V. DISCUSSION

The MOMP gene of *C. trachomatis* is the major gene specifying antigens. The type specific antigen coding sequences are clustered in the VD of the MOMP gene. The encoded antigens are located on the surface of the organisms and are the principal targets for neutralization by specific antibody, thus creating protection from re-infection. Differentiation of the *C. trachomatis* serotypes is necessary for the epidemiological studies in order to establish the regional serotype prevalence and prevention of infections. Serotype classification was originally based on the immunological reaction with type specific polyclonal and monoclonal antibodies, including micro-immunofluorescent and dot-ELISA techniques. However, these methods are now limited to the commercial availability of 15 type-specific monoclonal antibodies and the growing ability of *C. trachomatis* strains in a tissue culture. The molecular typing of *C. trachomatis* was carried out by using the RFLP analysis of the MOMP gene, which covered all VDs. However, they usually resulted in a complex banding pattern, which made them difficult to identify. The RFLP analysis performed in this study was based on the amplification of the VD4 region, the largest VD of the MOMP gene, and its digestion with 4 restriction endonucleases. The technique was successfully identified in at least 10 of the present *C. trachomatis* serotypes, which included all predominant genital serotypes circulating in different geographies.

Among 50 *C. trachomatis* positive samples, it was observed that the most prevalent serotype was D/Da/L1 (32%) followed by serotype F (18%). These serotypes accounted for 50% of the total isolates. However, the RFLP technique could not differentiate between some closely related *C. trachomatis* serotypes that had identical nucleotide sequences in the recognition sites of the appropriated restriction

endonucleases. Therefore, serotypes B/Ba, D/Da/L1 and H/J/Ia could not be differentiated in this study. After the nucleotide sequences of the VD4-MOMP gene were analysed, all 16 serotypes, originally identified as D/Da/L1 by the RFLP, were later classified as serotype D. However, their sequences were different from the earlier reported prototype, but similar to the D variant observed by Poole *et al.*(15) and Sayada *et al.*(44). The D variant genotypes were now increasingly observed along with the prototype, with 64% of the D genotype reported from France being the sequence variants (44). Similarly, 1 from 5 serotypes H/J/Ia was later classified as serotype J and the rest were serotype H or Ia after nucleotide sequencing analysis. However, the closely related serotype B and Ba, and serotype H and Ia could not be differentiated from each other since they had identical VD4 sequences. Serotype D comprising the five unidentified samples classified by RFLP accounted for 42 %. When placed together with the second most prevalent *C. trachomatis* serotype F, they accounted for 60 % of the isolates. The serotype distribution reported in this study was different from those previously shown by others. Serotype E was observed predominately by many investigators from different countries, e.g. Canada, Netherlands, France and Kenya (2,34,40,47). In those studies, however, serotype D was still found in high prevalence among the genital serotypes followed the serotype E or sometimes F (37,44,48). Nevertheless, a recent report from the National Institute of Health of Japan also revealed the highest prevalence of serotype D among the genital serotype (49). It was possible that the high prevalence of serotype D reported in this study was unique for this area.

The MOMP gene of *C. trachomatis* frequently exhibited DNA sequence variation, mainly in the VD regions. The sequence variation usually resulted in amino acid substitution, which led to some antigenic changes. It was reported that the MOMP variant could escape neutralization by both serotype specific monoclonal antibodies and human immune sera (50). At present, a large number of *C. trachomatis*

serovariants have been identified from clinical isolates by using monoclonal and polyclonal antibodies. Nucleotide sequence polymorphism in an antigen specifying gene (MOMP gene) of *C. trachomatis* could be most apparent among isolates originating from a high risk group of STD, presumably in response to the immune selection (40).

In this study, the VD4 nucleotide sequence was determined in 50 *C. trachomatis* positive samples collected from the STD high risk group. Twenty four (48%) had a sequence similar to the prototype while 26 (52%) were the sequence variants. None of the sequence variant isolates were detected by RFLP analysis. The high number of *C. trachomatis* variants observed in this study were similar to those reported from Kenya, by Brunham *et al.*(40) where 63% were variant sequences. Yang *et al.*(34) found only 31% with VD-MOMP variation of *C. trachomatis* in the STD low risk group. The individuals in a high risk group had likely experienced repeated infection, which might have necessitated a boost to antibody titers at the mucosal infection site, ultimately selecting a variant in the population of prevalent nonmutants (32,51).

Among the 26 nucleotide sequence variants, a single nucleotide substitution was observed in the VD4. It was found in 2 distinct serotypes, D and K. Twenty-one were serotype D variant, which had adenine (A) instead of guanine (G) at position 979 (Fig.17). These variants had a sequence identical to D/B-185 and D^{-}, as previously determined by Sayada *et al.*(44) and Poole *et al.*(15), respectively. In addition, this transition resulted in an amino acid substitution of threonine for alanine and caused the loss of reactivity with the subgroup-specific monoclonal antibody, BB-11 (45). In the meantime, 5 samples that were identified as K variants had guanine (G) instead of adenine (A) at position 973. This also resulted in an amino acid transition of alanine for threonine (Fig. 20). The K variants observed had identical sequences to prototype strain K/UW-31/Cx that was resequenced by Poole *et al.*(15) and Stohard *et*

al.(46). It was also similar to the K variant identified by Brunham *et al.*(40) and Yang *et al.*(34). However, all these K variants differed from the K prototype (K/UW-31/Cx) originally sequenced by Yuan *et al.*(16). It was suggested later that the K strain sequence reported earlier by Yuan *et al.* might actually be the variant one.

The genetic mechanisms for diversification of the MOMP locus are unclear. Stephens *et al.*(17) suggested that the accumulation of a clustered base substitution, deletion and insertion of distantly related serotypes, and genetic recombination for closely related serotypes probably accounted for MOMP diversification. Detailed analysis of the nucleotide changes in the VD4-MOMP sequence observed in this study suggests that all of the 26 variant serotypes were mutational drift to transition-type with purine transition (21 were G⟶A and 5 were A⟶G). Similarly, Brunham *et al.*(40) showed that 13 of 17 nucleotide substitutions were a transitional type in which purine transition predominated. It was suggested that *Tag* DNA polymerase could also produce base substitutions that were predominately a transitional type, and could be the origin of some of these mutations (52,53,54). However, the occurrence of phenotypic selection cannot be ruled out, since identical mutations were detected in different geographies and epidemiologically distinct infections on multiple occasions as well as in the repeated cycle sequencing of MOMP variants that yielded identical sequences.

It has been accepted that the selection host immune pressure plays an important role in the adaptability of organisms, perhaps in an effort to evade the host immune surveillance. This is supported by observations in this study and others, where most of the nucleotide substitutions resulted in amino acid substitution. However, the occurrence of random point mutations with the accumulation of amino acid changes may be structurally and functionally tolerant. It is not known at present what level of antigenic structure integrity or degree of antigenic drift is important for immune recognition.

The data in this study provide information for prospective studies to evaluate the rate of MOMP gene mutations and the significance of these mutations in relation to the antigenic diversity, severity of diseases and probably the emergence of antimicrobial resistance. The identification of antigenic variation of the *C. trachomatis* circulated in a certain environment are of great value for vaccine development.