
	M	0.252	32		0.219	32		0.104	1.6e-03	31	
Gastrointestinal tract	P	0.467	19	0.157	0.379	19	0.171	0.226	5.2e-03	17	0.099
	M	0.291	19		0.234	19		0.148	3.2e-03	19	
Spleen	P	0.761	19	0.001*	0.663	19	0.002*	0.379	9.9e-03	17	0.001*
	M	0.245	19		0.216	19		0.175	4.0e-03	18	
Digestive gland	P	0.369	17	0.957	0.350	17	0.709	0.274	8.6e-03	15	0.709
	M	0.374	17		0.315	17		0.254	6.6e-03	16	
Thyroid gland	P	0.594	7	0.074	0.484	7	0.097	0.283	1.3e-02	6	0.043*
	M	0.194	7		0.183	7		0.083	3.4e-03	7	

Notes. – P, primates. M, murids. SM-Li, simple mean using Li (1993) method. SM-ML, simple mean computed by means of a pairwise maximum likelihood method. SEWM-ML, standard error weighted mean computed by ML. SE, standard error. N, number of genes. Statistical methods are described in *Material and Methods*. Tissue definitions as specified in Table S1, Supplementary Material online.

286 H-TSGs, 551 HKGs, 9 Tissues



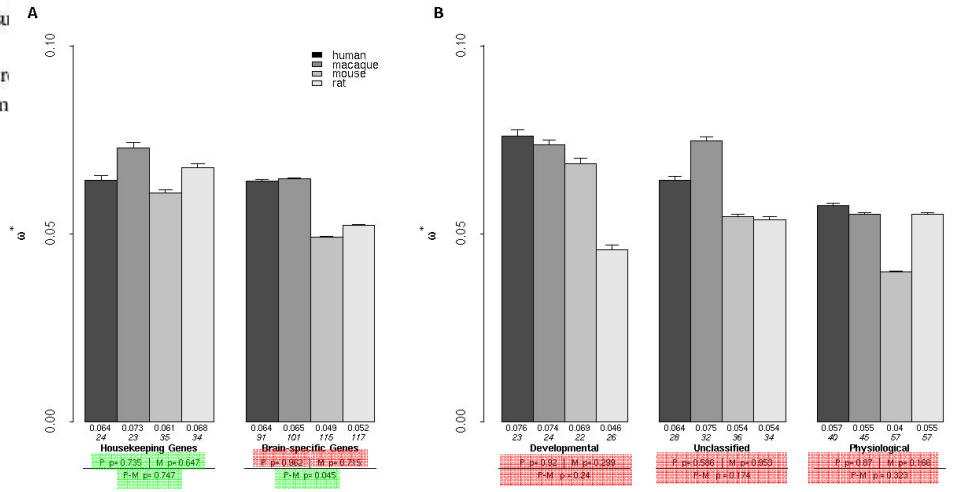
ω–SEWM estimation using a free-branch ML model

Re-Analysis of Dorus, et al. (2004) data

Table 2. Evolution of nervous system specific and housekeeping genes

Tissue-specific Genes		SM-Li			SM-ML			SEWM-ML			
		ω[b]	n	p	ω	n	p	ω[a]	SE	n	p
Housekeeping	P	0.059	55	0.137	0.078	55	0.593	0.052	7.5e-04	33	0.810
	M	0.071	55		0.066	55		0.055	7.6e-04	39	
Nervous system	P	0.123	156	0.002*	0.095	154	0.141	0.059	3.0e-04	122	0.108
	M	0.091	156		0.077	156		0.047	2.9 e-04	135	
Developmental	P	0.140	37	0.012*	0.108	37	0.094	0.065	1.2e-03	28	0.558
	M	0.084	37		0.065	37		0.053	1.2e-03	28	
Unclassified NS	P	0.099	47	0.080	0.086	47	0.527	0.070	9.7e-04	37	0.059
	M	0.074	47		0.072	47		0.044	5.8e-04	43	
Physiological	P	0.127	72	0.199	0.094	70	0.684	0.052	4.1e-04	57	0.588
	M	0.104	72		0.087	72		0.046	3.1e-04	64	

Notes. –The dataset consists of nervous system and housekeeping genes which are a su
and 58%, respectively) of those used by Dorus et al. (2004).
[b] ω was calculated as the ratio of average dN and average dS using values of gene classes r
Dorus et al. (2004). All other estimates of ω were pairwise-computed *de novo* from
alignments. See details in *Results* sections. Abbreviations defined in Table 1.



ω–SEWM estimation using a free-branch ML model

Some of the H-TSG's under PS (Test II)

Table 3
Human tissue-specific genes under positive selection

Tissue	N	CH	H	Gene Name	Chr	p[a]	Full Gene Name[b]	Function[b]
Brain	51	5	1	GRK1	13	*	G protein-coupled receptor kinase 1	Phosphorylates rhodopsin thereby initiating its deactivation. Defects in GRK1 are the cause of night blindness oguchi type-2.
Germ	33	1	2	TULP2	19	**	Tubby like protein 2	Unknown
				CST9L	20	*	Cystatin 9-like (mouse)	Some of the cystatin superfamily members are active cysteine protease inhibitors, while others have lost or perhaps never acquired this inhibitory activity.
Male Reprod. Gland	75	9	6	ADAM20	14	**	ADAM metallopeptidase domain 20	May be involved in sperm maturation and/or fertilization
				TMCO5	15	*	Transmembrane and coiled-coil domains 5	Muscle contraction
				BOLL	2	**	Bol-like (Drosophila)	This gene encodes an RNA-binding protein which may be required during spermatogenesis. Loss of this gene function results in the absence of sperm in semen (azoospermia).
				HIST1H1T	6	**	Histone 1, H1t	Histones h1 are necessary for the condensation of nucleosome chains into higher order structures. This histone is a testis-specific h1 variant that appears during meiosis in spermatogenesis.
				TCP11	6	**	T-complex 11 (mouse)	Unknown
				PEPP-2	X	*	Homeobox protein, PEPP Subfamily, 2	Regulates downstream genes and biological events in the testis
Kidney	33	2	1	SLC22A8	11	**	Solute carrier family 22 (organic anion	The protein is involved in the sodium-independent transport and excretion of organic anions, some of

Conclusions II

- **Estimates of average dN/dS ratios are sensitive to the methods used for combining estimates and the methods used for estimation of dN/dS.**
- **Brain specific genes show no evidence for acceleration in the primate lineage compared to many other tissue specific gene categories.**
- **There is no evidence for an elevation in the dN/dS ratio in brain-specific genes in humans compared to chimpanzees, neither between primates and murids taxa.**
- **The number of brain-specific genes showing evidence for positive selection is higher in the chimpanzee than in the human evolutionary lineage.**
- **While there undoubtedly has been much positive selection relating to brain function during the evolution of modern man, this selection has not been so pervasive that is has resulted in a detectably accelerated rate of molecular evolution.**

Opinion TRENDS in Genetics Vol.22 No.11

Did brain-specific genes evolve faster in humans than in chimpanzees?

Peng Shi*, Margaret A. Bakewell* and Jianzhi Zhang

Department of Ecology and Evolutionary Biology, University of Michigan, 1075 Natural Science Building, 830 North University Avenue, Ann Arbor, MI 48109, USA

However...

TIG, november 2006

Table 1. Evolutionary rates of brain-specific genes and other genes in humans and chimpanzees

Tissue-specificity definitions	Genes	No. of genes	Human lineage			Chimpanzee lineage				
			n_H[a]	s_H[b]	w_H[c]	n_C[a]	s_C[b]	w_C[c]	w_H/w_C[d]	n_H/n_C[e]
Microarray (2×)	Brain-specific genes	249	286	571	0.205	318	655	0.198	1.03	0.90
	Other tissue-specific genes	1544	2897	3621	0.327	2833	3765	0.307	1.06	1.02
	Non-tissue-specific genes	12 162	21 710	28 241	0.314	20 394	29 620	0.281	1.12****	1.06
	All genes	13 955	24 893	32 432	0.313	23 546	34 040	0.282	1.11****	1.06
	Ratio of brain to other tissue-specific				0.626			0.646	0.97	0.88
	Ratio of brain to non-tissue-specific				0.652			0.706	0.92	0.84*
Microarray (4×)	Brain-specific genes	72	91	182	0.205	104	192	0.221	0.93	0.88
	Other tissue-specific genes	502	973	1178	0.337	962	1176	0.334	1.01	1.01
	Non-tissue-specific genes	13 381	23 829	31 071	0.313	22 480	32 671	0.281	1.11****	1.06
	Ratio of brain to other tissue-specific				0.607			0.662	0.92	0.87
	Ratio of brain to non-tissue-specific				0.653			0.786	0.83	0.83
EST	Brain-specific genes	165	294	430	0.279	324	493	0.268	1.04	0.91
	Other tissue-specific genes	819	1963	2083	0.385	1891	2313	0.334	1.15**	1.04
	Non-tissue-specific genes	12 971	22 637	29 920	0.309	21 331	31 234	0.279	1.11****	1.06
	Ratio of brain to other tissue-specific				0.725			0.804	0.90	0.87
	Ratio of brain to non-tissue-specific				0.903			0.962	0.94	0.85
SAGE	Brain-specific genes	209	356	494	0.295	368	550	0.273	1.08	0.97
	Other tissue-specific genes	632	1214	1485	0.334	1238	1580	0.320	1.04	0.98
	Non-tissue-specific genes	13 114	23 323	30 454	0.313	21 939	31 911	0.281	1.11****	1.06
	Ratio of brain to other tissue-specific				0.883			0.854	1.03	0.99
	Ratio of brain to non-tissue-specific				0.942			0.974	0.97	0.91
Nervous system genes[f]	Nervous system genes	146	196	341	0.235	193	407	0.193	1.22	1.02
	Developmental	37	53	91	0.237	47	113	0.169	1.40	1.13
	Physiological	61	59	135	0.178	66	170	0.159	1.12	0.89
	Unclassified	48	85	115	0.300	80	124	0.263	1.14	1.06
	Other genes	13 809	24 697	32 091	0.314	23 353	33 633	0.283	1.11**	1.06
	Ratio of nervous system to other genes				0.748			0.682	1.10	0.96
	Ratio of developmental to other genes				0.754			0.596	1.26	1.07
Overlapping sets[g]	Brain-specific genes	74	86	176	0.199	117	216	0.221	0.90	0.73
	Other genes	13 881	24 808	32 256	0.314	23 429	33 824	0.283	1.11****	1.06
	Ratio of brain to other genes				0.632			0.781	0.81	0.69**

Human Disease & Evolution

doi:10.1016/j.jmb.2006.02.067 J. Mol. Biol. (2006) 358, 1390–1404

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Selective Pressures at a Codon-level Predict Deleterious Mutations in Human Disease Genes

Leonardo Arbiza[1], Serena Duchi[1], David Montaner[2], Jordi Burguet[2] David Pantoja-Uceda[3], Antonio Pineda-Lucena[3], Joaquín Dopazo[2] and Hernán Dopazo[1]*

[1]*Pharmacogenomics and Comparative Genomics Unit Centro de Investigación Príncipe Felipe (CIPF). Autopista del Saler, 16-3, 46013 Valencia Spain*

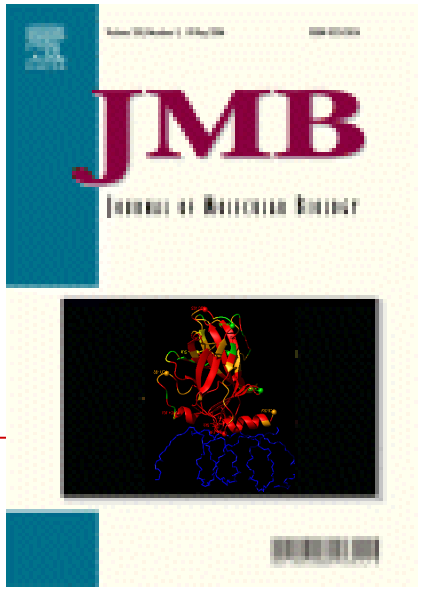
[2]*Functional Genomics Unit Bioinformatics Department Centro de Investigación Príncipe Felipe (CIPF). Autopista del Saler, 16-3, 46013 Valencia Spain*

[3]*Structural Biology Laboratory Medicinal Chemistry Department, Centro de Investigación Príncipe Felipe (CIPF). Autopista del Saler 16-3, 46013 Valencia, Spain*

*Corresponding author

Deleterious mutations affecting biological function of proteins are constantly being rejected by purifying selection from the gene pool. The non-synonymous/synonymous substitution rate ratio (ω) is a measure of selective pressure on amino acid replacement mutations for protein-coding genes. Different methods have been developed in order to predict non-synonymous changes affecting gene function. However, none has considered the estimation of selective constraints acting on protein residues. Here, we have used codon-based maximum likelihood models in order to estimate the selective pressures on the individual amino acid residues of a well-known model protein: p53. We demonstrate that the number of residues under strong purifying selection in p53 is much higher than those that are strictly conserved during the evolution of the species. In agreement with theoretical expectations, residues that have been noted to be of structural relevance, or in direct association with DNA, were among those showing the highest signals of purifying selection. Conversely, those changing according to a neutral, or nearly neutral mode of evolution, were observed to be irrelevant for protein function. Finally, using more than 40 human disease genes, we demonstrate that residues evolving under strong selective pressures ($\omega < 0.1$) are significantly associated ($p < 0.01$) with human disease. We hypothesize that non-synonymous change on amino acids showing $\omega < 0.1$ will most likely affect protein function. The application of this evolutionary prediction at a genomic scale will provide an *a priori* hypothesis of the phenotypic effect of non-synonymous coding single nucleotide polymorphisms (SNPs) in the human genome.

Keywords: comparative genomics; deleterious mutations; human diseases; purifying selection; codon-based models



Medicine believes in genetics, less in bioinfo and almost nothing in evolutionary biology!!

- **Medicine seldom takes into account evolutionary biology's conclusions**
  - Parasites should evolve towards a bening coexistence with their host…
  - *It strongly depends on mode of inheritance* (Paul Ewald, 1996)
- **Scientists working in biomedicine rarely recognize basic evolutionary biology concepts**
  - *PAML matrix*
  - *Positional homology*

Evolutionary Thinking in Biomedicine
Notable exceptions

- **1996. Paul M. Ewald. Evolution of infectious disease**
  - ”Parasites vertically inherited should evolve toward a bening coexistence with their host”.
- **1996. R. Nesse and G. Williams. Why we get sick? The new science of Darwinian Medicine**
  - “Why in a body of such exquisite design, are there a thousands flaws and fraities that make us vulnerable to disease?... They suggest new ways of addressing illness”
- **1998. Stephen C. Stearns. Evolution in health and disease**
  - “... our body was shaped by natural selection to maximize reproductive success in ancestral environments”.
- **2002. Steve A. Frank. Immunology and Evolution of Infectious disease**
  - Bring the gap between immunology and epidemiology.

Example: SNP's and Disease

SNPs can cause alterations of gene function by…

- **Alterations in expression level**
- **Alternative splicing**
- **Alteration (or loss) of gene product function**
  - **Changes in the stability of the protein**
  - **Functionally important residues**
  - **Phylogenetic conservation**

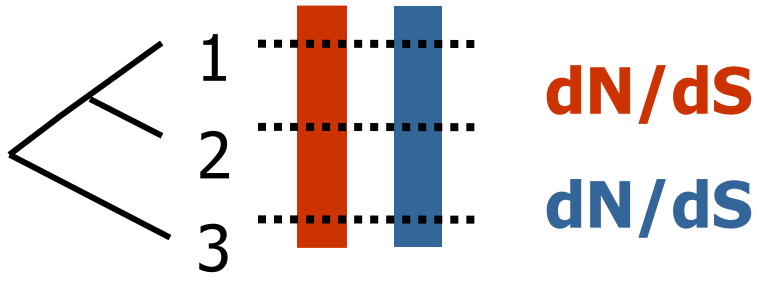
Natural selection working at codon level

Main Question

Could an estimator of the selective preassures acting at codon level (ω) be used as a predictor of the phenotype effect of SNP´s?

Detecting Positive & Negative Selection

- **Site-specific models** average dN/dS over lineages but differentiate over sites



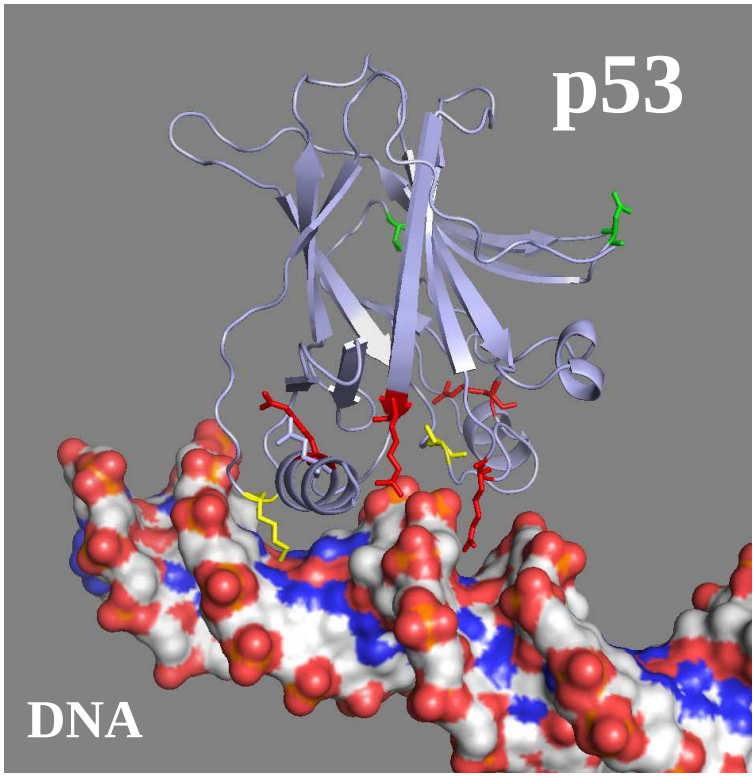
- **ML models for positive selection**
  - **M1a vs M2a; M7 vs M8**
- **Bayes Empirical Bayes (BEB)**
  - **M2a, M8**
  - (which ω class (h) each site is most likely to belong).
- **Sitewise likelihood-ratio method (SLR)**

Bayes Empirical Bayes (BEB)

Bayes Empirical Bayes (BEB) analysis
Positively selected sites (*: P>95%; **: P>99%)

		Pr(w>1)	post mean +- SE for w	
1	M	0.007	0.156 +- 0.298	
2	E	0.009	0.169 +- 0.353	
3	E	0.009	0.169 +- 0.353	
4	P	0.125	0.893 +- 1.263	Neutral
5	Q	0.010	0.182 +- 0.370	
6	S	0.015	0.212 +- 0.436	
7	D	0.010	0.180 +- 0.375	
8	P	0.368	2.207 +- 2.473	Positive
9	S	0.007	0.160 +- 0.310	
10	V	0.139	0.969 +- 1.480	
11	E	0.009	0.169 +- 0.353	
12	P	0.091	0.722 +- 1.043	
13	P	0.014	0.208 +- 0.450	Purifying
14	L	0.013	0.200 +- 0.411	
15	S	0.009	0.178 +- 0.371	
16	Q	0.010	0.182 +- 0.370	
17	E	0.011	0.186 +- 0.405	

Evolutionary models in action



A real case with the master protein of the cell

TP53 protein

- Probably **p53** is the main protein regulating cell division and apoptosis
- Many **mutant forms** are involved in different types of human cancer

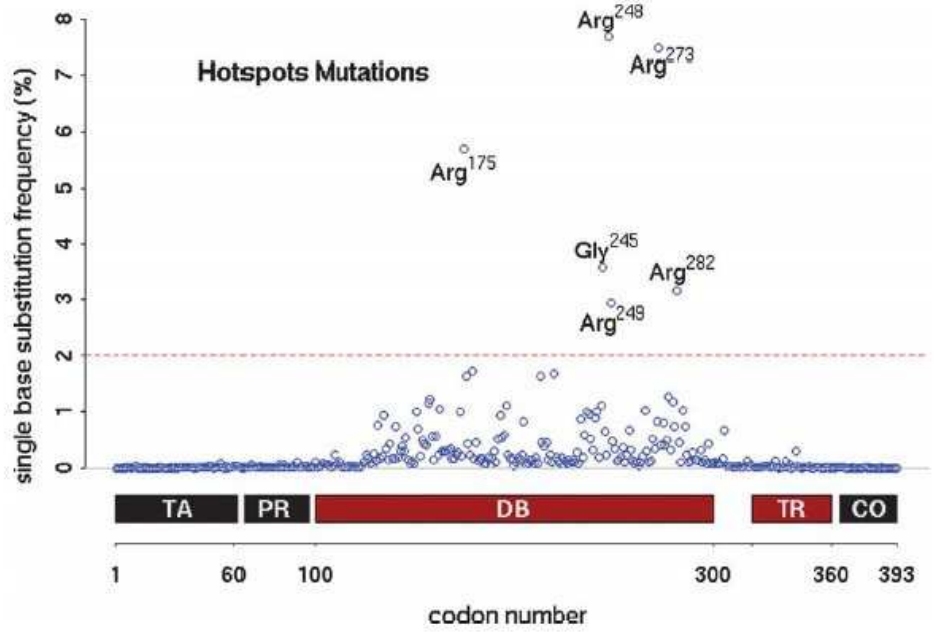
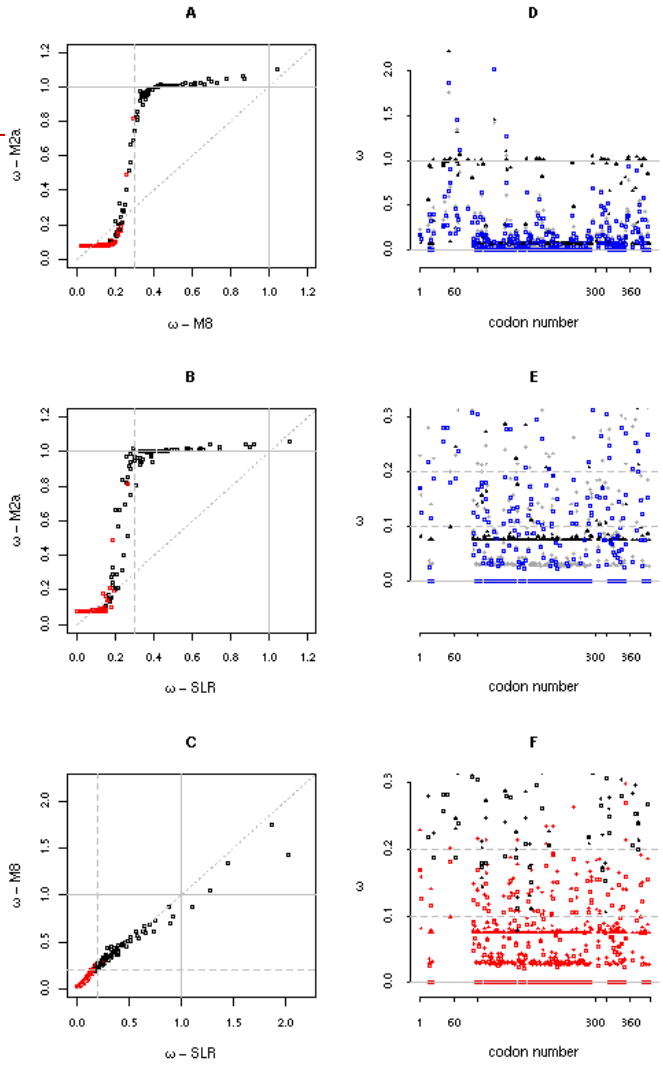


Figure 1. p53 mutations distribution. Mutation frequencies collected in the IARC TP53 R10 database (18,145 non-synonymous mutations) are plotted against the protein domains. The DNA-binding (p53DB) domain contains six residues considered mutational hotspots in cancer.

Evolutionary Models in TP53

- M1a vs M2a -> no PS
- M7 vs M8 -> no PS
- $logL_{(M2a)} > logL_{(M8)}$ -> **best
- SLR -> Other alternative method
- (multiple testing 95%, 99%)
- ---, ----
- +++, ++++



TP53 Evolutionary Analysis

DB and TR domains have the lowest ω value distribution

Table 1. Summary of p53 domains, mutations and ω statistics according to M8 and SLR models

p53 alignment			Mutations			ω statistics			
Domains	Codons	Indels[a]	Total	Mps[b]	Model	Min.	Median	Mean	Max.
TA	1-60	38	96	1.6	M8	0.030	0.334	0.379	1.747
					SLR	0.000	0.269	0.369	1.865
PR	61-97	22	151	4.2	M8	0.029	0.314	0.376	1.338
					SLR	0.000	0.307	0.376	1.447
DB	100-300	5	17,389	87.0	M8	0.027	0.039	0.116	1.423
					SLR	0.000	0.029	0.095	2.018
TR	325-355	0	178	5.1	M8	0.028	0.067	0.126	0.456
					SLR	0.000	0.068	0.103	0.379
CO	361-393	11	18	1.6	M8	0.027	0.216	0.255	0.878
					SLR	0.000	0.176	0.226	0.882

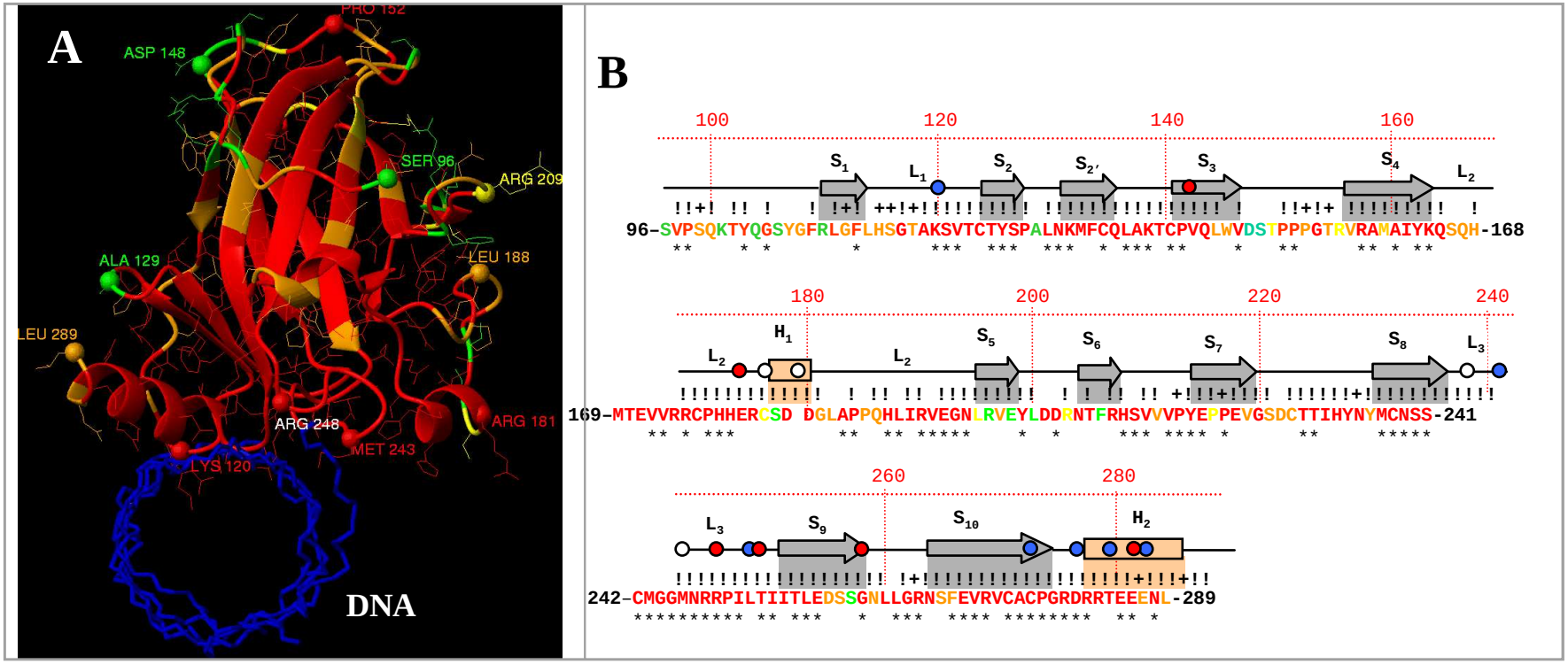
Mutations were deduced from the IARC TP53 database.

[a] Insertions/deletions.

[b] Mean number of mutations per site.

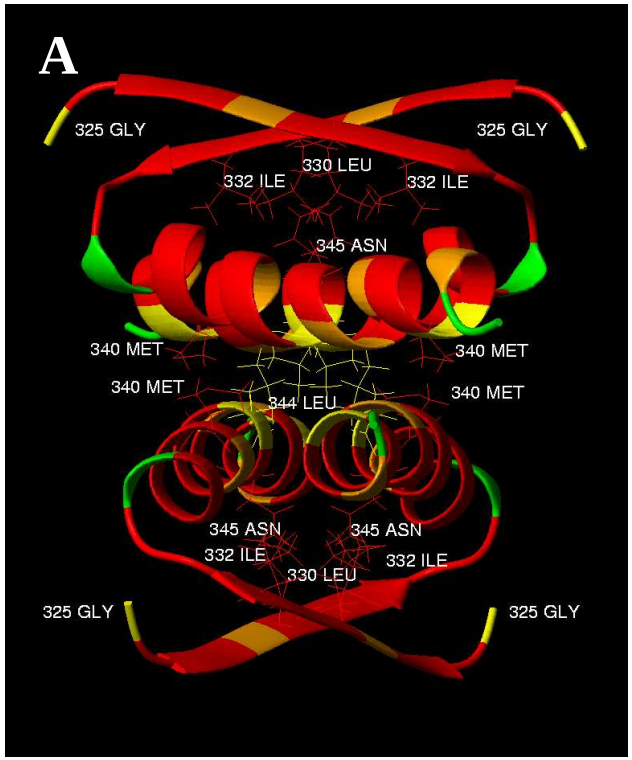
DNA-binding domain evolutionary and biological analysis

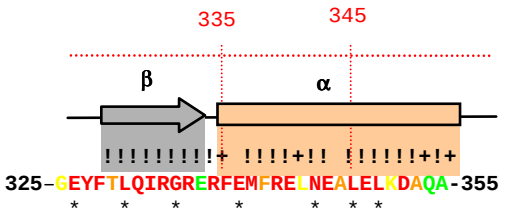
SLR: $\omega \leq 0.1$, $0.1 < \omega \leq 0.2$, $0.2 < \omega \leq 0.3$, $\omega > 0.3$



TR-domain evolutionary and biological analysis

SLR: ω ≤ 0.1, 0.1 <ω ≤ 0.2, 0.2 <ω ≤ 0.3, ω >0.3





How effective is natural selection ?

§ Evolutionary biologists recognized that natural selection works in proportion to the number of deleterious mutations in the population

Table 2. Summary of the number of mutations, and mutations per residue (bold), in p53 and 2 domains evaluated under different categories of selective constraints

p53	Model	$\omega > 0.3$	$0.2 \leq \omega < 0.3$	$0.1 \leq \omega < 0.2$	$\omega < 0.1$	[a]SLR*	[b]ω_{M8}	[c]ω_{SLR}	[d]PC
Full protein	M8	570	382	1714	15,165	16,883	13,028	12,992	13,152
		9.5	**13.2**	**35.0**	**87.7**				
	SLR	430	250	1495	15,656	**87.0**	**119.5**	**120.3**	**120.6**
		8.6	**11.4**	**25.3**	**87.0**				
DB	M8	437	337	1669	14,814	16,471	12,998	12,952	13,112
		23	**25.9**	**50.6**	**113.1**				
	SLR	306	223	1436	15,292	**99.2**	**139.6**	**140.8**	**141.0**
		20.4	**31.9**	**36.8**	**113.3**				
TD	M8	8	7	12	152	164	30	30	30
		2.0	**2.3**	**2.4**	**7.6**				
	SLR	6	8	6	158	**6.3**	**5**	**4.3**	**4.3**
		2.0	**2.5**	**1.5**	**7.5**				

Note that PC residues show higher values of mutations per residue (120.6 and 141.0 for p53 and p53DB) although SLR* contains a higher number of sites with statistical evidence of strong purifying selection (228 and 166). The increasing number of mutations per residue observed in ranges of ω with higher selective constraints demonstrates that natural selection works in proportion to the number of mutations in the population (see the text).

[a] Residues under the constraints of purifying selection evaluated by the SLR method at 95% and 99% statistical confidence.

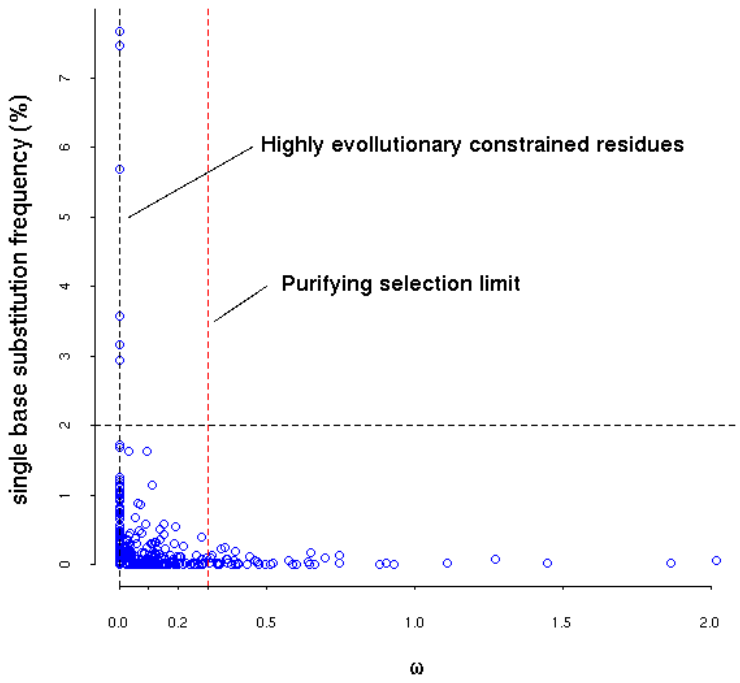
[b] Residues with $\omega_{M8} \leq 0.033$.

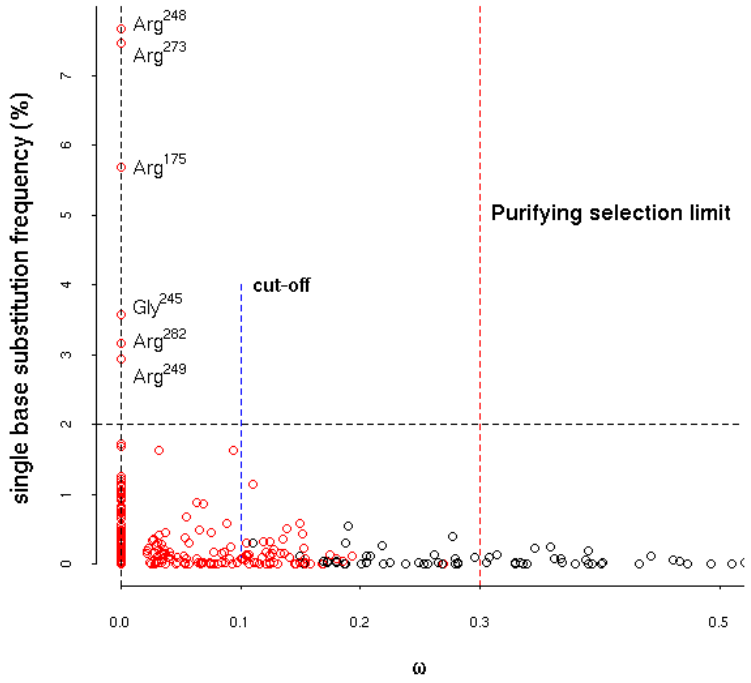
[c] Residues with $\omega_{SLR} = 0$.

[d] Residues phylogenetically conserved throughout the p53 alignment.

TP53 mutation freq. and selective constraints

According to the theory this will follow an “L” shape curve





The Main Question again

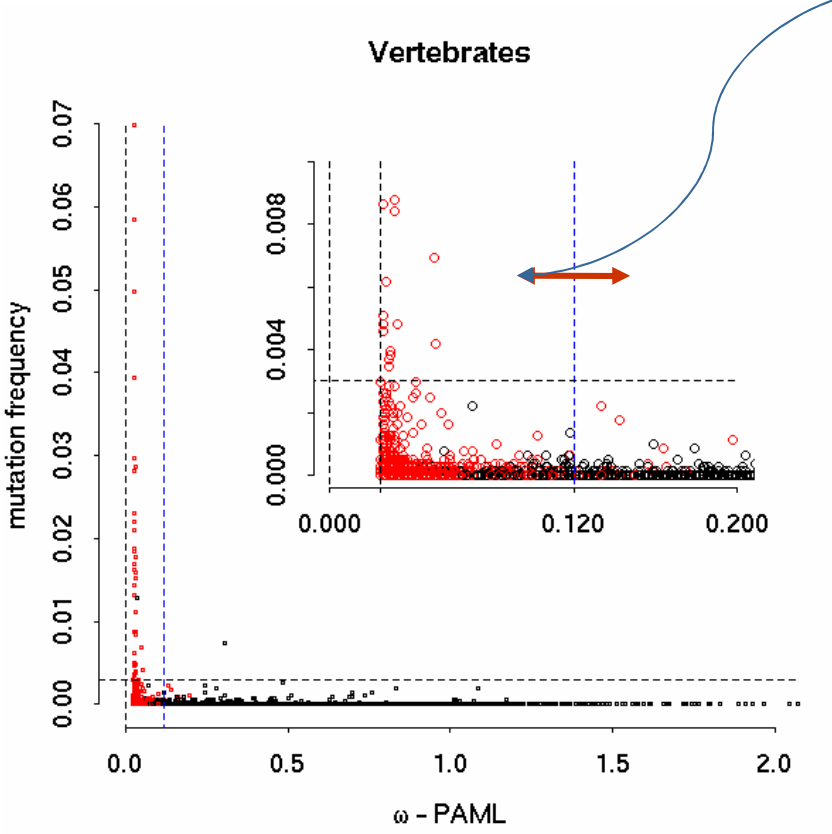
p53 results seem to show good signals, however,

Is it possible to obtain a specific predictor of the more frequent amino acid changes associated to human diseases?

Bioinformatics and evolutionary analysis

- **Analyse DB containing codon mutation frequencies for all the possible human diseases proteins**
  - **Immune deficiency and cancer (COSMIC) databases (approx. 250 genes)**
- **Ensembl-orthologous genes in different species**
  - **Mammals and Vertebrates**
- **Evolutionary ML analysis**
  - **(M1a, M2a, M7, M8, SLR)**
- **Statistical tests (KS)**
  - **reject genes with <10 mutations**

ω and frequency distribution of Immune and Cancer mutations



$\omega_{threshold} = 0.1$

Two sample Kolmogorov-Smirnov test

H_0: freq. $_{(lower\ \omega)}$ = freq. $_{(upper\ \omega)}$

H_A: freq. $_{(lower\ \omega)}$ > freq. $_{(uppper\ \omega)}$

Table 3. Evaluation of alternative $\omega_{cut\text{-}off}$ values and mutational frequencies in disease

	Mammals		Vertebrates	
$\omega_{cut\text{-}off}$	PAML	SLR	PAML	SLR
0.03	0.9748	0.0095	0.0504	0.0061
0.05	0.0114	0.0075	0.0026	0.0008
0.10	3.0×10^{-05}	0.0076	0.0016	0.0009
0.12	0.0007	0.0077	**0.0010**	0.0023
0.15	0.0025	0.0078	0.0012	0.0018
0.20	0.0715	0.0074	0.0019	0.0019
0.25	0.1938	0.0074	0.0044	0.0043
0.30	0.0188	0.0076	0.0035	0.0065
0.40	0.0486	0.0101	0.0176	0.0254
0.50	0.1849	0.0223	0.0534	0.1010
G[a]	43	43	43	43
R[b]	24,375	24,375	17,424	17,435
M[c]	8970	8970	8081	8083

One-tail K–S tests reject the null hypothesis, which considers that the frequency of mutations are not differentially distributed above and bellow the given $\omega_{cut\text{-}off}$. The alternative hypothesis, which considers that disease-associated mutations are preferentially associated with values below the $\omega_{cut\text{-}off}$, is accepted with the highest confidence using ω_{PAML} estimations on mammal ($\omega_{cut\text{-}off}$=0.10) and vertebrate ($\omega_{cut\text{-}off}$=0.12) datasets. The K–S test on SLR estimates reject the null hypothesis for all values of $\omega_{cut\text{-}off}$ evaluated. This is the consequence of the undesirable behaviour of the SLR method, which drops low values of ω to 0 (see the text and Figure 6 for explanation).

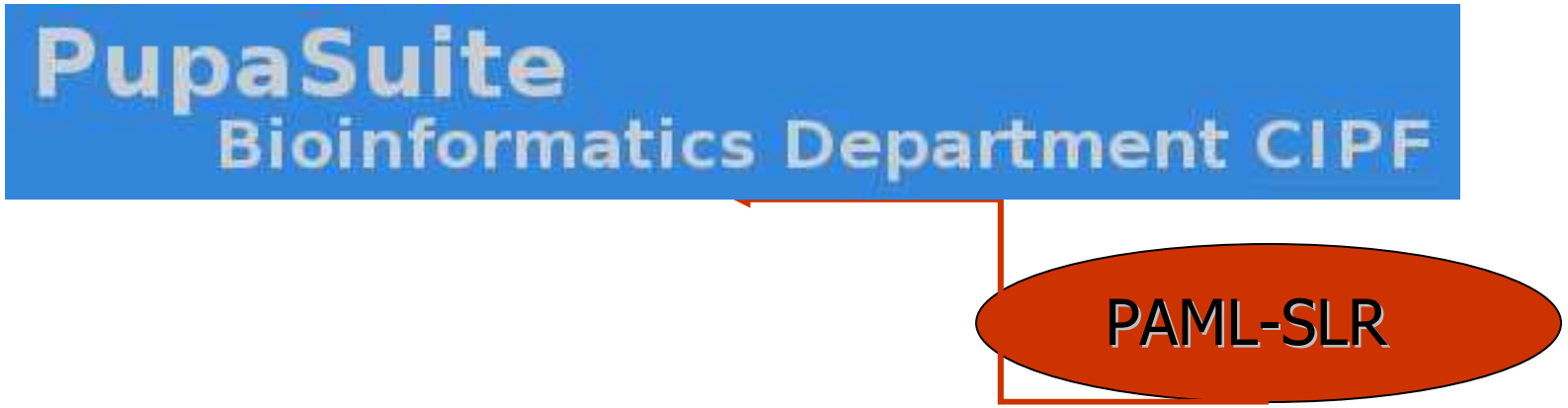
[a] Number of genes evaluated.
[b] Number of residues evaluated.
[c] Number of mutations evaluated.

Conclusions III

- **We have found an evolutionary parameter that allows to differentiate amino acids where disease is more frequent**
- **This parameter is a measure of the action of natural selection working on vertebrate species during million years**
- **We hypothesize that non-synonymous changes on amino acids showing $\omega < 0.1$ probably affects the normal function of proteins**

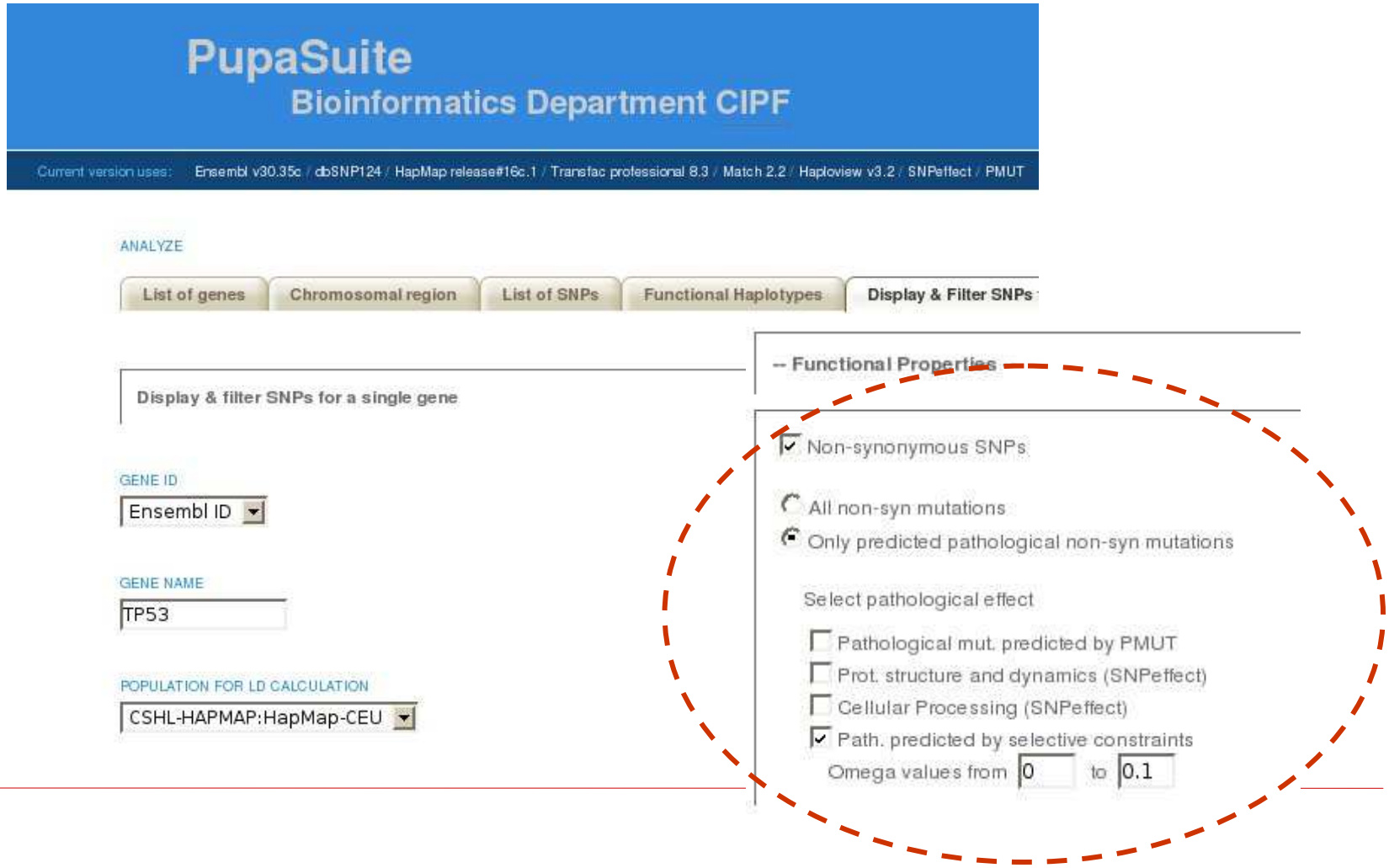
- **Recently we confirmed this results using more than 3,000 proteins**
- **Disease and polymorphisms are differentiated using ω values**

Selective constraints on all the cSNP's of the Human Genome



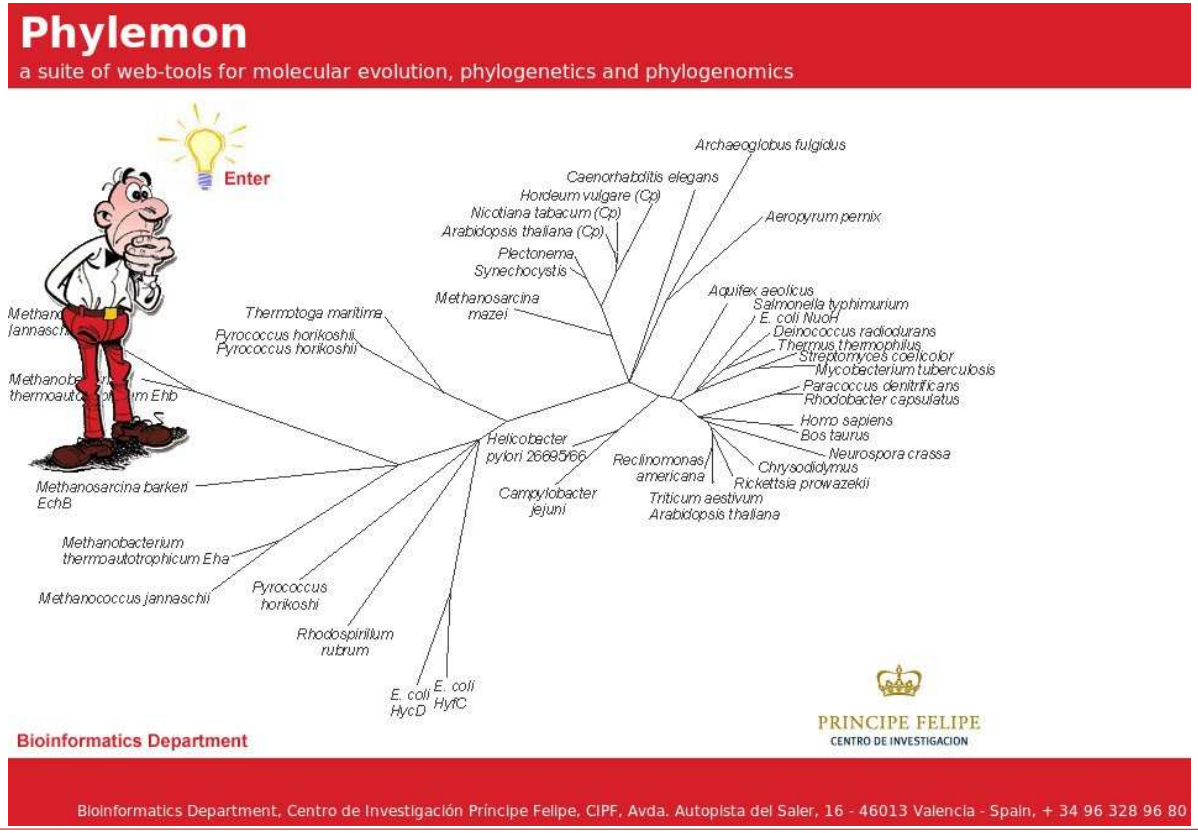
Evolutionary Models

Bioinformatic Tool: SNP's probable associated to mendelian diseases *(NAR, web issue 2006)*



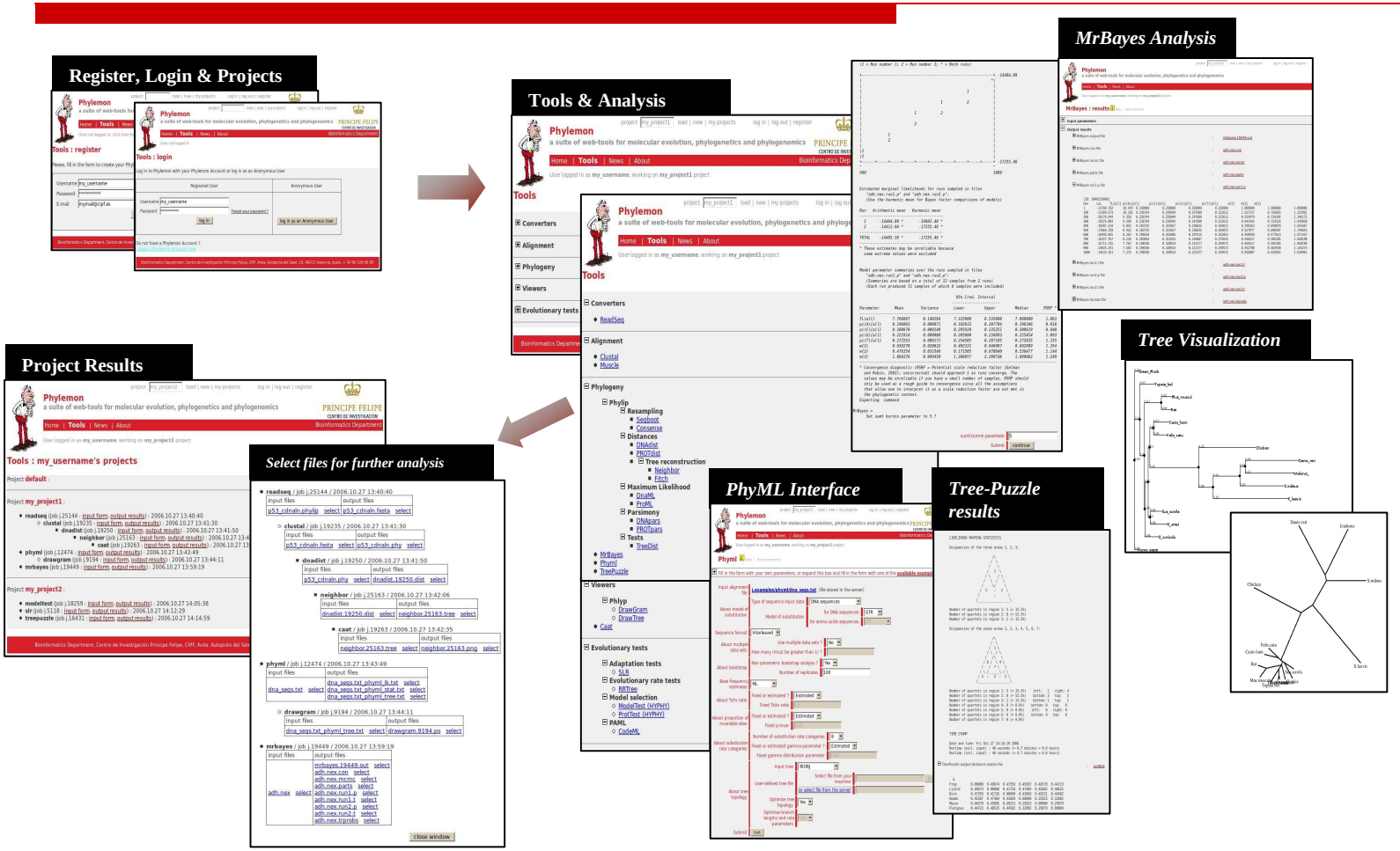
Phylogenetics/Phylogenomics

Phylemon server is at: **http://phylemon.bioinfo.cipf.es**

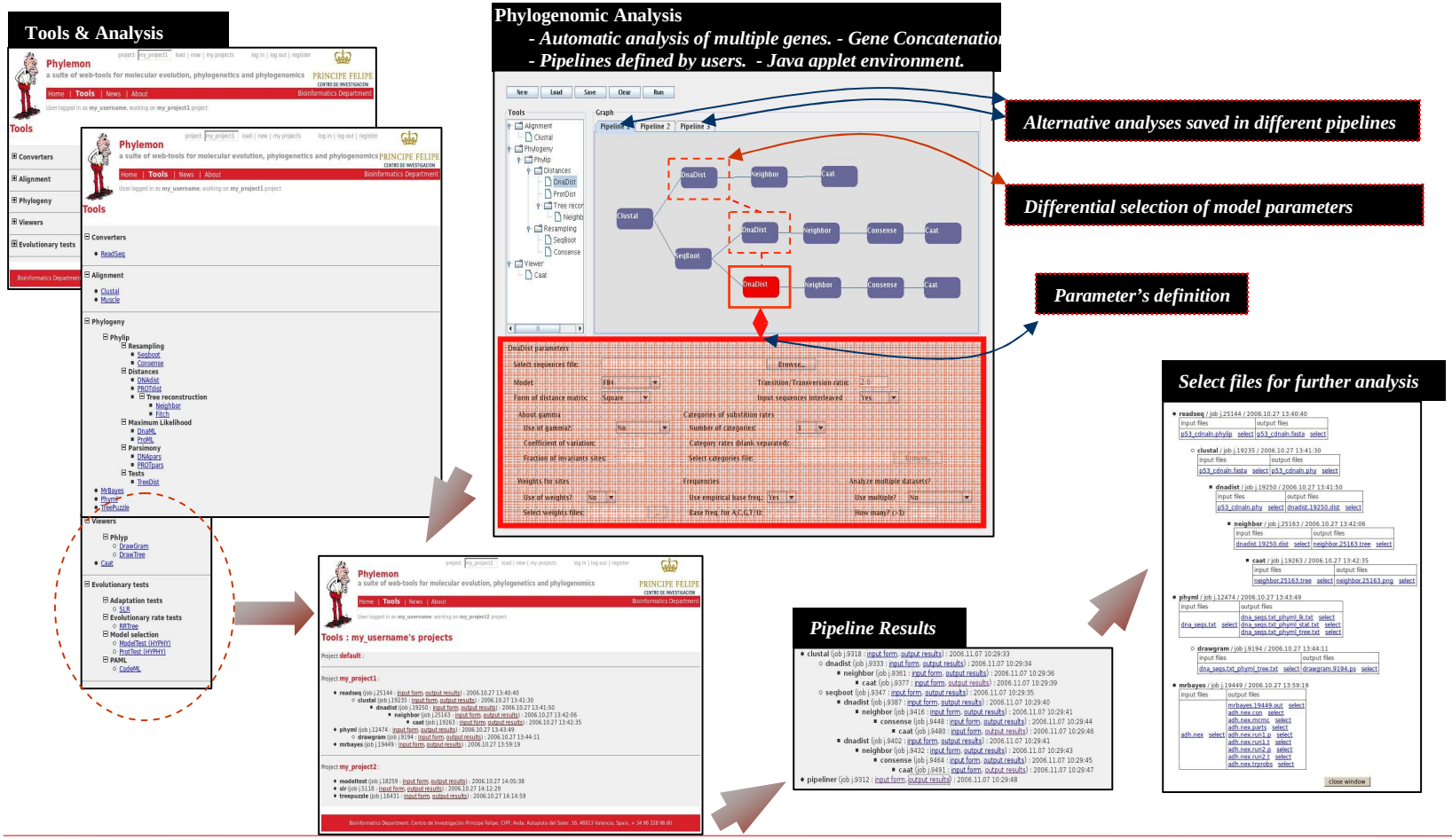


Phylemon characteristics I:

Phylogenetics Analysis



Phylemon characteristics II:
Evolutionary -tests/Phylogenomics/Pipelines



Muchas Gracias!!

Micostrium Vulgaris
Phylum: Chordata
Subphylum: Vertebrata
Class: Mammalia