

SHORT COMMUNICATION

Comprehensive Sequence Analysis of the Human *IL23A* Gene Defines New Variation Content and High Rate of Evolutionary Conservation

ELIZABETH A. Tindall and VANESSA M. Hayes*

Cancer Genetics Group, Children's Cancer Institute Australia for Medical Research, Lowy Cancer Research Centre, University of New South Wales, Sydney, Australia

*To whom correspondence should be addressed. Tel. +61 2-93851554. Fax. +61 2-93821850.
E-mail: vhayes@ccia.unsw.edu.au

Edited by Minoru Ko

(Received 10 November 2009; accepted 11 January 2010)

Abstract

A newly described heterodimeric cytokine, interleukin-23 (IL-23) is emerging as a key player in both the innate and the adaptive T helper (Th)17 driven immune response as well as an initiator of several autoimmune diseases. The rate-limiting element of IL-23 production is believed to be driven by expression of the unique p19 subunit encoded by *IL23A*. We set out to perform comprehensive DNA sequencing of this previously under-studied gene in 96 individuals from two evolutionary distinct human population groups, Southern African Bantu and European. We observed a total of 33 different DNA variants within these two groups, 22 (67%) of which are currently not reported in any available database. We further demonstrate both inter-population and intra-species sequence conservation within the coding and known regulatory regions of *IL23A*, supporting a critical physiological role for IL-23. We conclude that *IL23A* may have undergone positive selection pressure directed towards conservation, suggesting that functional genetic variants within *IL23A* will have a significant impact on the host immune response.

Key words: IL-23; *IL23A*; novel variants; genetic conservation

Interleukin (IL)-23 is a heterodimeric cytokine, comprised of a p40 subunit, which is common with IL-12, and a unique p19 subunit. Despite structural similarities between IL-12 and IL-23, they have vastly different immunological roles. The IL-23 complex is primarily secreted by monocyte-derived dendritic cells and macrophages¹ and is reported to play a critical role in survival of T helper (Th)17 cells and possibly the formation of Th17 memory cells.² Th17 cells have been implicated in resistance to fungal and extracellular bacterial pathogens as well as induction of various types of organ specific autoimmunity.³ In addition to the potential role in adaptive immunity, recent reviews have described IL-23 as being 'primitive in origin' and have proposed a primary role in the evolutionary older innate immune system,⁴ particularly in mediating

inflammatory attacks against pathogens of the intestine.^{5–9} Such pleiotropic cytokines may prove useful in determining an evolutionary link between the innate and adaptive immune pathways. Consistent with the innate immune system evolving prior to the adaptive immune response, it is reasonable to assume that molecules involved in this system may have evolved more rapidly and be genetically conserved across a broad spectrum of evolutionary distant species.

Since the discovery of IL-23, investigators have queried previous reports that target the shared p40 subunit to implicate IL-12 in several common autoimmune diseases.^{10–12} Subsequent studies have described a role for IL-23 (and in some instances exonerated IL-12) in the pathogenesis of many of these diseases.^{5–9}

The gene responsible for encoding the unique IL-23 p19 protein, later termed *IL23A*, was first described by Oppmann *et al.*¹ It is generally believed that the rate-limiting step of IL-23 production is expression of the p19 transcript.¹³ In light of the important role, IL-23 plays in both the innate and the adaptive immune response reportedly by enhancing the production of IL-17,¹⁴ it is within reason to hypothesize that regulation of *IL23A* expression, and therefore genetic variants within this gene, may be critical in determining the host immune response, as well as the development of autoimmune disease.

In this study, we determine the extent of *IL23A* sequence variation in two ethnically distinct human populations, including 48 Europeans from Australia and 48 Bantu from Southern Africa. Participants were

randomly selected, with no known disease status. Our linguistically mixed Southern African Bantu population (including Xhosa, Tswana and Sotho) more closely represents the origin of anatomically modern humans.¹⁵ We use the observed mutational spectrum and intra-species comparisons to determine the degree of evolutionary conservation and make predictions regarding functional implications.

Direct Sanger sequence analysis of *IL23A* (including ~2.9 kb 5' and 1.9 kb 3' of the transcription start and stop sites, respectively) in each of the two human populations, revealed a total of 33 different DNA variants present within the 6.36-kb region screened. All DNA variants underwent dual independent observations either as a result of bidirectional sequence validation and/or locality in overlapping fragments. With the

Table 1. Variant distribution of human *IL23A* in 96 individuals of Bantu and European ancestry

Population	RS number (if known)	Gene position	Variant ID	Variant frequency (n/48 samples)		
				Heterozygous	Homozygous	
Bantu (n = 48)		Promoter	-3009G>T	3	0	
		Promoter	-2186C>T	1	0	
		Promoter	-2572A>G ^a	7	1	
		Promoter	-2493T>G	2	0	
		Promoter	-2192C>T ^a	7	1	
		Promoter	-1982C>T	3	0	
		rs11835054	Promoter	-1774C>G	10	0
		rs73324334	Promoter	-1752C>G	20	18
			Promoter	-1538G>T ^a	7	1
			Promoter	-1316C>T	3	0
			Promoter	-1263C>T	1	0
		rs11465743	Promoter	-914C>T	3	0
			Promoter	-854A>G	1	0
		rs59094153	Promoter	-204A>G	3	0
			5'UTR/exon 1	-42G>A	1	0
		rs11465744	5'UTR/exon 1	-35G>A	2	0
			Exon 4	955A>G (Pro155Pro)	2	0
			Downstream	1792T>C	3	0
		rs11575253	Downstream	1945G>A	1	0
			Downstream	2136G>A	3	0
	rs7977692	Downstream	2421A>G	29	7	
		Downstream	3010T>G	1	0	
		Downstream	3176C>T	1	0	
European (n = 48)		Promoter	-2799C>T	1	0	
		Promoter	-2730T>C	1	0	
		rs73131077	Promoter	-2142G>A	1	0
			Promoter	-1956G>C	2	0
		rs73324334	Promoter	-1752C>G ^b	5	0
			Promoter	-1527T>C	1	0
			Promoter	-1517T>C	1	0
			5'UTR/Exon 1	-138G>A	1	0
		rs11171806	Exon 3	703G>A (Ser106Ser) ^b	5	0
		rs11575254	Downstream	1863C>T	1	0
	rs11575248	Downstream	3162G>T ^b	5	0	

Variants within a 6357-bp *IL23A* region were identified in both Bantu and Europeans for 48 individuals each, using direct Sanger sequencing (primer information and amplification conditions for the 10 overlapping fragments available upon request).

^aVariants in complete LD within the Bantu population.

^bVariants in complete LD within the European population.

exception of one variant at -1752 bp from the translation start site ($-1752C>G$; rs73324334), all variants observed were population specific (Table 1). The obvious distinction between variants in the two groups is important to note for investigating diseases with a strong correlation to ethnicity, particularly in this case, autoimmune-related diseases.^{16–18}

In addition to demonstrating the ethnic uniqueness of *IL23A*, we also highlight the limited genetic analysis of this gene to date. A total of 23 different genetic variants were identified in the Bantu, 16 (70%) of which had not previously been reported. Novel variants were also observed in the European population, and represented 7 (64%) of the 11 variants observed in this group. The increased observation of novel variation in the Southern African Bantu population collective can be explained by the exclusion of this largely diverse population grouping from current whole genome sequencing efforts defining current SNP database content. A recent report defines the Xhosa Bantu-speaking Southern Africans as genetically distinct from the peoples of Western African Yoruba.¹⁹

Haplotype predictive tools are not only important for identifying tag single-nucleotide polymorphisms (SNPs; relevant for disease association studies), but also for investigating the rate of gene evolution. As a result of population-specific variant distribution, linkage disequilibrium (LD) and haplotype analysis were performed on each the Bantu and European groups separately. Pairwise LD analysis on the Bantu samples, predicted a single haplotype block encompassing a 1-kb region towards the 5' end of the gene, which showed strong LD between four

variants, $-2572A>G$, $-2192C>T$, $-1982C>T$ and $-1538G>A$ (Fig. 1A). Three different combinations were predicted with the wild-type haplotype represented in 90.6% of the population (Fig. 1A). Overall, there were 25 allelic combinations predicted for the 23 variants identified in the Bantu (Supplementary Table S1). In Europeans, a single haplotype block within a 4-kb region towards the 3' end of the gene was determined, including three variants in complete LD, $-1752C>G$ (rs73324334), $703G>A$ (rs11171806, Ser106Ser) and $3162G>T$ (rs11575248) (Fig. 1B). The wild-type alleles were predicted to represent 94.8% of haplotypes (Fig. 1B). Frequency calculations for all 10 haplotype combinations identified in the Europeans demonstrate a high level of conservation within this group, with the wild-type alleles present in 85.4% of individuals (Supplementary Table S1). In accordance with genetic evolutionary data, the Southern African Bantu populations are among the genetically oldest populations in the world in comparison to the relatively young European populations,²⁰ thus the evolutionary time-line allows for more genetic variation (as depicted in this study) as well as more recombination events to have occurred. Although only a single haplotype block is depicted in each population (Fig. 1), the genetic distance within the Bantu haplotype (1 kb) is 4-fold smaller than the European haplotype block (4 kb). The importance of the identified variants and haplotypes in disease susceptibility studies would require further validation.

The majority of the 33 variants identified in this study occur outside the transcribed gene region. Of

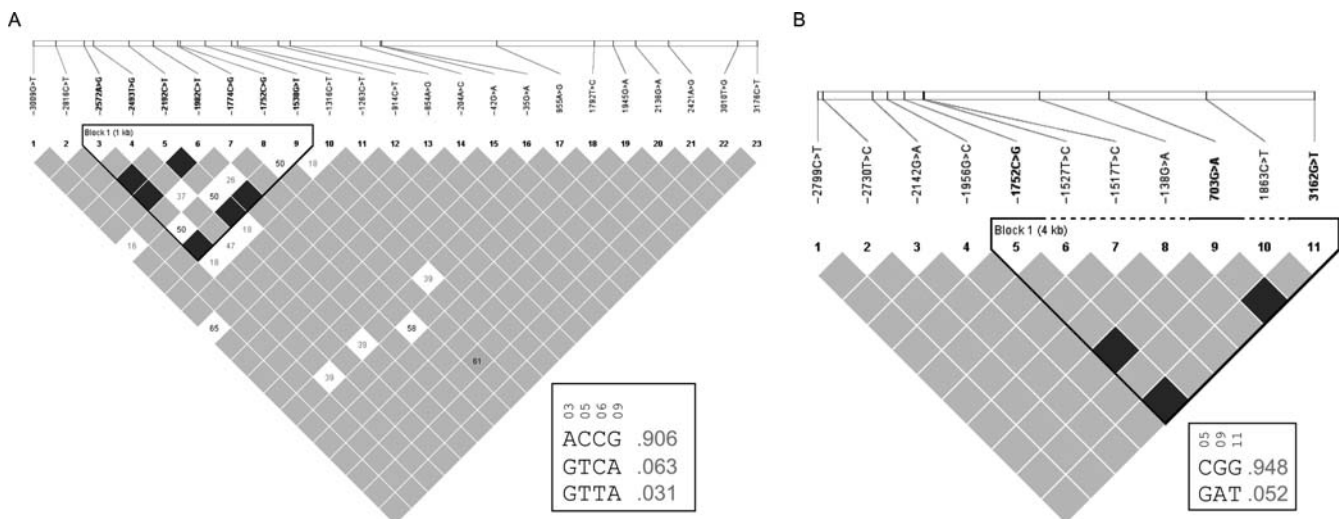


Figure 1. Pairwise LD plots and haplotype combinations of Bantu (A) and European (B) variants. (A) Variants 2572A>G, 2192C>T, 1982C>T and 1538G>T show strong LD within a 1-kb block in the Bantu population. Predicted haplotype frequencies for block 1 variants, depict the most common haplotype present in over 90% of the Bantu sequenced. (B) Variants 1752C>G, 703G>A and 3162G>T show strong LD within a 4-kb block in the Europeans. Predicted haplotype frequencies for block 1 reveal the most common haplotype is present in close to 95% of the Europeans sequenced. LD and haplotype information was generated using Hapview v4.1.

the five (15.6%) variants within the transcribed region, two are coding (both synonymous), and three occur in the 5' untranslated region (UTR) of exon 1. Similarly, of the 21 SNPs described on one or both EntrezSNP (<http://www.ncbi.nlm.nih.gov/snp/>) and SNPper (<http://snpper.chip.org>) SNP databases (nine of which are reported in this study), only four are located within the transcribed region of *IL23A*, three synonymous and one intronic. Only one of these previously reported transcribed SNPs was identified in our study. We conclude that the amino acid coding region of *IL23A* is highly conserved not only in the recently migrated (out of Africa) European population, but also the genetically diverse Southern African Bantu. Genomic regions, which exhibit a high degree of conservation, beyond neutral expectations, are often reported as having important functional roles.²¹ It may further be postulated that variants which might affect the function of *IL23A* may be found in the regulatory regions, rather than coding regions of this gene. A recent study, which used a genome wide scan to identify psoriasis susceptibility loci in a European cohort, was the first to implicate a genetic variant within the *IL23A* locus to autoimmune disease susceptibility.²² This variant, rs2066808, is located 3.7 kb 3' of *IL23A* (not within the 6.4-kb region of our analysis); however, any functional effect of this variant remains to be elucidated.

In this study, a total of 20 variants were observed 5' of the transcription start site. Initially, we assessed whether the variants we observed were located within known transcription factor binding sites (TFBSs), namely NF- κ B, ATF-2 and SMAD-3.^{23,24} We observed, no variants within these published sites. Assessing the entire 5' region using MatInspector (Genomatix) for the wild-type and variant promoter sequences for known *IL23A* TFBSs NF- κ B, ATF-2, SMAD-3 and Stat3,²⁵ no difference in transcription factor binding was predicted. Lack of direct impact on protein function or transcriptional activity is suggestive that *IL23A* variation defined in this study may be as a direct result of neutral contribution and that functionally important variants would have undergone selection pressure as a result of reduced fitness. Concluding that the regulatory region of *IL23A* is highly conserved is in concordance with the proposed hypothesis of conservation of regulatory regions of genes involved in highly complex, adaptive physiological processes.²⁶ As such genes are often not constitutively expressed, regulatory regions are considered more fundamental to maintaining an adequate response.

To determine the level of evolutionary conservation of *IL23A*, we performed sequence alignment of the transcribed region across the current content of

vertebrates within the UCSC Genome Browser (<http://genome.ucsc.edu>). Comparisons revealed a high level of genetic conservation within the coding region of *IL23A* across all species presenting with a human *IL23A* gene homolog (32 of 44) from the platypus to the chimpanzee (Supplementary Fig. S1). Of note is the large degree of conservation across the transcribed region within the extended primate family.

To further investigate intra-species conservation, the evolutionary relationship of *IL23A* between several mammalian species was predicted using Molecular Evolutionary Genetic Analysis version 4 (MEGA4) software.²⁷ All publicly available (<http://www.ncbi.nlm.nih.gov/>) *IL23A* mRNA sequences, including *Homo sapien* (Human, GenBank no. NM_016584), *Pan troglodytes* (Chimpanzee, XM_522436), *Sus scrofa* (Swine, NM_001130236), *Bos taurus* (Bovine, XM_588269), *Rattus norvegicus* (Norwegian rat, NM_130410) and *Mus musculus* (common house mouse, NM_031252), were aligned using ClustalW, and a phylogenetic tree was constructed using the neighbour-joining and maximum likelihood methods on the basis of nucleotide differences (Fig. 2A). This analysis revealed complete homology of the *IL23A* mRNA transcript between humans and the closely related chimpanzee indicating that in the estimated 5–7 million years since evolutionary divergence of these two species,²⁸ the transcribed sequence of *IL23A* has remained unaltered. In contrast, phylogenetic analysis of the gene, which encodes the p40 subunit of IL-23 and the binding partner of IL-23 p19, *IL12B*, revealed some (although minor) evolutionary divergence between the common transcript for both humans and chimpanzees (Fig. 2B). Similarly, the evolutionary distance between the human *IL23A* mRNA sequence and that of the remaining species was \sim 2-fold less than that observed between the human *IL12B* mRNA sequence and the alternative species (Fig. 2A scale bar represents 0.01 base substitutions per site compared with 0.02 in Fig. 2B). These results further illustrate the exceptional level of evolutionary conservation of *IL23A* compared with another closely related member of this cytokine family.

Genes that display limited intra- and inter-species variation are thought to have undergone positive selection pressure directed towards a functional conservation and may have played a pivotal role in species evolution.²⁸ Comparative genomics has revealed some of the most rapidly evolving genes are involved in reproduction and immune defense pathways.²¹ Genes which encode proteins with antibacterial, antiviral and antifungal activity are generally believed to be driven by co-evolution of host and pathogen and therefore pathogenic agents provide

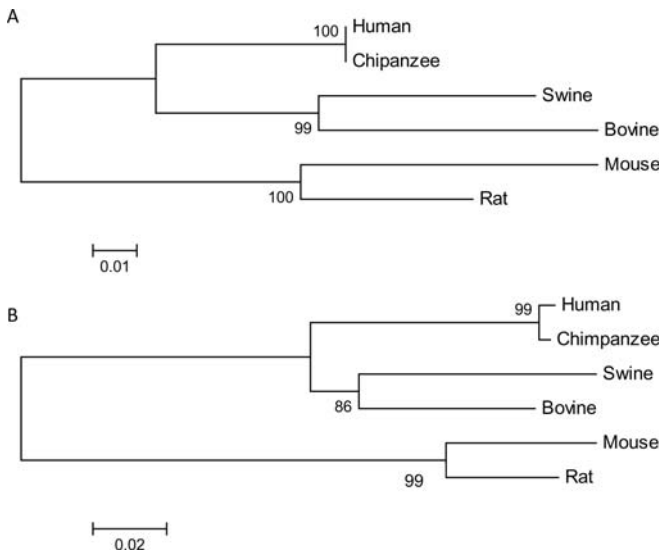


Figure 2. Phylogenetic tree representing the evolutionary relationship of six mammalian *IL23A* (**A**) and *IL12B* (**B**) mRNA sequences. Sequences were aligned using ClustalW and a neighbour-joining phylogenetic tree was generated using the maximum composite likelihood method and MEGA4. The tree is drawn to scale and the scale bar represents the same units as the evolutionary distance used to infer the phylogenetic tree; base substitutions per site. Reliability of the tree was assessed by bootstrapping, using 1000 bootstrap replications. Numbers at branch nodes represent the confidence level of 1000 bootstrap replications. (**A**) The human *IL23A* transcript (*H. sapiens*; NM_016584) is genetically identical to the predicted transcript of the chimpanzee (*P. troglodytes*; XM_522436). The next closest evolutionary related species is the swine (*S. scrofa*; NM_001130236), followed by the bovine (*B. taurus*; XM_588269), rat (*R. norvegicus*; NM_130410) and finally the mouse (*M. musculus*; NM_031252). The scale bar represents a genetic distance calculated by the number of base substitutions per site of 0.01. (**B**) The human *IL12B* transcript (NM_002187) is highly similar to the predicted transcript of the chimpanzee (XM_527101). The next closest evolutionary-related species is the swine (NM_214013), followed by the bovine (NM_174356), rat (NM_022611) and finally the mouse (NM_008352). The scale bar represents a genetic distance calculated by the number of base substitutions per site of 0.02.

some of the strongest selective pressures on human evolution.²⁹ A proposed role for IL-23 in mucosal immunity¹⁰ thus provides some explanation for the conservation of *IL23A*.

It has been suggested that when the immune response maintained by IL-23 persists in an uncontrolled manner, this protein may switch from being an integral part of the host defense to being a pro-inflammatory contributor to autoimmune disease.⁴ Thus, as a key component of the IL-23 complex, we suggest that *IL23A* may be vitally important to maintaining both innate and autoimmune responses. Although in this study we identify novel DNA variation at the *IL23A* locus, we also demonstrate the highly conserved nature of *IL23A*, at both amino acid and predicted regulatory levels. This study

therefore highlights the potential significance of genetic variants found within this gene locus. Identifying functional variants, whether they are common disease associated variants or rare non-synonymous mutations, may provide key information in determining a role for *IL23A* in disease progression.

Acknowledgements: We would like to thank Dr Desiree Petersen and Prof. Web Miller for helpful communications. Ethics approval for this study was provided by the Human Ethics Research Committee of the University of New South Wales (HREC #08244).

Supplementary data: Supplementary data are available at www.dnaresearch.oxfordjournals.org.

Funding

Funding sources include Cancer Institute New South Wales (CINSW) career development and support fellowship to V.M.H., CINSW scholar award and Australian Rotary Health PhD scholarship to E.A.T.

References

- Oppmann, B., Lesley, R., Blom, B., et al. 2000, Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with biological activities similar as well as distinct from IL-12, *Immunity*, **13**, 715–25.
- McGeachy, M.J. and Cua, D.J. 2008, Th17 cell differentiation: the long and winding road, *Immunity*, **28**, 445–53.
- Boniface, K., Blom, B., Liu, Y.J. and de Waal Malefyt, R. 2008, From interleukin-23 to T-helper 17 cells: human T-helper cell differentiation revisited, *Immunol. Rev.*, **226**, 132–46.
- Tato, C.M. and Cua, D.J. 2008, Reconciling id, ego, and superego within interleukin-23, *Immunol. Rev.*, **226**, 103–11.
- Hue, S., Ahern, P., Buonocore, S., et al. 2006, Interleukin-23 drives innate and T cell-mediated intestinal inflammation, *J. Exp. Med.*, **203**, 2473–83.
- Kullberg, M.C., Jankovic, D., Feng, C.G., et al. 2006, IL-23 plays a key role in Helicobacter hepaticus-induced T cell-dependent colitis, *J. Exp. Med.*, **203**, 2485–94.
- Uhlir, H.H., McKenzie, B.S., Hue, S., et al. 2006, Differential activity of IL-12 and IL-23 in mucosal and systemic innate immune pathology, *Immunity*, **25**, 309–18.
- Yen, D., Cheung, J., Scheerens, H., et al. 2006, IL-23 is essential for T cell-mediated colitis and promotes inflammation via IL-17 and IL-6, *J. Clin. Invest.*, **116**, 1310–6.
- Cua, D.J., Sherlock, J., Chen, Y., et al. 2003, Interleukin-23 rather than interleukin-12 is the critical cytokine for autoimmune inflammation of the brain, *Nature*, **421**, 744–8.

10. McGovern, D. and Powrie, F. 2007, The IL23 axis plays a key role in the pathogenesis of IBD, *Gut*, **56**, 1333–6.
11. Torti, D.C. and Feldman, S.R. 2007, Interleukin-12, interleukin-23, and psoriasis: current prospects, *J. Am. Acad. Dermatol.*, **57**, 1059–68.
12. Furuzawa-Carballeda, J., Vargas-Rojas, M.I. and Cabral, A.R. 2007, Autoimmune inflammation from the Th17 perspective, *Autoimmun. Rev.*, **6**, 169–75.
13. Lyakh, L., Trinchieri, G., Provezza, L., Carra, G. and Gerosa, F. 2008, Regulation of interleukin-12/interleukin-23 production and the T-helper 17 response in humans, *Immunol. Rev.*, **226**, 112–31.
14. Sutton, C.E., Lalor, S.J., Sweeney, C.M., Brereton, C.F., Lavelle, E.C. and Mills, K.H. 2009, Interleukin-1 and IL-23 induce innate IL-17 production from gammadelta T cells, amplifying Th17 responses and autoimmunity, *Immunity*, **31**, 331–41.
15. Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., et al. 2009, The genetic structure and history of Africans and African Americans, *Science*, **324**, 1035–44.
16. Boster, A.L., Endress, C.F., Hreha, S.A., Caon, C., Perumal, J.S. and Khan, O.A. 2009, Pediatric-onset multiple sclerosis in African-American black and European-origin white patients, *Pediatr. Neurol.*, **40**, 31–3.
17. Lee, H.S., Korman, B.D., Le, J.M., et al. 2009, Genetic risk factors for rheumatoid arthritis differ in Caucasian and Korean populations, *Arthritis Rheum.*, **60**, 364–71.
18. White, J.M., O'Connor, S., Winter, H.S., et al. 2008, Inflammatory bowel disease in African American children compared with other racial/ethnic groups in a multicenter registry, *Clin. Gastroenterol. Hepatol.*, **6**, 1361–9.
19. Patterson, N., Petersen, D.C., van der Ross, R.E., et al. 2009, Genetic structure of a unique admixed population: implications for medical research, *Hum. Mol. Genet.*, **19**, 411–9.
20. Campbell, M.C. and Tishkoff, S.A. 2008, African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping, *Annu. Rev. Genomics Hum. Genet.*, **9**, 403–33.
21. Ellegren, H. 2008, Comparative genomics and the study of evolution by natural selection, *Mol. Ecol.*, **17**, 4586–96.
22. Nair, R.P., Duffin, K.C., Helms, C., et al. 2009, Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways, *Nat. Genet.*, **41**, 199–204.
23. Al-Salleeh, F. and Petro, T.M. 2008, Promoter analysis reveals critical roles for SMAD-3 and ATF-2 in expression of IL-23 p19 in macrophages, *J. Immunol.*, **181**, 4523–33.
24. Mise-Omata, S., Kuroda, E., Niikura, J., Yamashita, U., Obata, Y. and Doi, T.S. 2007, A proximal kappaB site in the IL-23 p19 promoter is responsible for RelA- and c-Rel-dependent transcription, *J. Immunol.*, **179**, 6596–603.
25. Kortylewski, M., Xin, H., Kujawski, M., et al. 2009, Regulation of the IL-23 and IL-12 balance by Stat3 signaling in the tumor microenvironment, *Cancer Cell*, **15**, 114–23.
26. Lee, S., Kohane, I. and Kasif, S. 2005, Genes involved in complex adaptive processes tend to have highly conserved upstream regions in mammalian genomes, *BMC Genomics*, **6**, 168.
27. Tamura, K., Dudley, J., Nei, M. and Kumar, S. 2007, MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0, *Mol. Biol. Evol.*, **24**, 1596–9.
28. Kehrer-Sawatzki, H. and Cooper, D.N. 2007, Understanding the recent evolution of the human genome: insights from human-chimpanzee genome comparisons, *Hum. Mutat.*, **28**, 99–130.
29. Nakajima, T., Ohtani, H., Satta, Y., et al. 2008, Natural selection in the TLR-related genes in the course of primate evolution, *Immunogenetics*, **60**, 727–35.