

1 **Human papillomavirus type 16 variants in San Luis**
2 **Potosi City, Mexico**

3

4 Rubén López-Revilla,^{1,*} Marco A. Pineda,^{1,†} Julio Ortiz-Valdez,² Mireya Sánchez-
5 Garza,¹ and Lina Riego¹

6

7

8 ¹División de Biología Molecular, Instituto Potosino de Investigación Científica y
9 Tecnológica, Camino a la Presa San José 2055, 78216 San Luis Potosí, S.L.P.,
10 Mexico, ²Clínica de Colposcopia, Jurisdicción Sanitaria 1, Secretaría de Salud de San
11 Luis Potosí, Calzada de Guadalupe 530, 78339 San Luis Potosí, S.L.P., Mexico

12

13 *Corresponding author

14

15 Email: Rubén López-Revilla - rlopez@ipicyt.edu.mx; MAP - marco@ipicyt.edu.mx;

16 JOV - dr_ortiz_valdez@hotmail.com; MSG - mireya@ipicyt.edu.mx; LR -

17 lina@ipicyt.edu.mx

18

[†] Present address: M.C. Marco Antonio Pineda, Área de Ciencias Químico-Biológicas, Universidad del Noreste, Prol. Av. Hidalgo 6315, 89337 Tampico, Tams., Mexico. Phone +52(833)228-1156 extension 1103. Fax: +52(833)228-1182.

18

19 **Abstract**

20 **Background**

21 Human papillomavirus type 16 (HPV16) subtypes and variants from women with
22 cervical displasia and cancer from San Luis Potosí City, Mexico, were identified by
23 their E6 oncogene open reading frame (ORF) sequences. From the DNA of cervical
24 scrapings of 38 HPV16-infected women, E6-1 amplicons (~650 base pairs long)
25 generated with the LCRS/E7AS primer pair were used as templates for the E6F/PU-
26 2R16 primer pair to obtain E6-2 amplicons (~626 bp long) containing the E6-HPV16
27 ORF sequence.

28 **Results**

29 Three European (E) subtype variants (E-P, n = 27; E-T350G, n = 7; E-C188G, n = 2)
30 and one AA-a subtype variant (n = 2) were identified among the 38 E6-2 sequences.
31 E-P variant sequences contained 23 single nucleotide polymorphisms (SNP's), two of
32 which (A334G, A404T) had not been described before and allowed the phylogenetic
33 separation from the other variants. E-P A334G sequences predominated (22 cases,
34 57.9%), followed by the E-P Ref prototype (4 cases, 10.5%) and E-P A404T (1 case,
35 2.6%) sequences. The ratios of high to low grade squamous intraepithelial lesions
36 were 4/18 for E-P A334G sequences and 0/4 for E-P Ref sequences.

37 **Conclusions**

38 We conclude that in the women included in this study the HPV16 E subtype is 19
39 times more frequent than the AA subtype; that the circulating E subtype variants are
40 E-P (71.1%) > E-T350G (18.4%) > E-C188G (5.3%); and that 81.5% of the E-P
41 sequences carry the SNP A334G and appear to correspond to a HPV16 variant
42 characteristic of San Luis Potosi City.

43

44

45 **Background**

46 Human papillomavirus (HPV) types differ from each other by at least 10% of the L1
47 gene ORF sequence [1, 2]; differences among subtypes are 2%-10%, and smaller than

48 2% among variants [1, 3-6]. E6 gene sequences can also be used to identify HPV
49 types, subtypes and variants [7].

50 Persistent infection by high risk-HPV, among which HPV16 is the
51 predominant type, can progress to invasive cervical cancer (CC). The risk of CC
52 increases with certain HPV16 subtypes [8, 9] whose prototype is the European (E-P
53 Ref) subtype [7, 10]. Sixty-percent of invasive CC cases in Mexican women 35 years
54 old or younger are attributed to the Asian-American (AA) HPV16 subtype [11-13],
55 whereas in invasive CC cases of younger women in Mexico City the AA subtype is
56 21 times more frequent than the E subtype [14].

57 Invasive CC is a public health problem in the state of San Luis Potosí, whose
58 mortality rate in 2005 was above the national average and occupied the tenth place
59 among the 32 federated states [15]. In San Luis Potosí City, the state capital, infection
60 by HPV16 has the highest prevalence [16] and precancerous and cancerous lesions of
61 the cervix predominate in the youngest women (R. López-Revilla and L. Rosales-
62 Ortuño, unpublished data), suggesting that a more oncogenic HPV16 variant may be
63 circulating there.

64 In this work we identified the HPV16 subtypes and variants in cervical
65 precancerous and cancerous lesions from women residing in San Luis Potosí City by
66 comparing the E6 ORF sequences amplified from them.

67

68

69 **Results**

70

71 **Direct and nested PCR**

72 The intensity of E6-2 bands generated by nested PCR was highest in the presence of 2
73 mM MgCl₂, which was used in all subsequent amplification experiments. In a pilot
74 test E6-2 bands were amplified by direct PCR in positive controls (pHPV16 and
75 pHPV18) and in 23 out of 28 (82.1%) cervical samples. In view of these results we
76 decided to amplify all samples through nested PCR.

77 E6-1 was preamplified in PCR1 mixtures with the LCRS/E7AS primer pair
78 and E6-2 in PCR2 mixtures with the internal E6F/PU-2R16 primer pair. In this way

79 conspicuous E6-2 bands were obtained from all samples with yields sufficient (>
80 1500 ng) to sequence the amplicons (Fig. 2).

81

82 **Identification of HPV16 subtypes and variants**

83 The 5'-termini of the deposited sequences were aligned to start at nucleotide 83 of the
84 E-P genome, the first one of the ATG start codon of the E6 open reading frame
85 (ORF), and their upstream portions were ignored. Lengths of the deposited sequences
86 ranged from 337 to 576 bp (average = 541 bp). All sequences were longer than the
87 477 bp expected for the complete E6 ORF, except HPV16-27 (377 bp), HPV16-9
88 (436 bp), and HPV16-21 (465 bp). The GenBank/EMBL/DDBJ accession numbers of
89 the 38 sequences of this study, EU880235 to EU880272, are included in Table 2.

90 HPV16 subtypes and variants were identified by comparing the sequences
91 with that of the E6 ORF of the HPV16 E-P Ref prototype variant [7]. The HPV16
92 subtypes identified are listed in Table 2. The predominant one is the European
93 subtype (E) with 36 cases (94.7%); the two cases of the Asian-American (AA)
94 subtype represent 5.3% of the total (Table 2).

95 The frequency of HPV16 variants identified is shown in Fig. 3. For the E
96 subtype the most frequent was the E-P Ref variant with 27 cases (71.1%) followed by
97 the E-T350G variant (seven cases, 18.4%) and the E-C188G variant (two cases,
98 5.3%). AA-a was the only AA subtype variant (two cases, 5.3%).

99

100 **Novel polymorphisms identified**

101 In the E6-HPV16 ORF sequences 24 single nucleotide polymorphisms (SNP's) were
102 found that had not been described before (Table 3); 21 of them (13 substitutions, 6
103 insertions, 2 deletions) appeared once in nine sequences; the remaining three were
104 observed in two sequences. The A404T SNP was observed in a single E-P variant.

105 In 13 sequences there was a deletion of two neighboring bases (AC) located in
106 nucleotide positions 56 and 57, immediately before the ATG protein start codon. The
107 A334G synonymous substitution was found in 22 sequences, located next to
108 nucleotide 335, commonly used for subtype-variant identification.

109 Amino acid sequences of the E6 oncoprotein encoded by the samples analyzed
110 were compared with the E-P Ref sequence to identify non synonymous mutations

111 (Table 3). E-T350G variants had the expected substitution of leucine for valine at
112 position 83 (L83V). The expected amino acid changes were also found in AA-a
113 (Q14H; H78Y; L83V) and E-C188G variants (E20Q; L83V). Deletion of nucleotide
114 484, identified in one of the two AA-a variants implies the frameshift in the E6 ORF
115 starting at amino acid 127. The A404T SNP identified in a single sample, produced
116 the I101F change. The C37A SNP identified in an E-P variant produced the change
117 Q91K. The C206T SNP present in a E-T350G variant generated a stop codon instead
118 of amino acid 35 and T351A caused the L83E change in the same sample. The C173A
119 substitution produced the H24N change in one sequence, and the SNP A182T the
120 I27L change in another one.

121

122 **Phylogenetic analysis**

123 Identical dendrograms were generated with the 38 E6-2 nucleotide sequences using
124 the Phylip and MEGA programs. Comparing all sequences it was not possible to
125 resolve the A404T non synonymous mutation as a group independent of the E-P Ref
126 sequence. A tree was thus constructed with the six E6-HPV16 ORF sequences
127 representing each of the variants identified, including those containing the SNP's
128 A334G and A404T. The optimal tree with branch length = 0.01907 was drawn to
129 scale with the same evolutionary distance units used to infer the phylogenetic tree
130 (Fig. 4). In this way the AA-a variant and the three known E-P variants (E-P Ref, E-
131 T350G and E-C188G) could be related, with two new branches of the E subtype
132 corresponding to those carrying the A334G and A404T SNP's.

133

134 **Association of cervical lesions with E-P A334G sequences**

135 The presence of SNP A334G in 81.5% of the 27 E-P sequences makes the
136 corresponding HPV16 "variant" the predominant one and characteristic of San Luis
137 Potosi City. The E-P A334G variants appeared to be more oncogenic, because the
138 ratios of high to low grade squamous intraepithelial lesions (HSIL/LSIL) were 4/18
139 (22.2%) for them compared to 0/4 (0.0%) for the E-P Ref variants (Table 4).

140

141

142 **Discussion**

143 HPV16 is the viral type with the highest association to HSIL and invasive cervical
144 cancer and was one of the first HPV's to be sequenced [17], and it is well known that
145 infection by certain HPV16 subtypes and variants can lead to faster disease
146 progression in younger women [18-20].

147 HPV subtypes have been identified by comparing the sequences from the E6
148 and L1 genes and of the long control region (LCR). In the case of HPV16 the E6 gene
149 is frequently used because a short and continuous fragment of its sequence contains
150 sufficient information to identify all the subtypes and variants that have been
151 described [10, 21-24]. HPV16 subtypes differ in prevalence, biochemical and
152 biological properties (e.g., subtype AA replication and expression of E6 and E7
153 oncogenes is more efficient) with uncertain implications in CC aetiology [12, 25-27].

154 In this work, the first to approach HPV molecular phylogeny in Central
155 Mexico, we identified HPV16 subtypes and variants by comparing the E6 ORF
156 sequences [7]. Through the use of nested PCR enough E6-2 DNA was obtained to
157 sequence the amplified samples. Amplicon length was heterogeneous because the
158 length of the product ends is variable with the sequencing method used [28].
159 Sequence analysis confirmed that all samples correspond to HPV16, supporting the
160 specificity of the PCR-RFLP method [29] we used for genotyping [16].

161 Among the HPV16 E6-2 amplicon sequences, 36 corresponded to the E
162 subtype and two to the AA subtype. No African subtypes were identified, as has been
163 the case in previous studies performed in Mexico City [12, 14].

164 Berumen et al. [14] found a 1.1% prevalence of the HPV16 AA subtype in
165 controls and 23.2% in invasive CC cases in Mexico City, whereas del Refugio
166 Gonzalez-Losa et al. [12] did not find the HPV16 AA subtype. These differences are
167 probably due to the fact that the group analyzed by Berumen et al. [14] had invasive
168 CC, whereas in this work 27 lesions were LSIL (71.1%), eight were HSIL (21.1%)
169 and three (7.9%) corresponded to invasive CC (one adenocarcinoma associated to the
170 AA-a variant, and two squamous cell carcinomas: one associated to an E-C188G
171 variant and the other to an E-P Ref variant).

172 We identified four HPV16 variants in San Luis Potosi City: E-P (n = 27,
173 71.1%), E-T350G (n = 7, 18.4%), E-C188G (n = 2, 5.3%) and AA-a (n = 2, 5.3%),
174 whereas Berumen et al. [14] observed frequencies of 47% for the E-T350G and 8%
175 for the E-350T variants. The contrast between the subtype and variant frequencies

176 found by us probably derive from differences in the severity of the lesions in the
177 patients included and in the prevalence of the circulating types in the two populations
178 analyzed, since we have already found that the frequencies of HPV types circulating
179 in the neighboring Mexican states of San Luis Potosi and Guanajuato are different
180 [16] and differ from those observed in Mexico City [12, 14].

181 Besides the known point mutations characteristic of the HPV16 variants
182 identified by us, we detected 24 novel SNP's, two of which appeared in a
183 considerable proportion of the sequences analyzed. The non synonymous A404T
184 substitution, observed only in one sequence close to the E-P Ref, generates the I101F
185 amino acid change in the E6 protein. The second most frequent novel SNP, observed
186 in 13 cases (34.2%), consists in a deletion of two contiguous bases (AC) in the
187 nucleotide positions 56 and 57 located in the 5'-untranslated region of the E6 gene.

188 The synonymous A334G, the most frequent of the novel SNP's found in 22 E
189 subtype sequences (57.9%) phylogenetically close to the E-P Ref prototype and,
190 appear to identify an HPV16 variant characteristic of the region. To verify if it
191 corresponds to a new variant, the complete viral genome must be cloned and
192 sequenced [1].

193
194

195 **Conclusions**

196 HPV16 subtypes and variants infecting the cervix of 38 women from San Luis Potosí
197 City, Mexico, were identified by comparing the sequences of nested E6 ORF PCR
198 products; 36 sequences corresponded to the E subtype and two to the AA subtype.

199 Three variants of the E subtype were identified: E-P (n = 27, 71.1%), E-
200 T350G (n = 7, 18.4%), E-C188G (n = 2, 5.3%). The only AA variant identified was
201 AA-a (n = 2, 5.3%).

202 Besides the known point mutations of the E subtype variants identified, 24
203 novel SNP's were detected.

204 The most frequent of the novel SNP's, found in 22 (61.1%) E subtype
205 sequences, is the synonymous A334G which appears to identify an HPV16 variant
206 characteristic of the region.

207 The second most frequent novel SNP, observed in 13 E subtype sequences
208 (34.2%), is a deletion of two contiguous bases (AC) in nucleotide positions 56 and 57
209 of the E6 5'-untranslated region.

210

211

212 **Methods**

213 **DNA from cervical scrapings**

214 Thirty-eight women residing in San Luis Potosí City with precancerous or cancerous
215 lesions of the cervix and HPV16 infection diagnosed previously in our laboratory [16]
216 were included in this study. Cervical scrapings were obtained by one of us (JOV) at
217 the Colposcopy Clinic, Secretaría de Salud, San Luis Potosi City.

218 Each scraping was taken with an endocervical brush ('cytobrush') that was
219 immediately inserted into a 5 mL polypropylene tube (Nalge Nunc, Rochester, NY)
220 containing 1 mL phosphate buffered saline (PBS: 137 mM NaCl, 2.7 mM KCl, 10
221 mM Na₂HPO₄, 2 mM KH₂HPO₄, pH 7.4) supplemented with sterile 25 mM disodium
222 ethylene-diaminetetraacetate (EDTA), pH 8.0 (PBS-EDTA). Once detached from the
223 cytobrush and suspended in the PBS-EDTA vehicle, each sample was fixed by
224 addition of 1.5 mL 96% ethanol and processed to extract DNA on the same day or up
225 to 30 days after being kept at room temperature. Reagents were purchased from J.T.
226 Baker (Xalostoc, Mexico) unless other source is specified.

227 To extract the DNA, each fixed sample was mixed by vortexing and 1 mL
228 transferred to a 1.5 mL tube and spun in a Hettich Mikro 20 microcentrifuge
229 (Cologne, Germany) for 5 min at 13,000 rpm (16,250×g). The supernatant was
230 discarded by decantation and to each pellet were added 500 µL of Tris-EDTA-saline
231 (TES: 10 mM Tris-HCl; 2 mM disodium EDTA, 0.4 M NaCl, pH 8.0 at 25°C), 50 µL
232 of 10% sodium dodecyl sulphate and 20 µL of proteinase K (20 mg/mL). Mixtures
233 were incubated at 55°C for 3 h, at the end of which 150 µL of 5 M NaCl were added
234 and centrifuged again for 15 min. Each supernatant was aspirated and transferred to a
235 tube to which 577 µL cold isopropanol were immediately added and then left stand
236 for 10 min at 4°C to precipitate the nucleic acids. The tubes were centrifuged again
237 for 10 min and the supernatants discarded by decantation. Each pellet was washed by
238 vortexing with 1 mL of 70% cold ethanol and centrifuged for 10 min at 10,000 rpm

239 (9,615×g) and room temperature. Supernatants were discarded by aspiration and the
240 pellets dried out by inverting the tubes for 15 min on a paper towel. Each pellet was
241 dissolved with 50 µL TE (10 mM Tris-HCl, 1 mM disodium EDTA, pH 8.0 at 25°C).

242 DNA quality was verified by electrophoretic analysis in 1% agarose gels with
243 TAE buffer (40 mM Tris-acetate, 1 mM disodium EDTA, pH 8.2 at 25°C). Two-µL
244 from each sample were applied to gels which were run at 60 V for 90 min. λ-phage
245 DNA digested with *Hind* III (Sigma-Aldrich, Mexico) was used as marker. After
246 staining for 20 min with ethidium bromide (1 µg/mL) gels were transilluminated with
247 ultraviolet light and their fluorescence recorded with the Bio-Rad ChemiDoc EQ
248 (Hercules, CA) photodocumenter.

249 DNA was quantified by fluorometry with the PicoGreen dsDNA Quantitation
250 kit (Molecular Probes; Eugene, OR) by interpolation in a standard curve containing
251 up to 50 ng of λ-phage DNA. To each well of a black FIA 96 well plate (Greiner Bio-
252 One, Frickenhausen, Germany) 198 µL of the assay solution (PicoGreen diluted 1:400
253 in TE) and 2 µL of standard DNA or problem samples were added, and their
254 fluorescence determined using a 485 nm excitation filter and a 535 emission filter in
255 the GENios Pro fluorometer (Tecan Systems, San Jose, CA) with the Magellan 4
256 software.

257

258 **Direct and nested PCR**

259 Nested PCR was used to generate enough DNA to sequence the E6-HPV16 ORF [16].
260 The E6-1 product (~650 bp) was preamplified with the LCRS/E7AS primer pair in the
261 first reaction (PCR1), and the E6-2 product (~626 bp) with the E6F/PU-2R16 primer
262 pair in the second reaction (PCR2) (Fig.1, Table 1). PCR1 mixtures of 50 µL
263 contained 2 mM MgCl₂, the four deoxynucleotide triphosphates (0.4 mM each),
264 forward and reverse primers (0.6 µM each), 1.5 U of *Taq* DNA polymerase and 25 ng
265 of cervical DNA in 200 mM Tris-HCl 500 mM KCl, pH 8.4. Mixtures were
266 preamplified by incubation in a Touchgene Gradient (Techne) thermocycler with
267 initial denaturation at 94°C by 4 min, 40 cycles of amplification (1 min denaturation
268 at 94°C, 1 min annealing at 55°C 1 min extension at 72°C) and 10 min final extension
269 at 72°C. To generate E6-2, to PCR2 mixtures (same composition as PCR1 mixtures
270 except for primers and DNA) 1 µL of each preamplified PCR1 mixture was added as

271 template. Amplification products were electrophoresed in high-resolution sodium
272 borate (SB)1% agarose gels [30].

273 The pHPV16 and pHPV18 plasmids containing the complete genomes of
274 HPV16 and HPV18 respectively, donated by Dr. Alejandro García Carrancá (Instituto
275 Nacional de Cancerología, Mexico City), were used as positive controls. PCR
276 mixtures without DNA were used as negative controls.

277

278 **Identification and phylogeny of HPV16 subtypes and variants**

279 The DNA purified with the Wizard kit (PCR Preps DNA Purification Systems,
280 Promega, Madison, WI) from 40 µL of PCR2 mixtures was used to sequence E6-2
281 amplicons with the method of Sanger et al. [28] at the National Laboratory for
282 Genomic Biodiversity (Guanajuato campus of CINVESTAV, Mexico) and compared
283 with the E6 ORF of the E-P Ref HPV16 prototype by multiple alignment with the
284 ClustalW v1.82 software [31]. Viral subtypes and variants were identified by
285 comparing the E6-HPV16 ORF nucleotide sequences published by Yamada et al. [7].
286 Allocation of nucleotide and amino acid positions in the E6-HPV16 ORF was based
287 on the nucleotide sequences [32]. Amino acid sequences of the E6 oncoprotein were
288 predicted with the Translate program tool of the ExPASy database [33].

289 Their evolutionary history was inferred using the UPGMA method [13].
290 Evolutionary distances were computed using the 2-parameter method [34] whose
291 units are the number of base substitutions per site. Codon positions included were first
292 + second + third + noncoding. All positions containing gaps and missing data were
293 eliminated from the dataset (complete deletion option); there were a total of 477
294 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [35].

295

296 **Competing interests**

297 The authors declare that they have no competing interests.

298

299 **Authors' contributions**

300 JOV took care of the patients and selected and obtained the cervical samples. MAP
301 performed most of the molecular studies and the bioinformatics analyses. MSG

302 collaborated in PRC amplification experiments and sequencing. LR directed and
303 supervised the bioinformatics analyses. RLR conceived and designed the study,
304 obtained the funds to carry it out and drafted the manuscript. All authors read and
305 approved the final manuscript.

306

307

308 **Acknowledgements**

309 This work was partially supported by grants from Consejo Nacional de Ciencia y
310 Tecnología (CONACYT, Mexico) and the states of San Luis Potosí (FMSLP-4441)
311 and Guanajuato (FONINV-Gto-C01-5751) granted to RLR. MAP received
312 scholarships from CONACYT (182067) and IPICYT.

313

314

315 **References**

316

- 317 1. de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H:
318 **Classification of papillomaviruses.** *Virology* 2004, **324**(1):17-27.
- 319 2. Munger K, Baldwin A, Edwards KM, Hayakawa H, Nguyen CL, Owens M,
320 Grace M, Huh K: **Mechanisms of human papillomavirus-induced**
321 **oncogenesis.** *J Virol* 2004, **78**(21):11451-11460.
- 322 3. Ho L, Chan SY, Chow V, Chong T, Tay SK, Villa LL, Bernard HU:
323 **Sequence variants of human papillomavirus type 16 in clinical samples**
324 **permit verification and extension of epidemiological studies and**
325 **construction of a phylogenetic tree.** *J Clin Microbiol* 1991, **29**(9):1765-
326 1772.
- 327 4. Chan SY, Delius H, Halpern AL, Bernard HU: **Analysis of genomic**
328 **sequences of 95 papillomavirus types: uniting typing, phylogeny, and**
329 **taxonomy.** *J Virol* 1995, **69**(5):3074-3083.
- 330 5. Stewart AC, Eriksson AM, Manos MM, Munoz N, Bosch FX, Peto J, Wheeler
331 **CM: Intratype variation in 12 human papillomavirus types: a worldwide**
332 **perspective.** *J Virol* 1996, **70**(5):3127-3136.

- 333 6. Xi LF, Koutsky LA, Galloway DA, Kuypers J, Hughes JP, Wheeler CM,
334 Holmes KK, Kiviat NB: **Genomic variation of human papillomavirus type**
335 **16 and risk for high grade cervical intraepithelial neoplasia.** *J Natl Cancer*
336 *Inst* 1997, **89**(11):796-802.
- 337 7. Yamada T, Manos MM, Peto J, Greer CE, Munoz N, Bosch FX, Wheeler CM:
338 **Human papillomavirus type 16 sequence variation in cervical cancers: a**
339 **worldwide perspective.** *J Virol* 1997, **71**(3):2463-2472.
- 340 8. Lizano M, Garcia-Carranca A: **[Molecular variants of human**
341 **papillomaviruses types 16, 18, and 45 in tumors of the uterine cervix in**
342 **Mexico].** *Gac Med Mex* 1997, **133 Suppl 1**:43-48.
- 343 9. Villa LL, Sichero L, Rahal P, Caballero O, Ferenczy A, Rohan T, Franco EL:
344 **Molecular variants of human papillomavirus types 16 and 18**
345 **preferentially associated with cervical neoplasia.** *J Gen Virol* 2000, **81**(Pt
346 12):2959-2968.
- 347 10. Yamada T, Wheeler CM, Halpern AL, Stewart AC, Hildesheim A, Jenison
348 SA: **Human papillomavirus type 16 variant lineages in United States**
349 **populations characterized by nucleotide sequence analysis of the E6, L2,**
350 **and L1 coding segments.** *J Virol* 1995, **69**(12):7743-7753.
- 351 11. Casas L, Galvan SC, Ordonez RM, Lopez N, Guido M, Berumen J: **Asian-**
352 **american variants of human papillomavirus type 16 have extensive**
353 **mutations in the E2 gene and are highly amplified in cervical carcinomas.**
354 *Int J Cancer* 1999, **83**(4):449-455.
- 355 12. del Refugio Gonzalez-Losa M, Laviada Mier y Teran MA, Puerto-Solis M,
356 Garcia-Carranca A: **Molecular variants of HPV type 16 E6 among Mexican**
357 **women with LSIL and invasive cancer.** *J Clin Virol* 2004, **29**(2):95-98.
- 358 13. Ordonez RM, Espinosa AM, Sanchez-Gonzalez DJ, Armendariz-Borunda J,
359 Berumen J: **Enhanced oncogenicity of Asian-American human**
360 **papillomavirus 16 is associated with impaired E2 repression of E6/E7**
361 **oncogene transcription.** *J Gen Virol* 2004, **85**(Pt 6):1433-1444.
- 362 14. Berumen J, Ordonez RM, Lazcano E, Salmeron J, Galvan SC, Estrada RA,
363 Yunes E, Garcia-Carranca A, Gonzalez-Lira G, Madrigal-de la Campa A:
364 **Asian-American variants of human papillomavirus 16 and risk for**
365 **cervical cancer: a case-control study.** *J Natl Cancer Inst* 2001, **93**(17):1325-
366 1330.

- 367 15. [Women mortality (standardized by age) due to non-transmissible
368 diseases, by state of residence, 2005]
369 [http://sinais.salud.gob.mx/mortalidad/tabs/m_016.xls]
- 370 16. Lopez-Revilla R, Martinez-Contreras LA, Sanchez-Garza M: **Prevalence of**
371 **high-risk human papillomavirus types in Mexican women with cervical**
372 **intraepithelial neoplasia and invasive carcinoma.** *Infect Agent Cancer*
373 2008, **3**(1):3.
- 374 17. Seedorf K, Krammer G, Durst M, Suhai S, Rowekamp WG: **Human**
375 **papillomavirus type 16 DNA sequence.** *Virology* 1985, **145**(1):181-185.
- 376 18. Bosch FX, Manos MM, Munoz N, Sherman M, Jansen AM, Peto J, Schiffman
377 MH, Moreno V, Kurman R, Shah KV: **Prevalence of human papillomavirus**
378 **in cervical cancer: a worldwide perspective. International biological study**
379 **on cervical cancer (IBSCC) Study Group.** *J Natl Cancer Inst* 1995,
380 **87**(11):796-802.
- 381 19. Hildesheim A, Schiffman M, Bromley C, Wacholder S, Herrero R, Rodriguez
382 A, Bratti MC, Sherman ME, Scarpidis U, Lin QQ *et al*: **Human**
383 **papillomavirus type 16 variants and risk of cervical cancer.** *J Natl Cancer*
384 *Inst* 2001, **93**(4):315-318.
- 385 20. Swan DC, Rajeevan M, Tortolero-Luna G, Follen M, Tucker RA, Unger ER:
386 **Human papillomavirus type 16 E2 and E6/E7 variants.** *Gynecol Oncol*
387 2005, **96**(3):695-700.
- 388 21. Wheeler CM, Yamada T, Hildesheim A, Jenison SA: **Human papillomavirus**
389 **type 16 sequence variants: identification by E6 and L1 lineage-specific**
390 **hybridization.** *J Clin Microbiol* 1997, **35**(1):11-19.
- 391 22. Eriksson A, Herron JR, Yamada T, Wheeler CM: **Human papillomavirus**
392 **type 16 variant lineages characterized by nucleotide sequence analysis of**
393 **the E5 coding segment and the E2 hinge region.** *J Gen Virol* 1999, **80** (Pt
394 **3**):595-600.
- 395 23. Veress G, Szarka K, Dong XP, Gergely L, Pfister H: **Functional significance**
396 **of sequence variation in the E2 gene and the long control region of human**
397 **papillomavirus type 16.** *J Gen Virol* 1999, **80** (Pt **4**):1035-1043.
- 398 24. Kammer C, Warthorst U, Torrez-Martinez N, Wheeler CM, Pfister H:
399 **Sequence analysis of the long control region of human papillomavirus**

- 400 **type 16 variants and functional consequences for P97 promoter activity. *J*
401 *Gen Virol* 2000, **81**(Pt 8):1975-1981.**
- 402 25. Stoppler MC, Ching K, Stoppler H, Clancy K, Schlegel R, Icenogle J: **Natural**
403 **variants of the human papillomavirus type 16 E6 protein differ in their**
404 **abilities to alter keratinocyte differentiation and to induce p53**
405 **degradation. *J Virol* 1996, **70**(10):6987-6993.**
- 406 26. Zehbe I, Wilander E, Delius H, Tommasino M: **Human papillomavirus 16**
407 **E6 variants are more prevalent in invasive cervical carcinoma than the**
408 **prototype. *Cancer Res* 1998, **58**(4):829-833.**
- 409 27. Cruz MR, Cerqueira DM, Cruz WB, Camara GN, Brigido MM, Silva EO,
410 Carvalho LG, Martins CR: **Prevalence of human papillomavirus type 16**
411 **variants in the Federal District, Central Brazil. *Mem Inst Oswaldo Cruz*
412 2004, **99**(3):281-282.**
- 413 28. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating**
414 **inhibitors. *Proc Natl Acad Sci U S A* 1977, **74**(12):5463-5467.**
- 415 29. Fujinaga Y, Shimada M, Okazawa K, Fukushima M, Kato I, Fujinaga K:
416 **Simultaneous detection and typing of genital human papillomavirus DNA**
417 **using the polymerase chain reaction. *J Gen Virol* 1991, **72** (Pt 5):1039-
418 1044.**
- 419 30. Brody JRK, S.E.: **Sodium boric acid: a Tris-free, cooler conductive**
420 **medium for DNA electrophoresis. *BioTechniques* 2000, **36**(2):214-216.**
- 421 31. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the**
422 **sensitivity of progressive multiple sequence alignment through sequence**
423 **weighting, position-specific gap penalties and weight matrix choice.**
424 *Nucleic Acids Res* 1994, **22**(22):4673-4680.
- 425 32. Farmer AD, Calef CE, Millman K, Myers GL: **The Human Papillomavirus**
426 **Database. *J Biomed Sci* 1995, **2**(2):90-104.**
- 427 33. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A:
428 **ExPASy: The proteomics server for in-depth protein knowledge and**
429 **analysis. *Nucleic Acids Res* 2003, **31**(13):3784-3788.**
- 430 34. Kimura M: **A simple method for estimating evolutionary rates of base**
431 **substitutions through comparative studies of nucleotide sequences. *J Mol*
432 *Evol* 1980, **16**(2):111-120.**

433 35. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary**
434 **Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007,
435 **24(8):1596-1599.**
436

437 **Figures**

438 **Figure 1. Direct and nested PCR used to generate E6-1 and E6-2 amplicons**

439 The upper part of the figure is a diagram of the long control region (LCR) and the E6
440 and E7 genes, whose sequences partially overlap. The middle part describes the E6-1
441 amplicon of 650 bp generated by direct PCR with the LCRS/E7AS primer pair and
442 the E6-2 amplicon of 626 bp generated by nested PCR with the E6F/PU-2R 16 primer
443 pair. The lower part is a scale to mark the positions of the LCR, E6 and E7 genes and
444 the E6-1 and E6-2 amplicons in the HPV16 genome.

445

446 **Figure 2. Nested PCR increases the efficiency of E6 amplicon synthesis**

447 Examples of amplification from HPV16-positive cervical scrapings by direct and
448 nested PCR. **(A)** E6-1 amplification from cervical DNA by direct PCR with the
449 LCRS/E7AS primer pair. Lane 1, 100 bp ladder. Lanes 2-5, DNA of cervical samples.
450 Lane 6, Positive control 1 (pHPV16). Lane 7, Positive control 2 (pHPV18). Lane 8,
451 Negative control (water). **(B)** E6-2 amplification by nested PCR with the E6F/PU-
452 2R16 primer pair from direct PCR mixtures. Lanes 1-4, Amplicons generated using
453 direct PCR mixtures from Panel A lanes 2-5 as templates. Lane 5, Positive control 1
454 (pHPV16). Lane 6, Positive control 2 (pHPV18). Lane 7, Negative control (water).
455 Lane 8, 100 bp ladder.

456

457 **Figure 3. HPV16 variants identified In San Luis Potosi City**

458 The sequences of two cases corresponded to the Asian-American a subtype (AA-a
459 variant) whereas the 36 remaining cases corresponded to the European subtype (E).
460 Within the E subtype the predominant variant was E-P (n = 27), followed by variants
461 E-T350G (n = 7) and E-C188G (n = 2).

462

463 **Figure 4. Phylogenetic tree of the HPV16 variants formed with six** 464 **representative E6-2 ORF sequences**

465 Numbers to the right of each branch indicate the identity of the cervical samples; to
466 the right of the identity number the name of the variant is indicated. Capital letters to
467 the extreme right indicate the HPV16 European (E) or Asian-American (AA) subtype
468 to which the variants belong. This optimal phylogenetic tree was constructed to scale,
469 with branch lengths having the same units as the evolutionary distances used to infer
470 them. Note that sequences with the A334G and A404T SNP's (framed) identified for
471 the first time in this work, as well as the already known C188G, T350G variants are
472 close to the E-P Ref (P) prototype variant.
473

473

474 **Tables**

475 **Table 1. Primers used**

Pair	Primer	Sequence (5'→3')	Amplicon
1	LCRS (F)	AAGGGAGTAACCGAAAACGGT	E6-1 (~650 pb)
	E7AS (R)	TCATCCTCCTCCTCTGAG	
2	E6F (F)	CGTAACCGAAATCGGTTGAAC	E6-2 (~626 pb)
	PU-2R16 (R)	GAGCTGTCGCTTAATTGCTC	

476

477 **Table 2. Characteristics of the E6-2 amplicon sequences**

Cervical Sample	Characteristics of the sequences				
	Accession number ^a	Length (bp)	HPV16 subtype	Variant	Novel SNP
HPV16-1	EU880235	521	AA	AA-a	---
HPV16-2	EU880236	573	AA	AA-a	---
HPV16-3	EU880237	537	E	E-T350G	---
HPV16-4	EU880238	560	E	E-T350G	---
HPV16-5	EU880239	539	E	E-T350G	---
HPV16-6	EU880240	576	E	E-T350G	---
HPV16-7	EU880241	562	E	E-T350G	---
HPV16-8	EU880242	493	E	E-T350G	---
HPV16-9	EU880243	436	E	E-T350G	---
HPV16-10	EU880244	539	E	E-C188G	---
HPV16-11	EU880245	545	E	E-C188G	---
HPV16-12	EU880246	562	E	E-P	---
HPV16-13	EU880247	573	E	E-P	A334G
HPV16-14	EU880248	562	E	E-P	A334G
HPV16-15	EU880249	562	E	E-P	A334G
HPV16-16	EU880250	544	E	E-P	A334G
HPV16-17	EU880251	564	E	E-P	A334G
HPV16-18	EU880252	541	E	E-P	A334G
HPV16-19	EU880253	541	E	E-P	---
HPV16-20	EU880254	548	E	E-P	A334G
HPV16-21	EU880255	465	E	E-P	A334G
HPV16-22	EU880256	545	E	E-P	A334G
HPV16-23	EU880257	522	E	E-P	A334G
HPV16-24	EU880258	522	E	E-P	A334G
HPV16-25	EU880259	564	E	E-P	A334G
HPV16-26	EU880260	560	E	E-P	A334G
HPV16-27	EU880261	337	E	E-P	A334G
HPV16-28	EU880262	576	E	E-P	A334G
HPV16-29	EU880263	546	E	E-P	A334G
HPV16-30	EU880264	564	E	E-P	A404T
HPV16-31	EU880265	564	E	E-P	---
HPV16-32	EU880266	563	E	E-P	---
HPV16-33	EU880267	558	E	E-P	A334G
HPV16-34	EU880268	561	E	E-P	A334G
HPV16-35	EU880269	562	E	E-P	A334G
HPV16-36	EU880270	562	E	E-P	A334G
HPV16-37	EU880271	559	E	E-P	A334G
HPV16-38	EU880272	564	E	E-P	A334G

478 ^a GenBank/EMBL/DDBJ.479 ^b Nucleotide numbering of the HPV16 E-P Ref genome.480 ^c Described for the first time in this work.

481

481

482 **Table 3. Novel polymorphisms identified**

HPV16		Nucleotide change		Amino acid change	
Subtype	Variant	Type ^a	Location ^b	Substitution	Frameshift
AA	AA-a	D	484	---	127
E	350G	S	A182T	I27L	---
	350G	D	393	---	97
	350G	S	C206T	Q35stop	---
	350G	S	T351A	L83E	---
	350G	I	T, 206-207	---	35
	G188C	S	A97G	---	---
	E-P	S	A404T	I101F	---
	E-P	S	C374A	Q91K	---
	E-P	D	533	---	144
	E-P	I	G, 505-506	---	135
	E-P	S	C173A	H24N	---
	E-P	S	A330G	---	---
	E-P	S	T331A	---	---
	E-P	I	TG, 331-332	---	76
E-P	S	A354G	---	---	
E-P	S	C360G	---	---	
E-P	S	A361G	---	---	
E-P	S	A91C	Q-H	---	
E-P	I	C, 121-122	---	6	
E-P	I	C, 122-123	---	6	

483 ^a D, deletion. I, insertion. S, substitution.484 ^b Nucleotide numbering of the HPV16 E-P Ref genome.

485

485

486 **Table 4. Cervical lesions associated to the HPV16 variants identified**

HPV16 variant	Cervical lesions			HSIL frequency
	LSIL	HSIL ^a	Total	
E-P	4	0	4	0.0%
E-P (A334G)	18	4	22	18.2%
AA-a	1	1	2	50.0%
E-T350G	3	4	7	57.1%
E-C188G	0	2	2	100.0%

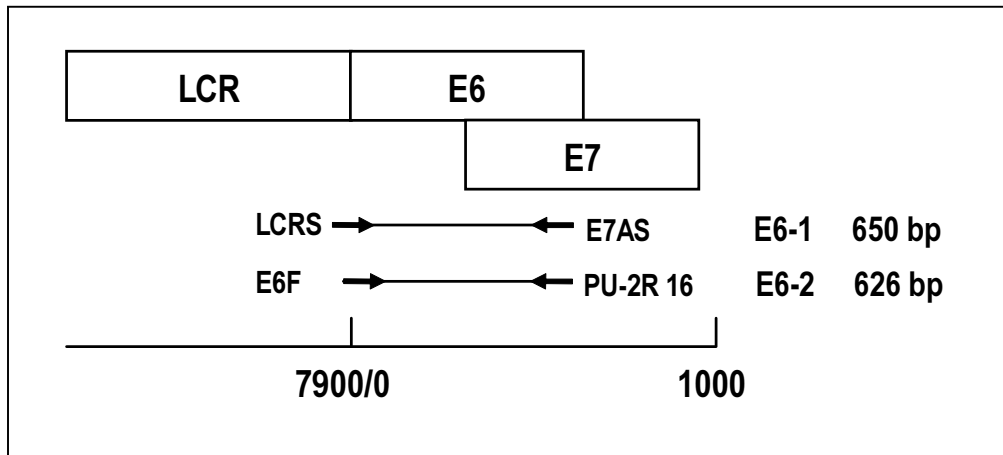
487 ^a Includes HSIL and invasive cervical cancer.

488

488

489 Fig. 1

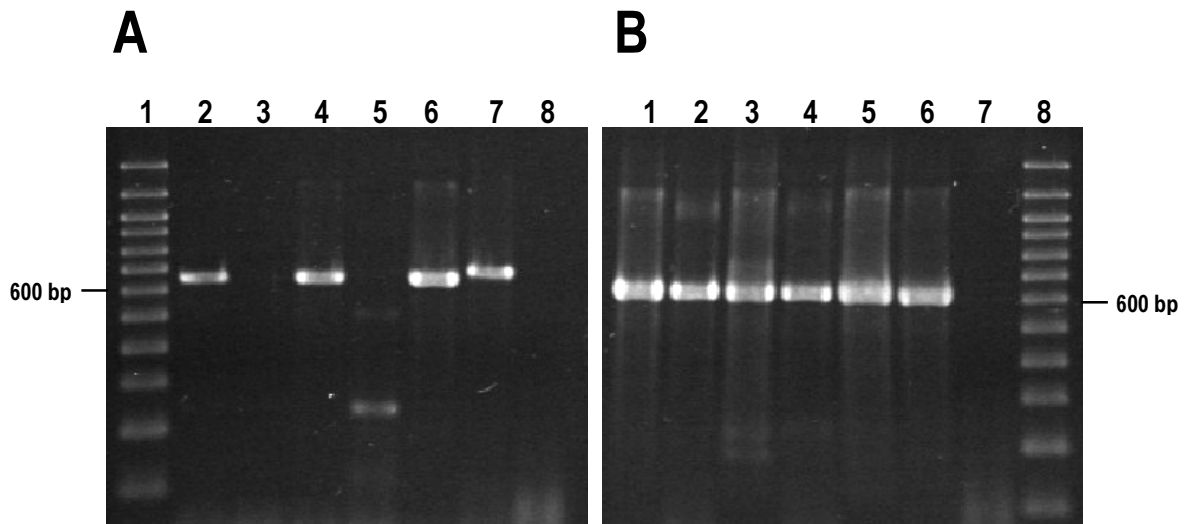
490



490

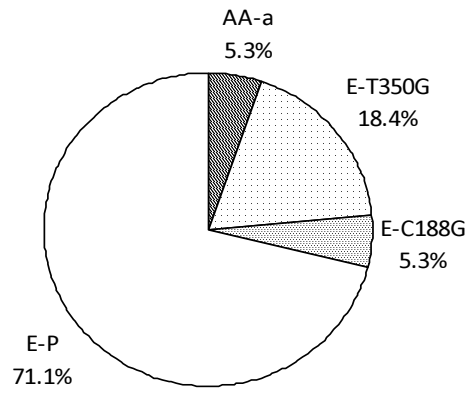
491 Fig. 2

492



492

493 Fig. 3



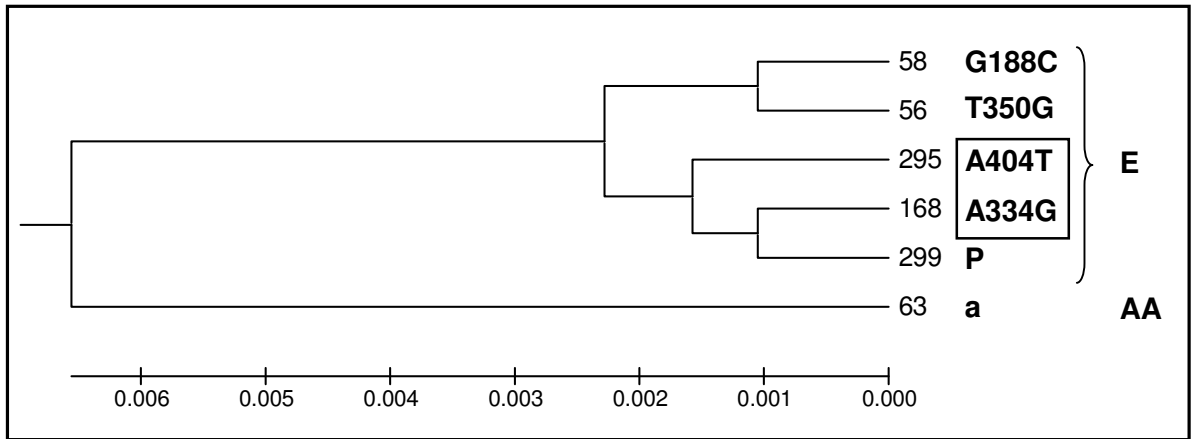
494

495

495

496 Fig. 4

497



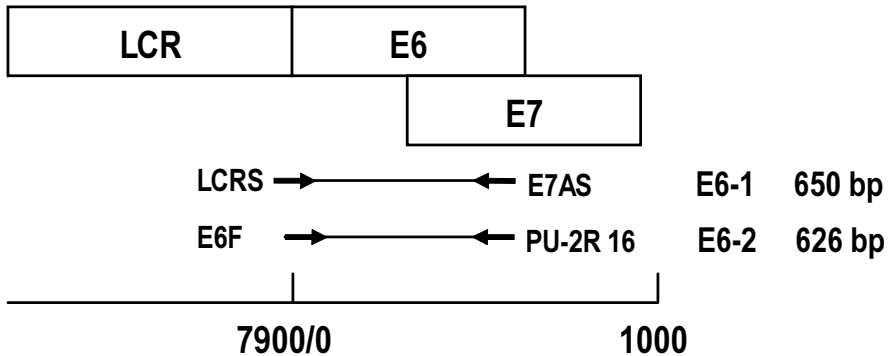


Figure 1

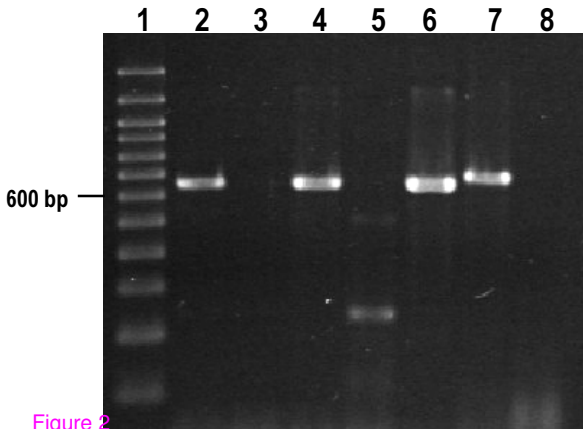
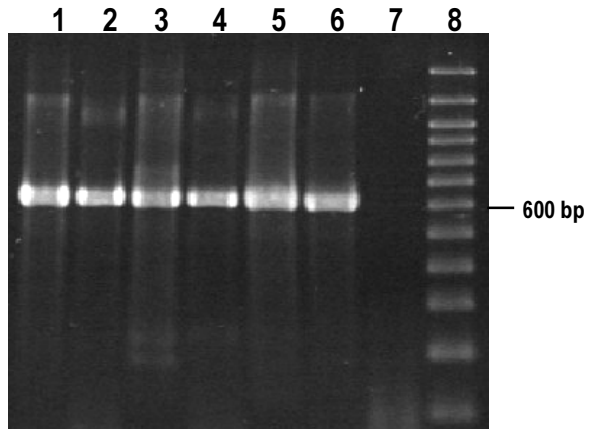
A**B**

Figure 2

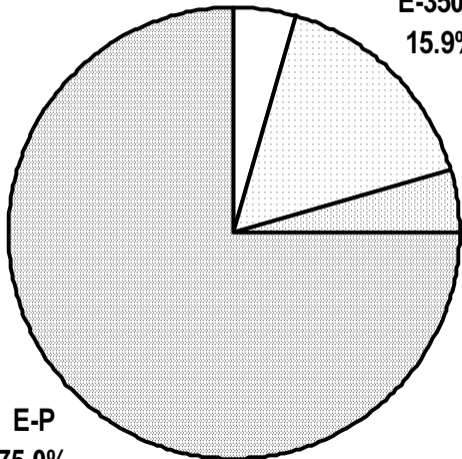
AA-a
4.5%

E-350G
15.9%

E-C188G
4.5%

E-P
75.0%

Figure 3



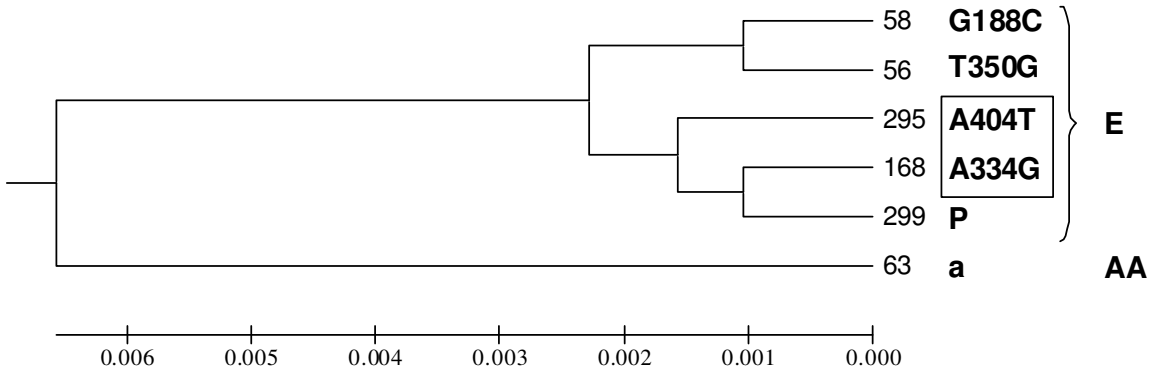


Figure 4