Molecular evolution of key receptor genes in primates and non-human primates

Barbara Picone[1, 2, *], Alan Christoffels[2]

[1]Department of Genetics, University of Stellenbosch, Stellenbosch, 7602, South Africa

[2]South African National Bioinformatics Institute (SANBI), SA Medical Research Council Bioinformatics Unit, University Western Cape, Bellville, 7535, South Africa

Email address:

barbara.picone@gmail.com (B. Picone)

Abstract: African primates remain an unexplored source of information required to complete the origin and evolution of many human pathogens. Current studies have shown the importance of several receptor human genes implicated in host resistance or susceptibility to tuberculosis. The validation of these genes in *Mycobacterium tuberculosis* infection makes them an excellent model system to investigate the mode of selective pressures that may act on pathogen defense genes. To trace the evolutionary history of these genes, the report describes preliminary results for eight receptors human genes having either a significant or a possible association with Tuberculosis (TB). By using a combination of maximum likelihood approaches, evidence of positive selection were detected for four genes. The analysis between species, nevertheless, shows a clear pattern of nucleotide variation mostly compatible with purifying selection.

Keywords: Receptor Genes, Tuberculosis, Phylogeny, Non-Human Primates, Adaptive Evolution

1. Introduction

Although the *Mycobacterium tuberculosis* complex infects a third of all humans, little is known regarding the prevalence of mycobacterial infection in non-human primates (NHP). For more than a century, tuberculosis has been regarded as a serious infection threat to NHP species. Just as humans, many mycobacteria are etiologic agents of disease in non-human primates with a diverse degree of susceptibility to the disease in different primate genera [1-3].

Comparisons with genomes of closely related species are necessary in order to achieve a more complete understanding of the evolution of the human genome. Thus far, differential gene expression, gene loss, gene duplication, and natural selection have received the most attention as the molecular mechanisms responsible for the differences in protein activity in these species [4, 5]. Current studies have showed the importance of numerous human genes implicated in host resistance or susceptibility to tuberculosis (i.e. receptors; transporters, cytokines, chemokines, etc.) [6]. The validation of these genes in *Mycobacterium tuberculosis* infection makes them an excellent model system to investigate the mode of selective pressures that may act on pathogen defense genes.

Frequent positive selection is a hallmark of genes involved in the adaptive immune system of vertebrate, but the prevalence of positive selection for genes underlying immunity in vertebrates has not been well studied. The identification of genes subjected to positive selection, can lead to predictions of putative functionally important regions of genes [7]. Receptor genes of the innate immune system represent the first line of defense against pathogens. Of particular interest is the identification of human or primate-specific events of adaptive evolution. Humans diverged from African apes (chimpanzee, gorillas and orangutans) about 5-7 million years ago. Genetically, humans and chimpanzee share nearly 99% DNA sequence similarity [9-10], which seems to deny the marked biological and phenotypic divergence between them, i.e., the highly developed cognitive abilities in humans. The major challenge in the human/chimpanzee genome sequence era is to determine the small subset of sequence differences that have phenotypic significance related to species-specific traits [11].

Interspecific comparisons provide information about evolutionary processes acting over different timescales. The history of pathogenic diseases during primate evolution undoubtedly played a role in shaping the present immune

system, and the forces acting on immune genes over this timescale can only be studied from inter-specific comparisons.

Current studies have showed the importance of numerous human genes implicated in host resistance or susceptibility to tuberculosis (*i.e, VDR, P2RX7, NOS2A, SP110, CD209, CCl2, CXL10, GC, IFNG, IL10, PTPN22, SP110, TIRAP, TLR2, TNFRSF1B*). The aim of the present study is to extend the analysis of these genes in non-human primates to thoroughly assess their evolutionary history and to identify possible amino acids under selective pressure. This analysis focuses on a fraction of these human genes interestingly having either a significant or a possible association with tuberculosis: (*VDR, P_2X7, TNFRSF1B, IL12RB1, CD209, MC_3R, TIRAP, IFNGR1*). In this scenario, future studies are fundamental to understand the evolution of other receptor.

2. Material and Methods

2.1. Sequences

The sequences of the primates genes used in the analyses were retrieved from the public database "Ensembl Genome Browser". For each gene, a subset of 4-11 species was used, that included species from the most representative primate groups (Old World primates, New World primates, and Prosimians).

2.2. Codon-Based Analyses of Positive Selection

To evaluate positive and negative selections at all the genes during primate evolution, we compared the rate per site of non-synonymous substitution (*dN*) to the rate per site of synonymous substitutions (*dS*) in a maximum likelihood (ML) framework (M7/M8, SLAC, FEL, REL and SLR). A ratio of dN/dS >1 is interpreted as strong evidence of positive selection, whereas a dN/dS <1 is evidence of purifying selection. For each gene, a neighbor-joining tree was used as the working topology, which was constructed using Mega v.5 with the option p-distance as the substitution model and complete deletion to gaps and missing data [12]. Two alternative models in CODEML M7 and M8 were then implemented (PAML v. 4) [13].

M7 allows only codons to evolve neutrally or under purifying selection while M8 adds a class of sites under positive selection. The two previous nested models were compared using a likelihood ratio test (LRT) with 2 degrees of freedom [14]. Amino acids under selection for M8 were indentified using Bayes Empirical Bayes approach (BEB) with posterior probability >90%.

It is more difficult to identify specific sites under selection, than to show that proportion of the sites are under selection. Therefore, only sites with a posterior probability >90% were considered as candidates. Next, a series of ML methods were applied to the data set in order to test for positive selection in individual codons of genes sequences: The Hyphy package implemented in the Data Monkey Web server [15] and the SLR method [16].

In the Data Monkey Web server, the best fitting nucleotide substitution model was searched for through the automatic model selection tool available on the server.

All sequences of each gene were analyzed under three distinct models, single likelihood ancestor counting (SLAC), fixed-effect likelihood (FEL) and random effect likelihood (REL). The SLAC model is based on the reconstruction of the ancestral sequences and the counts of d*s* and d*N* at each codon position of the phylogeny. The FEL model estimates the ratio *dN/dS* on a site-by-site basis, without assuming a priori distribution across sites. The REL model first fits a distribution of rates across sites and then infers the substitution rate for individual sites. Sites with P values <0.1 for SLAC and FEL, and Bayes factor >50 for REL were considered as candidates to be under positive selection.

The SLR method is complementary to the random sites model implemented in PAML. It performs an explicit likelihood-ratio test for selection at each site in the alignment, making few assumptions about the distribution of selection and allowing every site to be under a different level of evolutionary constraint. Further, a series of python scripts were written, in order to identify nucleotide changes positively selected for each gene (data not shown).

3. Results and Discussion

Genes of the immune systems and genes involved in the host-pathogen interaction have been shown to be highly prone to adaptive selection [17,18]. Despite several studies on the evolution of these genes in human and non-human primates, a clear picture of the evolution of these gene families has not emerged, probably because previous studies have considered samples within species or between species, but not both. Genes associated with the immune system are under constant evolutionary pressure to change as a result of host-parasite co-evolution, where advantageous mutations are heavily favored.

Using ML approaches, evidence of positive or negative selection was detected in all genes studied (Table 1). Nested models with or without positive selection were compared using LRTs. As result for four of eight genes (*IFNGR1, TIRAP, IL12RB1* and *CD209*), a model that includes sites with *dN/dS* >1 fits the data significantly better than a neutral model (Table 1). For each of these nine genes, the proportion of sites under selection according the M8 model in PAML was relatively low. The specific codons identified by the BEB approach with a posterior probability of 90% represent an even smaller fraction of that proportion.

The other ML methods also detected sites under selection, some of which coincide with the codons previously identified by M8. To identify robust candidates for sites under selection, the analysis considered sites with evidence of selection in at least one site that was concordant among methods (Table 1). *IFNGR1* presents the highest number of positive selected codons (13), whereas *VDR, P2RX7, TNFRS1, MC3R* genes showed evidence of being or having been under a regime of negative selection.

To examine the phylogenetic distribution of the inferred positively selected changes among primate clades (Prosimians, New World monkey, Old world monkey, and apes), the current study investigated the *dN/dS* ratio among lineages for each branches of gene phylogenetic tree. The one-ratio model, which assumes an equal ω-ratio for all branches in the phylogeny, was then compared with the free-ratio model, which assumes an independent ω-ratio for each branch. The difference between the two models was significant (0.01< P > 0.05), indicating that there is variable selective pressure in the phylogeny. Interestingly, the branches with dN/dS >1 was found among Old work monkeys and apes (data not shown).

When three-dimensional structures were available, it was possible to look at the functional significance and location of all positive selected sites in those genes identified by the ML methods. For two genes (*TIRAP* and *IFNGR1*) [19, 20] the analyses could detect several sites that fall in or immediately adjacent to regions or residues suggested to affect function such as sites involved in dimerization surface.

Table 1. Phylogenetic test of Positive Selection

Gene	No. Species	InL M7 (neutral)	InL M8 (selection)	-2lnΔL	Significance	ρs, ωs	PAML M8	SRL	SLAC	FEL	REL	Tot. sites
VDR	7	-2236,72	-2236,6	0.24	NS			194, 201, 204, 209, 215, 218		222		
P2RX7	6	-3179,55	-3179,56	0.02	NS			86, 348, 367 441				
TNFRS1B	6	-2754,38	-2753,05	2.66	NS			527, 491				
IFNGR1	6	-3416,13	-3412,95	6.36	**	0.007; 14.81	2,9,23,27,36, 39,41,48,65, 82,87,90,91, 107,112,135, 145,148,157, 161.179,197, 204,212,215, 217,246,265, 271,314,340, 353,360,372, 381,407, 424	13,31,52, 72,80,118, 133,135,143, 146,156, 172, 204,246, 259, 266,315, 376, 392,394, 406, 425,435, 464, 486		52,65,72,79, 96,113,124, 181,246,268, 233,323,368, 373, 394	31,52,56, 118,143 171,181 204, 223, 246, 268, 323,368 373, 394	13
MC3R	4	-1317,41	-1317,01	0.8	NS			350, 79			81	
TIRAP	7	-1229,6	-1226,34	6.52	**	0.04; 1.74	22, 44, 81	44, 51, 53, 54, 81,116, 127, 95, 102, 104, 108 109, 113, 116, 118, 125, 129, 131, 133,	81			2
IL12RB1	6	-2789,97	-2779,083	21.78	**	0.16; 3.54	3, 34, 35, 38,39,46,262, 266, 27, 279, 302	165, 169, 177, 179, 369, 403, 410, 424, 435	58,165			1
CD209	8	-1902,79	-1898,85	7.88	**	0.03;4.47	69, 176, 207	24, 35, 39, 40,45,91, 99,143,144, 158,285, 312, 350,367, 378, 423		53, 158, 211	158	1

To evaluate the level of selective constraint among genes, the global *dN/dS* were estimated for each gene over the primate phylogeny as well as for the human lineage (Table 2). Genes inferred as positive selected form the substitution pattern (IFNGR1, CD209, TIRAP, IL12) show that on average they evolve faster that those negative selected (*t* test P=0.007). The domain-specific *dN/dS* value further shows that the extracellular and transmembrane domains evolve faster that the cytoplasmic domain in most of the genes.

The present work describes levels of variation among primates and humans for each receptor gene discussed above with the aim to provide a general picture of their evolution over different timescales.

Signatures of positive selection were detected in the rates of substitution across primates in most of the genes analyzed. Several ML methods identified specific codons with high probability of being under selection. Undoubtedly, the identification of sites subjected to positive selection does

conclusively provide insights into the function of a molecule. The comparison of human-primate gene family repertoires is an important step in order to clarify and identify selective pressure driving adaptive evolution.

Future investigation such as functional consequences of possible polymorphisms needed to be explored to detect the functional relevance of these receptor genes.

4. Conclusions

The identification of genes and gene regions subjected to positive selection can lead to predictions regarding the putative functionally important regions of genes. African primates remain an unknown font of information fundamental to complete the origin of many human pathogens. The close relationship between humans and chimpanzees is an important link. It may be fundamental to further research and establish primate specific selective events, in order to understand the evolutionary history-giving rise to humans.

Additionally crystallographic studies would be helpful for assessing the functional relevance of positive selected codons detected.

Table 2. *Selective pressure*

	Human branch	Global	SP	EXT	TM	CYT
P2RX7	0.14	0.14	0.12	0.60	0.27	n/a
TNRF	0.09	0.28	0.40	0.10	n/a	0.5
MC3R	0.17	0.05	0.10	0.32	0.41	0.10
IL12RB1	0.28	0.38	n/a	0.40	0.50	0.04
CD209	0.51	0.44	n/a	0.31	n/a	0.16
TIRAP	0.49	0.11	0.25	0.42	0.11	0.12
IFNGR1	0.25	0.60	0.50	0.30	0.32	0.50
VDR	0.35	0.07	0.04	0.06	n/a	n/a

SP= signal peptide; EXT=extracellular domain; TM=transmembrane domain; CYT= cytoplasmatic domain

Acknowledgements

References

[1]   R. Alfonso, RE. Romero, A. Diaz, N. Calderon, G. Urdaneta, J. Arce, ME. Patarroyo, MA. Patarroyo, "Isolation and identification of mycobacteria in New World primates maintained in captivity". Vet. Micro. Vol. 98, pp. 285–295, 2004

[2]   SV. Capuano, DA. Croix, S. Pawar, A. Zinovik, A. Myers, PL. Lin, S. Bissel, C. Fuhrman, E. Klein, JL. Flynn, "Experimental *Mycobacterium tuberculosis* infection of cynomolgus macaques closely resembles the various manifestations of human *M. tuberculosis* infection". Infect. Immun. Vol. 71, pp. 5831–5844, 2003

[3]   JL. Flynn, SV. Capuano, D. Croix, S. Pawar, A. Myers, A. Zinovik, E. Klein, "Non human primates: a model for tuberculosis research". Tuberculosis Vol. 83, pp.116-118, 2003

[4]   MC. King, AC. Wilson, "Evolution at two levels in humans and chimpanzees". Science Vol.188, pp. 107–116, 1975

[5]   MV. Olson, A. Varki, "Sequencing the chimpanzee genome: Insights into human evolution and disease". Nat. Rev. Genet. Vol. 4, pp. 20–28, 2003

[6]   M. Möller, EG. Hoal, "Current findings, challenges and novel approaches in human genetic susceptibility to tuberculosis". Tuberculosis Vol.90, pp. 71-83, 2010

[7]   P. Andolfatto, "Adaptive evolution of non-coding DNA in Drosophila". Nature Vol. 437, pp. 1149–1152, 2005

[8]   I. Ebersberger, D. Metzler, C. Schwarz, S. Paabo, "Genome wide comparison of DNA sequences between humans and chimpanzees". Am. J. Hum. Genet. Vol. 70, pp. 1490-1497, 2002

[9]   I. Hellmann, I. Ebersberger, SE. Ptak, S. Paabo, M. Przeworski, " A neutral explanation of AIF decreases mithocondrial oxidative phosphorylation and protects from obesity and diabetes". Am. J. Hum. Genet. Vol.72, pp. 1527–1535, 2003

[10]   J. Shi, H. Xi, Y. Wang, C. Zhang, Z. Jiang, K. Zhang, Y. Shen, L. Jin, W. Yuan, J. Lin, Q. Hua, F. Wang, S. Xu, S. Ren, G. Zhao, Z. Chen, W. Huang, "Divergence of the genes on human chromosome 21 between human and other hominoids and variation of substitution rates among transcription units". Proc Natl. Acad. Sci. U S A Vol. 100, pp. 8331-8336, 2003

[11]   MW. Karaman, ML. Houck, LG. Chemnick, S. Nagpal, D. Chawannakul, D. Sudano, BL. Pike, VV. Ho, OA. Ryder, JG. Hacia "Comparative analysis of gene-expression patterns in human and African ape cultured fibroblasts". Genome Res. Vol.13, pp.1619-1630, 2003

[12]   K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, "MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods". Mol Biol. Evol. Vol. 28, pp. 2731-2739, 2011

[13]   Z. Yang, "PAML 4: phylogenetic analysis by maximum likelihood". Mol Biol Evol. Vol.24, pp. 1586-1591, 2007

[14]   R. Nielsen, ZH. Yang, "Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene". Genet. Vol. 148, pp. 929–936, 1998

[15]   SKL. Pond, SDW. Frost, "Site-to-site variation of synonymous substitution rates". Mol. Biol. Evol. Vol. 22, pp. 2375-2385, 2005

[16]   T. Massingham, N. Goldman, "Detecting amino acid sites under positive selection and purifying selection". Genet. Vol.169 (3), pp. 1753-1762, 2005

[17]   MT. Hamblin, EE. Thompson, A. Di Rienzo, "Complex signatures of natural selection at the Duffy blood group locus". Am. J. Hum. Genet. Vol. 70, pp. 369-383, 2002

[18]   EJ. Vallender, BT. Lahn, "Positive selection on the human genome". Hum. Mol. Genet. Vol.13, pp. 245-254, 2004

[19] RJ. Woo, S. Kim, SE. Shoelson, S. Park, "X-ray Crystollographic Structure of TIR Domain from the human TIR-Domain containg adaptor protein/MyD88-Adaptor-like protein (TIRAP/MAL)". Bull. Korean Chem. Soc. Vol. 33, pp. 3091-3094, 2012

[20] DJ. Thiel, MH. le Du, RL. Walter, A. D'Arcy, C. Chene, M. Fountoulakis, G. Garotta, FK. Winkle, SE. Ealick, "Observation of an unexpected third receptor molecule in the crystal structure of human interferon-γ receptor complex". Structure Vol.8, pp. 927-936, 2000