Published in final edited form as:
Int J Immunogenet. 2013 August ; 40(4): 261–269. doi:10.1111/iji.12021.

Variation in human β-defensin genes: new insights from a multi-population study

Rajeev K. Mehlotra[1], **Peter A. Zimmerman**[1], **Aaron Weinberg**[2], and **Richard J. Jurevic**[2,†]

[1] Center for Global Health and Diseases, Case Western Reserve University School of Medicine, Cleveland, OH, U.S.A.

[2] Department of Biological Sciences, Case Western Reserve University School of Dental Medicine, Cleveland, OH, U.S.A.

Summary

Human β-defensin 2 (hBD-2) and hBD-3, encoded by *DEFB4* and *DEFB103A*, respectively, have shown anti-HIV activity, and both genes exhibit copy number variation (CNV). Although the role of hBD-1, encoded by *DEFB1*, in HIV-1 infection is less clear, single nucleotide polymorphisms (SNPs) in *DEFB1* may influence viral loads and disease progression. We examined the distribution of *DEFB1* SNPs and *DEFB4/103A* CNV, and the relationship between *DEFB1* SNPs and *DEFB4/103A* CNV using samples from two HIV/AIDS cohorts from the United States (n = 150) and five diverse populations from the Coriell Cell Repositories (n = 46). We determined the frequencies of 10 SNPs in *DEFB1* by using a post-PCR, oligonucleotide ligation detection reaction-fluorescent microsphere assay, and CNV in *DEFB4/103A* by real-time quantitative PCR. There were noticeable differences in the frequencies of *DEFB1* SNP alleles and haplotypes among various racial/ethnic groups. The *DEFB4/103A* copy numbers varied from 2 to 8 (median, 4), and there was a significant difference between the copy numbers of self-identified whites and blacks in the US cohorts (Mann-Whitney *U* test $p = 0.04$). A significant difference was observed in the distribution of *DEFB4/103A* CNV among *DEFB1* -52G/A and -390T/A genotypes (Kruskal-Wallis $p = 0.017$ and 0.026, respectively), while not in the distribution of *DEFB4/103A* CNV among -52G/A_-44C/G_-20G/A diplotypes. These observations provide additional insights for further investigating the complex interplay between β-defensin genetic polymorphisms and susceptibility to, or the progression or severity of, HIV infection/disease.

Introduction

Human defensins are a family of small, β-sheeted, cationic, antimicrobial and immunoregulatory peptides, which belong to either the α or β subfamily. The former are found in azurophilic granules of phagocytic cells, while the latter are released from epithelial cells (Chen *et al.*, 2006; Ganz, 2003; Klotman and Chang, 2006). Recent studies have suggested that human β-defensins (hBDs) may play an important role in HIV-1 susceptibility and disease progression (Feng *et al.*, 2006; Quinones-Mateu *et al.*, 2003; Sun *et al.*, 2005; Zapata *et al.*, 2008). We have demonstrated that: (1) both R5 and X4 HIV-1 phenotypes induce *DEFB4* and *DEFB103A* mRNAs, encoding hBD-2 and -3, respectively, in normal human oral epithelial cells; (2) hBD-2 and -3 inhibit HIV-1 infection by both phenotypes, with greater activity against X4 viruses; and (3) this inhibition is due to a direct interaction with virions, and through down-modulation of the CXCR4 co-receptor (Quinones-Mateu *et al.*, 2003). A later study confirmed the anti-HIV activity of these hBDs

[†] Corresponding author Address: 10900 Euclid Avenue, School of Dental Medicine #3550, Cleveland, OH 44106-4905, U.S.A. Phone: +1-216-368-5218, Fax: +1-216-368-0145, rjj11@case.edu.

(Sun *et al.*, 2005). We subsequently discovered that hBD-3 acts as an antagonist of the CXCR4 co-receptor by promoting its internalization (Feng *et al.*, 2006). Concordant with the results of our *in vitro* study (Quinones-Mateu *et al.*, 2003), our group also found that HIV-exposed seronegative and HIV-seropositive Colombian individuals expressed significantly higher levels of *DEFB4* and *DEFB103A* transcripts in oral mucosa than did healthy controls (Zapata *et al.*, 2008).

Single nucleotide polymorphisms (SNPs) in *DEFB1*, encoding hBD-1, particularly -52G/A (660G/A, rs1799946), -44C/G (668C/G, rs1800972), and -20G/A (692G/A, rs11362), may also play a role in HIV-1 susceptibility and disease progression. By comparing the genotype frequencies between HIV-positive patients and comparable healthy controls, significant associations were observed between -52G/A (Braida *et al.*, 2004; Milanese *et al.*, 2006; Segat *et al.*, 2006), -44C/G (Braida *et al.*, 2004; Segat *et al.*, 2006), and/or -20G/A (Milanese *et al.*, 2006) genotypes and HIV-1 infection status. However, considering the variability in associations found in different (Braida *et al.*, 2004; Milanese *et al.*, 2006) and even the same (Milanese *et al.*, 2006; Segat *et al.*, 2009) populations, it appears that the *DEFB1* genotype associations may be cohort-specific. Other studies have shown that -52G/A, either singly or as haplotype with -44C/G, may influence viral loads in breast milk (Baroncelli *et al.*, 2008) and plasma (Ricci *et al.*, 2009), respectively. Recently, in HIV-1-infected children, -44CG genotype and -52G_-44G haplotype were significantly associated with a slower disease progression (Freguja *et al.*, 2012).

DEFB1, *DEFB4*, and *DEFB103A* are located in a cluster on chromosome 8p23.1, and are polymorphic. *DEFB1* has numerous SNPs (Jurevic *et al.*, 2002; Kim *et al.*, 2009; Prado-Montes de Oca, 2010), with at least 25 SNPs in the coding region (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=1672). Among these, -52G/A, -44C/G, and -20G/A in the exon-1 region have been the focus of a variety of studies related to functional consequences of variation in *DEFB1*, and *DEFB1* genetic associations where these SNPs were considered either singly or as haplotypes (Kalus *et al.*, 2009; Naslavsky *et al.*, 2010; Naslavsky *et al.*, 2009; Prado-Montes de Oca, 2010). In addition, with respect to size, gene content and numeric variability of *DEFB*, the 8p23.1 region represents one of the hotspots of copy number variation (CNV) (Abu Bakar *et al.*, 2009; Hollox *et al.*, 2008). *DEFB1* is generally considered a single-copy gene (2 copies per diploid genome [PDG]), and CNVs are very rare (Cagliani *et al.*, 2008). However, the 200 kb *DEFB* region, containing *DEFB4* and *DEFB103A*, varies *en bloc* in its copy number (Groth *et al.*, 2008). Individuals carry 2-12 copies PDG (Groth *et al.*, 2008; Hollox *et al.*, 2003; Linzmeier and Ganz, 2005). As with SNPs, CNVs may also affect gene expression and hence determine phenotypes (Ionita-Laza *et al.*, 2009; Zhang *et al.*, 2009). It is therefore not surprising that the *DEFB* CNV is associated with clinically important phenotypes (Hollox, 2008). In addition to CNV, sequence variations between copies (termed as multisite variations) represent a further level of complexity at *DEFB* (Groth *et al.*, 2010; Huse *et al.*, 2008). Furthermore, whether there is any relationship (association, linkage disequilibrium [LD]) between SNPs in *DEFB1* and CNV in *DEFB4/103A* is not clear, and there is some indication that there may not be (Hollox, 2008).

Herein, we report the frequencies of 10 *DEFB1* SNPs (Jurevic *et al.*, 2002) and their haplotypes in two admixed, HIV/AIDS patient populations from the United States, together with five diverse populations from the Coriell Cell Repositories. We also report the distribution of *DEFB4/103A* CNV, and show, for the first time, that there may be a relationship between *DEFB1* SNPs and *DEFB4/103A* CNV.

Materials and methods

Samples and genomic DNA extraction

De-identified packed blood pellets (n = 105), collected from HIV-infected adult subjects, were obtained from the Case Western Reserve University Center for AIDS Research (CFAR) specimen repository. The self-identified racial/ethnic distribution of these subjects was: whites (n = 37), blacks (n = 65), and Hispanics (n = 3). De-identified packed white blood cell pellets (n = 45), collected from HIV/AIDS patients, were obtained from the Multicenter AIDS Cohort Study (MACS) (Shepherd *et al.*, 2008). The self-identified racial/ethnic distribution of these subjects was: white, non-Hispanic (n = 39), white, Hispanic (n = 4), and black, non-Hispanic (n = 2). DNA samples from five diverse populations (Northern European [NE, n = 10], sub-Saharan African [AfSS, n = 9], African-American [Af-Am, n = 10], Chinese [CHN, n = 10], and Mexican [MEX, n = 7]) were obtained from the Coriell Cell Repositories, Coriell Institute for Medical Research, Camden, NJ.

Genomic DNAs were extracted from 200 μl of each of the pellets using the QIAamp DNA Blood Mini Kit (Qiagen Inc., Valencia, CA). The CFAR specimen collection protocol and genetic data storage and usage were approved by the Institutional Review Board of University Hospitals Case Medical Center.

DEFB1 polymerase chain reaction (PCR) and SNP genotyping

PCR primers were designed based on the *DEFB1* region of the sequence (GenBank accession #NT_023736, nucleotide coordinates 6,726,529-6,717,101) to selectively amplify the exon-1 and exon-2 regions. Sequence homology and specificity of all primer sequences were checked using the $BLAST_n$ program (http://www.ncbi.nlm.nih.gov). The primer sequences, PCR buffer, amplification conditions, and method used to perform agarose gel electrophoresis are described in supplementary Table S1.

Genotyping of six SNPs in the exon-1 region (-610G/A [102G/A, rs2741132], -390T/A [322T/A, rs2738182], -52G/A, -44C/G, -20G/A, and 79T/A [791T/A, rs2293958]), and four SNPs in the exon-2 region (1654G>A [Val38Ile, rs2738047], 1754G/A [rs1047031], 1836A/G [rs1800971], and 1873T/C [rs5743491]) (Jurevic *et al.*, 2002) was performed by using a high-throughput, post-PCR oligonucleotide ligation detection reaction-fluorescent microsphere assay (LDR-FMA) on the Bio-Plex™ multiplex suspension array system (Bio-Rad Laboratories, Hercules, CA) (Mehlotra *et al.*, 2007; Mehlotra *et al.*, 2006). The LDR primers were designed based on the *DEFB1* sequences available through GenBank (accession #NT_023736) and SNPper (http://snpper.chip.org/bio/export-sequence/6559); the primer sequences are provided in supplementary Table S2. The LDR conditions were 95°C for 1 min, 95°C for 15 sec and 60°C for 2 min (31×). The *DEFB1* SNP genotypes were determined as previously described (Mehlotra *et al.*, 2007; Mehlotra *et al.*, 2006). The mean and 95% confidence interval (CI) of log-transformed fluorescent values corresponding to each genotype, using the Coriell samples (n = 46), are presented in supplementary Table S3.

DEFB4/103A CNV determination

For the determination of *DEFB4* and *DEFB103A* copy numbers, the real-time quantitative PCR assay was used as described (Linzmeier and Ganz, 2005). Reference genes *TBP* (TATA-Box Binding Protein, GenBank accession #AL031259) and *DEFB1*, specific primer sets producing only one specific product of ~150 bp at 54°C annealing temperature, reaction mix, and conditions were used as described (Linzmeier and Ganz, 2005), and the Bio-Rad CFX96™ system (Bio-Rad Laboratories, Hercules, CA) was used for the PCR analysis. Each sample was amplified using 50 ng genomic DNA, and was run in triplicate. Data were analyzed by the comparative Ct method, and the copy numbers were calculated as described

(Linzmeier and Ganz, 2005). The assay results were validated by re-running 12 of the 46 randomly-selected Coriell samples.

Statistical analysis

HaploView v4.2 (Barrett *et al.*, 2005) was used to calculate *DEFB1* SNP frequencies, the Hardy-Weinberg (H-W) equilibrium p values, and Lewontin's D' and correlation coefficient (r^2) parameters as measures of LD between SNPs. PHASE v2.1.1 (http://www.stat.washington.edu/stephens/) was used to infer *DEFB1* 10-SNP haplotypes and diplotypes. The QI Macros Statistical Process Control Software for Excel (QI Macros 2012) was used to calculate descriptive statistics, and to perform Chi-square (χ^2) test of contingency table and Mann-Whitney U test. SAS v9.2 was used to perform the nonparametric Kruskal-Wallis test to compare *DEFB4/103A* CNV among *DEFB1* diplotypes and SNP genotypes, and to perform Bonferroni (Dunn's) t-test for multiple comparisons. For all statistical analyses, differences were considered significant at $p < 0.05$.

Results

DEFB1 SNP-allele frequencies and H-W equilibrium

The *DEFB1* SNP-allele frequencies in two HIV/AIDS cohorts, CFAR and MACS, and five Coriell populations are presented in Table 1. These results do not include the three Hispanic patient samples in the CFAR cohort, and the four white, Hispanic, and two black, non-Hispanic patient samples in the MACS cohort. The SNP-genotype data for all these samples (n = 9) are presented in supplementary Table S4. Taken collectively, higher frequencies of the six exon-1 SNPs were observed, compared with the frequencies of the four exon-2 SNPs (Table 1). The frequencies of the -44G and 79A alleles were noticeably higher in white HIV/AIDS patients, NE, and MEX, and those of the 1754A allele were noticeably higher in white HIV/AIDS patients, NE, and CHN. On the other hand, the frequencies of the 1654A and 1836G alleles were noticeably higher in black HIV patients, AfSS, and Af-Am. These results indicate that the *DEFB1* SNPs are distributed differently among different racial/ethnic groups, although a formal test to determine the statistical significance of these frequency differences was not performed due to low sample sizes of the Coriell populations. We calculated expected genotype numbers for each of the *DEFB1* SNPs in each population. For all SNPs, the expected genotype numbers did not differ significantly from the observed genotype numbers in any of the populations (data not shown), indicating no deviation from H-W equilibrium.

DEFB1 SNP-SNP LD patterns

We quantified the extent of LD among the *DEFB1* SNP pairs in all populations (Figure 1A-1H). Strong LD, defined by high values for both D' (≥ 0.8) and r^2 (≥ 0.5) parameters (Ferlin *et al.*, 2010), was observed between SNP pairs -390T/A – -52G/A and -44C/G – 79T/A in all populations. Inter-population differences in the LD patterns occurred: strong LD was observed between SNP pairs -610G/A – -390T/A and -610G/A – -52G/A in white HIV/AIDS patients, NE, CHN, and MEX, but not in black HIV patients, AfSS, and Af-Am; strong LD was also observed between a number of additional SNP pairs in CHN, which was not evident in any other population; and white HIV/AIDS patients and black HIV patients/Af-Am showed differences in pattern and extent of LD when compared with NE and AfSS, respectively. It is well-known that in the US populations, demographic factors, such as recent migration and admixture, have contributed to the distribution and extent of disequilibrium (Ardlie *et al.*, 2002).

DEFB1 haplotype profiles

We inferred *DEFB1* 10-SNP haplotypes using PHASE v2.1.1 (Table 2a). Assuming random associations among the eight polymorphic SNPs (Table 1), one can predict 256 different haplotypes (2^8). However, only nine haplotypes (8 + 1) would be possible if mutations are the only evolutionary forces acting to create new alleles, and other forces, such as recombination, recurrent mutation, and gene conversion, do not occur (Bonnen *et al.*, 2002). We inferred a total of 13 haplotypes in our study populations, which is closer to the theoretical minimum. Of these, four haplotypes, GTGCATGGAT (0.05-0.31), GTGCATGAAT (0.05-0.2), GTGGGAGGAT (0.05-0.3), and AAACGTGGAT (0.22-0.55), together accounted for the majority of total chromosomes examined. Considerable differences in the haplotype profiles were noticed between the populations of European ancestry (white HIV/AIDS patients and NE) and African ancestry (black HIV patients, AfSS, and Af-Am).

We then determined the frequencies of 3-SNP haplotypes (-52G/A_-44C/G_-20G/A) (Table 2b). As expected from the 10-SNP haplotype data, GCA (0.29-0.44), GGG (0.05-0.3), and ACG (0.28-0.6) were the predominant haplotypes among all populations.

DEFB4/103A CNV distribution

In all samples combined (total n = 187 [CFAR, n = 102; MACS, n = 39; Coriell, n = 46]), the copy numbers ranged from 2 to 8 (median, 4). Copy numbers 2 (n = 35), 3 (n = 54), 4 (n = 61), and 5 (n = 28) were highly prevalent, whereas 6, 7, and 8 were uncommon (n = 3 each). We compared *DEFB4/103A* CNV distribution between the two major racial groups in the US; i.e., self-identified whites and blacks. For this analysis, CNV results of CFAR white patients were combined with those of MACS white, non-Hispanic patients (total n = 76), and then compared with those of CFAR black patients (n = 65). We observed a significant difference in the CNV distribution between the two groups (white [median, 3], black [median, 4], Mann-Whitney *U* test $p = 0.04$). No difference was observed in the CNV distribution between CFAR white patients (n = 37) and MACS white, non-Hispanic patients (n = 39) (Mann-Whitney *U* test $p = 0.47$).

We then compared *DEFB4/103A* CNV distribution among *DEFB1* diplotypes and SNP genotypes. For this analysis, we considered -52G/A_-44C/G_-20G/A diplotypes and the SNP genotypes of all samples combined (total n = 187). We observed a total of nine 3-SNP diplotypes among all 187 samples. The distribution of *DEFB4/103A* CNV among these diplotypes is presented in Figure 2 and Table 3. A χ^2 test of contingency table revealed random distribution of CNV among the diplotypes ($\chi^2 = 34.87$, df = 48, $p = 0.922$). Comparisons of means by Kruskal-Wallis test also suggested no significant difference in the distribution of *DEFB4/103A* CNV among the *DEFB1* diplotypes ($p = 0.109$).

Lastly, while no significant differences were observed in the distribution of *DEFB4/103A* CNV among -44C/G and -20G/A genotypes (Kruskal-Wallis $p = 0.315$ and 0.716, respectively), a significant difference was observed in the distribution of *DEFB4/103A* CNV among -52G/A genotypes (Kruskal-Wallis $p = 0.017$). We then performed Bonferroni (Dunn's) t-test for multiple comparisons to identify which -52G/A genotype groups differ in the distribution of *DEFB4/103A* CNV. This analysis revealed that -52AA (mean CNV = 4.19, n = 36) was significantly different from -52GA (mean CNV = 3.44, n = 94) (Simultaneous 95% CI, 0.19-1.33). Since -52G/A is in complete LD with -390T/A in all populations (Figure 1A-1H), a significant difference was also observed in the distribution of *DEFB4/103A* CNV among -390T/A genotypes (Kruskal-Wallis $p = 0.026$). Bonferroni (Dunn's) t-test for multiple comparisons revealed that -390AA (mean CNV = 4.16, n = 37)

was significantly different from -390TA (mean CNV = 3.44, n = 93) (Simultaneous 95% CI, 0.15-1.29).

Discussion

In the present study, utilizing samples from two HIV/AIDS cohorts from the US and five diverse populations from the Coriell Cell Repositories, we made the following significant observations regarding β-defensin genetic variation: We observed that there were noticeable differences in the frequencies of *DEFB1* SNP alleles (Table 1) and haplotypes (Tables 2a and 2b) among various racial/ethnic groups. To the best of our knowledge, the information regarding comparative distribution of *DEFB1* polymorphisms among various populations is limited and, among admixed, HIV/AIDS patient populations from the US, is not available. The majority of studies have been conducted in European populations (Braida *et al.*, 2004; Carter *et al.*, 2010; Ricci *et al.*, 2009; Schaefer *et al.*, 2010). Limited information is available regarding the distribution of *DEFB1* polymorphisms in other populations, such as Asian (Chen *et al.*, 2007; Kim *et al.*, 2009; Leung *et al.*, 2006), African (Mozambican) (Baroncelli *et al.*, 2008), Brazilian (Milanese *et al.*, 2006; Segat *et al.*, 2009), American, either mixed (Jurevic *et al.*, 2003; Ozturk *et al.*, 2010) or Caucasian-American and Af-Am (Boniotto *et al.*, 2004), and Mexican (Prado-Montes de Oca *et al.*, 2006). Our *DEFB1* SNP-allele and haplotype frequency results are in agreement with those previously reported for comparable populations. Furthermore, our comparative *DEFB1* SNP-allele and haplotype frequency results indicate noticeable differences among diverse populations, thus providing highly valuable information for future genetic studies of the etiology of diseases where hBD-1 could play an important modulatory role.

We observed that in our study samples, *DEFB4/103A* copy numbers varied from 2 to 8 (median, 4). Using different methods, it has been consistently found that these genes vary from 2 to 12 copies PDG, and the copy numbers between 2 and 6 are common, with a modal copy number of 4 in most populations (Fode *et al.*, 2011a; Fode *et al.*, 2011b; Hollox, 2008; Linzmeier and Ganz, 2005; Taudien *et al.*, 2010). Thus, our results regarding *DEFB4/103A* CNV, in all samples combined, are in agreement with the published findings. We also observed that there was a significant difference between the copy numbers of two major racial groups in the US; i.e., self-identified whites and blacks. Although in all populations the modal *DEFB4/103A* copy number is 4, there appear to be differences in the mean copy numbers, with the distribution of Chinese and Yorubans (from Ibadan, Nigeria) shifted slightly towards higher copy numbers compared with European populations (Hollox, 2008). However, in a recent study, there was no significant difference in the mean copy numbers between European and Ghanaian populations (Fode *et al.*, 2011a). Using our Coriell samples, we compared the *DEFB4/103A* copy numbers between NE and the populations of African ancestry (AfSS and Af-Am, combined), and found no significant difference (Mann-Whitney U test $p = 0.13$). Thus, our finding that a significant difference in the copy numbers occurs between self-identified whites and blacks from the US could be unique to admixed populations or these particular sets of samples, and therefore requires further investigations.

We observed a significant difference in the distribution of *DEFB4/103A* CNV among *DEFB1* -52G/A genotypes, but not in the distribution of *DEFB4/103A* CNV among -52G/A_-44C/G_-20G/A diplotypes. Furthermore, since -52G/A is in complete LD with another *DEFB1* SNP -390T/A, a significant difference was observed in the distribution of *DEFB4/103A* CNV among -390T/A genotypes as well. To our knowledge, whether *DEFB4/103A* CNV has a relationship (association, LD) with *DEFB1* SNPs has not been reported elsewhere. However, evidence from family studies, HapMap phase II data from four ethnically diverse populations, and genomic characteristics of the 8p23.1 region suggests that *DEFB* CNV is likely not associated with neighboring SNP alleles (Hollox,

2008). It is important to recognize that our results, based on comparing the means of copy numbers by a nonparametric test, Kruskal-Wallis, are not indicative of an association between *DEFB4/103A* CNV and *DEFB1* SNPs. Nevertheless, they provide a basis for a more in-depth investigation of genetic variation in and around these loci: Determination of *DEFB* copy number diplotypes (Hollox, 2008) and application of LD/haplotype analysis approach (Conrad *et al.*, 2009; de Smith *et al.*, 2008; Menard *et al.*, 2009) may provide a better assessment of association between *DEFB4/103A* CNV and *DEFB1* SNPs.

In conclusion, our observations suggest that there may be an interracial difference in *DEFB4/103A* copy numbers between admixed populations, and a relationship between *DEFB1* SNPs and *DEFB4/103A* CNV. Given the emerging significance of *DEFB1* SNPs (Baroncelli *et al.*, 2008; Braida *et al.*, 2004; Freguja *et al.*, 2012; Milanese *et al.*, 2006; Ricci *et al.*, 2009; Segat *et al.*, 2009; Segat *et al.*, 2006) and *DEFB4*- and *DEFB103A*-encoded hBD-2 and hBD-3, respectively, (Feng *et al.*, 2006; Quinones-Mateu *et al.*, 2003; Sun *et al.*, 2005; Zapata *et al.*, 2008) in influencing HIV infection and disease, these observations deserve further exploration. Future studies will help unravel the complex interplay between β-defensin genetic polymorphisms and susceptibility to, or the progression or severity of, HIV infection/disease, complementing our current efforts to further lessen the morbidity and mortality due to this global killer.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are indebted to Michael Lederman and Benigno Rodriguez for providing the HIV-infected patient samples from the CFAR specimen repository, and to Janet Schollenberger for providing the HIV/AIDS patient samples from the MACS. We are thankful to Dave McNamara, Carolyn Myers, Tenisha Phipps, and Bangan John for thorough reading and constructive criticism of the manuscript. We sincerely thank Krufinta Bun and Kyle Logue for helping with some of the statistical analyses. This study was supported by a grant from the NIH/NIDCR (1P01DE019759, A.W.; Project 4, R.J.J.).

References

Abu Bakar S, Hollox EJ, Armour JA. Allelic recombination between distinct genomic locations generates copy number diversity in human beta-defensins. Proceedings of the National Academy of Sciences USA. 2009; 106:853.

Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. Nature Reviews Genetics. 2002; 3:299.

Baroncelli S, Ricci E, Andreotti M, Guidotti G, Germano P, Marazzi MC, Vella S, Palombi L, De Rossi A, Giuliano M. Single-nucleotide polymorphisms in human beta-defensin-1 gene in Mozambican HIV-1-infected women and correlation with virologic parameters. AIDS. 2008; 22:1515. [PubMed: 18614876]

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21:263. [PubMed: 15297300]

Boniotto M, Hazbon MH, Jordan WJ, Lennon GP, Eskdale J, Alland D, Gallagher G. Novel hairpin-shaped primer assay to study the association of the -44 single-nucleotide polymorphism of the DEFB1 gene with early-onset periodontal disease. Clinical and Diagnostic Laboratory Immunology. 2004; 11:766. [PubMed: 15242954]

Bonnen PE, Wang PJ, Kimmel M, Chakraborty R, Nelson DL. Haplotype and linkage disequilibrium architecture for human cancer-associated genes. Genome Research. 2002; 12:1846. [PubMed: 12466288]

Braida L, Boniotto M, Pontillo A, Tovo PA, Amoroso A, Crovella S. A single-nucleotide polymorphism in the human beta-defensin 1 gene is associated with HIV-1 infection in Italian children. AIDS. 2004; 18:1598. [PubMed: 15238780]

Cagliani R, Fumagalli M, Riva S, Pozzoli U, Comi GP, Menozzi G, Bresolin N, Sironi M. The signature of long-standing balancing selection at the human defensin beta-1 promoter. Genome Biology. 2008; 9:R143. [PubMed: 18817538]

Carter JG, West SK, Painter S, Haynes RJ, Churchill AJ. beta-Defensin 1 haplotype associated with postoperative endophthalmitis. Acta Ophthalmologica. 2010; 88:786. [PubMed: 19706017]

Chen H, Xu Z, Peng L, Fang X, Yin X, Xu N, Cen P. Recent advances in the research and development of human defensins. Peptides. 2006; 27:931. [PubMed: 16226346]

Chen QX, Lv C, Huang LX, Cheng BL, Xie GH, Wu SJ, Fang XM. Genomic variations within DEFB1 are associated with the susceptibility to and the fatal outcome of severe sepsis in Chinese Han population. Genes & Immunity. 2007; 8:439. [PubMed: 17508030]

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. Nature. 2009; 464:704. [PubMed: 19812545]

de Smith AJ, Walters RG, Coin LJ, Steinfeld I, Yakhini Z, Sladek R, Froguel P, Blakemore AI. Small deletion variants have stable breakpoints commonly associated with alu elements. PLoS One. 2008; 3:e3104. [PubMed: 18769679]

Feng Z, Dubyak GR, Lederman MM, Weinberg A. Cutting edge: human beta defensin 3 - a novel antagonist of the HIV-1 coreceptor CXCR4. Journal of Immunology. 2006; 177:782.

Ferlin A, Ganz F, Pengo M, Selice R, Frigo AC, Foresta C. Association of testicular germ cell tumor with polymorphisms in estrogen receptor and steroid metabolism genes. Endocrine-Related Cancer. 2010; 17:17. [PubMed: 19776291]

Fode P, Jespersgaard C, Hardwick RJ, Bogle H, Theisen M, Dodoo D, et al. Determination of beta-defensin genomic copy number in different populations: a comparison of three methods. PLoS One. 2011a; 6:e16768. [PubMed: 21364933]

Fode P, Stegger M, Andersen PS. Human beta-defensin 3 (DEFB103) and its influence on Staphylococcus aureus nasal carriage. International Journal of Infectious Diseases. 2011b; 15:e388. [PubMed: 21393042]

Freguja R, Gianesin K, Del Bianco P, Malacrida S, Rampon O, Zanchetta M, Giaquinto C, De Rossi A. Polymorphisms of innate immunity genes influence disease progression in HIV-1 infected children. AIDS. 2012:765. [PubMed: 22269973]

Ganz T. Defensins: antimicrobial peptides of innate immunity. Nature Reviews Immunology. 2003; 3:710.

Groth M, Szafranski K, Taudien S, Huse K, Mueller O, Rosenstiel P, Nygren AO, Schreiber S, Birkenmeier G, Platzer M. High-resolution mapping of the 8p23.1 beta-defensin cluster reveals strictly concordant copy number variation of all genes. Human Mutation. 2008; 29:1247. [PubMed: 18470942]

Groth M, Wiegand C, Szafranski K, Huse K, Kramer M, Rosenstiel P, Schreiber S, Norgauer J, Platzer M. Both copy number and sequence variations affect expression of human DEFB4. Genes & Immunity. 2010; 11:458. [PubMed: 20445567]

Hollox EJ. Copy number variation of beta-defensins and relevance to disease. Cytogenetics and Genome Research. 2008; 123:148. [PubMed: 19287149]

Hollox EJ, Armour JA, Barber JC. Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster. American Journal of Human Genetics. 2003; 73:591. [PubMed: 12916016]

Hollox EJ, Barber JC, Brookes AJ, Armour JA. Defensins and the dynamic genome: what we can learn from structural variation at human chromosome band 8p23.1. Genome Research. 2008; 18:1686. [PubMed: 18974263]

Huse K, Taudien S, Groth M, Rosenstiel P, Szafranski K, Hiller M, et al. Genetic variants of the copy number polymorphic beta-defensin locus are associated with sporadic prostate cancer. Tumor Biology. 2008; 29:83. [PubMed: 18515986]

Ionita-Laza I, Rogers AJ, Lange C, Raby BA, Lee C. Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. Genomics. 2009; 93:22. [PubMed: 18822366]

Jurevic RJ, Bai M, Chadwick RB, White TC, Dale BA. Single-nucleotide polymorphisms (SNPs) in human beta-defensin 1: high-throughput SNP assays and association with Candida carriage in type I diabetics and nondiabetic controls. Journal of Clinical Microbiology. 2003; 41:90. [PubMed: 12517831]

Jurevic RJ, Chrisman P, Mancl L, Livingston R, Dale BA. Single-nucleotide polymorphisms and haplotype analysis in beta-defensin genes in different ethnic populations. Genetic Testing. 2002; 6:261. [PubMed: 12537649]

Kalus AA, Fredericks LP, Hacker BM, Dommisch H, Presland RB, Kimball JR, Dale BA. Association of a genetic polymorphism (-44 C/G SNP) in the human DEFB1 gene with expression and inducibility of multiple beta-defensins in gingival keratinocytes. BMC Oral Health. 2009; 9:21. [PubMed: 19712472]

Kim E, Lee JE, Namkung JH, Kim PS, Kim S, Shin ES, Cho EY, Yang JM. Single nucleotide polymorphisms and the haplotype in the DEFB1 gene are associated with atopic dermatitis in a Korean population. Journal of Dermatological Science. 2009; 54:25. [PubMed: 19135873]

Klotman ME, Chang TL. Defensins in innate antiviral immunity. Nature Reviews Immunology. 2006; 6:447.

Leung TF, Li CY, Liu EK, Tang NL, Chan IH, Yung E, Wong GW, Lam CW. Asthma and atopy are associated with DEFB1 polymorphisms in Chinese children. Genes & Immunity. 2006; 7:59. [PubMed: 16435024]

Linzmeier RM, Ganz T. Human defensin gene copy number polymorphisms: comprehensive analysis of independent variation in alpha- and beta-defensin regions at 8p22-p23. Genomics. 2005; 86:423. [PubMed: 16039093]

Mehlotra RK, Bockarie MJ, Zimmerman PA. Prevalence of UGT1A9 and UGT2B7 nonsynonymous single nucleotide polymorphisms in West African, Papua New Guinean, and North American populations. European Journal of Clinical Pharmacology. 2007; 63:1. [PubMed: 17115150]

Mehlotra RK, Ziats MN, Bockarie MJ, Zimmerman PA. Prevalence of CYP2B6 alleles in malaria-endemic populations of West Africa and Papua New Guinea. European Journal of Clinical Pharmacology. 2006; 62:267. [PubMed: 16506047]

Menard V, Eap O, Harvey M, Guillemette C, Levesque E. Copy-number variations (CNVs) of the human sex steroid metabolizing genes UGT2B17 and UGT2B28 and their associations with a UGT2B15 functional polymorphism. Human Mutation. 2009; 30:1310. [PubMed: 19572376]

Milanese M, Segat L, Pontillo A, Arraes LC, de Lima Filho JL, Crovella S. DEFB1 gene polymorphisms and increased risk of HIV-1 infection in Brazilian children. AIDS. 2006; 20:1673. [PubMed: 16868452]

Naslavsky MS, Crovella S, Lima Filho JL, Rocha CR. The sound of silence: human beta-defensin-1 gene untranslated SNPs change the predicted mRNA secondary structure in a length-dependent manner. Immunology Letters. 2010; 129:53. [PubMed: 20060856]

Naslavsky MS, Rocha CR, Lima Filho JL, Crovella S. Predicting alternative candidates as binding sites to DEFB1 668 (-44) SNP: a long way from statistical association with multifactorial diseases. Infection, Genetics and Evolution. 2009; 9:1129.

Ozturk A, Famili P, Vieira AR. The antimicrobial peptide DEFB1 is associated with caries. Journal of Dental Research. 2010; 89:631. [PubMed: 20371866]

Prado-Montes de Oca E. Human beta-defensin 1: a restless warrior against allergies, infections and cancer. International Journal of Biochemistry and Cell Biology. 2010; 42:800. [PubMed: 20100591]

Prado-Montes de Oca E, Rangel-Villalobos H, Gallegos-Arreola MP, Sandoval L, Figuera LE. SNPs in human beta-defensin 1 gene (DEFB1): frequencies in a Mexican population and new PCR-RFLPs assays. International Journal of Immunogenetics. 2006; 33:339. [PubMed: 16984277]

Quinones-Mateu ME, Lederman MM, Feng Z, Chakraborty B, Weber J, Rangel HR, et al. Human epithelial beta-defensins 2 and 3 inhibit HIV-1 replication. AIDS. 2003; 17:F39. [PubMed: 14571200]

Ricci E, Malacrida S, Zanchetta M, Montagna M, Giaquinto C, De Rossi A. Role of beta-defensin-1 polymorphisms in mother-to-child transmission of HIV-1. Journal of Acquired Immune Deficiency Syndrome. 2009; 51:13.

Schaefer AS, Richter GM, Nothnagel M, Laine ML, Ruhling A, Schafer C, et al. A 3' UTR transition within DEFB1 is associated with chronic and aggressive periodontitis. Genes & Immunity. 2010; 11:45. [PubMed: 19829306]

Segat L, Brandao LA, Guimaraes RL, Crovella S. Are defensin beta 1 gene polymorphisms associated with HIV infection and virus replication? AIDS. 2009; 23:647. [PubMed: 19516112]

Segat L, Milanese M, Boniotto M, Crovella S, Bernardon M, Costantini M, Alberico S. DEFB-1 genetic polymorphism screening in HIV-1 positive pregnant women and their children. Journal of Maternal-Fetal and Neonatal Medicine. 2006; 19:13. [PubMed: 16572527]

Shepherd JC, Jacobson LP, Qiao W, Jamieson BD, Phair JP, Piazza P, Quinn TC, Margolick JB. Emergence and persistence of CXCR4-tropic HIV-1 in a population of men from the multicenter AIDS cohort study. Journal of Infectious Diseases. 2008; 198:1104. [PubMed: 18783316]

Sun L, Finnegan CM, Kish-Catalone T, Blumenthal R, Garzino-Demo P, La Terra Maggiore GM, et al. Human beta-defensins suppress human immunodeficiency virus infection: potential role in mucosal protection. Journal of Virology. 2005; 79:14318. [PubMed: 16254366]

Taudien S, Groth M, Huse K, Petzold A, Szafranski K, Hampe J, Rosenstiel P, Schreiber S, Platzer M. Haplotyping and copy number estimation of the highly polymorphic human beta-defensin locus on 8p23 by 454 amplicon sequencing. BMC Genomics. 2010; 11:252. [PubMed: 20403190]

Zapata W, Rodriguez B, Weber J, Estrada H, Quinones-Mateu ME, Zimermman PA, Lederman MM, Rugeles MT. Increased levels of human beta-defensins mRNA in sexually HIV-1 exposed but uninfected individuals. Current HIV Research. 2008; 6:531. [PubMed: 18991618]

Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. Annual Review of Genomics and Human Genetics. 2009; 10:451.

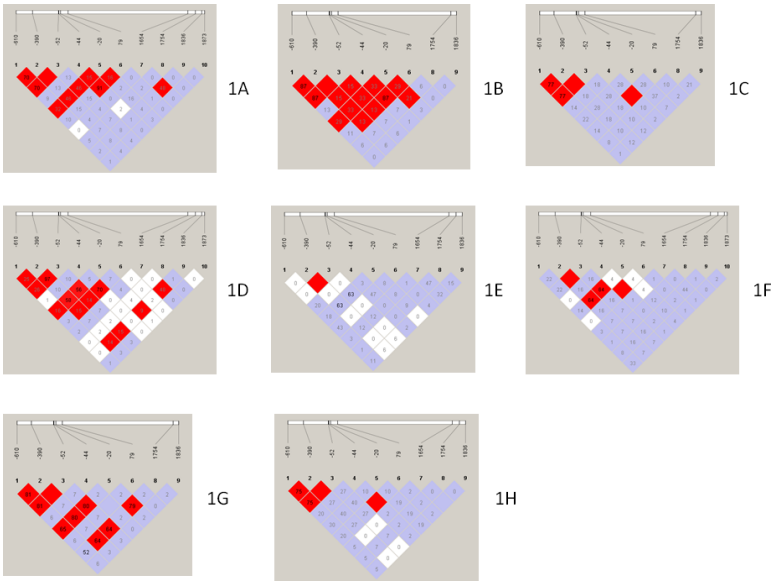

Fig. 1A-1H.

Heat maps of pairwise LD measurements for the 10 SNPs in *DEFB1*. 1A, CFAR white patients; 1B, MACS white, non-Hispanic patients; 1C, NE; 1D, CFAR black patients; 1E, AfSS; 1F, Af-Am; 1G, CHN; and 1H, MEX. Color scheme: bright red, D' = 1.0 and LOD score ≥ 2.0; blue, D' = 1.0 and LOD score < 2.0; and white, D' < 1.0 and LOD score < 2.0. Numbers represent r^2 values. r^2 values of 1.0 are not shown (the box is empty).

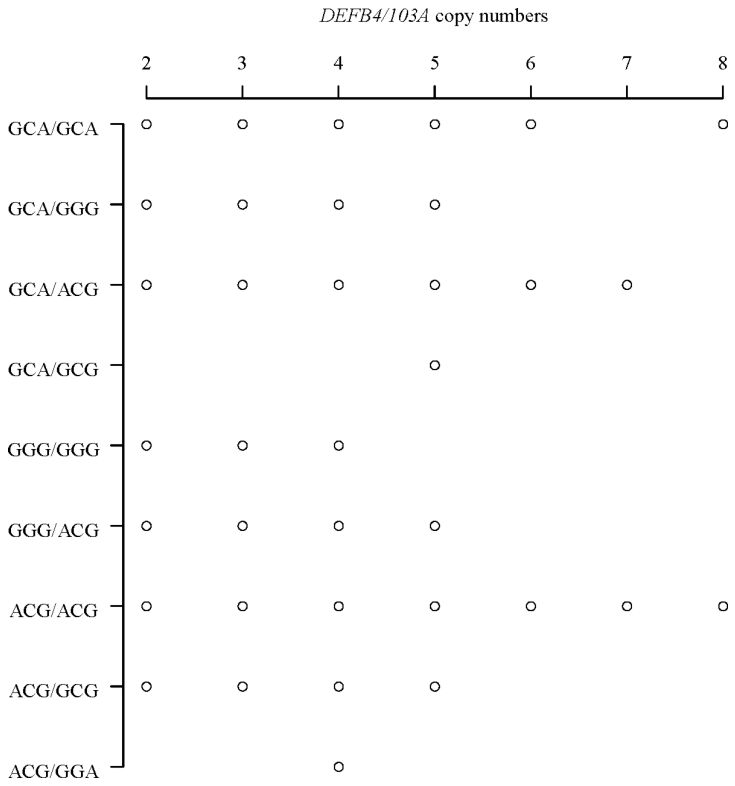

Fig. 2.

Dot plot showing distribution of *DEFB4/103A* copy numbers among *DEFB1* 3-SNP diplotypes.

Each circle represents number of individuals, which are provided in Table 3.

Table 1

DEFB1 SNP-allele frequencies in various groups of samples

Region	SNP	*f* (Allele)	Group of samples†							
			CFAR		MACS	Coriell				
			White(n = 37)	Black (n = 65)	(n = 39)	NE (n = 10)	AfSS (n = 9)	Af-Am (n = 10)	CHN (n = 10)	MEX (n = 7)
Exon-1	-610G/A	*f*(G)	0.69	0.77	0.74	0.75	0.78	0.75	0.45	0.57
		f(A)	0.31	0.23	0.26	0.25	0.22	0.25	0.55	0.43
	-390T/A	*f*(T)	0.61	0.46	0.72	0.70	0.50	0.40	0.40	0.50
		f(A)	0.39	0.54	0.28	0.30	0.50	0.60	0.60	0.50
	-52G/A	*f*(G)	0.61	0.47	0.72	0.70	0.50	0.40	0.40	0.50
		f(A)	0.39	0.53	0.28	0.30	0.50	0.60	0.60	0.50
	-44C/G	*f*(C)	0.82	0.92	0.72	0.70	0.94	0.90	0.95	0.79
		f(G)	0.18	0.08	0.28	0.30	0.06	0.10	0.05	0.21
	-20G/A	*f*(G)	0.58	0.65	0.54	0.60	0.61	0.70	0.65	0.71
		f(A)	0.42	0.35	0.46	0.40	0.39	0.30	0.35	0.29
	79T/A	*f*(T)	0.81	0.88	0.74	0.70	0.89	0.90	0.95	0.79
		f(A)	0.19	0.12	0.26	0.30	0.11	0.10	0.05	0.21
Exon-2	1654G>A	*f*(G)	0.97	0.92	1.00	1.00	0.89	0.90	1.00	0.93
		f(A)	0.03	0.08	0	0	0.11	0.10	0	0.07
	1754G/A	*f*(G)	0.89	0.95	0.85	0.80	0.94	0.95	0.70	0.93
		f(A)	0.11	0.05	0.15	0.20	0.06	0.05	0.30	0.07
	1836A/G	*f*(A)	0.95	0.85	0.97	0.95	0.72	0.80	0.95	0.93
		f(G)	0.05	0.15	0.03	0.05	0.28	0.20	0.05	0.07
	1873T/C	*f*(T)	0.97	0.96	1.00	1.00	1.00	0.90	1.00	1.00
		f(C)	0.03	0.04	0	0	0	0.10	0	0

CFAR patients of white and black races were self-identified.

MACS patients were white, non-Hispanic.

†NE = Northern European, AfS S = African sub- Sahara, Af-Am = African-American, CHN = Chinese, MEX = Mexican.

Table 2a

DEFB1 10-SNP haplotype frequencies in various groups of samples

Haplotype	Group of samples†							
	CFAR		MACS	Coriell				
	White (n = 37)	Black (n = 65)	(n = 39)	NE (n = 10)	AfSS (n = 9)	Af-Am (n = 10)	CHN (n = 10)	MEX (n = 7)
GTGCATGGAT	0.31	0.30	0.28	0.20	0.28	0.25	0.05	0.21
GTGCATGAAT	0.11	0.05	0.15	0.20	0.06	0.05	0.30	0.07
GTGGGAGGAT	0.18	0.08	0.26	0.30	0.06	0.10	0.05	0.21
AAACGTGGAT	0.31	0.23	0.26	0.25	0.22	0.25	0.55	0.43
GAACGTAGGT	0.03	0.08	0	0	0.11	0.10	0	0.07
GAACGTGGGT	0.03	0.06	0.03	0.05	0.11	0.10	0.05	0
GAACGTGGAT	0	0.12	0	0	0.06	0.05	0	0
GAACGTGGAC	0.03	0.04	0	0	0	0.10	0	0
GTGCGAGGAT	0.01	0.03	0	0	0.06	0	0	0
GTGCATGGGT	0	0	0	0	0.06	0	0	0
GAGCGTGGGT	0	0.01	0	0	0	0	0	0
GAACATGGAT	0	0.01	0	0	0	0	0	0
GTGGATGGAT	0	0	0.03	0	0	0	0	0

†As described in Table 1.

Table 2b

DEFB1 3-SNP haplotype frequencies in various groups of samples

Haplotype*	Group of samples†							
	CFAR		MACS	Coriell				
	White (n = 37)	Black (n = 65)	(n = 39)	NE (n = 10)	AfSS (n = 9)	Af-Am (n = 10)	CHN (n = 10)	MEX (n = 7)
GCA	0.42	0.35	0.44	0.40	0.39	0.30	0.35	0.29
GGG	0.18	0.08	0.26	0.30	0.06	0.10	0.05	0.21
ACG	0.39	0.52	0.28	0.30	0.50	0.60	0.60	0.50
GCG	0.01	0.04	0	0	0.06	0	0	0
ACA	0	0.01	0	0	0	0	0	0
GGA	0	0	0.03	0	0	0	0	0

†As described in Table 1.

*-52G/A_-44C/G_-20G/A

Table 3

Distribution of *DEFB4/103A* CNV among *DEFB1* 3-SNP diplotypes

Diplotype	Copy number							Total
	2	3	4	5	6	7	8	
	Number of individuals							
GCA/GCA	6	9	10	4	1	0	1	31
GCA/GGG	3	5	7	4	0	0	0	19
GCA/ACG	14	17	22	4	1	2	0	60
GCA/GCG	0	0	0	1	0	0	0	1
GGG/GGG	2	2	2	0	0	0	0	6
GGG/ACG	6	10	6	4	0	0	0	26
ACG/ACG	3	9	11	9	1	1	2	36
ACG/GCG	1	2	1	2	0	0	0	6
ACG/GGA	0	0	2	0	0	0	0	2
Total	35	54	61	28	3	3	3	187