Journal of General Virology (2000), **81**, 2959–2968. Printed in Great Britain

Molecular variants of human papillomavirus types 16 and 18 preferentially associated with cervical neoplasia

Luisa L. Villa,[1] Laura Sichero,[1,5] Paula Rahal,[1,6]† Otavia Caballero,[1,5] Alex Ferenczy,[2] Tom Rohan[3]‡ and Eduardo L. Franco[4]

[1] Department of Virology, Ludwig Institute for Cancer Research, R. Prof. Antonio Prudente 109, 4 Andar, 01509-010 São Paulo, SP, Brazil

[2] Department of Pathology, Jewish General Hospital, 3755 Cote-Ste-Catherine Road, Montreal, Canada H3T 1E2

[3] Public Health Sciences Department, University of Toronto, 12 Queen's Park Crescent West, Toronto, Ontario, Canada M5S 1A8

[4] Departments of Epidemiology and Oncology, McGill University, 546 Pine Avenue West, Montreal, Canada H2W 1S6

[5,6] Department of Biochemistry, Chemistry Institute[5] and Department of Microbiology[6], University of São Paulo, Cidade Universitária, São Paulo, CP20.780, Brazil

In order to determine geographically related intratypic variation in human papillomavirus (HPV) type 16 and 18 isolates that could be associated with lesion development, data were analysed from an ongoing cohort study of the natural course of infection of HPVs and cervical neoplasia. Testing for HPVs was carried out by PCR and molecular variants of these HPVs were characterized by sequence analysis of the long control region and by dot blot hybridization of the E6 and L1 genes. Tests for HPV were done in multiple first-year specimens from 1690 women enrolled in a cancer screening program from 1993 to 1997. Subjects were followed-up by cytology and cervicography for detection of cervical lesions. Seven variants of HPV-16 and four of HPV-18 were detected in one or more specimens from 65 subjects. The same variant was found in specimens taken on different visits from each case of persistent infection. Overall, non-European variants tended to persist more frequently [odds ratio (OR) = 4·5; 95% confidence interval (CI), 1·6–12·4] than European (E) variants (OR = 2·5; 95% CI, 1·3–4·9), relative to the risk of persistence for non-oncogenic HPVs. In addition, non-E variants were more strongly associated with risk of both prevalent (age- and race-adjusted OR = 172·2; 95% CI, 47·1–630·1) and incident [relative risk (RR) = 22·5; 95% CI, 6·0–83·9] high-grade lesions than E variants (prevalent lesions OR = 46·3; 95% CI, 15·5–138·0 and incident lesons RR = 6·1; 95% CI, 1·3–27·4), relative to the risk for HPV-negative women. Although consistent, the latter differences were not statistically significant. If confirmed in other populations, measurement of intratypic variation of HPV-16 and -18 has the potential to serve as an ancillary tool in cervical cancer screening.

Introduction

Studies using DNA sequencing analyses of hundreds of isolates of human papillomaviruses (HPVs) from clinical specimens and cervical cancer-derived cell lines have shown considerable intratypic diversity for HPV-16 (Chan *et al.*, 1992; Eschle *et al.*, 1992; L. Ho *et al.*, 1991, 1993; Icenogle *et al.*, 1991; Xi *et al.*, 1993, 1995; Yamada *et al.*, 1995), HPV-18 and -45 (Ong *et al.*, 1993), and HPV-6 and -11 (Heinzel *et al.*, 1995). These molecular variants or lineages differ in nucleotide sequence by no more than 2% in the coding regions and 5% in the non-coding regions of the viral genome with respect to the prototype isolates for that strain (Bernard *et al.*, 1994*a*).

Through nucleotide sequence comparisons, it has been possible to reconstruct the spread of these viruses in human populations (Bernard *et al.*, 1994*a*). The ancient spread of papillomaviruses and their low rates of evolution suggest that

Author for correspondence: Luisa Villa.
Fax +55 11 270 7001. e-mail llvilla@node1.com.br

† **Present address:** Department of Biology, Ibilce, Universidade Estadual Paulista, R. Cristóvão Colombo 2265, São José do Rio Preto, São Paulo 15054-000, Brazil.

‡ **Present address:** Department of Epidemiology and Social Medicine, Albert Einstein College of Medicine, Bronx, New York, NY, USA.

HPVs have co-evolved with their natural hosts over a period of several million years. Data obtained from 25 different geographical regions in the world indicate that HPV-16 has evolved along five major branches, two being present mainly in Africa, two in Asia, and one in Europe and India (L. Ho *et al.*, 1993; Ong *et al.*, 1993). These studies revealed that colonization of the New World by Europeans and Africans is reflected in the distribution of HPV-16 variants in the American continent. It also became clear that these variants represent genuine HPV genomes since they have been found in different individuals from different regions around the world. Furthermore, variation studies performed on the long control region (LCR), E6, L1 and L2 genes of HPV-16 indicate that recombination between variants is either rare or non-existent (L. Ho *et al.*, 1993; Yamada *et al.*, 1995, 1997).

Few nucleotide differences found in HPV variants correspond to changes in amino acids. Nevertheless, it is of great interest to define alterations that may interfere with the functional or antigenic properties of specific viral proteins. For instance, variant 114K of HPV-16 assembles into virus-like particles (VLP) in a heterologous expression system, whereas the reference (prototype) clone of HPV-16 does not have the same property (Kirnbauer *et al.*, 1992, 1993). This difference has been attributed to a single amino acid change at residue 202 (Asp to His) of the L1 gene. Other changes in the L1/L2 regions of the HPV genomes may be important for discriminating between the infectious potential of different variants as well as in defining epitopes relevant to vaccine design. It has also been reported that a specific variation in the HPV-16 E6 protein, isolated from cervical cancers of HLA-B7 individuals, interferes with the T-cell cytotoxic immune response (Ellis *et al.*, 1995).

In studies examining the course of HPV infection and the role of the virus in the genesis of cervical cancer, issues such as polyclonality of lesions and frequency of particular HPV variants may be critical to our understanding of virus infectivity and pathogenicity (Zehbe & Tommasino, 1999). Nucleotide sequence variation has been used as an important tool for epidemiological studies of virus transmission and persistence (Franco *et al.*, 1994). The utility of this approach was indicated in a study of the sexual transmission of HPV-16 (G. Ho *et al.*, 1993). In a cohort study of young women, Xi *et al.* (1995) found 16 different HPV-16 variants, one of which persisted over time, while the other variants were transiently detected. In addition, intratypic variation of HPV-16 has also been shown to be an important predictor of progression to clinically relevant cervical lesions (Xi *et al.*, 1997).

In this report, we present results from an ongoing study of the natural course of HPV infection and cervical neoplasia in Brazil showing the association between variants as defined by geographical relatedness and characteristics of the infection and lesion development. Unlike most previous studies, we have analysed variants of both HPV-16 and -18 and assessed their correlation with virus persistence and virus burden as well as with risk of prevalent and incident pre-invasive lesions.

Methods

■ **Study cohort.** Details of the Ludwig–McGill cohort study have been presented elsewhere (Franco *et al.*, 1999*a*, *b*). Since 1993, women enrolled in a cancer screening program catering for low-income families in São Paulo, Brazil have been followed-up in scheduled returns every 4 months in the first year post-enrolment (0, 4, 8 and 12 months) and once every 6 months thereafter. In each of these visits, a questionnaire-based interview was carried out. Subjects also had a cervical specimen taken for Pap cytology and HPV testing and a blood sample was drawn for serologic testing for HPV antibodies. A cervicography was performed once in the first year and every 2 years thereafter. This investigation was approved by ethical review boards of the authors' institutions.

■ **Cervical cell specimens.** An Accelon biosampler (Medscand) was used to collect a sample of ectocervical and endocervical cells at each of the visits. After the smear was prepared on a glass slide and fixed in 95% ethanol, the exfoliated cells remaining in the sampler were immersed in a tube containing Tris–EDTA buffer pH 7·4, kept at 4 °C at the clinic for at most 5 days, and then frozen until testing. Cervical smears were sent to Montreal and were evaluated by the Bethesda system. Cytology readings were done blindly (A.F.).

■ **HPV DNA detection and typing.** Cervical specimen DNA was extracted and purified following standard techniques. In brief, cells were digested with 100 μg/ml proteinase K for 3–18 h at 55 °C and the DNA was purified by spin-column chromatography. Specimens were tested for the presence of HPV DNA by a previously described PCR protocol amplifying a highly conserved 450 bp segment in the L1 viral gene (flanked by primers MY09/11). Typing of the amplified products was performed by hybridization with individual oligonucleotide probes specific for 27 HPV genital types (Bauer *et al.*, 1991; Hildesheim *et al.*, 1994). Amplified products hybridizing to the generic probe but not to any of the type-specific probes were further tested by restriction fragment length polymorphism analysis (Bernard *et al.*, 1994*b*). This extends the range of identifiable HPVs to over 40 genital types. In order to check the integrity of the host DNA material extracted from the specimens, assays also included an additional set of primers (GH20 and PC04), which amplify a 268 bp region of the human β-globin gene (Saiki *et al.*, 1988). In most analyses, infections with HPV-16 and -18 were considered together and the remaining HPV types were grouped according to their oncogenic potential. We followed the classification scheme of Liaw *et al.* (1999), which grouped HPV-31, -33, -35, -39, -45, -51, -52, -56, -58, -59 and -68 as oncogenic types and considered all other HPV types (except HPV-16 and -18) as non-oncogenic. Unknown types were included in the latter group.

■ **Determination of virus load.** We measured virus load in the cervical specimens by a quantitative PCR protocol as previously described (Caballero *et al.*, 1995). This test consists of employing low stringency conditions to allow co-amplification of the specific HPV DNA fragment along with certain DNA sequences from the human genome present in the starting mixture. After amplification, the ratio of the specific HPV band to that of the host genome internal band was measured by densitometry and followed by quantification via interpolation from a standard curve; the standard curve was prepared with pre-selected amounts of reference HPV plasmid added to a constant background of normal human DNA (Caballero *et al.*, 1995; Villa *et al.*, 1996).

■ **Molecular variant analysis.** We tested for molecular variants of HPV-16 and -18 using a PCR sequencing method (Ho *et al.*, 1991; Ong *et al.*, 1993). A 364 bp segment (nt 7478–7841) or a 321 bp segment (nt 7464–7825) within the HPV-16 or HPV-18 LCR, respectively, was amplified as previously described (Franco *et al.*, 1994; Villa *et al.*, 1997). The PCR products were cloned using the SureClone ligation kit (Pharmacia) and transformed into *E. coli*, strain XL1-Blue. Recombinant plasmid DNAs were isolated from positive clones and sequenced by automated DNA sequencing in an Alfie Express sequencing machine (Pharmacia). To prevent errors due to possible misincorporation of bases by *Taq* DNA polymerase a minimum of five clones was sequenced for each isolate. This also allowed the identification of multiple co-existing variants. Variants were grouped within branches of geographical relatedness based on the nucleotide difference profiles (L. Ho *et al.*, 1993; Ong *et al.*, 1993). These variant branches are designated as E (European), As (Asian), AA (Asian–American) and Af (African).

Most specimens were also tested by a recently described hybridization procedure that allows the detection of all previously described variant groups of HPV-16 (Wheeler *et al.*, 1997). This approach focuses on the joint variability in the E6 and L1 genes and allows the classification of these variants into two additional subclasses within the E and AA branches (Yamada *et al.*, 1995).

■ **Statistical analysis.** We analysed persistence of HPV infection throughout the first year of follow-up (up to four visits) of HPV-16 and -18 variants and other grouped types using logistic regression. Persistent infection was defined as a case with two or more positive visits for the same HPV type, consecutive or not. We computed two separate logistic models including co-variant adjustment for age and race to estimate odds ratios (OR) and their respective 95% confidence intervals (95% CI) for persistence: one in which variants of both HPV-16 and -18 were classified together and another in which only variants of HPV-16 were grouped by geographical relatedness.

We used Student's *t*-test to compare the mean virus burden in HPV-positive cervical specimens by grouped HPV-16 and -18 variants and other types based on oncogenicity. To obtain normally distributed groups for comparison, the logs were taken of individual copy numbers per cell obtained by quantitative PCR. We contrasted the mean in each group separately against those of non-oncogenic types and against those of E branch variants of HPV-16 and -18. We computed geometrical mean copy numbers and respective 95% CI per group by pooling the results from all four visits using a weighted method to combine the variances in order to account for non-independence of observations.

Risk of cervical lesions according to grouped HPV-16 and -18 variants and other types based on oncogenicity was analysed in two ways. First, we analysed the association between HPV-infection status as cumulative exposure during all first year visits and any detection of squamous intraepithelial lesions (SILs) during the same period, but not during subsequent follow-up visits. Logistic regression models were computed to estimate OR (and 95% CI) of SILs. Second, we analysed the risk of post-enrolment occurrence of SILs as an incident finding by considering only women free of SILs at the enrolment visit in relation to HPV-infection status at enrolment. Cox proportional hazards regression models were computed to estimate relative risks (RR) and 95% CI of incident SILs by HPV-infection status. Three non-mutually exclusive outcomes were used in both logistic and Cox analyses: any grade SIL, high-grade SIL (HSIL) and an augmented HSIL definition that included detection by cervicography in addition to cytology. As in the analysis of persistence above, estimates of effect (OR and RR) were computed in separate models: one in which variants of both HPV-16 and -18 were classified together and another in which only variants of HPV-16 were grouped by geographical relatedness. OR and RR values were adjusted for age and ethnicity.

Results

The 1690 participants included in this analysis represented mostly those women admitted early in the study (a total of 2528 women were enrolled in the cohort with a 70% response rate). At the time of the analysis only cervical specimens from the first year post-enrolment had been tested for HPV. A total of 6077 tests for HPV, an average of 3·6 tests per subject, were performed. Of these tests, 173 were non-informative because of failure to amplify the β-globin gene (2·8%). The median age of participants was 33 years (mean = 33; 1st and 3rd quartiles, 26 and 39 years, respectively). Data were available until November 1999 for lesion surveillance via cytology and cervicography, for a total of 69 190 woman-months of follow-up (average of 40·9 months per subject).

A total of 443 women had one or more positive results for HPV (26·2%), 92 of which were positive for variants of HPV-16 or -18, 175 were positive for other oncogenic types and 176 were positive exclusively for non-oncogenic HPVs. Using LCR sequence analysis, we have characterized 97 isolates of HPV-16 from 54 subjects, and 25 HPV-18-positive samples from 12 subjects (one of them with both HPV-16 and -18). We identified seven different molecular variants of HPV-16 belonging to all four branches of geographical and phylogenetic relatedness. The prototype of HPV-16, which is classified in the E branch, was the most frequent isolate (29 of 54 subjects, 54%), followed by variant B-2 from the AA branch (12 of 54 subjects, 22%). Two subjects harboured novel variants with nucleotide patterns comparable to the latter variant. Regarding HPV-18 isolates, we found four variants that were classifiable only in the E and Af branches. The B18-2 variant (E branch) was the most common finding (8 of 12 subjects, 67%). The one subject infected with both HPV-16 and -18 for whom variant analysis was completed harboured variants that were classifiable in the E branch for both of these types. There was only one subject with HPV-18 classifiable in the Af branch (variant T18-7), and no subjects were infected with the prototype strain. This precluded a separate presentation of results for this type in the analysis of persistence and lesion risk.

The above analysis based on the LCR region was extended by a search for mutations in the E6 and L1 genes (Fig. 1). This analysis allowed us to corroborate the variant classification based on LCR (data not shown) and permitted the classification of the two novel HPV-16 isolates into the North-American-1 sub-class of the AA branch, given their pattern of nucleotide substitutions in the E6 and L1 genes.

Cases of persistent infection with HPV-16 or -18 always had the same molecular variant in all positive specimens, irrespective of the number of clones that were tested. We also searched for multiple infections of different variants of the same HPV type in all isolates. We did not find a single case

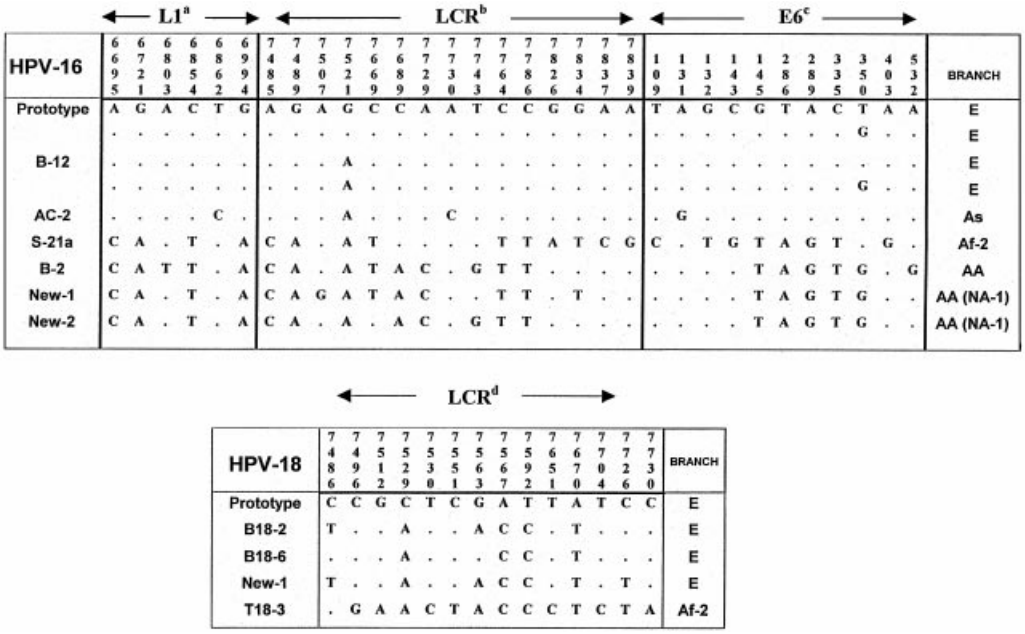

Fig. 1. Nucleotide sequence variation among HPV-16 and -18 isolates. L1, LCR and E6 nucleotide positions where variations were detected are written vertically across the top. Positions at which no variation was found relative to the prototype are marked with periods, whereas a letter indicates a variant nucleotide at this position. Variants based on the analysis of the LCR are defined as in (b) L. Ho *et al.* (1993) and Wheeler *et al.* (1997) for HPV-16 at nt 7477–7842 and (d) Ong *et al.* (1993) for HPV-18 at nt 7484–7805. The variability in genes (a) L1 at nt 6598–7021 and (c) E6 at nt 56–640 were analysed according to Wheeler *et al.* (1997).

Table 1. Age- and race-adjusted OR (and respective 95% CI) of persistent HPV infection during the first year visits

OR and CI values are from logistic regression analyses of the probability of having two or more positive visits for the same HPV type. HPV-18 infections were included among other oncogenic types when variants were considered only for HPV-16. Values exclude untested HPV-16 and -18 isolates. Variants of HPV-16 and -18 are grouped according to branches of geographical relatedness.

HPV-infection status	No. persistent/total (%)	OR (95% CI)	
		HPV-16 and -18 variants combined	HPV-16 variants only
E branch	26/46 (56·5)	2·53 (1·3–4·9)	2·19 (1·1–4·5)
AA/Af/As branches	13/19 (68·4)	4·47 (1·6–12·4)	5·35 (1·8–15·8)
Other oncogenic types	86/175 (49·1)	1·92 (1·2–3·0)	1·98 (1·3–3·1)
Only non-oncogenic types	59/176 (33·5)	1·0 (reference value)	1·0 (reference value)

with more than one variant of HPV-16 or -18 based on a search of a minimum of five clones per isolate.

Table 1 shows the age- and race-adjusted OR values of persistent infection by HPV-infection status based on up to four results during the first year of follow-up. Between HPV-16- and -18-positive cases, 39 (60%) were classified as persistent. Among individuals infected exclusively with low-risk types, persistent infection was less common (34%). Taking the latter as the reference category, persistent infections were significantly more common among women harbouring HPV-16 and -18 or other oncogenic types. A somewhat stronger association with persistence seemed to emerge for HPV-16 and -18 variants or, when considered alone, for HPV-16 variants of non-E branches (AA, Af, As) as compared to E branch variants, although the 95 % CI values overlapped extensively.

Table 2 compares the mean virus load across categories of

Table 2. Virus burden according to HPV-infection status in the cohort

Values given here are the geometric mean and exclude untested HPV-16 and -18 isolates. Variants of HPV-16 and -18 are grouped according to branches of geographical relatedness.

HPV-infection status (no. isolates)	Mean no. of virus copies per cell (95% CI)	Significance (*P* value)	
		Compared to E branch	Compared to non-oncogenic HPVs only
16 and 18, E branch (85)	5·8 (3·2–10·8)	–	0·0041
16 and 18, AA/Af/As branches (36)	2·8 (1·4–5·7)	0·1737	0·6487
Other oncogenic types (272)	9·5 (6·5–13·9)	0·2169	< 0·0001
Only non-oncogenic types (267)	2·3 (1·7–3·1)	0·0041	–

Table 3. Age- and race-adjusted OR (and respective 95% CI) of prevalent cervical lesions by HPV-infection status at enrolment and in first year follow-up visits

OR and CI values are calculated by logistic regression analysis. Models for any grade SIL exclude atypical squamous cells of undetermined significance (ASCUS) and models for HSIL exclude ASCUS and LSIL. Calculated values exclude untested HPV-16 and -18 isolates. HPV-18 infections were included among other oncogenic types when variants were considered only for HPV-16. Variants of HPV-16 and -18 are grouped according to branches of geographical relatedness.

Lesion outcome, ascertainment method	HPV-infection status (no. cases/subjects)	OR (95% CI)	
		HPV-16 and -18 variants combined	HPV-16 variants only
Any grade SIL, cytology only	Negative (10/1211)	1·0 (reference value)	1·0 (reference value)
	E branch (12/44)	45·48 (18·3–113·3)	55·47 (21·5–143·3)
	AA/Af/As branches (5/23)	99·65 (32·1–308·9)	110·05 (34·9–347·1)
	Other oncogenic types (46/164)	48·06 (23·4–98·9)	44·95 (22·0–92·0)
	Only non-oncogenic types (12/159)	9·08 (3·8–21·8)	9·06 (3·9–21·7)
HSIL, cytology only	Negative (2/1203)	1·0 (reference value)	1·0 (reference value)
	E branch (7/39)	151·25 (29·7–771·3)	173·94 (32·8–923·5)
	AA/Af/As branches (6/16)	529·81 (88·6–3167)	573·30 (95·1–3457)
	Other oncogenic types (12/130)	74·79 (16·2–345·6)	71·49 (15·6–326·7)
	Only non-oncogenic types (3/150)	8·90 (1·2–63·9)	8·88 (1·2–63·7)
HSIL, cytology or cervicography	Negative (7/1203)	1·0 (reference value)	1·0 (reference value)
	E branch (8/40)	46·31 (15·5–138·0)	54·89 (17·5–172·0)
	AA/Af/As branches (7/17)	172·24 (47·1–630·1)	186·21 (50·2–690·1)
	Other oncogenic types (13/131)	21·62 (8·3–56·6)	20·59 (8·0–53·1)
	Only non-oncogenic types (3/150)	2·51 (0·5–12·2)	2·50 (0·5–12·2)

Table 4. Age- and race-adjusted RR (and respective 95% CI) of incident cervical lesions by HPV-infection status at enrolment in the cohort

Values were calculated using Cox proportional hazards regression analyses including women free of the stated lesion grade at enrolment but excluding untested HPV-16 and -18 isolates. HPV-18 infections were included among other oncogenic types when variants were considered only for HPV-16. Variants of HPV-16 and -18 are grouped according to branches of geographical relatedness.

Lesion end-point, ascertainment method	HPV-infection status (no. cases/subjects)	RR (95% CI)	
		HPV-16 and -18 variants combined	HPV-16 variants only
Any grade SIL, cytology only	Negative (26/1145)	1·0 (reference value)	1·0 (reference value)
	E branch (5/35)	7·37 (2·8–19·2)	8·27 (2·9–23·7)
	AA/Af/As branches (4/13)	13·69 (4·7–40·0)	14·92 (5·1–43·5)
	Other oncogenic types (42/154)	14·24 (8·6–23·5)	13·73 (8·4–22·5)
	Only non-oncogenic types (16/173)	4·31 (2·3–8·0)	4·31 (2·3–8·1)
HSIL, cytology only	Negative (3/1148)	1·0 (reference value)	1·0 (reference value)
	E branch (1/37)	11·37 (1·2–109·9)	16·37 (1·7–157·9)
	AA/Af/As branches (2/14)	58·50 (9·3–367·3)	61·99 (9·9–388·1)
	Other oncogenic types (11/163)	26·46 (7·2–96·9)	26·47 (7·3–95·6)
	Only non-oncogenic types (2/174)	4·71 (0·8–28·2)	4·74 (0·8–28·4)
HSIL, cytology and cervicography	Negative (11/1148)	1·0 (reference value)	1·0 (reference value)
	E branch (2/37)	6·06 (1·3–27·4)	8·61 (1·9–39·0)
	AA/Af/As branches (3/14)	22·49 (6·0–83·9)	23·91 (6·4–89·1)
	Other oncogenic types (14/163)	8·66 (3·9–19·5)	8·54 (3·9–18·9)
	Only non-oncogenic types (3/174)	1·85 (0·5–6·6)	1·85 (0·5–6·7)

HPV-infection status. Specimens with other oncogenic types had the highest average viral genome counts (9·5 per cell). Despite the higher copy frequency for E branch variants as compared to non-E branch variants of HPV-16 and -18, the difference in virus load did not reach statistical significance ($P = 0{\cdot}174$). However, infections with E branch variants were significantly more productive than those with only non-oncogenic types ($P = 0{\cdot}0041$), whereas virus load for the latter infections and that for non-E branch variants were virtually indistinguishable (2·3 versus 2·8, respectively, $P = 0{\cdot}649$).

Table 3 shows the logistic regression analyses of prevalent cervical lesions by HPV-infection status using cumulative data from the entire first year of follow-up. Strong and significant associations for all HPV-infection categories were found for any grade SIL and HSIL, regardless of the outcome definition. It is noteworthy that the magnitude of the associations was greatest for non-E branch variants of combined HPV-16 and -18 or HPV-16 alone in all outcome definitions. In most combinations, however, the 95% CI values overlapped substantially. The sole exception was the contrast based on HSIL assessed jointly by cytology and cervicography for which the OR for non-E branch variants of HPV-16 and -18 was significantly greater than that for E branch variants ($P = 0{\cdot}045$).

Table 4 also shows associations between HPV-infection status and lesion outcomes based on similar comparisons to those presented in the previous analysis except that lesion outcomes were defined prospectively in a true cohort fashion and estimates of effect (RR) were assessed via proportional hazards regression models. All prevalent lesions at enrolment were excluded. As above, RR values of lesion development were substantially elevated for HPV-infection categories involving oncogenic HPV types, with the greatest increase in risk being observed for non-E branch variants. As above, however, the increased risk for non-E branch variants was not statistically significant (at the 5% level) as compared to that calculated for the E branch variants in any of the combinations analysed.

Discussion

In this study we have detected several variants of HPV-16 and -18 in cervical specimens from asymptomatic women living in a high-risk area for cervical cancer in Brazil. We

searched for genetic variation in a fragment of the LCR by DNA sequencing, as well as in the E6 and L1 genes by dot blot hybridization. The LCR is a hypervariable region of the viral genome, thus theoretically enabling the detection of a greater diversity of variants than when conserved genes are analysed (Bernard *et al*., 1994*a*). Analysis of defined segments of other genes, such as E2, E5, E6, L1 and L2, has also been shown to indicate substantial diversity in molecular variants of HPV-16 (Eriksson *et al*., 1999; Icenogle *et al*., 1991; Yamada *et al*., 1995; Wheeler *et al*., 1997).

All cases with persistently detected HPV-16 and -18 had the same variant in multiple specimens collected at different points in time. Although we always tested a minimum of five clones per isolate we did not find multiple variants in the same specimen, which is in disagreement with previous reports (L. Ho *et al*., 1991, 1993; Xi *et al*., 1995; Wheeler *et al*., 1997). Interestingly, Xi *et al*. (1995) noticed that one variant prevails over the other in women with multiple infections of HPV-16 variants. Polyclonality of variants of HPV-16 has been more frequently found in cancers (L. Ho *et al*., 1993; Nindl *et al*., 1999). In fact, in a survey of HPV-16-positive invasive cervical cancers from Belém (Northern Brazil) we have been able to detect more than one variant in about 10% of the cases (Junes *et al*., 2000).

Although little is known about the direct biological consequences of variations in the HPV genomes on cervical carcinogenesis, it is reasonable to anticipate that they may be relevant to different aspects of this process. Several reports suggest that some variants of HPV-16 are associated with risk of cervical neoplasia. Londesborough *et al*. (1996), studying genetic variability in the HPV-16 E6 gene in a cohort of women from England, described a T to G substitution at nt 350 that was predominantly associated with virus persistence and risk of cervical neoplasia. Another study reported that a nucleotide change at nt 647 of the E7 gene of HPV-16 was more frequently found in cervical carcinomas (Song *et al*., 1997). Modification of this residue, involved in the binding of E7 to pRb, could have important functional implications.

Of direct relevance to our ongoing investigation were the findings of Xi *et al*. (1997). These authors analysed a fragment of the HPV-16 LCR by single-strand conformation polymorphism analysis in cervical specimens from young American women. They observed that non-prototype-like (NPL) variants were more strongly associated with the development of cervical lesions than prototype-like (PL) variants. Similar results were observed in a series of anal carcinomas where NPL variants were more prevalent than the HPV-16 PL variants (Xi *et al*., 1998). Since the HPV-16 prototype is classified as an E branch variant, based on geographical and phylogenetic relatedness, we hypothesized that the differences in risk attribution between NPL and PL variants found in the latter studies could be indicative of broad differences in the ability to induce cervical carcinogenesis by oncogenic HPVs from different geographical areas. We sought to test this hypothesis by searching for molecular variants in different genome regions using sequencing (LCR) corroborated with direct oligonucleotide probing (E6 and L1). Unlike Xi *et al*. (1997), we grouped our variant findings according to geographical relatedness based on a well-established classification system (L. Ho *et al*., 1993; Ong *et al*., 1993; Yamada *et al*., 1995). We also extended the analysis to include HPV-18 in addition to HPV-16.

Our results are consistent with the aforementioned hypothesis and tend to corroborate the results of Xi *et al*. (1997). We found that infections with non-E branch variants of HPV-16 and -18 have a general tendency to persist more frequently and to be more associated with pre-invasive lesions, both cross-sectionally and prospectively. These results were not a consequence of follow-up bias since women with E branch and non-E branch infections were comparable in terms of follow-up time ($P = 0{\cdot}697$), number of HPV tests ($P = 0{\cdot}896$) and number of Pap tests ($P = 0{\cdot}889$) or cervicographies ($P = 0{\cdot}407$) during follow-up.

An important limitation of our study was the relatively low ability to detect these associations. An additional caveat is the fact that our definition of lesion outcomes was based on cytology and/or cervicography and not on histology. Our ongoing cohort study relies on cytologic and cervicography follow-up with biopsy required only of HSIL lesions that are seen on colposcopy in association with lesional tissue. However, the fact that the findings were consistent across definitions of lesion outcomes and analysis layouts (cross-sectional versus prospective) and that we have observed the increased risk associated with non-E variants since earlier phases of our cohort study bolsters our confidence that the association may be real and worthy of consideration for disease prevention purposes. It is also conceivable that the magnitude of the associations would have been much greater if we had used histological ascertainment of all lesions detected in the study, an observation that we will make at a later phase of the investigation after substantially more cases of HPV-16 and -18 infections are accrued in the cohort.

The distribution of HPV variants defined on the basis of geographical relatedness may be expected to be associated with ethnicity. Since the latter variable is a correlate of cervical cancer risk, a confounded association between variant grouping and lesion risk may occur empirically. In the present analysis, 15 of 46 women (33%) infected with E branch variants were non-white (mostly black and mulatto ethnicity) as compared to 11 of 19 women (58%) infected with non-E branch variants in the same ethnic category. Although the difference fails to reach statistical significance ($P = 0{\cdot}058$), the imbalance in ethnic distribution is sufficient to warrant confusion assuming that race is a correlate of the outcome in our cohort (data not shown). This observation underscores the importance of conducting analyses that are appropriately adjusted for race when gauging the association between variant grouping and lesion risk. We controlled for both ethnicity and age in our

models of the relationships under investigation. We cannot completely rule out, however, a residual confounding influence of race on the results because of the extensive admixture of ethnic groups in the Brazilian population.

The differential oncogenic potential associated with certain variants has been addressed (Hecht *et al.*, 1995; Conrad-Stöppler *et al.*, 1996; Veress *et al.*, 1999; Kämmer *et al.*, 2000). Research in this area is anticipated to provide important information concerning the biological significance of HPV intratype genomic variability, which could ultimately be used to address the control of these infections. Changes in the L1/L2 regions of the HPV genome may be important in discriminating the infectious potential of different variants, as well as in defining epitopes relevant to vaccine design. Our knowledge about the amino acid residues that are involved in antigenic recognition of HPV is very limited. It has been shown that naturally occurring antibodies against L1, the major viral capsid protein, are directed to conformational epitopes (Kirnbauer *et al.*, 1994; Rose *et al.*, 1994). However, little is known about the influence of HLA haplotypes on the host immune response. HLA haplotypes seem to be important determinants of susceptibility to cervical cancer and may represent an important aspect for understanding cervical carcinogenesis (Maciag & Villa, 1999). Recent research on the role of both HLA (Ellis *et al.*, 1995; Terry *et al.*, 1997; Bontkes *et al.*, 1998) and p53 polymorphisms (van Duin *et al.*, 2000) indicates that susceptibility to cervical cancer may depend on the particular variant of HPV-16 to which the woman was originally exposed.

Cervical cancer incidence is much higher in Africa and Latin American countries than in Europe and North America. There is little doubt that ineffective or non-existent screening, lack of access to health care, high fertility and poor nutrition explain many of the differences in rates between these underdeveloped and developed areas. It is also important to realize that non-E variants of HPV-16 and -18 may have increased oncogenic potential and may represent an additional factor contributing to the disproportionately high burden of cervical cancer in underdeveloped populations in different regions of the world.

We are grateful to M. Luiza Baggio and Lenice P. Galan for conducting the epidemiological interviews and collecting samples, to Stella Leme for data management and to Dr Eliane Franco for overseeing the quality control of the cervical pathology information. Excellent technical support was provided by Joao S. Sobrinho, J. Carlos Prado, Silvaneide Ferreira and Monica dos Santos (Ludwig Institute) and by Juliette Robitaille (Jewish General Hospital). Supported in part by grants from the National Institutes of Health (CA70269) and from the Medical Research Council of Canada (13647). Additional support was provided by the São Paulo branch of the Ludwig Institute for Cancer Research. E.L.F is a recipient of a Distinguished Scientist Award from the Medical Research Council of Canada and of a National Research Scholar Award from the Fonds de la recherche en santé du Québec. L.S. is a recipient of a fellowship from Fundação de Amparo à Pesquisa do Estado de São Paulo and P.R. is a recipient of a fellowship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico.

References

Bauer, H. M., Ting, Y., Greer, C. E., Chambers, J. C., Tashiro, C. J., Chimera, J., Reingold, A. & Manos, M. M. (1991). Genital human papillomavirus infection in students as determined by a PCR-based method. *Journal of the American Medical Association* **265**, 472–477.

Bernard, H. U., Chan, S.-Y. & Delius, H. (1994*a*). Evolution of papillomaviruses. *Current Topics in Microbiology and Immunology* **186**, 33–53.

Bernard, H. U., Chan, S.-Y., Manos, M. M., Ong, C. K., Villa, L. L., Delius, H., Peyton, C. L., Bauer, H. M. & Wheeler, C. M. (1994*b*). Identification and assessment of known and novel human papillomaviruses by polymerase chain reaction amplification, restriction fragment length polymorphisms, nucleotide sequence, and phylogenetic algorithms. *Journal of Infectious Diseases* **170**, 1077–1085.

Bontkes, H. T., van Duin, M., de Gruij, I. T. D., Duggan-Keen, M. F., Walboomers, J. M. M., Stukart, M. J., Verheijen, R. H. M., Helmerhorst, T. J. M., Meijer, C. J. L. M., Scheper, R. J., Stevens, F. R. A., Dyer, P. A., Sinnott, P. & Stern, P. L. (1998). HPV-16 infection and progression of cervical intra-epithelial neoplasia: analysis of HLA polymorphism and HPV-16 sequence variants. *International Journal of Cancer* **78**, 166–171.

Caballero, O. L., Villa, L. L. & Simpson, A. J. G. (1995). Low stringency-PCR (LS-PCR) allows entirely internally standardized DNA quantitation. *Nucleic Acids Research* **23**, 193–203.

Chan, S.-Y., Ho, L., Ong, C. K., Chow, V., Drescher, B., Dürst, M., ter Meuler, J., Villa, L. L., Luande, J., Mgaya, H. N. & Bernard, H. U. (1992). Molecular variants of human papillomavirus-16 from four continents suggest ancient pandemic spread of the virus and its co-evolution with humankind. *Journal of Virology* **66**, 2057–2066.

Conrad-Stöppler, M. C., Ching, K., Stöppler, H., Clancy, K., Schlegel, R. & Icenogle, J. (1996). Natural variants of human papillomavirus type 16 E6 protein differ in their abilities to alter keratinocyte differentiation and to induce p53 degradation. *Journal of Virology* **70**, 6987–6993.

Ellis, J. R. M., Keating, P. J., Baird, J., Hounsell, E. F., Renouf, D. V., Rowe, M., Hopkins, D., Duggan-Keen, M. F., Bartholomew, J. S., Young, L. S. & Stern, P. L. (1995). The association of an HPV-16 oncogene variant with HLA-B7 has implications for vaccine design in cervical cancer. *Nature Medicine* **1**, 464–470.

Eriksson, A., Herron, J. R., Yamada, T. & Wheeler, C. M. (1999). Human papillomavirus type 16 variant lineages characterized by nucleotide sequence analysis of the E5 coding segment and the E2 hinge region. *Journal of General Virology* **80**, 595–600.

Eschle, D., Dürst, M., ter Meulen, J., Luande, J., Eberhardt, H. C., Pawlita, M. & Gissman, L. (1992). Geographical dependence of sequence variation in the E7 gene of human papillomavirus type 16. *Journal of General Virology* **73**, 1829–1832.

Franco, E., Villa, L. L., Rahal, P. & Ruiz, A. (1994). Molecular variant analysis as an epidemiological tool to study persistence of cervical human papillomavirus infection. *Journal of the National Cancer Institute* **86**, 1558–1559.

Franco, E., Villa, L., Rohan, T., Ferenczy, A., Petzl-Erler, M. & Matlashewski, G. (1999*a*). Design and methods of the Ludwig–McGill longitudinal study of the natural history of human papillomavirus infection and cervical neoplasia in Brazil. Ludwig–McGill Study Group. *Revista Panamericana De Salud Publica* **6**, 223–233.

Franco, E. L., Villa, L. L., Sobrinho, J., Prado, J. M., Rousseau, M. C., Désy, M. & Rohan, T. E. (1999*b*). Epidemiology of acquisition and clearance of cervical human papillomavirus infection in women from a high-risk area for cervical cancer. *Journal of Infectious Diseases* **180**, 1415–1423.

Hecht, J. L., Kadish, A. S., Jiang, G. & Burk, R. D. (1995). Genetic characterization of the human papillomavirus (HPV)-18 E2 gene in clinical specimens suggests the presence of a subtype with decreased oncogenic potential. *International Journal of Cancer* **60**, 369–376.

Heinzel, P. A., Chan, S.-Y., Ho, L., O'Connor, M., Balaram, P., Campo, M. S., Fujinaga, K., Kiviat, N., Kuypers, J., Pfister, H., Steinberg, B. M., Tay, S.-K., Villa, L. L. & Bernard, H. U. (1995). Variation of human papillomavirus type 6 (HPV-6) and HPV-11 genome samples throughout the world. *Journal of Clinical Microbiology* **33**, 1746–1754.

Hildesheim, A., Schiffman, M. H., Gravitt, P. E., Glass, A. G., Greer, C. E., Zhang, T., Scott, D. R., Rush, B. B., Lawler, P. & Sherman, M. E. (1994). Persistence of type-specific human papillomavirus infection among cytologically normal women. *Journal of Infectious Disease* **169**, 235–240.

Ho, G. Y., Tay, S., Chan, S.-Y. & Bernard, H. U. (1993). Sequence variants of human papillomavirus type 16 from couples suggest sexual transmission with low infectivity and polyclonality in genital neoplasia. *Journal of Infectious Diseases* **168**, 803–809.

Ho, L., Chan, S.-Y., Chow, V., Chong, T., Tay, S., Villa, L. L. & Bernard, H. U. (1991). Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and construction of a phylogenetic tree. *Journal of Clinical Microbiology* **29**, 1765–1772.

Ho, L., Chan, S.-Y., Burk, R. D., Das, B. C., Fujinaga, K., Icenogle, J. P., Kahn, T., Kiviat, N., Lancaster, W., Mavromara-Nazos, P., Labropoulou, V., Mitrani-Rosenbaum, S., Norrild, B., Pillai, M. R., Stoerker, J., Syrjaenen, K., Syrjaenen, S., Tay, S., Villa, L. L., Wheeler, C. M., Williamson, A. L. & Bernard, H. U. (1993). The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations. *Journal of Virology* **67**, 6413–6423.

Icenogle, J. P., Sathy, P., Miller, D. L., Tucker, R. A. & Rawls, W. E. (1991). Nucleotide and amino acid sequence variation in the L1 and E7 open reading frames of human papillomavirus type 6 and type 16. *Virology* **184**, 101–107.

Junes, K. S., Sichero, L., Mello, W., Noronha, V. & Villa, L. L. (2000). Intratypic variability of HPV-16 and -18 in tumour biopsies from a high risk area for cervical cancer. *18th International Papillomavirus Conference* (Barcelona, Spain, 23–28 July, 2000).

Kämmer, C., Warthorst, U., Torrez-Martinez, N., Wheeler, C. M. & Pfister, H. (2000). Sequence analysis of the long control region of human papillomavirus type 16 variants and functional consequences for P97 promoter activity. *Journal of General Virology* **81**, 1975–1981.

Kirnbauer, R., Booy, F., Cheng, N., Lowy, D. R. & Schiller, J. T. (1992). Papillomavirus L1 major protein self-assembles into virus-like particles that are highly immunogenic. *Proceedings of the National Academy of Sciences, USA* **89**, 12180–12184.

Kirnbauer, R., Tabú, J., Greenstone, H., Roden, R., Dürst, M., Gissman, L., Lowy, D. R. & Schiller, J. T. (1993). Efficient self-assembly of human papillomavirus type 16 L1 and L1–L2 into virus-like particles. *Journal of Virology* **67**, 6929–6936.

Kirnbauer, R., Hubbert, N. L., Wheeler, C. M., Becker, T. M., Lowy, D. R. & Schiller, J. T. (1994). A virus-like particle enzyme-linked immunosorbent assay detects serum antibodies in a majority of women infected with human papillomavirus type 16. *Journal of the National Cancer Institute* **86**, 494–499.

Liaw, K. L., Glass, A. G., Manos, M. M., Greer, C. E., Scott, D. R., Sherman, M., Burk, R. D., Kurman, R. J., Wacholder, S., Rush, B. B., Cadell, D. M., Lawler, P., Tabor, D. & Schiffman, M. (1999). Detection of human papillomavirus DNA in cytologically normal women and subsequent cervical squamous intraepithelial lesions. *Journal of the National Cancer Institute* **91**, 954–960.

Londesborough, P., Ho, L., Terry, G., Cuzick, J., Wheeler, C. & Singer, A. (1996). Human papillomavirus genotype as a predictor of persistence and development of high-grade lesions in women with minor cervical abnormalities. *International Journal of Cancer* **69**, 364–368.

Maciag, P. & Villa, L. L. (1999). Genetic susceptibility to HPV infection and cervical cancer. *Brazilian Journal of Medical and Biological Research* **32**, 915–922.

Nindl, I., Rindfleisch, K., Teller, K., Schneider, A. & Dürst, M. (1999). Cervical cancer, HPV-16 E6 variant genotypes, and serology. *Lancet* **353**, 152.

Ong, C. K., Chan, S.-Y., Campo, M. S., Fujinaga, K., Mavromara-Nazos, P., Labropoulou, V., Pfister, H., Tay, S.-K., ter Meulen, J., Villa, L. L. & Bernard, H. U. (1993). Evolution of human papillomavirus type 18: an ancient phylogenetic root in Africa and intratype diversity reflect co-evolution with human ethnic groups. *Journal of Virology* **67**, 6424–6431.

Rose, R. C., Reichmann, R. C. & Bonnez, W. (1994). Human papillomavirus (HPV) type 11 recombinant virus-like particles induce the formation of neutralizing antibodies and detect HPV-specific antibodies in human sera. *Journal of General Virology* **75**, 2075–2079.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. S., Higuchi, H., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487–497.

Song, Y. S., Kee, S. H., Kim, J. W., Park, N. H., Kang, S. B., Chang, W. H. & Lee, H. P. (1997). Major sequence variants in E7 gene of human papillomavirus type 16 from cervical cancerous and non-cancerous lesions of Korean women. *Gynaecologic Oncology* **66**, 275–281.

Terry, G., Ho, L. & Cuzick, J. (1997). Analysis of the E2 amino acid variants of human papillomavirus types 16 and 18 and their associations with lesion grade and HLA DR/DQ type. *International Journal of Cancer* **73**, 651–655.

van Duin, M., Snijders, P. J., Vosen, M. T., Klaassen, E., Voorhorst, F., Verheijen, R. H., Helmerhorst, T. J., Meijer, C. J. & Walboomers, J. M. (2000). Analysis of human papillomavirus type 16 E6 variants in relation to p53 codon 72 polymorphism genotypes in cervical carcinogenesis. *Journal of General Virology* **81**, 317–325.

Veress, G., Szarka, K., Dong, X.-P., Gergely, L. & Pfister, H. (1999). Functional significance of sequence variation in the E2 gene and the long control region of human papillomavirus type 16. *Journal of General Virology* **80**, 1035–1043.

Villa, L. L., Franco, E. L., Caballero, O., Rahal, P., Ferenczy, A. & Rohan, T. E. (1996). Virus load, persistent cervical HPV infection, and cumulative risk of cervical intraepithelial neoplasia in a high-risk area. *15th International Papillomavirus Conference* (Gold Coast, Australia).

Villa, L. L., Rahal, P. & Franco, E. L. (1997). Molecular variant analysis as a tool in natural history studies of human papillomavirus infection and cervical neoplasia. In *New Developments in Cervical Cancer Screening and Prevention*, pp. 379–385. Edited by E. L. Franco & J. Monsonego. Oxford: Blackwell Science.

Wheeler, C. M., Yamada, T., Hildesheim, A. & Jenison, S. A. (1997). Human papillomavirus type 16 sequences variants: identification by E6 and L1 lineage-specific hybridization. *Journal of Clinical Microbiology* **35**, 11–19.

Xi, L. F., Demers, W., Koutsky, L. A., Kiviat, N. B., Kuypers, J., Watts, D. H., Holmes, K. K. & Galloway, D. A. (1995). Analysis of human papillomavirus type 16 variants indicates establishment of persistent infection. *Journal of Infectious Diseases* **172**, 747–755.

Xi, L. F., Koutsky, L. A., Galloway, D. A., Kuypers, J., Hughes, J. P., Wheeler, C. M., Holmes, C. & Kiviat, N. B. (1997). Genomic variation of human papillomavirus type 16 and risk for high-grade cervical intraepithelial neoplasia. *Journal of the National Cancer Institute* **89**, 796–802.

Xi, L. F., Critchlow, C. W., Wheeler, C. M., Koutsky, L. A., Galloway, D. A., Kuypers, J., Hughes, J. P., Hawes, S. E., Surawicz, C., Goldbaum, G., Holmes, K. K. & Kiviat, N. B. (1998). Risk of anal carcinoma in situ in relation to human papillomavirus type 16 variants. *Cancer Research* **58**, 3839–3844.

Yamada, T., Wheeler, C. M., Halpern, A. L., Stewart, A. C. M., Hildesheim, A. & Jenison, S. A. (1995). Human papillomavirus type 16 lineages in the United States populations characterized by nucleotide sequence analysis of the E6, L2 and L1 coding segments. *Journal of Virology* **69**, 7743–7753.

Yamada, T., Manos, M., Peto, J., Greer, C. E., Muñoz, N., Bosch, X. & Wheeler, C. M. (1997). Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective. *Journal of Virology* **71**, 2463–2472.

Zehbe, I. & Tommasino, M. (1999). The biological significance of human papillomavirus type 16 variants for the development of cervical neoplasia. *Papillomavirus Report* **10**, 105–116.

Received 28 April 2000; Accepted 29 August 2000